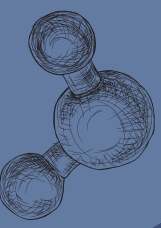
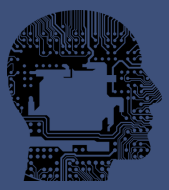
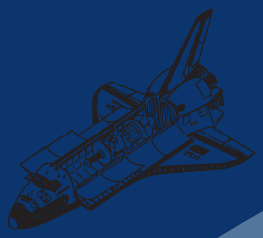
NIKOLA TESLA STEM HIGH SCHOOL RESEARCH JOURNAL

Redmond, Washington







2022-2023



Introduction

Welcome to the eighth edition of our annual student research publication. The Nikola Tesla STEM Research Journal enables our students to share their research with the academic and business communities. The investigations included in this edition demonstrate both the increasing depth and diversity of our studies. In addition, we have included a list of the awards and honors earned by Tesla STEM students this academic year.

Our thanks to the Tesla STEM staff for ensuring all students are afforded every opportunity to showcase their knowledge and talents and to our students, for their dedication to STEM Literacy. Many thanks to the dedicated student researchers and mentors, your long hours and hard work made this publication possible.

Editor: Heidi Cooperrider
Cover Design: Saachi Sharma
Journal Promotion: Kate Allender, Rebecca Townsend
Journal Support: Debi Welcome, Kate Allender, Bethany Kankelborg, Andy Christensen,

Current and past Research Journals are available online at https://tesla.lwsd.org/academics

---

Nikola Tesla STEM High School

Lake Washington School District

4301 228th Ave NE, Redmond, Washington 98053
(425) 936-2770 * tesla.lwsd.org

Principal: Amie Karkainen
Associate Principal: Sumeyye Cardakli

Table of Contents

Biomedical/Cellular Biology

Identification of Genomic Breast Cancer Biomarkers: A Computerized Approach at Predicting Mutant Proteomic Prognosis . 1

Seeking CVD: An Exploration of the Links Between the Gut Microbiome, Sex, and Cardiovascular Disease.. 7

Development of Pollen Extract-Coated Microneedles to Assess their Future Potential in Allergy Treatment. 14

Domain Knowledge vs AI: Detecting Cervical Spine Fractures from CT Scans 20

A 3D-Printed Scoliosis Brace Designed to Increase User Comfort and Mobility. 25

OHCT: One Hair Cell at a Time – Using ATOH1 Molecules to Reverse Hearing Loss. 31

The Effect of Vitamin C Treatment on the Growth of Mouse Squamous Cell Carcinoma Cells. 38

Environmental Science

Using Machine Learning to Predict the Power Conversion Efficiency of Perovskite Solar Cells 57

A Novel Integrated Approach to Illegal Logging Detection. 66

BioPGAN a Novel Way of Generating Novel Bioplastics Using Generative Adversarial Networks 80

Using Crop Simulation Models and Machine Learning to Find Sustainable Alternative to Common Industrial Farming Practice. 87

Neurosciences

Reconnect: A Computational Approach to Using DBS To Combat Neurodegenerative Disease Through Induced Plasticity (CBIO-0349). 103

Identification of Novel Diagnostic Neuroimaging Biomarkers for Autism Spectrum Disorder Through Convolutional Neural Network-Based Analysis of Functional, Structural, and Diffusion Tensor Imaging Data Towards Enhanced Autism Diagnosis. 115

EEG-Based Analysis of Brain Activity in Individuals with Varying Levels of Social Media Usage. 127

Engineering/Mathematics

Parameterizing Binary Neutron Star Postmergers Using Vision Transformers. 137

A Low-Cost Approach to the Filtration of Particulate Matter and Carbon Monoxide Produced from the Partial Combustion of Biomass. 148

Engineering a Microplastic Filter for Indoor Taps 157

An Everting Vine Robot With A Pneumatic Soft Gripper 166

Software Applications/Computational Sciences

A Quantum Optimization Algorithm to Efficiently Route Public Transportation in Cities. 176

DHEBNet – a Novel Deep Heterogenous Ensemble-Based Learning Model for Deepfake Detection. . . .184

Finding The Source: A School Shooting Prediction Program. 196

Stroke Prediction with Logistic Regression Based on Various Risk Factors. 202

Using VR In-vivo Exposure Therapy to Identify Startle Responses in Phobic and Non-Phobic Individuals Through Body Wave Analysis. 208

Testing the Effects of the Smiling Mind Mobile App on Reducing Anxiety Symptoms. 217

Virtual Reality for ADHD: Using Technology to Enhance Learning and Attention 224

Behavioral & Social Sciences

Light Therapy as a Biological Intervention for Irregular Circadian Rhythms 232

Correlation between Empathy and Communication with Highly Emotional People 242

The Effect of Different Modalities and Frequencies of Exercise on Mental Health 252

Self-esteem and Social Anxiety of South Asian Adolescents with Varying Acne Severity 256

Examining the Efficacy of Dark Chocolate on Test-Taking Anxiety in High School Students. 265

The Relationship Between Different Genres of Media Consumed to Burnout Symptoms Reported . . 269

Quantifying Depression: An art-based approach to reducing the self-report bias in diagnosing depression Using Koch's Baum Test and color psychology. 278

Autism and Moral Criteria 286

Adolescents vs. Adults: Upward Social Comparison in a Standardized Test Setting 293

Comparison of Lung Capacity in Competitive Swimmers vs Non-Competitive Swimmers 306

Elaboration on the Effect Maternal Education has on Fertility, Age at First Birth, and Contraceptive Use. . .313

Impact of Stigma on Indian Parental Detection of Generalized Anxiety Disorder in Adolescents 321

Identifying the Levels of Susceptibility of Law Students to Racial Bias 326

Evaluating the Effect of Beats Per Minute of Techno Music on Reaction Times to Auditory Stimuli in High School Students 339

Red Light for Sleep Quality and Performance in Gymnasts. 347

Exploring The Effects of Music Genre on Memorization 353

Evaluation of Horse Comfort in English and Western Riding Style Saddle 359

Awards and Achievements 364

Biomedical/ Cellular Biology

Identification of Genomic Breast Cancer Biomarkers: A Computerized Approach at Predicting Mutant Proteomic Prognosis

Talah Albadawi

Nikola Tesla STEM High School

Abstract:

Breast cancer patients with an identified genomic mutation can benefit from a target therapy that attacks the mutant protein, such therapies have been developed for mutations like BRCA1 and BRCA2. This study aims to compare the accuracy of two kinds of machine learning model in XGBoost that both serve to identify new mutations for therapies to address, one (models A), considered the standard, fed mutation data (BARD1, BRCA1, BRCA2, PALB2, and PIK3CA) to predict classification data (breast cancer type, luminality, and infiltration status), and one, novel and new (models B), fed classification data to predict mutation status. It was hypothesized that models B would outperform models A. Mutations data regarded as NEG (including nonsense mutations), missense, truncated, or spliced was plugged into XGBoost alongside classification under cancer type, infiltration status and luminality. The predictive power of the model when it was using mutation presence to prove classification was compared to its predictive power when it was using classification to prove mutation presence. Results showed that the model fed classifications data to predict mutation presence had a 0.314 higher accuracy than the standard mutation to prove classification model. This confirms the hypothesis that breast cancer classification is a better indicator towards mutation presence than mutation presence is toward classification. However, due to the large source of error that was imbalanced data across classes, which is shown by the average Cohen Kappa Coefficient of models A being 0.047 and models B being 0, we regard the hypothesis as neither true nor false.

Thesis Statement

This study aims to identify the best machine learning approach for identifying breast cancer biomarkers – the standard in which the model is fed mutation data to predict classification, or a novel approach in which the model is fed classification data to predict mutation presence.

Introduction

1.1 Background:

1.1.1 Breast Cancer

Breast Cancer [BC] is the most commonly diagnosed cancer for women in the United States, killing over 40,000 of the 230,000 diagnosed each year (Ziegler et al).

To prevent these deaths and treat BC patients, a variety of therapies are used. These therapies fall under the categories of patient-specific target therapies and general, cheaper, broad spectrum therapies.

1.1.2 Current Solutions

1.1.2.1 Broad Spectrum Therapies

Widely available treatments to those with BC in the US include broad spectrum therapies. This includes chemotherapy [CT], radiotherapy [RT], hormone therapy [HT]/endocrine therapy [ET], immunotherapy [IT] and surgeries. CT uses drugs to stop the growth of cancer kills by terminating their ability to divide or killing them directly. RT uses intense energy beams, usually from X-rays, but protons and other energies can be used, to kill cancer cells, most often externally. HT/ETs block hormones, such as estrogen, from binding to receptors. These therapies are only useful in cases where patients have hormone receptors on their BC cells (HR+). IT uses drugs to stimulate the immune system by guiding it to recognize and kill BC cells.

1.1.2.2 Target Therapies

Most efficient at killing breast cancer – measured by

means of being the safest (low chance of harming other non-cancerous cells), having the lowest re-occurence rate, and causing the least physical pain – are target therapies [TT]. TT treat BC by targeting proteins that help cancer grow. There are a few different kinds of TT, inclduing, but not limited to, monoclonal antibodies, antibody-drug conjugates [ADC], and kinase inhibitors.

Monclonal antibodies, sometimes referred to as a type of IT, are immune system imitation proteins (antibodies) made in a lab with the goal of attacking a specific protein. One monoclonal antibody, trastuzumab, is used to target the BRCA2 mutant. A healthy BRCA2 protein is a tumor surpressor, but the mutation can be a BC causer.

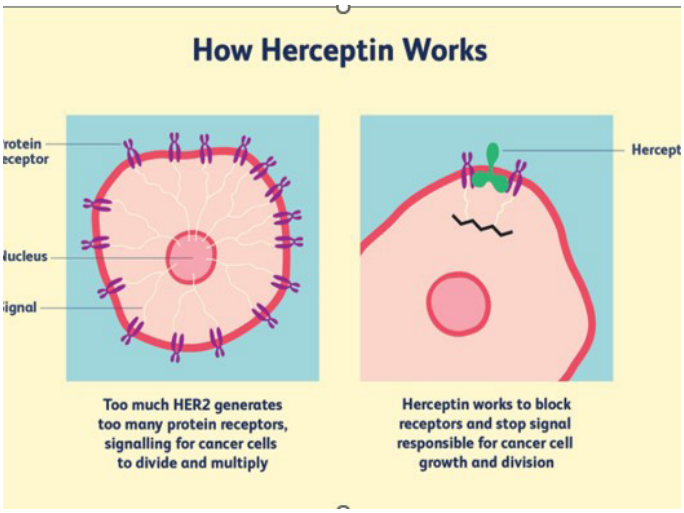


Figure I. How Monoclal Antibody Herception Works: This diagram displays how monoclal antibody herceptin binds to overexpressed HER2 mutant proteins, block cell receptors, stop cell signaling for growth, and help contain BC.

Image taken from Targeted Therapies for Breast Cancer (verywellhealth.com).

ADCs are a monoclonal antibody linked to a CT drug. The antibody binds to the protein and sends a signal the CT drug can pick up on and use to identify the protein it needs to attack. One ADC is Ado-trastuzumab emtansine, where the antibody binds to the mutant HER2 and signals to the CT emtansine.

Healthy kinase proteins are ones that send signals to other cells, such as the signal to grow. Kinase inhibitors are drugs that block this signal. One such drug is lapatinib, which stops mutant HER2 growth, which increases the length of BC progression free patient survival.

As wonderful as TT are, they are less often used, because only patients with an identified mutation relevant to their BC would benefit from them. And, unfortunately, most BC patients do not have one of these known relevant mutations. So, increase usability of TT, since it would help the survivability of the BC population, more BC relevant mutations must be identified.

1.1.2.3 Cumilative Treatment Plans

Most effective by far are cumilative treatment plans, which take a patient-specific treatment approach where a personalized mix of target and broad spectrum therapies are used to treat the BC. However, these tailored plans are expensive, which presents an equity issue, as well as difficult to develop with as little therapies as we have. And so, more therapies must be developed. And, for the new TT, specific mutations must be identified.

1.1.3 Appplications of Machine Learning

To identify new BC relevant mutations, a correlation must be found between protein presence and and the occurrence of a classification of BC. The best way to do this is by using machine learning (ML). ML refers to computer systems that can "learn" by adapting without following specific instructions.

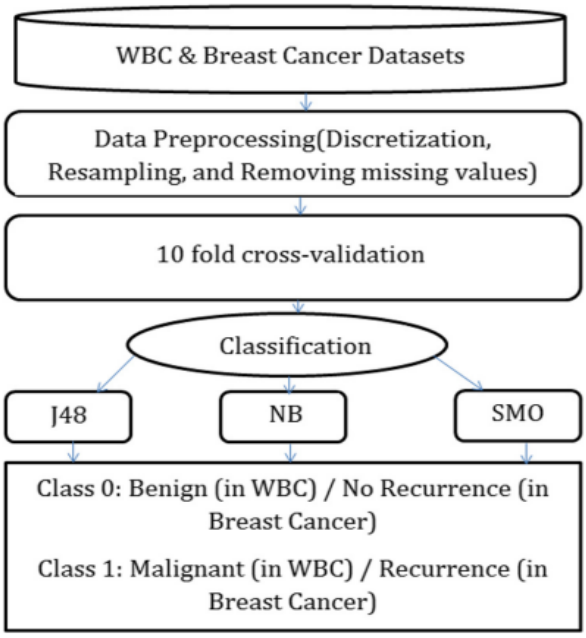


Figure II. Machine Learning Application for Biomarker Identification: This figure shows one method of using ML to identify BC biomarkers in which wild breast cancer [WBC] and BC datasets are preprocessed and run through 10 fold cross validation during training and classified and tested as class 0 [benign] or class 1 [malignant]. Figure taken from: https://doi.org/10.1007/978-981-15-7205-0_10.

To identify mutant biomarkers, ML models are fed large datasets of gene health. Genes determine proten creation, a healthy gene will lead to a healthy protein, and a mutated gene to a mutated protein. Gene data is cheaper to collect than protein data, which is why it is used. The gene data is paired with BC status/presence metadata. The classification and gene data is used by the model for it to learn to predict classification from protein information.

The most common type of models used for biomarker identification are random forest [RF] models – or those that use decision trees.

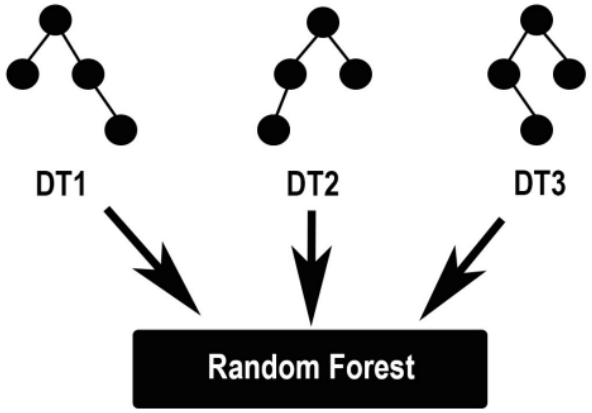


Figure III. How Decision Tree Models Work: This figure displays how models such as RF use multiple separate decision trees.

Figure taken from: https://doi.org/10.1109/ICEDSA.2016.7818560

Models can be retrained to get a higher accuracy, and then key features are identified that helped the model in making its decisions. These key features are significant input points, which, in the case of biomarker identificaion, are proteins, most often mutated ones.

1.2 Research Approach:

1.2.1 Current Problem Statement

However, this ML approach at identifying mutations is overdone. Nothing new is getting identified, and medical demands are not getting kept up with.

1.2.2 Suggested Approach

And so, a new, novel ML approach is suggested in which instead of treating protein data as input and classification data as output, classification is the inputed data and mutated protein data is the output. Key features of the model will not need to be identified, the most accurately predicted mutations will just need to be identified.

1.2.3 Novelty of Study

This study proposes using BC classification to predict mutation presence, as compared to the standard, using mutation presence to predict classification. The reasoning for this stems from BC's ability to evolve and stimulate the growth of mutant proteins that will help it.

1.2.4 Purpose of Study

This study aims not to create the best model, but to compare the accuracies of the two biomarker identification methods – the standard and the novel.

1.2.5 Research Question

Will a ML model fed BC classification data to predict mutation data outperform a model fed mutation data to predict classification data's accuracy, and therefore, ability to identify BC mutant biomarkers?

1.2.6 Hypothesis

Because of BC's ability to stimulate the growth of mutations that will help its growth, a ML model inputed classfication data to predict outputed mutation presence will yield a higher accuracy than a model fed mutation presence data to output classification data, meaning that it will serve better to assist in BC biomarker identification.

1.2.7 Variables

1. Standard accuracy of A models – standard: Fed mutation data to predict classification.

2. Standard accuracy of B models – novel: Fed classification data to predict mutation data.

Methodology

2.1 Data Pre-Processing:

2.1.1 Data Collection

25 open source datasets were accessed from cBioPortal available online. These datasets included gene presence and BC classification data on 11,633 tumor samples from 10,001 people. One tumor was selected at random [using a random number generator] from the people who donated multiple tumor samples so that just 10,001 data points were considered, since these mutiple tumors from the same people had the same gene expression and BC classfication.

2.1.2 Data Selection

Because of the type of super computer needed to filter through such a large database of 10,001 patients, 522 classifications, and 18,638 genes being unavailable, 5 genes and 3 classifications were selected to be focused on in this study. In a usual study attempting to identify novel biomarkers, this would be harmful, since many proteins and classifications would be left unconsidered. But, since this study focuses on comparing the accuracy of the 2 kinds of model, this should not be significant.

2.1.3 Mutation Data Selection and Pre-Processing

The 5 mutations selected as a focus of this study were ones already known to have a correlation with breast cancer, since this would lead to clearer results, as higher model accuracy would mean a higher predictive power. These mutations were (1) BARD1, (2) BRCA1, (3) BRCA2, (4) PALB2, and (5) PIK3CA.

2.1.4 Classification Data Selection and Pre-Processing

The 3 classifications considered in this study were (1) cancer type, (2) infiltration status, and (3) luminality.

Infiltration status refers to the invasivity of the BC.

Luminality refers to the hormone receptor status of the tumor.

2.2 Development of Models:

2.2.1 Development of Models A:

3 [standard] models [inputed mutation data and outputs classification] were developed in Extreme Gradient Boosting [XGBoost].

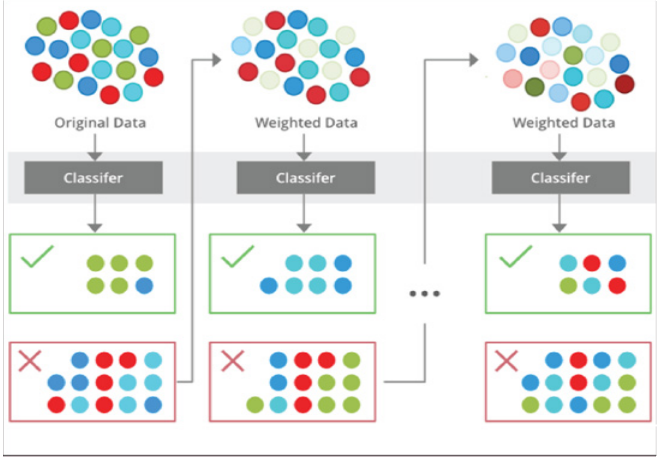


Figure IV. Extreme Gradient Boosting: This figure displays how XGBoost models work – when the model makes an incorrect prediction, a random classifier is boosted. If this helps model accuracy, the classifier is kept boosted. If it is harmful or has no effect, it is unboosted. All the classifiers are ran through until the most relevant ones for boosting to increase model accuracy have all been identified and boosted.

Figure taken from: XGBoost - GeeksforGeeks

2.2.2 Development of Models B:

5 [novel] models [inputed classification data and outputs mutation information] were developed in Extreme Gradient Boosting [XGBoost]. Mutations were classified as NEG [negative – meaning mutation not present – including nonsense mutations, since nonsense mutations, by definition, are ones that have no effect], missense, truncated, or spliced.

Results and Conclusions:

3.1 Raw Results:

Model	Accuracy
Mutation → cancer type	61.45%
Mutation → infiltration status	30.61%
Mutation → luminality	88.46%

Table 1. Raw Model A Results: This table shows the accuracy of each model of type A.

Model	Accuracy
Classification → BARD1	99.54%
Classification → BRCA1	98.06%
Classification → BRCA2	97.38%
Classification → PALB2	98.99%
Classification → PIK3CA	64.92%

Table 2. Raw Model B Results: This table shows the accuacy of each model of type B.

Models	Average accuracy
A	0.602
B	0.916

Table 3. Models A and B Averages: This table shows the average accuracies of models A and B side by side for comparison.

3.2 Conclusions:

3.2.1 Analysis of Results

Models B having a 0.314 higher accuracy than models A, proving the hypothesis that because of BC's ability to stimulate the growth of mutations that will help its growth, a ML model inputed classfication data to predict outputed mutation presence will yield a higher accuracy than a model fed mutation presence data to output classification data, meaning that it will serve better to assist in BC biomarker identification to be correct.

This is significant.

3.2.2 Impactful Limitations

Although this 0.314 accuracy increase is significant, a significant limitation of the study can be seen in the Cohen Kappa Coefficient of models A being 0.047 and models B being 0. Cohen Kappa Coefficient measures inter-rater reliability, and the coefficients being 0 and so close to 0 mean that the model was practically guessing. This is due to unbalanced data being fed to the model [for reference – this unbalanced data comes from the fact that mutations are rare; the most common BC mutation, BRCA2, is only present in less than 4% of the BC population].

Because of this, the research question of will a ML model fed BC classification data to predict mutation data outperform a model fed mutation data to predict classification data's accuracy, and therefore, ability to identify BC mutant biomarkers can be considered unanswered – neither true nor false.

3.3 Future Research:

3.3.1 Improving Upon Current Study

Because of the significance this research could hold in proposing a better approach to identifying biomarkers, this should be retested with significant data manipulation and balancing.

3.3.2 Potential Applications

3.3.2.1 Primary Application

The discovery of the most effective approach for biomarker identification would have a practical approach in determining the most effective research tactic scientists should be conducting.

3.3.2.2 Secondary Application

Once new biomarkers have been identified, this research will help in assisted diagnoses the development of new personalized treatment plans.

References and Acknowledgments

admin. (2020, June 25). *Machine Learning is a tool to achieve Artificial Intelligence.* Educlasses.co.in. http://educlasses.co.in/machine-learning-is-a-tool-to-achieve-artifical-intelligence.html

API and API Clients. (n.d.). Docs.cbioportal.org.

Retrieved January 17, 2023, from https://docs.cbioportal.org/web-api-and-clients/

Beretov, J., Wasinger, V. C., Millar, E. K. A., Schwartz, P., Graham, P. H., & Li, Y. (2015). Proteomic Analysis of Urine to Identify Breast Cancer Biomarker Candidates Using a Label-Free LC-MS/MS Approach. *PLOS ONE*, *10*(11), e0141876. https://doi.org/10.1371/journal.pone.0141876

Bozic, I., Reiter, J. G., Allen, B., Antal, T., Chatterjee, K., Shah, P., Moon, Y. S., Yaqubie, A., Kelly, N., Le, D. T., Lipson, E. J., Chapman, P. B., Diaz, L. A., Jr, Vogelstein, B., & Nowak, M. A. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *ELife*, *2*, e00747. https://doi.org/10.7554/eLife.00747

Breast Cancer Risk Factors You Can't Change. (n.d.). https://www.cancer.org/cancer/breast-cancer/risk-and-prevention/breast-cancer-risk-factors-you-cannot-change.html

Liu, M., Xie, F., Liu, M., Zhang, Y., & Wang, S. (2021). Association between BRCA mutational status and survival in patients with breast cancer: a systematic review and meta-analysis. *Breast Cancer Research and Treatment, 186(3), 591–605.* https://doi.org/10.1007/s10549-021-06104-y

Rose, S. (2020, July 27). *A BRCA Gene Test Saved My Life, and My Sister's. Healthline.* https://www.healthline.com/health/breast-cancer/brca-gene-test-saved-my-life

Yadav, S., Ladkany, R., Yadav, D., Alhalabi, O., Khaddam, S., Isaac, D., Cardenas, P. Y., & Zakalik, D. (2018). Impact of BRCA Mutation Status on Survival of Women With Triple-negative Breast Cancer. *Clinical Breast Cancer, 18(5), e1229–e1235.* https://doi.org/10.1016/j.clbc.2017.12.01

Ziegler, Y. S., Moresco, J. J., Tu, P. G., Yates, J. R., & Nardulli, A. M. (2014). Plasma Membrane Proteomics of Human Breast Cancer Cell Lines Identifies Potential Targets for Breast Cancer Diagnosis and Treatment. *PLoS ONE*, *9*(7), e102341. https://doi.org/10.1371/journal.pone.0102341

A special thank you to Mrs. Kate Allender, the teacher who supported me through the logistics of this research process, and Dr. Andie Doak of Fred Hutch, my mentor who guided my understanding of what the industry needs.

Seeking CVD: An Exploration of the Links Between the Gut Microbiome, Sex, and Cardiovascular Disease

Anjana Punniamoorthy

Nikola Tesla STEM High School

Abstract:

As the leading cause of death globally, the need for early identification of cardiovascular disease remains a prevalent issue. The goal of this study was to use machine learning as a way to find a relationship between the gut microbiome, sex, and cardiovascular disease using. To accomplish this, random forest, k-nearest neighbor, support vector machine, and adaptive boosting algorithms were built using python, with a final goal of 70% accuracy. Additionally, the algorithms aimed to correctly identify sex-specific, highly contributing taxa.

Two models were built for each algorithm – one to run on female patient data and one for male. Each model took input from a CSV file containing columns on a patient's disease status (present or absent), and hundreds of nucleotide sequences. These sequences served as identifiers for the bacteria, and had data on frequencies. The most efficient model was the Random Forest model. The model run on female data had scores of about 89.9% for accuracy, 88.3% for area under the curve, while the male model had an 87% score for accuracy and 85% for area under the curve.

There was a clear relationship between the OTUs and propensity for cardiovascular disease, as the model was able to use OTU data to accurately predict disease. Additionally, highly contributing taxa were identified from the model. Further research would aim to use a larger dataset, alongside the identification of taxa down to the species level for all sex-specific, highly contributing taxa.

Introduction

Earlier, personalized detection of cardiovascular disease is integral to successfully lowering its extreme morality rate. In the United States, one person dies every 34 seconds from cardiovascular disease (CVD). This statistic extends beyond the United States - cardiovascular diseases are the leading cause of death globally, with 17.9 million people dying from CVDs annually. In large part, this is due to late detection of the disease, a common issue in many countries. Late detection occurs often, and allows the disease to progress to irredeemable stages. This project seeks to provide an alternative method to disease detection through analysis of the gut microbiome and patient sex.

1. Development and Management of Cardiovascular Disease

Cardiovascular disease is extremely prevalent because development of the disease is relatively easy. Major risk factors emerge from common behaviors such as a lack of exercise, unhealthy diet, tobacco use, or excessive alcohol consumption. Alongside making lifestyle changes to avoid the risks brought on by this behavior, individuals with cardiovascular disease have to manage common comorbidities such as obesity, diabetes, or high blood pressure.

2. Ambiguity of Symptoms

The need for earlier detection is further highlighted by the nature of cardiovascular disease symptoms. These symptoms may present themselves as part of the natural aging process, with symptoms such as indigestion or extreme fatigue. Alternatively, it's common that symptoms won't appear until late in development. This allows the disease to go undetected for decades.

In people with heart valve disease - a type of cardiovascular disease - symptoms may not be shown even once the condition reaches a severe stage. Similarly, coronary artery disease - the most

common heart disease in the United States - can develop decades before someone forms symptoms. Examination of young adults in their late teens and early twenties has shown that the arteries start thickening very early, while symptoms may only be revealed once a patient reaches their fifties, sixties, or even seventies.

3. Addressing Medical Bias

Alongside the need for early detection is a necessity to understand the impacts of sex within this area. Historically, women have been underrepresented in the medical field. This leads to a considerable deficit in the resources available to women. Without conducting research on women, less is known, leading to inadequate healthcare options.

The gaps in cardiovascular research are clear - while 49% of people with cardiovascular disease are women, only 41.9% of participants in this area of research are female, an alarming discrepancy. (Sosinsky et al., 2022). Limited representation in this clinical trials has led to unpredicted differences in how women absorb, metabolize, and excrete treatments. Additionally, women feel adverse side effects to drugs and experience these effects at a higher rate than male counterparts. For example, certain medications used to treat cardiovascular disease have been noted to increase the risk of developing a potentially fatal arrhythmia (torsades de pointes) from use in women. (Pinn, 2003). Although several have since been pulled from the market, the initial release of these drugs into the market clearly depicts the ignorance of sex in research and testing.

3.1 Differences Between Sexes in CVD Development

Current research has found that sex hormones protect women from cardiovascular disease, an advantage provided for women until they reach menopause. (Harvey et al., 2016). Women often receive the worst forecasting for disease intensity of heart failure, and the leading cause of hospital admissions and deaths for women worldwide is cardiovascular disease. Treatments that have been useful in aiding men have not led to a significant decrease in fatality rates for women. The article, *"Sex/Gender-Specific Imbalance in CVD: Could Physical Activity Help to Improve Clinical Outcome Targeting CVD Molecular Mechanisms in Women?"*, suggests that the current model surrounding cardiovascular disease is a multifactorial system, with sex hormones contributing significantly through their relationships with risk factors. Additionally, sex-specific hormones have been identified to result in sex-related differences in the molecular mechanisms that take part in CVD development. (Vaccarezza et al., 2020). Recent studies have identified differences in the levels of potential soluble CVD markers. The A-FABP (a fatty acid binding protein), was identified as an independent determinant of carotid atherosclerosis only for Chinese women, and not male counterparts. (Yeung et al., 2007). This discovery points towards the potential for sex-specific identification markers.

4. The Gut Microbiome and Cardiovascular Disease

The gut microbiome is a relatively unexplored area of the human body that plays a very important role in health. As it is intimately connected to various aspects of the body (such as the brain or immune system), exploring and understanding the gut microbiome could reveal correlations between the propensity for a disease, based on gut composition. For decades, a connection between gut health and heart health has been suspected. Recent research has identified associations between gut dysbiosis – the loss of organisms in the gut microbiome – and cardiovascular disease. Major risk factors are linked to this loss in the gut microbiome, "Most CVD risk factors, including aging, obesity, dietary patterns, and a sedentary lifestyle, have been shown to induce gut dysbiosis" (Novakik et. al, 2020). Additionally, dysbiosis has been linked to inflammation of the intestines, alongside reduced gut barrier integrity. This increases certain levels of bacterial structures and microbial metabolites that may aid in the development of cardiovascular disease. Furthermore, the gut microbiome is able to impact the immune system, as leaks in the intestinal wall allow components of the microbiome to slip through.

Bacterial taxa observed in atherosclerotic plaques (plaque that builds up in arteries) have also been found in the gut of patients with plaques, and the gut microbiome of patients with cardiovascular disease have been proposed to foster inflammation by producing pro-inflammatory molecules. An example is the TMA metabolite (a waste product produced by microbial enzymes from the gut) which has been linked to promotion of atherosclerosis and cardiometabolic disease. Overall, it is clear the gut microbiome plays a significant role in the development of cardiovascular disease through various interactions, such as the manipulation of the immune system or production of certain metabolites.

Summary

Understanding the correlation between sex, the microbiome, and cardiovascular disease could play a vital role in predicting cardiovascular disease in a patient. Using the gut microbiome to address the gaps in early detection of cardiovascular disease could prevent a great number of deaths. Additionally, the gut microbiome has potential as a way to produce sex-specific identifiers for disease identification. We hypothesize that machine learning algorithms utilizing individual's data on sex and the flora of the gut microbiome will work as a reliable predictor of an individual's potential for cardiovascular disease, reaching at least 70% accuracy. Additionally, at least 6 sex-specific highly contributing taxa will be identified through this project.

5. Project Goals

This project seeks to create a machine learning model for each sex that utilizes data on patient's gut microbiome taxa to formulate a prediction of an individual's potential for cardiovascular disease. Additionally, the machine algorithm (Random Forest, Extreme Gradient Boosting, K-Nearest Neighbor, or Support Vector Machine) best suited to this problem will be identified.. This is based on values retrieved from data analysis, with the most importance given to accuracy scores. Last, this project aims to identify sex-specific taxa that have a significant contribution to an individual's potential for cardiovascular disease.

It is expected that a machine learning model with a score of at least 70% accuracy will be developed. This model will likely be the Random Forest model, and sex-specific, highly contributing taxa will be identified.

Methodology

Design Protocol

Data Download and Data Frame Creation

Data is available through NCBI SRA Run selector using code PRJEB11419. The metadata from this website was downloaded as a text file. This file will then be filtered so that only sex, gut microbiome taxa, and disease status were included in the table. A total number of samples being used will be determined from this file, inputted into Data Table 1. Based on an 80/20 split, the number of training and testing samples will subsequently be determined and inputted into Data Table 1. An excel sheet containing samples of only the male sex will be created. The same is done for female samples. Files are created based on these sex files, containing the OTUs (operational taxonomic units) for each bacteria.

Using the pandas library, a data frame is created from the male file containing OTUs. This will be referred to as the 'male file' in later steps. The same will be done for the female file containing OTUs. This will be referred to as the 'male file' in later steps. Both files are to be manipulated using the pandas library, first by transposing the data so that the columns and rows are switched (transpose data). The sex column will then be removed, and the disease status column is moved to be the last column in the data frame. Finally each respective data frame is exported as a CSV file. Figure 1 shows what the data frames should look like.

AGGGUACA...	ACCGUACA...	GAUCGUACA...	...	CVD status
16.0	0	0	...	0
0	0	0	...	1
0	0	2.0	...	1
1.0	27.0	0	...	1

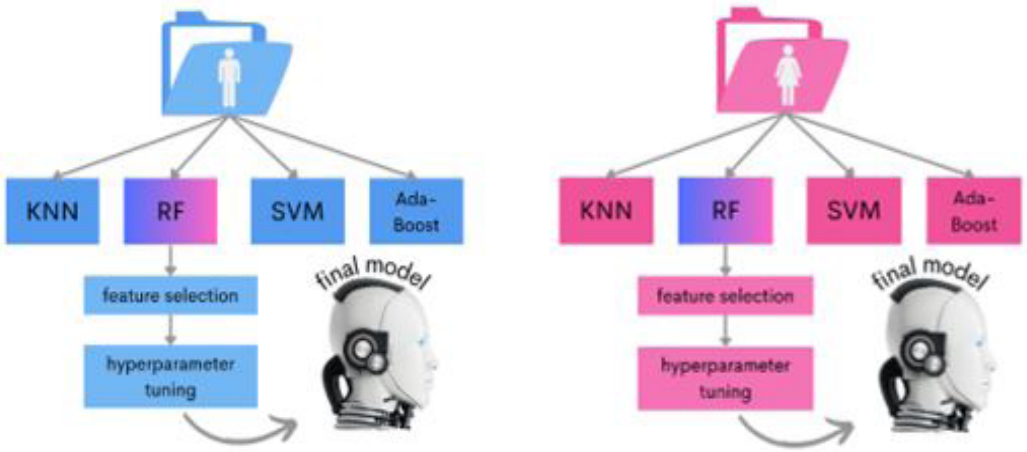
Figure 1 | ***Sample Table***

Model Creation

Identification of best model for this problem occurs at the first tier of this diagram, evaluated on accuracy scores from initial data.

Further tuning is then employed.

Figure 2 | ***General process for creation of final male and female models***



First import scikit learn and use the train/test split function to assign 80% of the data to the training dataset and 20% to the testing dataset. Using scikit learn as a base, create Random Forest, Support Vector Machine (with linear kernel), K-Nearest Neighbor, and an Adaptive Boosting model in separate Jupyter notebooks. A total of 8 notebooks will be created because model creation has to occur for females and males.

The following steps were repeated for a female set and male set of model(s).

Establish a baseline accuracy for the dataset using a dummy classifier – input this score into Data Table 2. Fit each model to the current data frame and evaluate base accuracy scores for each model run on the unfiltered male data frame, and female data frame respectively. Input these results into Data Table 2.

Perform dimensionality reduction on each file (male and female) using the model that had the highest accuracy from the base accuracy scores yielded from each model in the previous step. (This 'best' model should be the Random Forest (RF) model). Complete dimensionality reduction of the data using the feature_importances_ function from the RF model. Make sure this is done once for the 'male' file, and once for the 'female' file. Create a new data frame with features (bacterial sequences) sorted by the importance score yielded by this function. Using a for loop, run 100 trials with different numbers of 'top' features on the RF model to see what number of features yields the highest accuracy. Top features are the left-most features (ones with the highest importance score) in the new data frame.

After performing dimensionality reduction, create a new 'final' data frame that contains only the optimal number of 'top' features. Run hyperparameter tuning on the model using this final data frame. Hyperparameter tuning was accomplished using randomized search. From sklearn.model_selection import RandomizedSearchCV, from sklearn.ensemble import RandomForestClassifier, import random, and from scipy.stats import randint. Set up a parameter grid including the number of estimators, as a random integer from 10-300, minimum samples split as a random integer from 2-100, minimum samples per leaf as a random integer from 1-50, max_depth as a number from 1-10, and bootstrap as either true or false. Then create a random search object that uses the 'RandomizedSearchCV' function from scikit learn, fitted to 'final' data frame and set to run 200 times with the maximum number of jobs allowed by the PC currently being utilized. Print the best parameters and the score they yielded using functions from the random search object.

Experimental Testing of Prototype

Fit each male and female model to the 'final' data frames created earlier. Ensure parameters are set to the 'best' parameters, yielded from hyperparameter tuning trials. Then return a classification report, accuracy score, and AUC score for each model. Repeat this process twice

more for a total of three trials. In running new trials, create a new model at the start of each trial to ensure the model is not overfitted and does not retain knowledge from previous trials.

Record average accuracy, AUC scores, f1 scores, precision scores, and recall scores for each model into data table 3. Run the top three features from each sex in NCBI BLAST to identify the sex-specific, highly contributing taxa. Record these results in Data Table 4.

Calculations of Metrics

Average accuracy was evaluated by taking the total accuracy (a sum of the scores from each trial) divided by 3. Recall is the proportion of true positives out of actual positives from each trial. The average recall was evaluated by taking the total recall score (a sum of the scores from each trial) divided by 3. Precision is the total true positives divided by sum of the false positives and true positives. The average precision scores was evaluated by taking the total precision score (a sum of the scores from each trial) divided by 3. The average AUC-ROC score (Area under ROC curve) was evaluated by taking the total AUC score (a sum of the scores from each trial) divided by 3. The average F1 score was evaluated by taking the total F1 score (a sum of the scores from each trial) divided by 3.

Data Analysis:

Data Tables

Data Import

Random Sort | Data Table 1

	Number of testing samples	Number of training samples	Total number of samples
Male	71	285	356
Female	71	285	356

Accuracy Testing

Accuracy Analysis of Individual Algorithms | Data Table 2

	% Accuracy Female	% Accuracy Male
Ada-boost	59	60
RF	64	64
KNN	44	47
SVM	61	60
Baseline (Dummy Classifier)	48	52

* Numbers rounded to the whole

The Random Forest model yielded the highest accuracy for both female and male data when compared to other models. All models besides KNN provided a better prediction than the dummy classifier, indicating these models were employing complex decision making to come to a decision.

Accuracy of Final Models | Data Table 3

	Accuracy	AUC	F1	Recall	Precision
Male Average	0.870	0.850	0.900	0.823	0.813
Female Average	0.899	0.883	0.923	0.933	0.914

* Numbers rounded to the thousandth, average taken from three trials

The high recall score is especially important because it demonstrates the model was able to identify positive cases – patients that had the disease. Within this problem, a high recall score is especially important because it means patients with the disease are not being falsely diagnosed as negative.

Sex-Specific Highly Contributing Taxa | Data Table 4

	Bacteria 1	Bacteria 2	Bacteria 3
Male	Firmicutes	Firmicutes	Anaerostipes hadrus
Taxonomic Rank	Phylum	Phylum	Family
Female	Firmicutes	Firmicutes	Parasutterella excrementihominis
Taxonomic Rank	Phylum	Phylum	Species

Figure 1. The 4 different methods of drug delivery using microneedles. The diagram shows how dif

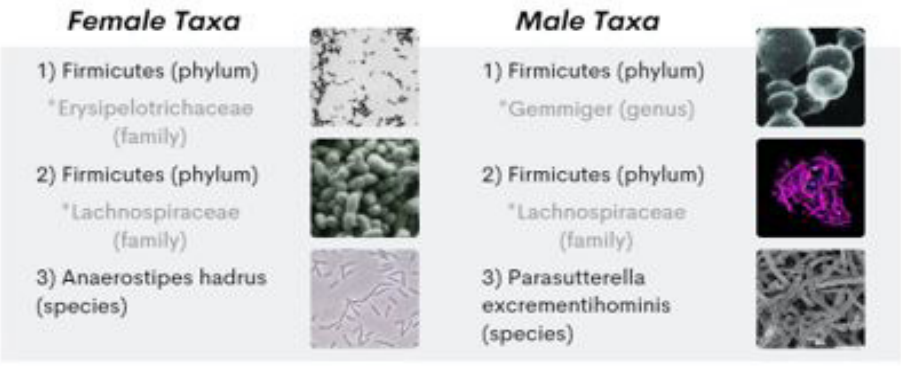
Results

The hypothesis that an accurate machine learning algorithm utilizing individual's data on sex and their gut microbiome will be developed was correct. Additionally, the hypothesis that sex-specific identifiers for cardiovascular disease will be achieved through gut microbiome analysis was correct.

Each engineering goal was met. Random Forest was identified as the best suited for this problem, as it had the highest raw accuracy out of all the models tested. After further tuning of the dataset and Random Forest models, the final model run on female data had average scores of about 89.9% for accuracy, 88.3% for area under the curve over three trials, while the final male model had an 87.0% score for accuracy and 85.0% for area under the curve over three trials.

Sex-specific, highly contributing taxa were identified for each gender, but at varying taxonomic ranks. Some bacteria were identified to the species level, while others stayed at the phylum level. However, all top bacteria identified had different 16S rRNA sequences, indicating that certain bacteria are more contributive towards disease development for one sex while a different set is more contributive for the other. Additionally, for bacteria who could only be definitely identified to the phylum level, there were still hits at subsequent taxonomic levels, but returning more than one result for the same rank. The majority result was used as a result, and this is indicated in Figure 3.

Figure 3 | *Diagram of Identified taxa*



Grey text indicates an identification that had multiple hits, but represents the majority result returned from BLAST run.

Given that the industry standard for machine learning models is an accuracy score of 70%, results are significant and prove that gut microbiome analysis has potential as a simple diagnosis method for cardiovascular disease. Future research would aim to test on a larger dataset, and identify bacteria for each sex down to the species level.

Conclusion

The gut microbiome has important potential as an identifier for cardiovascular disease. Through analysis utilizing the Random Forest algorithm, the microbiome was a data source that could be used to provide accurate predictions of disease. The final models created for each sex provided high recall, precision, f1, and AUC scores alongside high accuracy.

Results from this project also provide evidence that certain bacteria are more predictive of cardiovascular disease in women than in men, and vice versa. The top taxa identified for each sex were unique, indicating they were specific to each sex. This provides rationale for further analysis of sex differences within the gut microbiome, and use of the gut microbiome to identify sex-specific identifiers of cardiovascular disease.

Bibliography

Alqahtani, A., Alsubai, S., Sha, M., Vilcekova, L., & Javed, T. (2022, August 16). Cardiovascular disease detection using ensemble learning. Computational Intelligence and Neuroscience. Retrieved November 3, 2022, from https://www.hindawi.com/journals/cin/2022/5267498/

Aryal, S., Alimadadi, A., Manandhar, I., Joe, B., & Cheng, X. (2020). Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. Hypertension, 76(5), 1555–1562. https://doi.org/10.1161/hypertensionaha.120.15885

Harvey, R. E., Coffman, K. E., & Miller, V. M. (2015). Women-specific factors to consider in risk, diagnosis and treatment of cardiovascular disease. Women's health (London, England), 11(2), 239–257. https://doi.org/10.2217/whe.14.64

Mosca, L., Barrett-Connor, E., & Wenger, N. K. (2011, November 8). Sex/gender differences in cardiovascular disease prevention: What a difference a decade makes. Circulation. Retrieved November 3, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3362050/

Muhammad, Y., Tahir, M., Hayat, M., & Chong, K. T. (2020). Early and accurate detection and diagnosis of heart disease using intelligent computational model. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-76635-9

Novakovic, M., Rout, A., Kingsley, T., Kirchoff, R., Singh, A., Verma, V., Kant, R., & Chaudhary, R. (2020). Role of gut microbiota in cardiovascular diseases. World journal of cardiology, 12(4), 110–122. https://doi.org/10.4330/wjc.v12.i4.110

Peters, S. A. E., Muntner, P., & Woodward, M. (2019). Sex differences in the prevalence of, and trends in, cardiovascular risk factors, treatment, and control in the United States, 2001 to 2016. Circulation, 139(8), 1025–1035. https://doi.org/10.1161/circulationaha.118.035550

Pinn, V. W. (2003). Sex and Gender Factors in Medical Studies. JAMA, 289(4), 397. doi:10.1001/jama.289.4.397

Sosinsky, A. Z., Rich-Edwards, J. W., Wiley, A., Wright, K., Spagnolo, P. A., & Joffe, H. (2022). Enrollment of female participants in United States drug and device phase 1-3 clinical trials between 2016 and 2019. Contemporary clinical trials, 115, 106718. https://doi.org/10.1016/j.cct.2022.106718

Vaccarezza, M., Papa, V., Milani, D., Gonelli, A., Secchiero, P., Zauli, G., Gemmati, D., & Tisato, V. (2020). Sex/Gender-Specific Imbalance in CVD: Could Physical Activity Help to Improve Clinical Outcome Targeting CVD Molecular Mechanisms in Women?. International journal of molecular sciences, 21(4), 1477. https://doi.org/10.3390/ijms21041477 jama.289.4.397

Sosinsky, A. Z., Rich-Edwards, J. W., Wiley, A., Wright, K., Spagnolo, P. A., & Joffe, H. (2022). Enrollment of female participants in United States drug and device phase 1-3 clinical trials between 2016 and 2019. Contemporary clinical trials, 115, 106718. https://doi.org/10.1016/j.cct.2022.106718

Vaccarezza, M., Papa, V., Milani, D., Gonelli, A., Secchiero, P., Zauli, G., Gemmati, D., & Tisato, V. (2020). Sex/Gender-Specific Imbalance in CVD: Could Physical Activity Help to Improve Clinical Outcome Targeting CVD Molecular Mechanisms in Women?. International journal of molecular sciences, 21(4), 1477. https://doi.org/10.3390/ijms21041477

Development of Pollen Extract-Coated Microneedles to Assess their Future Potential in Allergy Treatment

Mitra Junu

Nokola Tesla STEM High School

Abstract

Many people suffer from pollen allergies and currently, the main form of treatment, other than allergen avoidance, is immunotherapy through subcutaneous injections and sublingual administration and subcutaneous injections are the most common approach used in allergy immunotherapy because of its effectiveness but can cause side effects including skin irritation, swelling, itching, and more. Due to the large number of potential side effects, this project aims to develop a pollen-extract coated microneedle patch and assess its future potential in allergy treatment. The microneedle array was coated by combining a protein solution, which included scots pine pollen and phosphate buffer saline, with the excipients solution which included carboxymethyl cellulose to increase the viscosity, F-68 which was used as a surfactant to increase surface tension, and rhodamine B which was used as a fluorescent dye. The solution was coated on the microneedle array using a dip coating process. The microneedle array was then observed through a scanning electron microscope prior to injection and post injection in the silicone skin. The scanning electron microscope images were used to determine the difference in length prior to injection and post injection and the results showed that there was at least a 96% reduction in the length. The microneedles penetrated the silicone with a 100% success rate and a 99% accuracy rate. Overall, the pollen-extract coated microneedle patch exhibited significant efficacy, and further testing is warranted to explore its potential as a treatment option for allergy immunotherapy.

Introduction

This project aims to develop a microneedle patch and asses its future potential in allergy treatment. Pollen allergies affect over 24 million people in the United States (AAFA 2018). Pollen allergies can be so severe that they are fatal for some children, and currently there are a few methods that can be used to alleviate the allergic reactions. Some methods to treat allergic reactions include taking antihistamines; antihistamines work by blocking histamine in the patient's body. Histamines are released when the body detects something harmful, like an allergy. But the most effective form of treatment is simply avoiding the allergy and using epinephrine in case of exposure (Shakya 2022). Epinephrine, also known as adrenaline, is commonly used in allergic reactions and is injected into the patient during an anaphylaxis reaction. Epinephrine is used only in emergency situations. Prior to an anaphylaxis reactions, there is a preventative method that has proved to be successful against allergy symptoms. This method is allergy shots, also known as immunotherapy. Immunotherapy is a type of preventative treatment that uses small doses of the allergen, which is a substance that causes a patient to have an allergic reaction. It repeatedly injects increasing amounts of allergen extract over a certain period. Once it achieves maximum tolerance, the allergen injections are discontinued (Durham 2022). This process can take anywhere from a few months to a few years. By increasing the dosage of the allergen, the immune system becomes less sensitive to the allergen over time. Currently, the only form of immunotherapy is taking shots of the allergen. Although this is very effective, almost 2 in 3 children may have a fear of needles, which makes it difficult for them to take the

dose (CDC). A potential solution for this problem is using microneedles to deliver the allergen doses. Research has shown that microneedles increase skin permeability and have been used to deliver a variety of different drugs and vaccines. They have also proven to have fewer side effects compared to regular injections including less pain, skin irritation, and skin infections (Kim, Park, Prausnitz 2013). There are 4 methods of drug delivery using microneedles: solid, coated, dissolving, and hollow. Solid microneedles are used as a pretreatment to the drug. Once the microneedles pierce the skin and leave microscopic holes, the drug can then be applied to the skin using a topical formula and diffuse through the holes and into the skin. Coated microneedles involve the drug being coated on the microneedles and once penetrated, the drug dissolves into the skin. Dissolving microneedles are made using water-soluble or biodegradable materials. The material used encapsulates the drugs and release it into the skin once it is dissolved. Hollow microneedles inject the liquid into the skin (Kim, Park, Prausnitz 2013). See Figure 1 showing the 4 different methods of drug delivery using microneedles.

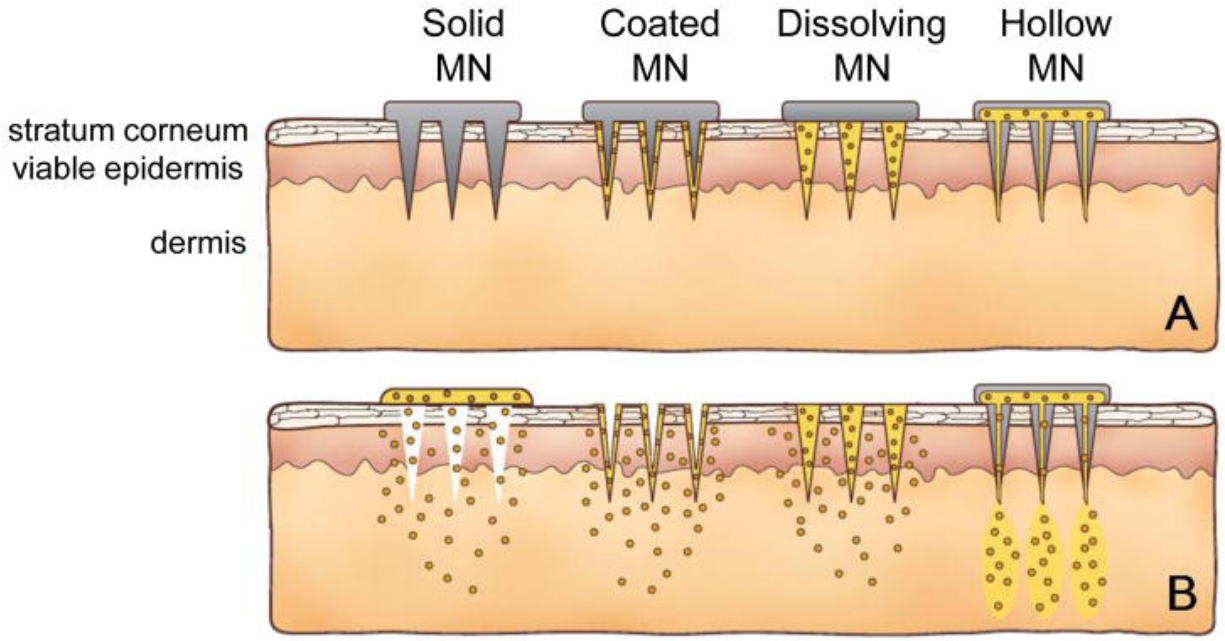


Figure 1. The 4 different methods of drug delivery using microneedles. The diagram shows how different methods deliver drugs into the skin. Solid microneedles allow the drug to be delivered using the holes from injection. Coated microneedles allow the coated drug to be dissolved into the skin. The dissolving microneedles encapsulate the drug in the microneedle and are released into the skin after the material has dissolved. Hollow microneedles inject the drug into the skin.

This project focuses on using coated microneedles. A similar study conducted at Texas Tech University by Akhilesh Kumar Shakya researched peanut coated microneedles and their effectiveness of desensitizing patients through the process of immunotherapy. Dr. Shakya coated his microneedles using a dip coating process. After coating his microneedles, he texted it on mice. Dr. Shakya made the mice allergic to peanuts and then used the microneedle patch overtime to observe its effect on the mice's symptoms. The results showed that 65.1% of the coated material was delivered into the mice's skin and that the microneedles were "safe and showed enhanced desensitization to peanut allergies" (Shakya, 2022). Dr. Shakya's research is a huge step towards a painless method for immunotherapy. This project uses the dip coating technique to create coated microneedles, similar to Dr. Shakya's research. This project seeks to develop a pollen-coated microneedle patch and asses its future potential in allergy treatment.

Methodology

This project seeks to investigate if a pollen-coated microneedle patch can effectively administer the required dosage to a patient. All steps in this experiment will be done under a fume hood at Tesla STEM High School. The AdminPatch 600 microneedle array containing eighty-five 800 um tall microneedles from AdminMed will be the array used throughout the testing process. After putting on a pair of latex gloves, the coating solution will be made using a 1:1 ratio of the protein solution and the excipients solution. The protein solution is prepared using 10 µg of pollen extract, dissolved in 1 ml of phosphate buffer saline. The excipients solution is prepared by combining 3 µg of carboxymethyl cellulose and 3 µg of F-68. Combine the protein solution and the excipients solution using a micropipette. Once combined, add 1 µg of Rhodamine B into the solution. Rhodamine B will be used in the testing procedure. Using the combined mixture, fill the dam board with the solution. Use a dip coating technique to coat the needles, making sure to evenly dip the microneedle array into the pollen coating formula. Refer to Figure 2 for a diagram on the dip coating process. Refer to Figure 3 for a diagram of how a coated and uncoated microneedle should look like.

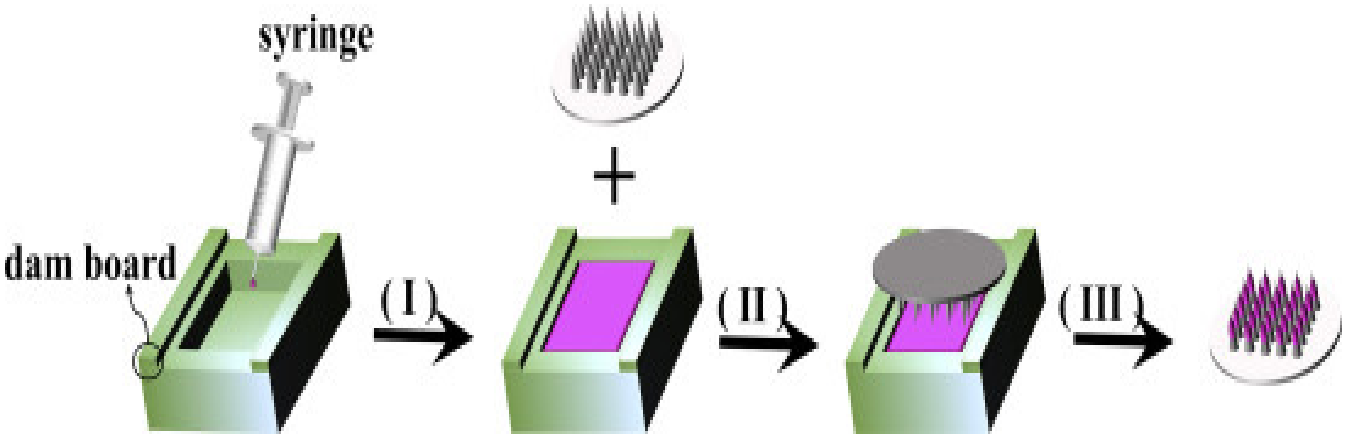


Figure 2. Dip coating method for microneedle array. Step 1 is filling the dam board with the coating solution. Step 2 and 3 display the process of dipping the microneedle array upside down to coat the solution on the microneedles.

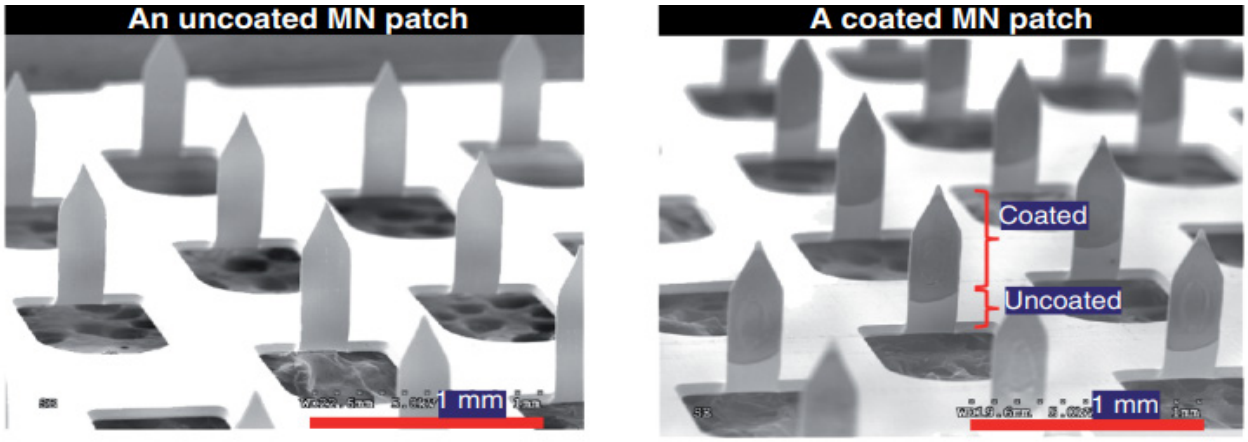
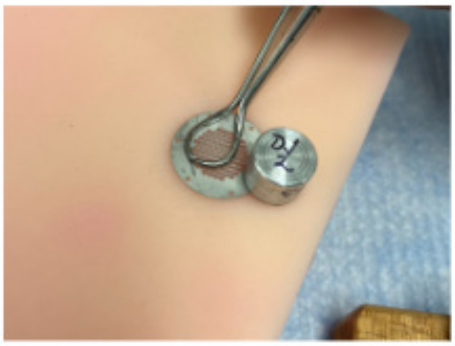


Figure 3. Scanning electron microscopy images of an uncoated microneedle patch (left) and coated microneedle patch (right).

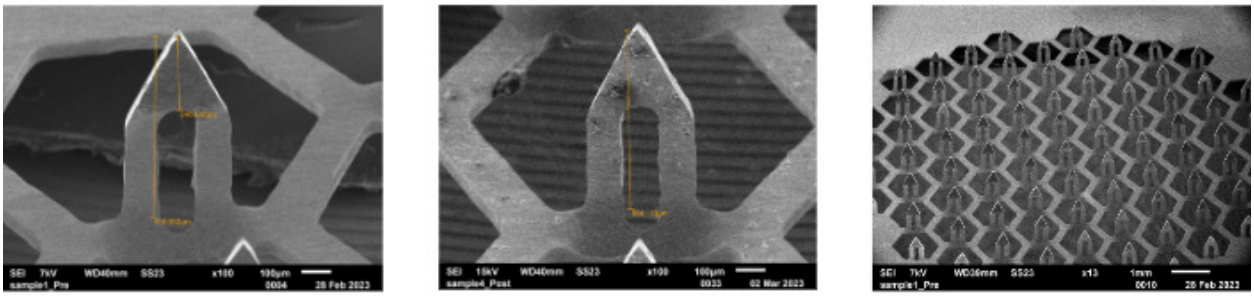
Once the microneedle array is coated, suspend the microneedle array upside down and let the coating dry for at least 15 minutes. Once the coating on the microneedle array is dry, the testing procedure can begin. Start off by putting on a pair of latex gloves. Then, mount the microneedle arrays onto aluminum stubs with carbon tape. The samples were imaged on a JSM 6610 LV scanning electron microscope at an accelerating voltage of 15kV (JEOL Ltd, Tokyo, Japan). Electron microscopy image data were generated with assistance from the Fred Hutchinson Cancer Center Electron Microscopy Shared Resource (Seattle, WA). Using the SEM images, measure the coating of the solution prior to the insertion and input that into column 1 of data table 1. Repeat this process and observe a total of 12 microneedles; 3 microneedles from 4 different microneedle arrays. Then, clean an area on the silicone skin (at least 1 in by 1 in) using isopropyl alcohol. Then, flip the microneedle array upside down and insert it into the silicone skin, making sure to keep even pressure on the microneedle array and hold it for 5 seconds. Once releasing the microneedle array, measure the length of the needle that has coating formula on it using a scanning electron microscope (in μm) and record it in column 2 of data table 1.



Injecting Microneedle Array

Figure 4. Technique used to inject the microneedle array into the silicone skin.

SEM Images Collected



Microneedle Array Prior to Injection

Microneedle Array Post the Injection

Microneedle Array

Figure 2. SEM Images Collected from prior to injection, post injection, and microneedle array.

Data Tables

Data Table 1: Length of pollen-extract coating prior and post injection

Microneedle #	Length of Coating Prior to Insertion (µm)	Length of Coating Post the Insertion (µm)
A 1, MN 1		
A 1, MN 2		
A 1, MN 3		
A 2, MN 1		
A 2, MN 2		
A 2, MN 3		
A 3, MN 1		
A 3, MN 2		
A 3, MN 3		
A 4, MN 1		
A 4, MN 2		
A 4, MN 3		

Data Analysis

To analyze the data given, calculate 4 different factors. These factors are the difference in length, the percent difference, the accuracy rate, and the success rate. The formula used for the difference in length is as follows:

$$\text{length of coating prior to insertion} - \text{length of coating post the insertion}$$

The difference in length was calculated for all 12 microneedles observed. Then, the percent difference was calculated using the formula as follows:

$$\frac{\text{length prior to insertion} - \text{length post the insertion}}{\text{length prior to insertion}} \times 100$$

The percent difference shows the change in length of the coating prior and post the insertion. Once this is calculated, the accuracy rate was determined by calculating the average of the percent differences. The accuracy rate shows the precision of the change in the coating. Once the accuracy rate was calculated, the success rate was determined by the number of microneedles that achieved a percent difference above 90%.

Conclusions

After collecting all my data, I was able to determine the difference in length and the percent difference as seen below:

Microneedle #	Length of Coating Prior to Insertion (µm)	Length of Coating Post the Insertion (µm)	Difference In Length	Percent Difference
A 1, MN 1	268.03	0	268.03	100.00000000
A 1, MN 2	248.032	0	248.032	100.00000000
A 1, MN 3	234.009	0	234.009	100.00000000
A 2, MN 1	305.823	8.349	297.474	97.26998950
A 2, MN 2	277.382	0	277.382	100.00000000
A 2, MN 3	247.023	0	247.023	100.00000000
A 3, MN 1	318.402	5.723	312.679	98.20258667
A 3, MN 2	260.031	0	260.031	100.00000000
A 3, MN 3	272.184	0	272.184	100.00000000
A 4, MN 1	257.875	9.365	248.51	96.36839554
A 4, MN 2	244.673	0	244.673	100.00000000
A 4, MN 3	221.394	0	221.394	100.00000000

I can conclude that my pollen extract coated microneedle array has a 100% success rate and a 99% accuracy rate. This proves my research question correct and achieves both of my engineering goals. I was able to make a pollen extract coating solution and the percent difference of the microneedle array was at least 90% across all 12 microneedles observed. Some limitations in this research include letting the coating solution dry for too long. When I was allowing the microneedle array to dry, the times varied which could have affected the transfer of the coating solution. If the coating solution was too dry, it would stick onto the microneedle and wouldn't transfer as well. Another limitation was the thinness of the microneedle array. The metal of the microneedle array was extremely thin and handling the microneedle array incorrectly could have resulted in a bend in the microneedle array. This could have made the microneedle array not flat and could have affected the evenness of the coating on the microneedle array. Another factor that could have affected the evenness of the coating was not having access to an automated dipping system. The automated dipping system would have ensured that the coating was even across the entire microneedle array.

Discussion

I could apply my research to conduct mice clinical trials. This would allow me to test the immunotherapy aspect of my microneedle array. Before I conduct mice clinical trials, I would also like to change the pollen extract to red cedar. I wasn't able to purchase red cedar extract and it would have been ideal to use it because it's most native to Washington State. I would also like to create a "patch" around the microneedle array. The current microneedle array has a thin outer metal area that can be held but it is difficult to use. Creating a patch around the microneedle array allows the microneedle array to be more user friendly.

Bibliography

Durham, S. R., & Shamji, M. H. (2022). Allergen immunotherapy: Past, present and future. *Nature Reviews Immunology. https://doi.org/10.1038/s41577-022-00786-1*

Ingrole, R. S. J., & Gill, H. S. (2019). Microneedle Coating Methods: A Review with a Perspective. Journal of Pharmacology and Experimental Therapeutics, 370(3), 555–569. https://doi.org/https://doi.org/10.1124%2Fjpet.119.258707

Kim, Y.-C., Park, J.-H., & Prausnitz, M. R. (2012). Microneedles for drug and vaccine delivery. Advanced Drug Delivery Reviews, 64(14), 1547–1568. https://doi.org/https://doi.org/10.1016/j.addr.2012.04.005

Landers, J. J., Janczak, K. W., Shakya, A. K., Zarnitsyn, V., Patel, S. R., Baker, J. R., Gill, H. S., & O'Konek, J. J. (2022). Targeted allergen-specific immunotherapy within the skin improves allergen delivery to induce desensitization to peanut. Immunotherapy, 14(7), 539–552. https://doi.org/10.2217/imt-2021-0206

Oh, N. G., Hwang, S. Y., & Na, Y. H. (2022). Fabrication of a PVA-based Hydrogel Microneedle Patch. Fabrication of a PVA-Based Hydrogel Microneedle Patch, 7(29), 25179–25185. https://doi.org/10.1021/acsomega.2c01993

Pine Needles Extract | Safety data sheet . Selleckchem.com. (2014, May 1)., https://www.selleckchem.com/msds/pine-needles-extract-MSDS.html

Safety Data Sheet - Carboxymethyl Cellulose. ThermoFisher Scientific. (2021, December 25)., https://www.fishersci.com/store/msds?partNumber=AC332640010&productDescription=CARBOXYMETHYL+CELLULOSE%2C+1KG&vendorId=VN00032119&countryCode=US&language=en

Safety Data Sheet - Pluronic F-68 Solution. ThermoScientific. (2010, July 20). Retrieved from https://static.thermoscientific.com/images/D02113~.pdf

Domain Knowledge vs AI: Detecting Cervical Spine Fractures from CT Scans

Vasudha Narayanan

Nikola Tesla STEM High School

Abstract

1.5 million cervical spine fractures occur each year in the United States, and it is critical that these fractures are detected accurately, as misdiagnosis can lead to paralysis and death. Diagnostic radiologists are medically trained to detect fractures from radiographs, but in rural emergency clinics the task may fall to a lone clinician, who is less trained and offered no second opinion to confirm their diagnosis, which is detrimental to patient outcomes. This project seeks to investigate if convolutional neural networks (CNNs) are a viable means of providing a second opinion in these rural clinics, by comparing the performance of various CNNs to the performance of diagnostic radiologists in detecting cervical spine fractures from CT scans. Radiologists were administered tests, in which they were asked to determine whether a given scan was fractured or not, and their performance accuracy was calculated. Then, multiple CNNs (untrained CNN, ResNet, DenseNet, InceptionNet-V4, EfficientNetV2) were each implemented on a dataset of cervical spine CT scans and trained to determine the fracture status of each scan. Results revealed, respectively, a 67.3%, 87.9%, 85.6%, 93.2%, and 95.3% accuracy of each model in detecting fractures, while the average accuracy of all radiologist participants was 86.3%. With 3 out of 5 implemented CNNs outperforming the radiologists, these results reveal CNNs strong performance in radiographic fracture detection. These results can additionally be used to facilitate the implementation of convolutional neural networks in radiological clinics, as their high accuracies can benefit both clinician performance and patient outcomes alike.

Keywords: Cervical spine; convolutional neural networks; CT scans; fracture detection; radiologists

1. INTRODUCTION

1.1 Background: The Cervical Spine

The cervical spine is the area of the spine located in the neck. It is composed of the C1 through C7 vertebrae, and is vital for many important neck movements, as well as holding the head up. Most importantly, the cervical spine is responsible for protecting the spinal cord. If the spinal cord is injured in any way, it almost always leads to paralysis, or even death. Thus, it is imperative to keep the cervical spine healthy, and to be able to quickly know when it is not.

The most common way the cervical spine is damaged is when it is fractured. Any one of the 7 vertebrae can be fractured, and in severe cases multiple vertebrae can be fractured at one time. Cervical spine fractures occur due to high impact accidents, such as falling from an extreme height, or a high speed car crash. While severe, patient outcomes from cervical spine fractures will be positive, if the spinal cord is undamaged, and the fractures are detected quickly and accurately, and treated correctly. However, this quality of detection cannot always be assured.

1.2 Current Issues in Fracture Detection

Diagnostic radiologists are the professionals medically trained to detect bone fractures from

radiographs, such as X-rays, MRIs, and CT scans (Gan et al). However, radiologists make upwards of $350,000 USD per year, which limits the areas in which they can be employed.

Radiologists are commonplace in large urban hospitals, yet in in rural emergency clinics, which serve only a couple hundred people, a radiologist may not be present for miles. Thus, the task of fracture detection would fall to a lone clinician, who is much less experienced than a radiologist, and thus is at an immediate disadvantage. These factors have led missed fractures to account for up to 79.7% of diagnostic errors in the United States, which is detrimental to patient outcomes (Lindsey et al).

Clinicians will not be able to provide the quick and accurate fracture detection required for patients to recover fully from cervical spine fractures. This is especially problematic due to the highly independent culture of rural towns in America. People are much more inclined to do manual labor themselves, and thus the risk of high trauma incidents that cause cervical spine fractures is exponentially higher than in the cities.

1.3 Past Research

While not specifically for the purpose of rural clinics and cervical spine fractures, research has gone into the automated detection of fractures from radiographs, through utilizing machine learning models. These models are image detection models, which can be fed scans and predict whether a fracture is present or not. The class of models most popular for image detection are convolutional neural networks, or CNNs. CNNs have been trained to detect fractures, with notably high accuracies (Adams et al). Yet, these numbers are standalone, and thus do not have much meaning in the real world. One way to give these accuracies more concrete meaning is to establish a baseline of radiologist performance to compare model performance to, in order to deepen the understanding of exactly how proficient convolutional neural networks are in detecting fractures.

1.4 Study Goals

This project seeks to investigate the performance of diagnostic radiologists versus the performance of five different convolutional neural networks (Generic CNN, DenseNet, ResNet, InceptionNetV4, EfficientNetV2), in detecting Cervical spine fractures from CT scans. Through this comparison, the real power of CNNs could be discovered. If convolutional neural networks end up performing better, i.e., more accurate, than a sample of diagnostic radiologists, then they may be an ideal solution to the issues of rural clinics. CNNs could be implemented to provide rural clinicians with an expert second opinion, making them more confident about their diagnosis, and making patients more confident about the care that they will receive.

2. METHODOLOGY

2.1 Dataset Investigation and Analysis

CT scans of 2019 different cervical spines were obtained from the RSNA 2022 Cervical Spine Fracture Dataset, an open-source dataset made publicly available via Kaggle. The dataset was organized by patient ID, with each patient containing a label of whether their cervical spine was fractured or not, along with images of the cervical spine.

However, these images could not easily be viewed initially, as they were in the DICOM medical image format. This format, rather than the traditional JPEG or PNG image format, caused the dataset to be extremely large, at 398 GB, which limited what could be done with the dataset at that point. Laptops do not support the DICOM format, which meant it was impossible to simply view any of the images in the dataset.

Additionally, the dataset was not structured in a way that there was one image of the cervical spine per patient. CT scans are taken by travelling down the spine, and taking several pictures, called slices, in quick succession. Thus, there were anywhere from 200 to 600 slices present for a given patient.

2.2 Dataset PNG and 3D Conversion

In order to view the data, it has to be converted from DICOM into a more common image format. This was accomplished through the dicom2png helper class. Every slice in the dataset was converted into a PNG image, which took around 26 hours.

After conversion, the images in the dataset could be easily viewed. This immediately revealed an issue with the slices. Since they were taken in such quick succession, many slices closely resembled one another, which would lead to overfitting of the model. Additionally, for patients labelled as fractured, not every slice contained a fracture, since not every vertebra of each patient was fractured. To overcome this, the slices of each patient were taken and converted into a 3D sagittal view of the cervical spine, which displayed every fracture at once. This was accomplished through the RSNACervicalSpineFracture TorchIO class, which converted scans into either a sagittal, coronal, or axial view. After 3D conversion, scans were regulated with the Hounsfield Medical Image Scale in order to make the background darker, and the white bone stand out more. The dataset was then split into 70 percent training, 20 percent validation, and 10 percent testing.

2.3 Radiologist Recruitment

Diagnostic radiologists were recruited through an interest form made with Google Forms. The questions on the form included: Participant Age, Participant Email, Participant Biological Sex, and Previous Radiology Experience. Flyers were then created, with a QR code to the interest form on them, and posted in Overlake Emergency Clinic, Evergreen Hospital, Overlake Hospital, Redmond Library, and Sammamish Library. Radiological clinics were also called and given information about the study to pass on to the radiologists practicing at that center. Once responses were collected, each respondent was assigned a number. Then, a random number generator was used to randomly select 10 participants, who were all sent an informed consent form, an information flyer, and a parental consent form in the case the participant was under the age of 18.

2.4 Radiologist Data Collection

Data was collected from the radiologist participants through a Microsoft forms test composed of 100 questions. Each question included a sagittal, coronal, and axial view of the cervical spine, along with a multiple choice simply asking whether the radiologist believed the scan was fractured or unfractured. Data was collected 100% offline, but radiologists were given guidelines to complete the test in one sitting, and to refrain from receiving outside help. The answers radiologists inputted were collected as a 0 (unfractured) or 1 (fractured) and compared to the correct answer key. From this comparison, the sensitivity, specificity, and accuracy of each radiologist was calculated and inputted into Table 1.

2.5 CNN Training and Testing

5 different convolutional neural networks were utilized in this study, in order to give a range of performance to compare to the radiologist baseline. The 5 models implemented were a Generic CNN (not pretrained), DenseNet, ResNet, InceptionNetV4, and EfficientNetV2. All models were trained, validated, and tested on the same sets of images. Additionally, all models were only evaluated on the sagittal view of the cervical spine, rather than the slices, to avoid overfitting and confusion.

After training and subsequent validation was completed for a model, it was evaluated on the testing set. Beyond just receiving the simple accuracy score, if a model was given a specific scan, it would output a number, along with what it believed the classification for the scan was. For any given model, if a value less than 0.5 was predicted, then the scan was labeled as fractured. If a value greater than 0.5 was predicted, then a scan was labeled as unfractured.

Figures and Tables

Figure 1: Similarity of scans across one patient

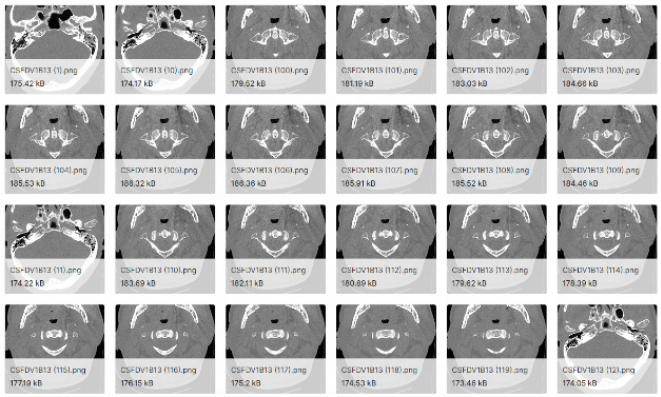


Figure 2: Sagittal View of the Cervical Spine

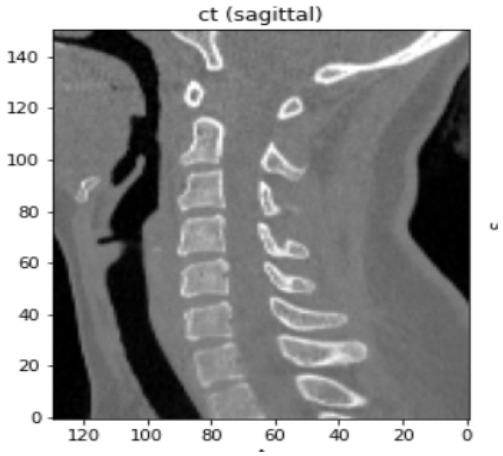


Table 1 –Sensitivities, Specificities, and Accuracies of each Radiologist Tested

	Rd 1	**Rd 2**	**Rd 3**	**Rd 4**	**Rd 5**	**Rd 6**	**Rd 7**	**Rd 8**	**Rd 9**	**Rd 10**
Sensitivity	44.2%	84.6%	94.4%	92.1%	85.4%	81.8%	85.5%	98.1%	96.2%	90.9%
Specificity	86%	78.7%	87.5%	80.6%	91.8%	84.4%	84.4%	89.4%	89.4%	88.9%
Accuracy	68%	81%	90%	85%	91%	83%	85%	94%	93%	90%

Table 2 – Accuracy of Each CNN Model

	Generic CNN	**ResNet**	**DenseNet**		**InceptionNetV4**	**EfficientNetV2**
Accuracy	67.3%	87.9%	85.6%		93.2%	95.3%

3. DATA ANALYSIS

3.1 Statistical Analysis and Comparison

The average detection accuracy of radiologists tested came out to 86%. One major outlier existed, with an accuracy of only 66%, but data was kept in the study due to the small sample size of radiologists available. The most highly accurate radiologist had an accuracy of 94%. In comparison, the accuracies of the models are as follows: Generic CNN(67.3%), ResNet(87.9%), DenseNet(85.6%), InceptionNetV4(93.2%), and EfficientNetV2(95.3%). Generic CNN was a low outlier for the models, likely because it had not been pretrained on a large image dataset like all of the other models.

On average the radiologists outperformed both Generic CNN and DenseNet, but underperformed ResNet, InceptionNetV4, and EfficientNetV2. The lowest performing radiologist was outperformed by every single model, while the highest performing radiologist was only outperformed by EfficientNetV2.

4. CONCLUSIONS

4.1 Relevance of Results

These results reveal the applicability of convolutional neural networks in fracture detections. If models can indeed be created that are consistently more accurate than radiologists and clinicians, then they can easily be promoted as a tool. While the sample size of radiologists tested here was very small, and thus not representative of the total population, the results of this study still hold up when compared to similar research. While in similar research, in which radiologist accuracies are calculated, the same type of fracture detection is not taking place, it is similar enough that the results can be referenced as a sub-baseline to this research. Compared to other studies, at least one model still is more accurate than the average radiologist, and in some cases, all but one model are more accurate. Thus, these results can still hold true, even against a more representative radiologist population.

With results holding up against similar research, the findings of these study are extremely relevant to both groups that automated fracture detection can benefit: doctors and patients.

4.2 Future Applications

As presented multiple times in this study, a future application of fracture detecting machine learning models is in rural imaging clinics. With a highly accurate model always available to provide a qualified second opinion on fracture existence, clinicians will be able to provide higher quality care, no matter what resources they previously had to work with.

Another future application of this study would be to edit the outputs of the models, so not only do they output whether a fracture is present in a scan, but also create a bounding box around the exact location of the fracture. This would make the model infinitely times more helpful to the clinician.

5. ACKNOWLEDGMENTS

I would like to thank my mentor, Mr. Cheng Wu, for his support and guidance throughout this entire project. Without his expertise and suggestions, this project would not have been possible. I would also like to thank my teacher, Mrs. Kate Allender, who's resources and encouragement were invaluable.

REFERENCES

Adams, M., Chen, W., Holcdorf, D., McCusker, MW., Howe, P., Gaillard, F. (2018). *Journal of Medical Imaging and Radiation Oncology,* 63(1), 27-32.

Gan, K., Xu, D., Lin, Y., Shen, Y., Zhang, T., Hu, K., Zhou, K., Bi, M., Pan, L., Wu, W., Liu, Y. (2019). *Acta Orthopaedica,* 90(4), 394-400

Hallas, P. Ellingsen, T. (2006). *BMC Emerg Med* 6:4.

Kaiume, M., Suzuki, S., Yasaka, K., Sugawara, H., Shen, Y., Katada, Y., Ishikawa, T., Fukui, R., Abe, O. (2021). *Medicine,* 100(20).

Kim D H, MacKinnon T. (2018). *Clin Radiol,* 73(5): 439-45.

Kitamura, G., Chung, C., Moore, B. (2019). *Journal of Digital Imaging,* 32, 672 – 677.

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., Potter, H. (2018). *PNAS,* 115(45), 11591 – 11596.

A 3D-Printed Scoliosis Brace Designed to Increase User Comfort and Mobility

Inaya Kamal

Nikola Tesla STEM High School

Abstract

Scoliosis is a condition where the spine twists and curves, affecting 200 million people worldwide, and bracing is the most common form of treatment but is often uncomfortable for the patient. This case-study tests the comfortability of a novel, 3D-printed scoliosis brace with a survey and the effectiveness of the brace by measuring the rotation of the spine. The goal of the brace is to be more comfortable than a Wood-Cheneau Rigo brace while maintaining the necessary amount of pressure on the spine needed to correct the rotation of the scoliosis. The novel brace was created with computer-aided design (CAD) and fitted once for adjustment. The rotation of the patient's spine (dependent variable) was measured with and without the novel brace (independent variable), and the patient completed a comfort survey to self-report data about the comfort of the brace. The overall spine correction for rotation was 39% and statistically significant, with a p-value of 0.0025. The greatest amount of correction was measured to be 5˚ and 100% improvement. The brace was deemed comfortable, scoring 13/16 on the survey. This novel brace provides evidence that 3D printed braces with less coverage can effectively apply the pressure needed to correct a patient's spine, meeting the design criteria and also offers a more affordable option compared to conventional braces.

Introduction

The effectiveness and comfortability of a newly designed scoliosis brace were tested by a comfort survey and observation of the rotational difference in each vertebra when the patient did and did not wear the brace.

Scoliosis is a condition where the spine unnaturally curves and twists, usually in a period of growth. In the total world population, 200 million people, have been diagnosed with scoliosis between the ages of 10 and 15 years. The curvature and rotation of the spine causes patients daily pain, which can make daily or physical tasks difficult to complete, and greatly decreases a patient's perception of their body and self-esteem (NHS, 2022). If the curve or rotation is severe enough, it can cause a patient's ribcage to push against their heart and lungs, making it difficult to pump blood and provide oxygen throughout the body, and can result in frequent lung infections such as pneumonia. Most scoliosis patients, however, lead relatively normal lives and only experience minor pain on a daily basis.

To measure the severity of scoliosis, most doctors and therapists use the Cobb angle of the spine. The Cobb angle is found by using an X-ray of the patient's spine to identify the endplates of the curve and drawing lines parallel to the endplates creates an angle, the Cobb angle. The greater the angle, the more severe the condition. An alternative method of measuring the severity of scoliosis is measuring the rotation of the vertebra. Rotation caused by scoliosis can cause drastic changes in appearance and affect the patient's ability to breathe comfortably. A spine without scoliosis typically has 0° of rotation and scoliotic spine has an average rotation greater than 5°.

The most common treatment method for

scoliosis is bracing. Brace treatment involves the patient wearing a rigid piece of plastic around their torso. The brace applies pressure to the spine in specific points and directions in order to push the spine in a position where it is less curved and rotated. Braces are prescribed mostly to those in a period of growth, adolescents and children, because scoliosis progresses the most when the spine is growing. However, adults do use braces to offer support and correction for their scoliosis. Bracing is used when the Cobb angle of the spine is 25° or more, an indication of severe scoliosis, and prevents 75% of wearers from needing spinal fusion surgery for permanent correction (Boston Children's Hospital).

Patients find their braces uncomfortable to wear for 16-22 hours each day. Due to the discomfort, patients avoid wearing their braces as much as possible, which poses a problem to the effectiveness of the treatment. The comfort of the patient in-brace should be high in priority, because the curve reduction is greater when the patient wears the brace longer (Lou, Ng & Hill, 2022). Patients usually find issue with the brace firmly digging into their sides at all times, and their thighs when they sit down. Braces also restrict the patient's movements and retain a lot of heat due to the rigid nature and corset-like design. One brace that has strayed a bit from the traditional design of a brace is the Rigo-Cheneau brace, which has been rising in popularity because of the unique holes that orthotists design into the brace. These holes allow cool air to reach the patient's body while in-brace and create space for the patient's torso to expand when breathing. It has been found to be even more effective that most other braces available, so the idea of reducing the plastic coverage doesn't seem to take away from the effectiveness of the brace.

3D printed braces were reported to be lighter and feel less clunky than a traditional brace and were cheaper to manufacture than most other braces. (Lou et al, 2022). Instead of a corset design, the experimental brace only incorporates plastic where pressure needs to be applied to the spine, each one designed to fit the patient's form exactly through CAD. The different plastic components were connected using belts so that the patient had a wider range of motion, had spaces in the brace where their chest could expand, and to allow cool air under the brace and prevent heat build-up.

The goal of the experimental brace was to lessen the discomfort of the patient and the restriction of movement in comparison to the Rigo-Cheneau brace, while still applying enough pressure to reduce the rotation of the scoliosis in-brace. The rotation of each vertebrae was used to test the efficacy of the experimental brace in comparison to not wearing the brace. When using the brace, the average rotational correction of the spine should be 50% and the score on the comfort survey should be at 10 to deem the brace more comfortable than a Rigo-Cheneau brace and effective at treating scoliosis.

Methodology

Sampling Procedure

Dr. Elizabeth Apple contacted her patients with scoliosis to ask if they were interested in participating in the experiment. The possible participants confirmed whether they were an adult or a minor, whether they used a scoliosis brace or not, and provided an email address to contact if they were selected to participate. If multiple eligible participants had shown interest in the study, a random number generator would have been used to select a participant to ensure that the selection was completely random. Only one participant was to be selected to ensure that the curve was the same for all testing conditions, and only one showed interest and agreed to participate. When the participant agreed to participate in this case study, they were asked for informed consent.

Design Process

As shown in Fig. 1, the design process began with identifying the points on the participant's where pressure needed to be applied for correction and what points should aid in the pressure application. Each of these points needed a form fitting component, which was designed in Autodesk Inventor as shown in Fig. 2. All of the pieces were printed using PLA plastic from a soybean printer. Each piece was tested for fit and additional comfort feedback from the participant. All observations from the fitting were used to adjust the measurements and design of the components, which were reprinted for testing. Foam was attached to each piece to ensure that gaps

were filled and that the brace didn't cause discomfort during use. Right before testing, the components were attached to heavy cotton webbing belts that were fitted to apply pressure while being adjustable with plastic side release curved buckles. The first prototype of the brace was tested after the belts were fitted and tightened.

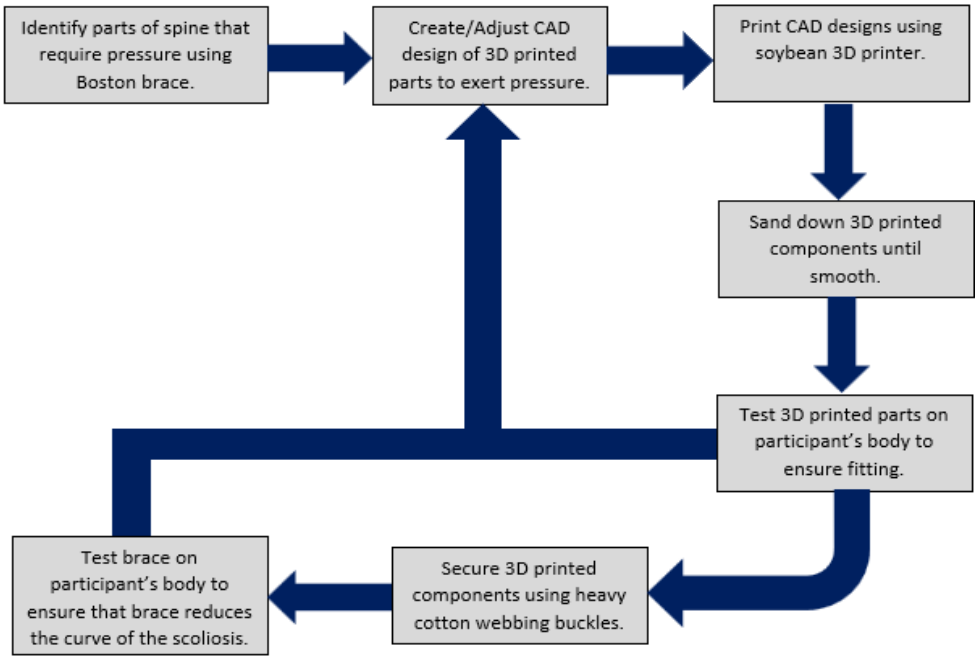


Figure 1. The scoliosis brace design process. The design of each component must be tested to find any possible flaws with the design that could decrease the comfort of the end product. This reiteration must also happen with the final product for the same reason.

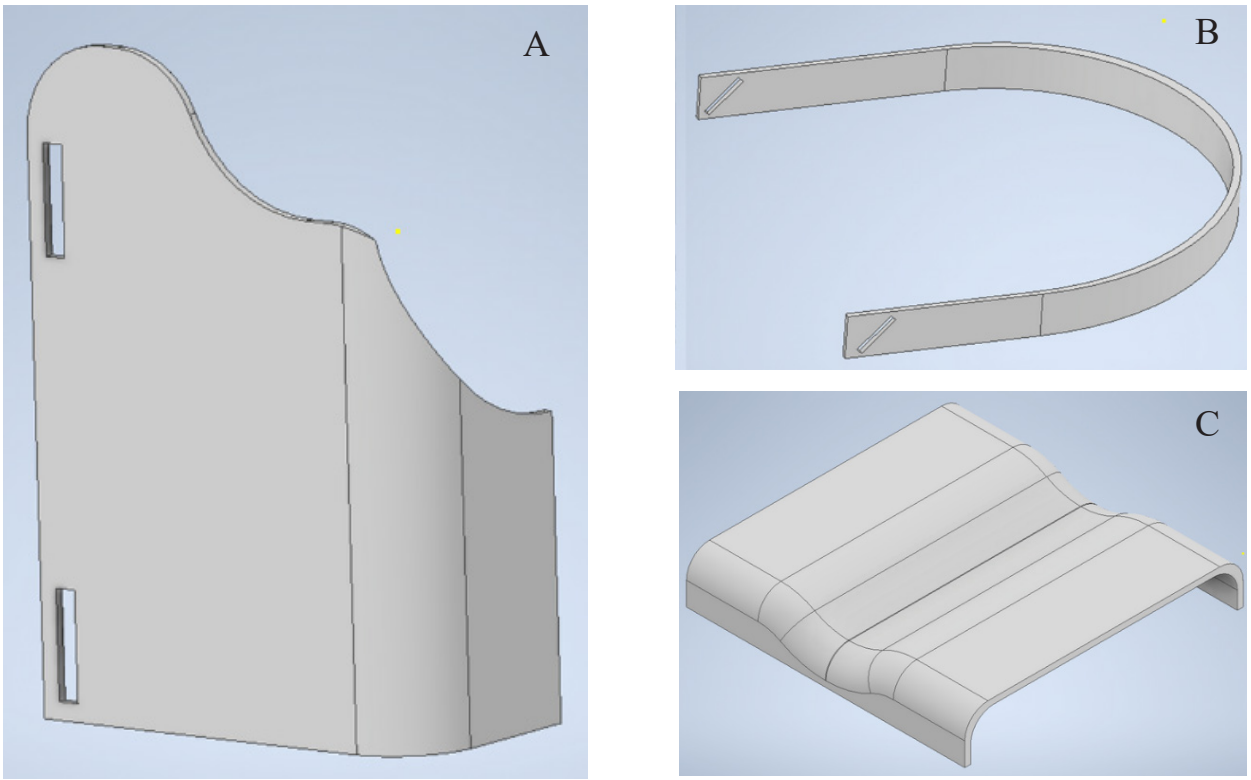


Figure 2. Initial CAD design of the brace. A) The thoracic piece is meant to apply pressure to the right ribcage, underneath the arm. B) The pelvic piece is meant to act as a basis for the rotational corrective pressure. C) The shoulder piece is meant to aid in pressure application to the thoracic spine for rotation and elongation.

Experimental Procedure

Once the brace had been fully fitted for the patient, the patient was instructed to bend over forward to measure the rotation for each vertebra. A scoliometer was placed over each vertebra to measure which direction it was rotated and by how many degrees. If the vertebrae were rotated to the right, the rotation was considered to be positive, and vertebrae rotated to the left were recorded as negative rotation. Once the rotation of the vertebrae were measured with the brace, the same was done without the brace present. The overall comfort was assessed with a comfort survey created for this study. Typically, comfort is measured with compliance rate, however, compliance requires long-term testing of the brace, while this study only measures immediate effect. The questions were produced based off of discomforts caused by other braces so that the results could compare the experimental brace to the general idea of other braces in terms of comfort alone. Answers in favor of the brace's comfort added 2 points to the total score, answers not in favor of the brace's comfort added 0 points to the total score, while a middle option added 1 point to the total score.

The average percent difference and difference in degrees between the rotation of the vertebrae with and without the brace were calculated. To analyze the statistical significance of the efficacy, a paired T-test on GraphPad was performed. The p-value resulting from the calculation is the amount of uncertainty in the data collection and is considered significant if it is below 0.05.

Results

Brace Design

The final brace design consisted of 3 components: a thoracic component, a pelvic component, and a shoulder component. The thoracic piece and pelvic piece were 3D printed with 35% filling. For the pelvic piece, the low filling allowed the flexibility needed to account for minor changes in the participant's physique over time. The shoulder piece was replaced by a piece of foam since there was no pressure application needed for the upper thoracic and cervical spine.



Figure 3. Final design of the brace, shown on participant from the back. Pelvic piece is worn on the left side of the lower torso. Thoracic piece is worn on the right, below the arm.

The participant had a right thoracic curve, so the pressure application was focused onto the spine right underneath the shoulder and arm. The thoracic component, shown in Fig. 3, is worn under the right arm in order to apply pressure to the right ribcage connected to the thoracic spine. Its main purpose was to pull the right ribcage forward to reduce the rotation of the spine in comparison to the pelvis, resulting in the left ribcage also moving backwards.

The pelvic piece, also shown in Fig. 3, was worn around the left side of the lower torso to aid in the pressure application of the thoracic piece. The pelvic piece is essential to scoliosis correction because the pelvis is always at 0˚, and acts as the basis for any corrections done by the brace. The two pieces were connected to each other with heavy cotton webbing belts and side release buckles so that the participant could adjust the fitting at any point. The length of the belts determined the pressure application of the brace, so the belts in front of the participant were shorter and the belts behind the participant were longer. This was to pull the right ribcage forward and

adjust the vertebrae closer to 0˚.

Brace Efficacy

While in use, the brace reduced the rotation of the curve up to 5˚ on the vertebra T5, which was equivalent to 100% correction. The lowest points of correction were 0˚ and 0% correction at T7. The vertebrae T1, T2, L4, and L5 were all measured to have 0˚ of rotation without the brace, so they were not included in the calculation of efficacy. On average, the brace reduced the rotation of the spine by 2.583˚, which was a 39% correction. The p-value resulting from the T-test was 0.0025, indicating that the results were significantly significant.

Table 1. The rotation of each vertebra with the experimental brace and without it. The difference in degrees and percent change between the two conditions

Vertebrae	Rotation Without Brace (°)	Rotation With Brace (°)	Difference (°)	Difference (%)
T1	0	0	0	0
T2	0	0	0	0
T3	3	0	3	100
T4	5	0	5	100
T5	9	4	5	56
T6	11	8	3	27
T7	10	10	0	0
T8	11	8	3	27
T9	10	5	5	50
T10	9	5	4	44
T11	7	6	1	14
T12	6	5	1	17
L1	-5	-	-	-
L2	-3	-4	1	33
L3	-2	-2	0	0
L4	0	0	0	0
L5	0	0	0	0
		Average:	2.583	39%

Brace Comfort

Table 2 shows the results of the comfort survey given the participant. The total score was a 13/16, so the brace was considered to be extremely comfortable by the participant. The participant verbally confirmed that the brace could not earn points for aesthetic value, which was expected due to the fact that efficacy and physical comfort were prioritized over the mental comfort of wearing the brace. The other two points were lost due to the fact that the participant felt a piece of hot glue digging into their side during the testing and that they were unsure of how the brace would feel on a daily basis due to lack of time spent in-brace.

Table 2. Comfort survey and participant responses (highlighted)

Question	Points (Y / S / N)
Does the brace support your spine when being used?	Y (2) / S (1) / N (0)
Does it feel like the brace is digging into your body when wearing it?	Y (0) / S (1) / N (2)
When wearing the brace, do you feel uncomfortably warm?	Y (0) / S (1) / N (2)
Can you bend over enough to touch your toes?	Y (2) / S (1) / N (0)
Can you lean side to side in a comfortable manner?	Y (2) / S (1) / N (0)
Is this brace more comfortable than the Rigo-Cheneau brace?	Y (2) / S (1) / N (0)
Does the brace look too awkward to wear in public?	Y (0) / S (1) / N (2)
Would you wear the brace on a daily basis?	Y (2) / S (1) / N (0)
Total Score:	13/16

Conclusion

The brace failed to reach 50% average correction but succeeded in outperforming the participant's previous Rigo-Cheneau brace. The participant of this study was not in a period of growth, so the lower average correction in the brace would've sufficed to provide comfort and support on a daily basis. This brace would not be recommended for adolescents or children because higher correction rates would be needed to prevent curve progression in larger periods of growth.

The brace broke during testing due to thinness of the thoracic component. Weight was not reported to be a major concern of scoliosis braces, so future brace design would be thicker in order to accommodate higher amounts of pressure. The 3D printed components also need further fitting to ensure that they fit completely. Once the pieces have been redesigned and have undergone a final fitting, the plastic components will be attached together with a different form of adhesive, since the hot glue used in this study was not effective at holding the brace together for more than a few hours. The aesthetic component of the brace could also be considered in the design, since many adolescents avoid wearing the brace because of the mental discomfort of standing out among others.

Due to this being a case study, the results cannot be assumed to be generalizable. Future research would also include creating braces for new participants after the brace design is improved. This would involve creating braces that could be used for lumbar, double curve, and thoracolumbar scoliosis. All three of these designs would require very different components and pressure application than the thoracic brace designed for this study and may have varying efficacies.

References

Advanced Care Physical Therapy. (2022, October 24). *Scoliosis*. Advanced Care Physical Therapy. Retrieved October 27, 2022, from https://advancedcarephysicaltherapy.com/conditions/scoliosis/

Boston Children's Hospital. (2023). Scoliosis Bracing for Teens and Tweens. Boston Children's Hospital. Retrieved from https://www.childrenshospital.org/treatments/scoliosis-bracing

Chalmers, E., Hill, D., Donauer, A. *et al.* How does pressure configuration affect Cobb angle during AIS brace casting? *Scoliosis* **10** (Suppl 1), O60 (2015). https://doi.org/10.1186/1748-7161-10-S1-O60

Lou, E., Ng, K., Hill, D. (2022). Immediate Outcomes and Benefits of 3D Printed Braces for the Treatment of Adolescent Idiopathic Scoliosis. *Frontiers in Rehabilitation Sciences*, 3, 2673-6861. https://www.frontiersin.org/articles/10.3389/fresc.2022.840286. Accessed on 9/22/2022.

Meng, Zeng-Dong MD; Li, Tian-Peng MM; Xie, Xu-Hua MM*; Luo, Chong MM; Lian, Xing-Ye MM; Wang, Ze-Yu MD. Quality of life in adolescent patients with idiopathic scoliosis after brace treatment: A meta-analysis. Medicine: May 2017 - Volume 96 - Issue 19 - p e6828 doi: 10.1097/MD.0000000000006828

NHS. (2022, December 7). *Scoliosis. NHS inform. Retrieved January 20, 2023, from* https://www.nhsinform.scot/illnesses-and-conditions/muscle-bone-and-joints/conditions/scoliosis

Xu L, Yang X, Wang Y, Wu Z, Xia C, Qiu Y, Zhu Z. Brace Treatment in Adolescent Idiopathic Scoliosis Patients with Curve Between 40° and 45°: Effectiveness and Related Factors. World Neurosurg. 2019 Jun;126:e901-e906. doi: 10.1016/j.wneu.2019.03.008. Epub 2019 Mar 11. PMID: 30872192.

OHCT: One Hair Cell at a Time – Using ATOH1 Molecules to Reverse Hearing Loss

Aubrey Grove

Nikola Tesla STEM High School

Abstract

40 million people in the U.S. and hundreds of millions worldwide experience sensorineural (inner ear) hearing loss from continuous exposure to loud sounds, ototoxic drugs (drugs that are toxic for hearing loss), or aging (Winn). The purpose of OHCT is to reverse this hearing loss and regenerate the hair cells via a non-invasive and generalized solution: eardrops that can be administered by the patient themself. The eardrops send ATOH1 molecules through the ear drum and to the cochlea where the presence of the molecules induces inactive progenitor cells to differentiate into healthy cochlear hair cells. The drops' ability to reach the cochlea efficiently are tested using human membrane samples from throughout the ear canal, arranged in a columnar model. OHCT implies a more accessible solution to a wider population of the hearing loss community and focuses on administration without surgery, easy application, and affordability compared to hearing aids or cochlear implants.

Present Technology

40 million people in the U.S. and hundreds of millions worldwide experience sensorineural (inner ear) hearing loss (Winn). It's a natural part of aging and can also be onset through continuous exposure to loud noise – meaning, it affects everybody in some way. The goal of this product is to serve as a permanent, noninvasive, general solution for sensorineural hearing loss. The human ear consists of three main regions: outer, middle, and inner ear (Figure 1). The outer ear (pinna and ear canal) channels soundwaves from the environment onto the eardrum. Next the eardrum vibrates the hammer, anvil, and stirrup which vibrates the fluids in the cochlea. The cochlea is lined with hair cells which are responsible for converting the vibrations into electrical impulses that are transmitted to the brain via the auditory nerve. The electric impulses from the right ear go to the left auditory cortex and vice versa. The auditory cortices organize and interpret the signals to focus and suppress certain sounds. Damage to any of these systems (mainly the hair cells) can lead to permanent or temporary hearing loss. Hearing loss can be caused by a variety of factors including age, disease, exposure to loud noises over long/short time periods, genetic disposition, medications, illness, ear wax, bone tumor/growth, etc. (Mayo Clinic Staff). Additionally, humans are born with a set amount of hair cells, and they do not regenerate naturally (Winn 2022).

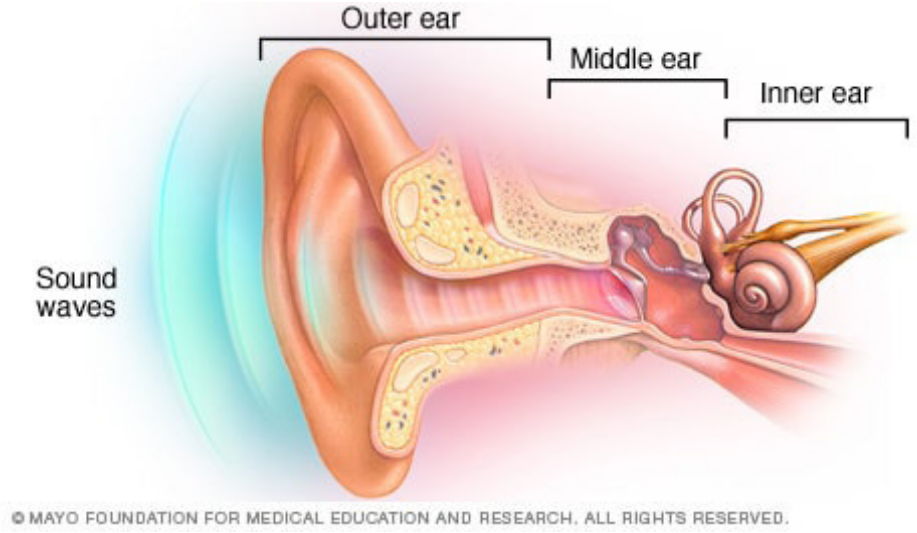


Figure 1. [Anatomical Ear Diagram]. (n.d.). https://www.mayoclinic.org

There are three main types of hardware that can be used to enhance hearing abilities but not regenerate hair cells: hearing aids, cochlear implants, and assistive devices. Hearing aids work by amplifying the sound vibrations entering the ear, but only work if there are some hair cells remaining. There are also implantable hearing aids known as MEI (Middle Ear Implants) or BAHA (Bone Amplified Hearing Aid). They bypass the middle ear structures or move the ear bones directly. In comparison, cochlear implants have an inner and outer ear part. The inner part directly stimulates the auditory nerve, and the outer part converts sounds from the environment into electrical impulses which are transmitted to the inner part. This bypasses the inner ear and damaged hair cells (NIH). Lastly, there are assistive devices including but not limited to telephone amplifiers, flashing/ vibrating alarms, infra-red listening devices, portable sound amplifiers, and audio loop systems all of which serve as aids for hearing loss on a low scale (CDC).

There has also been innovation regarding regenerative therapy. Frequency Therapeutics are running clinical trials to research if cochlea injections, with molecules that stimulate progenitor cells to become hair cells, could be a generalized solution (Winn). Cochlear gene therapy has also been successful in regenerating hair cells and is in the beginning phases of being another treatment option (Santaolalla et al.).

Stem cells are the undifferentiated precursor to all cell types and are found in embryos, bone marrow, and most other human tissue. Specialized cells with different functions, like muscle or skin cells, differentiate from stem cells based on environmental or genetic influence. Examples of such an influence would be surrounding tissue type or patterns of gene expression. Progenitor cells are early descendants of stem cells. They are slightly differentiated from the stem cell and have a smaller range of differentiation, as they are usually unipotent or oligopotent meaning they are only able to differentiate into one type of specialized cell or a specific line of specialized cells, in comparison to stem cells that are usually pluripotent meaning they can differentiate into many types of specialized cells. Progenitor cells function for the repair or replacement of damaged tissue. Stem cells function for the creation of tissues in specific placements to create functioning organs, bones, and blood in the body as well as for repair and fighting diseases in tissue. Stem cells can divide via mitosis without limit given enough nutrients and correct environmental conditions however progenitor cells cannot divide indefinitely. Stem cells are activated through genetic or environmental factors while progenitor cells are usually inactive and will activate from growth factors, or cytokines (Lakna). The limitations of using progenitor cells for hair cell regeneration are its lesser differentiation capabilities in terms of cell division and differentiation as well as the effect of age on the efficiency of the cell (Picerno et al.).

However, with all these solutions to hearing loss, there are some limitations. Cochlear implants do not restore hearing, they can cause hearing to change with reports of electronic or synthetic sounds, they can destroy any residual hearing in the ear, they can just not work, and they can prevent medical tests from being run due to their electrical interference. They are extremely expensive, which sets up an accessibility problem. Hearing aids do not restore hearing, they only amplify sound for patients. They are expensive and are not always accessible (FDA). Gene therapy is effective only case-by-case, can be expensive and is only effective on hereditary hearing loss and not environmental or sound-related hearing loss (NIH).

History

The concepts of modern sensorineural hearing loss (SNHL) treatments date back to the early Renaissance period, with "hearing trumpets," whose horn shape naturally amplified sound for the user (Pickett). In the late 1800's, the capacity to amplify sound with electricity was introduced, leading to the first electric hearing aid called "Akouphone" in 1898 (Traynor). Hearing devices in the 1940's were efficient transistor systems but inconvenient for transport. Over the years, smaller microphones allowed these devices to become smaller, and the first partially digital hearing aids came out in the 1960's. Now, hearing aids are discreet and efficient, equipped with programmable computers to better adapt to individuals (Saperstein). For mild to severe cases, surgically inserted BAHA devices and cochlear implants can help amplify sounds. Though cochlear implants have come a long way from single wire systems in the 1960's to single electrode systems in the 1970's, they can only stimulate nerves to sense more auditory stimuli, which often ends in "crude and distorted patterns" (Hoboken) and don't restore hearing. Genetic therapy including amplifying the expression of genes linked to hair cell regeneration is currently being researched in birds and some mammals but is yet to be applied to humans (Hoboken).

Future Technology

Hearing loss is a constantly expanding area of research; our solution plans to take advantage of a recent breakthrough in knowledge on the process of regenerating bird hair cells and progenitor cells. Our product is ear drops that contain ATOH1 molecule that will help transition progenitor cells into specialized hair cells, therefore mimicking the regenerative properties of the avian auditory system.

Humans are born with roughly 15,000 hair cells and do not have a regenerative auditory system (Winn 2022). This means that as sensorineural damage occurs it is irreversible. However, there is a unique property about cochleae that show promise for regeneration capabilities. In utero humans have an abundance of progenitor cells on the cochlea which become dormant prior to birth. In previous studies with mice, ABCG2 had been an indicator for presence of cochlear stem cells. One study looked for the presence of kin humans as a sign for progenitor cells. Their results were showed one 33-year-old female out of a sample of three had positive results and while the others were inconclusive (Massucci-Bissoli et al., 2017). Methods for identifying presence of progenitor and stem cells is limited and we see in the future, new methods/testing-protocols will be defined.

In contrast to humans, birds have a unique mechanism that allows them to regenerate hair cells after exposure to trauma (loud noises). Supporting cells (progenitor cells) can become hair cells through mitosis or transdifferentiation (specialization of a cell into a specific cell type). Our product takes advantage of the transdifferentiation process in which progenitor cells will transform into a hair cell when exposed to transcription factors such as ATOH1. However, results have shown that ATOH1's effectivity is greatest in younger mammals and decreases in impact for adults. This can be combatted by a decrease/inhibition of notch signaling which promotes hair growth patterns and essentially blocks hair growth in alternating cells (Lewis & Stone, 2019).

We see the future of sensorineural hearing loss being transformed with more general solutions such as ours, which is also non-invasive and applicable to younger and older populations. In avian populations, regeneration was complete after 4 weeks and similar results would be expected with our product. Additionally, we see this being a prescription product that has differing treatment

plans depending on if the patient is trying to restore long term or short-term sensorineural hearing loss. For example, it could be used a few times a month for someone who had was briefly exposed to loud noises or multiple times a week for someone born with deteriorating hearing.

Breakthroughs

There are several necessary breakthroughs for the success of this eardrop technology: One would be the confirmation of the presence of cochlear progenitor or stem cells in all humans as there is very little research affirming this fact. Another breakthrough necessary is the transport of the ATOH1 molecules to the cochlea via a peptide that would facilitate the movement of a phage containing the ATOH1 molecules. The peptide would facilitate the movement through the usually impermeable membrane of the ear drum (Kurabi et al.). However, the combination of this specific molecule transported with a phage has not been documented due to the research thus far being new and only exacted on mice. The next breakthrough necessary is the ability to destroy hair cells obtained from cell line UB-OC2 via soundwaves to imitate the same environmental damage done to hair cells in humans and the reason this most likely doesn't exist is that demand for such a machine is not high in testing as some hair cell damage can be imitated through physical manipulation of the hair cells.

One of the necessary breakthroughs that can be tested is the permeability of the membranes throughout the ear canal. The drops would have to pass through these to get to their destination in the cochlea in the inner ear. A model of the membranes can be created with a simple column with layers of the corresponding cells, then later in an anatomical version, complete with the spiral shape of the cochlea. This can be done with the following investigation, done in a laboratory setting:

1. Obtain human cell samples from each membrane layer between the outer ear and cochlea, including the tympanic and mucous membranes. These can be in the form of the UB-OC2 cell line, kept in the following conditions: MEM with 10% FCS, 50 units/ml y-IFN, L-glutamine, 33° C (Holley).
2. Set up a semi-permeable membrane model with each layer of cells in the order the drops would take from the outer ear to the cochlea.
3. Add a fluorescent dye to a predetermined amount of the ear drop solution and administer to the outer ear portion of the membrane.
4. Record the time it takes in minutes for the solution to pass through the entire model.
5. Record amount of solution lost between each membrane in ml.
6. Depending on the ease with which the solution can pass through the membranes, the concentration may have to be diluted accordingly, and the procedure must be repeated with the new concentration of the solution.

After passing through this testing, the hair cell regeneration progress would have to be tracked regularly to better understand the effectivity and permanence of the treatment. If the molecules can effectively pass through the membranes, the solution must be tested again in mice with transplanted human hair cells.

Design Process

Our team came up with three alternative ideas to the eardrops: permanent injection sites, cochlear grafts, and nano technology.

The idea of permanent injection sites uses the concept of insulin injections for the treatment of diabetes. The idea was that the patient could inject themselves with a dose of progenitor cells on a schedule via a site provided through surgery that removes a small part of the skull so that the injection would directly access the cochlea. This idea was rejected because it is incredibly invasive and continuously injecting a sensitive area could lead to infection or damage to surrounding parts of the cochlea. Our eardrop product would be less invasive, easier and safer application, and would not affect surrounding areas, only the targeted ear parts.

The cochlear graft idea was based on skin grafts being taken from different parts of the body to then regenerate into the new spot. Our idea was to get avian cochlear hair cell grafts to then place into the human cochlea. The avian hair cells regenerate triggered by trauma, and we believed implementing them along with damaged hair cells would stimulate

the dormant human hair cells to become live by increasing their ATOH1 production (Cotanche). Gaining access to the cochlea to place a skin graft would require surgery, possibly even cutting through the temporal bone (Massucci-Bissoli et al., 2017). This is invasive and not generalizable, which is why we rejected the idea.

The idea of nano technology was to implant a device that could be remote controlled or programmed schedule to release ATOH1 molecules on the cochlea. It would be an invasive option and have potential for the body to reject the device. Additionally, nano technology is relatively new for permanent internal applications and there is evidence of nanostructure toxicity which could cause hesitancy in patients (Patra et al.). We rejected this idea because nano technology is a newer idea that could make patients hesitant and having to go through surgery to implant the device would not be preferred or as accessible as ear drops which are a more common concept.

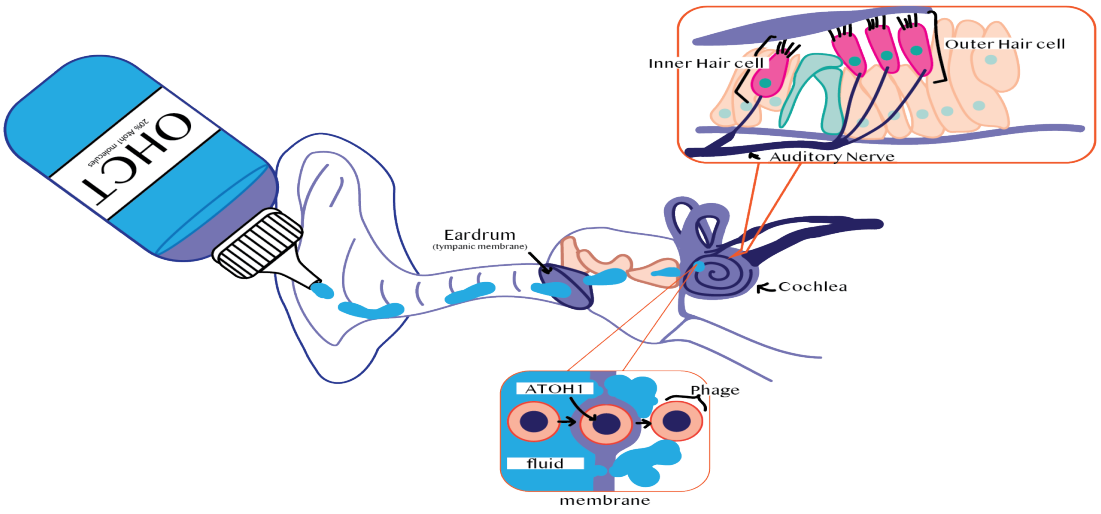
Consequences

There are people in the deaf community who identify with community in the cultural sense. There is "A language, a history, institutions such as schools and clubs, sports, art, and movies" (Cooper, 2019) established within the deaf community. Of those who have embraced the cultural identity and don't desire to have their hearing return, some have found a threat in hearing loss aids such as cochlear implants and might find OHCT to be harmful in the same way. OHCT's goal is not to regenerate hearing for everyone with hearing loss, but to regenerate and restore hearing to those who want it. Another possible negative consequence could be the lack of caution people would have in retaining hearing when introduced with a simple solution. People may not use earplugs or make sure audio levels are at safe decibel levels and end up damaging their cochlear cells beyond regeneration.

However, there are many general positive consequences to the success of our proposed technology. People with sensorineural hearing loss can have less invasive alternatives that can still help improve or restore hearing. The eardrops would not require any surgeries and are a longer-lasting solution that would not require as much maintenance. Administration of the drops can be done by the patient themselves with training, without having to go into a clinic. The drops are meant to be generalizable to anyone who may need them, making them more accessible to the population.

Regenerative medicine is a large area of research and if ATOH1 molecules are proven capable for inducing transdifferentation, this could open a gateway for more parallels between human and bird auditory systems. More specifically, if there was a breakthrough allowing phages to transport inorganic molecules across the tympanic membrane it could prove useful for administration of other forms of medications and provide a less invasive or toxic option to nano technology. Additionally, if it was discovered that progenitor cells are present in adult human patients, the opportunities for exploring other treatment options would expand and demonstrate other methods such as Frequency Therapeutics and would have the capacity to be clinically effective (Massucci-Bissoli et al., 2017).

Conceptual Sketch of Product



Bibliography

Bock, E. (2020, September 3). *NIDCD lab applies gene therapy to hearing loss*. National Institutes of Health. Retrieved January 18, 2023, from https://nihrecord.nih.gov/2020/09/04/nidcd-lab-applies-gene-therapy-hearing-loss

Cortada, M., Levano, S., & Bodmer, D. (2021). mTOR Signaling in the Inner Ear as Potential Target to Treat Hearing Loss. *International Journal of Molecular Sciences*, *22*(12), 6368. https://doi.org/10.3390/ijms22126368

Cotanche, D. A., Lee, K. H., Stone, J. S., & Picard, D. A. (1994). Hair cell regeneration in the bird cochlea following noise damage or ototoxic drug damage. *Anatomy and embryology*, *189*(1), 1–18. https://doi.org/10.1007/BF00193125

Crane, Brent. "How Chickens Could Inspire New Treatments for Deafness." Audubon, 7 Nov. 2019, www.audubon.org/news/how-chickens-could-inspire-new-treatments-deafness. Accessed 12 Dec. 2022.

D, A. (2023). *Structure and Function of Organ of Corti, Auditory Nerve, and Hair Cells*. Study.com. https://study.com/cimages/videopreview/videopreview-full/7.149_110655.jpg

De Jong, W. H., & Borm, P. J. A. (2008). Drug delivery and nanoparticles:applications and hazards. *International Journal of Nanomedicine*, *3*(2), 133–149. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2527668/

Dooling, Robert. "Hair Cell Regeneration - Functional Significance." Grantome, NIH, 1997, grantome.com/grant/NIH/R01-DC001372-04. Accessed 12 Dec. 2022.

Dunleavy, B. P. (2022, December 5). *Hearing loss treatment moves forward despite early disappointment*. UPI. Retrieved December 11, 2022, from https://www.upi.com/Health_News/2022/12/05/hearing-loss-drug-injection/9301669138413/

Eshraghi, A. A., Nazarian, R., Telischi, F. F., Rajguru, S. M., Truy, E., & Gupta, C. (2012, November). The cochlear implant: Historical aspects and future prospects. Retrieved January 18, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4921065/

FDA. (n.d.). *Benefits and risks of cochlear implants*. U.S. Food and Drug Administration. Retrieved January 18, 2023, from https://www.fda.gov/medical-devices/cochlear-implants/benefits-and-risks-cochlear-implants#b

FDA. (n.d.). *Hearing aid benefits and limitations*. U.S. Food and Drug Administration. Retrieved January 18, 2023, from https://www.fda.gov/medical-devices/hearing-aids/hearing-aid-benefits-and-limitations#2

Frequency Therapeutic. (2023). *FX 322 | Regenerative Medicine | Hearing Restoration*. Frequency Therapeutics. https://www.frequencytx.com/

Holley, M. C. (n.d.). *UB-OC2 cell line*. Ximbio. Retrieved January 18, 2023, from https://ximbio.com/reagent/153623/ub-oc2-cell-line#datasheet

Jaslow, R. (2019, December 4). *Reprogramming the Inner Ear to Regrow Hair Cells Shows Promise to be an Effective Target for Hearing Loss Treatments*. Mass Eye and Ear. https://www.masseyeandear.org/news/press-releases/2019/12/reprogramming-the-inner-ear-to-regrow-hair-cells-shows-promise-to-be-an-effective-target-for-hearing-loss-treatments

Lakna. (2017, September 20). *Difference between progenitor cells and stem cells: Definition, characteristics, function, examples*. Pediaa. Com. Retrieved January 18, 2023, from https://pediaa.com/difference-between-progenitor-cells-and-stem-cells/#Progenitor%20Cells

Lee, Y. (2022, May 26). *Increasing our understanding of avian hair cell regeneration* . Hearing Health Foundation. Retrieved December 11, 2022, from https://hearinghealthfoundation.org/blogs/increasing-our-understanding-of-avian-hair-cell-regeneration

Lewis, R., & Stone, J. (2019, December 4). *Journey into the world of hearing*. Cochlea. Retrieved January 18, 2023, from http://www.cochlea.eu/en/research-lines/regeneration

Massucci-Bissoli, M., Lezirovitz, K., Oiticica, J., & Bento, R. F. (2017). Evidence of progenitor cells in the adult human cochlea: sphere formation and identification of ABCG2. *Clinics*, *72*(11), 714–717. https://doi.org/10.6061/clinics/2017(11)11

Matthew, Kelley. "Progress toward Hair Cell Regeneration and the Restoration of Hearing: NIDCD beyond the Lab Series." Videocast. nih.gov, 28 June 2022, videocast.nih.gov/watch=45266. Accessed 12 Dec. 2022.

Mayo Clinic. (2008). *Hearing loss - Symptoms and causes*. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/hearing-loss/symptoms-causes/syc-20373072#:~:text=Aging%20and%20exposure%20to%20loud

Patra, J. K., Das, G., Fraceto, L. F., Campos, E. V. R., Rodriguez-Torres, M. del P., Acosta-Torres, L. S., Diaz-Torres, L. A., Grillo, R., Swamy, M. K., Sharma, S., Habtemariam, S., & Shin, H.-S. (2018). Nano based drug delivery systems: recent developments and future prospects. *Journal of Nanobiotechnology*, *16*(1). https://doi.org/10.1186/s12951-018-0392-8

Picerno, A., Stasi, A., Franzin, R., Curci, C., di Bari, I., Gesualdo, L., & Sallustio, F. (2021, November 26). *Why Stem/progenitor cells lose their regenerative potential*. World journal of stem cells. Retrieved January 18, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8641024/

Reprogramming the inner ear to regrow hair cells shows promise to be an effective target for hearing loss treatments. Mass Eye and Ear. (n.d.). Retrieved December 11, 2022, from https://www.masseyeandear.org/news/press-releases/2019/12/reprogramming-the-inner-ear-to-regrow-hair-cells-shows-promise-to-be-an-effective-target-for-hearing-loss-treatments#:~:text=Hair%20cells%20are%20the%20specialized,then%20relayed%20to%20the%20brain.

Rodríguez, A. (2022, December 8). *Research illuminates approaches for long-term hearing loss treatment*. Baylor College of Medicine Blog Network. https://blogs.bcm.edu/2022/12/08/from-the-labs-research-illuminates-approaches-for-long-term-hearing-loss-treatment/

Russ Pickett, A. D. (2022, September 13). *History of hearing loss treatment*. History of Hearing Loss Treatment. Retrieved January 18, 2023, from https://greentreeaudiology.com/blog/history-of-hearing-loss

Segeritz CP, Vallier L. Cell Culture: Growing Cells as Model Systems In Vitro. Basic Science Methods for Clinical Researchers. 2017:151–72. doi: 10.1016/B978-0-12-803077-6.00009-6. Epub 2017 Apr 7. PMCID: PMC7149418.

Shargorodsky, J. (2023). Sensorineural deafness. Retrieved January 18, 2023, from https://www.mountsinai.org/health-library/symptoms/sensorineural-deafness

Stone, J. S., & Cotanche, D. A. (2007). Hair cell regeneration in the avian auditory epithelium. *The International journal of developmental biology*, *51*(6-7), 633–647. https://doi.org/10.1387/ijdb.072408js

U.S. Department of Health and Human Services. (n.d.). *Line up! to reverse hearing loss, new hair cells need to stand in formation*. National Institute of Deafness and Other Communication Disorders. Retrieved December 11, 2022, from https://www.noisyplanet.nidcd.nih.gov/have-you-heard/new-hair-cells-need-to-stand-in-formation-to-reverse-hearing-loss

Winn, Z. (2022, March 29). *Reversing hearing loss with regenerative therapy*. MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2022/frequency-therapeutics-hearing-regeneration-0329#:~:text=Humans%20are%20born%20with%20about

Xu, S., & Yang, N. (2021). *Research Progress on the Mechanism of Cochlear Hair Cell Regeneration*. Frontiers in Cellular Neuroscience. https://www.frontiersin.org/articles/10.3389/fncel.2021.732507

Zhang, L.-W., Cang, X.-H., Chen, Y., & Guan, M.-X. (2019). *In vitro culture of mammalian inner ear hair cells*. Journal of Zhejiang University. Science. B. Retrieved December 11, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6380999/

The Effect of Vitamin C Treatment on the Growth of Mouse Squamous Cell Carcinoma Cells

Aditi Kalluru

Nikola Tesla STEM High School

Abstract

Latest figures show that more than 15,000 people die of squamous cell carcinoma (skin cancer) in the U.S. each year. Vitamin C has been shown to reduce cancer in many ways. However, there is a lot about the mechanisms used by vitamin C to kill cancer that remain unknown. This study seeks to identify the effects of vitamin C (sodium ascorbate) on 3 mutant cell lines of squamous cell carcinoma, namely CK2, CK4, and CK103 (mouse cells with mutations) and P3C1 (normal mouse cells used as a control) with the hypothesis that when varying concentrations of vitamin C (independent variable) are added to the CK103, CK2, and CK4 cell lines, the growth of the tumor cell lines (dependent variables) will be inhibited. Each cell line was treated with 5 different concentrations of vitamin C ranging from 0.05 millimolar (mM) to 0.5 mM. Untreated cells were used as control. Cell behavior was observed over the course of 4 days and data (amount of cell growth) was collected through a device called Incucyte. IC50 was calculated for each cell line (CK2 had lowest IC50 of 0.04 mM). ANOVA tests showed that growth of each cell line at each concentration of vitamin C was different. Paired T-tests comparing the growth of each cell line to the untreated cells showed with statistical significance that vitamin C stopped or reduced the growth of all tumorous cell lines investigated. It was also found that as the concentration of vitamin C increased, cell growth decreased.

Introduction

This study seeks to identify the effects of vitamin C (sodium ascorbate) on 3 mutant cell lines of squamous cell carcinoma, namely MSCC-CK2 (Hras mu p53wt), MSCC-CK4 (Hras mu p53null), and MSCC-CK103 (Hras wt p53 wt p19null) and P3C1 (non-mutant 3T3 mouse fibroblasts).

Squamous cell carcinoma (SCC) is the second most common form of skin cancer that develops in the epidermis (outermost layer) of the skin. An estimated 1.8 million cases of SCC are diagnosed in the U.S. each year. The latest figures suggest that more than 15,000 people die of squamous cell carcinoma (a type of skin cancer) in the U.S. each year (Skin Cancer Foundation, 2020). It is caused by long exposures to sunlight and tanning beds, both of which involve increased exposure to ultraviolet (UV) radiation, which can lead to DNA damage causing mutations in the RAS family.

The RAS family, which consists of three members, HRAS, KRAS and NRAS, is a family of GTPases that act as molecular switches to regulate cell proliferation, differentiation, apoptosis, and other processes (Moser et al., 2022). Mutations in the members of the RAS family cause cancer, but mutations in HRAS in particular are associated with SCC. RAS and the tumor suppressor gene p53 often co-mutate and cause a more malignant cell state in human cancer. $p19^{Arf}$ is a tumor suppressor gene that, when mutated or deleted, causes SCC. RAS, HRAS p53, and p19Arf play vital roles in the formation of SCC.

Studies have shown that vitamin C can target 3 vulnerabilities that many cancer cells share: redox imbalance, epigenetic reprogramming, and oxygen-sensing regulation (Reczek & Chandel, 2015). In terms of redox imbalance, there are two ways vitamin C uses reactive oxygen species (ROS, also known as free radicals) to promote cell death. Based

on the environment of the cell, vitamin C can act as either an antioxidant and reduce the amount of ROS in the cell and thereby reduce cancer, or act as a pro-oxidant and excessively increase the amount of ROS, causing cell death. It is popularly known that free radicals cause cancer. Although this is true, excessive amounts of ROS can also be detrimental to cancer cells and can cause cell death, depending on the environment and the type of cancer cell.

One study testing the effect of vitamin C on colon cancer showed that the colon cancer cells express more of the glucose transporter 1 (GLUT1). GLUT1 transports DHA (Dehydroascorbic acid, an oxidized form of vitamin C) into the cell. So increased GLUT1 leads to increased amounts of DHA in the cell, leading to an increase in the amount of ROS in cell. This is because DHA reduces intracellular antioxidants and thereby inducing cell death in KRAS and BRAF cells (cells being studied in the experiment). The study also revealed that epigenetic reprogramming in cancer cells includes DNA hypermethylation, which is the process by which tumor suppressor genes are silenced and therefore causing cancer. Vitamin C is known to increase the activity of the proteins that can reduce DNA hypermethylation.

Studies have also shown that vitamin C can cause cell death through the increased production of H2O2 outside the cell. This is because vitamin C acts as an electron donor to O2, creating H2O2. This increased amount of H2O2 in extracellular space can seep into the cell through the process of diffusion and cause cell death (7).

Despite its many known benefits, there is still a lot about the mechanisms that vitamin C uses to fight cancer that remains unknown or poorly understood. For example, how do various concentrations of vitamin C impact growth of cancer cells? How do various types of cancer cells behave with different concentrations of vitamin C? Does vitamin C impact healthy cells? This study seeks to answer these questions by better understanding the mechanisms of vitamin C in fighting cancer by using mouse tumor cells of squamous cell carcinoma as well as its effect on normal cells. Since mutations in the RAS family is known to play vital roles in the formation of SCC, this project uses different mutations (namely HRAS mutant(Q61L);p53 wildtype, HRAS mutant; p53 deleted, and RAS wildtype; p53 wildtype; p19Arf deleted cell lines) of SCC and different concentrations of vitamin C to understand its mechanisms. The hypothesis of this study was that when varying concentrations of sodium ascorbate is added to the P3C1, CK103, CK2, and CK4 cell lines, the growth of the tumor cell lines CK103, CK2, and CK4 will be inhibited, but will have no effect on the P3C1 (non-mutant control) cell line.

Materials and Methods

Making the stock solution of sodium ascorbate: 247 mg of sodium ascorbate powder was mixed with 15 mL of pure water to make a 250 mM stock solution. The stock solution was then filtered to prevent bacterial, yeast, or other microbial growth. After being filtered, the stock solution was transferred to smaller plastic tubes and stored in the freezer.

Cell passaging: DMEM or Dulbecco's Modified Essential Medium was the medium used for this experiment. DMEM and PBS (Phosphate Buffered Saline) were given a warm water bath heated to 37° Celsius before being used on the cells to prevent the cells from getting a temperature shock. The spent medium from the cell culture was aspirated using a sterile pipette. The cells were then rinsed with 10 mL of PBS (Phosphate Buffered Saline), a solution free from salt and magnesium. Using a solution with calcium or magnesium may inhibit trypsin, a solution that helps the cells lift off plate. After rinsing the cells, the PBS solution was aspirated using a sterile pipette. Then, 2.5 mL of trypsin was added using a micro pipettor to help the cells detach from the plate. The trypsin completely covered the cells. The cells were then incubated for 5 minutes at 37° Celsius. After they were taken out of the incubator, the cells were checked under the microscope to make sure they had detached completely from the plate. The plate was tapped on the table to break up any remaining clumps of cells. 4 mL of fresh medium was added and mixed well with repeated pipetting. Repeatedly pipetting warm medium over cells can also help break up clumps. Cells were then transferred to a conical tube and centrifuged at 1000 G for 5 minutes to remove residual trypsin and to make the cells settle down and form a pellet at the bottom. The medium was then removed from the centrifuge tube and discarded without disturbing the pellet using a sterile pipette. 10 mL of fresh medium (10 mL for 400,000 cells) was added to the cells in

the tube using a pipette. Gentle, repeated pipetting will disperse the cells to create a homogenous solution of single cells. From this, 4 dilutions were made (1:5, 1:10, 1:20, 1:40; the original dilution was 1:5) to determine the optimal seeding density.

Cell counting: The spent medium from the cell culture was aspirated using a sterile pipette. The cells were then rinsed with 10 mL of PBS. After rinsing the cells, the PBS solution was aspirated using a sterile pipette. Once PBS solution is aspirated off the cells, 2.5 mL of trypsin was added to help the cells detach from the plate. The trypsin completely covered the cells. The cells were then incubated for 5 minutes at 37° Celsius. After they were taken out of the incubator, the cells were checked under the microscope to make sure they had detached completely from the plate. The plate was tapped on the table to break up any remaining clumps of cells. 4 mL of fresh medium was added and mixed well with repeated pipetting. A 1:1 dilution was prepared by mixing 10 µL of cells and 10 µL of trypan blue. 10 µL of the trypan blue – cell mixture was then transferred to one side of the cell counting chamber. This was then inserted into the hemocytometer to count the cells. To count the cells manually, the 10 µL of the trypan blue – cell mixture was inserted into the Neubauer chamber. The Neubauer chamber has a grid that allows for the counting of the cells under the microscope (Figure 1.A).

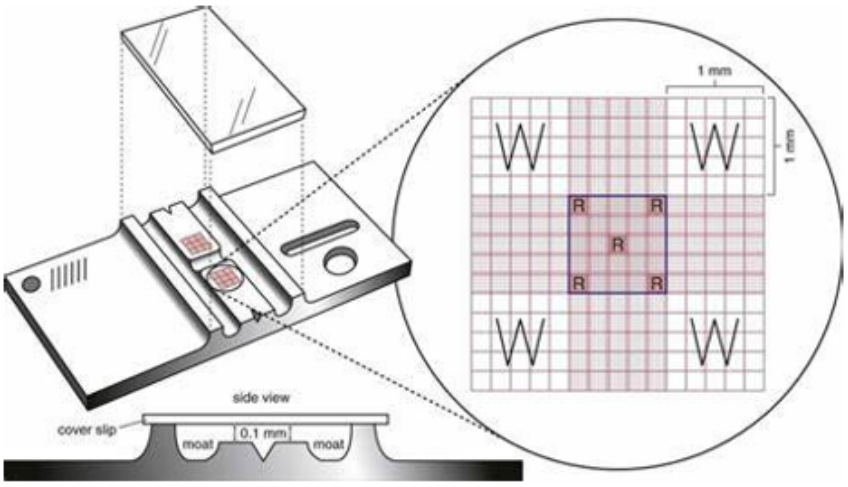


Figure 1.A: Neubauer chamber under the microscope. Only cells present in the 4 quadrants labeled with a 'W' are counted.

Clark, K; Hippel, T (Photographers). (2016, July, United States of America. Oncohema Key)

$$\text{Concentration of cells (cells/mL)} = \frac{(\text{sum of cells in 4 quadrants}) \times (2 \times 10^4)}{4 \text{ (squares)}}$$

Total # of cells = concentration of cells (cells/mL) x volume of sample (mL)

Plating the cells: Based on the results from the dilutions created to test the optimal seeding density, the 1: 10 dilution (20,000 cells/mL; 10,000 cells for 0.5 mL) was found to be the best suited for all 4 cell lines (P3C1 - the control, CK103, CK2, CK4). The 250 mM stock solution of sodium ascorbate was diluted 1:5 in media to get a 50 mM stock. Based on previous research, it was determined that the suitable range of concentration of sodium ascorbate was 0.00 mM, 0.05 mM, 0.10 mM, 0.20 mM, 0.30 mM, 0.40 mM, and 0.50 mM. A 24-welled plate was used for each of the four cell lines, in which each of the above concentrations occupied 3 wells. There were 3 replicates for each concentration of sodium ascorbate. Each cell line had a plate dedicated to it. Hence, there were 4 plates in total. A 24 -welled plate was used to keep all replicates on the same plate. Cells were washed and counted before being diluted and pipetted into the wells. In

previous experiments, 12-welled plates were used and 1mL of a cell-medium mixture was pipetted into each well. However, the wells of the 24-welled plate were smaller than that of the 12 welled plate. Hence, only ½ mL of the cell-medium mixture was pipetted. This equates to 10,000 cells per well. The last three wells of each of the four 24-welled plates were pipetted with medium only to serve as a negative imaging control (denoted as "blank" in Figure 1.B). The 4 plates (Figure 1.B) were then placed in the Incucyte for imaging at 8-hour intervals over the course of 4 days. Data (Table 1.A, Table 1.B, Table 1.C, Table 1.D) collected by the Incucyte was exported to Excel for further analysis.

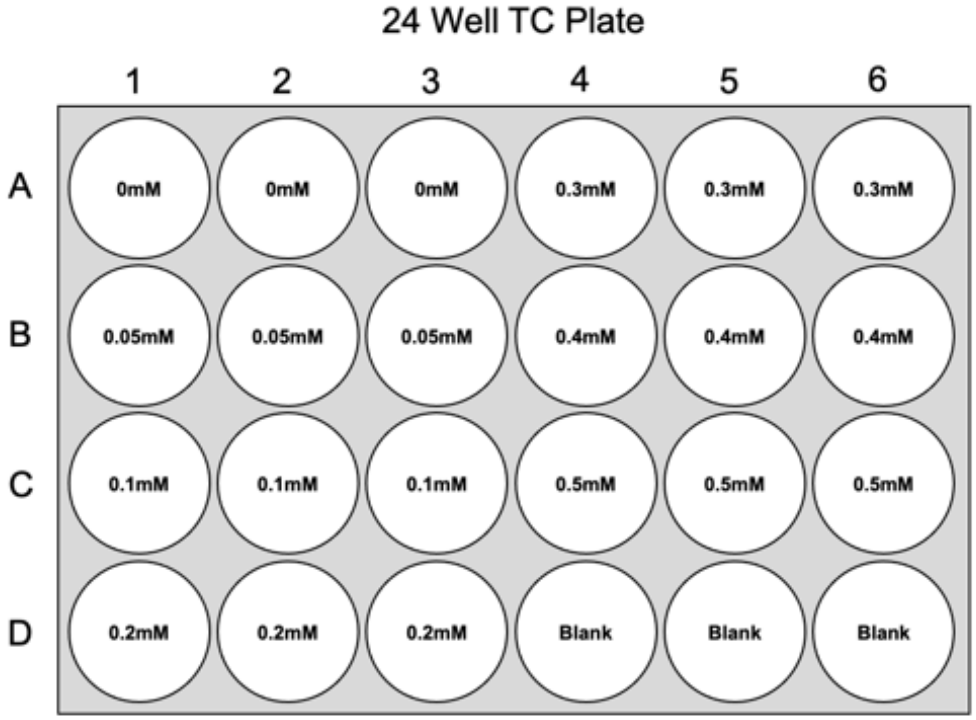


Figure 1.B: Diagram of the plating of cells for each of the four cell lines

Results

Images captured by the Incucyte -

[1]GTP stands for guanosine triphosphate

[2]The HRAS gene provides instructions for making a protein called H-Ras that is involved primarily in regulating cell division

[3]tumor suppressor mutation

[4]tumor suppressor mutation

[5]Hemocytometer – a machine that counts cells.

[6]Optimal seeding density – the number of cells that can be in a well without overcrowding.

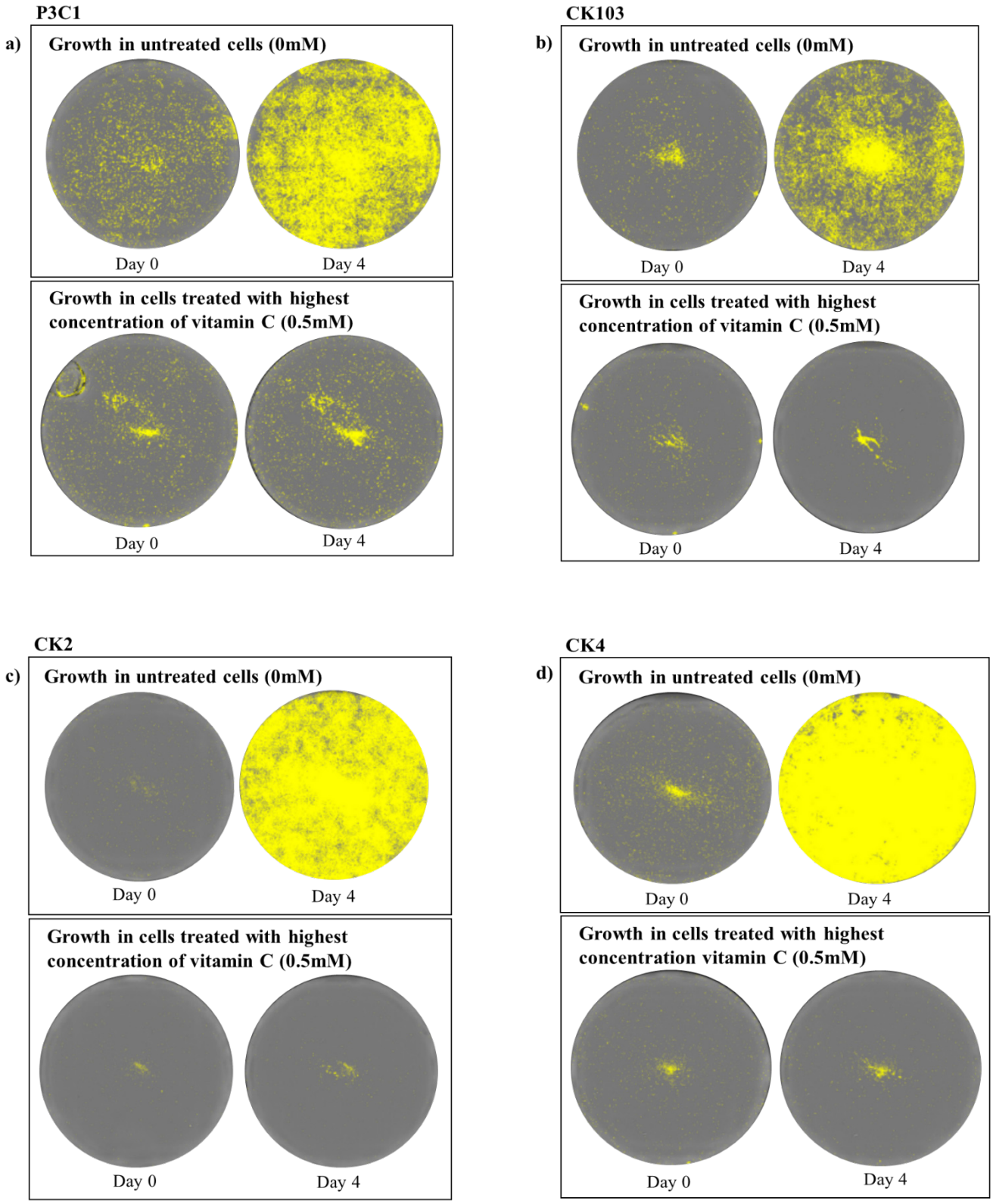


Figure 2 a, b, c, d: The images captured by the Incucyte show the percent confluency (number of cells in the well) of each cell line for the untreated cells (0mM vitamin C) and the highest concentration (0.5mM vitamin C) at day zero and day four. These images reveal that day four, the untreated cells have grown exponentially and almost cover the entire cell well. However, the cells dosed with the highest concentration of vitamin C (0.5mM) show little to no growth since the images captured on day zero and day four are almost identical.

Vessel Name: AK CK103 123022

Date Time	Elapsed	0 mM	0.05 mM	0.1 mM	0.2 mM	0.3 mM	0.4 mM	0.5 mM	**BLANK**
12/30/2022 7:21:00 PM	0	3.208142	3.547223	2.480882	2.459187	2.226799	1.852285	1.794703	0.08817902
12/31/2022 3:21:00 AM	8	4.649767	5.033744	2.329737	2.045528	1.756302	1.539926	1.639126	0.1233372
12/31/2022 11:21:00 AM	16	6.252533	6.562821	2.76963	2.026215	1.628974	1.469335	1.528004	0.09559392
12/31/2022 7:21:00 PM	24	7.656613	8.121145	3.258162	2.181365	1.747083	1.422952	1.373733	0.08876801
1/1/2023 3:21:00 AM	32	9.492157	10.18647	3.639633	2.272979	1.659203	1.363456	1.307927	0.06686168
1/1/2023 11:21:00 AM	40	11.69038	12.23243	4.411011	2.533166	1.608235	1.322509	1.245868	0.05564042
1/1/2023 7:21:00 PM	48	14.11417	14.62014	5.34555	2.964402	1.636129	1.271567	1.219061	0.05294609
1/2/2023 3:21:00 AM	56	16.60349	16.6105	6.472707	3.454375	1.678086	1.258487	1.146186	0.05973687
1/2/2023 11:21:00 AM	64	19.84693	19.18061	7.550889	3.975343	1.760405	1.20454	1.136701	0.04550919
1/2/2023 7:21:00 PM	72	23.12027	22.50371	8.674582	4.615009	1.855174	1.17607	1.101244	0.03978209
1/3/2023 3:21:00 AM	80	26.22459	24.59097	10.63874	5.456099	2.070081	1.19	1.096942	0.05414162
1/3/2023 11:21:00 AM	88	29.23216	27.252	11.77392	6.327923	2.205323	1.181433	1.078011	0.04198415
1/3/2023 7:21:00 PM	96	33.22675	30.85005	13.83816	7.176198	2.448112	1.168343	1.066886	0.06210596
1/4/2023 3:21:00 AM	104	36.64915	32.70643	15.20712	8.257142	2.709261	1.16931	1.046452	0.04657288

Table 1.A: Shows the data collected by the Incucyte for the CK103 cell line. Note that only one of the three replicates for each of the sodium ascorbate concentrations is listed in this table.

Vessel Name: AK P3C1 123022

Date Time	Elapsed	0 mM	0.05 mM	0.1 mM	0.2 mM	0.3 mM	0.4 mM	0.5 mM	**BLANK**
12/30/2022 7:59:00 PM	0	9.257054	9.939314	5.209105	4.416124	5.062003	5.002151	5.397478	0.3812417
12/31/2022 3:59:00 AM	8	14.24285	14.85293	5.822347	5.23246	6.001578	5.92771	6.85007	0.329535
12/31/2022 11:59:00 AM	16	19.2014	20.25772	7.761957	6.556591	8.136018	8.009401	8.500852	0.293511
12/31/2022 7:59:00 PM	24	23.89273	21.70515	9.328972	7.945562	9.64803	9.376909	9.639729	0.7208642
1/1/2023 3:59:00 AM	32	28.85117	26.80554	10.59	7.94508	10.0147	9.643949	10.34674	0.9407271
1/1/2023 11:59:00 AM	40	32.25	28.9238	10.62617	7.492489	9.544491	8.564187	9.484861	1.022111
1/1/2023 7:59:00 PM	48	34.68029	32.64593	10.45404	7.254945	8.592954	7.840842	8.91279	0.7416056
1/2/2023 3:59:00 AM	56	41.11228	34.91528	11.65833	6.993068	9.052424	7.933961	8.650798	0.9635257
1/2/2023 11:59:00 AM	64	46.51301	41.51875	11.75861	6.851184	8.506689	7.615226	8.133229	1.058719
1/2/2023 7:59:00 PM	72	50.47283	44.8361	10.94002	6.579624	7.993414	7.929834	7.992685	0.8231609
1/3/2023 3:59:00 AM	80	57.02773	49.53218	11.48945	6.178617	7.712223	7.386652	8.224862	0.962752
1/3/2023 11:59:00 AM	88	59.49795	51.39554	10.19793	6.218856	7.987832	7.607633	8.46981	0.8945806
1/3/2023 7:59:00 PM	96	64.56669	55.51294	11.00704	5.897332	8.234968	7.802373	8.08497	0.7287275
1/4/2023 3:59:00 AM	104	68.36809	59.81666	10.19027	6.012927	8.212517	7.336018	8.272265	0.9573852

Table 1.B: Shows the data collected by the Incucyte for the P3C1 cell line. Note that only one of the three replicates for each of the sodium ascorbate concentrations is listed in this table.

Vessel Name: AK CK2 123022

Date Time	Elapsed	0 mM	0.05 mM	0.1 mM	0.2 mM	0.3 mM	0.4 mM	0.5 mM	**BLANK**
12/30/2022 6:03:00 PM	0	0.6468552	0.5727066	0.4509341	0.3062398	0.5158527	0.2397035	0.4534392	0.4170725
12/31/2022 2:03:00 AM	8	2.028111	0.4941445	0.1790875	0.2107383	0.3248541	0.1851971	0.6532422	0.330779
12/31/2022 10:03:00 AM	16	3.01753	0.5702233	0.1961765	0.2200343	0.3197688	0.2037235	0.636601	0.311633
12/31/2022 6:03:00 PM	24	3.933177	0.6663756	0.2212255	0.1733429	0.3592564	0.2278053	0.3041786	0.1857333
1/1/2023 2:03:00 AM	32	5.72016	0.8383813	0.2116392	0.1979961	0.3906565	0.2694158	0.3310162	0.1751581
1/1/2023 10:03:00 AM	40	8.114723	1.04388	0.1936625	0.190256	0.391061	0.2599352	0.3295043	0.1854344
1/1/2023 6:03:00 PM	48	11.73516	1.373109	0.2004884	0.197627	0.3862436	0.2798416	0.3463298	0.1770569
1/2/2023 2:03:00 AM	56	17.7388	1.851415	0.17071	0.1991259	0.37024	0.2830149	0.3370819	0.1689607
1/2/2023 10:03:00 AM	64	26.98645	2.47975	0.1700508	0.1855442	0.3756816	0.2603219	0.3397674	0.155722
1/2/2023 6:03:00 PM	72	39.5799	3.327795	0.1640862	0.2082242	0.3255837	0.2654073	0.3249289	0.1854212
1/3/2023 2:03:00 AM	80	56.49169	4.443132	0.1569703	0.1920317	0.3356929	0.2930099	0.3296494	0.1569043
1/3/2023 10:03:00 AM	88	73.65044	5.929448	0.1709475	0.1894781	0.3190304	0.2705893	0.3318075	0.1646006
1/3/2023 6:03:00 PM	96	86.65119	8.023535	0.1695453	0.1820455	0.3220981	0.2653547	0.3366471	0.1745999
1/4/2023 2:03:00 AM	104	94.23405	10.8553	0.1661785	0.1882167	0.314556	0.2520588	0.3218696	0.1635237
1/4/2023 10:03:00 AM	112	97.38647	14.59773	0.1752372	0.1732682	0.3151494	0.2512632	0.3243046	0.1735186

Table 1.C: Shows the data collected by the Incucyte for the CK2 cell line. Note that only one of the three replicates for each of the sodium ascorbate concentrations is listed in this table.

Vessel Name: AK CK4 123022

Date Time	Elapsed	0 mM	0.05 mM	0.1 mM	0.2 mM	0.3 mM	0.4 mM	0.5 mM	**BLANK**
12/30/2022 6:45:00 PM	0	3.538114	3.535718	2.164097	1.564506	1.791471	1.530737	1.312027	0.0982004
12/31/2022 2:45:00 AM	8	8.407809	7.90273	2.652611	0.8403594	1.15696	1.087636	1.025363	0.1182431
12/31/2022 10:45:00 AM	16	11.50242	10.75658	3.530194	0.8499382	1.172945	1.089404	1.012494	0.109804
12/31/2022 6:45:00 PM	24	14.95074	14.20162	4.429863	1.033152	1.425106	1.178624	1.047454	0.0773314
1/1/2023 2:45:00 AM	32	21.42971	19.76999	5.491057	1.033196	1.550077	1.178418	1.065753	0.055702
1/1/2023 10:45:00 AM	40	29.34825	26.72084	6.91847	1.044356	1.463225	1.154103	1.01856	0.0805971
1/1/2023 6:45:00 PM	48	38.33307	34.69322	8.573271	1.060742	1.384318	1.108787	0.980633	0.0548844
1/2/2023 2:45:00 AM	56	49.76555	45.78113	10.58948	1.080331	1.395696	1.119379	0.9546075	0.0593808
1/2/2023 10:45:00 AM	64	62.84402	58.11563	13.74534	1.158973	1.45706	1.02979	0.9294488	0.0651299
1/2/2023 6:45:00 PM	72	75.82278	71.96132	17.27341	1.194688	1.373689	1.022771	0.930561	0.0620532
1/3/2023 2:45:00 AM	80	86.51682	84.82169	21.96839	1.312321	1.378778	1.014143	0.9085752	0.0571085
1/3/2023 10:45:00 AM	88	93.98927	93.57334	27.33498	1.40821	1.472095	1.009681	0.8867569	0.0599435
1/3/2023 6:45:00 PM	96	96.87081	97.45972	34.31768	1.58647	1.497259	0.9979367	0.8827047	0.0647783
1/4/2023 2:45:00 AM	104	97.63957	98.60181	41.85439	1.874263	1.598897	0.9778364	0.8390412	0.0733536
1/4/2023 10:45:00 AM	112	98.15648	98.73965			1.587543	0.9382002	0.8409489	0.0837529

Table 1.D: Shows the data collected by the Incucyte for the CK4 cell line. Note that only one of the three replicates for each of the sodium ascorbate concentrations is listed in this table.

Data Analysis

The data from the Incucyte (shown in **Tables 1.A – 1.D**) was first graphed to observe the general trends in growth, through which it was observed that: (1) as the concentration of vitamin C increased, cell growth decreased. (2) CK2 (Hras mu p53wt) displayed the most sensitivity to vitamin C whereas (3) P3C1 (non-mutant 3T3 mouse fibroblasts) displayed the least sensitivity to vitamin C. However, to prove the above observations and the hypothesis that vitamin C will inhibit the growth of the tumorous cell lines CK103, CK2 and CK4, statistical analyses were conducted.

ANOVA tests were done to see if there was a statistically significant difference between the average cell growth at each concentration of vitamin C for each cell line. A P-value greater than 0.05 showed that there was a statistically significant difference between the average cell growths at each concentration of vitamin C at each cell line. Paired T-tests were conducted, comparing the mean percent confluency (average cell growth) at each concentration of vitamin C to the mean percent confluency of the untreated cells. All concentrations of vitamin C had a p-value less than 0.05 except the 0.05mM concentration. This was true for all cell lines except CK2. This meant that all concentrations of vitamin C caused a statistically significant reduction in cell growth except 0.05mM, for all cell lines except the CK2 cell line. Because CK2 showed statistically significant sensitivity to all concentrations of vitamin C including 0.05mM, it was concluded that CK2 was the most sensitive to vitamin C. Detailed observations and data analysis for each cell line can be found below.

CK103 (Hras wt p53 wt p19 null)

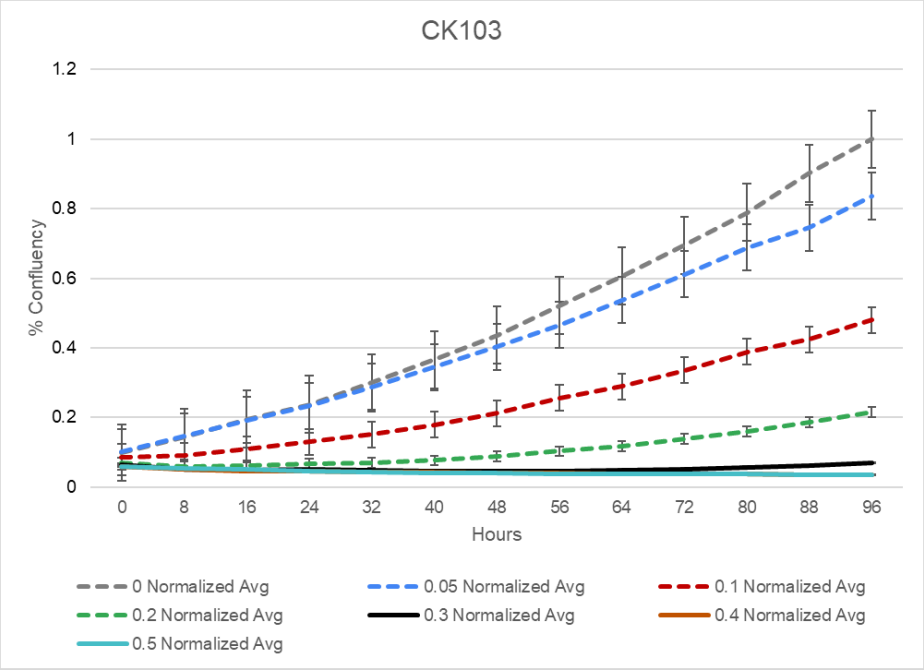


Figure 3.A: *Summary of trends of various concentrations* - As evident in the adjacent graph, percent confluency at treatment below 0.3 mM remained stable or decreased. The percent confluency at concentrations 0.05 – 0.2 mM increased; however, they were still lesser than the increase in the percent confluency of the untreated cells.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	***Average***	*Variance*
0 Normalized Avg	13	6.29118211	**0.483937**	0.08770199
0.05 Normalized Avg	13	5.59373514	**0.430287**	0.05741387
0.1 Normalized Avg	13	3.13751001	**0.241347**	0.01778053
0.2 Normalized Avg	13	1.41771242	**0.109055**	0.00265639
0.3 Normalized Avg	13	0.70088902	**0.053915**	5.7702E-05
0.4 Normalized Avg	13	0.55498473	**0.042691**	4.7341E-05
0.5 Normalized Avg	13	0.55718567	**0.04286**	6.1852E-05

ANOVA: single factor (CK103)

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	2.78733788	6	0.46455631	19.6228626	**3.47753E-14**	2.20855381
Within Groups	1.98863597	84	0.02367424			
Total	4.77597385	90				

Figure 3.B: *ANOVA* test for CK103 shows that the means are different (Low P value highlighted in green). So, paired comparison T-tests were performed using the 0 mM vitamin C concentration as a baseline to statistically validate whether each concentration of Vitamin C had an impact on confluency.

Concentration	**% Change in confluency after 96 hrs**
0 mM	908.04%
0.05 mM	725.41%
0.1 mM	456.50%
0.2 mM	203.48%
0.3 mM	6.66%
0.4 mM	-40.02%
0.5mM	-41.12%

Figure 3.C: Final change in % confluency - The above table shows the change in percent confluency after 96 hrs (4 days).

t-Test: Two-Sample Assuming Equal Variances: CK103

	0 Avg	*0.05 Avg*
Mean	**0.483937086**	**0.430287318**
Variance	0.087701985	0.057413868
Observations	13	13
Pooled Variance	0.072557927	
Hypothesized Mean Difference	0	
df	24	
t Stat	0.507787599	
P(T<=t) one-tail	**0.31**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.616237479	
t Critical two-tail	2.063898562	

	0 Avg	*0.1 Avg*
Mean	**0.483937086**	**0.241346924**
Variance	0.087701985	0.017780529
Observations	13	13
Pooled Variance	0.052741257	
Hypothesized Mean Difference	0	
df	24	
t Stat	2.693113291	
P(T<=t) one-tail	**0.01**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.012705626	
t Critical two-tail	2.063898562	

	0 Avg	*0.4 Avg*
Mean	**0.483937086**	**0.042691133**
Variance	0.087701985	4.73405E-05
Observations	13	13
Pooled Variance	0.043874663	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.370695333	
P(T<=t) one-tail	**8.15E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	1.63055E-05	
t Critical two-tail	2.063898562	

	0 Avg	*0.3 Avg*
Mean	0.483937086	0.05391454
Variance	0.087701985	5.7702E-05
Observations	13	13
Pooled Variance	0.043879844	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.233778851	
P(T<=t) one-tail	**1.15E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	2.30169E-05	
t Critical two-tail	2.063898562	

	0 Avg	*0.2 Avg*
Mean	0.483937086	0.109054801
Variance	0.087701985	0.002656387
Observations	13	13
Pooled Variance	0.045179186	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.49658073	
P(T<=t) one-tail	**7.47E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.000149403	
t Critical two-tail	2.063898562	

	0 Avg	*0.5 Avg*
Mean	**0.483937086**	**0.042860436**
Variance	0.087701985	6.1852E-05
Observations	13	13
Pooled Variance	0.043881919	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.368190772	
P(T<=t) one-tail	**8.20E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	1.64084E-05	
t Critical two-tail	2.063898562	

Figure 3.D: *T test summary* - The mean percent confluency of CK103 cells at each concentration of vitamin C were compared the percent confluency of CK103 cell line at 0 mM vitamin C with a t-test. A P-value less than 0.05 for all concentrations except 0.05 mM shows that the mean confluency for all concentrations of Vitamin C is significantly lower than the percent confluency at 0mM vitamin C. With the 0.05 concentration of vitamin C, however P>0.05. Therefore, the null hypothesis for 0.05 mM vitamin C (the mean of the percent confluency at 0 mM is different from the mean of the percent confluency at 0.05 mM) cannot be rejected, indicating that all concentrations except 0.05 mM had a statistically significant effect on the percent confluency of the CK103 cell line.

CK2 (Hras mu p53 wt)

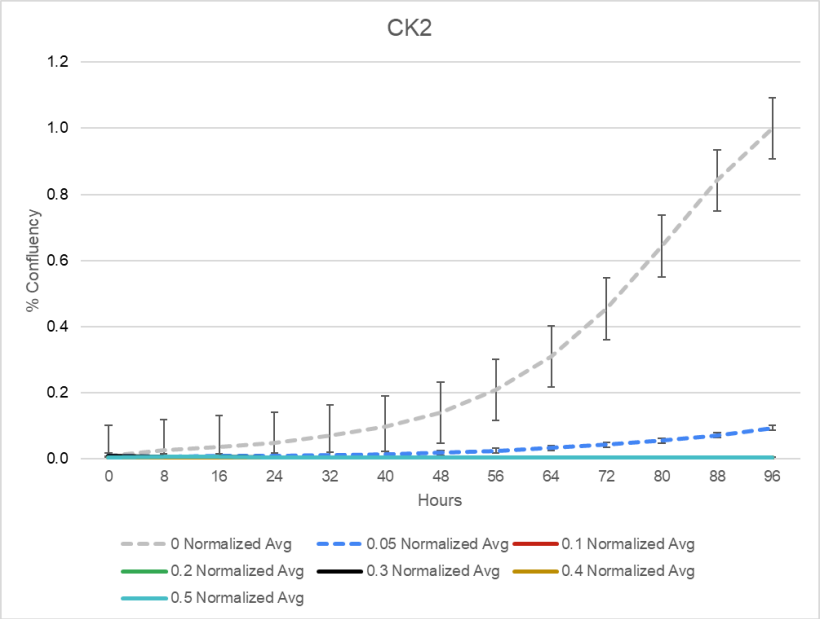


Figure 4.A: Summary of trends of various concentrations - As seen in the adjacent graph, Vitamin C did have an impact on the CK2 cell line at concentrations below 0.05 mM. The percent confluency for 0.1, 0.2, 0.3, 0.4 and 0.5 was either stable or decreased slightly. At 0.05 mM, the percent confluency increased from 0.008 to 0.092. However, this increase was not as significant as the increase in the percent confluency in the cells that were not treated with vitamin C.

Anova: Single Factor (CK2)

SUMMARY

Groups	*Count*	*Sum*	***Average***	*Variance*
0 Normalized Avg	13	3.875257829	**0.298096756**	0.112439131
0.05 Normalized Avg	13	0.387090511	**0.029776193**	0.000760466
0.1 Normalized Avg	13	0.035627502	**0.002740577**	9.17004E-07
0.2 Normalized Avg	13	0.033063752	**0.002543366**	3.48714E-07
0.3 Normalized Avg	13	0.049885883	**0.003837376**	1.7114E-06
0.4 Normalized Avg	13	0.043019567	**0.003309197**	7.15107E-08
0.5 Normalized Avg	13	0.049997486	**0.00384596**	5.2218E-08

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	0.947477861	6	0.157912977	9.76470397	**3.80611E-08**	2.208553806
Within Groups	1.35843238	84	0.016171814			
Total	2.305910241	90				

Figure 4.B: ANOVA test for CK103 shows that the means are different (Low P value highlighted in green). So, paired comparison T-tests were performed using the 0 concentration as a baseline to statistically validate whether each concentration of Vitamin C had an impact on confluency.

Concentration	% Change in confluency after 96 hrs
0 mM	11535.66%
0.05 mM	1037.26%
0.1 mM	-56.28%
0.2 mM	-50.88%
0.3 mM	-59.13%
0.4 mM	1.71%
0.5mM	-1.27%

Figure 3.C: *Final change in % confluency* - The above table shows the change in percent confluency after 96 hrs (4 days).

t-Test: Two-Sample Assuming Equal Variances: CK2

	0 Avg	*0.05 Avg*
Mean	**0.298096756**	**0.029776193**
Variance	0.112439131	0.000760466
Observations	13	13
Pooled Variance	0.056599799	
Hypothesized Mean Difference	0	
df	24	
t Stat	2.875432594	
P(T<=t) one-tail	**0.0042**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.00832673	
t Critical two-tail	2.063898562	

	0 Avg	*0.1 Avg*
Mean	**0.298096756**	**0.002740577**
Variance	0.112439131	9.17004E-07
Observations	13	13
Pooled Variance	0.056220024	
Hypothesized Mean Difference	0	
df	24	
t Stat	3.175829847	
P(T<=t) one-tail	**0.002**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.004071879	
t Critical two-tail	2.063898562	

	0 Avg	*0.2 Avg*
Mean	**0.298096756**	**0.002543366**
Variance	0.112439131	3.48714E-07
Observations	13	13
Pooled Variance	0.05621974	
Hypothesized Mean Difference	0	
df	24	
t Stat	3.177958404	
P(T<=t) one-tail	**0.002**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.004050991	
t Critical two-tail	2.063898562	

	0 Avg	*0.3 Avg*
Mean	**0.298096756**	**0.003837376**
Variance	0.112439131	1.7114E-06
Observations	13	13
Pooled Variance	0.056220421	
Hypothesized Mean Difference	0	
df	24	
t Stat	3.164025297	
P(T<=t) one-tail	**0.0021**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.004189613	
t Critical two-tail	2.063898562	

	0 Avg	*0.4 Avg*
Mean	**0.298096756**	**0.003309197**
Variance	0.112439131	7.15107E-08
Observations	13	13
Pooled Variance	0.056219601	
Hypothesized Mean Difference	0	
df	24	
t Stat	3.16972765	
P(T<=t) one-tail	**0.0021**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.004132336	
t Critical two-tail	2.063898562	

	0 Avg	*0.5 Avg*
Mean	**0.298096756**	**0.00384596**
Variance	0.112439131	5.2218E-08
Observations	13	13
Pooled Variance	0.056219592	
Hypothesized Mean Difference	0	
df	24	
t Stat	3.163956333	
P(T<=t) one-tail	**0.0021**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.004190311	
t Critical two-tail	2.063898562	

Figure 4.D: *ANOVA* The mean percent confluency of CK2 cells at each concentration of vitamin C were compared the percent confluency of CK2 cells at 0 mM vitamin C with a t-test. A P-value less than 0.05 for all concentrations shows that the mean confluency for all concentrations of Vitamin C are significantly lower than the percent confluency at 0mM vitamin c, indicating that the all concentrations had a statistically significant effect on the percent confluency of the CK2 cell line.

CK4 (Hras mu p53 null)

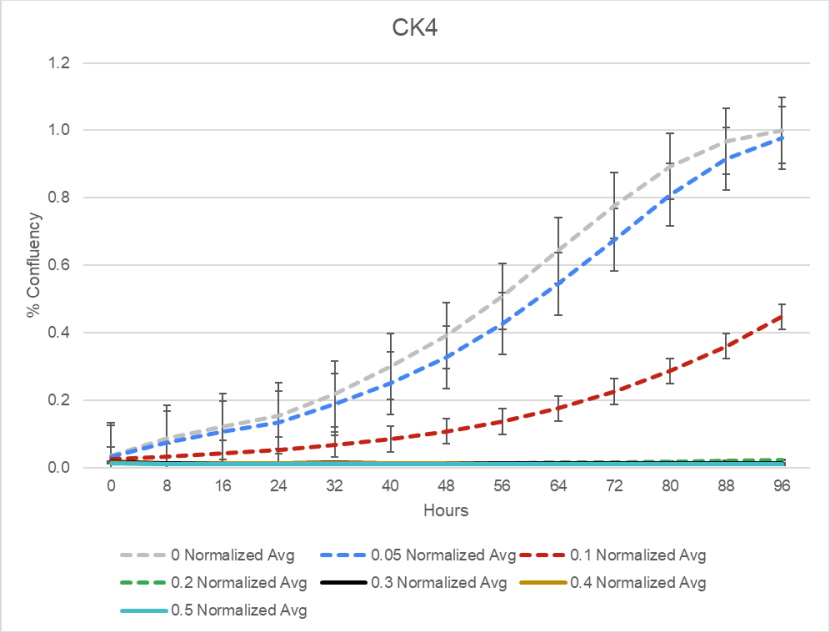


Figure 5.A: *Summary of trends of various concentrations* - As seen in the adacent graph, all concentrations of vitamin C below 0.1 mM caused either a stabilization or slight decrease in in the percent confluency of CK cells. At 0.05 mM and 0.1 mM concentrations, the percent confluency of CK4 cells did increase, but this was still lesser compared to the untreated cells.

SUMMARY

Groups	*Count*	*Sum*	***Average***	*Variance*
0 Normalized Avg	13	6.0914499 3	**0.468573**	0.124203331
0.05 Normalized Avg	13	5.4588014 9	**0.419908**	0.110895702
0.1 Normalized Avg	13	2.0423976 7	**0.157108**	0.018369107
0.2 Normalized Avg	13	0.1930147 8	**0.014847**	1.40571E-05
0.3 Normalized Avg	13	0.1700206 8	**0.013079**	2.80521E-06
0.4 Normalized Avg	13	0.1475023	**0.011346**	1.95972E-06
0.5 Normalized Avg	13	0.1300336	**0.010003**	1.42176E-06

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	3.2492184 3	6	0.5415364	14.95435318	**1.47E-11**	2.2085538 1
Within Groups	3.0418606 2	84	0.0362126 3			
Total	6.2910790 4	90				

Figure 5.B: *ANOVA* test for CK103 shows that the means are different (Low P value highlighted in green). So, paired comparison T-tests were performed using the 0 concentration as a baseline to statistically validate whether each concentration of Vitamin C had an impact on confluency.

Concentration	% Change in confluency after 96 hrs
0 mM	2716.05%
0.05 mM	2881.34%
0.1 mM	1771.31%
0.2 mM	40.83%
0.3 mM	-31.07%
0.4 mM	-34.41%
0.5mM	-35.01%

Figure 4.C: *Final change in % confluency* - The above table shows the change in percent confluency after 96 hrs (4 days).

t-Test: Two-Sample Assuming Equal Variances: CK4

	0 Avg	*0.05 Avg*
Mean	**0.468573071**	**0.419907807**
Variance	0.124203331	0.110895702
Observations	13	13
Pooled Variance	0.117549517	
Hypothesized Mean Difference	0	
df	24	
t Stat	0.361880635	
P(T<=t) one-tail	**3.6E-01**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.720608019	
t Critical two-tail	2.063898562	

	0 Avg	*0.1 Avg*
Mean	**0.468573071**	**0.157107513**
Variance	0.124203331	0.018369107
Observations	13	13
Pooled Variance	0.071286219	
Hypothesized Mean Difference	0	
df	24	
t Stat	2.974157179	
P(T<=t) one-tail	**3.3E-03**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.006598176	
t Critical two-tail	2.063898562	

	0 Avg	*0.2 Avg*
Mean	**0.468573071**	**0.01484729**
Variance	0.124203331	1.40571E-05
Observations	13	13
Pooled Variance	0.062108694	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.641666525	
P(T<=t) one-tail	**5.2E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.000103287	
t Critical two-tail	2.063898562	

	0 Avg	*0.3 Avg*
Mean	**0.468573071**	**0.013078514**
Variance	0.124203331	2.80521E-06
Observations	13	13
Pooled Variance	0.062103068	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.659972371	
P(T<=t) one-tail	**4.9E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	9.85878E-05	
t Critical two-tail	2.063898562	

	0 Avg	*0.4 Avg*
Mean	**0.468573071**	**0.01134633**
Variance	0.124203331	1.95972E-06
Observations	13	13
Pooled Variance	0.062102645	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.67770953	
P(T<=t) one-tail	**4.7E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	9.42388E-05	
t Critical two-tail	2.063898562	

	0 Avg	*0.5 Avg*
Mean	**0.468573071**	**0.010002584**
Variance	0.124203331	1.42176E-06
Observations	13	13
Pooled Variance	0.062102377	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.691467034	
P(T<=t) one-tail	**4.5E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	9.09984E-05	
t Critical two-tail	2.063898562	

Figure 5.D: The mean percent confluency of CK4 cells at each concentration of vitamin C were compared the percent confluency of CK4 cell line at 0 mM vitamin C with a t-test. A P-value less than 0.05 for all concentrations except 0.05 mM shows that the mean confluency for all concentrations of Vitamin C are significantly lower than the percent confluency at 0mM vitamin C. With a 0.05 concentration of vitamin C, however P>0.05. Therefore, the null hypothesis for 0.05 mM vitamin C (the mean of the percent confluency at 0 mM is different from the mean of the percent confluency at 0.05 mM) cannot be rejected, indicating that all concentrations except 0.05 mM had a statistically significant effect on the percent confluency of the CK4 cell line.

P3C1 (non-mutant 3T3 mouse fibroblasts)

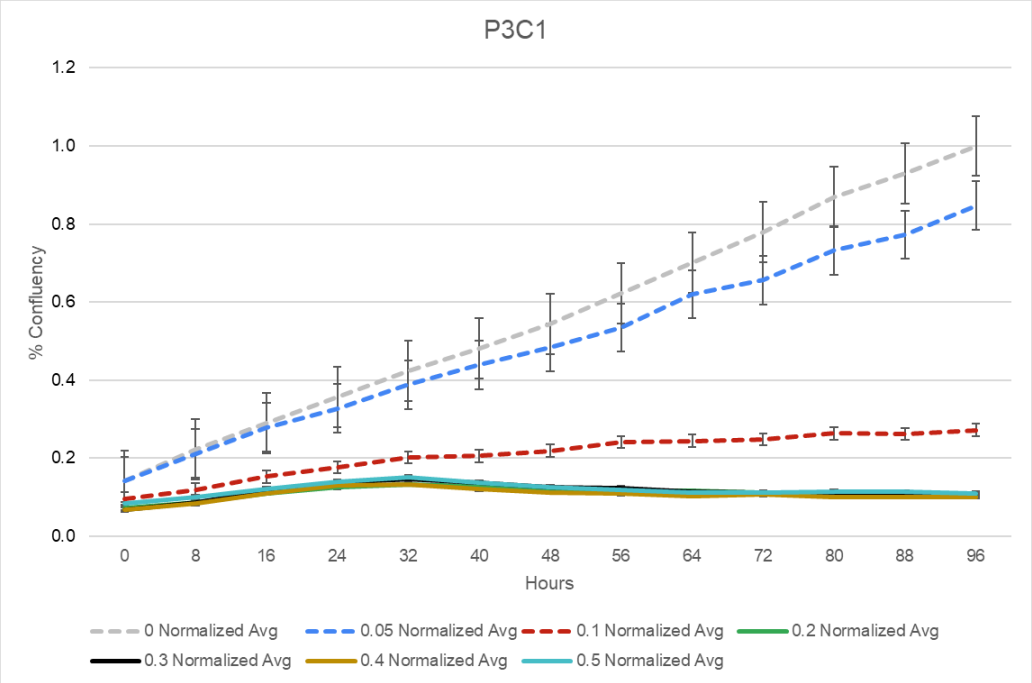


Figure 6.A: *Summary of trends of various concentrations* - As seen in the adjacent graph, there was a slight increase followed by a decrease in the percent confluency with all concentrations below 0.1 mM. At 0.05 mM and 0.1 mM concentrations, the percent confluency of CK4 cells did increase, but this was still lesser compared to the untreated cells.

Anova: Single Factor: P3C1

SUMMARY

Groups	*Count*	*Sum*	***Average***	*Variance*
0 Normalized Avg	13	7.361	**0.566**	0.0771
0.05 Normalized Avg	13	6.435	**0.495**	0.0496
0.1 Normalized Avg	13	2.704	**0.208**	0.0032
0.2 Normalized Avg	13	1.434	**0.110**	0.0003
0.3 Normalized Avg	13	1.493	**0.115**	0.0004
0.4 Normalized Avg	13	1.385	**0.107**	0.0003
0.5 Normalized Avg	13	1.542	**0.119**	0.0003

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	3.084731653	6	0.514	27.425	**6.92644E-18**	2.209
Within Groups	1.574728987	84	0.019			
Total	4.65946064	90				

Figure 6.B: *ANOVA* test for CK103 shows that the means are different (Low P value highlighted in green). So, paired comparison T-tests were performed using the 0 concentration as a baseline to statistically validate whether each concentration of Vitamin C had an impact on confluency.

Concentration	% Change in cell growth after 96 hrs
0 mM	606.53%
0.05 mM	500.81%
0.1 mM	180.96%
0.2 mM	45.30%
0.3 mM	60.11%
0.4 mM	47.21%
0.5mM	30.26%

Figure 5.C: *Final change in % confluency* - The above table shows the change in percent confluency after 96 hrs (4 days).

t-Test: Two-Sample Assuming Equal Variances: P3C1

	0 Avg	*0.05 Avg*
Mean	**0.566229637**	**0.494982823**
Variance	0.077064013	0.049625242
Observations	13	13
Pooled Variance	0.063344627	
Hypothesized Mean Difference	0	
df	24	
t Stat	0.721717497	
P(T<=t) one-tail	**2.4E-01**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.47744133	
t Critical two-tail	2.063898562	

	0 Avg	*0.1 Avg*
Mean	**0.566229637**	**0.208020486**
Variance	0.077064013	0.003213372
Observations	13	13
Pooled Variance	0.040138692	
Hypothesized Mean Difference	0	
df	24	
t Stat	4.55839278	
P(T<=t) one-tail	**6.4E-05**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	0.000127661	
t Critical two-tail	2.063898562	

	0 Avg	*0.2 Avg*
Mean	**0.566229637**	**0.110284415**
Variance	0.077064013	0.000270407
Observations	13	13
Pooled Variance	0.03866721	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.911503189	
P(T<=t) one-tail	**2.1E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	4.2373E-06	
t Critical two-tail	2.063898562	

	0 Avg	*0.3 Avg*
Mean	**0.566229637**	**0.114810625**
Variance	0.077064013	0.000441173
Observations	13	13
Pooled Variance	0.038752593	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.846367895	
P(T<=t) one-tail	**2.5E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	4.9772E-06	
t Critical two-tail	2.063898562	

	0 Avg	*0.4 Avg*
Mean	**0.566229637**	**0.106522409**
Variance	0.077064013	0.000296835
Observations	13	13
Pooled Variance	0.038680424	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.959260868	
P(T<=t) one-tail	**1.9E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	3.76664E-06	
t Critical two-tail	2.063898562	

	0 Avg	*0.5 Avg*
Mean	**0.566229637**	**0.118631733**
Variance	0.077064013	0.000316374
Observations	13	13
Pooled Variance	0.038690194	
Hypothesized Mean Difference	0	
df	24	
t Stat	5.801553094	
P(T<=t) one-tail	**2.8E-06**	
t Critical one-tail	1.71088208	
P(T<=t) two-tail	5.56131E-06	
t Critical two-tail	2.063898562	

Figure 6.D: *T test summary* - The mean percent confluency of P3C1 cells at each concentration of vitamin C were compared the percent confluency of P3C1 cell line at 0 mM vitamin C with a t-test. A P-value less than 0.05 for all concentrations except 0.05 mM shows that the mean confluency for all concentrations of Vitamin C are significantly lower than the percent confluency at 0mM vitamin C. With a 0.05 concentration of vitamin C, however P>0.05. Therefore, the null hypothesis for 0.05 mM vitamin C (the mean of the percent confluency at 0 mM is different from the mean of the percent confluency at 0.05 mM) cannot be rejected, indicating that all concentrations except 0.05 mM had a statistically significant effect on the percent confluency of the P3C1 cell line.

The IC50, which is the concentration of vitamin C that will kill 50% of the cell population was calculated for each cell line. The final percent confluency (cell growth) for each cell line was graphed on a scatterplot. A line of best fit was found for these plots. The concentration of vitamin C when the population was 0.5 was found and using the equation for these lines of best fit.

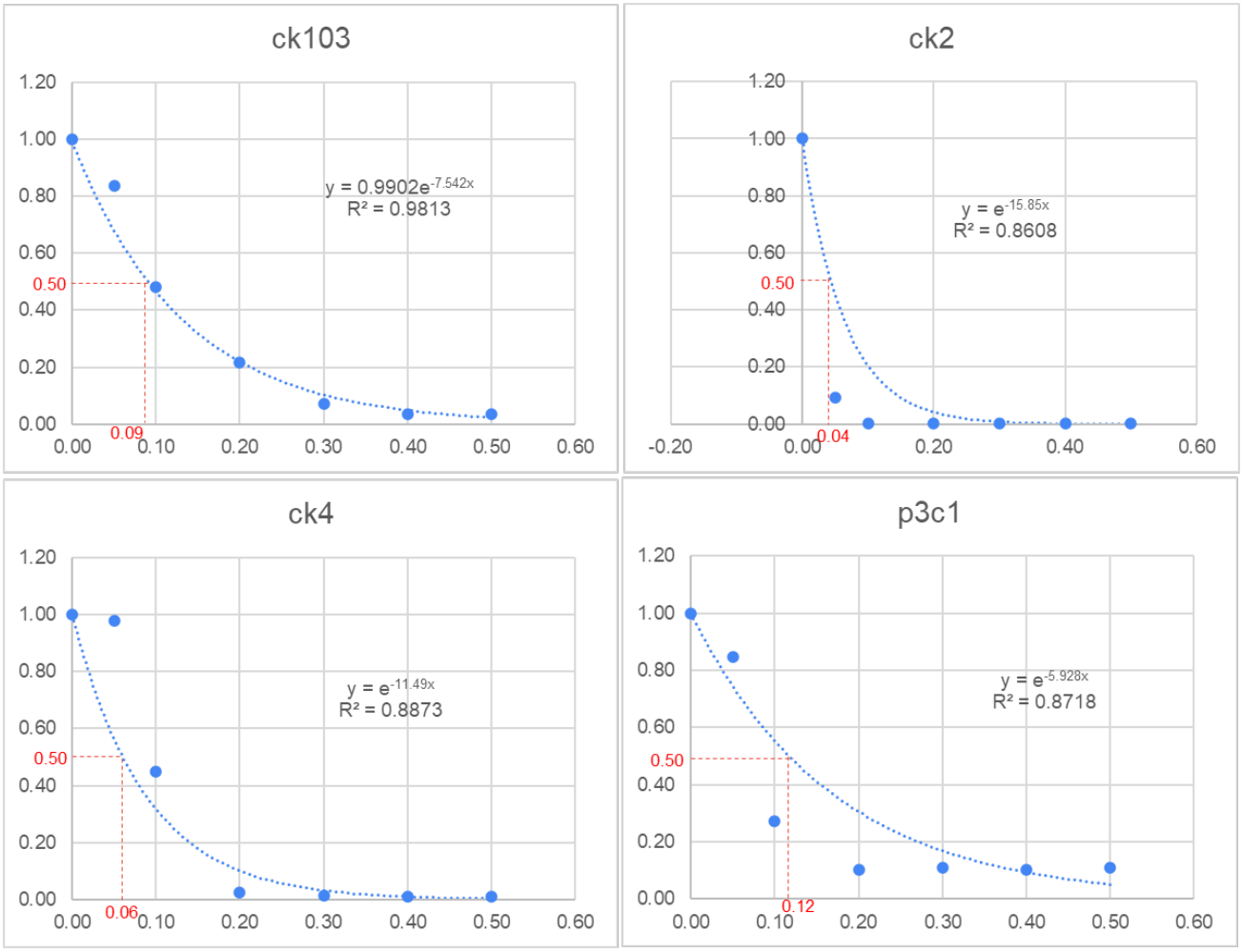


Figure 7: Graphs depict the IC50 for each for each cell line.

Conclusions

The statistical analysis mentioned above helped form the following conclusions: (1) Vitamin C decreased or stopped the growth of all cell lines, proving the hypothesis that vitamin C will inhibit the growth of cancerous cell lines (CK2, CK4, and CK103). (2) Increasing dose of vitamin C decreased cell growth in a dose dependent manner (meaning that as the concentration of vitamin C increased, the cell growth decreased). (3) CK2 (Hras mu p53wt), which was one of the most invasive cancer cell lines tested as a part of this project displayed the most sensitivity to vitamin C whereas P3C1 (non-mutant 3T3 mouse fibroblasts) displayed the least sensitivity to vitamin C.

This study showed that vitamin C can indeed control the growth of squamous cell carcinoma tumors. This can be further researched to find alternative treatments for cancer at a fractional cost without the adverse effects of chemotherapy and radiation, which has been shown to cause collateral damage to patients.

References

1. Moser, R., Gurley, K. E., Nikolova, O., Qin, G., Joshi, R., Ashley, A., Grandori, C., & Kemp, C. J. (2022, May 10). Synthetic lethal kinases in Ras/p53 mutant squamous cell carcinoma.

2. Ngo, B., Van Riper, J. M., Cantley, L. C., & Yun, J. (2019, May). Targeting cancer vulnerabilities with high-dose vitamin C [Review of Targeting cancer vulnerabilities with high-dose vitamin C].

3. Skin Cancer Foundation. (2019). *Squamous cell carcinoma - the skin cancer foundation*. The Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/squamous-cell-carcinoma/

4. Reczec, C., & Chandel, N. (2015). Revisiting Vitamin C and Cancer [Review of *Revisiting Vitamin C and Cancer*]. *Science*, *350*(6266), 1317–1318. Fred Hutch Cancer Research Center. https://doi.org/10.1126

5. Villagran, M., Ferreira, J., Martorell, M., & Mardones, L. (2021). The Role of Vitamin C in Cancer Prevention and Therapy: A Literature Review [Review of *The Role of Vitamin C in Cancer Prevention and Therapy: A Literature Review*]. *Antioxidants*, *10*(1894), 1–15. MDPI. https://doi.org/10.3390/antiox10121894

6. Skin Cancer Foundation. (2020, April). *Skin Cancer Facts & Statistics - The Skin Cancer Foundation*. The Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

7. Du, J., Martin, S. M., Levine, M., Wagner, B. A., Buettner, G. R., Wang, S., Taghiyev, A. F., Du, C., Knudson, C. M., & Cullen, J. J. (2010). Mechanisms of Ascorbate-Induced Cytotoxicity in Pancreatic Cancer. *Clinical Cancer Research*, *16*(2), 509–520. https://doi.org/10.1158/1078-0432.ccr-09-1713

Environmental Science

Using Machine Learning to Predict the Power Conversion Efficiency of Perovskite Solar Cells

Divij Aswinkumar

Nikola Tesla STEM High School

Abstract

Currently, one of the main drawbacks of solar energy is low efficiency due to the intrinsic properties of the semiconductors being used. This project seeks to strengthen the knowledge of a new type of semiconductor, perovskites. Since perovskites have varying stability and synthetically testing each one is both time and cost inefficient, this work investigates whether machine learning can be used to predict if a perovskite will have high efficiency using the inherent properties such as atomic radii of elements in the material or layers in perovskite solar cell (PSC). The goal of this project is to construct an ML model that can predict the power conversion efficiency (PCE) with an accuracy above 70% using the Goldschmidt Factor, a formula based on the radii of specific elements in the perovskite compound, as a stability indicator. Using linear regression on data from the National Perovskite Database, an inverse relationship was found to predict PCE (dependent variable) using the Goldschmidt Factor (independent variable) of each individual perovskite. When the Goldschmidt Factor value increased the PCE decreased. This relationship was compared to other ML models to confirm the superior strength. The Goldschmidt Predictor model had an accuracy of 73.7% in predicting the efficiency of perovskites and a root mean square error of 26.29. With these results, the target accuracy (> 70%) was achieved and a baseline was established for predicting the efficiency of perovskite-based solar cells solely based on inherent properties of the material in a time and cost-efficient manner.

Keywords: solar cells, machine learning, perovskites, semiconductor, Goldschmidt Factor, stability factors, power conversion efficiency

1. INTRODUCTION

1.1 How Solar Panels Work

Solar panels use the sun's rays to create energy by absorbing high energy particles called photons. The photons are absorbed by a solar panel using something known as a photovoltaic cell. These cells are made of thin-layered semiconductors. Once a photon is absorbed, its energy can be transmitted to an electron. This excites the electron to a higher energy state. The electron leaves behind a hole, which can be treated as a positively charged particle, and creates something known as an electron-hole pair.

The electron in the n-type layer of the semiconductor attempts to recombine with the hole in the p-type layer. This recombination is a process where energy is lost. However, an existing electric field in the photovoltaic cell prevents the electron from recombining immediately. As a result, the electron travels through an external circuit in order to meet up with the hole in the p-type layer.

This process, however, must take place within a certain amount of time because the electron begins to lose energy and drift back towards the hole if another pathway is not provided. By connecting the two ends of the p-n junction externally, a circuit is created which provides a way for electrons to avoid going through the electric field within the p-n junction. This causes electrons to move to one end of the semiconductor known as a contact while the holes move to the other contact. As a large number

of electron-hole pairs go through this process, a current is generated in the circuit.

However, to generate power in a solar cell voltage is necessary in addition to the current. These vital processes to create energy is the photovoltaic effect. Since there are many negatively charged electrons in the n-type side and positively charged holes in the p-type side there is a buildup of charge in the p-n junction of the semiconductor which can be used to produce energy. As the electrons and holes move in opposite directions through the external circuit, the voltage is then usable outside the semiconductor as well.

Figure 1 below shows the basic functionality of a solar cell where a photon hits a semiconductor which converts the light energy into usable electrical energy.

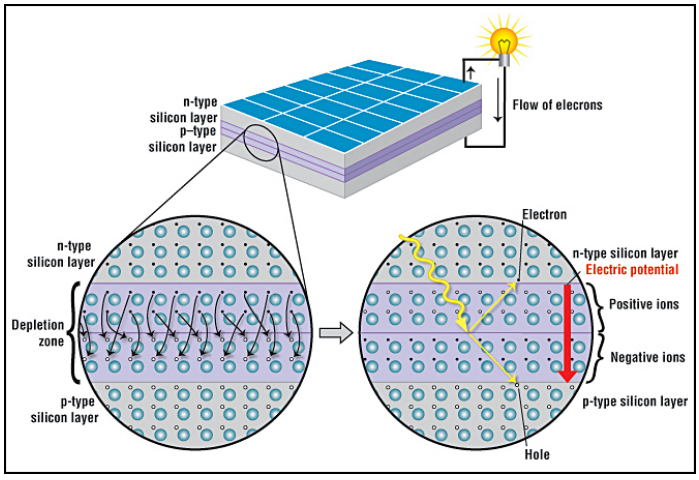


Figure 1. Schematic representation of a solar cell, showing the n-type and p-type layers, with a close-up view of the depletion zone around the junction between the n-type and p-type layers; from Fernandez, 2020

1.2 Replacing Low Efficiency Semiconductors

Unfortunately, current semiconductors do not convert enough energy from the sun into electrical energy which is a key reason why solar panels are not as popular as non-renewable energy alternatives like natural gas. During this process, some energy is lost and can't be used. These solar panels have a typical efficiency of 17-24% which is a key reason why people don't switch to solar power.

Along with this, previous research has proven that over time solar panels begin to degrade, which means their efficiency reduces even more and over the course of a solar panel's life it produces lower and lower amounts of energy.

Perovskites solve these problems because they show the promise of much higher starting efficiencies than existing semiconductors and last much longer. Perovskites are not just one specific compound but a group of multiple compounds that follow a certain structure. This repeating crystalline structure of ABX_3 includes three different elements that fit into the A, B, and X ion sites respectively and these elements can come from anywhere on the periodic table. This leads to thousands of combinations for new perovskite materials for solar panel development. Figure 2 below shows 6 possibilities for $SrTiO_3$ perovskite with the ABX_3 structure.

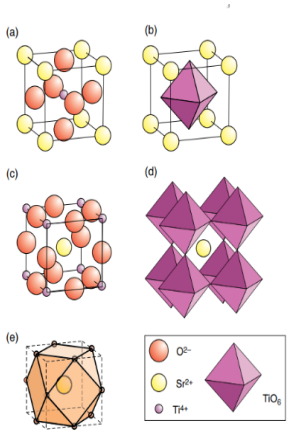


Figure 2. Different possible structures of Perovskites for all following repeating crystalline structure; from Tilley, 2016

1.3 Testing Perovskites Viability

The new problem that arises is determining which perovskite to choose for building the solar cell because perovskites have varying stability which means it is important to choose the perovskite that is going to remain stable and maintain a high efficiency over time.

Currently, the process used for determining which perovskites materials should be used for building solar panels involves synthesizing the solar cell and experimentally determining the power conversion efficiency, or the PCE. This is both time and cost inefficient. With so many possible perovskites to test, the total time for testing increases significantly. As more perovskites continue to be discovered a new strategy must be used. This strategy is using machine learning to predict PCE based on other measurements for a semiconductor. Figure 3 below shows the current synthesis and testing process which takes multiple hours and must be replaced.

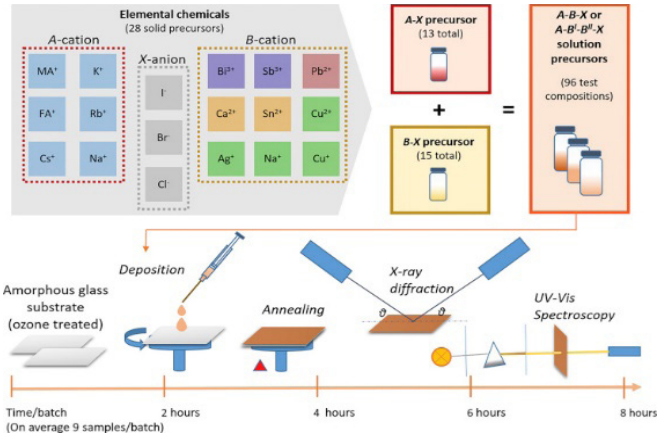


Figure 3. Streamlined process for synthesis & testing of perovskite-based materials taking 8 hours vs. neural network taking 5.5 minutes; from Dunlap-Shohl et al., 2019

1.4 Use of Machine Learning in Solar Industry

In the past people have tried to use stability factors like the T80 Lifetime which determines the time it takes for the PCE to degrade to 80% of its original value to predict the PCE but this is still a post-synthesis measurement. Other post-synthesis measurements include outdoor vs indoor performance and current-voltage curves. Studies like these produced an accuracy of about 70% in predicting the PCE of a semiconductor.

The key aspect of this research is that it focuses on using a pre-synthesis stability indicator to predict the PCE. A stability indicator is determined before synthesis using the inherent properties of the material so you can create a prediction without having to build every solar cell.

As more perovskites are discovered this research will become more crucial because it can help with choosing which material is viable enough for experimental testing based on the predicted PCE from this ML model.

2. HYPOTHESIS AND GOALS

This research aims to identify a stability indicator to predict the PCE at an accuracy higher than the current industry standard. This research also aims to employ machine learning to predict PCE without requiring synthesis for any perovskite semiconductors.

The question that this project attempted to answer was “can a machine learning algorithm be built utilizing stability indicators of perovskite solar cells to predict expected power conversion efficiency with an accuracy above 70%?”

The hypothesis was that a model will be able to use stability indicators because stability indicators will have a correlation with the power conversion efficiency. The reason for this is that synthetically tested perovskite solar cells, or PSCs, have shown a trend between the measured stability factors and the PCE. Since stability factors like the T80 and stability indicators are both stability-based values the trend should persist.

This goal of 70% was set because the current industry standard of predictions using the T80 lifetime had an accuracy of 70%. For this research to be useful, the new model would have to have a similar or higher accuracy to justify replacing the existing T80 prediction system with a new machine learning based model.

3. METHODOLOGY

3.1 Choosing Stability Indicator

The stability indicator that is used in this research is the Goldschmidt Factor which is calculated using the inherent atomic radii for the 3 elements that compose the perovskite material. This factor was chosen for the initial model because it included a value for each unique element in the perovskite and also had a straightforward calculation method. The formula used is as follows:

$$t = \frac{r_a + r_x}{\sqrt{2}(r_b + r_x)}$$

Where r_a is the radius of the element in the A ion site, r_b is the radius of the element in the B ion site, and r_x is the radius of the element in the X ion site.

3.2 Calculating Model Accuracy

The model uses one variable, the Goldschmidt Factor, to predict PCE. Then the accuracy of the ML model is tested using the root mean square error. The root mean square error is commonly used in ML to test the performance of a model. The formula used is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted - Actual)^2}{N}}$$

Where Predicted is the predicted PCE value, Actual is the actual PCE value, and N is the total number of data points that were tested.

4. DETAILED METHODOLOGY

4.1 Data Acquisition and Cleanup

To train and test the machine learning models, data was collected from the MaterialsZone National Perovskite Database Project. Researchers from around the world submit their results here for any new perovskite synthesized and tested, and the data includes measurements and values for 410 different factors. These measurements are grouped based on what type of information they contain about each perovskite. Since this research was only going to use pre-synthesis measurements and factors all other columns were removed from the dataset.

4.2 Building Goldschmidt Univariate ML Model

To determine the Goldschmidt Factor, data was extracted about the elements used in each perovskite and then cross-referenced with a table of each element and their ionic radius to perform the calculation. Figure 4 below shows the code used to do this for each of the 3 ion sites:

Then a simplified data table was created using the X values as the Goldschmidt value and Y values as the PCE of the same perovskite semiconductor. Using sklearn a test_train split was created with 80% of the data for training and 20% for testing.

This dataset also provided the identified target factor for our machine learning models which is the Stabilized End of Experiment PCE. This was the only post-synthesis column that was kept in the dataset to use as the y_train and y_test values.

4.3 Comparing Other ML Models

To confirm the strength of these results, the Goldschmidt Predictor's RMSE value was compared to the results obtained using two more Multivariable Linear Regression Models. The first model was simple multivariate regression on all the columns from the MaterialsZone dataset. This included 306 factors in total. Usually machine learning performs better when more factors are inputted for making a prediction, but that is only true when the factors have a direct correlation with the value being predicted. The purpose of this test was to prove that the Goldschmidt Factor was a better indicator because despite being a univariate model it would perform better because of its greater relevance to the PCE.

The second model was a more unique aspect of this research because it used Principal Component Analysis which helps with dimensionality

```
if elementBreakdown.iloc[1, 2].find(";") != -1 :
  bIon = elementBreakdown.iloc[1, 2].split(";")[0]
else:
  bIon = elementBreakdown.iloc[i, 2]


bIonIndex = elementRadii[elementRadii["Symbol"]==bIon].index.values
bIonRadius = elementRadii.iloc[bIonIndex, 1].values
if bIonRadius.size == 0:
  bIonRadius = [0]


if elementBreakdown.iloc[i, 4].find(";") != -1:
  cIon = elementBreakdown.iloc[i, 4].split(";")[0]
else:
  cIon = elementBreakdown.iloc[1, 4]

cIonIndex = elementRadii[elementRadii["Symbol"]==cIon].index.values
cIonRadius = elementRadii.iloc[cIonIndex, 1].values
if cIonRadius.size == 0:
  cIonRadius = [0]

compoundInfoRough = [compound, aIonRadius[0], bIonRadius[0], cIonRadius[0], validCompound]
if 0 not in compoundInfoRough:
  validCompound = 1
  validCompoundCount += 1
else:
  validCompound = 0
compoundInfo.loc[len(compoundInfo.index)] = [compound, aIonRadius[0], bIonRadius[0], cIonRadius[0], validCompound, pceEnd]
```

Figure 4. Code snippet from ML Model used to calculate Goldschmidt Factor

reduction. In essence, with PCA analysis, instead of considering 100s of variables, the dimensionality is reduced to fewer components by using covariance matrices based on assigning weightage to the effect of each factor on the PCE. Principal component analysis was implemented to improve the predictive power of the multivariate linear regression model. This means covariance matrices of all factors from the MaterialsZone database were created, and weightages were assigned to factors that had more variance across the dataset. Like the regular multivariate linear regression model, this PCA enhanced model was also tested to see how it would compare to the Goldschmidt Factor's performance.

5. RESULTS

5.1 Univariate Linear Regression with Goldschmidt Factor

The ML analysis first developed a relationship between Goldschmidt Factor and PCE by running a Linear Regression on 80% of the data as training data. This training set developed an equation to predict the PCE using just the Goldschmidt Factor value. The equation is as follows:

$$y = -85.19x + 202.39$$

Where y is the predicted PCE value and x is the Goldschmidt Factor value.

Then the model used the equation it had developed to predict the PCE for the 20% testing data and this resulted in an RMSE of 26.29 and an accuracy of 73.71%. Figure 5 below shows the predicted and actual PCE for each perovskite that was tested using the Goldschmidt Factor Model:

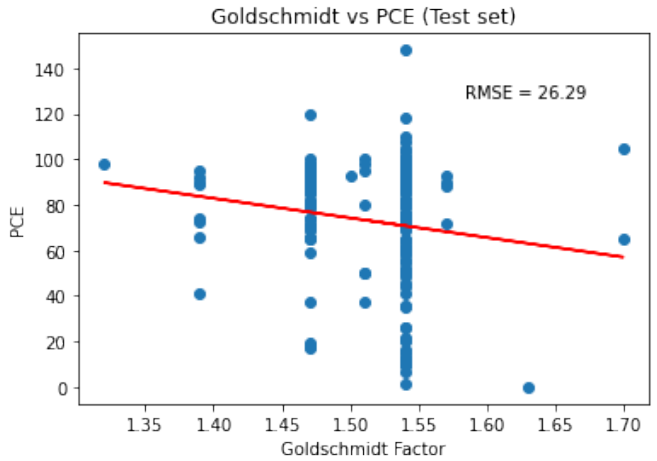


Figure 5. Results from univariate linear regression model using Goldschmidt Factor to predict PCE

The red line shows the linear regression equation that was created and used by the ML model. At first it may look like there are many data points that do not match up with the red line, but for every point along the red line there were many perovskites that had the same Goldschmidt Factor value and thus were predicted to have the same PCE as well. As a result, the datapoints are stacked on top of each other.

Furthermore, if we analyze the distribution of errors, we see that even though the RMSE value is 26.29 more than 50% of the datapoints had an error under 20 and a few outliers are causing the RMSE value to increase. Figure 6 below shows this distribution:



Figure 6. Distribution of error for univariate linear regression with Goldschmidt Factor

5.2 Multivariate Linear Regression (MVLR) with MaterialsZone Data

As mentioned earlier, to increase the confidence in the Goldschmidt Predictor Model a coefficient table was created for all 306 factors from the MaterialsZone database to see if any of them had a higher correlation than the Goldschmidt Factor. Figure 7 below shows a truncated version of that correlation coefficient table:

	Attribute	Coefficient
0	Cell_stack_sequence	0.065956
1	Cell_area_total	0.004405
2	Cell_area_measured	0.009790
3	Cell_number_of_cells_per_substrate	-0.012889
4	Cell_architecture	0.042748
...	...	...
231	Add_lay_back_deposition_reaction_solutions_con...	0.008844
232	Add_lay_back_deposition_reaction_solutions_vol...	-0.003355
233	Add_lay_back_deposition_reaction_solutions_age	0.000995
234	Add_lay_back_deposition_reaction_solutions_tem...	0.000995
235	Add_lay_back_deposition_substrate_temperature	0.000995

Figure 7. Truncated correlation coefficient table for 306 factors from MaterialsZone database

In order to not assume whether a factor would have a positive or negative correlation with PCE both the largest positive and largest negative correlation coefficient factors were graphed against PCE to see if any possible relationship would exist. Both factors had correlation coefficients below 0.1 so even without looking at their graphs it would be fairly easy to expect that they would have low predictive power. Regardless they are provided below in Figure 8:

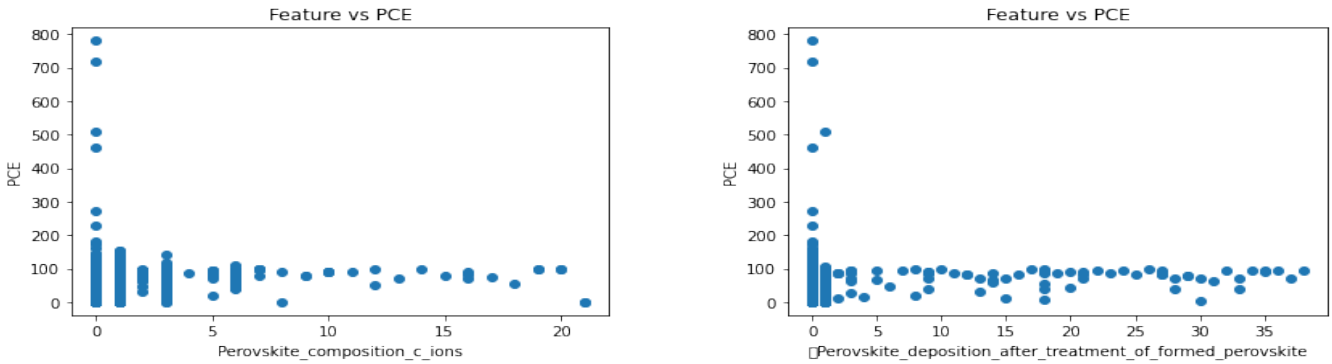


Figure 8. Individual graphs of perovskite C site ions vs. PCE and perovskite deposition after treatment vs PCE showing low predictive power

The basic multivariate linear regression model where all 306 factors were combined with weighted coefficients increased the predictive power for determining PCE compared to the highest positive/negative coefficients, but it was still lower than the Goldschmidt Factor Model. The RMSE obtained was 30.21 which is higher than it was for the Goldschmidt Predictor Model meaning the Multivariable Linear Regression Model has a lower performance, contrary to typical expectation. Figure 9 below shows a visualization of these results:

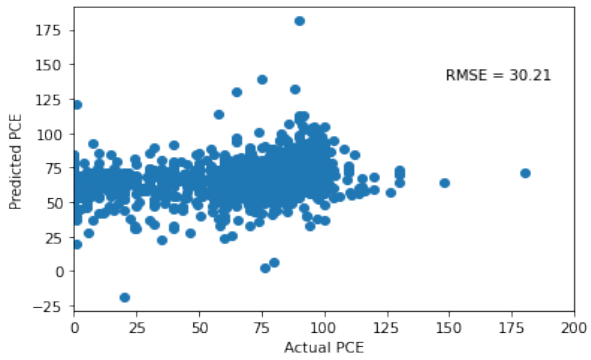


Figure 9. Results from multivariate linear regression model using 306 MaterialsZone factors to predict PCE

5.3 MVLR Model Enhancement with Principal Component Analysis

As mentioned before, principal component analysis was implemented to improve the predictive power of this multivariate linear regression model. To include PCA analysis in the model it had to be set up so that each factor would be measurable in terms of variance rather than original value using StandardScaler. To expedite this process a pipeline system was set up with the steps shown in Figure 10 to make each of the 8000 datapoints from the MaterialsZone database run through the process.



Figure 10. Simplified view of the pipeline system setup for each datapoint with following steps: 1) Standard Scaler 2) PCA 3) MVLR

The PCA process determined that the model would require 25 components to measure the full variance in the dataset. Figure 11 below shows the scree plot of variation below:

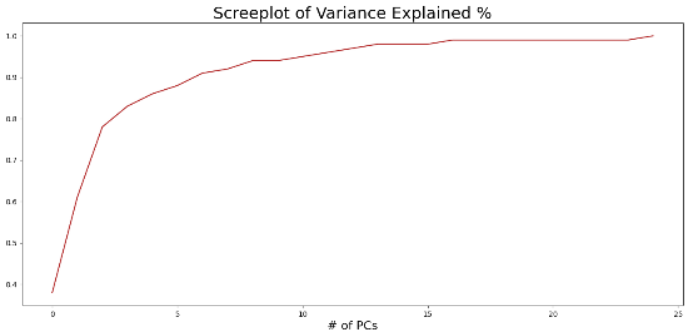


Figure 11. Scree plot of variation measured by each inclusion of new principal component (PC) into PCA enhanced model until 25 components measuring 100% variation

The PCA analysis also highlighted that the factor with the most influence on the model was the cell stack sequence but Figure 12 below shows how the

predictive power of just this variable to predict PCE would have also been low as indicated by the red line very close to the x-axis.

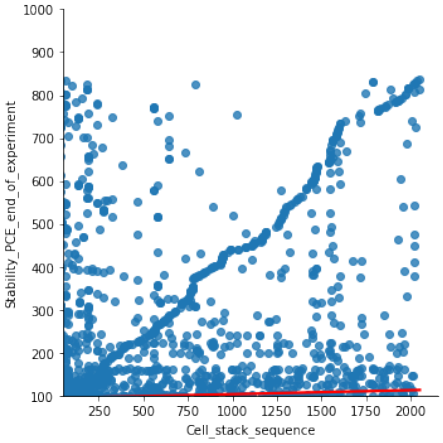


Figure 12. Individual graph of cell stack sequence vs PCE showing low predictive power

Even when combined with all the other factors that measured the variance in 25 dimensions the RMSE value for the PCA model was 34.66 which is again larger than that of the Goldschmidt Factor. Figure 13 below shows a visualization of these results:

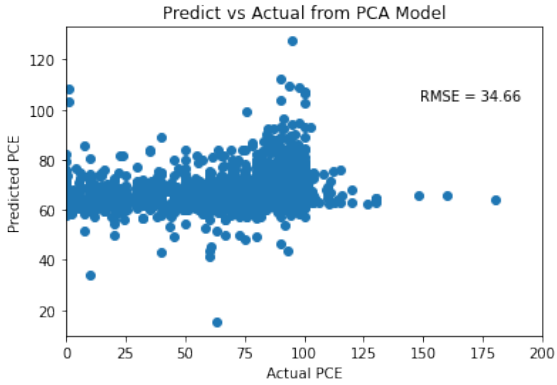


Figure 13. Results from multivariate linear regression model after dimensionality reduction to 25 principal components using PCA analysis

An important thing to consider: Without using machine learning, predicting the PCE for these 8000+ perovskite would have taken 64000+ hours. Instead with these models the predicted PCE was obtained in just minutes.

5.4 Limitations of Current Goldschmidt Factor Model

One limitation of this methodology is that the Goldschmidt Factor is only usable for single perovskites that have singular elements in each site of the ABX3 structure. The Goldschmidt Factor could not be calculated for other types of perovskites such as double or organic perovskites with more complex structures. This may have introduced unintentional selection bias between the datasets used for the Goldschmidt Model and the other two ML models tested.

5.5 Further Enhancement of Goldschmidt Factor Model

Since Machine Learning predictions tends to work best with multiple factors that have a direct relation to the target value, the limitations of this model could be overcome by using other stability indicators such as the Octahedral or Tau factors which use varying levels of complexity and different properties of perovskites like oxidation state to predict PCE. Another possible enhancement is to determine the range of Goldschmidt Factor values where the accuracy of predicting PCE will be highest.

5.5a Applicability Range for Goldschmidt Factor Model

To determine where the Goldschmidt Factor had the highest accuracy in predicting a limit was set as to what was considered low error and high error. This boundary was set at an error value of 15. As Figure 14 below shows, green data points are where the Goldschmidt Factor Model predicted the PCE with the highest accuracy while red shows the remaining points.

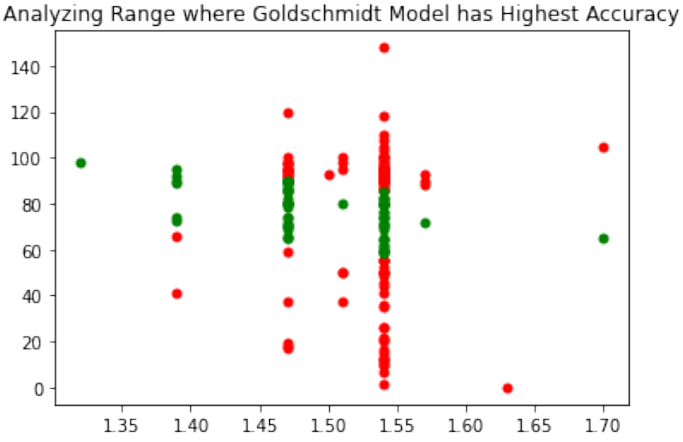


Figure 14. Graph showing areas where Goldschmidt Factor Model had highest accuracy (green)

From this graph, it's evident that the greatest success occurred when the Goldschmidt Factor value was from 1.45 to 1.55. This means that when a new perovskite developed has a Goldschmidt Factor value in this range the prediction can be viewed with higher confidence because the error in that prediction will be lower.

5.5b Combining Goldschmidt and Octahedral Factor

To begin the process of developing a model with multiple factors with high relevance to PCE the Octahedral Factor was added as a second variable Figure 15a and 15b shows the results from the combination of both variables in predicting PCE along with the distribution of error.

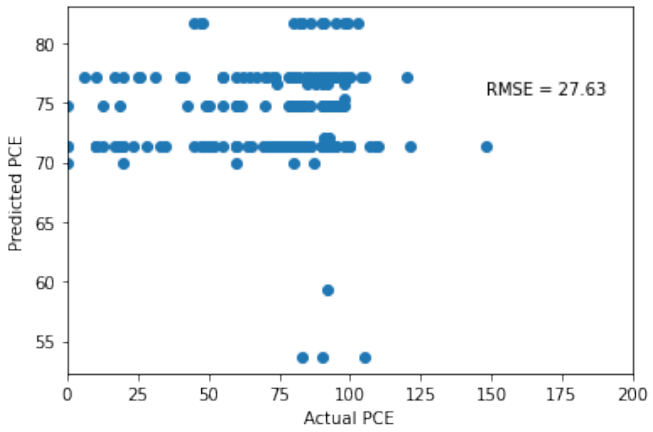


Figure 15a. Results from multivariate linear regression model using Goldschmidt Factor and Octahedral Factor to predict PCE



Figure 15b. Distribution of error for multivariate linear regression with Goldschmidt Factor and Octahedral Factor

This addition produced a similar RMSE value of 27.63. Since this value is close to the RMSE of the Goldschmidt Factor on its own, it proves that Octahedral Factor also is relevant in predicting the PCE. Additionally, the distribution of error shifted to the left and is now under 15 as opposed to 20 when using only the Goldschmidt Factor.

6. Conclusion

To conclude, the current Goldschmidt Predictor Model, with a 73% accuracy, achieved the engineering goal of surpassing the industry standard of 70% accuracy. The confidence in this predictor is also high because it performed better than other multivariate ML models proving that the Goldschmidt Factor on its own had more predictive power than a combinations of factors from the MaterialsZone Database. This shows that stability indicators can and should be used to predict PCE and also avoid synthesizing thousands of perovskites before testing for viability.

To simplify, all the results gathered in Table 1 below show that the Goldschmidt Factor Model had the lowest RMSE and the highest accuracy of 73.71%.

ML Model	RMSE	Accuracy
Goldschmidt Factor	26.29	73.71%
Goldschmidt + Octahedral	27.63	72.37%
MVLR	30.21	69.79%
PCA + MVLR	34.66	65.44%

Table 1. Summary of RMSE and accuracy values for machine learning models tested

Based on the RMSE and accuracy obtained upon inclusion of the Octahedral Factor, we see that a combination of multiple stability indicators highly correlated to the PCE might be something to explore. For example, including the Tau Factor might increase the accuracy to more desirable levels and provide applicability over a wider range of perovskites.

7. Future Industry Application

This current model is germane for application in industry to identify new viable single perovskite materials. This can help streamline the process of determining which perovskite to use for a new solar panel by seeing if it meets desired PCE requirements. To further improve the usability of the model the next steps would be to research more factors that impact PCE and implement them into the model so that overall accuracy rises to 90% desirably.

8. Acknowledgments

I would like to take this time to thank my mentors, Ashutosh Joshi and Nitin Agrawal who are machine learning experts at Amazon and Microsoft respectively for helping me come up with different models to use for my project along with suggesting

visualizations that would help me represent my data effectively.

Citations

Bartel, C. J., Sutton, C., Goldsmith, B. R., Ouyang, R., Musgrave, C. B., Ghiringhelli, L. M., & Scheffler, M. (2019). New tolerance factor to predict the stability of perovskite oxides and halides. Science advances, 5(2), eaav0693. https://doi.org/10.1126/sciadv.aav0693

Chen, Y., Zhang, L., Zhang, Y., Gao, H., & Yan, H. (2018, March 14). *Large-area perovskite solar cells – a review of recent progress and issues*. RSC Advances. Retrieved February 28, 2023, from https://pubs.rsc.org/en/content/articlehtml/2018/ra/c8ra00384j

Connelly, B. (2021, May 13). *Principal Component Analysis (PCA) For Dummies*. Bill Connelly. Retrieved February 28, 2023, from http://www.billconnelly.net/?p=697

Dunlap-Shohl W, Meng Y, Sunkari P, Beck D, Meila M, Hillhouse H. Physiochemical Machine Learning Models Predict Operational Lifetimes of CH3NH3PbI3 Perovskite Solar Cells. ChemRxiv. Cambridge: Cambridge Open Engage; 2022; This content is a preprint and has not been peer-reviewed.

Fernandez, A. (2023). *How a solar cell works*. American Chemical Society. Retrieved February 28, 2023, from https://www.acs.org/education/resources/highschool/chemmatters/past-issues/archive-2013-2014/how-a-solar-cell-works.html

Jacobsson, T. J. (2020). *The Perovskite Database Project*. The Perovskite Database. Retrieved February 28, 2023, from https://www.perovskitedatabase.com/Resources

Lang, Felix & Shargaieva, Oleksandra & Brus, Viktor & Neitzert, Heinz & Rappich, Joerg & Nickel, Norbert. (2017). Influence of Radiation on the Properties and the Stability of Hybrid Perovskites Perovskite Solar Cells. Advanced Materials. 30. 10.1002/adma.201702905.

Li, W., Jacobs, R., & Morgan, D. (2018, May 8). *Data and supplemental information for predicting the thermodynamic stability of perovskite oxides using machine learning models*. Data in Brief. Retrieved November 17, 2022, from https://www.sciencedirect.com/science/article/pii/S2352340918305092

Ngo, L. (2018, December 5). *Principal component analysis explained simply*. BioTuring's Blog. Retrieved February 28, 2023, from https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/

M.I. Asghar, J. Zhang, H. Wang, P.D. Lund, Device stability of perovskite solar cells – A review, Renewable and Sustainable Energy Reviews,

Volume 77, 2017, Pages 131-146, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2017.04.003.(https://www.sciencedirect.com/science/article/pii/S1364032117304756)

Saliba, Michael & Stolterfoht, Martin & Wolff, Christian & Neher, Dieter & Abate, Antonio. (2018). Measuring Aging Stability of Perovskite Solar Cells. Joule. 2. 10.1016/j.joule.2018.05.005.

Sun, S., Hartono, N. T. P., Ren, Z. D., Oviedo, F., Buscemi, A. M., Layurova, M., Chen, D. X., Ogunfunmi, T., Thapa, J., Ramasamy, S., Settens, C., DeCost, B. L., Kusne, A. G., Liu, Z., Tian, S. I. P., Peters, I. M., Correa-Baena, J.-P., & Buonassisi, T. (2019, June 5). *Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis*. Joule. Retrieved February 28, 2023, from https://www.cell.com/joule/fulltext/S2542-4351(19)30257-0

Tilley, R.J.D. (2016). The ABX_3 Perovskite Structure. In Perovskites, R.J.D. Tilley (Ed.). https://doi.org/10.1002/9781118935651.ch1

A Novel Integrated Approach to Illegal Logging Detection

Saraswathy Amjith

Tesla STEM High School

Abstract

Illegal selective logging is devastating to the world's biodiversity and forests, and many countries struggle with funding to monitor these forests. 94% of the deforestation in the amazon is done illegally and occurs at selective logging sizes which hasn't been researched as much as clear cut deforestation. This study seeks to improve the accuracy of detecting selective logging, through a novel machine learning approach that integrates sentinel 1 (radar) imagery and sentinel 2 (optical) imagery, as current solutions use only one form of data. Additionally, an interactive tool was created for forest protection agencies to identify any logging activities. Requisite data was collected through Google Earth Engine for sentinel 1 and sentinel 2 and used to create 3 new datasets of 134.606 samples. Then, all data sets were used for prediction using three different machine learning architectures for classification: Convolutional Neural Network (CNN), Random Forest, and Gradient Boosted Trees. The CNN performed best on the integrated data set with an F1 score of 94.73% and an accuracy of 95.08% compared to Sentinel 1 and 2 on their own achieving accuracies of 57.51%, and 91.95% respectively. The XGB had a 94.39% accuracy/94.03 F-1 and required less computing power, therefore, the XGB model was used in the website to allow users to select any region in the world to identify logged and stable forest. This study demonstrated that the integration of optical and radar data can help forest protection agencies in predicting where selective logging is happening worldwide.

1 Introduction

Thesis: The goal of this project is to use machine learning models to detect illegal logging occurrences through exploring different models (CNN, Random Forest, Gradient Boosted Trees) using a novel integration of both optical (satellite) and radar data, and develop a tool to track logging for any location at real-time.

According to the United States Agency of International Development, illegal logging is estimated to be a $150 billion industry, destroying the world's forests, and stealing resources from local communities. More than half of all tropical deforestation is illegal, eventually contributing towards the 1.5 gigatons of carbon resulting annually from deforestation (WWF). These illegal logging practices destroy the world's unique biodiversity and forests, including the Amazon, Central Africa, and Southeast Asia. Nearly 50-90% of tropical timber is illegally logged. The once net carbon sink of the Brazilian Amazon has been transformed into a carbon dioxide (CO2) producer due to deforestation, emitting 3x as much CO2 as New York City (Monitoring of the Andean Amazon Project (MAAP). These alarming statistics contributed to various countries doing their part in setting aside protected areas, and making logging practices illegal, though weak governance and corruption cause these laws to not be enforced. However, without the human resources, aid, or funding to adequately monitor these protected areas, these third-world countries struggle to protect their forests.

The advent of machine learning methods has allowed a wide range of stakeholders, including governments, NGOs, private sector actors, and communities, to recognize the value of satellite-based disturbance alert products in land management and law enforcement actions against illegal forest activities (Lynch et al., 2013; Finer et al. 2018, Weisse et al. 2019, Tabor and Holland 2020). Current land management systems rely on optical medium-scale resolution optical satellite data, namely Landsat. The Peruvian Geobosques stem (Vargas et al. 2019) and the Global Land Analysis and Discovery (GLAD) pan-tropical forest disturbance alerts (Hansen et al., 2016) use Landsat data to provide disturbance alerts. However, the crucial flaw of optical imagery is the limited availability of cloud-free data year-round, hindering the ability to track forest change events consistently and in real-time. The areas most prone to deforestation are in the humid tropics, which experience a very wet monsoon season, in which cloud-free observations could be several months apart.

Consequently, many selective logging occurrences go undetected due to smaller logging disturbance signs often disappearing within weeks due to the fast regeneration of understory vegetation. Other deforestation detection systems include the DETER project and the PRODES project, which both map deforestation and forest loss through Landsat data, though both require manual visual interpretation to reach needed accuracy levels. Small-scale selective logging disturbances also require higher resolution satellite data more significant than that provided by Landsat. Artificial Intelligence (AI) models such as the one by Stanford HAI Lab collected data on charcoal burning sites as pinpoints for deforestation sites. They trained the model on the data to detect charcoal-burning sites. However, if discovered to be a known driver for detection, patterns such as charcoal burning sights could incentivize loggers to change their logging methods to avoid fires or burning sites.

Additionally, although methods for clear-cut deforestation detection are well developed and identified with accuracies > 90% consistently in many alert systems, forest degradation or selective logging has been less intensely studied. Although it has no concrete definition, forest degradation can be summed up as various intensities of selective logging, small fires, fuelwood extraction, and more. Rather than clearance of entire forest or harvesting an entire clearing leaving barren land, most loggers opt for more subtle methods of harvesting only a small amount of trees, selecting the most profitable ones. In doing so, measures such as canopy cover, biodiversity, or carbon content are loosely preserved. However, though unable to be directly quantified, selective logging activities are an estimated significant source of carbon emissions and the most common form of forest degradation in the tropics. In order to detect this scale of forest degradation, this research seeks to develop a machine learning model using the newer data sets providing higher quality data than the coarse Landsat imagery.



(Courtesy of Getty Images and Mongabay respectively)

Figures 1.1 and 1.2 - Clear Cut Logging Example and Selective Logging Example

Radar sensors transmit their signals at a frequency that can penetrate clouds, offering greater timeliness and imagery at higher resolution. The European Space Agency's Sentinel 1A and 1 BC Band radar satellites were launched in 2014 and 2016 and offer high-resolution radar data at 10m spatial scale updating every 12 days over most

tropics. Recent studies have shown some success in accurately detecting forest disturbance of selective logging practices in near real-time. For example, Reiche et al., using only sentinel 1 data, produced a user (1- false detection rate) and producer accuracy (detection rate) of 97.6% and 95.0%, respectively, for forest disturbances more significant than or equal to 0.2 ha. However, when including samples representing disturbance events <0.2 ha, the producer's accuracy was 83.5%, indicating a high rate of fine-scale disturbance events which were not detected (Reiche et al.), showing room for improvement. However, radar doesn't capture quite everything, and optical still provides additional signals.

In order to improve the accuracy of deforestation monitoring, this study will integrate the use of both Sentinel 1 (radar) and Sentinel 2 (optical) imagery. The integration of sentinel 1 and 2 will also provide more timeliness since both operate on a 12-day orbit schedule, but combining the two can cut down to 6 days. Numerous algorithms have been tested for remote sensing image analysis, namely Convolutional Neural networks (CNNs), Deep Neural Networks (DNNs), Support Vector machines (SVMs), and random forests. SVMs provide high accuracy in fewer data training sets, whereas neural networks tend to perform better as the training input becomes larger. Ortega et al.'s evaluation of deep learning techniques found Siamese CNN to have greater accuracy than SVM. However, other factors arise, such as run time or cost of operation. A variety of factors affect the success of one model, so to find the optimal model for integrating radar and optical imagery, a variety of machine-learning models will be built and tested. The first is a convolutional neural network, a neural network that uses an architecture built upon convolution, convolutional, pooling, and fully connected layers that trigger activations similar to neurons to analyze visual imagery effectively. The second is the random forest, a supervised learning method that uses an ensemble-based decision tree framework that averages the decisions of various trees at the end. Gradient boosted trees will also be tested, which combine the decisions of the various "decision trees" throughout the process from the beginning instead of only at the end.

This project seeks to create an integrated approach that utilizes both radar and optical sensing data to create a detection algorithm that predicts 95% of identified logging (logging area > 0.1 ha) in real-time in the Amazon rainforest with greater accuracy and timeliness than current models offer, demonstrating the benefits of combining optical and radar imagery through testing various deep learning models ranging from CNN, Random Forest, and Gradient Boosted trees.

2 Methodology

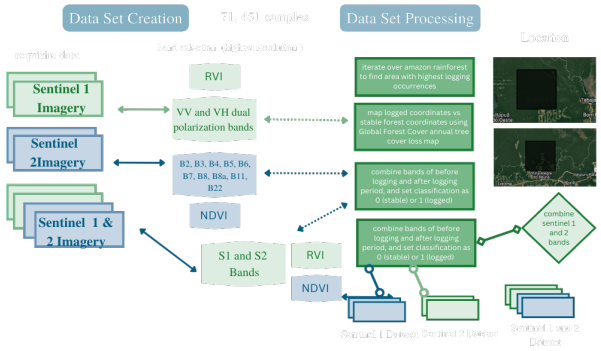


Figure 2.1 Methodology Chart for Data Processing/ Data Set Creation

2.1 Data Set Location / Forest Loss Events

Data Set Location Selection

The region selected was the Jamari National Park which has been used in previous studies (Kuck et al) for selective logging studies. The area is managed by the Brazilian forest ministry in sustainable logging practices, using selective logging rather than clear cut deforestation. Hence, the deforestation in this region is selective logging and can be directly compared to current research as it is using same data region. Kuck et Al had access to logging records to verify selective logging sizes, I however did not, but by selecting same region, can classify my data points as selective logging.

2.2 Data Pre-Processing

Sentinel 1 Data Processing:

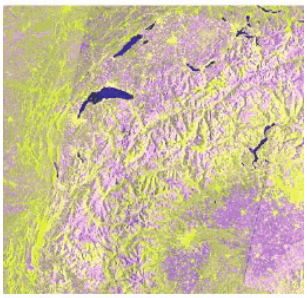


Image Credited to Google Earth Engine

Sentinel-1 GRD images available in the Google Earth Engine collection (Google Earth Engine 2020) have already undergone pre-processing of orbit, thermal noise and GRD border noise removal, radiometric calibration to sigma naught, and range-Doppler terrain correction.

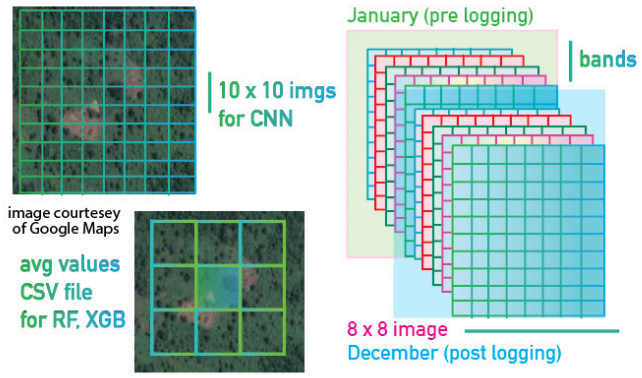
Sentinel 2 Data Processing:



Image Credited to Google Earth Engine

The Sentinel-2 satellite uses a multi-spectral optical sensor to offer an array of image resolution ranging 10 m, 20 m and 60 m spatial resolution. The 10 bands at 10m or 20m resolution were selected: B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12. No additional preprocessing was done to sentinel 2 imagery obtained from Google Earth Engine.

Data Set Creation for Random Forest/ Gradient Boost:



Using the global forest cover map, logged pixels in the region were identified and a subset of the stable forest pixels were also selected to have a dataset with around equivalent amounts of logged and stable to prevent the model from becoming biased towards one classification. Afterwards, a 3 x 3 region around the logged pixel was selected and averages of the various S1 band values were taken for the 9 pixels. The same process was repeated for S2 band values. This process was repeated for both the December and January data to contrast the differences in band values before and after logging. Afterwards, the January and December imagery and accompanying band values were combined using concatenation. For the combined Sentinel 1 and 2 datasets, the sentinel 1 and 2 imagery was combined as well.

Three different data sets were produced: Sentinel 1, Sentinel 2, and Sentinel 1 & 2 combined. The data sets had 134. 606 samples and compromised 4 band values (VV December, VH December, VV January, VH January for sentinel 1 and 20 band values for Sentinel 2 (B2 Dec, B2 Jan, B3 Dec, B3 Jan, B4 Dec, B4 Jan, B5 Dec, B5 Jan, B6 Dec, B6 Jan, B7 Dec, B7 Jan, B8 Dec, B8 Jan, B8A Dec, B8A Jan, B11 Dec, B11 Jan, B12 Dec B12 Jan) and 24 band values for the combined Sen 1 and Sen 2 data set.

Sentinel 1 Features

	Resolution
VH	10m
VV	10m

Sentinel 2 Features

	Resolution	Wavelength	Description
B2	10m	443.9nm (S2A) / 442.3nm (S2B)	Blue
B3	10m	496.6nm (S2A) / 492.1nm (S2B)	Green
B4	10m	560nm (S2A) / 559nm (S2B)	Red
B5	20m	664.5nm (S2A) / 665nm (S2B)	Red Edge 1
B6	20m	703.9nm (S2A) / 703.8nm (S2B)	Red Edge 2
B7	20m	740.2nm (S2A) / 739.1nm (S2B)	Red Edge 3
B8	10m	782.5nm (S2A) / 779.7nm (S2B)	NIR(Near Infra-red)
B8A	20m	835.1nm (S2A) / 833nm (S2B)	Red Edge 4
B11	20m	864.8nm (S2A) / 864nm (S2B)	SWIR 1
B12	20m	945nm (S2A) / 943.2nm (S2B)	SWIR 2

In addition, to accompany the band values NDVI for Sentinel 2 imagery and RVI calculations for Sentinel 1 imagery were calculated and used as additional data values accompanying the band values. NDVI is a quantified measure of vegetation using the difference between near infrared which vegetation reflects back and red light which the vegetation absorbs. Meanwhile RVI is similar, but using radar data and measures the scattering randomness of vegetation.

NDVI is calculated as follows:

RVI is calculated as follows :

σVH and σVV represent different measured linear backscattering intensities

classifcation	VVdec	VHdec	VVjan	VHjan	B2dec	B3dec	B4dec	B5dec	B6dec	B7dec	B8dec
1.0	1733.111111111110	1512.111111111110	1285.5	1458.8333333333300	2799.222222222220	3569.3888888888900	3470.5	4001.722222222220	2189.3333333333300	1288.888888888890	1143.888888888890
1.0	1761.3333333333300	1603.388888888890	1484.0555555555600	1603.3333333333300	2481.3333333333300	3005.222222222220	2851.4444444444400	3347.1666666666700	2510.6666666666700	1567.7777777777800	1077.388888888890
1.0	2098.5555555555600	1926.2777777777800	1752.3333333333300	1839.6666666666700	2720.1111111111100	3304.1111111111100	3252.777777777780	3640.222222222220	2521.3333333333300	1588.7777777777800	1020.8888888888900
1.0	970.7777777777780	813.0	545.7777777777780	869.7777777777780	2085.777777777780	2667.8888888888900	2517.1111111111100	3103.6666666666700	1877.3333333333300	773.4444444444450	980.1111111111110
1.0	927.5555555555560	723.0	422.8888888888890	743.1111111111110	2385.5555555555600	3253.4444444444400	3062.8888888888900	3725.6666666666700	1579.2222222222200	536.6666666666670	1029.0
1.0	962.8888888888890	723.4444444444450	454.7777777777800	733.0	2328.777777777780	3208.777777777780	3022.0	3649.4444444444400	1611.0	561.6666666666670	961.0
1.0	1096.6666666666700	926.8888888888890	754.7777777777780	1067.6666666666700	1882.0	2437.777777777780	2488.8888888888900	2925.4444444444400	2083.8888888888900	1082.3333333333300	949.6111111111110
1.0	1483.888888888890	1140.111111111110	1005.5555555555600	1167.6666666666700	1762.0	2252.0	2372.5555555555600	2749.222222222220	2630.3333333333300	1451.2222222222200	976.9444444444450
1.0	1469.111111111110	1300.0	1152.111111111110	1265.7777777777800	1893.2222222222200	2350.6666666666700	2568.3333333333300	2706.5555555555600	2477.3333333333300	1504.5555555555600	893.9444444444450

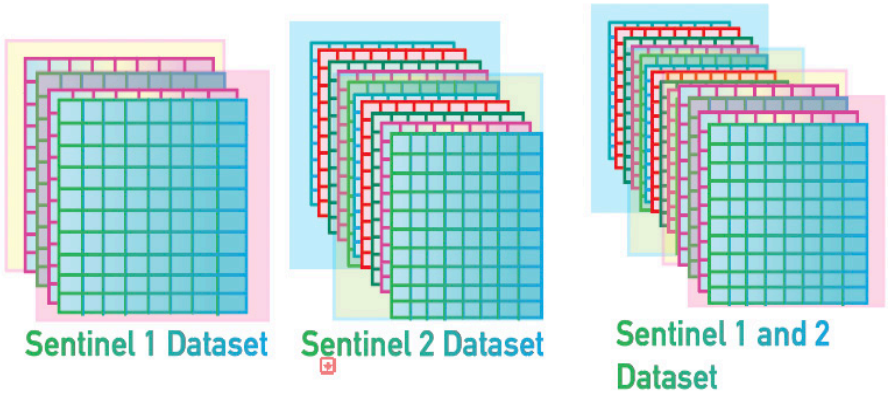
B8dec	B8Adec	B11dec	B12dec	B2jan	B3jan	B4jan	B5jan	B6jan	B7jan	B8jan	B8Ajan
1143.888888888890	979.8888888888890	956.6666666666670	998.7777777777780	1331.7777777777800	1561.888888888890	1453.5555555555600	1813.5555555555600	2726.6666666666700	1912.2222222222200	-8.031335079499840	-14.36777957272390
1077.388888888890	1058.7222222222200	953.4444444444450	1248.111111111110	2280.3333333333300	2866.5555555555600	2608.277777777780	3277.3888888888900	2562.222222222220	1262.7222222222200	-7.380337953360340	-15.11890245544130
1020.8888888888900	871.8333333333330	791.7777777777780	897.4444444444450	1359.2222222222200	1663.9444444444400	1566.7777777777800	1945.7777777777800	2354.6666666666700	1396.5555555555600	-5.693484920394210	-11.61737398108999
980.1111111111110	839.4444444444450	692.5555555555560	876.7777777777780	1594.111111111110	1999.2222222222200	1756.4444444444400	2269.9444444444400	2228.722222222220	1288.7222222222200	-9.298665344980830	-15.58982191105930
1029.0	857.1666666666670	802.7777777777780	868.8333333333330	1107.2777777777800	1324.5555555555600	1221.4444444444400	1527.7777777777800	2467.222222222220	1732.9444444444400	-7.417905053348420	-13.79415711903270
961.0	776.8888888888890	671.2777777777780	748.2222222222220	1188.0555555555600	1445.3333333333300	1230.2777777777800	1688.0	1973.8333333333300	1185.8333333333300	-5.900054616091620	-13.59145849021770
949.6111111111110	772.6111111111110	628.7222222222220	718.0	1226.611111111110	1550.0	1447.6666666666700	1741.4444444444400	1822.888888888890	1167.5555555555600	-8.006683738099020	-15.4202522363617
976.9444444444450	811.0555555555560	642.0555555555560	748.9444444444450	1356.4444444444400	1696.611111111110	1642.2222222222200	1879.4444444444400	1724.0555555555600	1100.888888888890	-9.584315973429980	-14.80273835238510
893.9444444444450	819.7222222222220	501.6666666666670	868.4444444444450	2664.8333333333300	3655.722222222220	3468.222222222220	4040.3888888888900	1992.7777777777800	870.1111111111110	-9.405793906724950	-17.51330165824630
1899.388888888890	1784.8333333333300	1713.2222222222200	1808.388888888890	2608.5555555555600	3117.722222222220	2826.3333333333300	3433.6111111111100	2789.277777777780	1843.9444444444400	-8.778563429401390	-15.03112866341110
1398.6666666666700	1197.6666666666700	992.9444444444450	1153.6666666666700	2005.0555555555600	2518.8333333333300	2359.0	2797.8888888888900	1989.4444444444400	1054.7222222222200	-9.613349662646700	-15.94046891339550
2270.722222222220	2135.722222222220	2055.0	2105.8888888888900	2876.277777777780	3358.5555555555600	3178.4444444444400	3662.3333333333300	2907.6111111111100	1846.1666666666700	-6.767975734356340	-10.73410436304180

Figure 2.2 - Example slice of data points (CSV file- comma separated file) of averaged data values for Random Forest and XGB models data.

Data Set Creation for CNN:

Unlike the Random Forest and Gradient boosted models which took the averages of the band values, CNN is a deep learning model which uses imagery (typically as 3d array) as input. Using the global forest cover map, logged pixels in the region were once again identified and a subset of the stable forest pixels were also selected to have a dataset with around equivalent amounts of logged and stable to prevent the model from becoming biased towards one classification. Afterwards, a 10 x 10 pixel slice/ region of the Sentinel 1 and 2 imagery around the logged/ stable forest pixel was selected. To keep data sets nearly identical, to count as “logged” at least 3 pixels needed to be logged in a 2 x 2 region of the GFC dataset, as single logged pixels in the GFC dataset are often incorrect. This process was repeated for both the December and January data to contrast the differences in band values before and after logging and the 10 x 10 slice images were combined to have two layers for December and January. In addition, for the combined Sentinel 1 and 2 datasets, the sentinel 1 and 2 imagery was combined as well resulting in a dataset of 134606 10 x 10 pixel images.

Note, all data set creating processes are nearly identical for both types of models in order to eliminate biases and attribute changes in accuracy to the model itself and not data.



2.3 Machine Learning Methods

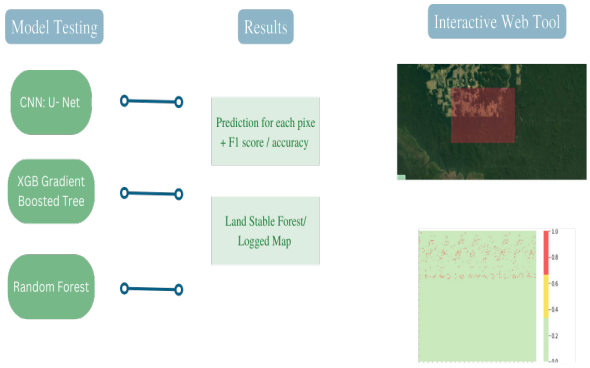


Figure 2.3 Methodology Chart for Model Testing/ Web Tool

The various models explored will be CNN (U- Net), Random Forest, Gradient Boosted Trees. The models will be built using python libraries and trained and tested on the Sentinel 1, Sentinel 2, and Sentinel 1 and 2 datasets.

CNN adaption of U- Net Architecture:

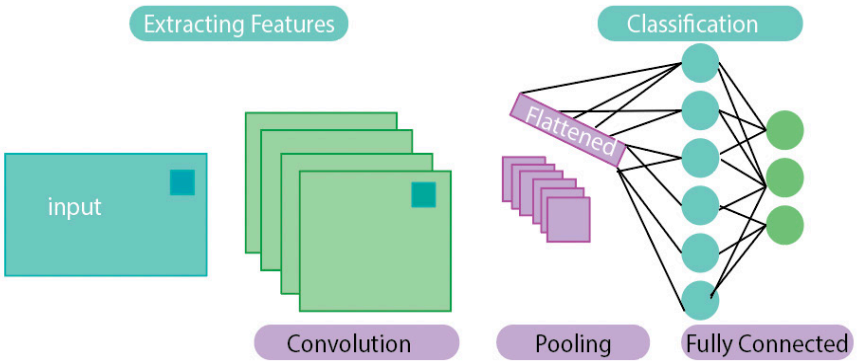


Figure 2.4 Convolutional Neural Network Diagram

The U-Net architecture [Ronneberger et al., 2015] was originally designed for segmentation of biomedical images using little training data. The U-Net is a fully convolutional neural network which can take in input images of any sizes. The U net architecture is able to do semantic segmentation using repeated application of 3 x 3 unpadded convolutions, batch normalization, and a 2 x 2 max pooling operation followed by doubling feature channels. An adaption of the U-Net architecture will be applied to all data sets.

A PyTorch implementation of the U-Net for image semantic segmentation with high quality images was implemented using https://github.com/milesial/Pytorch-UNet. However, this implementation was intended for larger images and thus reduced the image size past it's 10 x 10 size. The U – Net model was modified to remove one of the down sampling layers and one of the up-sampling layers.

The main hyperparameters for tuning is the learning rate and batch size. Higher learning rate is optimized for smaller batch sizes and vice versa. The model's performed best loss and accuracy wise when using a batch size of 64, and learning rate of 1.1e-4. The model also performed better under an Adam optimizer compared to gradient descent with the loss function as Cross Entropy Loss.

Random Forest:

The random forest model will be built using the scikit-learn machine learning library, a model that uses an ensemble method to identify whether the pixel was logged or not based on the surface reflectance values. Random forest models construct multiple decision trees (similar to a forest) and each tree casts a vote on which class the input data should be labeled as.

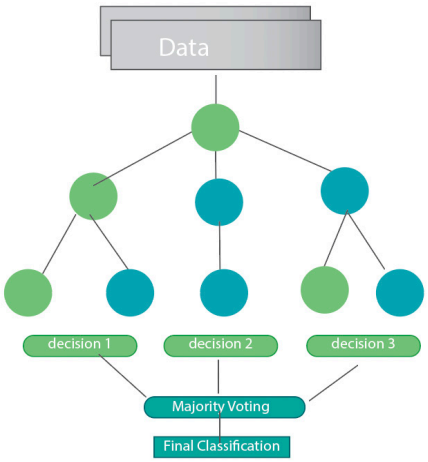


Figure 2.5 Random Forest Diagram

Gradient Boosted Tree:

Gradient boosted trees are similar to random forest in that they're both ensemble methods, combining decision trees. The difference lies in random forests building independent decision trees and combining them at the end whereas gradient boosted trees use a method called boosting to learn from the errors of the past trees. The Gradient boosted tree used an implementation of XGBBoost, a popular implementation known for its fast training time and high performance capabilities due it's advanced regularization features, improving the models generalization abilities.

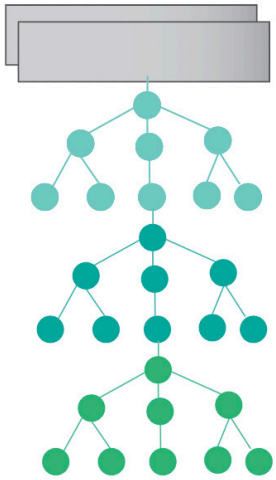


Figure 2.6 Gradient Boosted Tree Diagram

The gradient boosted tree models have a variety of parameters including learning rate, loss function, number of boosting rounds, and max depth of the tree created, The model performed the best under log loss, a max depth of 7, learning rate of 0.1, 400 rounds, cross entropy loss/ log loss function.

The training and testing data sets were split 90-10. In addition, small regions were selected for use as testing locations for logged /stable forest prediction maps. Using MATLAB and seaborn library, the prediction for each pixel was used to create map, see below.

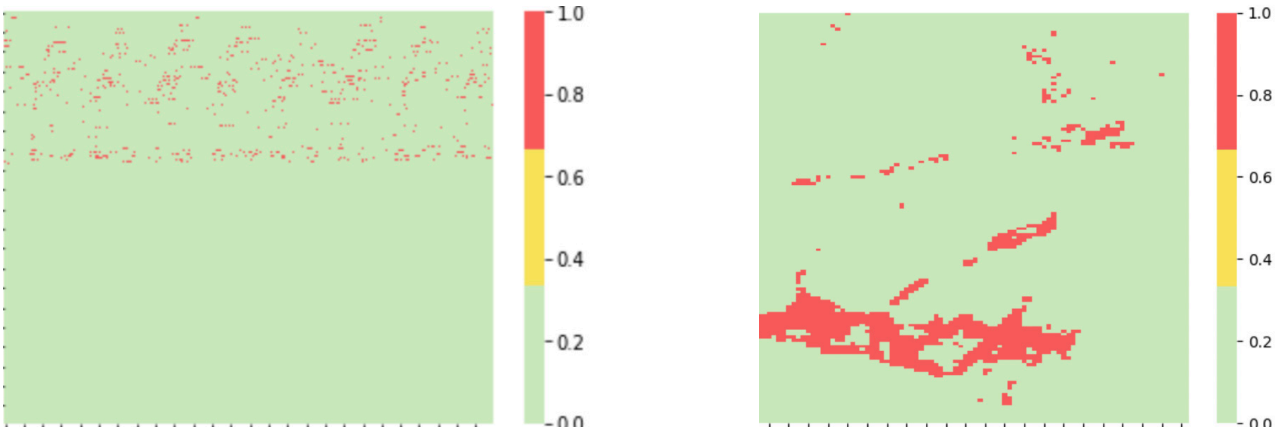


Figure 2.4 and 2.6 - logging maps for two different regions (red indicates logged, green indicates stable forest)

2.3 Model Performance in non – Amazon Region.

The model was trained on 71,451 samples from the amazon region of the Rondônia forest, which is the area where most logging occurs and highly applicable region. However, to prove the models can be used worldwide, the models will be validated on a region in the Congo forest to demonstrate worldwide applicability.

2.4 Webtool for Logging Detection for anywhere in the world

A webtool has been created in addition for forest protection agencies to use as an interactive tool to detect logging occurrences for any location in the world. In order to achieve this tool, a website wwascreated using HTML, CSS, JS with Node and Python for the backend.

First using the Google Maps open-source API, a Google Maps map will be integrated into the website. Using various google maps markers, users will be able to select 4 different coordinates. Once they've selected 4 different coordinates, this will output a layer map of the region the users have selected. In addition, users can delete coordinates or redo their region selections. Once they have selected 4 coordinates, the saved coordinates are used to send a request to node (backend). The node server uses the 4 passed in coordinates to select a region using the Google earth engine API and then select the appropriate bands for two different time frames (12 days before and after). The node program starts an export task for the google earth engine imagery which then is uploaded to cloud bucket after around 20 minutes of processing time. Once the imagery is in the bucket, using a webhook to detect the loaded imagery, the data processing program runs using a call to flask and uploads a csv file of the data to the cloud bucket. Finally, another flask program runs the csv file using a trained XGB model and outputs a logged map to which the user can download.

3.1 Data Evaluation

The data ia analyzed using each type of ML model (random forest, CNN, gradient boosted tree) and then calculated on accuracy, precision, recall, and F1 score for each type of model.

TP = True positives, i.e. the number of deforested areas classified as deforested.

TN = True negatives, i.e. the number of forested areas classified as forested.

FP = False positives, i.e. the number of forested areas classified as deforested.

FN = False negatives, i.e. the number of deforested areas classified as forested.

Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
Recall = TP / (TP + FN)
F1 = 2 x (Precision x Recall) / (Precision + Recall)

Figure 3.1 Tested Models' Accuracy, Precision, Recall, F1

Models Sentinel 1	Accuracy	Precision	Recall	F1
CNN	78.2			
Random Forest	57.52			
Gradient boost	58.11			

Models Sentinel 2	Accuracy
CNN	91.95
Random Forest	86.91
Gradient boost	90.40

Models Sentinel 1 and 2	Accuracy	Precision	Recall	F1
CNN	95.08	93.32	96.28	94.73
Random Forest	92.46	92.64	91.09	92.46
Gradient Boost	94.39	93.58	94.49	94.03

The CNN typically performed the best compared to XGB and Gradient boost in terms of all performance metrics (accuracy, precision, recall, and F1). The XGB is close in terms of performance, and less computationally intensive and was thus selected for use in the web tool.

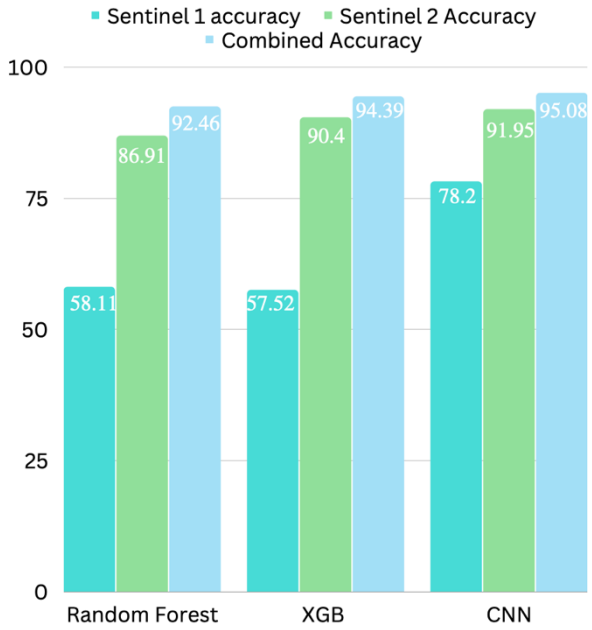


Figure 3.2 Integrated vs Stand Alone Accuracy For All tested Models

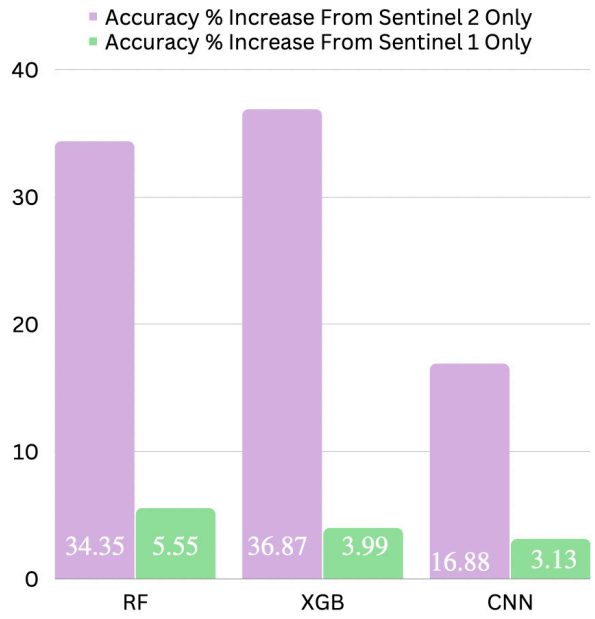


Figure 3.3 Accuracy Increase After Integration

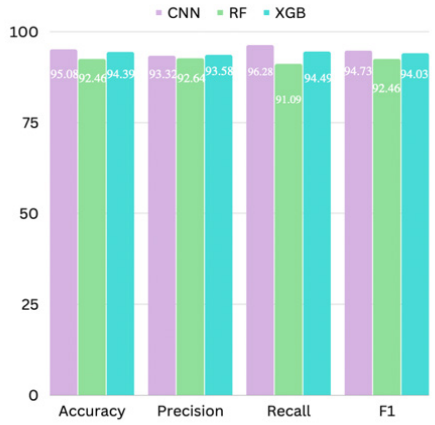


Figure 3.4 Overall Accuracy, Precision, Recall, F1 for all Models After Integration

Figure 3.1 – accuracy graphs for RF, XGB, CNN before and after sentinel 1 and 2 integration

The results show that the integration of sentinel 1 and 2 RF resulted in an accuracy increase of 5.55% for the Random Forest model, 3.99% for the XGB model, and 3.13% for the CNN model.t



Validation in Non- Amazon Region

A small region in Congo was used to test the XGB Model which is the model used for the web tool, in a region different to what it's trained on to demonstrate worldwide applicability. The validation resulted in an accuracy of 87.51, similar to the XGB model's accuracy on the amazon of 90.79.

4.1 Web tool

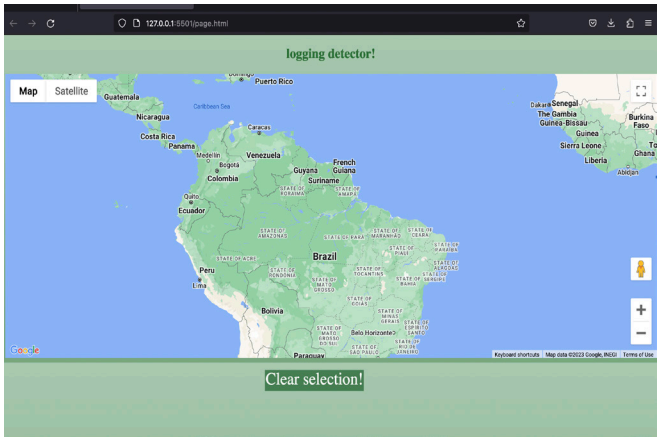


Figure 4.1 – Google Map interface for users to zoom in and out for selecting location

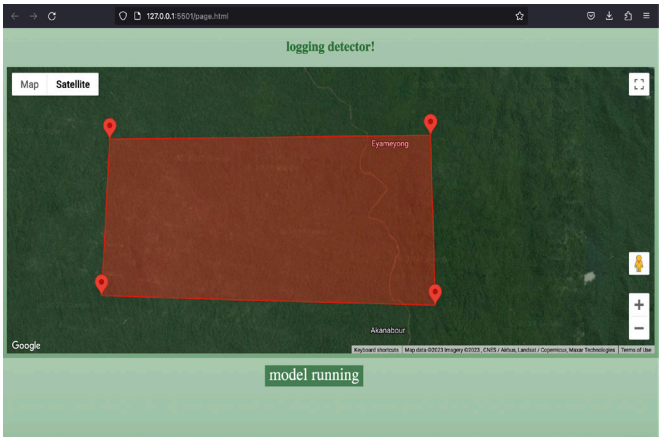


Figure 4.2 – after user has selected 4 coordinates, selected region displayed as a layer and "model running'



Figure 4.4 – after data has been acquired, processed, and tested logging map is ready to download

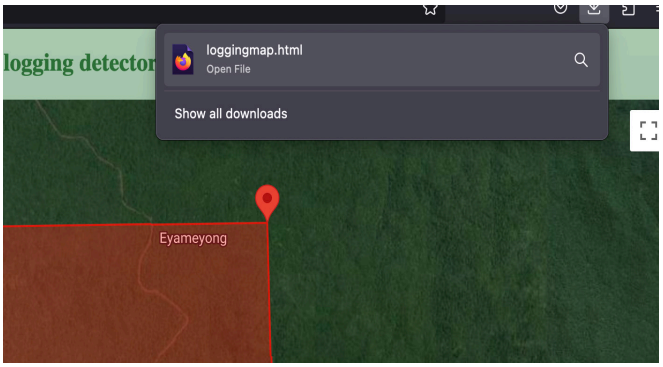


Figure 4.5 – map available in downloads

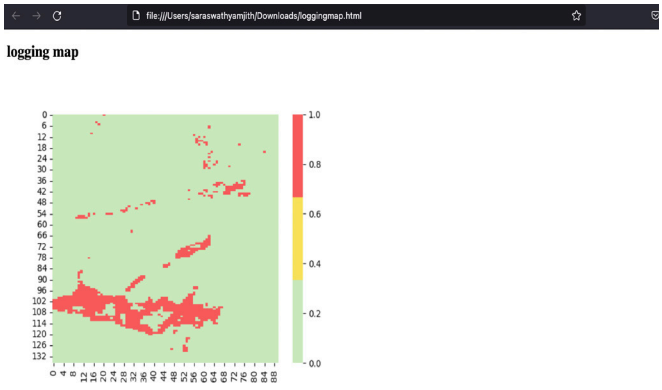


Figure 4.6 – logging map ready for users

5.1 Conclusion

The new integrated data set saw massive improvements compared to existing models hitting a breakthrough of 95.08%, significantly higher than the 88% recorded by previous researchers in this same region for selective logging (Kuck et al.). The previous standards were 90%, so that the increases can be attributed to the large data set and integration of multiple forms of data. The integration of optical and radar data through Sentinel 1 and 2 has the potential to produce promising advancements in selective logging deforestation detection. All current models only use one form of data or the other so that the integration can improve the existing model's accuracy by a minimum of 3% or greater. The SAR and optical bands complement each other and are advantageous in classifying logged vs. stable forests. Especially as illegal logging activities show an upward trend of decreasing in size but increasing in frequency, it becomes ever more significant to classify and detect occurrences accurately. The new detail and performance improvements of integrating sentinel 1 and 2 can be used to detect selective logging better and shows a promising future towards combatting selective logging. The various methods tested for machine learning revealed that a CNN architecture typically performs the best, specifically a U-Net implementation. In addition, the interactive web tool created can aid forest protection agencies by giving them a hands-on tool to detect logging occurrences without having to do the arduous, lengthy, and quiet complex process of data collecting, processing, and testing. The web tool combines all the steps into an easy-to-use tool that can be used for any region worldwide to detect logging occurrences. In addition, the models demonstrated applicability worldwide rather than location specific to the area it was trained on, achieving an accuracy of 87.51 in the Congo region.

5.2 Future Work

Currently, the web tool is hosted on an Ip server rather than the Internet. Future Work includes hosting the website as a live website for anyone to access and use, though it requires additional funding for cloud hosting, Google Earth Engine used, and domain. The researcher had no access to logging records and hence could only rely on the Global Forest Change for classifications, which could result in errors. The researcher will contact the Brazilian Forest Ministry for access to logging records as a better classification system. The researcher will also reach out to forest protection agencies for feedback and testing of the web tool to make improvements and demonstrate functionality by contacting Amazon Trust, Rainforest Watch, and Rainforest Action Network in the future. The tool can also not differentiate illegal vs. legal logging (allowed vs. illegal logging). In the future, access to a location's level of security ex. (National park – no logging permitted or Mixed Use zone) can be used to classify illegal/ legal. In a mixed-use zone, logging might have a likeliness of being legal, whereas, in the national park, it is most definitely illegal.

6.1 Bibliography

de Araujo, M., Andrade, E., & Machida, F. (2021). Performance analysis of machine learning-based systems for detecting deforestation. 2021 XI Brazilian Symposium on Computing Systems Engineering (SBESC). https://doi.org/10.1109/sbesc53686.2021.9628252

de Bem, P. P., de Carvalho Junior, O. A., Fontes Guimarães, R., & Trancoso Gomes, R. A. (2020, March 11). Change detection of deforestation in the Brazilian amazon using landsat data and Convolutional Neural Networks. MDPI. Retrieved November 8, 2022, from https://www.mdpi.com/2072-

4292/12/6/901

Evaluation of deep learning techniques for deforestation detection in ... (n.d.). Retrieved November 9, 2022, from https://www.researchgate.net/profile/Mabel-Ortega-3/publication/335843495_EVALUATION_OF_DEEP_LEARNING_TECHNIQUES_FOR_DEFORESTATION_DETECTION_IN_THE_AMAZON_FOREST/links/5d977d0b299bf1c363f8cfb4/EVALUATION-OF-DEEP-LEARNING-TECHNIQUES-FOR-DEFORESTATION-DETECTION-IN-THE-AMAZON-FOREST.pdf

Forest disturbance alerts for the Congo Basin using sentinel-1. (n.d.). Retrieved November 28, 2022, from https://iopscience.iop.org/article/10.1088/1748-9326/abd0a8

Hethcoat, M. G., Edwards, D. P., Carreiras, J. M. B., Bryant, R. G., França, F. M., & Quegan, S. (2018). A machine learning approach to map tropical selective logging. https://doi.org/10.1101/451856

Illegal logging and deforestation. U.S. Agency for International Development. (2022, February 8). Retrieved November 27, 2022, from https://www.usaid.gov/biodiversity/illegal-logging-and-deforestation

MAAP #144: The amazon & climate change: Carbon sink vs carbon source. MAAP. (2021, August 30). Retrieved November 27, 2022, from https://www.maaproject.org/2021/amazon-carbon-flux/

Mayfield, H., Smith, C., Gallagher, M., & Hockings, M. (2016, October 29). Use of freely available datasets and machine learning methods in predicting deforestation. Environmental Modelling & Software. Retrieved November 8, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S1364815216308428

Pacheco-Pascagaza, A. M., Gou, Y., Louis, V., Roberts, J. F., Rodríguez-Veiga, P., da Conceição Bispo, P., Espírito-Santo, F. D. B., Robb, C., Upton, C., Galindo, G., Cabrera, E., Pachón Cendales, I. P., Castillo Santiago, M. A., Carrillo Negrete, O., Meneses, C., Iñiguez, M., & Balzter, H. (2022, February 2). Near real-time change detection system using sentinel-2 and machine learning: A test for Mexican and Colombian forests. MDPI. Retrieved November 8, 2022, from https://www.mdpi.com/2072-4292/14/

Illegal logging and deforestation. U.S. Agency for International Development. (2022, February 8). Retrieved November 27, 2022, from https://www.usaid.gov/biodiversity/illegal-logging-and-deforestation

Butler, R. A. (2015, April 28). Selective logging leaves more dead wood in rainforests. Mongabay Environmental News. Retrieved March 2, 2023, from https://news.mongabay.com/2015/04/selective-logging-leaves-more-dead-wood-in-rainforests/

Global Forest Change. Google Crisis Map. (n.d.). Retrieved March 2, 2023, from

https://earthenginepartners.appspot.com/science-2013-global-forest

Google. (n.d.). Sentinel collections in Earth engine | Earth Engine Data catalog | google developers. Google. Retrieved March 2, 2023, from https://developers.google.com/earth-engine/datasets/catalog/sentinel

Photography, T. (n.d.). Ariel View of trees being cut down in the Pine Forest Stock Photo. Getty Images. Retrieved March 4, 2023, from https://www.gettyimages.com/photos/cut-down-forest

BioPGAN a Novel Way of Generating Novel Bioplastics Using Generative Adversarial Networks

Yutaro Urata

Tesla STEM High School

Abstract

The environmental impacts that result from the creation of plastics, starting at drilling for oil and ending with plastic degradation are detrimental to the environment. An alternative to these non-renewable bioplastics must be found to stop further pollution in our environment. In this project, a novel way of discovering new renewable and easy to degrade polymers was proposed; BioPGAN is a GAN (Generative Adversarial Network) to generate novel molecules. A GAN (a machine learning model that can create unique outputs that mimic the input data) was implemented to accomplish this goal. The training data consisted of organic (carbon based), which the model trained on first; and polymers (organic but chain-based molecules) which the model was transferred over to. Once trained on the more than 1 million organic molecules and 100 polymers, the model can generate multiple molecules with low formation energies (stable); Additionally, when 10,000 molecules were generated using the BioPGAN, 98.87% of the molecules were novel (not in the dataset) and included multiple promising bioplastic candidates that could potentially replace conventional plastics. This project shows a novel and efficient way of discovering novel bioplastics, allowing material scientists to search for alternative materials more efficiently than before.

Introduction

A Generative Adversarial Network (hereon referred to as GAN Fig 5) is used to create novel bioplastics. Every year, the world has a larger and larger problem as millions of tons of plastic are produced, and only 1-2% of the plastic are biopolymers that are made from biomass. The rest of the plastics are produced from petroleum, a nonrenewable source. These plastics are making their way into the ocean. These plastics are ruining fish habitats, suffocating animals, and ruining large marine ecosystems. When larger pieces of plastics degrade in the ocean, they leave behind what is known as microplastics. These plastics end up in fish, vegetables, and water that we use on a daily basis (WHO 2022). What makes this problem even worse is the fact that microplastics can build up inside of the body of humans and other organisms. There is no easy way of solving this, but scientists have come up with a idea which involves creating plastics that can be degraded by natural processes such as light, water, microbes, etc. These new types of plastics are known as bioplastics. However, not all bioplastics can be made sustainably.

For example, bioplastics created out of corn require mass production of corn which uses large amounts of nitrogen-based fertilizers which seep into the water, causing algal blooms downstream of the river, resulting in the proliferation of dead zones. Additionally, bioplastics can be over 20-50% more expensive than their petroleum based non-biodegrading counterparts, making them unfavorable for mass production especially in developing countries. (phys.org 2017) To solve this problem, scientists have created a new family of bioplastics known as PHA/PHB/PHBV that try to solve this issue by making GMO bacteria produce bioplastics. The rationale behind this is the fact that bacteria populations can be multiplied with small investments of time and cost. However, the main problem that these scientists discovered when creating these plastics is the fact that the plastics that these bacteria produce are not too usable on their own, not having suitable properties for

manufacturing products. This is where copolymers come into play. Copolymers are a set of two monomers combined to form one polymer. Making copolymers out of already existing PHAs can result in different properties that are not achievable by the original PHAs. However, creating polymers one at a time and testing their properties manually requires time, effort, and money.

This paper aims to address the problem through the generation of new potential bioplastics and compare them to currently existing bioplastics to reduce the time it takes for a new bioplastic to be discovered using deep learning methods such as a GAN. This is a method that is commonly seen in drug discovery, however there are almost no examples of it being performed in bioplastic discovery. One of the few articles that attempt this uses a technique known as NLP (**N**atural **L**anguage **P**rocessing). This method however is not the best since the deep learning network may have been paying attention to something that is not relevant in the SMILES (**S**implified **M**olecular **L**ine **E**ntry **S**ystem) string therefore generating imitated molecule data on something that has nothing to do with bioplastics. This problem is made worse by the fact that SMILES is not 100% robust and some molecules that would be valid chemically become invalid. This is where the alternative chemical encoding method called SELFIES (**SELF**-referenc**I**ng **E**mbedded **S**trings) is used. This new method of encoding molecules is suited for generative models (Mario Krenn et al.) and ensures that the structure is encoded along with the chemical elements. This process is achieved using a process known as one-hot encoding. Using this encoder also makes it impossible to get an invalid molecule. This method of encoding is a perfect candidate for this paper due to its robustness and ability to output encoded data as matrices. After the molecule is generated, the results must be validated to see if the molecules will be stable enough to be used in everyday applications. This will be done using a process known as DFT (Density functional theory) (Preferred Networks R&D). DFT is a mathematical process that predicts the properties of a molecule such as potential energy and the forces acting on each atom.

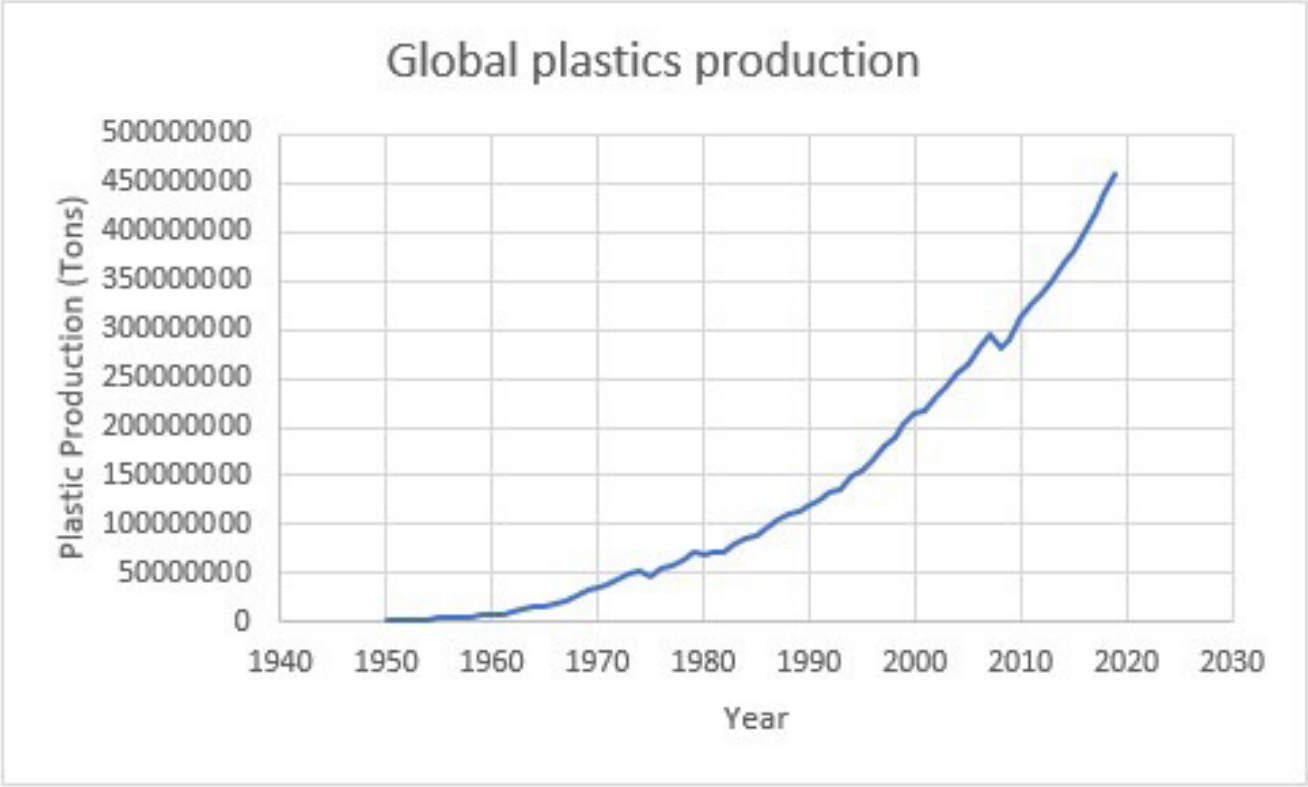


Fig 1 Global plastic Production Image Courtesy of Our World in Data CC-BY

Methodology

This project aims to propose novel bioplastics from preexisting bioplastics in hopes of creating novel bioplastics using machine learning. To do this, massive amounts of data about organic molecules, and molecules were required. To read these files, and convert them into readable formats, a few open-source datasets and libraries were used. The libraries that were used include RDKit; a python wrapper for a C++ library used for cheminformatics; Pytorch, a machine learning framework; SELFIES, a novel molecular encoding method that uses molecular graphs; Matplotlib, a python based graphing and plotting utility. The data used comes from a few databases. These databases include ZINC15; an organic compounds database which consists of approximately 230 million molecules; QM9, a database consisting of 133 thousand small organic molecules; and PI1M, a database consisting of 1 million polymers.

All of the data obtained from the datasets were encoded using the SLEFIES library into a machine readable, one-hot format (Fig 9), which is a form of categorical encoding. This was then loaded using a custom dataloader (Fig 2) from the pytorch tensor files they were saved to.

```
class customDataset(Dataset):
    def __init__(self,data_dir,transform=None):
        self.data_dir = data_dir
        self.transform = transform

    def __len__(self):
        return(len(os.listdir(self.data_dir)))

    def __getitem__(self, idx):
        tensor = torch.load(self.data_dir+str(idx)+".pt")
        if self.transform:
            tensor = self.transform(tensor)
        return tensor
```

Fig 2: The dataloader that was used to load the data.

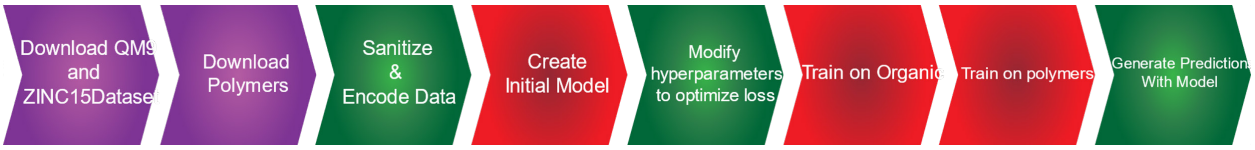


Fig 3 Flow chart of steps taken to train the model

All the data from the datasets were combined according to what type of data they were (organic molecules or polymers). An unsupervised machine learning method was used. The organic molecule data was first used to train the model to be able to generate molecules. After the results had been verified, the model was saved, then loaded again and trained with the data regarding polymers. BCE (Binary Cross Entropy) Loss was used to measure loss in this model (Fig 4).

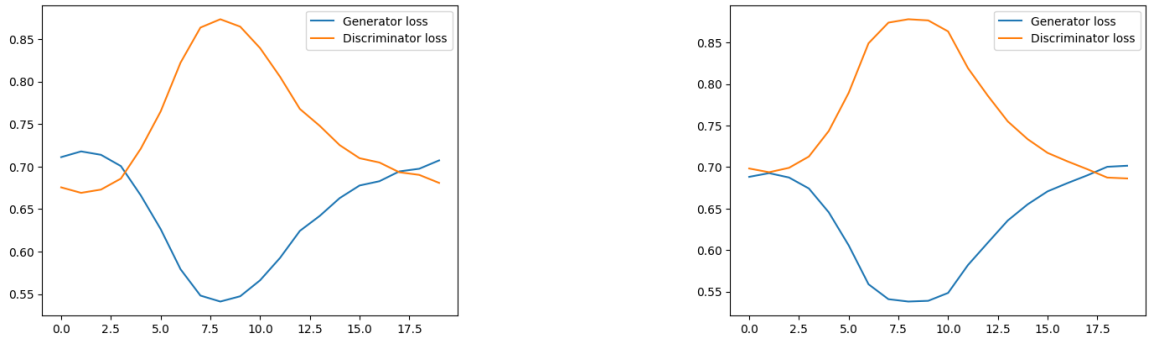


Figure 4 Loss graphs for when the model is trained at different learning rates (2e^-6 (left), 2e^-7(right))

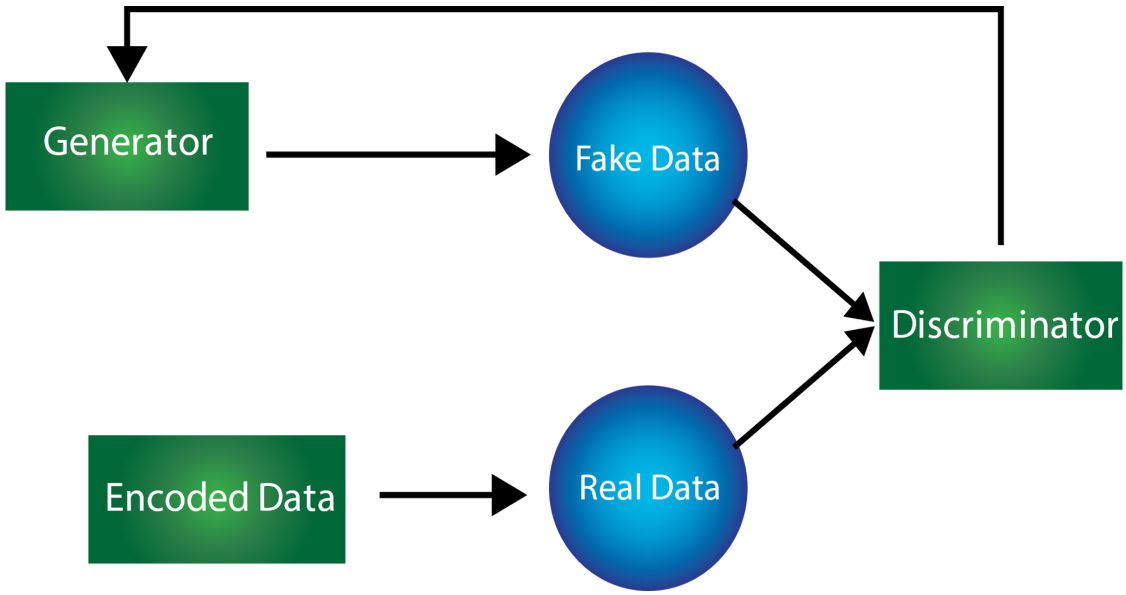


Fig 5 GAN Working principle

Datasets

Zinc 15: https://zinc15.docking.org/
QM9: http://quantum-machine.org/datasets/
PI1M: https://github.com/RUIMINMA1996/PI1M

Results and Data Analysis

10,000 molecules were generated using the trained model to verify that the model was outputting proper molecules. To do this, ElemNet was used to calculate the formation energies of each of the molecules that were generated. This was used to validate that the molecules could be created and were stable after they were created. Next, the formation energy of 10,000 polymers in the polymer dataset was calculated. These two results were then graphed on a heat map (Fig 6) and the clustering was visually confirmed to make sure that the machine learning network viewed the chemical structure and reactivity of the molecules. To make sure that the molecules were novel, a search was run on the entire dataset for the generated molecules. This returned that 98.87% (Fig7) of the generated molecules were not in the dataset that was used to train the algorithm. The top 3 molecules were selected from the generated molecules based on formation energy, and they were visualized (Fig 8 exhibits a-c, original molecules Fig 9 exhibits a-c). Due to the theoretical nature of this project, a physical validation of the molecules must be done to verify their manufacturability, structural properties, and their chemical properties. Computational steps such as DFT can also be applied to make this process more efficient. After testing, these plastics can be used to replace conventional petroleum-based plastics. The results of this study can be improved in a few ways in subsequent studies. The number of molecules that were used can be increased to give the model a better reference for what a polymer is. Employ better molecular analysis methods to better predict the properties of the generated molecule, add more layers/test different machine learning methods.

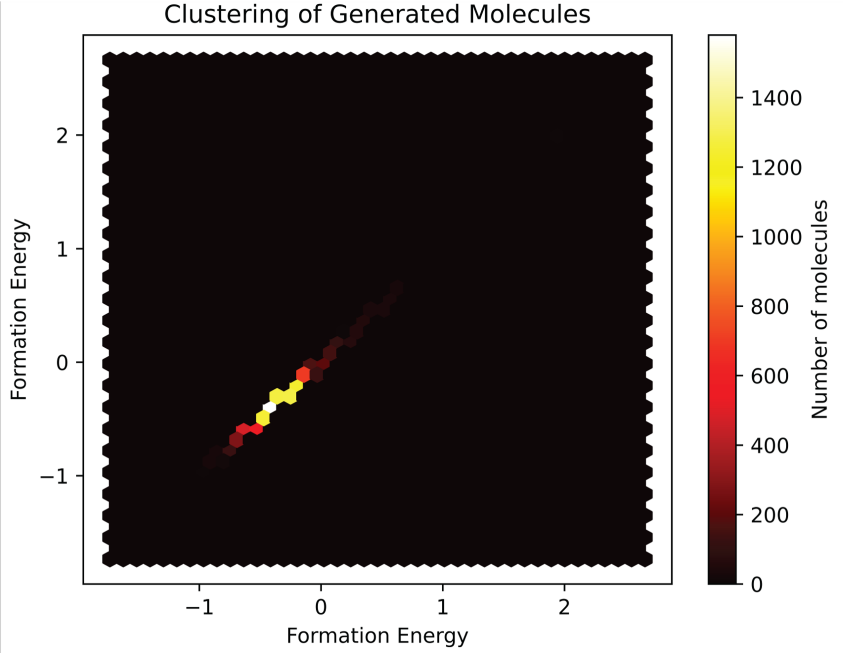


Fig 6 Clustering of the Generated Molecules in terms of formation energy

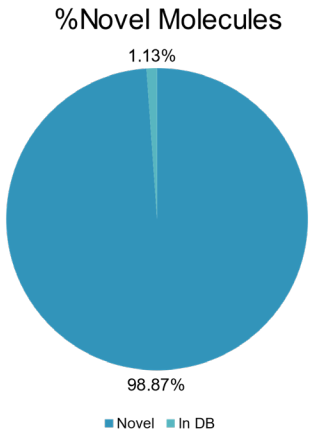


Fig 7 Percent of Novel Molecules

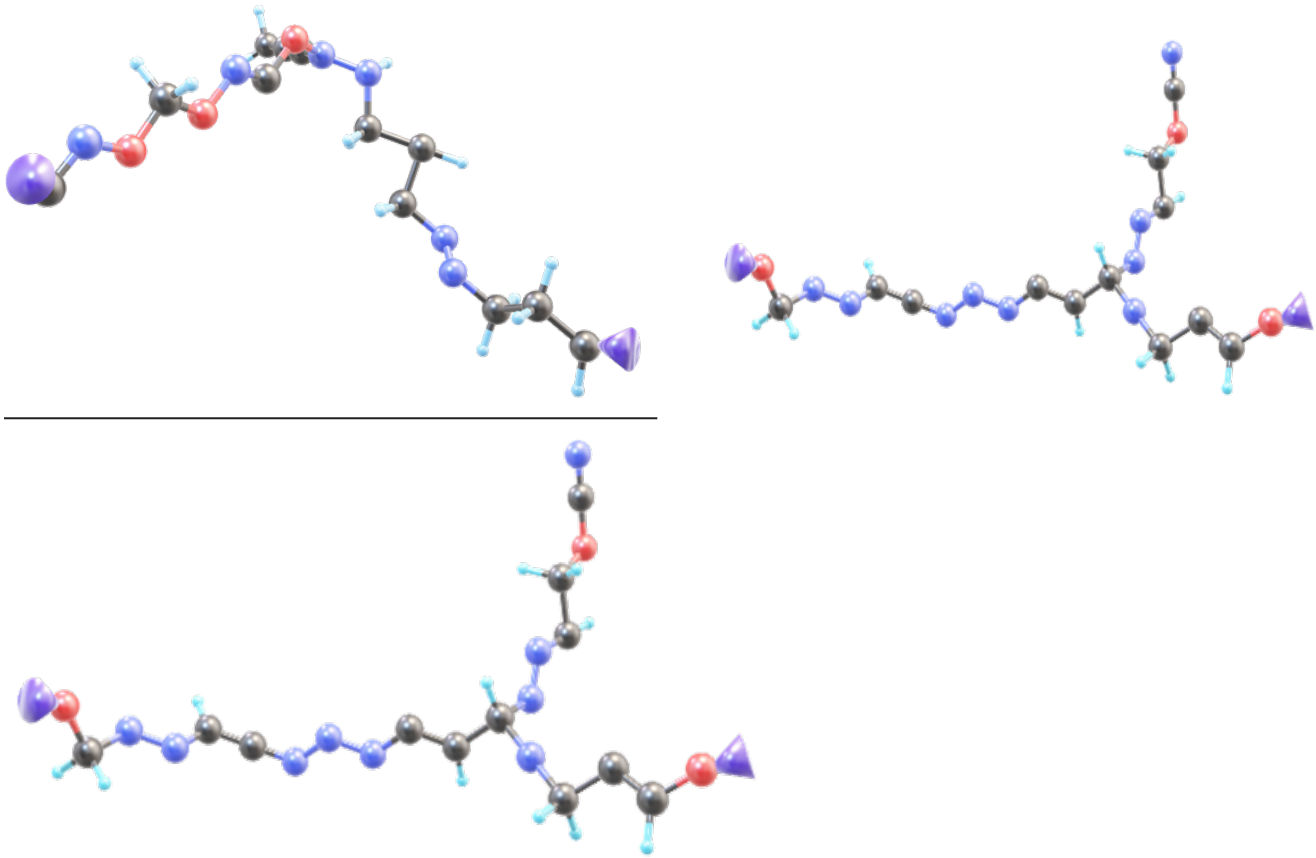


Fig 8 Exhibit a-c 3D visualizations of the generated molecules

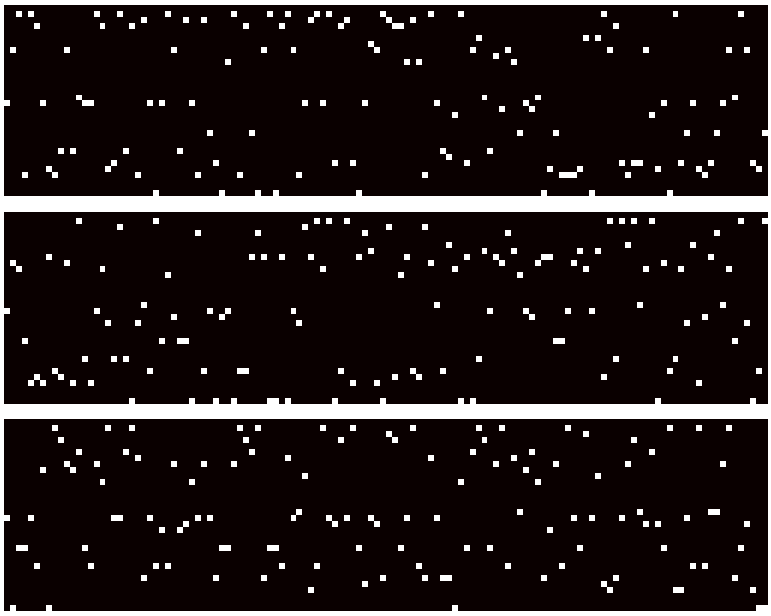


Fig 9 exhibits a-c Generated data as a one-hot format

Charts, Graphs

Number of Epochs	Accuracy %	Loss	Log P
5			
10			
15			
20			
30			
40			
50			
60			

Dataset name	Mean Potential Energy	Mean Forces	Mean Log P
QM9			
ZINC 15			
Bioplastics			

References

Cho, R. (2017, December 14). The truth about bioplastics. Phys.org. Retrieved April 24, 2023, from https://phys.org/news/2017-12-truth-bioplastics.html

Dietary and inhalation exposure to nano- and microplastic particles and potential implications for human health. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO

Dipendra Jha, Logan Ward, Arindam Paul, Wei-keng Liao, Alok Choudhary, Chris Wolverton, and Ankit Agrawal, "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition," Scientific Reports, 8, Article number: 17593 (2018) [DOI:10.1038/s41598-018-35934-y]

L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, J. Chem. Inf. Model. 52, 2864–2875, 2012.

R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Scientific Data 1, 140022, 2014.

Wu, S., Kondo, Y., Kakimoto, Ma. et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. npj Comput Mater 5, 66 (2019). https://doi.org/10.1038/s41524-019-0203-2

Lim, J., Ryu, S., Kim, J.W. et al. Molecular generative model based on conditional variational autoencoder for de novo molecular design. J Cheminform 10, 31 (2018). https://doi.org/10.1186/s13321-018-0286-7

Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik

ACS Central Science 2018 4 (2), 268-276

DOI: 10.1021/acscentsci.7b00572

Takamoto, S., Shinagawa, C., Motoki, D., Nakago, K., Li, W., Kurata, I., … Ibuka, T. (2022). Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements. Nature Communications, 13, 2991. doi:10.1038/s41467-022-30687-9

Using Crop Simulation Models and Machine Learning to Find Sustainable Alternative to Common Industrial Farming Practice

Arpan Agarwal

Nikola Tesla STEM High School

Abstract

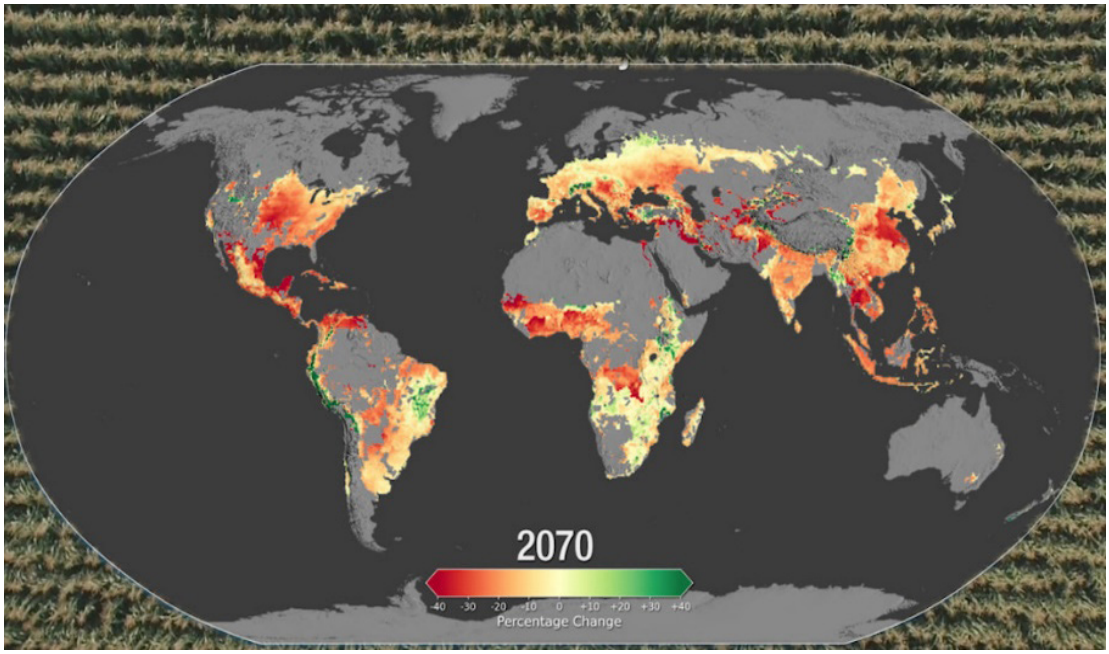
Under modern-day agricultural practices, climate change will reduce corn production in the Midwest USA by more than 20% over the next 30 years leading to devastating economic and social impacts. Therefore, farming must become more resilient towards climate change, and one approach is to abandon the monoculture cropping systems in favor of a sustainable alternative cropping system that preserves land, soil moisture content, and uses fewer nonrenewable resources. This study aims to develop a sustainable alternative cropping system for growing corn in the Midwest, USA by analyzing and comparing previously proposed practices. The study examined and compared the effects of 10 cropping sequences, 2 row spacings, and 5 nitrogen rates, resulting in the analysis of 100 distinct cropping systems. Data was collected from the Agricultural Production Systems sIMulator, which simulated 35 years of data on variables measuring the impact on soil quality and crop yield for each cropping system. The data was then analyzed using principal component analysis and data aggregation techniques. The results revealed that the most resilient cropping system was planting corn in a Maize-Winter Wheat cover cropping sequence, with a nitrogen rate of 225 kilograms/hectare, and a row space of either 37.5 or 75 centimeters. This cropping system helped the soil retain more moisture, while also having higher crop yields on average compared to the monoculture cropping system. The findings of this study should provide farmers with more clarity on how to make corn production more resilient towards climate change, a goal that must be achieved.

Introduction

This project seeks to find the best cropping sequence and management techniques to maximize the resiliency of corn production towards global warming impacts and to maximize the sustainability of corn production. According to NASA projections, annual corn production in the Midwest, USA – the highest corn producing region in the world – could decrease by 30-40% within the next 50 years, due to the effects of global warming. This is because global warming will increase the frequency of extreme weather causing a lot of the land in the Midwest, USA to become very dry and have a lack of moisture, which will significantly impact the production of corn. This reduction in corn production will have serious implications both economically and on the wellbeing of people as corn production is a $72 billion industry (USDA) in the United States and is relied upon for a multitude of other food and industrial products such as starch, sweeteners, corn oil, beverages, and alcohol. Ultimately the US economy will lose billions of dollars and many people will loose their jobs and occupations such as farmers and factory workers if corn production decreases according to NASA projections. Therefore, to prevent this there must be a change to many current farming practices for corn production. In addition to not being resilient towards global warming impacts, current farming practices for corn production are also highly unsustainable as according to Cornell University, 140 gallons of

fossil fuels are required to grow and produce just one acre of corn. As fossil fuels are a nonrenewable resource that exist in finite supply, this rate of consumption will lead to fossil fuels being fully depleted before they can be replenished. As agriculture as a whole greatly relies on fossil fuels, if fossil fuels are to be fully depleted then overall annual agricultural production will significantly drop, which will send the world in worldwide famine crisis due to the exponential population growth, which will lead to millions of people dying. Therefore, farming practices for corn production must change to make corn production more sustainable.

Figure 1: This is a map highlighting how climate change will reduce corn production in different parts of the world by 2070. For the Midwest, USA it is 30-40%

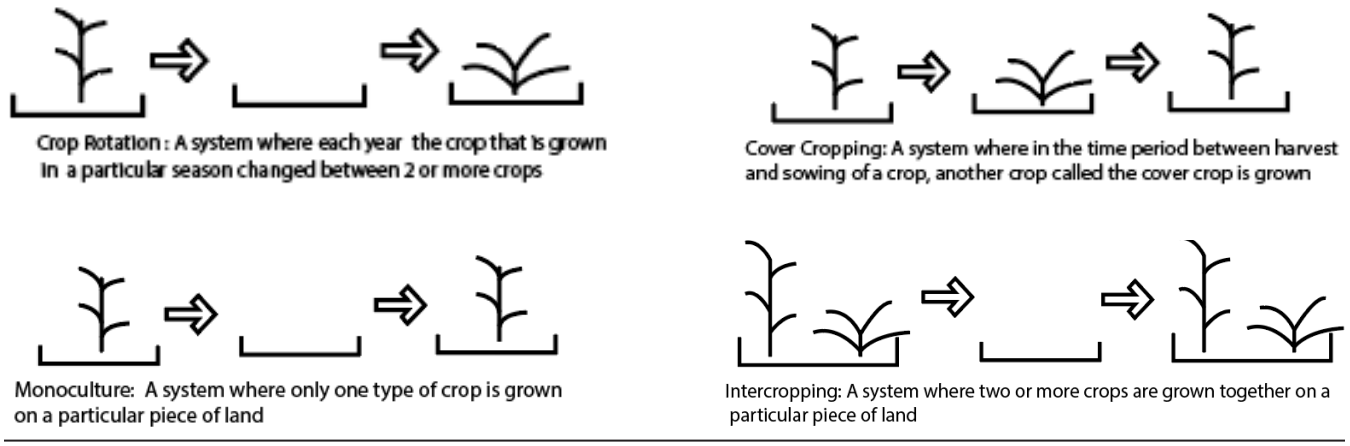


Speidel, J. J. (2007, July). Estimated World Population Growth: 1750 to 2050 [Chart]. Research Gate. https://www.researchgate.net/figure/Estimated-World-Population-Growth-1750-to-2050-20_fig1_225463481

Finding an alternative cropping sequence to the continuous corn (monoculture) cropping sequence would be a potential solution to make corn production more sustainable and resilient towards global warming impacts. The continuous corn cropping sequence is one that grows corn year after year on the same piece of land. The main problem with it is that it consumes a large amount of nitrogen fertilizer, which leads to a large amount of N2O emissions being emitted. Synthetic fertilizer also requires fossil fuels to be made, so consuming large amount of nitrogen fertilizers leads to more fossil fuels being consumed in the production of corn. Another big problem with continuous corn cropping is that it significantly depletes the quality of soil, since corn takes more nutrients and moisture out of the soil than it replenishes the soil with. This makes corn production more susceptible towards the impact of global warming because global warming will increase the frequency of extreme weather, and if the soil has a lack of moisture and nutrients it won't be able to withstand the impacts. Farmers also can't rely on continuously finding new farmland as land is a nonrenewable resource with 38.6% of worldwide land already being consumed in agriculture, according to National Geographic. The Earth can't afford to keep creating more land for agriculture. Therefore, to make corn production more sustainable and more resilient towards climate change it is imperative for farmers to stop using the continuous corn cropping sequence and for them to switch to a more sustainable alternative cropping sequence – one that preserves soil quality and reduces fossil fuel consumption.

Encouragingly, there has been a lot of research into finding sustainably alternative cropping sequences for producing corn in the Midwest, USA. The first main alternative cropping sequence is intercropping, which is the practice of simultaneity growing 2 or more crops with a synergic relationship are grown together, and it is practiced because it can help reduce demand for synthetic fertilizers and better preserve the soil quality (Fig. 2). The corn-soybean intercropping sequence in particular was found to emit significantly lower N2O emissions than the continuous corn cropping sequence (Fung et al., 2019). Other common intercropping sequences for corn are corn-lablab intercropping and corn-perennial grass intercropping. Next, there is crop rotation, which the practice of sequentially planting crops of a specific growing season in a particular plot, and it is practiced because it can also help significantly preserve the soil quality and demand use of fertilizer (Fig. 2). Common corn crop rotation systems are corn-soybean and corn-oat crop rotation. Next, there is cover cropping, which is like crop rotation, where except sequentially alternating crops every specific growing season, a cover crop is planted in between every growing season for corn, which is practiced preventing soil erosion and preserve soil quality (Fig. 2). Field peas are a common cover crop for growing corn, and it was found that using field peas as a cover crop helps the soil have a higher nitrogen content and carbon content (Lavergne et. al., 2021). Other common cover crops for corn are vetch, barley, faba beans, and winter wheat. Lastly there is the fallow system, where a specific plot of land is left arable every other year so it can recover, store organic matter, retain moisture, and disrupt the life cycle of pests. For management techniques, the first main one is nitrogen rate, which is the amount of nitrogen fertilizer applied on a field. It has been found that higher nitrogen rates such as 300 and 225 kilograms/hectare (Kg/Ha) can be better for getting high crop yields, but lower nitrogen rates such as 0 Kg/Ha and 65 Kg/Ha can be better for maintain the soil quality. There has also been research done into row spacings, which is the distance between rows of crops on a field. For example, Raeann Lyn Huffman from Illinois State University has stated that, "in narrow-planted rows, the rate of photosynthesis can increase due to majority of the sunlight being utilized by the canopy…assisting in an increase in yield potential" (Huffman, 2021). However, he has also stated that "when the row spacing is too narrow it can cause too much competition" (Huffman, 2021).

Figure 2: This is a diagram describing all the specific cropping sequences

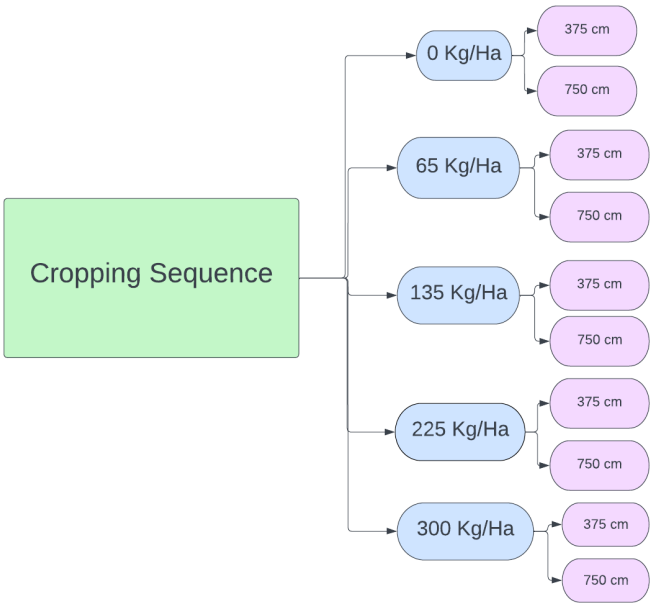


Despite all this research there has been hardly any research into comparing each of these specific cropping sequences and management techniques and developing/finding one clear specific cropping system – the specific cropping sequence and management techniques used on a piece of land – that is the best for growing corn in the Midwest, USA, considering factors such as growth productivity, environmental friendliness, and sustainability. Due to this, it is hard for farmers to maximize sustainability and global warming resilience of corn production since there is a lack of clarity over what cropping sequences and management techniques are the best for this purpose. This project sought to change this by evaluating 10 different cropping sequences, 5 specific nitrogen rates, and 2 row spacings commonly used for producing corn in the Midwest, USA, in order to develop one clear best alternative cropping system through using crop simulation models and machine learning. Findings for this project aim present Midwest farmers with a clear best alternative cropping system for producing corn and help make more farmers switch away from continuous cropping and maximize sustainability and global warming resilience of corn production. Overall, this project could be a big step forward towards the big goal of sustainable and resilient food production, which is a goal that the United States must achieve for the future of the economy, environment, and society in general.

Methodology

As discussed in the introduction there were 10 cropping sequences, 5 nitrogen rates, and 2 row spacings that were evaluated in this project. Each cropping system was composed of 1 cropping sequence, 1 nitrogen rate, and 1 row space put together. Figure 3, down below shows that for 1 cropping sequence, 5 nitrogen rates, and 2 row spacings there are 10 possible cropping systems. Since there are 10 cropping sequences evaluated, there were therefore 100 distinct cropping systems that this project analyzed. To account for the cofounding variable of weather, each cropping system was simulated 3 times per crop model, each time under a different weather data file. To simulate each scenario and collect data the APSIM (Agriculture Production Systems sIMulator) was used, an internationally recognized platform for modeling and simulating agricultural systems. The model simulates biophysical processes in agricultural systems accurately predicts the crop production knowing climate, genotype, and soil. For this project, APISM710-r4420 version of the software was used. The DSSAT (Decision Support System for Agrotechnology Transfer) model was also used in this project, which is a model that simulated agricultural systems and collected data on the greenhouse gas emissions emitted throughout each system's simulation. For this project, the DSSAT Version 4.8 was used.

Figure 3: Diagram showing all the possible cropping systems for one cropping sequence.



For each cropping system, before they could be run and simulated, the models had to be parameterized accordingly. For parametrization the models had to be parametrized on data about the weather, the soil, and management rules – sowing rules, harvesting rules, fertilizer application, manure application, tillage – imposed for the specific cropping system being simulated. To parametrize the model DAYMET weather data files, was used where each file contained data about the daily weather in a specific region for the past 30 years. In this study, data about Ames, Iowa weather, northwest Iowa Data, and Southeast Iowa data were used. For the soil data, the USDA WBB Clarion Soil data file, was used which contains data about initial water, soil organic matter, initial nitrogen contained in the soil. For management rules literature analysis was conducted to find information about each of the management rules imposed for each specific cropping system being tested. Once the models

were parametrized accordingly, variables that the model had to collect data on were specified based on the data variables shown in table 1. The model was then run, and data was collected. These set of steps were repeated for each cropping system being simulated.

Table 1: Table showing all the variables that data was collected on in this investigation.

Variables	Units
Cropping System	No units
Carbon Dioxide (CO2)	Million Metric Tons
Biomass Carbon	Kg/Ha
Biomass Nitrogen	Kg/Ha
Biomass Phosphorus	Kg/Ha
Total Carbon	Kg/Ha
Net Biomass Nitrogen Mineralized	Kg/Ha
Extractable Soil Water	mm
Total Nitrogen	Kg/Ha
Organic Carbon	%
Soil Water Storage	mm
Soil Temperature	Degrees Celsius
Yield	Kg/Ha
Biomass	Kg/Ha
Soil Water	Mm
N2O Emissions	Kg/Ha
CH4 Emissions	Kg/Ha
CO2 Emissions	Kg/Ha

Once the data was collected, data analysis occurred to find out which sustainable cropping system was the best. First, the data was preprocessed where columns that had many missing values were removed. Additionally, labels were added to the dataset that described the cropping sequence, row spacing, and nitrogen rate used by each cropping system. Table 2 shows part of the final preprocessed dataset. Once the data was preprocessed, the data was then aggregated. For data aggregation, the GroupBy function in Pandas library was used where the data was split into groups based on the cropping sequences each cropping system used. Therefore, the data was split into 10 separate groups since there were 10 cropping sequences evaluated. Each variable in each group was then aggregated down to just one value describing the whole group, by taking either the mean and median value of the variable column. The groups of data were then merged back together in one data frame. Table 3 shows the data aggregation table. After that a bar graph for each variable in the data aggregation table was made, to visually see how each variable's values differed for each cropping sequence (Fig. 6). This data aggregation process was then repeated for both row spacing and nitrogen rates, where the GroupBy function split and aggregated the data based on these 2 parameters just like it did for cropping sequence. Lastly, PCA (Principal Component Analysis) an unsupervised machine learning model was used. This model is a dimensionality reduction algorithm that is useful for clustering and visualizing high-dimensional data such as soil quality, since the impacts of each cropping system on the soil was represented by multiple data variables. The PCA model basically states, how similar cropping systems where in terms of their impacts on the soil quality. Cropping systems plotted close together were more similar in terms of their impacts on soil quality, whereas cropping systems plotted farther away for each other were less similar (Fig. 5). Each cropping system was color coded on the PCA plot based on the cropping sequence used in the cropping system. This made it easier to specifically compare the effects of each cropping sequence on soil quality.

Figure 4: Flow Chart summarizing the whole experimental procedure.

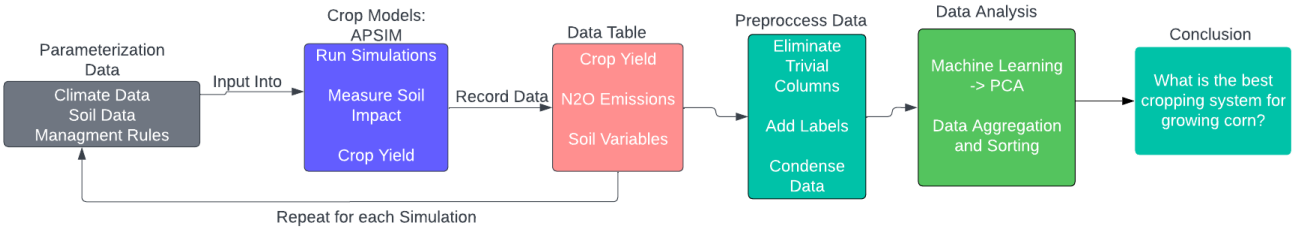


Table 2: Table showing how part of the data looked after preprocessing.

	Cropping System	year	Biomass Carbon	Biomass Nitrogen	Total carbon	Total nitrogen	Organic Carbon	Soil Water Storage	Soil Water	N2O emissions	Cropping Sequence	Nitrogen rate	row spacing
0	Continuous Maize (375cm) 0kg fertilizer.csv	1979	1234.4	154.3	132691.1	9235.9	5.634	1.246	1.245	0.000	Monoculture	0 kg/ha	375cm
1	Continuous Maize (375cm) 0kg fertilizer.csv	1980	1728.7	216.1	134063.1	9270.4	5.632	1.239	1.239	0.000	Monoculture	0 kg/ha	375cm
2	Continuous Maize (375cm) 0kg fertilizer.csv	1981	1776.5	222.1	132481.7	9229.7	5.610	1.259	1.259	0.000	Monoculture	0 kg/ha	375cm
3	Continuous Maize (375cm) 0kg fertilizer.csv	1982	2148.4	268.5	132766.2	9260.9	5.610	1.193	1.193	0.000	Monoculture	0 kg/ha	375cm
4	Continuous Maize (375cm) 0kg fertilizer.csv	1983	1866.2	233.3	131664.7	9200.1	5.578	1.218	1.218	0.000	Monoculture	0 kg/ha	375cm
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4668	Maize-Wheat Cover Crop 65kg.csv	2011	2067.8	258.5	122553.7	9523.4	5.164	1.119	0.539	0.488	Maize-Wheat Cover Crop	65 kg/ha	750cm
4669	Maize-Wheat Cover Crop 65kg.csv	2012	1921.8	240.3	122255.7	9488.9	5.149	1.123	3210.858	0.476	Maize-Wheat Cover Crop	65 kg/ha	750cm
4670	Maize-Wheat Cover Crop 65kg.csv	2012	2000.5	250.1	121963.4	9513.3	5.142	1.091	0.527	0.420	Maize-Wheat Cover Crop	65 kg/ha	750cm
4671	Maize-Wheat Cover Crop 65kg.csv	2013	2066.7	258.3	122234.1	9485.2	5.139	0.858	2434.984	0.470	Maize-Wheat Cover Crop	65 kg/ha	750cm
4672	Maize-Wheat Cover Crop 65kg.csv	2013	2217.0	277.1	122061.0	9509.6	5.139	1.227	0.606	0.442	Maize-Wheat Cover Crop	65 kg/ha	750cm

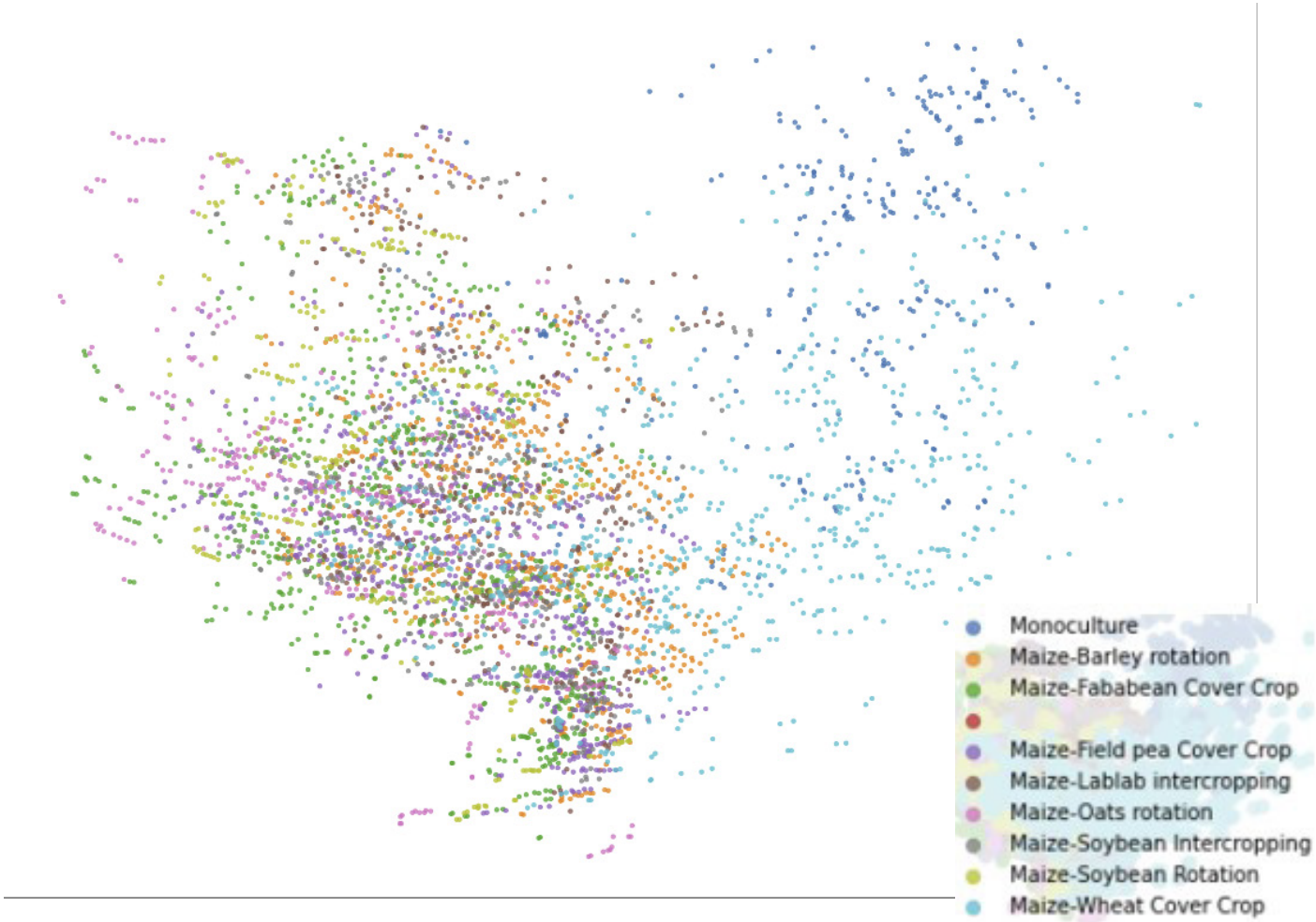
Table 3: Table Showing the combined data after aggregation.

	Biomass Carbon		Biomass Nitrogen		Total carbon		Total nitrogen		Organic Carbon		...	
	nanmean	nanmedian	nanmean	nanmedian	nanmean	nanmedian	nanmean	nanmedian	nanmean	nanmedian	...	nanmean
Cropping Sequence												
Maize fallow	900.881111	885.15	112.611111	110.60	123461.987222	122940.25	9562.361667	9522.60	5.232644	5.2100	...	0.722367
Maize-Barley rotation	1905.312705	1891.20	238.163527	236.40	126755.300676	126969.15	9924.613092	9942.35	5.355482	5.3625	...	0.446078
Maize-Fababean Cover Crop	1418.326352	1439.25	177.290815	179.90	125053.520000	125014.65	9687.578584	9696.30	5.280695	5.2800	...	0.402580
Maize-Field pea Cover Crop	1665.946995	1685.70	208.243399	210.70	126266.558030	126223.45	9784.081478	9790.55	5.318341	5.3160	...	0.408388
Maize-Lablab intercropping	1751.752095	1702.50	218.972857	212.80	127050.151905	127129.30	9829.668190	9837.50	5.360779	5.3630	...	0.756730
Maize-Oats rotation	1107.735645	1093.30	138.469171	136.70	123882.739499	123696.05	9592.743642	9573.60	5.244319	5.2360	...	0.878365
Maize-Soybean Intercropping	1643.576571	1670.45	205.446857	208.90	126143.126714	126154.75	9799.485000	9813.85	5.330659	5.3280	...	0.810756
Maize-Soybean Rotation	1362.022929	1305.50	170.253214	163.15	125393.971429	125298.10	9718.860000	9717.20	5.302671	5.3005	...	0.801662
Maize-Wheat Cover Crop	2637.072837	2485.30	329.636684	310.60	129372.599468	130591.40	10093.968004	10175.40	5.455340	5.5040	...	0.465136
Monoculture	3637.399237	3762.40	454.672545	470.30	133846.337274	133857.40	9498.658341	9490.10	5.625163	5.6370	...	0.002955

Results

Figure 5 shows the PCA plot down below for the data, where each dot represents one cropping system. The PCA plot shows that most of the cropping sequences had similar impacts on the soil quality compared to one another, since a lot of the cropping sequences are tightly clustered together especially in the middle-left section of the plot. The only cropping sequences that had somewhat unique impacts on the soil quality, are corn-wheat cover cropping and monoculture (continuous corn cropping) cropping sequences since they are further apart and separated from the other cropping sequences. These results were very helpful for simplifying the whole final analysis process as discussed later.

Figure 5: PCA plot of the Northwest Iowa Weather Aggregated Data. PCA plots for other weather data followed similar pattern.



For my final analysis of cropping sequences, only 3 cropping sequences of Monoculture (continuous-cropping), Maize-Wheat Cover Cropping, and Maize-Faba Bean Cover Cropping were analyzed. The Monoculture and Maize-Wheat Cover Cropping were analyzed since they both made somewhat unique impacts on the soil quality as shown by the PCA plot. The Maize-Faba Bean Cover Cropping was analyzed since it had the highest crop yield. All the other cropping sequences had similar impacts on the soil that Corn-Faba beans cover cropping had (shown by PCA plot), but less crop yield as well as higher N2O emissions, so there wasn't a need to analyze them in depth. To find the best cropping sequence, bar graphs were made for each variable in the data aggregation table (table 2), which is shown in figure 6 and a table that listed the pros and cons of each of the 3 cropping sequences was also made which is shown below. Based on the pros and cons of each cropping sequence, it was concluded that the Maize-Wheat Cover Cropping system is the best cropping sequence for growing corn, since it allows the soil to retain high moisture, allows the soil to retain decently rich content of carbon and nitrogen, emitted small amounts of N2O, and lead to decent production in crop yields. The other 2 cropping sequences have major flaws, such as the Monoculture cropping sequence making the soil have bad water retaining efficiency and moisture as well has high N2O emissions, and the Corn-Faba Bean cover cropping sequence leading the soil to not being rich in nutrients.

Figure 6: Figure showing the bar graphs made based on each variable in Table 2. Yellow is Maize-Faba bean cover cropping, blue is Maize-Wheat cover cropping, and green is monoculture

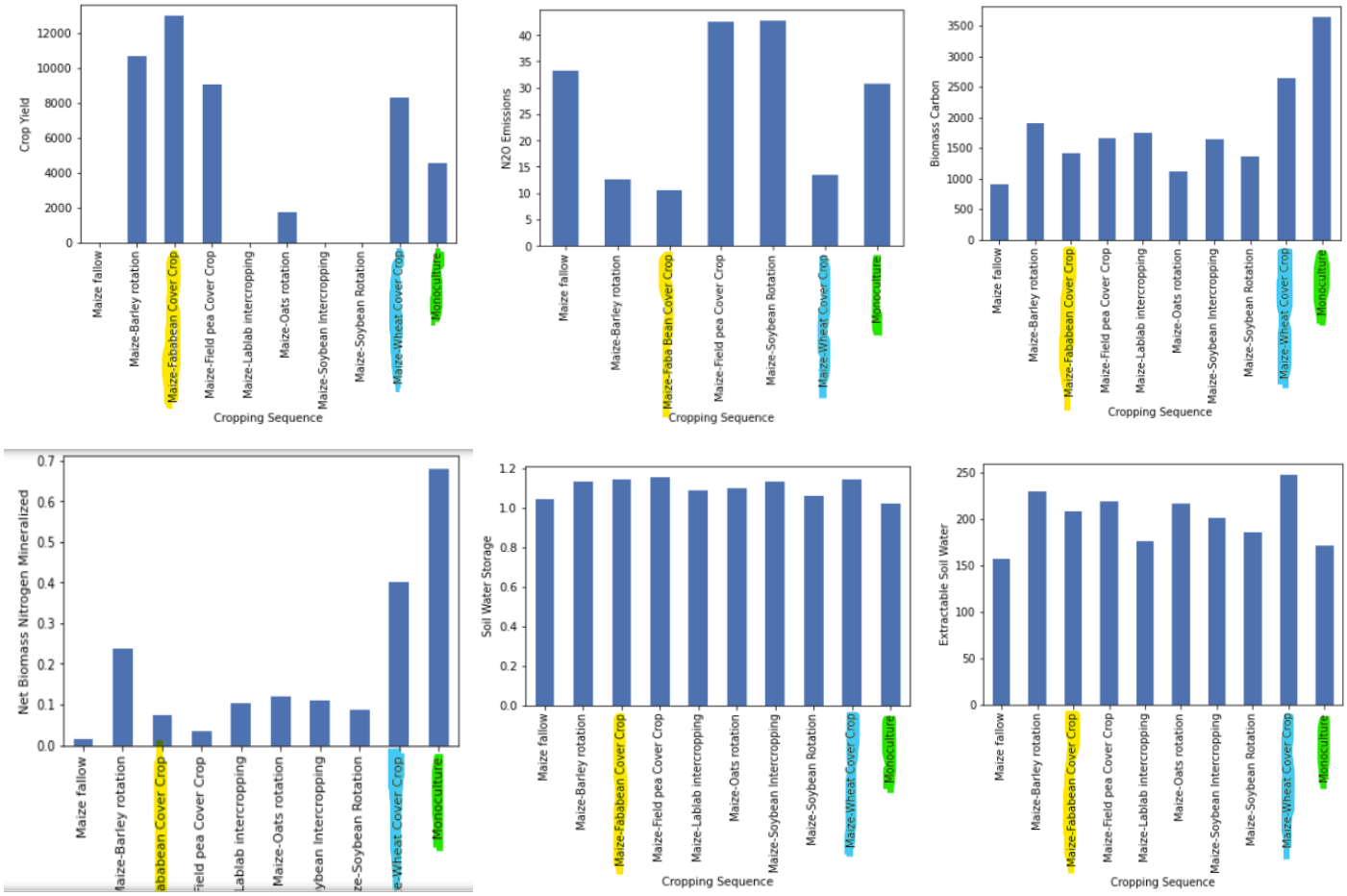


Table 4: Table showing the pros and cons for each of the 3 cropping sequences.

Summary	Monoculture	Maize-Winter Wheat Cover Crop	Maize-Faba Bean Cover Crop
Pros	• High carbon quantity in soil • High nitrogen quantity in soil • Low N2O emissions	• Decent carbon quantity in soil • Decent nitrogen quantity in soil • High soil moisture • High water retaining efficiency • Relatively low N2O emissions	• Highest Crop Yield • Average/Low N2O Emissions • Decent soil moisture • Decent water retaining efficiency
Cons	• Low soil moisture • Low water retaining efficiency. • Mediocre crop yields • High N2O emissions	• Relatively high CO2 emissions	• Low carbon quantity in soil • Low nitrogen quantity in soil

Next the different nitrogen rates were analyzed, where bar graphs were made for each variable in the nitrogen rate data aggregation table. Looking at the bar graphs there was a general trend that as nitrogen rate increased, crop yield, soil carbon content, and soil nitrogen content increased. However, as nitrogen rate increased soil moisture and soil water holding capacity, tended to decrease and N2O emissions tended to increase. Therefore, it was concluded that a nitrogen rate of 135 Kg/Ha is the best since it is the balance of all the extremes that allows farmers to have decent crop yield, soil nitrogen content, soil carbon content, soil moil moisture, water holding capacity, and not to high N2O emissions. However, if the crop yield is very important to keep high, then a 225 Kg/Ha is also an acceptable nitrogen rate. Lastly, the 2 different row spacings were analyzed. It was concluded that there was no significant difference between either row spacing and that either is acceptable.

Figure 7: Figure showing all the bar graphs made based on each variable in the data aggregation table of nitrogen rate.

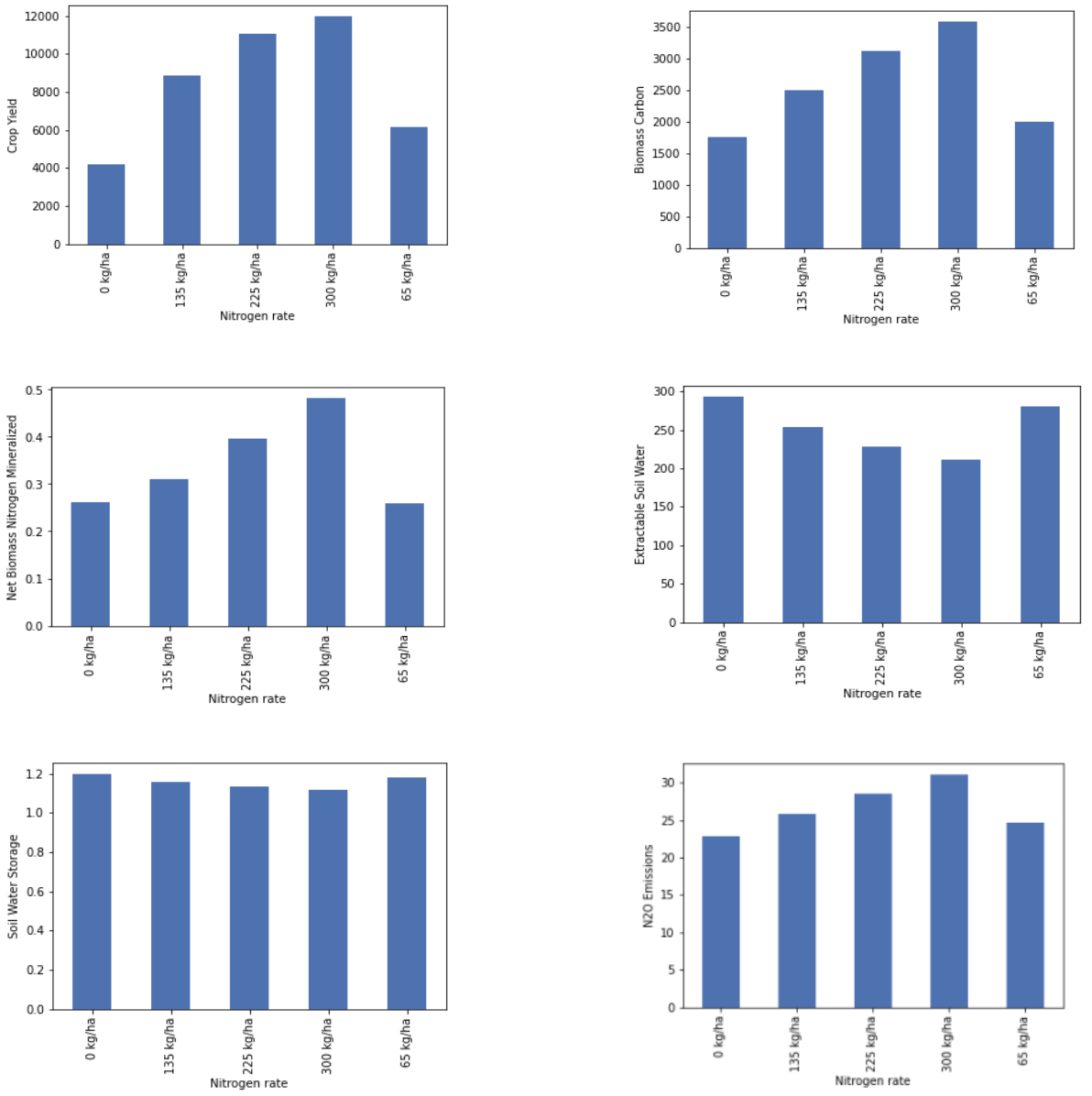


Table 4: Table Showing the pros and cons of the nitrogen rates of 0, 135, and 300 Kg/Ha

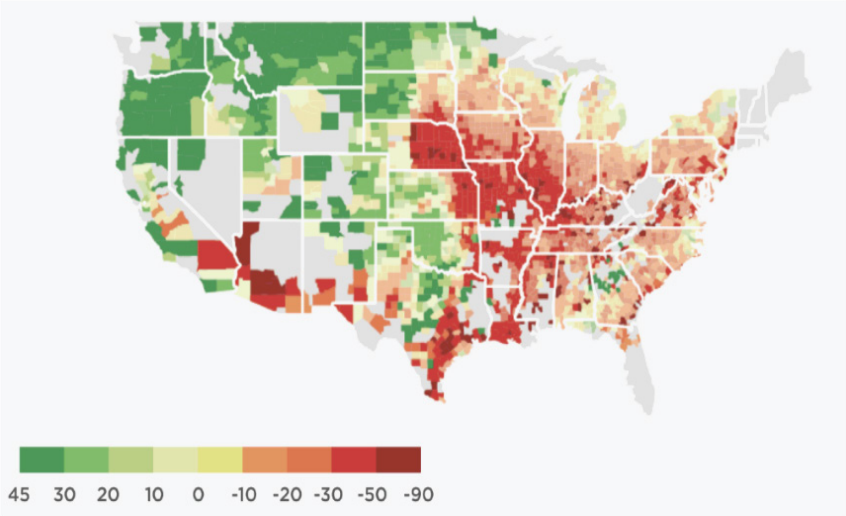
Summary	0 Kg/Ha	135 Kg/Ha	300 Kg/Ha
Pros	• High soil moisture • High water retaining efficiency • Low N2O emissions	• Decent soil moisture • Decent water retaining efficiency • Moderate N2O emissions	• High carbon quaintly in soil • High nitrogen quaintly in soil • High Crop yield • High N2O emissions
Cons	• Low carbon quantity in soil • Low nitrogen quantity in soil • Low crop yield	• Moderate carbon quantity in soil • Moderate nitrogen quantity in soil • Moderate crop yield	• Low soil moisture • Moderate/low water retaining efficiency

Conclusion

After conducting the final research and analysis, it can be concluded that the best sustainable and global warming resilient cropping system for growing corn in the Midwest, USA was a cropping system that implements a corn-winter wheat cover cropping sequence, uses a nitrogen rate of 135 kg/ha or 225 kg/ha, and either a row spacing of 37.5cm or 75cm. This cropping system is significantly more resilient towards the impacts of global warming as it is significantly better for retaining the long-term moisture of the soil. This is demonstrated by the fact that this system had an extractable soil water content that was 90mm higher than the mean extractable soil water for fields that implemented a continuous cropping sequence. As shown by the map below, the land in the Midwest is going to be extremely dried out and therefore it is essential for farmers to implement a cropping system that allows

the soil to retain high moisture. This system is also highly more sustainable as its crop yield is 1900 kg/ha higher than the mean crop yield of fields that implanted a continuous cropping sequence, and its N2O emissions were 17.28 Kg/Ha lower than the mean N2O emissions of fields that implemented a continuous corn cropping sequence. The future for this project is to verify these results on the field and spread the message to farmers.

Figure 8: Map showing how each area in the United States is going to be agriculturally impacted by climate change.



Hsiang, Kopp, Jina, Rising, et al./Science

Acknowledgements

I would like to express a special thanks to my mentor, Deirdre Griffin Lahue, who helped me a lot during the duration of this research project. She really helped me complete this project and learn a lot which I am very thankful for.

I would also like to express gratitude towards Dr. Kyungdahm, who provided me valuable resources and advice about crop modeling.

Bibliography

Beck, K. (2022, June 22). Corn [Photograph]. TastingTable.https://www.tastingtable.com/905203/which-us-state-produces-the-most-corn/

Edith, R. S. (2020). Modelling the effects of maize/lablab intercropping on soil water content and nitrogen dynamics using APSIM-Model [Doctoral dissertation, University of Venda]. UnivenIR. https://univendspace.univen.ac.za/handle/11602/1568

Estimated World Water Use [Chart]. (n.d.). The Food and Agriculture Organization. http://www.fao.org/fileadmin/user_upload/faowater/images/graphs_maps/waterusegraph.jpg

Etemadi, Fatemeh Hashem, Masoud Zandvakili, Omid Dolatabadian, Aria Sadeghpour, Amir; 2018/01/11; Nitrogen Contribution from Winter-Killed Faba Bean Cover Crop to Spring-Sown Sweet Corn in Conventional and No-Till Systems; 110; 10.2134/agronj2017.08.0501;Agronomy Journal

Ethanol fuel from corn faulted as 'unsustainable subsidized food burning' in analysis by Cornell scientist. (2001, August 6).

Cornell University. https://news.cornell.edu/stories/2001/08/ethanol-corn-faulted-energy-waster-scientist-says

Foley, J. (n.d.). The Future of Food. National Geographic. Retrieved November 9, 2022, from https://www.nationalgeographic.com/foodfeatures/feeding-9-billion/

Friedland, A., & Relyea, R. (2015). Feeding the World. In A. Friedland & R. Relyea (Authors), Enviornmental Science (Third ed., pp. 373-413). Bedford, freeman & worth high school publishers.

Hoogenboom, G., C.H. Porter, K.J. Boote, V. Shelia, P.W. Wilkens, U. Singh, J.W. White, S. Asseng, J.I. Lizaso, L.P. Moreno, W. Pavan, R. Ogoshi, L.A. Hunt, G.Y. Tsuji, and J.W. Jones. 2019. The DSSAT crop modeling ecosystem. In: p.173-216 [K.J. Boote, editor] Advances in Crop Modeling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, United Kingdom (http://dx.doi.org/10.19103/AS.2019.0061.10).

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, J.W. White, W. Pavan, F.A.A. Oliveira, L.P. Moreno-Cadena, J.I. Lizaso, S. Asseng, D.N.L. Pequeno, B.A. Kimball, P.D. Alderman, K.R. Thorp, M.R. Jones, S.V. Cuadra, M.S. Vianna, F.J. Villalobos, T.B. Ferreira, J. Koo, W.D. Batchelor, L.A. Hunt, and J.W. Jones. 2021. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.8 (https://dssat.net). DSSAT Foundation, Gainesville, Florida, USA.

Hoss, M., Behnke, G. D., Davis, A. S., Nafziger, E. D., & Villamil, M. B. (2018). Short Corn Rotations Do Not Improve Soil Quality, Compared with Corn Monocultures. Agronomy Journal, 110(4), 1274-1288. https://doi.org/10.2134/agronj2017.11.0633

Huffman, Raeann Lyn, "Intercropping Corn And Soybean: Profitability Of Different Agronomic Management Techniques" (2021). Theses and Dissertations. 1375. https://ir.library.illinoisstate.edu/etd/1375

Jagadamma, S., Lal, R., Hoeft, R. G., Nafziger, E. D., & Adee, E. A. (2008). Nitrogen fertilization and cropping system impacts on soil properties and their relationship to crop yield in the central Corn Belt, USA. ScienceDirect, 98(2), 120-129. https://doi.org/10.1016/j.still.2007.10.008

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, J.W. White, W. Pavan, F.A.A. Oliveira, L.P. Moreno-Cadena, J.I. Lizaso, S. Asseng, D.N.L. Pequeno, B.A. Kimball, P.D. Alderman, K.R. Thorp, M.R. Jones, S.V. Cuadra, M.S. Vianna, F.J. Villalobos, T.B. Ferreira, J. Koo, W.D. Batchelor, L.A. Hunt, and J.W. Jones. 2021. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.8 (https://dssat.net). DSSAT Foundation, Gainesville, Florida, USA.

Ka ming fung, 2019, Co-Benefits of intercropping as a sustainable farming method for safeguarding both food security and air quality, Environmental Research Letters, 14, 044011, pages: 2-12

Lavergne S, Vanasse A,Thivierge M-N, Halde C. Nitrogen content ofpea-based cover crop mixtures and subsequent organiccorn yield. Agronomy Journal. 2021;113:3532–3547. https://doi.org/10.1002/agj2.20727

Liebig, M.A., Varvel, G.E., Doran, J.W. and Wienhold, B.J. (2002), Crop Sequence and Nitrogen Fertilization Effects on Soil Properties in the Western Corn Belt. Soil Sci. Soc. Am. J., 66: 596-601. https://doi.org/10.2136/sssaj2002.5960

Liebig, M. A., Varvel, G. E., Doran, J. W., & Wienhold, B. J. (2002). Crop Sequence and Nitrogen Fertilization Effects on Soil Properties in the Western Corn Belt. Agronomy Journal, 66(2), 596-601. https://doi.org/10.2136/sssaj2002.5960

Moser, W. (2017, July 12). Agricultural Yields (% Change) [Map]. Chicago Magazine. https://www.chicagomag.com/city-life/july-2017/global-warming-could-be-worse-for-the-illinois-economy-than-you-think/

#67 Monocropping. (n.d.). HealThePlannet. https://healtheplanet.com/100-ways-to-heal-the-

planet/monocropping/

Speidel, J. J. (2007, July). Estimated World Population Growth: 1750 to 2050 [Chart]. Research Gate. https://www.researchgate.net/figure/Estimated-World-Population-Growth-1750-to-2050-20_fig1_225463481

SubbaRao, M. (2022, September 14). Maize Yield Projections [Map]. NASA. https://svs.gsfc.nasa.gov/4925#section_credits

Trends in food production and nitrous oxide emissions from the agriculture sector in India: environmental implications - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Cereal-production-and-nitrogen-fertilizer-consumption_fig15_225575493 [accessed 11 Mar, 2023]

Trends in food production and nitrous oxide emissions from the agriculture sector in India: environmental implications - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Cereal-production-and-nitrogen-fertilizer-consumption_fig15_225575493 [accessed 11 Mar, 2023]

University of Wisconsin Cooperative Extension Service, The University of Minnesota Extension Service, & The Center for Alternative Plant and Animal Products. (1992). Alternative Field Crops Manual. Corn Agronomy. http://corn.agronomy.wisc.edu/Crops/

Wade, J., Culman, S.W., Logan, J.A.R. et al. Improved soil biological health increases corn grain yield in N fertilized systems across the Corn Belt. Sci Rep 10, 3917 (2020). https://doi.org/10.1038/s41598-020-60987-3

Wolday, K., & Hruy, G. (2015, November 21). A Review on: Performance Evaluation of Crop Simulation Model (APSIM) in Prediction Crop Growth, Development and Yield in Semi Arid Tropics (Research Report No. 5). Journal of Natural Sciences Research.

Neurosciences

Reconnect: A Computational Approach to Using DBS To Combat Neurodegenerative Disease Through Induced Plasticity (CBIO-0349)

Shivani Sama

Tesla STEM High School

Abstract

Restoring weakened connections in the brain could save millions of people suffering from neurodegenerative diseases worldwide. Neurodegenerative disease is caused by the loss of synapses (the connection between neurons) and leads to cognitive and motor decline. This project seeks to determine if targeted deep brain stimulation can be used to restore motor function in someone with neurodegenerative disease through an adaptive deep brain stimulation technique. This was done through a computational simulation of the Subthalamic Nucleus (STN) and the Globus Pallidus (GPi) - two regions of the brain responsible for motor control. Where the independent variable was the strength of the synapses in the STN and the dependent variable was the mean firing rate of the GPi with and without stimulation. This simulation was used to derive simulated data with four features: STN strength, the firing rate of the GPi with and without stimulation, and percent change (improvement in motor function). This data was trained using the K-means and K-Nearest Neighbors machine learning algorithms to determine the optimal parameters for deep brain stimulation. There was a significant increase in motor function in the STN, with a percent change of 103% at the maximum of all trials. The machine learning algorithm used to predict if stimulation in the STN would increase synaptic firing to improve motor control had a 91% accuracy. This project was able to create an adaptive deep brain stimulation algorithm that restored weakened regions of the brain to a healthy state.

Introduction

Is it possible to reconnect the brain? Deep brain stimulation or electrical stimulation of different brain regions might provide some insight. Adaptive deep brain stimulation or personalized deep brain stimulation can be used to induce artificial connectivity in the brain of patients with neurodegenerative disease through personalized stimulation parameters and can be used as a treatment to improve cognitive function. This could help over 55.2 million people across the world that suffer from neurodegenerative disease such as Alzheimer's and Parkinson's disease (World Health Organization). This potential treatment could lower cognitive decline at a much faster rate than existing medication, and significantly improve the lives of patients with the condition.

Neurodegenerative Disease

Neurodegenerative disease or the slow decay of neurons in the brain is a disease that impairs a patient's ability to think, remember, and make everyday decisions (Berman 1). More severe forms of neurodegenerative disease can lead to loss in motor movement or body tremors, such as Parkinson's disease. It affects adults greater than 65 years of age and is not a normal part of aging. The causes of neurodegenerative disease are not clear, but it has been linked to age, genetics (family history), ethnicity, poor heart health, and previous head traumas (The Center for Disease Control and Prevention). Neurodegenerative disease can be absolutely devastating as the patient slowly loses their cognitive abilities.

Neurodegenerative disease is correlated with

the loss of synapses in the brain. Think of the brain as a large network of neurons. These neurons send signals to each other through neurotransmitters such as potassium and sodium that flow through a channel across the neurons known as synapses. These synapses help stimulate certain neurons to help move a part of the body or to remember something. As these connections between parts of the brain are made more and more frequently, synaptic strength increases. That's why as you practice something over and over, you're able to recall it more quickly than before.

Synapses naturally weaken over time as connections between neurons are made less and less frequently – how our memory works. However, in patients with neurodegenerative disease this synapse loss is at alarmingly high rates, and this is associated with memory loss in patients with Alzheimer's disease. In these situations, the connections between these synapses weaken rapidly and sometimes altogether gone (Subramanian 1). Eventually, a patient will face severe cognitive decline that leads to more concerning conditions such as memory loss and loss of motor control.

In fact, synapse loss is used as a common diagnosis method for dementia, "synapse loss is the best correlate of dementia severity" (Clare 1). Other forms of diagnosis for neurodegenerative disease include blood tests for dementia related biomarkers, MRI scans to determine structure loss in the brain, and other electrical signals to determine synaptic strength in the brain. Dementia is often diagnosed much later than when synapses in the brain begin to weaken rapidly, and this lowers the efficiency of medication and treatment that can be provided.

Current Treatments

Current treatments for synapse loss related dementia include cholinesterase inhibitors, memantine, and other medication that can be used to slow down synapse loss in the brain. These medications stop the breakdown of chemical messengers in the brain that are responsible for memory and learning. These chemical messengers break down neurotransmitters, so the brain isn't stimulated too often. This is usually healthy, but in patients with neurodegenerative disease this breakage is done at a much higher rate which can lead to synapse loss.

Deep Brain Stimulation

A new potential treatment for this synapse loss in patients with neurodegenerative disease is Deep Brain Stimulation (DBS). DBS is a process where electrodes are inserted into the brain that produce electrical impulses in the brain that are usually used to combat abnormal impulses. These electrodes are placed into specific brain areas, such as the basal ganglia or the thalamus where neurons are connected. This is a current treatment for patients with Parkinson's disease as it is used to combat the synchronization of neurons in the brain that can lead to body tremors and have been used in over 160,000 patients worldwide and have proven to be effective (Lozano 1). Deep brain stimulation has also been used to treat types of epilepsy and even some types of dementia such as Alzheimer's, however, there needs to be more research done. Deep brain stimulation has also opened new frontiers in medicine as a modulator for malfunctioning neural circuits.

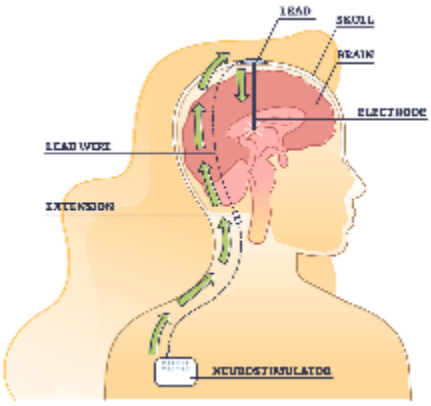


Figure 1*: Diagram of electrode placement in DBS*

Deep brain stimulation may even be used to artificially connect different regions of the brain, something known as artificial connectivity. The theory that neurons that fire together wire together may provide insight into using this information to potentially connect different parts of the brain that can be used to treat more types of neurodegenerative disease with more efficiency. This is known as engineered neuroplasticity, or the ability to change neural circuits in the brain to be able to regain cognitive and physical function (Falowski 1). This kind of engineered connectivity can be done through deep brain stimulation, by stimulating two different regions of the brain at the same time to have them wire at the same time.

Deep brain stimulation can be used to target stimulation in different regions of the brain, therefore improving neuroplasticity and artificially connecting different regions of the brain. This can then be applied to a variety of neurodegenerative diseases, such as Alzheimer's and is not just limited to the basal ganglia like current deep brain stimulation is. Although promising, some current problems with deep brain stimulation is the variance of treatment results from person to person. For instance, deep brain stimulation in Parkinson's patients has led to over 95% of patients showing successful results such as alleviation of symptoms and improve in motor and cognitive function. However, these effects of stimulation don't last very long, and patients need to come back frequently to improve their abilities. Additionally, a large amount of stimulation is required to achieve these results which has significant consequences on surrounding brain regions. Some of these symptoms can lead to change in personality and often cognitive decline in other areas of the brain (Zarzycki 1).

Adaptive Deep Brain Stimulation

A type of deep brain stimulation technique to combat this problem is closed loop deep brain stimulation. This type of deep brain stimulation involves electrodes being placed in targeted areas and only firing at specific rates in specific neural populations instead of random fires (Dan Pina Fuentes 1). This closed loop stimulation method lowers the amount of stimulation required to achieve the same results, which can significantly lower the symptoms of deep brain stimulation. Research has shown that this method of deep brain stimulation is just as effective if not more effective at treating Parkinson's disease.

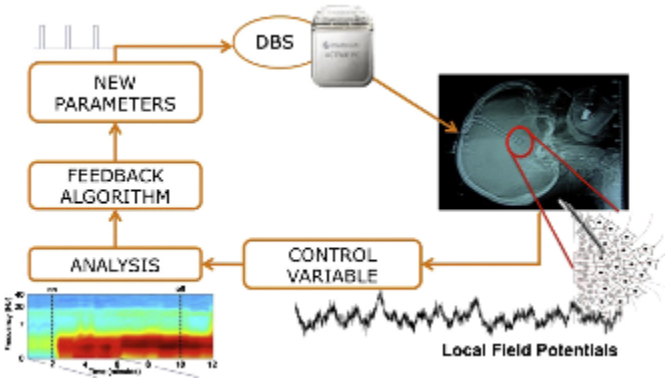


Figure 2: *Diagram of Adaptive Deep Brain Stimulation Methodology*

Expected Outcomes

This study aims to explore a computational approach to a closed loop deep brain stimulation method as a treatment for neurodegenerative disease using the concept of artificial connectivity and induced neuroplasticity. Current practices exist for deep brain stimulation as a treatment for Parkinson's disease but there is little research in using deep brain stimulation for other neurological disorders.

Creating simulations and models of healthy and unhealthy neural populations of the basal ganglia and the thalamus (the intersection of multiple neurons across the brain) outputs simplified EEG data (gives us the voltage over time) that can be used to visualize this process. This EEG data can help us determine the synchronization between the different neurons and neuron populations. By applying input on these unhealthy neuron populations that have lower synaptic strength we can improve the synaptic strength and also determine the synchronization between populations.

This simulation will give us the baseline. By applying machine learning algorithms, such as the k-nearest neighbors model, we can determine the amount and voltage of deep brain stimulation needed to bring an unhealthy neural population to a healthy population through artificial connectivity. The same way we can also determine the appropriate deep brain stimulation parameters that give us sustained change in neuroplasticity for a long time even after stimulation.

By using the KNN unsupervised machine learning model, we can predict the deep brain stimulation parameters needed to bring simulated unhealthy neural population to a healthy population through induced connectivity and neuroplasticity. This type of adaptive deep brain stimulation could be applied to patients with a variety of neurodegenerative diseases and can be used as a future treatment to improve and sustain cognitive abilities.

Methodology

Creating the Simulation

Use the BRIAN interface (a neural spiking simulator) to input the Hodgkin – Huxley mathematical equation that will simulate a realistic

and accurate brain-physics model showing neuron connectivity (this should be in the library). Start in the code by creating variables for each of the different ion charges in the model.

$$I = C_m \frac{\mathrm{d}V_m}{\mathrm{d}t} + \bar{g}_{\mathrm{K}} n^4 (V_m - V_K) + \bar{g}_{\mathrm{Na}} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_l),$$

$$\frac{dn}{dt} = \alpha_n(V_m)(1-n) - \beta_n(V_m)n$$

$$\frac{dm}{dt} = \alpha_m(V_m)(1-m) - \beta_m(V_m)m$$

$$\frac{dh}{dt} = \alpha_h(V_m)(1-h) - \beta_h(V_m)h$$

Hodgkin Huxley model for simulating neuron populations: V is voltage of different ions (K, Na, m, l). This is a derivative to determine firing rate.

Then input the synaptic plasticity equations that are derived from spike-timing dependent plasticity. Do this in the BRIAN python program in the form of if-statements. If the change in time is greater than the threshold, then increase the change in plasticity otherwise (else) decrease the change in plasticity.

$$\Delta w_{ij}(\Delta t) = \begin{cases} \lambda e^{-\frac{|\Delta t|}{\tau_+}} & \Delta t > 0, \\ -\lambda \beta e^{-\frac{|\Delta t|}{\tau_-}} & \Delta t \leq 0. \end{cases}$$

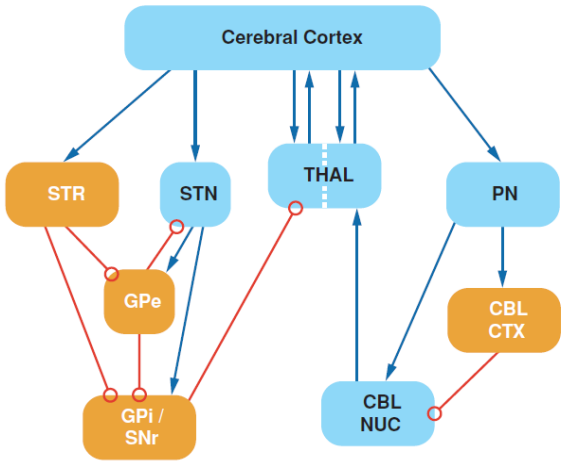
Mathematical model for synaptic plasticity, increases ionization above a threshold and decreases below.

Next add the equation for electrical stimulation of the neurons through deep brain stimulation. This equation will be conditional, so that the amount of deep brain stimulation can change periodically to create an adaptive DBS simulation. In the BRIAN python code, make stimulation a variable. Have an if-statement that will run if the Ik and Ina variables are below a threshold or completely turned off. In this if-statement increase the simulation variable, by adding the following equation – the effects should be shown on the neural clusters in the BRIAN interface

$$P(t) = \begin{cases} \kappa & t_k \leq t < t_k + \omega \\ -\kappa/p_s & t_k + \omega \leq t < t_k + \omega(1 + p_s) \\ 0 & \text{else} \end{cases}$$

Mathematical model for adaptive stimulation input, increase ionization above a threshold and apply no input if reached.

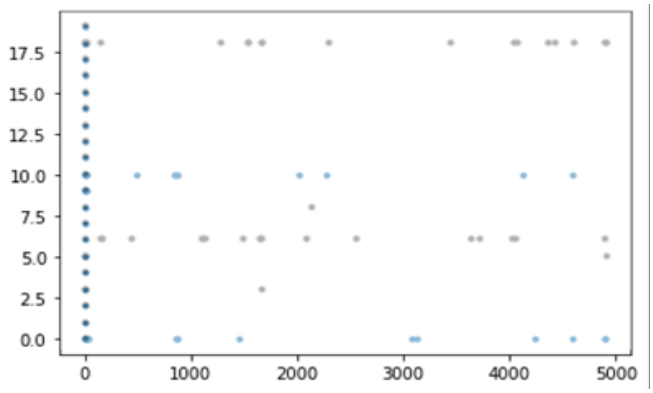
Test the equations in BRIAN, create firing distributions. If there is no linear pattern in the models, then it is accurate in terms of neuro-physics. This means there is noise in the model. Use the spike time plots to visualize oscillations in the data, this shows that DBS is modeled accurately as it is sinosodial.



Model of Brain Connectivity, stimulation in the Thalamus is connected to the GPi neuron group and this is the model for artificial connectivity. The model describes how the different parts of the brain and populations of neurons are connected. This model shows what the simulation of the equations represents.

Generating Simulated Data

Generate simulated data using the simulation created. Change the weight of the STN so it is approaching zero and record the effects in both the STN and the GPi. Count the spikes in the GPi plot and write down the number of spikes over the milliseconds the simulation ran for.



Example spike time plot of the GPi, blue is the GPi without stimulation and the gray is the GPi with stimulation

Then using the number of spikes in the GPi without and with stimulation, find percent change.

*(Spikes in GPi with stimulation – Spikes in GPi without stimulation) / Spikes in GPi without stimulation * 100 = percent change*

This also gives us the improvement in motor function due to stimulation. Record all this data in *Table 1*.

Table 1: Simulated data derived from running the simulation.

STN Weight	GPi without stimulation	GPi with stimulation	Percent change/improvement in motor function

K-Means Clustering

Use the K-means machine learning algorithm to cluster the data to figure out any irregularities and the threshold for healthy neuron stimulation. Write the algorithm in Google Collaboratory using the simulated data that was 4 features: STN Weight, GPi without and with stimulation, and percent change/improvement in motor function. Find the average distance from these points from each other to group the data into 3 clusters. Look at the algorithm and the raw data to determine what these clusters could mean.

K-Nearest Neighbors Machine Learning

Label the simulated data as unhealthy or healthy based on the thresholds from the clustering algorithm. Record this data in *Table 2*. Determine healthy or unhealthy based on the firing rate of GPi after stimulation.

Table 2: The simulated data with the "unhealthy" or "healthy" label

STN Weight	GPi without stimulation	GPi with stimulation	Percent change/improvement in motor function	Label

Using this data use the KNN unsupervised machine learning algorithm to predict whether or not DBS will increase firing rate to a healthy state. The KNN algorithm clusters data just like k-means but based on the labels and uses these clusters to predict. This would create an adaptive DBS method that would make sure DBS stimulation results in improvement in motor function. This would make DBS more personalized and effective.

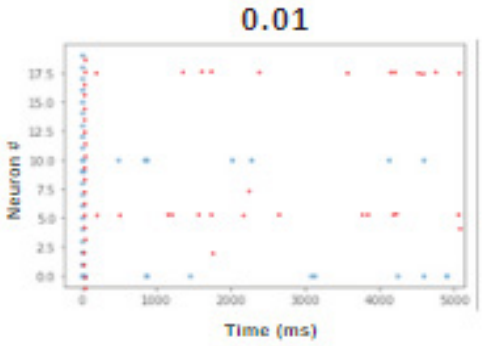
Results

Spike Time Plots

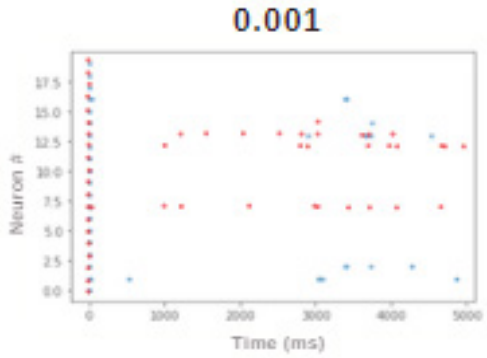
Lowered the weight of the synaptic strength between the Cerebral Cortex and the STN (Subthalamic Nucleus) by 10% each time to simulate an unhealthy brain such as one of a stroke or other types of dementia.

Recorded the spike time plots in the GPi before and after stimulation to see if there was any improvement. This was specifically recorded because the STN applies input to the GPi and red is the spiking with stimulation while blue is the spiking without stimulation.

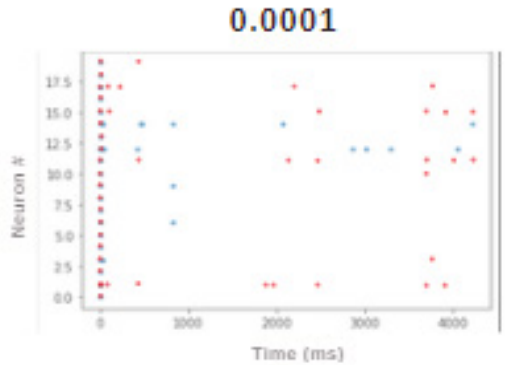
Spike time plot at synaptic strength of 0.01:



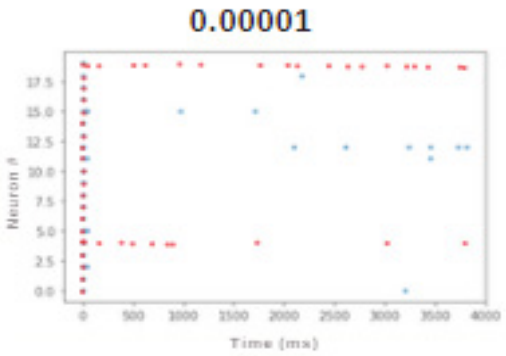
Spike time plot synaptic strength of 0.001:



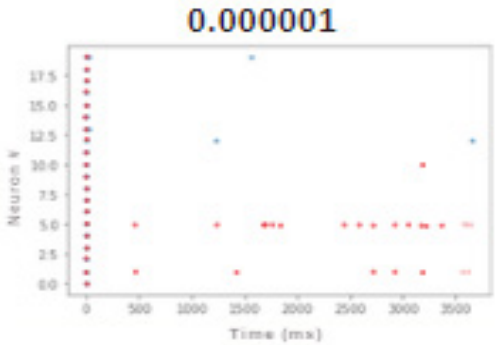
Spike time plot the synaptic strength of 0.0001:



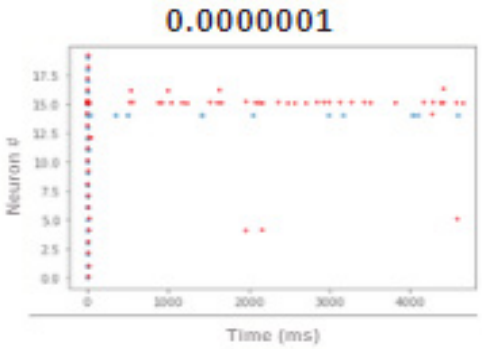
Spike time plot at the synaptic strength of 0.00001:



Spike time plot at the synaptic strength of 0.000001:



Spike time plot at the synaptic strength of 0.0000001:



All of these spike time plots show significantly more red dots than blue dots, showing significant improvement in synaptic strength due to stimulation of the STN.

Heat Maps

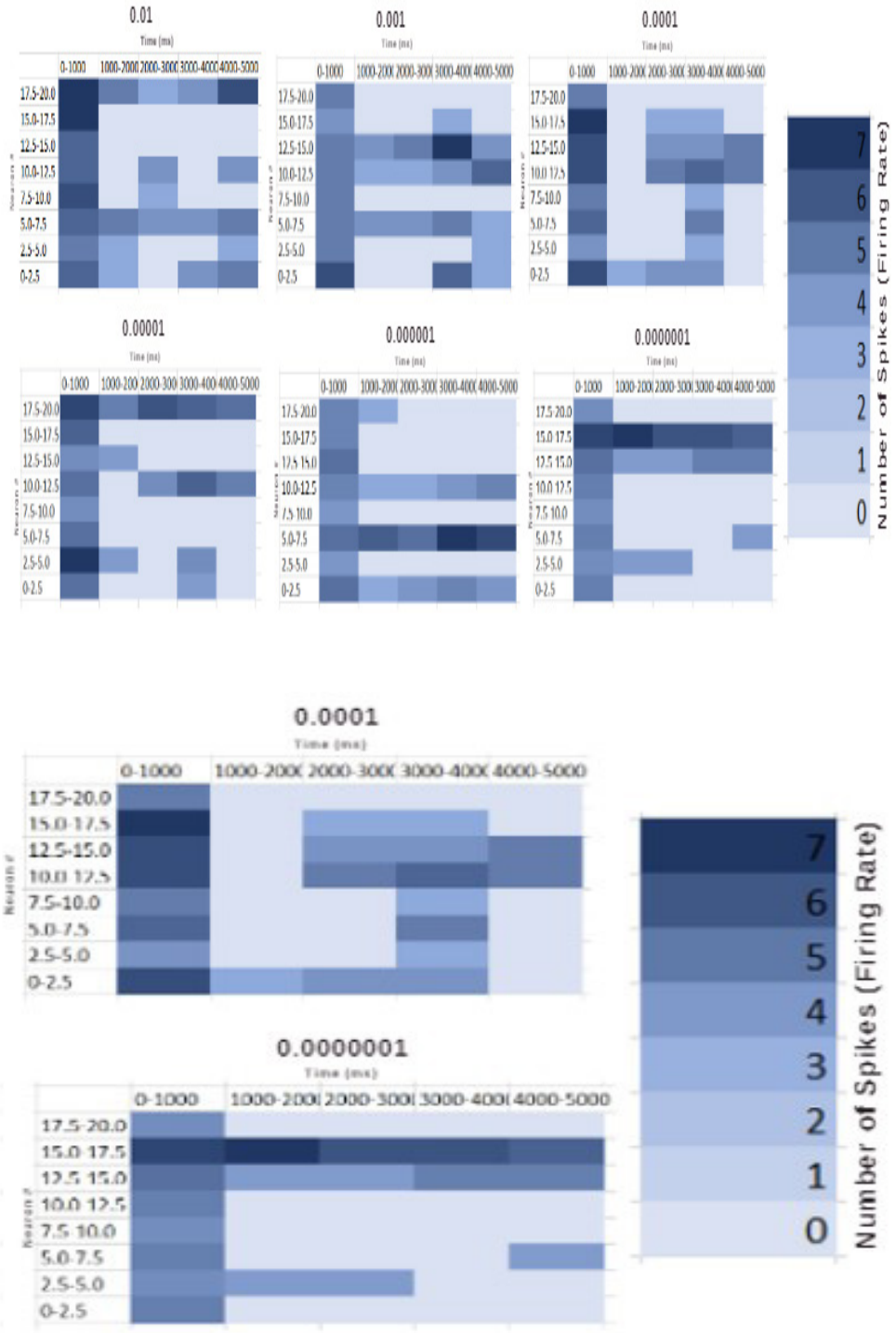


Figure 1: Heatmaps showing the firing rate of the GPi at different STN weights
The darker the region, the more firing there was (a higher mean firing rate).

These heatmaps show the effects of plasticity on the STN due to stimulation. The darkened horizontal lines show plasticity taking effect as the neurons fire more and more at some neuron populations over others. This increased and repeated firing is known as plasticity and shows that these changes in the synaptic strength of the GPi are going to be sustained for longer than the stimulation takes place.

Summary

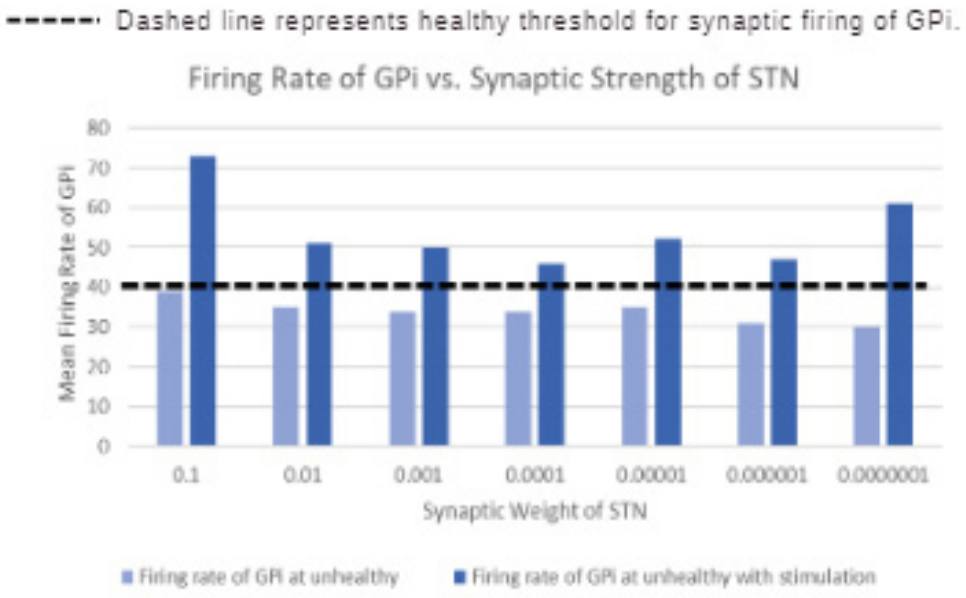


Figure 2: Bar graphs showing the relationship between Synaptic Strength and DBS effects on firing rate of GPi for seven trials (dashed line is the healthy threshold)

From these seven trials there was a clear increase in synaptic firing. So much so that the firing rate was able to improve above a healthy threshold for all of the trials. This shows that DBS can be used to improve synaptic strength to such a level that motor function can be restored in a patient who has lost it.

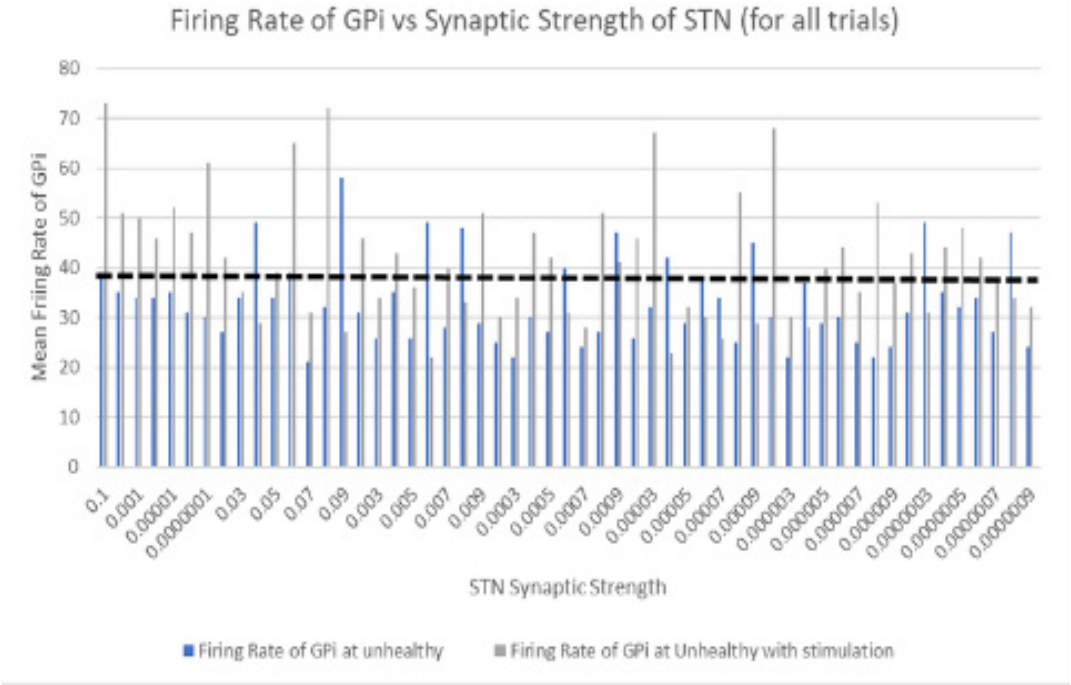


Figure 3: Bar graphs showing the relationship between Synaptic Strength and DBS effects on firing rate of GPi for 55 trials (dashed line is the healthy threshold)

Another 55 trials were generated from the simulation. However, these trials had more variability than the previous 7. 80% of the trials had an improvement in synaptic strength, however, about 10% did not have enough of an increase to get over the healthy threshold. In the other 20% of the trials there was a decrease in synaptic firing – most likely due to the stimulation of an already healthy GPi.

DBS Patterns

STN Synaptic Strength	Improvement in Motor Function due to DBS (%)
0.1	87%
0.01	48%
0.001	47%
0.0001	35%
0.00001	49%
0.000001	50%
0.0000001	**103%**

Table 3: Relationship between STN weight/synaptic strength and improvement in motor function (%) for seven trials

Significant improvement in motor function due to DBS. In each of the 7 trials there was a significant increase in the synaptic firing rate of the GPI due to stimulation showing that DBS can be used to improve motor function.

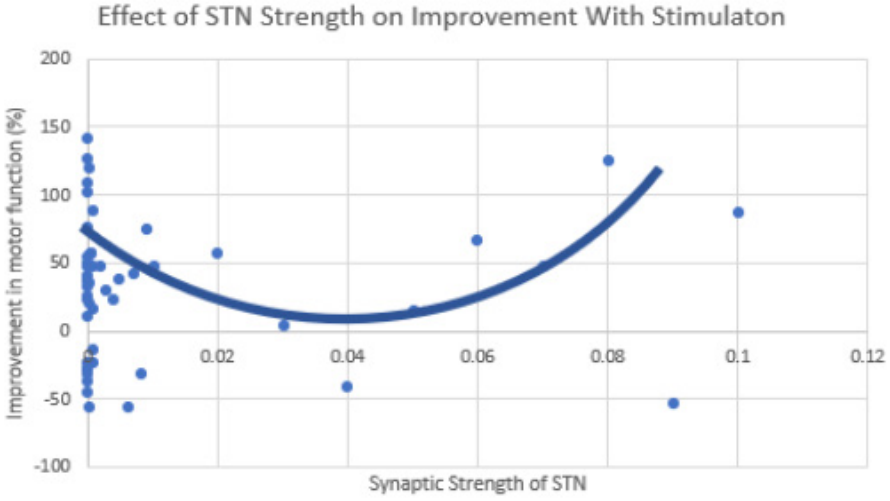


Figure 4: Graph of improvement in firing rate based on STN weight for all 55 trials; novel parabolic shape

The improvement in motor function compared to STN strength (a measure of how unhealthy a patient is) had a parabolic shape to it. This is an incredibly novel finding since current effects of DBS are unknown! But this means there is variability in DBS effects so machine learning analysis is vital.

Machine Learning

K-Means Clustering

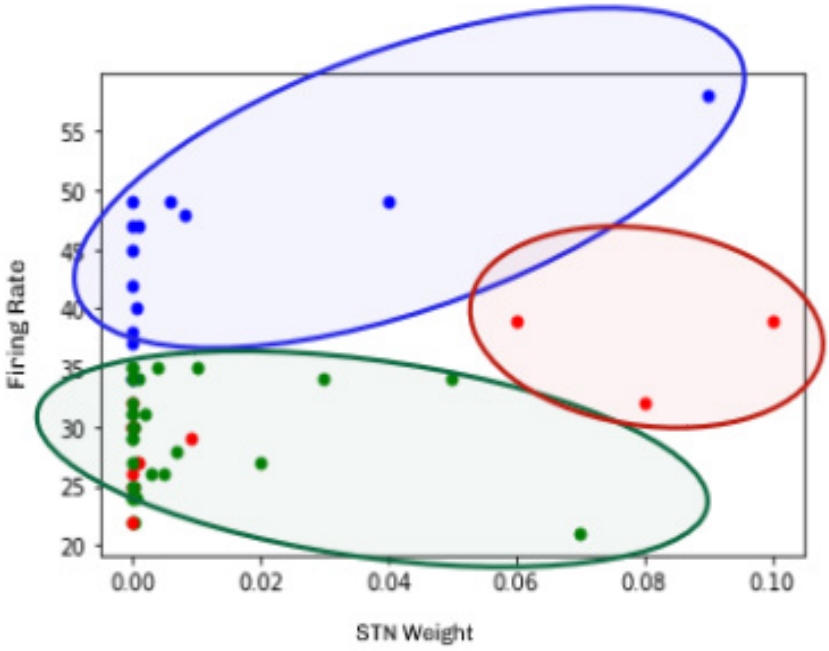


Figure 5: K-means clustering on the 110 generated data points

Blue: *Healthy (High)*
Green: *Unhealthy but had improvement in motor function (Moderate)*
Red: *Healthy to Unhealthy, decrease in motor function (Low)*

Generated 110 data points using simulation that had 4 features each: STN weight, GPi before and after stimulation, and % change. From these clusters I noticed some irregularities: some neurons had an increase in motor function but not above healthy threshold, and others had *decreased* motor function.

K-Nearest Neighbors

Trained a k-nearest neighbors machine learning algorithm on the 110 data points with the four features: STN weight, GPi before and after stimulation, and % change.

This algorithm can be used to predict whether a population of neurons will be healthy after stimulation to make sure that the synaptic firing isn't decreased. This significantly reduces the variability of DBS and makes the treatment more reliable and predictable.

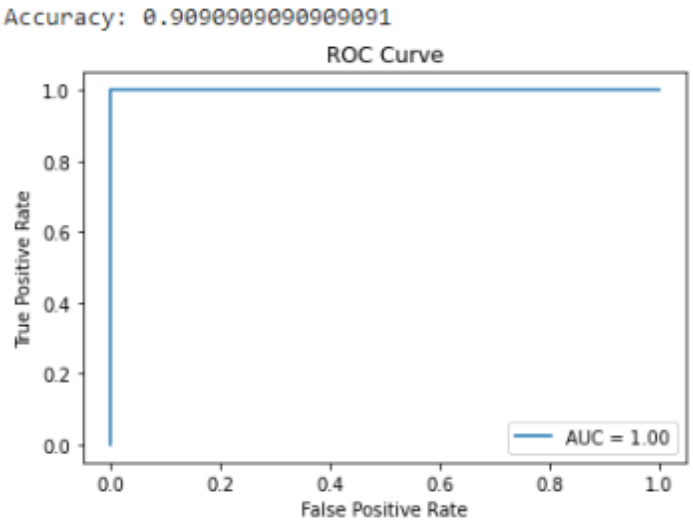


Figure 6: ROC Curve with a 91% accuracy showing highly reliable results

Adaptive Deep Brain Stimulation

The machine learning algorithm can be translated into a novel adaptive deep brain stimulation framework. This can regulate the amount of stimulation applied to remove varied effects and unwanted side effects of current stimulation techniques.

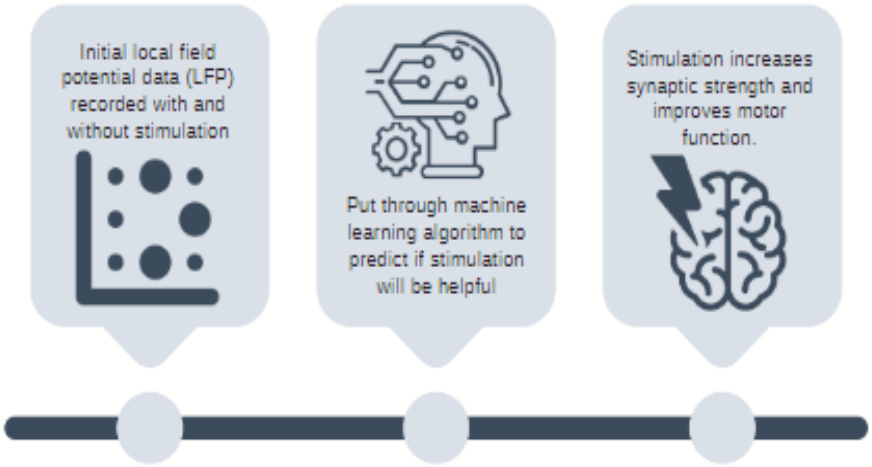


Figure 7:Diagram of Adaptive DBS Framework/Netwo

Conclusion

This study showed the efficacy of using DBS to increase synaptic firing to improve motor function. The hypothesis that DBS can be used to artificially increase the synaptic firing of an unhealthy GPi to restore motor function was supported. In this experiment there were 110 different trials of varying STN weights to mimic unhealthy brain activity and replicate the brain of someone with neurodegenerative disease. In all of these trials the greatest increase in motor function was 103-150% showing that DBS can be used to improve motor function at an incredibly rate and bring an unhealthy patient to a healthy state. There was an 80% efficiency rate with using DBS to improve motor function from all the trials tested. DBS patterns were also realized as being parabolic which is incredibly novel since current effects of DBS are unknown and this can be incorporated in the future to make DBS treatment more reliable. To reduce this variability a machine learning algorithm

for an aDBS framework was created to predict whether or not stimulation would increase synaptic firing to a healthy state in unhealthy patients. This algorithm had a 91% accuracy and proved to be reliable for doing so and also accounted for the variability in the efficacy of DBS. Some next steps include trying this DBS technique on other regions of the brain to see if the increase in synaptic firing can work to improve memory and cognitive function instead of just motor function. Additionally, testing on real patients would help show the true potential of this potential treatment. This research is incredibly novel since there is currently no means to improve and restore motor function in those with neurodegenerative disease. This is also the first time machine learning has been applied to DBS treatment to create an adaptive DBS framework for improved efficiency.

References

Bruno Bonnechère, Authors Ana Spasojevic, Authors Roni Sarel-Gallily, & Noeline Subramaniam. *(n.d.). What are neurodegenerative diseases and how do they affect the brain?* Frontiers for Young Minds.

Clare R, King VG, Wirenfeldt M, Vinters HV. Synapse loss in dementias. J Neurosci Res. 2010 Aug 1;88(10):2083-90. doi: 10.1002/jnr.22392. PMID: 20533377; PMCID: PMC3068914.

Dan Piña-Fuentes, J. Marc C. van Dijk, Jonathan C. van Zijl, Harmen R. Moes, Teus van Laar, D.L.Marinus Oterdoom, Simon Little, Peter Brown, Martijn Beudel, Acute effects of adaptive Deep Brain Stimulation in Parkinson's disease, Brain Stimulation, Volume 13, Issue 6, 2020, Pages 1507-1516, ISSN 1935-861X.

Falowski SM, Sharan A, Reyes BA, Sikkema C, Szot P, Van Bockstaele EJ. An evaluation of neuroplasticity and behavior after deep brain stimulation of the nucleus accumbens in an animal model of depression. Neurosurgery. 2011 Dec;69(6):1281-90. doi: 10.1227/NEU.0b013e3182237346. PMID: 21566538; PMCID: PMC4707959.

Lozano AM, Lipsman N, Bergman H, Brown P, Chabardes S, Chang JW, Matthews K, McIntyre CC, Schlaepfer TE, Schulder M, Temel Y, Volkmann J, Krauss JK. Deep brain stimulation: current challenges and future directions. Nat Rev Neurol. 2019 Mar;15(3):148-160. doi: 10.1038/s41582-018-0128-2. PMID: 30683913; PMCID: PMC6397644.

Subramanian J, Tremblay MÈ. Editorial: Synaptic Loss and Neurodegeneration. Front Cell Neurosci. 2021 Apr 21;15:681029. doi:

10.3389/fncel.2021.681029. PMID: 33967701; PMCID: PMC8097031.

World Health Organization. (n.d.). Dementia. World Health Organization. Retrieved January 24,2023, from https://www.who.int/news-room/fact-sheets/detail/dementia

Zarzycki MZ, Domitrz I. Stimulation-induced side effects after deep brain stimulation - a systematic review. Acta Neuropsychiatr. 2020 Apr;32(2):57-64. doi: 10.1017/neu.2019.35. Epub 2019 Oct 14. PMID: 31452489.

Identification of Novel Diagnostic Neuroimaging Biomarkers for Autism Spectrum Disorder Through Convolutional Neural Network-Based Analysis of Functional, Structural, and Diffusion Tensor Imaging Data Towards Enhanced Autism Diagnosis

Annie Adhikary

Nikola Tesla STEM High School
Redmond, Washington, United States

Keywords: Autism spectrum disorder, neuroimaging biomarkers, brain regions, diagnostics, magnetic resonance imaging, convolutional neural network, neurodevelopment

Abstract

Autism Spectrum Disorder is one of the leading neurodevelopmental disorders in our world, present in over 1% of the population and rapidly increasing in prevalence, yet the condition lacks a robust, objective, and efficient diagnostic. Clinical diagnostic criteria relies on subjective behavioral assessments, which are prone to misdiagnosis as they face limitations in terms of their heterogeneity, specificity, and biases. This study proposes a novel convolutional-neural-network based classification tool that aims to identify the potential of different neuroimaging features as autism biomarkers. The model is constructed using a set of sequential layers specifically designed to extract relevant features from brain scans. Trained and tested on over 300,000 distinct features across three imaging types, the model shows promising results, achieving an accuracy of 95.4% and outperforming metrics of current gold standard diagnostics. 32 optimal features from the imaging data were identified and classified as candidate biomarkers using an independent samples t-test, in which functional features such as neural activity and connectivity in various brain regions exhibited the highest differences in the mean values between individuals with autism and typical control subjects. The p-values of these biomarkers were < 0.001, proving the statistical significance of the results and indicating that this research could pave the way towards the usage of neuroimaging in conjunction with behavioral criteria in clinics. Furthermore, the salient features discovered in the brain structure of individuals with autism could lead to a more profound understanding of the underlying neurobiological mechanisms of the disorder, which remains one of the most substantial enigmas in the field even today.

Introduction

1.1 Background and Current Diagnostic Methods

This study aims to identify robust neuroimaging biomarkers for Autism Spectrum Disorder (ASD) with the goal of developing a more precise and effective diagnostic tool that proves effective for both male and female individuals on the autism spectrum. ASD is a complex neurodevelopmental disorder currently lacking objective diagnostic tools, which can lead to inconsistencies in diagnosis and treatment. The existing gold standard, the Autism Diagnostic Observation Schedule (ADOS), suffers from a variety of limitations such as subjective evaluations by clinicians, an inability to accurately distinguish between varying symptom severities, and inadequate early detection capabilities as it primarily assesses existing conditions in children. Objective diagnosis of autism based on unbiased medical data, like neuroimaging, holds the potential to provide a more accurate and dependable assessment of an individual's condition compared to subjective diagnostic methods that are prone to bias and variability among clinicians. Non-invasive

whole-brain scans may offer valuable insights for diagnosing neuropsychiatric disorders like autism, due to the neuroscientific nature of the disorder.

The assessment of ASD is filled with challenges due to the limitations of publicly available imaging datasets and the subjectivity of dominant diagnostic methods such as the ADOS or the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). In order to address these challenges and enhance the accuracy and reliability of ASD diagnosis, recent research has turned to machine learning and deep learning techniques for identifying unique neural patterns in brain imaging data from individuals with autism. With the use of advanced computational methods to investigate potential patterns in the neural underpinnings of autism, this study aims to contribute to the development of a more reliable and specific diagnostic tool for individuals on the autism spectrum.

Surveillance Year	Birth Year	Number of ADDM Sites Reporting	Combined Prevalence per 1,000 Children (Range Across ADDM Sites)	This is about 1 in X children
2020	2012	11	27.6 (23.1-44.9)	1 in 36
2018	2010	11	23.0 (16.5-38.9)	1 in 44
2016	2008	11	18.5 (18.0-19.1)	1 in 54
2014	2006	11	16.8 (13.1-29.3)	1 in 59
2012	2004	11	14.5 (8.2-24.6)	1 in 69
2010	2002	11	14.7 (5.7-21.9)	1 in 68
2008	2000	14	11.3 (4.8-21.2)	1 in 88
2006	1998	11	9.0 (4.2-12.1)	1 in 110
2004	1996	8	8.0 (4.6-9.8)	1 in 125
2002	1994	14	6.6 (3.3-10.6)	1 in 150
2000	1992	6	6.7 (4.5-9.9)	1 in 150

Figure 1. Autism rates over the years, according to the Centers for Disease Control and Prevention (CDC), showing the increasing prevalence of the disorder and relevancy of the problem.

The issue this study seeks to address is the challenge of accurately diagnosing ASD and overcoming the limitations of current methods, particularly those that depend on behavioral observations and psychological questionnaires, which are prone to resulting in false negatives and require significant time to administer. Given the rising prevalence of ASD (Figure 1), which is now estimated to affect more than 1% of the population (Chen, Chia-Min et al., 2019), the need for efficient and reliable diagnostic methods has become increasingly urgent.

ASD is a complex and heterogeneous disorder characterized by a variety of symptoms, including social impairments, repetitive behaviors, and difficulties with speech and communication. Due to its heterogeneity and wide spectrum, ASD often presents diagnostic challenges that may result in delayed or inaccurate identification (Bernhardt, Boris C et al., 2017). Traditional evaluation methods rely heavily on behavioral observations and psychological assessments, which are not only time-consuming but also subject to potential inaccuracies (Bone, Daniel et al., 2015). Given the increasing ASD prevalence and the limitations of existing diagnostic approaches, there is a pressing need to explore alternative methods that can more efficiently and effectively identify individuals on the autism spectrum. By focusing on the development of innovative diagnostic tools that harness advanced computational techniques and integrate multiple sources of data, this research aims to contribute to the optimization of ASD diagnosis and, ultimately, the enhancement of support and treatment options for those affected by the disorder.

1.2 Neuroimaging Biomarkers: What is Known

A diagnostic neuroimaging biomarker is a specific type of neuroimaging biomarker that is used to identify the presence of a particular disease or condition in an individual. It is a measurable feature detected through neuroimaging techniques, such as magnetic resonance imaging (MRI), that can differentiate between healthy individuals and those with the disease or condition of interest. Neuroimaging biomarkers have emerged as a promising area of research for understanding the neural correlates of neurodevelopmental disorders such as ASD. These biomarkers involve a variety of imaging techniques that provide valuable insights into the structural and functional organization of the brain and its role in the presence of ASD.

Functional magnetic resonance imaging (fMRI) has been widely used to investigate functional connectivity between brain regions in individuals with ASD. By measuring changes in blood flow as an indicator of brain activity, fMRI could identify abnormal patterns of connectivity that may significantly influence the core symptoms of the disorder. Structural magnetic resonance imaging (sMRI) examines brain morphology, such as gray and white matter volumes, cortical thickness, and surface area, in order to detect structural differences in the brains of individuals with ASD. Comparing these structural features across groups could enable the identification of potential biomarkers associated with the disorder. Diffusion tensor imaging (DTI) is another essential imaging technique that assesses the integrity of white matter by measuring water diffusion along axonal tracts. Previous studies investigating DTI have revealed abnormalities in white matter microstructure and connectivity in individuals with ASD, offering further insights into the neural substrates of the disorder.

As research progresses, the combination of multiple imaging modalities, such as fMRI, sMRI, and DTI, the most widely available and prevalent imaging methods in the field of neuroimaging, has immense potential to provide a comprehensive view of the structural and functional organization of the brain in individuals with ASD. This integration of techniques and tools can, therefore, lead to the development of an undeniably accurate and reliable diagnostic tool, which in turn produces targeted interventions for those affected by the disorder at hand. Recent advancements in computational methods, specifically neural networks, may also enable the analysis of large amounts of complex neuroimaging data, potentially leading to the discovery of novel biomarkers that were not identifiable before.

1.2.1 Why Neuroimaging Biomarkers are Relevant

Neuroimaging biomarkers are highly relevant in the context of neurodevelopmental disorders, such as ASD, because they provide a unique window into the underlying neural mechanisms that contribute to the wide range of symptoms and behaviors associated with these conditions. Traditional diagnostic

methods, which rely heavily on subjective behavioral observations and psychological assessments, can be time-consuming, less reliable, and subject to inaccuracies. Neuroimaging biomarkers, on the other hand, offer a more objective, quantifiable, and data-driven approach to understanding the complex biological underpinnings of neurodevelopmental disorders.

The identification of reliable neuroimaging biomarkers has the potential to revolutionize the diagnosis and treatment of ASD. By providing a more accurate and earlier diagnosis, these biomarkers can enable timely and targeted interventions, leading to improved outcomes for affected individuals and their families. In addition, the discovery of specific neural correlates associated with these disorders can guide the development of novel, personalized therapeutic strategies that take into account the unique neurobiological profile of each individual.

Neuroimaging biomarkers have the potential to significantly advance our understanding of the etiology and pathophysiology of neurodevelopmental disorders such as ASD. Unraveling the neural networks of the brain and their alterations in certain disorders may facilitate the identification of novel targets for intervention and prevention strategies. Furthermore, the integration of neuroimaging data with other sources of information, such as clinical diagnostic evaluations, can lead to a more comprehensive understanding of the complex interconnection between multiple determinants of neurodevelopmental disorders. The relevance of neuroimaging biomarkers in the study of neurodevelopmental disorders lies in their ability to provide objective, data-driven insights into the neural substrates of these conditions. This information holds tremendous potential for improving diagnostic accuracy, guiding personalized treatment approaches, and ultimately enhancing the quality of life for individuals affected by these disorders and their families.

1.2.2 Challenges with Neuroimaging Biomarkers

Despite the immense potential of neuroimaging biomarkers, there are several challenges that should be acknowledged and addressed. One of the primary issues is the heterogeneity of ASD itself, which manifests in diverse symptoms, cognitive profiles, and underlying genetic factors. This complexity makes it difficult to establish universally applicable biomarkers, as they may not capture the full range of variability present in the disorder.

Another challenge arises from the variability in neuroimaging data acquisition and processing. Different imaging sites employ varying imaging protocols, hardware, and analytical techniques, leading to inconsistencies in the findings and making it difficult to establish reliable and reproducible biomarkers. Standardization of these protocols and the development of harmonization techniques can help mitigate this issue and enhance the comparability of results across different studies.

The size and representativeness of study samples also pose significant challenges. Many neuroimaging studies have relatively small sample sizes, which can limit the generalizability of their findings. ASD research often underrepresents certain demographic groups, such as females and individuals from diverse cultural and socioeconomic backgrounds. Ensuring more inclusive and representative study samples is crucial for the identification of robust neuroimaging biomarkers that are applicable to the broader ASD population.

The integration of multimodal data and the application of advanced computational methods, such as machine learning and deep learning, present both opportunities and challenges. While these approaches hold great promise for the discovery of novel biomarkers and the refinement of our understanding of the neural underpinnings of ASD, they also require the development of appropriate analytical frameworks and the careful interpretation of results to avoid overfitting and falsified correlations. Addressing these challenges in the context of neuroimaging biomarkers for ASD is essential for improving diagnostic accuracy, guiding personalized treatment approaches, and advancing our understanding of the complex neural mechanisms underlying this heterogeneous disorder.

1.2.3 Past Research

The application of machine learning and deep learning techniques, particularly convolutional neural networks (CNNs), in neuroimaging research has shown potential in identifying biomarkers and

predicting autism based on MRI data (Dichter, Gabriel S., 2012). However, despite these promising developments, there remain several challenges and limitations that warrant further investigation and refinement.

Many of the previously developed models using machine learning and deep learning techniques have limitations in their performance metrics, including accuracy, precision, recall, and F1-score (El-Baz, Ayman, and Jasjit S Suri., 2020). These limitations may be partly attributed to the fact that most previous models have only considered shallow features from MRI data, rather than extracting more complex, higher-level features that may be more indicative of autism (Gabrielsen, T.P., Anderson, J.S., Stephenson, K.G. et al., 2018). Furthermore, the heterogeneity of ASD poses significant challenges for the identification of reliable and generalizable neuroimaging biomarkers. Many studies have focused on specific subgroups or aspects of autism, making it difficult to draw overarching conclusions about the disorder as a whole. Additionally, research samples are often small and lack representation from diverse demographic groups, limiting the generalizability of findings (Rahman, Md Mokhlesur et al., 2020).

Another critical issue is the need for standardization of data acquisition and processing methods across studies. Variability in imaging protocols may additionally lead to inconsistencies in results, hindering the establishment of reliable and reproducible biomarkers (Rashid, M, Singh, H, Goyal, V., 2020). While advanced computational methods such as deep learning have demonstrated the ability to extract complex patterns from neuroimaging data, they also present challenges in terms of interpretability and potential underfitting or overfitting. Ensuring the validity and robustness of these methods requires the development of appropriate analytical frameworks and cautious interpretation of results (Abraham et al., 2017). Although machine learning and deep learning techniques have made significant contributions to the field of neuroimaging biomarkers for ASD, there remains ample room for improvement and further exploration. Novel research that addresses these challenges and limitations can enhance our understanding of ASD's neural substrates and contribute to the development of more reliable, specific, and inclusive diagnostic tools for individuals on the autism spectrum.

1.3 Study Goals

To address these limitations and improve the accuracy of autism diagnosis, this study aims to build a novel deep learning model involving the use of 3D MRI data and preprocessing techniques such as slice time correction and normalization to extract features from a lower level to a higher level and improve classification accuracy for autism. By training the model on this data, the model aims to achieve an accuracy of over 95% in classifying autism versus typical controls, a significant improvement over previous solutions that have struggled to achieve similar levels of accuracy.

To address the exclusion of women in previous ASD research and improve classification accuracy, this model will incorporate data from equal populations of males and females with ASD and apply statistical techniques to identify features specific to each gender. It is important to use both male and female data in studies because gender can have significant effects on a wide range of biological, psychological, and social phenomena. This is particularly important in fields such as neuroimaging, where there may be significant differences in brain structure and function between males and females (Shahrjooihaghighi, Aliasghar et al., 2017). By including data from both males and females, this study will help to identify and control for these differences, ensuring that the findings from this study are more representative of the general population and are not biased towards one gender such as existing solutions, including ADOS, that were developed using a largely male sample and consequently make it more difficult to develop useful interventions or provide accurate diagnoses for girls and women (Thabtah, Fadi et al., 2019).

To address the biases and limitations in the current diagnostic methods, it is crucial to develop a more objective, specific, and functional diagnostic tool for ASD. To achieve this, this study aims to ensure that the model is trained on a diverse and representative dataset that includes a balanced representation of both ASD and typical control (TC) subjects. The dataset aims to encompass a variety of individuals, with equal proportions of males and females and adequate representation across different cultural and

linguistic backgrounds. By incorporating diversity within the dataset, the model aims to capture the variability in the general population, effectively avoiding overfitting to certain demographic subgroups.

In addition to creating objectivity, this study aims to enhance the specificity of the diagnostic tool created by leveraging novel, in-depth neuroimaging analysis techniques. This approach allows for the extraction of a comprehensive set of features that can accurately differentiate ASD from TC subjects. By examining a large number of features derived from three types of neuroimaging data, the model aims to provide a more nuanced and specific understanding of the neurobiological underpinnings of ASD. The diagnostic tool also aims to analyze the extracted neuroimaging features effectively. If successfully analyzed, the model results will help identify candidate diagnostic biomarkers for ASD, which can then be employed to improve the accuracy, specificity, and objectivity of ASD diagnostics. Ultimately, these advancements will lead to better identification and support for individuals affected by ASD.

Methodology

FPCN_vox125	0.78	0.75	0.8	0.85	0.78	0.78	0.82	0.75	0.78
DMN_conn67	0.72	0.65	0.72	0.78	0.72	0.74	0.75	0.65	0.72
AMYG_vox204	0.66	0.82	0.85	0.74	0.66	0.68	0.68	0.82	0.66
SMA_conn156	0.8	0.7	0.78	0.68	0.8	0.8	0.83	0.79	0.82
PCC_vox98	0.73	0.68	0.74	0.83	0.73	0.72	0.75	0.71	0.68
DLPFC_conn72	0.89	0.6	0.68	0.75	0.89	0.85	0.79	0.7	0.83
ACC_vox189	0.82	0.8	0.83	0.79	0.82	0.78	0.71	0.65	0.75
PPC_conn57	0.75	0.72	0.75	0.71	0.68	0.74	0.7	0.84	0.79
INS_vox314	0.68	0.85	0.79	0.7	0.83	0.68	0.65	0.77	0.71
TPJ_conn105	0.82	0.78	0.71	0.65	0.75	0.83	0.84	0.73	0.7
VMPFC_vox215	0.76	0.74	0.7	0.84	0.79	0.75	0.77	0.66	0.65
IFG_conn88	0.78	0.68	0.65	0.77	0.71	0.79	0.73	0.88	0.84
MTL_vox112	0.71	0.83	0.84	0.73	0.7	0.71	0.66	0.8	0.77
SPL_conn93	0.83	0.75	0.77	0.66	0.65	0.7	0.88	0.77	0.73
V1_vox301	0.77	0.79	0.73	0.88	0.84	0.65	0.8	0.7	0.66
STG_conn76	0.7	0.71	0.66	0.8	0.77	0.84	0.72	0.76	0.88
AG_vox137	0.63	0.7	0.88	0.77	0.73	0.77	0.88	0.67	0.8
OFC_conn114	0.79	0.65	0.8	0.7	0.66	0.73	0.81	0.81	0.77
PHG_vox212	0.72	0.84	0.72	0.76	0.88	0.66	0.76	0.74	0.7
IPL_conn99	0.88	0.77	0.88	0.67	0.8	0.88	0.69	0.85	0.76
SFG_vox178	0.81	0.73	0.81	0.81	0.77	0.8	0.74	0.78	0.74
FFA_conn85	0.76	0.66	0.76	0.74	0.7	0.77	0.67	0.77	0.67
V2_vox305	0.69	0.88	0.69	0.85	0.76	0.7	0.87	0.7	0.87
MTG_conn103	0.74	0.8	0.74	0.78	0.74	0.8	0.74	0.78	0.74
CAU_vox224	0.67	0.77	0.67	0.77	0.67	0.77	0.67	0.77	0.67
THA_conn91	0.87	0.7	0.87	0.7	0.87	0.7	0.87	0.7	0.87
CB_vox248	0.79	0.76	0.79	0.76	0.79	0.76	0.79	0.76	0.7
PUT_conn80	0.84	0.67	0.84	0.67	0.84	0.67	0.84	0.67	0.65
STS_vox176	0.76	0.81	0.76	0.81	0.76	0.81	0.76	0.81	0.84
IC_conn78	0.71	0.74	0.71	0.74	0.71	0.74	0.71	0.74	0.77
S1_vox285	0.65	0.85	0.65	0.85	0.65	0.85	0.65	0.85	0.65
M1_conn95	0.85	0.78	0.85	0.78	0.85	0.78	0.85	0.78	0.85

Figure 2. Data table depicting metrics retrieved from the CCS-NYU site in the Autism Brain Imaging Data Exchange.

2.1 Aggregating and Understanding the Data

2.1.1 Collecting Magnetic Resonance Images

The raw MRI data utilized in this study was sourced from the Autism Brain Imaging Data Exchange I (ABIDE-I) dataset, obtained via the Connectome Computation System (CCS) pipeline. As a publicly accessible dataset, ABIDE-I is widely recognized and utilized by the neuroimaging community, containing a substantial sample of subjects diagnosed with ASD as well as typical controls (TC). The focus of this investigation is centered on the CCS-New York University (NYU) Langone Medical Center site, which is one of 17 available sites within the dataset. To ensure consistency and comparability across the entire sample, the CCS ABIDE data underwent preprocessing using standardized pipelines that were applied uniformly to all subjects. The initial stage of data processing involved loading the three-dimensional (3D) MRI data and converting it into two-dimensional (2D) images.

Slice time corrections and normalizations were implemented to normalize the data, effectively adjusting the images to meet the necessary specifications for further analysis. The use of standardized preprocessing pipelines and parameters is a crucial component of the study's

methodology, as it guarantees that any variations observed in the data are attributable to the inherent biological differences between subjects rather than disparities in the preprocessing steps. In addition, necessary steps were taken on each of the three distinct types of imaging used to prepare them for analysis. By adhering to these meticulous standards, this study ensures the validity of the data and the validity of the subsequent analyses, thus contributing valuable insights into the complex world of neuroimaging and our understanding of autism.

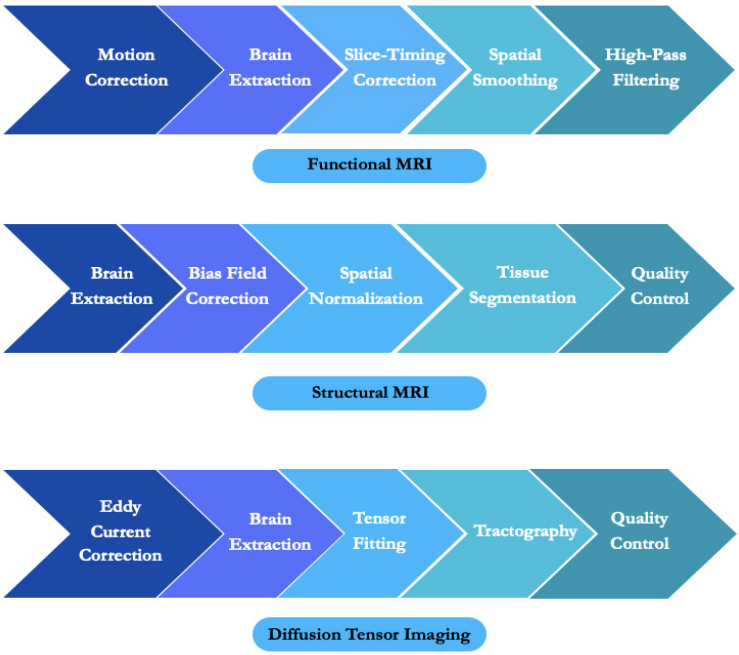


Figure 3. Diagram illustrating the distinct preprocessing procedures used on the imaging data.

2.1.2 Data Preprocessing

A total of 1,111 samples was aggregated, drawing from an extensive array of seven university study databases. For each individual, the dataset includes one fMRI scan, one sMRI scan, and one DTI scan. This dataset comprised a total of 3,333 neuroimaging files, 1 of each type of imaging per sample, creating a large and diverse dataset that significantly expands on the amount of patients used in previous neuroimaging studies concerning ASD. An equal amount of ASD and TC samples were selected along with an equal amount of male and female samples, resulting in a representative selection of subjects.

To maintain consistency and optimize the comparability of results across various imaging modalities, a unique preprocessing pipeline has been employed for each type of neuroimaging data. This standardization process ensures that the data are normalized and standardized, allowing for more accurate and reliable analyses. The widely-used FMRIB Software Library (FSL) was utilized in this procedure, providing a robust and efficient platform for visualizing and preprocessing the neuroimaging scans. Access to FSL was facilitated through the use of a terminal interface, streamlining the overall processing workflow.

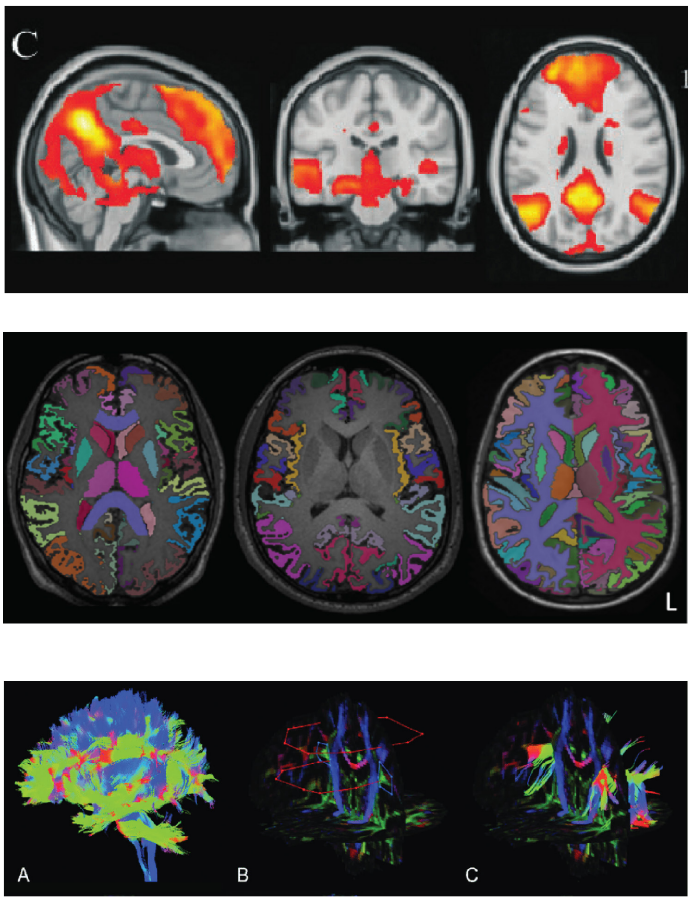


Figure 4. Visualization of the standardized neuroimages used for feature extraction.

2.1.3 Feature Extraction

The application of mathematical algorithms in the field of neuroimaging analysis has revolutionized the way we can interpret and understand complex brain activity patterns. Among these tools is the computational library SimpleITK, which uses advanced algorithms to rapidly extract hundreds of features from various aspects of neuroimages, such as activity in specific brain regions and the connectivity between these areas. This comprehensive extraction process enables a more in-depth investigation into the intricacies of brain function and the underlying mechanisms of various neurological disorders.

Following the extraction of an extensive array of features from the neuroimaging data, it is crucial to identify the most relevant and informative features for inclusion in the analytical model. This step is essential to minimize the risk of overfitting or underfitting the model, which could lead to inaccurate predictions and hinder the overall effectiveness of the analysis. To accomplish this, the Recursive Feature Elimination (RFE) technique was employed, a widely-used and efficient algorithm that systematically determines the most informative features for a given classification problem.

The RFE algorithm accomplishes this task by iteratively evaluating the importance of each feature and removing the least significant ones, ultimately refining the feature set to include only those that contribute the most to the model's predictive power. Based on the feature rankings generated by the RFE algorithm, a decision was made to utilize the top 106 most important features from each imaging type for the development of the analytical model as they were among the top 10% of features as ranked by the algorithm. This selection process ensures that the model is both robust and reliable, maximizing its potential to provide valuable insights into the complex world of neuroimaging data and contribute to our understanding of brain function and dysfunction.

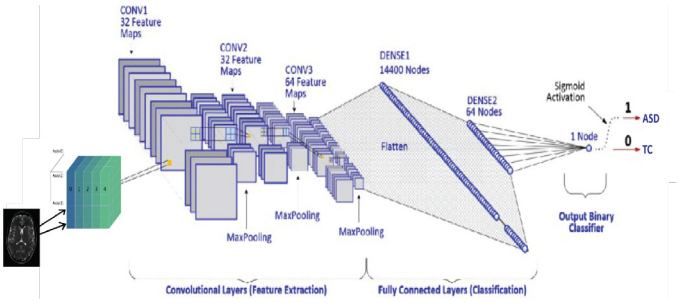


Figure 5. Visualization of CNN transforming the input feature values into an output prediction.

2.2 Deep Learning

2.2.1 Development, Training, and Testing of Models

The deep learning methodology employed in the model building process is a crucial aspect of any research study, as it serves as the foundation upon which the validity and reliability of the results are built. In the present analysis, a total of 777 samples were utilized for the construction of the model, comprising approximately 70% of the entire dataset. This sizable subset of the data was carefully selected to ensure that the model is both representative and generalizable, thus increasing the likelihood of obtaining accurate and meaningful results.

The independent variables (X) in the model consisted of neuroimaging features extracted from the various types of scans. These features provided a comprehensive overview of the different aspects of brain activity and connectivity, allowing for a more in-depth investigation into the relationships between these variables and the dependent variable (Y), the diagnostic group to which each subject belongs.

To enhance the robustness and reliability of the model, the analysis was iterated three times, each time utilizing different input values derived from the fMRI, sMRI, and DTI scans. This iterative approach allowed for a more thorough evaluation of the model's performance across various combinations of input values, ensuring that the final model is well-equipped to handle the complexities and nuances of the neuroimaging data. By incorporating a large sample size, a diverse array of neuroimaging features, and an iterative model-building process, the methodology employed in this study was designed to maximize the potential for obtaining meaningful and actionable insights. These results not only contribute to our understanding of the underlying mechanisms governing brain function and dysfunction but also hold promise for the development of more accurate diagnostic tools and targeted interventions for individuals affected by neurological disorders.

In this study, the performance of the proposed models was thoroughly assessed using four benchmark metrics, providing a comprehensive evaluation of the models' effectiveness in classifying subjects based on their neuroimaging data. To ensure the validity and reliability of the results, a 10-fold cross-validation procedure was conducted, which involved partitioning the data into ten equal subsets and iteratively training and testing the models on these subsets.

The objective of the analysis was to identify the best performing imaging type for classification, with a focus on the average accuracy, F1-score, precision, and recall. These metrics were carefully chosen to provide a well-rounded assessment of the models' performance, taking into account factors such as the proportion of correct predictions (accuracy), the balance between precision and recall (F1-score), the ability to correctly predict the ASD class (true positives) and typical controls (true negatives), and the rates of incorrect predictions for ASD (false positives) and typical controls (false negatives). An in-depth search was conducted to learn the optimal hyperparameters for each model, leading to numeric results (Figure 6).

Performance Measures	fMRI	sMRI	DTI
Accuracy	95.40%	74.70%	69.30%
Precision	94.50%	73.60%	68.90%
Recall	92.50%	73.30%	66.50%
F1-Score	95%	74.50%	69.10%
AUC Score	94.70%	74.33%	68.93%

Figure 6. Performance metrics of the model tested on different types of imaging data.

2.2.2 Results

The evaluation of model performance metrics suggests that fMRI may be the most informative type of neuroimaging data for autism classification. This finding underscores the potential of fMRI as a valuable diagnostic tool in the field of autism research and clinical practice. Consequently,

further in-depth analysis was conducted on the highest performing model, focusing on the fMRI data.

Several methods were employed to better understand and evaluate the performance of the fMRI model, including a confusion matrix. This matrix provides a summary of the model's performance by displaying the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). By presenting these values in a comprehensive format, the confusion matrix offers valuable insights into the model's ability to correctly classify subjects and identify potential areas for improvement. In addition, a receiver operating characteristic (ROC) curve was used to model the metrics of the model. The ROC curve illustrates the trade-off between the TP rate (TPR) and the FP rate (FPR) as the threshold for classification is varied. This graphical representation allows us to assess the overall performance of the model across different classification thresholds and determine the optimal balance between sensitivity and specificity. A train versus validation accuracy model was also employed. This analysis involves examining the trend of the model's accuracy during the training and testing phases. The training accuracy measures how well the model fits the training data, while the validation accuracy evaluates the model's ability to generalize to new, unseen data. By comparing these two measures, future research can better understand the model's performance and identify potential issues such as overfitting or underfitting.

Based on the evaluation of the five benchmark metrics, it is evident that the fMRI model outperforms the vast majority of similar models in the field, but additionally has the potential to improve upon the metrics of clinical practices in theory if the results remain constant throughout larger independent datasets. 10-fold-cross-validation was additionally performed to validate, in theory, how the model would perform on an independent dataset, and the results remained largely constant with a variance in metrics less than 1%.

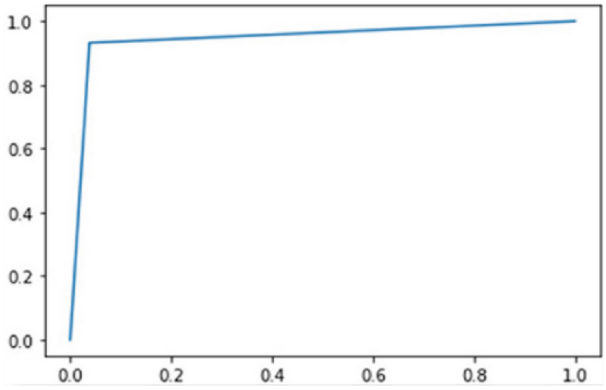


Figure 7. Displays ROC curve of fMRI model, generated through TensorFlow software.

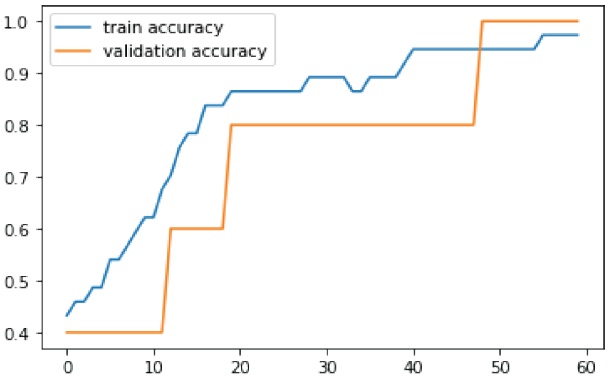


Figure 8. Displays train accuracy versus validation accuracy of fMRI model, generated through TensorFlow software.

This high performance highlights the potential of fMRI as a powerful tool for the accurate classification of autism and highlights the importance of continued research and development in the area of neuroimaging-based diagnostics and interventions for individuals affected by ASD.

Data Analysis and Results

3.1 Statistical Analyses for Biomarker Selection

In this study, the optimal features extracted from the functional imaging data were identified using independent sampling t-tests, a statistical method designed to compare the means of two groups. The top 32 features exhibiting the highest differences in mean values between individuals with autism and typical control subjects were chosen as candidate biomarkers as these potential biomarkers exhibited p-values less than 0.001, indicating a high degree

of statistical significance and supporting their relevance in distinguishing between the two groups.

3.2 Male and Female Layer Separation

Notably, the analysis revealed distinct biomarkers for each sex group, underscoring the importance of considering sex differences when examining the neural underpinnings of autism. For biological females, the most significant biomarkers included activity in the amygdala, activity in the ventromedial prefrontal cortex, activation in the visual cortex, connectivity between the default mode network and the frontoparietal network, and activation in the superior temporal gyrus. For biological males, the most prominent biomarkers included activation in the dorsolateral prefrontal cortex, activity in the caudate nucleus, connectivity between the cerebellum and the prefrontal cortex, activation in the anterior cingulate cortex, activity in the fusiform gyrus. In addition to these gender-specific biomarkers, 22 other features were identified for each gender group, further emphasizing the unique neural characteristics associated with autism in males and females. By retrieving features and metrics separately for data labeled as male and female, this study was able to identify distinct biomarkers for each gender group. The highly differential results between groups of this approach not only highlights the importance of considering sex differences in autism research but also holds promise for the development of more targeted and effective diagnostic tools and interventions tailored to the unique needs of individuals affected by ASD.

3.3 Differences in Imaging Protocols

The observed slight variation in mean metrics between datasets highlights the potential influence of differences in imaging protocols between sites on the results. To combat this issue, future studies may investigate the effect of the methodology used for getting MRI scans on the results and accuracy of resulting biomarkers, leading to a potential standardized approach for medical imaging collection.

Conclusions and Discussion

The study presented several key findings and implications. Among its accomplishments included the extraction of 106 neuroimaging features from 1,111 unique patients and the successful assembly of a robust dataset suitable for deep learning applications. The initial study demonstrated that the proposed model achieved an impressive 95.4% accuracy, outperforming previous statistical analyses conducted on the DSM-5 and ADOS-2 diagnostics. Moreover, t-tests suggested that functional magnetic resonance imaging (fMRI) might be the most informative type of neuroimaging data for autism classification, offering valuable insights for future research and clinical practice. The study's findings have significant implications for various aspects of autism research and treatment. For instance, early detection and diagnosis can be greatly improved by using the identified candidate biomarkers. Detecting autism at an early stage is crucial for enhancing treatment outcomes and improving the quality of life for affected individuals and their families. Furthermore, the identification of neuroimaging biomarkers for autism can contribute to a better understanding of the underlying biology of the condition. This increased understanding can, in turn, lead to the development of novel treatments and therapies that target the root causes of autism. Additionally, neuroimaging biomarkers can be employed to identify patient subgroups for clinical trials, which can enhance the efficiency and effectiveness of these trials. This can result in faster development of new treatments and therapies for autism, ultimately benefiting those affected by the condition and their support networks.

Acknowledgments

I would like to thank my mentors: Dr. Jill E. Howard, Assistant Professor in Psychiatry and Behavioral Sciences at Duke University's Child & Family Mental Health & Community Psychiatry Division, and Mr. Silas Busch, PhD Candidate at the University of Chicago, for their generous critiques, comprehensive answers to my inquiries, and unwavering support of my research undertakings in recent months. I would also like to thank Ms. Kate Allender of Nikola Tesla STEM High School for equipping me with the necessary resources for rigorous scientific exploration and consistently nurturing my intellectual pursuits with unwavering encouragement.

References

Bernhardt, Boris C et al. "Neuroimaging-Based Phenotyping of the Autism Spectrum." Current topics in behavioral neurosciences vol. 30 (2017): 341-355. doi:10.1007/7854_2016_438

Bone, Daniel et al. "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." Journal of autism and developmental disorders vol. 45,5 (2015): 1121-36. doi:10.1007/s10803-014-2268-6

Chen, Chia-Min et al. "Deriving and validating biomarkers associated with autism spectrum disorders from a large-scale resting-state database." Scientific reports vol. 9,1 9043. 21 Jun. 2019, doi:10.1038/s41598-019-45465-9

Dichter, Gabriel S. "Functional magnetic resonance imaging of autism spectrum disorders." Dialogues in clinical neuroscience vol. 14,3 (2012): 319-51. doi:10.31887/DCNS.2012.14.3/gdichter

El-Baz, Ayman, and Jasjit S Suri. Neurological Disorders and Imaging Physics. Vol. 3, IOP Publishing, 2020.

Gabrielsen, T.P., Anderson, J.S., Stephenson, K.G. et al. Functional MRI connectivity of children with autism and low verbal and cognitive performance. Molecular Autism 9, 67 (2018). https://doi.org/10.1186/s13229-018-0248-y

Rahman, Md Mokhlesur et al. "A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder." Brain sciences vol. 10,12 949. 7 Dec. 2020, doi:10.3390/brainsci10120949

Rashid, M, Singh, H, Goyal, V. The use of machine learning and deep learning algorithms in functional magnetic resonance imaging—A systematic review. Expert Systems. 2020; 37:e12644. https://doi.org/10.1111/exsy.12644

Shahrjooihaghighi, Aliasghar et al. "An Ensemble Feature Selection Method for Biomarker Discovery." Proceedings of the ... IEEE International Symposium on Signal Processing and Information Technology. IEEE International Symposium on Signal Processing and Information Technology vol. 2017 (2017): 416-421. doi:10.1109/ISSPIT.2017.8388679

Thabtah, Fadi et al. "A machine learning autism classification based on logistic regression analysis." Health information science and systems vol. 7,1 12. 1 Jun. 2019, doi:10.1007/s13755-019-0073-5

EEG-Based Analysis of Brain Activity in Individuals with Varying Levels of Social Media Usage

Venkata Pranay Kumar Nookala & Sri Sai Tejas Verma Pakalapati

Nikola Tesla STEM High School, Redmond, WA, 98052, USA

Dr. Eric H. Chudler

Center for Neurotechnology, Gates Center Bldg., 3800 E Stevens Way NE

Abstract:

Social media's widespread usage in current generation Z, has raised questions regarding its impact on human brain activity. This study had aimed to investigate the differences in electroencephalogram (EEG) brain activity of individuals with varying levels of social media usage in order to test where there is a negative correlation between social media usage and brain activity. The study recruited 65 participants from ages 14-18, the participants filled out a form compiled with questions regarding their social media usage per week. The responses were categorized as low (0-14 hours), medium (14-30 hours), or (high 30+ hours) per week. 4 trials were run on each participant, where the EEG activity was recorded with their eyes closed and eyes opened for 5 minutes, and when they were viewing preset social media content, and their own social media content for 15 minutes. The results showed that heavy social media users have significantly decreased brain activity in the alpha(3-7hz) and beta (12.5-30hz) wave frequency compared to infrequent users during the tasks. By using EEGLAB to perform group analysis of the time-frequency graphs we obtained a P value of 0.023 proving our results were statistically significant and allowing us to reject our null hypothesis that different levels of social media usage for individuals, don't have any effect on brain activity. This study concludes that there is a negative correlation between social media usage and brain activity, meaning as social media increases brain activity decreases.

Thesis:

This research study seeks to understand and find brain activity changes in individuals with differing levels of social media usage ranging from casual/low social media usage to addiction/high social media usage, the study would be done using an electroencephalogram, a scientific measuring device which measures electrical activity of the brain using electrodes which are placed on the scalp.

Introduction:

An overwhelming, 75% of kids between the ages of thirteen and seventeen actively use social media (Woodward et al). In today's world, social media is key and needed to keep in touch with everything occurring in the world. From political information to news about TV-shows, social media is the go-to, it's the form of communication for young adults. Social media apps use specific strategies to keep users hooked onto their app, with features such as for you page, suggested posts, etc. Companies' trap young adults and start impacting them in ways they're too unconscious and immature to understand. As more information is being transported via social media apps by the day, teenagers are increasing their usage within these apps and becoming more addicted.

One of the many negative impacts of social media is its effect on teenagers' mental health as human beings are naturally social, known to be as social creatures (Robinson et al). We require the company of others to succeed in life and the strength of our bonds has a significant influence on our happiness and mental health. By being socially linked to others it can, "ease stress, anxiety and depression, boost self-worth, provide comfort and joy, prevent loneliness, and even add years to your life." (Robinson et al, 2022). The Robinson study and research was identifying how being socially linked to others, and having face to face interaction has many positives and is what keeps us healthy and alive. However, on the other side, having a lack of social relationships and strong bonds can endanger your mental and emotional wellness. Where in today's world teenagers are missing out the strong relationships we once had earlier and replacing them with social media platforms and the key qualities it offers. Many teenagers rely on social media platforms like TikTok, Twitter, YouTube, Instagram, and Snapchat. Social media has its own advantages, however it's crucial to remember that social media can never fully replace face to face human interaction. Many teenagers change their mentality and fall for the illusion that social media can replace face-to-face interaction. Face to face interaction is what allows for hormones which reduce stress and make you feel happier to be activated. Social media does the opposite, and it leads to increased stress and less situations where an individual is happy. Specifically, many surveys have been conducted to see the real effects of all the extensive use of social media. In 2017, a survey conducted by the Royal Society for Public Health (United Kingdom) showed that Instagram and Snapchat were ranked the worst platform for mental health, and it led to all sorts of negative influences on the teen.

While everyone talks about how great social media is, as it provides key information not many individuals address the negative impacts it holds on to people, specifically teenagers. Young adults are more feeble-minded and as mentioned before are easily trapped in the depths of social media. A research article that was published in 2019, reported that 71% of teenagers surveyed reporting that they experienced anxiety when they were unable to access their social media account (Vannuci et al., 2017). The negative effects on them include depression and anxiety which both influence an individual's mental health. Social media addiction can be defined as a type of psychological dependence that develops through cognitive, sensual, and behavioral processes and produces social, academic, or professional negative results in an individual's life (Serdar). As a result, occupational and recreational activities may be lost, which inevitably leads to social isolation. This also corresponds to the brain's cognitive decisions, as evidence from studies done by researchers, Dr. Trujillo and Stanford explore how functional brain networks interact on local and global dimensions produce behavior (Trujillo et al). Many symptoms from social can also lead to said isolation such as being left out, like count, and bullying (Almond Solutions et al).

The usage of social media varies from person to person but as our world is turning more technological every second users are spending more time than ever. As mentioned by many previous research studies, specifically the one done by JLG, where the electroencephalogram was able to detect biomarkers for depression and could detect other waves the research states, "The advancements in electroencephalography (EEG) make it a powerful tool for non-invasive studies on neurological disorders including depression" (JLG et al., 2016). Those teenagers suffering with anxiety and depression, are often linked to their extensive use in social media. Most of these teenagers who are suffering aren't even aware of their current situation as it happens unconsciously. By being able to record EEG brain activity using the electrodes we can conclude and detect depression biomarkers (JLG et al). This can provide teenagers with valuable information about how social media is affecting them. So far in the scientific world here hasn't been research done that quantifies the brain patterns/ activity that may occur with EEG data. A possible solution would be using the EEG to take data recordings, analyze them and make interpretations on the brain patterns and activity that may be similar across the board, for those with varying levels of social media usage. Similar studies which have been conducted in the past, were able to support the hypothesis that electroencephalograms are able to recognize stress.

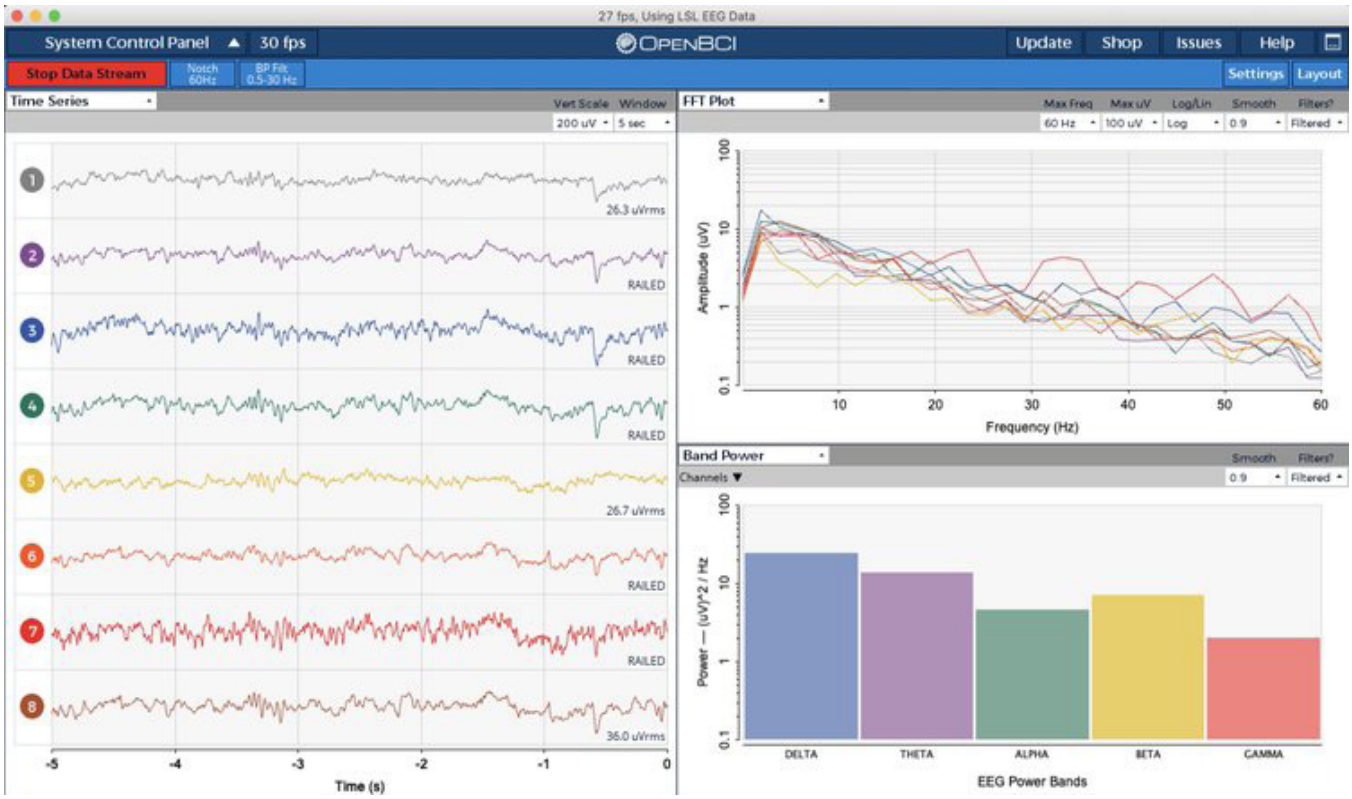


FIGURE 1: ELECTROENCEPHALOGRAM BRAIN ACTIVITY PLOTTED ON THE OPENBCI SOFTWARE, EACH GRAPH INDICATES AN ELECTRODE AND A FFT PLOT IS THE COMBINATION OF ALL OF THE ELECTRODES BRAIN ACTIVITY

An electroencephalogram (EEG) is a medical device that uses tiny metal discs known as electrodes which are connected to the scalp in order to measure the electrical activity in an individual's brain. Brain cells interact by electrical impulses and are constantly active, even while sleeping. On an EEG recording, this activity appears as wavy lines. One of the primary diagnostic tests for epilepsy is an EEG. An EEG can also be used to diagnose various types of brain diseases (Mayo Clinic). Some of these brain diseases include anxiety and depression. As mentioned before anxiety and depression are heavily associated with increased levels of social media usage and our EEG data could be a leading factor in finding the correlation between a disorder or brain pattern and activity on social media (Zander et al). Also, "Studies show that patients with depression have hemispheric asymmetry in their brain signals when compared to non-depressed subjects" (Acharya et al., 2015).

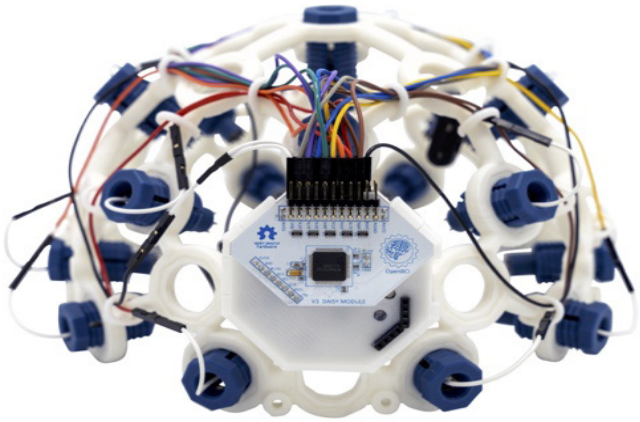


FIGURE 2: IMAGE OF AN ELECTROENCEPHALOGRAM FROM THE OPENBCI COMPANY, 3-D PRINTED WITH THE ELECTRODES BEING ABLE TO ATTACH ONTO THE SCALP. ON THE BACK OF THE EEG IS A CYTON ARDUINO BOARD, WHICH ALLOWS IT TO COMMUNICATE TO THE COMPUTER WHEN ATTACHED TO THE DONGLE.

By being able to identify the science behind it, an individual's brain patterns and activity changes when addicted to social media it can help bring a change upon teenagers in the current world. As a whopping 24 percent of teenagers are addicted to social media (JLG et al), we can help bring a change upon them as they most likely are exposed to depression and anxiety symptoms. By preforming this research study, the expected outcomes are hoped to be brain patterns or activity that is similar between those with varying levels of social media usage to prove the effects of social media on the brain.

Methodology:

For the research study, individuals were recruited through an interest form made with Google Forms. The questions on this form contained the following information about the participants: Age, Email, Biological Sex, Any Recent Surgeries, Any Medical Conditions, and any Significant Information such as, if they have a disability, recent brain surgery, any issues with mental health etc. Along with the QR code of this form, flyers and other posts were created and posted at local libraries: Redmond Library and Sammamish Library as well as nearby schools: Redmond High School, Eastlake High School, Tesla STEM High School to recruit participants. Once all the responses were collected, they were filtered according to whether they were qualified to perform in our study. The next step was to assign a random number, using a random number generator to the qualified participant. They were sent the informed consent form, an informational flyer, parents' consent form in case the participant was under-age (below 18 years of age). Also based off their forms response they will be separated into three groups accordingly, to their usage on social media the groups are: higher levels of social media usage, medium levels of usage, lower levels of social media usage. This study is a single blind study and we received 60 participants (n = 60).

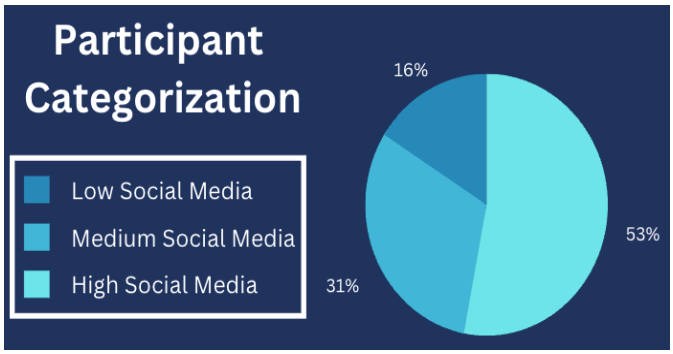


Figure 3: This research study's categorization of participants based on the weekly hours on social media

Next, the participants arrived at the location to take a recording of their brain waves and activity through the EEG. They reviewed a document outlining what will happen during their appointment. After the participant had confirmed that they have washed their hair the night before for their scalp to be clean for the best and accurate results, the EEG testing process had begun, it started with the cleaning of the EEG cap and ensuring the safety and comfort of the patient. When everything had been settled, OpenBCI GUI was launched, and we started the system control panel and followed the steps to accurately setup the recording of the EEG. Then we placed the headset on the human participant and connected the reference electrodes to the participants ears. We then started data stream and started recording the data.

Once taken to the testing and recording phase of the experiment this part included the preprocessing of the EEG data which was received from the patient's test. To do this we saved the EEG recording to the designated folder and then put this into a GitHub. Which we were able access to MATLAB a software. We modified the code within the GitHub's files to fit the needs of the EEG recording which had just been taken. We had the software plot the graph, this contained data which had eye blinks indicated by the drops of voltage in the graphs. To fix this and further preprocess the data, we used the MATLAB commands to create a detrended EEG Recording. It gave us the d200 data which included the raw EEG Signal from the electrodes. Along with these eye blinks, some of these EEG recordings will also have noise as a part

of them and there's a separate MATLAB command that filters out the noise. It took the data and then outputted three graphs: Raw EEG Signal, De-Trended Signal, and the After Moving Average Filter (Noise Reduction) Filter. Along with this there are also peaks in the code and the graphs which we had to detect and to add the variance peaks we used another command which gave us another graph putting us at 4 graphs with the addition of the After Moving Variance Filter. After preprocessing the data, they were used for data analysis after the social media stimulus testing.

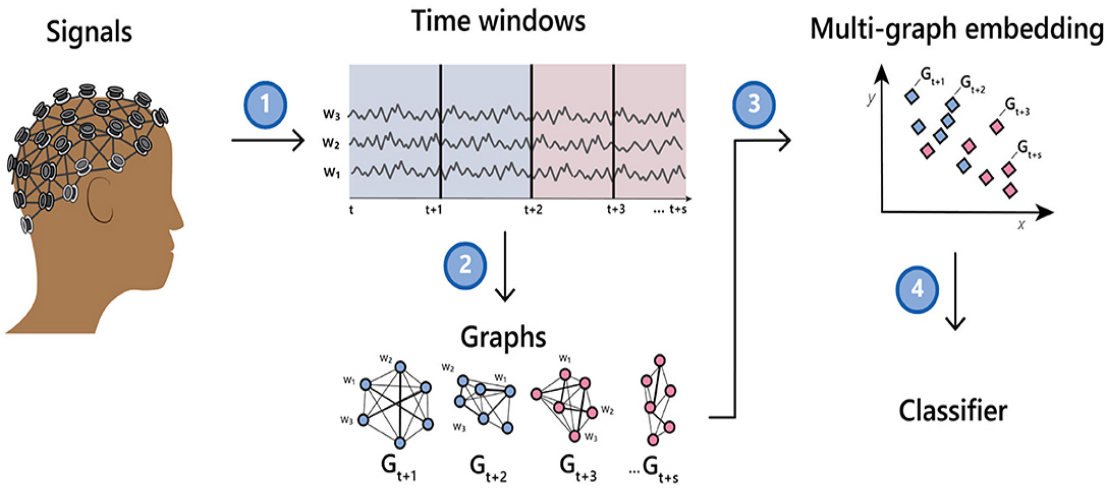


FIGURE 4: EXAMPLE OF HOW THE FOUR GRAPHS (THE TIME WINDOWS) ARE GENERATED THROUGH THE SIGNALS AND CHANGED TO THE MULTIGRAPH EMBEDDING FOR CLASSIFICATION AND DATA ANALYSIS

The social media stimulus testing was what we had the participants to do, which was to pull out their phone and open TikTok and start browsing content on them for you page, then their Instagram and browse content on their suggested posts and Instagram Reels, after was Snapchat, where they were asked to browse content on Snapchat spotlight. We had these steps repeated according to the application they were using, and then we saved the recordings to a different file for further analysis. Having done this, the work with the participant was over and so we helped them clean up by disconnecting all the pieces, removing the headset, and cleaning up the cap.

After this was the data analysis in which we analyzed the data that we had received from the EEG recordings. We were able to analyze the data through the different waves: Alpha, Beta, Gamma, Theta, Delta waves. These waves would be on the graph and can be read through a model and analyzed to make accurate predictions/correlations (Abhang et al). To the right, is an example of how the Gamma, Beta, Alpha, Theta, and Delta waves appear on graphs, with the y-axis representing the hertz (brain activity).

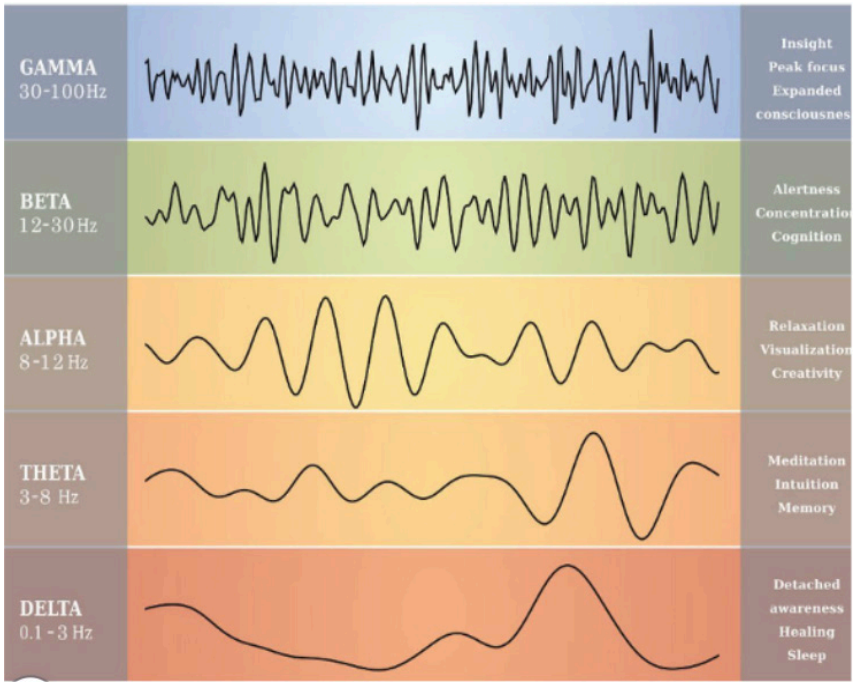


FIGURE 5: HOW GAMMA, BETA, ALPHA, THETA, AND DELTA WAVES APPEAR ON GRAPHS THE Y AXIS REPRESENTING THE FREQUENCY OF BRAIN ACTIVITY. THE RIGHT OF THE FIGURE IDENTIFIES THE USE OF EACH BRAIN WAVE.

After achieving all of the participants results on the form, they were inputted onto a data table for organization and categorization.

TABLE 1: CATEGORIZATION OF PARTICIPANTS FOR EXPERIMENTAL PROCEDURE, RANKING EACH INDIVIDUAL ON A SCALE OF LOW, MEDIUM OR HIGH SOCIAL MEDIA USAGE BASED ON THEIR HOURS ON SOCIAL MEDIA PER WEEK.

Individual (Name):	Hours on Social Media (Weekly):	Categorization (Low, Medium or High Social Media Usage)
Ex: John Doe	10.1	Low
Ex: Pranay Nookala	22.3	Medium
Ex: Tejas Pakalapati	34.7	High

Results:

The most critiqued part in this study was the results, to see if the hypothesis for this research study was statistically significant. After extensive background research on the raging dilemma across teenagers in generation Z, the hypothesis stated was that there was a negative correlation between the two variables: brain activity in individuals and weekly hours on social media. After the hundreds of hours poured into, conducting Electroencephalogram scans on social media users and participants of this study, results were achieved. The results for every single participant were grouped based upon their weekly social media usage (hourly), the three groups were low social media usage being 0-14 hours, medium social media usage 14-30 hours and high social media usage 30+ hours per week. To be able to understand if there was truly, a negative correlation the groups data was averaged into time frequency graphs, from where brain activity could be determined. Additionally, it would be later used to determine the impacts on the beta and alpha waves in the human brain. However, to see the overall shape and impact of how certain things affected the scan, the trials and conditions for the

EEG scans were averaged out. The first two conditions, where the individual had their eyes opened and closed were done to set a baseline for the scans, to have a reference point to compare the social media effect too. Also, to be able to make sure the participant was in a state of homeostasis and to find out how each type of condition differs from the state of homeostasis. When comparing between conditions three and four where the individual is on social media, whether it be on their own for you page or the preset social media content; the brain activity changes are huge and are comparably much higher than when the individual is in the state of homeostasis.

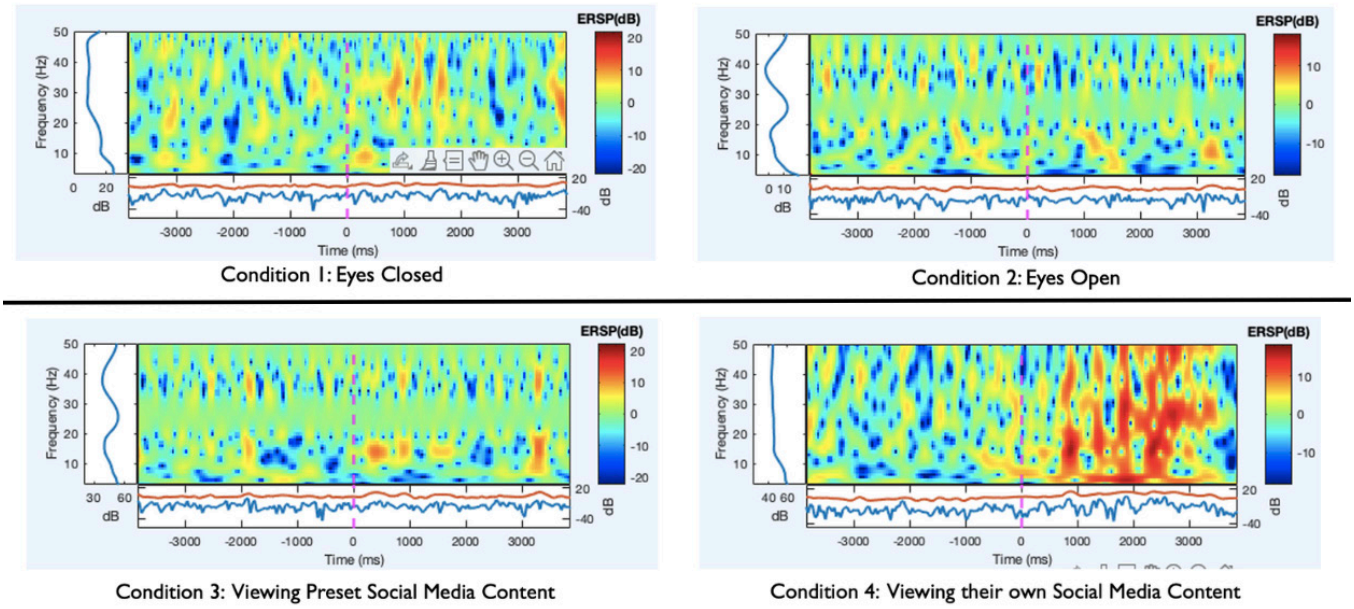


FIGURE 6: AVERAGE TIME-FREQUENCY GRAPHS FOR THE CONDITIONS TESTED ON EACH PARTICIPANT FOR THIS STUDY. THE RIGHT OF EACH GRAPH INCLUDES A SCALE OF THE EXTENT OF BRAIN ACTIVITY. THE Y-AXIS IS THE FREQUENCY (HERTZ) AND X-AXIS BEING TIME (MILLISECONDS).

Further analysis of these graphs was able to determine that individuals (teenagers) who are active on social media constantly experienced diminished frequency. Depicted in the graphs below, the red shows that an individual's brain activity is highly prominent. As seen towards the right of each graph a scale of the ERSP (dB) of where Red HIGH brain activity is while green and blue are rather LOW brain activity. The clear differences between low and high social media usage time frequency graphs supports the idea of a negative correlation. To be able to understand the cognitive impact on these individuals, the rate and the time of which the brain activity spiked/decreased was analyzed. The region of frequency that is correlated with the brain activity that's being impacted is 4-7 hertz which corresponds to brain wave: alpha and another frequency of which cognitive utilization was impacted was 12.5-30 hertz which corresponds with brain wave: beta.

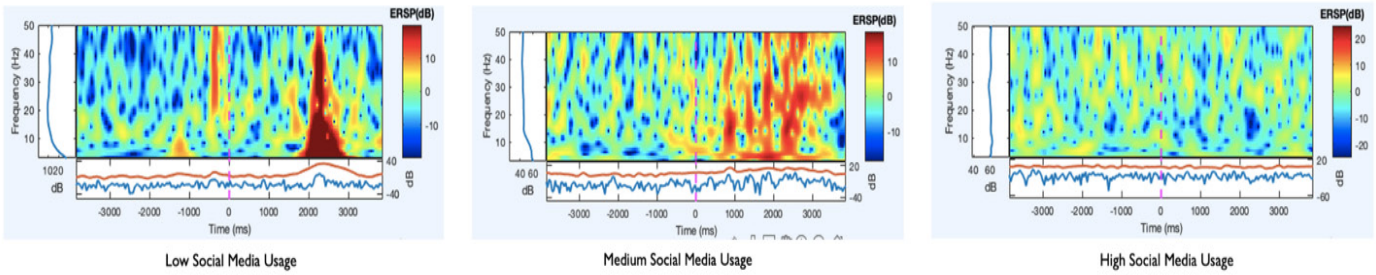


FIGURE 7: AVERAGE TIME FREQUENCY GRAPHS FOR ALL THE CATEGORIES OF THIS RESEARCH STUDY (LOW, MEDIUM, HIGH). SHOWS SPIKES AT 2000 MILLISECONDS, WHERE THE CHANGE FOR EACH CATEGORY CAN BE SEEN. Y- AXIS REMAIN FREQUENCY (HERTZ), X-AXIS REMAINS TIME (MILLISECONDS).

To deem this experiment as statistically significant, an ANOVA P-test had to be done. The data which we received from all the EEG scans were set and filtered through a software known as the EEGLab. In this software, there was a plugin attached (MATLAB), which allowed for data to further be processed. The data had been pre-processed, and was ran through an ICA, which is an independent component analysis. After this data, had been processed it was able to be ran through multiple statistical tests. We ran the ANOVA P Value test to be able to determine if the data we received, and to support our hypothesis for a negative correlation was statistically significant. After running the test on our data, the P Value achieved was: 0.021. This suggests that there is in fact a statistically significant difference in brain activity between all the social media users (low, medium, high social media usage). Further, the probability of obtaining the observed difference in the brain activity by just chance and not pure data is less than 2.5 percent; meaning there is strong evidence that supports the hypothesis referring to the negative correlation with brain activity and social media activity. Another step, that was done to be able to deem the project statistically significant was to find a percent change in the difference of brain activity between the individuals who had low social media usage to medium social media usage being 32.32% and the percent change in the difference of activity between individuals who had low social media usage to high social media usage being 68.93%. The last measure we took for data, was within our actual participants, we were able to find the mean time of an individual's weekly social media hours to be: 18.4 hours per week. After all data analysis, the data was converted into graphs to be able to see if there's a negative correlation.

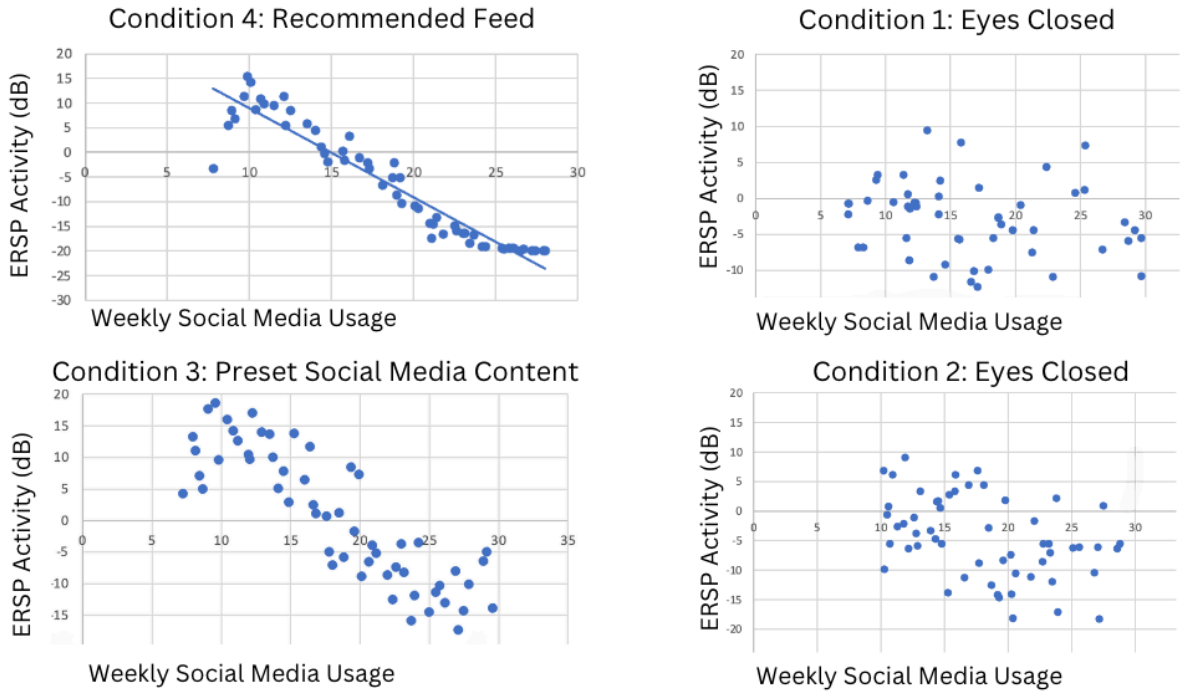


FIGURE 8: CONDITIONS AND THEIR DATA POINTS GRAPHED TO SEE IF THERE IS A NEGATIVE CORRELATION. CONDITION 1 AND 2: BASELINE. CONDITION 3: SHOWS A NEGATIVE CORRELATION, NOT STRONG. CONDITION 4: HIGH NEGATIVE CORRELATION, VERY STRONG DEPICTED BY THE LINE OF BEST FIT.

In the graphs depicted above, condition 1 and 2 the baseline testing shows random activity as the individual is in a state of homeostasis. In condition 3, a downward trend starts to appear as the social media usage increases the brain activity on average decreases for all the viewers. Lastly, the most of it all, in condition 4, where individuals viewed their own social media content (for-you page), the line of best fit indicates how there is indeed a negative correlation as most of the data points to the extensive downwards trend unlike any other of the correlation graphs for different conditions.

The future of this project could be implicated with an Electroencephalogram with more channels on the scalp. This would allow for further readings on the brain, and a more diverse range of what can be analyzed. Second, with the number of participants we had, having an increased number of participants would show more significant value and would be able to hold more credibility in determining if the study is statistically significant. Lastly, the future of this experiment includes holding a longitudinal study, this would allow for precise measurements in terms of alpha and beta waves to see the detrimental effects social media addiction has on teenagers in generation Z.

Conclusion:

In conclusion, the Electroencephalogram (EEG) based analysis of brain activity in individuals with varying levels of social media usage has revealed significant differences in brain activity patterns between high to low social media users. Through statistical analysis, high social media users displayed decreased brain activity, in the frequencies of alpha and beta waves. On the other hand, low social media users showed more activity in the brain. With the extent of the lowered brain activity, there are extreme detrimental effects in these teenagers in the long-term scale. The findings concluded from this study suggest that excessive social media use may have immense effects on cognitive and emotional functioning. The study was also able to determine that viewers who view their own social media content, their recommended feed and such have much more brain ERSP (measure of brain activity) present. Overall, the EEG-Based approach used in this study offers a valuable tool for investigating the mechanisms of social media and provides insights into the underlying mechanisms on its effect of the brain specifically, in the brain frequencies of alpha and beta.

Bibliography:

Abhang, P. A. (n.d.). *Brain waves. Brain Waves - an overview | ScienceDirect Topics. Retrieved January 25, 2023, from https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/brain-waves*

Acharya, U. R., Sudarshan, V. K., Adeli, H., Santhosh, J., Koh, J. E. W., & Adeli, A. (2015, May 14). Computer-aided diagnosis of depression using EEG signals. European Neurology. Retrieved January 25, 2023, from https://www.karger.com/Article/Fulltext/381950

Almond Solutions. (2022, November 19). *Parenting tips if your teenager is social media addict. Almond Solutions. Retrieved January 25, 2023, from https://www.almondsolutions.com/blog/parenting-tips-if-your-teenager-is-social-media-addict*

Houtan Jebelli, Sungjoo Hwang, SangHyun Lee, EEG-based workers' stress recognition at construction sites, Automation in Construction, Volume 93, 2018, Pages 315-324, SSN 0926-5805, https://doi.org/10.1016/j.autcon.2018.05.027. Retrieved Dec 1, 2022 from https://www.sciencedirect.com/science/article/pii/S092658051830013

JLG;, de A. N. F. S. R. (2016, December 8.). *Depression biomarkers using non-invasive EEG: A Review. Neuroscience and biobehavioral reviews. Retrieved January 25, 2023, from https://pubmed.ncbi.nlm.nih.gov/31400570/*

Matthew Woodward. (2022, August 31). *Social Media Addiction Statistics: Who is addicted and what are the consequences?* Matthew Woodward. Retrieved December 4, 2022, from https://www.matthewwoodward.co.uk/work/social-media-addiction-statistics/

Mayo Clinic, Mayo Foundation for Medical Education and Research. (2022, May 11). *EEG (electroencephalogram). Mayo Clinic. Retrieved January 25, 2023, from https://www.mayoclinic.org/tests-procedures/eeg/about/pac-20393875*

Robinson, L., & Smith, M. (2022, December 22). *Social Media and Mental Health. HelpGuide.org. Retrieved January 25, 2023, from https://www.helpguide.org/articles/mental-health/social-media-and-mental-health.htm*

Trujillo, L. T., Stanfield, C. T., & Vela, R. D. (2017, July 7). *The effect of electroencephalogram (EEG) reference choice on information-theoretic measures of the complexity and integration of EEG signals. Frontiers. Retrieved January 25, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2017.00425/full*

Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., Picht, B., & Nijboer, F. (2011, March 30). *A dry EEG-system for scientific research and brain–computer interfaces. Frontiers. Retrieved January 25, 2023, from https://www.frontiersin.org/articles/10.3389/fnins.2011.00053/full*

Engineering/Mathematics

Parameterizing Binary Neutron Star Postmergers using Vision Transformers

Chegu Vijay "Aashray" Aashray

Tesla STEM High School, Redmond, WA 98053, USA

Abstract

Binary Neutron Star Mergers are incredible celestial events, where two ultra-high-density objects collide, releasing massive amounts of energy through gravitational waves. These waves contain crucial information about what kind of matter exists at the neutron star's core, and the Equation of State (EOS). The Equation of State is a mathematical model that governs particle interactions and properties of neutron stars like radii and density. Gravitational waves from these mergers are picked up by Laser Interferometer Gravitational-Wave Observatories and analyzed. This study predicted the inherent equation of state and the masses of the individual neutron stars in a merger from its gravitational waves using machine learning techniques like 2D/1D-Transformers, Convolutional Networks, and Artificial Neural Networks. The results were two algorithms: a 2D-Transformer classifier that achieved 99% accuracy in classifying one EOS out of 19 in noisy spectrogram data, and a 2D-Transformer regressor that predicted the masses of the individual neutron stars and achieved a mean error of 0.009 solar masses, a tenfold decrease in error from previous models. Understanding gravitational waveforms and inferring their equations of state will help us gain deeper insights into the nature of matter and physical processes at the subatomic scale, especially whether matter at supranuclear densities will remain as hadrons or undergo a phase transition to quark-gluon plasma, a question that has defied the most powerful accelerators worldwide. Improving our knowledge of matter in extreme conditions will enhance current physical models' predictive power and accuracy, helping scientists generate more comprehensive representations of our universe.

Introduction

Binary Neutron Stars are compact stellar objects that are crucial in understanding high-energy physics, and the gravitational waves they release carry important information about the internal mechanics of neutron stars that can be parameterized and understood by machine learning algorithms. This paper seeks to develop machine learning algorithms that can be used to accurately parameterize gravitational waves from binary neutron star mergers to better understand the physics of neutron stars.

In 1915, Albert Einstein first posited that the fabric of spacetime could be made to vibrate by a powerful enough force. To try and detect these waves, the LIGO (Laser Interferometer Gravitational-wave Observatory) detectors were created. Using high-precision laser interferometry to measure the stretching in spacetime, these detectors provide important information about the objects and processes that vibrate spacetime. Current detectors are sensitive enough to identify the highly energetic collisions between neutron stars and black holes. In 2015, the first binary black hole merger (GW170817) was observed, setting off a chain of discoveries that have resulted in 90 black hole mergers and five binary neutron star mergers (Whittaker et al). However, identifying mergers from the raw LIGO strain data is a difficult task.

The current method, matched filtering, requires extensive data banks of gravitational waveforms templates and only searches in a limited part of the parameter space that gravitational waves have (Babak et al. 2012, Huerta et al. 2017). As more LIGO detectors come online, and detectors become more sensitive as technology advances, current methods are too computationally expensive to scale easily for the rapid detection, parameterization, and localization needed to confirm the short-lived merger and learn more about it with conventional telescopes. By learning more about compact object mergers, like binary neutron star (BNS) mergers and binary black hole (BBH) mergers, it is possible to learn about the equation of state of neutron stars, test general relativity, explore how matter behaves at ultra-high densities and temperatures, and find how gamma-ray bursts are formed. CBCs can also provide new insights into the Hubble constant, the maximum mass of neutron stars and the principle behind the creation and processes of intense electromagnetic outbursts (Whittaker et al. 2022).

Much work has been done on identifying gravitational waves from black holes and neutron stars in LIGO data (Krastev et al. 2021). Current approaches rely on electromagnetic observations for more detailed information about the CBCs, and due to the short time span of mergers, rapid detection is needed for following up on the mergers through conventional methods. In order to identify and parameterize compact binary coalesces (CBCs) faster, many have looked to machine learning because of its ability to offset computationally intensive calculations to the training stage and run quickly and effectively when scanning for gravitational wave signals on easier to find hardware. Machine learning approaches are preferred as they can interpolate and generalize between given training samples and don't require the extensive template banks that the current approach, Matched Filtering requires. They are also much more effective than Matched filtering as Matched filtering requires probing the 15-dimension parameter space that CBCs (and their gravitational waves) can have in order to get accurate results, a process that can take weeks (George & Huerta, 2018).

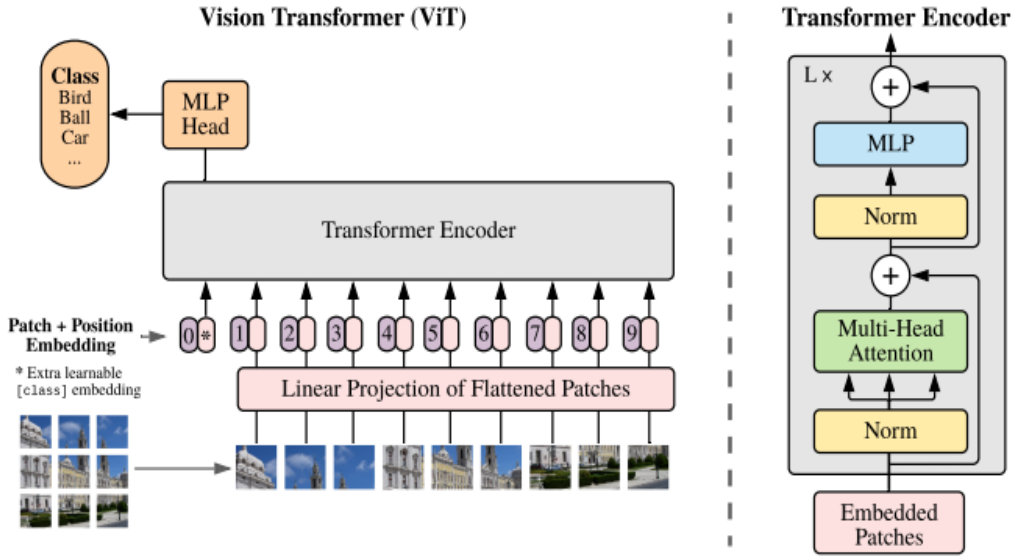
Groups have also developed methods that can parameterize black hole mergers, often finding out the masses of the two black holes and the location of the merger. However, not as much research has been done on parameterizing BNS mergers. Current machine learning approaches all use simulated data, as there isn't enough real GW data available. However, simulating BNS mergers is very complex as they must account for the magnetohydrodynamical effects of the ultra-dense matter interacting in the neutron stars and additional factors like turbulence. This isn't a problem for black holes, as they have no matter in the conventional sense. Due to this, there aren't many highly accurate simulated datasets available for BNS post-merger waveforms. Previous approaches have had to use workarounds and fiducial data in order to be able to train their machine learning models (Whittaker et al. 2022). The database used in this project, the Computational Relativity database, is a bleeding edge database with 2245 simulations created using 150 million CPU hours on supercomputers across the world (Dietrich et al.).

This study seeks to parameterize 2 important factors – the Equation of State and the component masses of the neutron stars from the simulated binary neutron star mergers. Equations of State describe the internal physics of neutron stars and relate properties of neutron stars, like mass, radius, temperature, and states of matter. Understanding equations of state and finding which human-made equations of state are closest to the inherent equations of state of real neutron stars are important steps in better understanding physics at ultra-high energies, as these conditions cannot be replicated on Earth in particle accelerators. Here, the machine learning algorithm tries to predict which theoretical equation of state was used in the simulation of a certain merger from the gravitational waveform.

Similar accuracy for predictions using contemporary deep learning methods was first demonstrated in George and Huerta's 2017 paper, where they showed that deep learning could have similar sensitivities to matched filtering. In previous studies done by Krastev et al. , their approaches in using convolutional networks to parametrize the masses of neutron stars were very successful, with an average mean squared error of .1 solar mass. In Whittaker et al., similar results were found, although the small size of the dataset led to overfitting.

The Model

For the machine learning model, many models were used as described in Table A. The best model was the Deep Vision Transformer. Vision transformers are a powerful and useful tool for image analysis due to their ability to use self-attention. Self-attention helps the machine learning model decide what areas are most important and allows it to focus on these areas. There were significant problems in using CNNs with this dataset, primarily due to the distribution of data in the spectrograms. The spectrograms have most of their information in one section, with the rest of the image being noise. This causes problems for CNNs, as it is hard for them to identify the small features needed for accurate classification. The nature of CNNs means that convolutional layers at the 'bottom', which receive the image first, pay attention to smaller scale details, while layers at the 'top', which receive the image after it has been processed by other convolutional layers, pay attention to larger scale detail that the 'lower' convolutional layers have picked out. This system works well for images with an even data distribution, but not for images with a heavily skewed data distribution, like the BNS GW Spectrograms. The training was also problematic, as CNNs with very few layers were too computationally expensive to train at higher resolutions due to the large MLP backend needed. CNNS with lots of layers washed out too much detail, getting only 70% to 80% accuracy. Transformers were useful, as they could pay global attention to the images, and focus on where the data was without extra transformations.



This is a diagram of how the Vision Transformer works, and the multi-headed attention that allows for significant increases in accuracy. Credit: *Dosovitisky Et Al.*

Section II

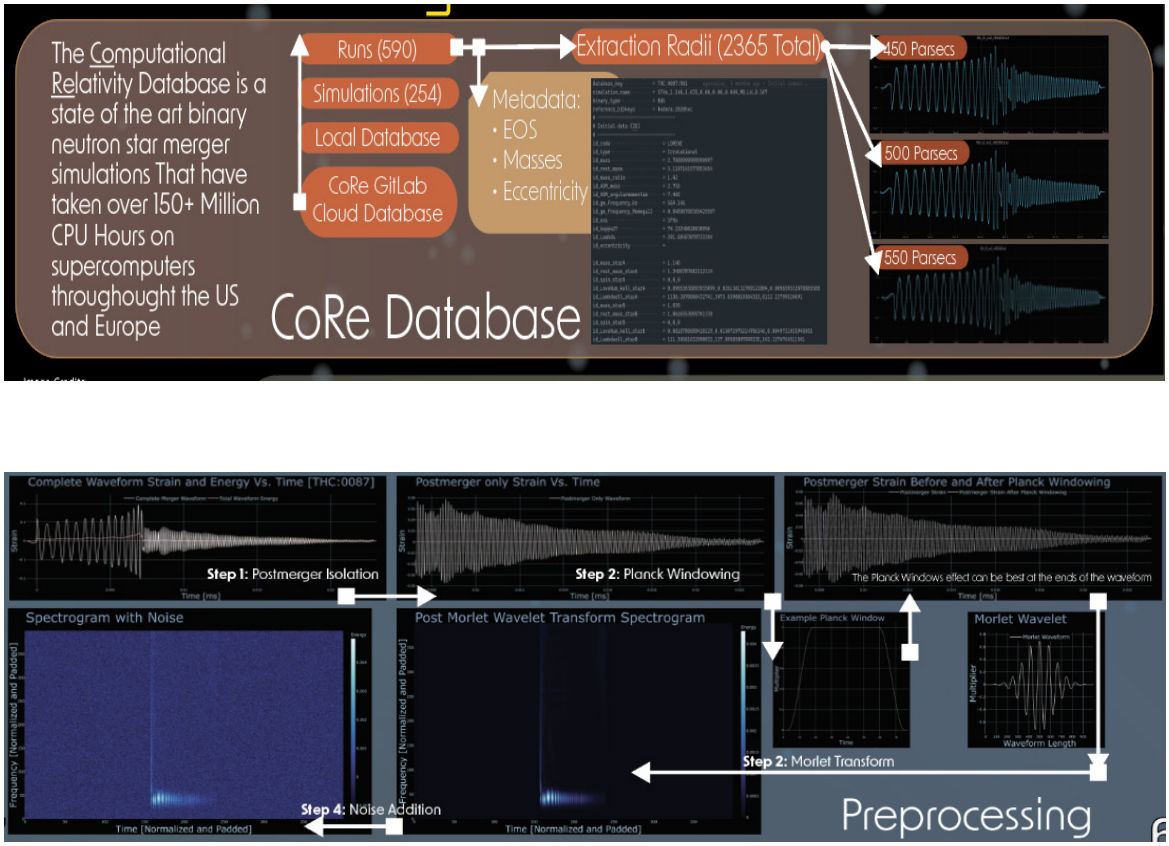
Data Collection

All of the data for this project is from the *Co*mputational *Re*lativity (CoRe) database. The CoRe database is composed of 367 simulations of binary neutron star mergers. Each simulation is composed of different runs. Per run, there are different waveforms corresponding to the different extraction radii and metadata. The gravitational wave data is formatted as Numpy arrays formatted in H5 files and is a time series of 64-bit floating-point numbers. The metadata, which is composed of different floating point and string values corresponding to data about the simulation. Here,

I will be calling the gravitational wave data as *strain*, and the different parameters associated with a certain strain simply as *parameters*.

Data Processing

Before the data could be used for training the machine learning model, it had to be processed so that it would be compatible with the model. First, all of the raw GW data was cloned to the training machine to speed up training. From here, only the post-merger waveform was extracted by taking the minimum of the energy of the envelope of the waveform. The post-merger waveform is the gravitational waveform that is released after the two neutron stars touch each other. The parameters corresponding to this strain data were then also extracted. Then, a Planck window is applied to prepare the data for the next step, the Morlet transform. The Morlet transform converts the one-dimensional time-energy data into two-dimensional time-frequency energy data. Some models were trained using the 2D data, while others were trained with the 1D raw strain data. For both types of data (1D and 2D), the data was resampled and padded to be of an appropriate length. For the 2D data, random gaussian noise was added at different signal to noise ratios from 0.001 to 5. This helped to better simulate realistic detector conditions. The data is stored in a PyTorch dataset to speed up the ease of training the machine learning algorithms. The data was then randomly split into three datasets for training, testing, and validation.



This is a diagram of the CoRe Database, and some of my preprocessing steps. The preprocessing steps are in blue, with the CoRe Database in orange.

The Model

Throughout the study, many different machine learning models were used. After extensive validation, the Deep Vision Transformer was selected as the best model due to its high accuracy, low overfitting, and ability to generalize on small datasets.

Model Name	Theory	Results
1D-ANN (Artificial Neural Network)	Artificial neural networks have been successfully used to solve many problems, and their ability to generalize and theoretically learn any problem as well as their ease of creation and use makes them very easy to create. This ease also allows them to test the training infrastructure and data preprocessing and loading.	Low AUC Low Accuracy Low Generalizability and poor training characteristics
1D-LSTM (Long Short Term Memory Networks)	LSTM neural networks use a hidden state that can store information about previous parts of the input data. This would theoretically allow the ML algorithm to understand the waveform better with less computational resources.	Little to no Generalizability with a 41% accuracy
1D-Encoder Transformers	Encoders use self-attention to better understand what sections of the input data is useful in order to gain a higher accuracy and lower computational resource use	Little to no Generalizability with a 41% accuracy
2D-ANNs	Similar to 1D-ANNs, 2D-ANNs are simple to use and proven, but on the larger input sizes that I used, they were too computationally expensive to run.	Poor accuracy
2D-CNNs	Convolutional Neural Networks are state of the art image classification algorithms, which use convolutions to extract important features from input images.	Low Accuracy High AUC Poor training Characteristics
2D-Simple Vision Transformer	Vision transformers use self-attention to better understand what parts of the image are important to better extract important information	High AUC High Accuracy Very easy to overfit
2D-Small Dataset Vision Transformer	Small Dataset Vision Transformers using patch shifting and changes to the attention mechanism to better encode images	High AUC High Accuracy
2D-Deep Vision Transformer	Deep Vision Transformers use the Re-Attention mechanism to make sure that each Transformer layer pays attention to unique features, preventing saturation, and allowing for accuracy to scale better with the number of layers	High AUC High Accuracy Resistant to overfitting

This table describes the different models, the theory behind their operation, and what promise they held for further development.

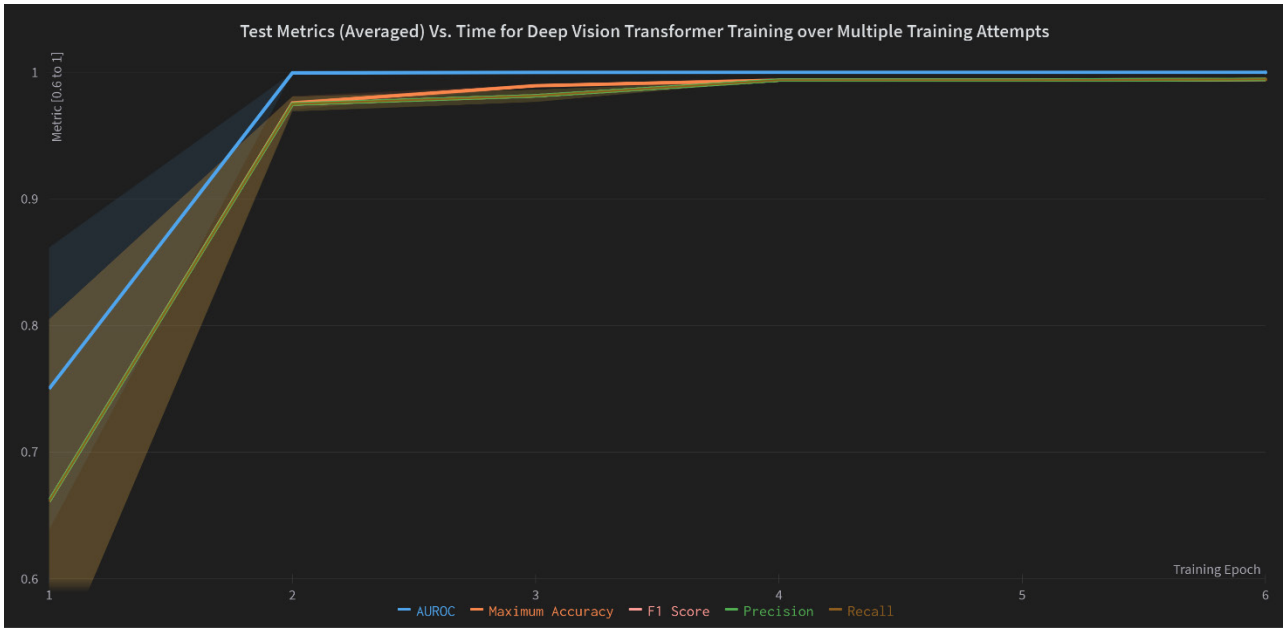
Training

Each model was trained on the training dataset for 500 epochs, and after each epoch, statistics such as the models' accuracy, specificity, recall, area under receiver operator curve, loss and F1 score were logged using Weights and Biases. Hyperparameter sweeps were also conducted to optimize the hyperparameters used in the training of the algorithm like layer sizes, block numbers and the optimizer parameters.

Testing

The final model was tested 10 times, and the average Accuracy and Area Under Curve were collected. These were then averaged between each of the runs to provide a more accurate demonstration of the capabilities of the neural network.

1)



This table tracks test metrics for the EOS classifier model over the training to track how the model is evolving. The limit of 6 epochs is due to training time requirements.

2)

Parameter	Accuracy (MAE for all but EOS)
Mass of Neutron Star A	0.001
Mass of Neutron Star B	0.001

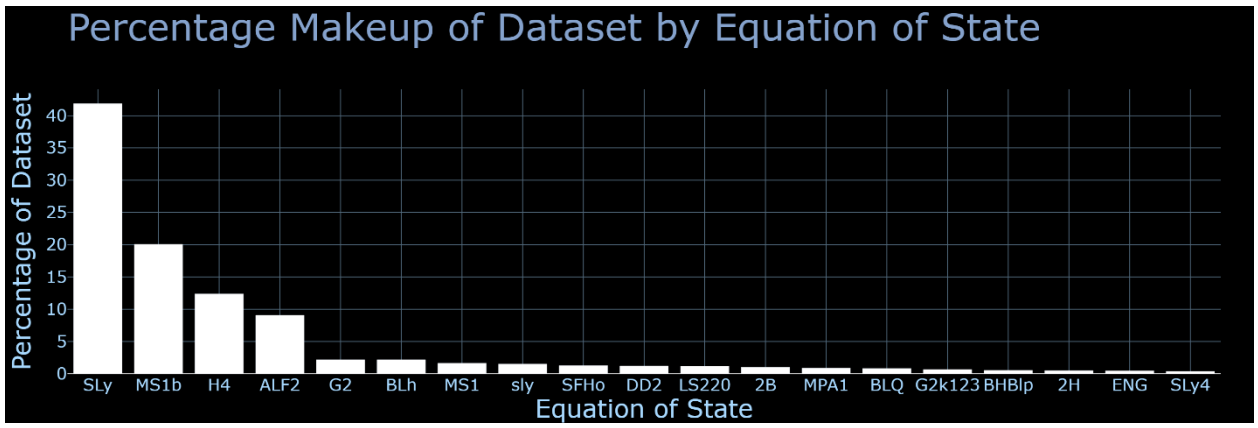
This table tracks mean error for the mass predictor model.

3)

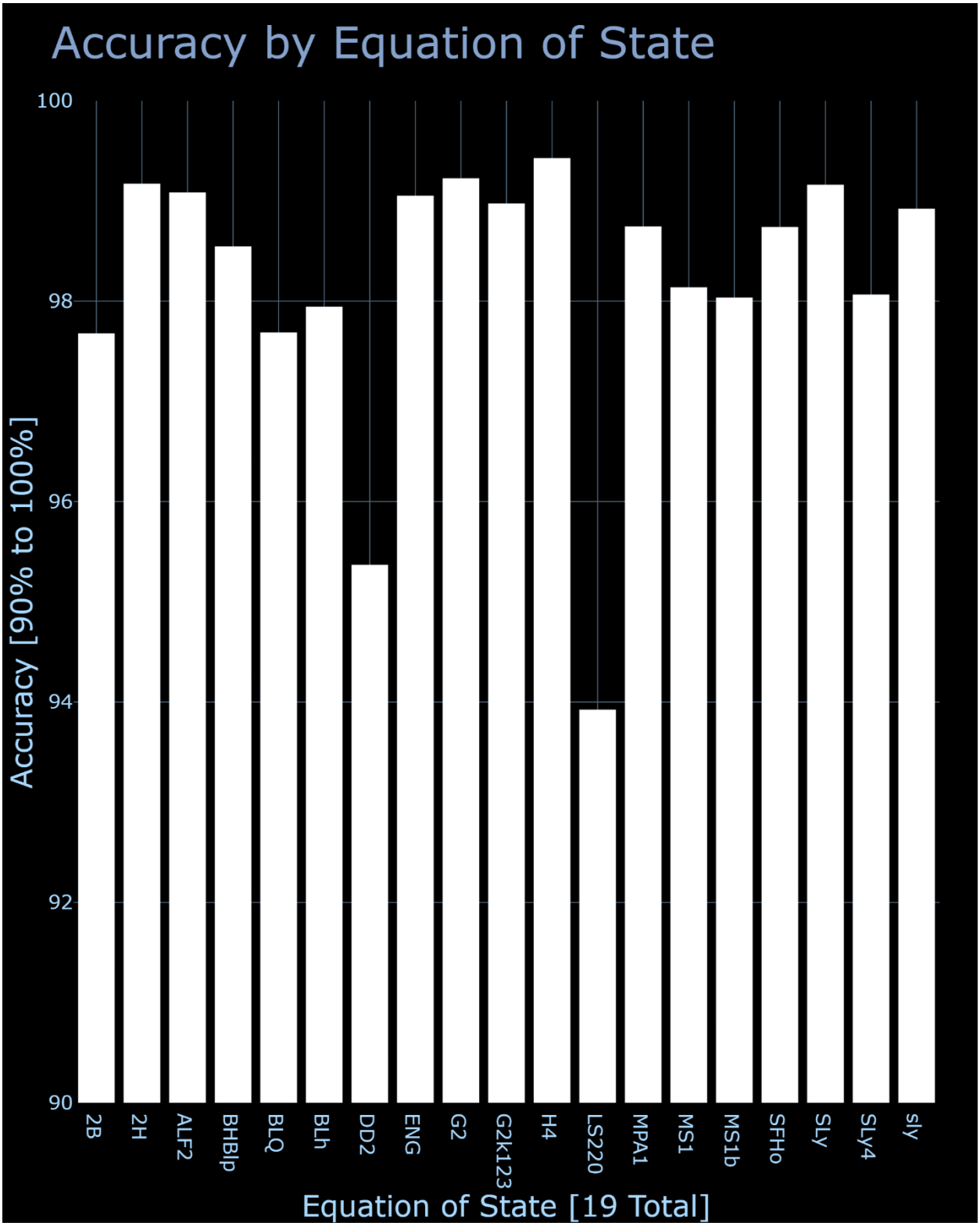
Average Accuracy, Loss, Specificity between my models:

Machine Learning Model	Accuracy	Area Under Curve
CNN	89%	.95
Vision Transformers	99.9%	.9999
DNN	70%	Not Tracked
LSTMs	41%	Not Tracked
1D Transformers	41%	Not Tracked

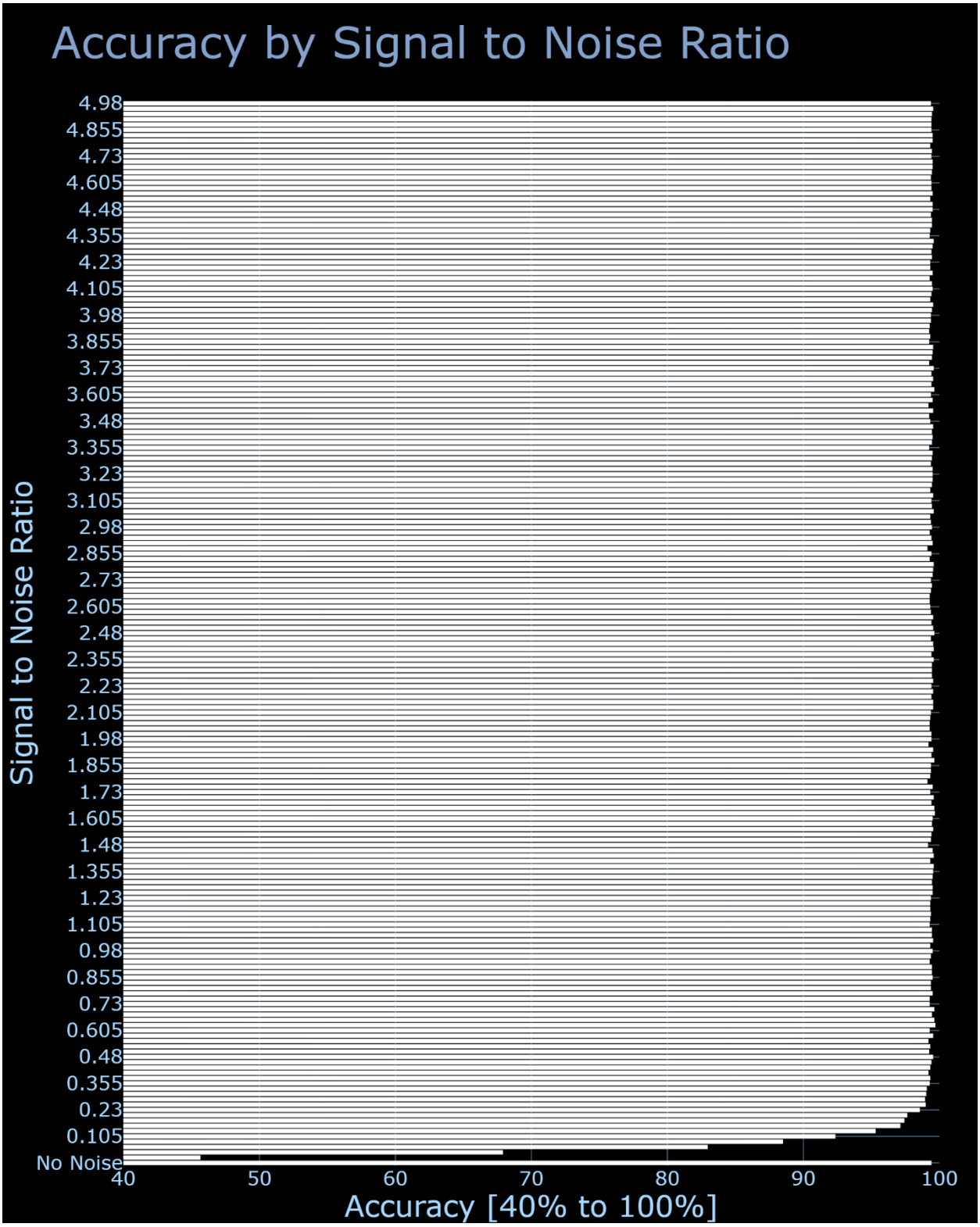
This table tracks how accurate different versions of my model are, and what parameters they are best at.



This is a breakdown on what percentage of the dataset each equation of state makes up. Nearly 81% of the data is contained in the SLy, MS1b, H4, and ALF2, leading to skewed results, with some algorithms entirely giving up, and only predicting SLy, which has 41% of the data, like 1D Transformers and LSTMs.



This is a plot of the accuracy of the model relative to different equations of state, showing how the accuracy changes with the different equations.



This plot shows how the accuracy changes with different equations of state, with the machine learning algorithm having a 90%+ accuracy at signal to noise ratios greater than 0.13.

4)

Comparison of different models:

Machine Learning Model	Equation of State (Accuracy)	Average Component Neutron Star Mass Error (MAE)
Gravitas Prime	99%	0.009
Whittaker et al.	Not Tested	0.15
Krastev et Al	Not Tested	0.1

This table tracks the accuracy of my model in predicting parameters relative to other models by other authors.

Conclusion:

Understanding gravitational waves from binary neutron stars is key to increasing our knowledge of the physics of neutron stars. This paper introduced a novel method of using Vision Transformers and Morlet Transformed gravitational waveforms to create machine-learning models capable of discerning different equations of state from simulated binary neutron star mergers with high accuracy and specificity. With a 99% accuracy and a .99 area under the curve, these machine learning algorithms show great promise, and with further improvements in model parameters, model types, and optimizations, it is possible to get even higher model accuracies. Many of the current limitations of the model come from a lack of training data. With more training data available, it is possible to make more accurate, and more robust models. Improvements also need to be made to ensure a more even distribution for different equations of state. Another way to remedy this input data problem is to use Generative Adversarial models to try and create synthetic spectrograms to better train models. Even without simple improvements, like patch autoencoders, this model is very promising and can be applied to real LIGO data.

References

Aveiro, J., Freitas, F. F., Ferreira, M., Onofre, A., Providência, C., Gonçalves, G., & Font, J. A. (2022). Identification of binary neutron star mergers in gravitational-wave data using object-detection machine learning models. Physical Review D, 106(8). https://doi.org/10.1103/PhysRevD.106.084059

Corizzo, R., Ceci, M., Zdravevski, E., & Japkowicz, N. (2020). Scalable auto-encoders for gravitational waves detection from time series data. Expert Systems With Applications, 151, 113378. https://doi.org/10.1016/j.eswa.2020.113378

Dietrich, T., Radice, D., Bernuzzi, S., Zappa, F., Perego, A., Brueugmann, B., Chaurasia, S. V., Dudi, R., Tichy, W., & Ujevic, M. (n.d.). CoRe database of binary neutron star merger waveforms and its application in waveform development. Arxiv. https://doi.org/10.48550/arXiv.1806.01625

George, D., & Huerta, E. (2018). Deep neural networks to enable real-time multimessenger astrophysics. Physical Review D, 97(4). https://doi.org/10.1103/PhysRevD.97.044039

Kim, K., Li, T. G., Lo, R. K., Sachdev, S., & Yuen, R. S. (2020). Ranking candidate signals with machine learning in low-latency searches for gravitational waves from compact binary mergers. Physical Review D, 101(8). https://doi.org/10.1103/PhysRevD.101.083006

Krastev, P. G. (2020). Real-time detection of gravitational waves from binary neutron stars using artificial neural networks. Physics Letters B, 803, 135330. https://doi.org/10.1016/j.physletb.2020.135330

Sasaoka, S., Hou, Y., Somiya, K., & Takahashi, H. (2022). Localization of gravitational waves using machine learning. Physical Review D, 105(10). https://doi.org/10.1103/PhysRevD.105.103030

Schäfer, M. B., Ohme, F., & Nitz, A. H. (2020). Detection of gravitational-wave signals from binary neutron star mergers using machine learning. Physical Review D, 102(6). https://doi.org/10.1103/PhysRevD.102.063015

Whittaker, T., East, W. E., Green, S. R., Lehner, L., & Yang, H. (2022). Using machine learning to parametrize postmerger signals from binary neutron stars. Physical Review D, 105(12). https://doi.org/10.1103/PhysRevD.105.124021

A Low-Cost Approach to the Filtration of Particulate Matter and Carbon Monoxide Produced from the Partial Combustion of Biomass

Sophia Rodriguez

Tesla STEM High School

ABSTRACT

One third of the world cooks on open fires and simple stoves that only partially combust materials, making exposure to deadly pollutants a reality for 3 billion people worldwide. This population is mainly concentrated in developing countries, because they lack access to proper cooking devices/fuels. This project aims to reduce indoor air pollution emitted by these fires and stoves by creating a low-cost air filtration device to filter particulate matter and carbon monoxide. This device will have a cost value 75% lower than the standard High Efficiency Particulate Air (HEPA) filter, an efficiency rate of 60%, and will reduce particulate matter and carbon monoxide levels by 60%. The filter utilizes inexpensive naturally occurring materials like coconut fiber, coconut activated carbon, and mosquito netting to increase accessibility. The device was tested using a 3 stone fire (a makeshift woodstove seen in many developing countries) with and without the filter (independent variable), to test for carbon monoxide in parts per million and particulate matter in micrograms per meter cubed (dependent variable). The total cost of the filter is 82% cheaper than the standard HEPA device, with an efficiency rating of 62.18%, and a p-value of 0.0085, indicating the results are statistically significant. It also reduced carbon monoxide levels by 76% relative to the control with a p-value of 0.03, and particulate matter emissions by 60.9%. This project achieved all 4 engineering goals, making it a possible solution to reduce indoor air pollution, and improve the quality of life for so many people.

INTRODUCTION

This project aims to reduce levels of indoor particulate matter and carbon monoxide pollution produced from the partial combustion of biomass from cookstoves through a low-cost air filtration system. One third of the world's population is exposed to toxic air pollutants every day just from cooking in their own home. Most families in this group live in developing countries and lack access to clean-burning fuels and stoves. This leaves them to make do with what they have access to, and for most, this means cooking on open fires or very basic stoves with biomass fuels such as dung or wood. National Geographic stated that these fires produce about, "400 cigarettes' worth of smoke an hour" (M. Nijhuis). These basic fires and stoves only partially combust materials, resulting in the release of pollutants such as carbon monoxide and particulate matter, which both have very harmful effects on the human body. Particulate matter can get stuck in peoples' lungs and blood stream, and has been linked to things like heart attacks, asthma, and chronic obstructive pulmonary disease (COPD). Carbon monoxide is a colorless and odorless gas which can inhibit the blood's ability to carry oxygen to organs and tissues and can cause permanent heart and brain damage as well as issues with fetal death and miscarriages. The dangers of these pollutants are why the Environmental Protection Agency (EPA) publishes recommended exposure levels for harmful pollutants that

households should not exceed. However, exposure studies in developing countries have shown that, "particulate matter levels in these homes are almost 2000 times more than the levels recommended by the EPA" (N. Bruce). Indoor air pollution is a major public health threat around the world, and a growing problem that needs to be solved. The World Health Organization stated that indoor air pollution was responsible for 3.2 million deaths in 2020, including over 237,000 deaths of children under the age of 5. These pollutants are disproportionately affecting women and young children, because they are typically the people who are present while cooking and the burning of fuel takes place. The public health issue associated with indoor air pollution is mainly one of economic disparity. We are not facing these issues in wealthy communities in developed countries because our society has moved passed the use of biomass fuels. However, in these developing areas, these are not available to most people either because of economic reasons or accessibility. The United Nations published a report in 2022 saying that over 1 billion people, about 12 percent of the population, is living on less than one U.S. dollar per day. Majority of this population is concentrated to Sub-Saharan Africa where the poverty rates are among the highest in the world, and these countries are also coincidently the ones that have the highest deaths attributed to indoor air pollution. This is not an issue of obtaining new technologies, but rather making the technology we already have available to those who need it the most. Both pollution levels and human health could be greatly improved if these families had access to proper ventilation and filtration systems. In Guatemala, a randomization study was done where the researchers implemented the use of clean-burning stoves, and carbon monoxide levels reduced by over 40 percent in the first year, and participants reported a decrease in phlegm, a symptom associated with respiratory infections such as bronchitis (T. Smith-Sivertsen). While these clean-burning stoves successfully reduced pollution levels in Guatemala, this is not an attainable item for most people living in rural areas. The price of clean stoves can range anywhere from hundreds to thousands of dollars, but in some of these lower-income countries, people are living on a few dollars per day and can barely afford food for their families, let alone a stove. Thus, we look towards cheaper and more accessible alternatives for reducing pollution levels. Another study was done in California during the COVID-19 pandemic with the goal of making a DIY air filtration device as efficient as a HEPA (High Efficiency Particulate) filter for classrooms that cannot afford proper ventilation systems. The idea of air changes was a big deal during the pandemic, because COVID is an airborne illness, and this level of airflow is typically achieved by the use of HEPA filters. These filters are specially designed to remove 99.99% of airborne particles, but their high efficiency comes with a high cost as well. However, the California researchers achieved their goal, and they managed to create a filtration device with 85% efficiency for one fifth of the price of a standard HEPA filter (D. Srikrishna). This was a major achievement, and very applicable to these classrooms during the pandemic. Though, more applicable to indoor air pollution is a study done in rural Montana which tested the use of two filter devices in homes that use wood stoves. The first was a high-efficiency air filter, and the second was a placebo filter with a low-efficiency rating. The researchers were testing levels of particulate matter and endotoxin. The experimental high-efficiency group showed a 66% reduction in particulate matter concentrations relative to the placebo. Furthermore, both the experimental and the lower-cost placebo group showed a reduction in endotoxin and larger particulates (M. McNamara). The results of the lower-cost filter being active in reducing pollutants was very significant in the movement towards accessible alternatives to air filtration devices, but there is still yet to be a low-cost solution to the pollution of both particulate matter and carbon monoxide produced by biomass stoves. This project will seek to solve this problem by designing an air filter made from low-cost materials that are accessible to rural communities. This filter will utilize activated carbon technology and fiber filters for the filtration of particulate matter, as well as a fan for ventilation and air circulation. This project aims to see how designing a low-cost affordable air filtration device made with activated carbon and natural fiber has a role in reducing levels of particulate matter and carbon monoxide pollution caused by the incomplete combustion of biomass. Compared to the conditions with no filter installed, the air filtration device will reduce average carbon monoxide gas ppm by 60 percent, average particulate matter μ/m^3 by 60 percent

and will have a filtration efficiency rate of over 60 percent and have an overall cost that is at least 75 percent lower than the standard HEPA air purifier.

METHODOLOGY

In this project, the activated carbon material will be bought for safety reasons, but this can also be made from natural materials that are rich in carbon like coconut or bamboo. This is done through a process called carbonization which involves the cooking down of the material to its pure carbon form. The porosity of the pure activated carbon allows for the adsorption of air pollutants like carbon monoxide, and that is why it is used as a filtration material in this filter. This activated carbon material is held in a honeycomb structure which is designed using the Fusion360 online platform and saved as a Makerbot file type onto a USB drive. Start by 3D printing this design by loading the filament into the 3D printer and uploading the Makerbot file into the printer and hitting print. Wait for the structure to print – this process will take multiple hours – and then carefully remove the structure from the printer. Note: all 3D printing will be done at Tesla STEM High School in a classroom with proper ventilation. See figure 1 for example of 3D printed design. This structure is where the activated carbon material will be placed.

After the honeycomb structure is printed out, cut two layers of mosquito netting that are 12” by 12” squares. Place a layer of glue onto the honeycomb structure, and place one of the cut-out squares on top of this so that the mosquito netting glues to the structure. Once the glue has dried, flip the structure over so that the open side is facing up, and place equal parts of activated carbon material into each honeycomb tube of the structure until each tube is full. This honeycomb design allows for high efficiency filtration because there is a high carbon-to-structure ratio. This also creates a very durable and long-lasting structure that is user-friendly and very easy to open and replace the carbon pellets if needed. When all the parts are filled with carbon, place another layer of glue on the structure and place the second cutout square on top so that the carbon is stuck in place by the two netting layers. See Figure 2 for more detail on filling the carbon. Make sure that this ensures stability and keeps the carbon inside before proceeding. This will filter carbon monoxide and thin particulates.

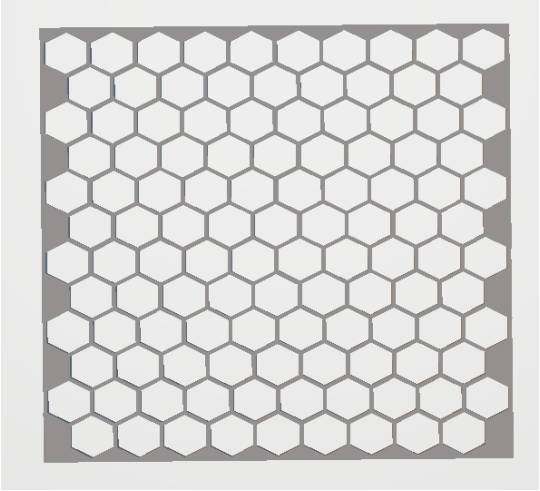


Figure 1. 3D Design of Honeycomb structure to hold activated carbon (12” x 12” x 0.75”)

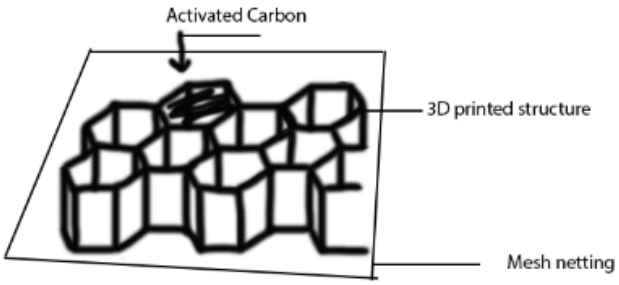


Figure 2. Model of 3D printed honeycomb design for activated carbon filter including diagram of activated carbon and mesh netting structure.

For the filtration of the larger particulates, a coconut fiber filter will be used. To make this, start off with two 12” by 12” squares of mosquito netting. Connect these two by sewing them together in a quilted-like structure shown in Figure 3. This can be done with a sewing machine, or by hand-sewing with a needle and thread. Connect the squares in a 4 by 4 grid, making each square 3” by 3”. Sew each square one row at a time so that the square structure is sewn in but leave the top portion of each square open in order to place the coconut fiber inside. Throughout the process, after sewing the sides of each square in each row, place coconut fiber inside each square so that the entire square is full. Then,

in a straight line, sew the top of each row, and do the next row until finished with all 4. See figure 3 for more detail.

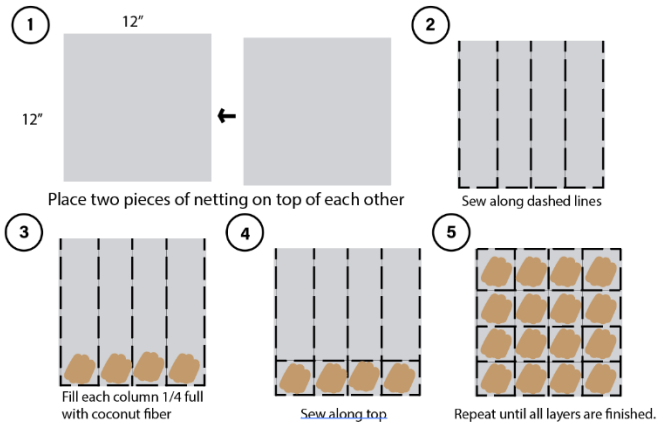


Figure 3. Step by step diagram of how to put together the particulate matter coconut fiber filter.

For the process of putting together the entire device, the quilted particulate matter structure will be placed first on the very bottom, then the activated carbon honeycomb structure behind that, and then an exhaust fan and a housing behind that (see Figure 4). First, place the particulate matter quilt flat on a surface. Behind that, place the honeycomb structure directly on top of the quilt. Cut out four 12" by 1.5" cardboard strips, and tape or glue them together to make a cube-outline structure. Place the stack of both filters into the cardboard open cube so that the cardboard creates a stabilizing ring around the filters. Attach the filters to the cardboard gently with duct tape. Place the exhaust fan behind the stack of the filters directly on top. Duct tape this to the cardboard edges that hold the other two filters. Finally, cut out four 12" by 6" squares of cardboard or foam core. Place these pieces on each side of the filter above the fan and attach them with duct tape. This will create a housing that can direct the air flow out of whatever area cooking is taking place in.

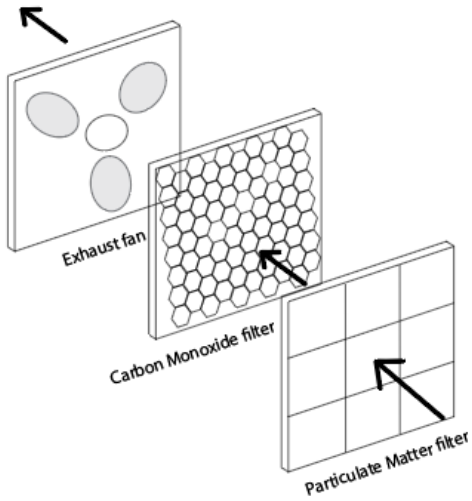


Figure 4. Visual display of the order of filters as well as the air flow through the entire filtration device.

For testing, the independent variable for this experiment will be whether or not the filter is present. There will be two groups: one experimental group with the filtration device present, and one control group without the filter present. The dependent variable will be the amounts of carbon monoxide and particulate matter measured in parts per million (ppm) and micrograms per meter cubed (μ/m^3).

First, safety precautions must be taken by wearing a KN95 mask, protective eyewear, and a burn-safe apron. Because testing cannot be done in actual houses, it will take place at the Burn Design Lab in Vashon, Washington. To test the filter, two types of stoves will be used. The first is the Natural Draft Rocket Stove which is a clean cook stove designed at the BURN Design Lab (Figure 5). The next is a 3 stone fire which is a makeshift fire made with 3 bricks and wood (Figure 5) This is a commonly seen makeshift fire throughout developing countries. There will be one control trial on each stove and 3 experimental trials for 30 minutes each. For the data collection, the BURN Design Lab standard data collection methods will be used along with handheld readings for carbon monoxide. For particulate matter, particulate paper will be used (see Figure 6). This will be weighed before and after every trial and placed in a tube that is held in the pipe connected to the fume hood

during testing. For the carbon monoxide readings, as the air goes through the fume hood, there is a carbon monoxide sensor in the pipe that is detecting carbon monoxide levels every second and recording them into an excel spreadsheet. This will be used to measure the overall carbon monoxide levels of each trial, but a handheld carbon monoxide monitor will be used to take measurements upstream and downstream of the filter (Figure 7) in order to calculate efficiency.



Natural Draft Rocket Stove



3-Stone Fire

Figure 5. Images of the improved cookstove designed at BURN Design Lab and 3-stone fire made from stones and wood used for the testing of the filter. The 3-stone fire is more commonly seen amongst developing countries, and has much more emissions than the improved stove.



Particulate paper before trial



Particulate paper after trial



Contraption that holds filter paper (circled)

Figure 6. Image of filter paper used to measure particulates, paper was weighed before trial and placed in the contraption (right), and then weighed after trial (lower left). The mass of the paper after trial minus the mass of the initial paper showed the mass of the particulates captured over each trial.

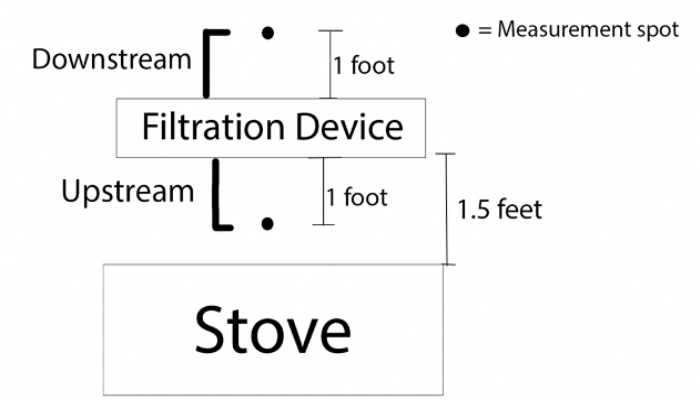


Figure 7. Model of upstream and downstream measurements for testing.

To begin the set up for each trial, place the stove in the center of the fume hood area and place the wood inside. Place a large pot large enough to fit over the stove on top of the stove and fill it with 5 liters of water. Take the temperature of this water by sticking the thermometer inside and leave it inside to keep a running temperature throughout the trial. If running an experimental trial, place the filter above the stove on the two pieces of wood suspended above the filter (see figure 8). Weigh the filter paper and take note of the weight in Table 1. Place the filter paper in the holder and place this in its spot in the pipe. Turn on the fume hood, start online data log using computer, light the wood, turn on the filter if running an experimental trial, and start a timer for 30 minutes. At 10 minutes, 20 minutes, and 30 minutes in, take 3 measurements of carbon monoxide levels above and below the filter (for experimental trial only) and record these in data Tables 2 and 3. Once 30 minutes are over, turn off the fume hood, save online data file, and put out the fire by using tongs to grab each wood stick and place it in the ash pile below the fume hood. Take out the particulate paper and weigh it again, and record this in Table 1. Repeat the entire process with the rest of the trials making sure to record data in the data tables. The online data will log everything on its own into online data tables in Excel – it will be taking carbon monoxide and carbon dioxide measurements in parts per million, and temperature levels of the water in the pot in degrees Celsius throughout the entire trial.

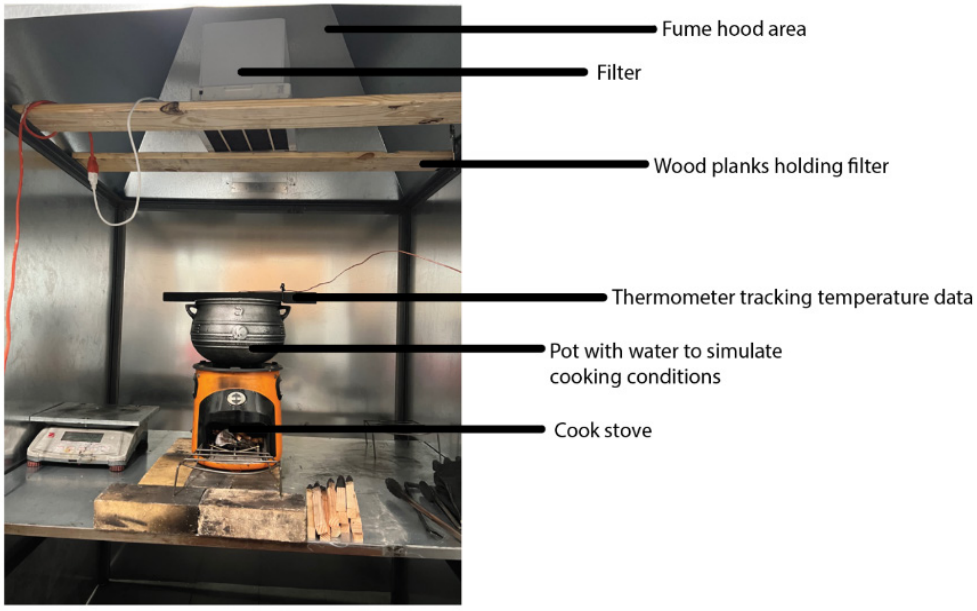


Figure 8. Diagram of set up for testing under fume hood. Includes filter, stove, pot, and thermometer (does not include pipe system after fume hood involved in data collection).

DATA TABLES

Note: Throughout the testing process, there were complications with lighting the cookstove, so instead, 3 trials were done with the 3 stone fire (trials 1-3), and one control was done with the 3 stone fire (trial 4)

Table 1. Results for particulate matter paper weights before trial, after trial, the difference, and the calculated result of micrograms per meter cubed for each trial.

Trial number	Weight of PM paper before (mg)	Weight of PM paper after (mg)	Difference (after-before) (mcg)	Mcg/m^3
1	653.938	655.528	1.590	5.68
2	653.788	655.021	1.233	4.41
3	655.491	657.177	1.686	6.02
4 (control)	654.323	658.173	3.850	13.75

Table 2.1. Carbon monoxide upstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 1

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	189	170	181	180
20	219	172	185	192
30	189	179	193	189

Table 2.2. Carbon monoxide upstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 2

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	151	175	154	160
20	196	181	169	182
30	141	159	168	156

Table 2.3. Carbon monoxide upstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 3

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	198	176	190	188
20	159	155	178	164
30	178	168	185	179

Table 3.1 Carbon monoxide downstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 1

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	48	57	54	53
20	51	56	61	56
30	72	69	72	71

Table 3.2 Carbon monoxide downstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 2

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	67	54	53	58
20	72	76	59	69
30	67	56	54	59

Table 3.3 Carbon monoxide downstream readings at 10 minutes, 20 minutes, and 30 minutes for trial 3

Time Elapsed in Trial (min)	CO Reading 1 (ppm)	CO Reading 2 (ppm)	CO Reading 3 (ppm)	Average
10	87	70	77	78
20	63	64	59	62
30	74	65	62	69

RESULTS/CONCLUSIONS

After the collection of all the data at the BURN Design Lab, data analysis was done. Throughout the testing stage, there were some complications with keeping the Rocket stove lit, and so some modifications were made to the data collection procedure, and 4 trials were done with just the 3 stone fire. Thus, for the data analysis, the 3 stone fire data was primarily used, but this data is more accurately representative of the worst-case-scenario of a cooking situation, whether as the rocket stove is an improved cookstove that shows less emissions. This means the 3 stone fire is more accurate to what is available to the larger population, and a better test to the filter's ability to filter out large levels of pollution. First, a cost analysis was done of the entire filter by adding up all the costs of the materials involved in the building of the filter. This came out to be a total of $26.49, which is 82% cheaper than the standard High Efficiency Particulate Air (HEPA) filter which averages around $150. This meets the engineering goal of being at least 75% cheaper than a HEPA filter. Next, the efficiency of the filter was calculated by comparing the upstream and downstream carbon monoxide measurements of each of the 3 experimental trials. This was calculated as a percentage, and then the mean of the three percentages was taken for an average efficiency rate of 62.18%. This means that the filter is 62.18% efficient in filtering out pollutants. This measurement was tested for significance with a two-sample t-test which compares the means of two data sets to see if the difference between the two is significant. The mean of the upstream data points and downstream data points were compared using this test, and they had a p-value of 0.0085, meaning the efficiency rate of 62.18% is statistically significant with a significance level of 0.05. This means the efficiency rate is generalizable to a larger population, and thus if the filter was to be replicated and used in the setting of a home, the filter would have a similar efficiency rate, and would most likely be effective at filtering pollution. This also means the engineering goal of having an efficiency rate of at least 60% was successfully achieved.

Next, the analysis of the reduction in particulates and carbon monoxide was calculated by taking the mean percent difference of carbon monoxide parts per million and particulate matter micrograms per meter cubed in the control vs experimental group. This data was collected by the weight of the particulates and the computerized carbon monoxide data. The particulate matter weight was measured in grams, and this was converted to micrograms per meter cubed with the help of the engineers at the BURN Design Lab. The average was found for the carbon monoxide and particulate levels amongst all three trials, and this was compared to the average in the control trial using a percentage difference. This showed a 76% decrease in carbon monoxide, significant with a p-value of 0.03 from the two-sample t-test. This means the reduction in carbon monoxide is also generalizable, and so is the reduction in particulate matter which was 60.9% and had a p-value of 0.042. This means the final two engineering goals of having a 60% reduction in both particulate matter and carbon monoxide were both achieved as well.

While all engineering goals were achieved, there were still some limitations to the experiment. The first of which was the fume hood that the filter was placed under in the testing area which was used for safety reasons. This fume hood had a fan that was 16 square feet with a cubic feet per minute (cfm) rating of 471, which was much larger and much more powerful than my filter fan which was only 1 square foot with a rating of only 176 cfm. This means that there is a high chance that a lot of polluted air was bypassing the filter and still going through the fume hood system to be measured for data. There was also not a consistent burn rate throughout the trials which was shown through the carbon dioxide data measured on the computer system. This means that there were some trials in which more pollutants were released than others. Both these things could have caused some inconsistencies throughout the data. In the future, I would like to continue this project by improving my design so it has even higher pollution reduction and efficiency rates. This can be done by implementing a more powerful fan, and different materials that allow for a higher rate of filtration. I would also like to alter my design so the overall cost is less than $10, because that will make the filter even more accessible to the intended audience. I would also like to do more testing with different

kinds of fuel sources like dung or coal, because this experiment only utilized wood, but many families cooking on open fires are using other types of fuel that can emit different levels of pollution. There is no way of telling what kind of health benefits could come from this filter being used in homes from this data alone, because the households that need a filter like this are typically facing so many other issues like poverty, malnutrition, lack of access to clean water or healthcare, and so many other things. Thus, I would like to implement this filter into peoples' homes and do a longitudinal study to watch how it can improve the health of the families involved. There are so many people who are slowly dying from something they do not even know is harming them. With the implementation of this filter into homes, and the increase in awareness and education surrounding indoor air pollution, there is a chance to change the lives of billions of people for the better.

BIBLIOGRAPHY

3D printer safety. 3D Printer Safety | UVM Risk Management and Safety | The University of Vermont. (n.d.). Retrieved December 14, 2022, from https://www.uvm.edu/riskmanagement/3d-printer-safety

Bruce, N., Perez-Padilla, R., & Albalak, R. (2000). Indoor air pollution in developing countries: a major environmental and public health challenge. Bulletin of the World Health Organization, 78(9), 1078–1092.

Environmental Protection Agency. (n.d.). EPA. Retrieved November 2, 2022, from https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm

McNamara, M. L., Thornburg, J., Semmens, E. O., Ward, T. J., & Noonan, C. W. (2017). Reducing indoor air pollutants with air filtration units in wood stove homes. The Science of the total environment, 592, 488–494. https://doi.org/10.1016/j.scitotenv.2017.03.111

Nijhuis, M. (2021, May 4). Three billion people cook over open fires — with deadly consequences. Photography. Retrieved December 11, 2022, from https://www.nationalgeographic.com/photography/article/guatemala-cook-stoves

Servi-Tek. (2021, March 26). Air Filtration and covid-19: Indoor Air Quality expert explains. Servi. Retrieved December 11, 2022, from https://servi-tek.net/hvac-filtration/

Srikrishna D. (2022). Can 10× cheaper, lower-efficiency particulate air filters and box fans complement High-Efficiency Particulate Air (HEPA) purifiers to help control the COVID-19 pandemic?. The Science of the total environment, 838(Pt 1), 155884. https://doi.org/10.1016/j.scitotenv.2022.155884

Smith-Sivertsen T, Díaz E, Pope D, Lie RT, Díaz A, McCracken J, Bakke P, Arana B, Smith KR, Bruce N. Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: the RESPIRE Randomized Trial, Guatemala. Am J Epidemiol. 2009 Jul 15;170(2):211-20. doi: 10.1093/aje/kwp100. Epub 2009 May 14. PMID: 19443665; PMCID: PMC8889940

Tyrell, C. (2020, May 27). Easy guide to make activated carbon at home. Survivopedia. Retrieved December 14, 2022, from https://www.survivopedia.com/how-to-make-activated-carbon/

Engineering a Microplastic Filter for Indoor Taps

Shruthika Saravanakumar

Nikola Tesla Stem High School

Abstract

The goal of this project is to address the various problems of microplastics being ingested through drinking water on a daily basis, by designing and testing a microplastic filter which attaches onto indoor taps. The prototype filter attachment is made to remove microplastics from tap water, making it easier and faster for customers to use. The methodology includes an engineering procedure and experimental testing procedure. The engineering process includes 3D rendering and printing of the filter attachment, while the experimental procedure assesses the filter attachment's and the membrane filter's efficiency through water leakage tests and membrane filtration. The filter's ability to remove microplastics and low leakage are demonstrated by data collection and analysis. The results of the project show a mean efficiency of 97.403% when it comes to filtering microplastics. At the end of the project all the engineering goals were accomplished, and this project was determined a success.

Introduction

This project seeks to find a solution for the problem of humans ingesting microplastics unknowingly through everyday food consumption by designing a microplastic filter that can be attached on indoor taps. The average person ingests around 1769 microplastics every week. Tap water contains most of the microplastics that are ingested by us, humans. "94% of the water samples from the United States contained microplastics," (Kosuth et al., 2017), there were 159 water samples collected from five regions of the world which tested showing that 83% of the water contained microplastics. They were mostly microplastics that were between 0.1mm to 5mm in size.

Ingested microplastics enter human blood and cause many health problems. A recently done study on the health effects of microplastics discovered that microplastics can lead to cell damage and allergic reactions (Carrington, 2020). Since the consumption or inhalation of microplastics cannot be controlled in many situations it could end up being fatal for some if their body reacts in a dramatic way.

Microplastics have been found in humans for a long time now, but recently in a study done in 2020, microplastics were discovered in unborn babies, meaning these kids have ingested microplastics before being born (Carrington, 2020). The babies were said to have gotten the microplastics through the mother, since there was microplastics discovered in the mother's placenta, where the fetus is essentially growing. After scanning 12 different microplastics were discovered, scattered throughout the mother's placenta (Ragusa et al., 2021).

A potential solution for this problem is a filtration device that can be attached to any household tap, to filter out the microplastics from the tap water. Recently, household membrane filters, that filter out microplastics out of drinking water, have been created by the company, Life Straw. They have made a pitcher which uses a mix of modern technology and membrane filter

to filter the tap water that is poured into the water pitcher (Lundy, 2022). However, this method of filtration can only be done in a specific pitcher and with stagnant water which will need to be refilled continuously. The purpose of this investigation is to design and engineer a prototype that filters out microplastics out of tap water through a household filter which is more accessible and time efficient for the consumers. The prototype would essentially attach onto any household tap and when the water is turned on it filters the water by going through the filter.

Methodology

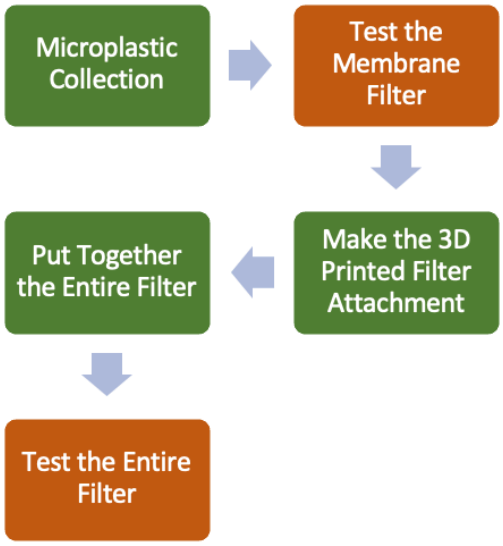


Figure 1: A summary of the two types of procedures conducted throughout the duration of this project: (1) The engineering procedure (green), and (2) the experimental procedure (orange).

Engineering Procedure

Start by first collecting all the materials needed for this experiment, laptop, pencil, engineering notebook, fusion3d printer, 3d printing filament, Simpure MCE gridded membrane filter sterile, timer, membrane filtration system, tweezer, fluorescent green polyethylene microspheres, laboratory spoon, 2 5l buckets, fountain pump – 90gph, cups, microscope, microscope slides and covers, spoons, shapr3d(software), 1/2in vinyl tubing, and stirring stick. After collecting all the materials make sure to follow the steps for Testing the Membrane Filter Sheet in Part B of the procedure.

When you complete the first part of the testing procedure download and open Shapr3D on your laptop. Using the schematic sketch of the filter, render the filter attachment on Shapr3D. Export the rendered 3D print as an STL file. Using a Fusion360 3D printer, print the STL filter attachment file.

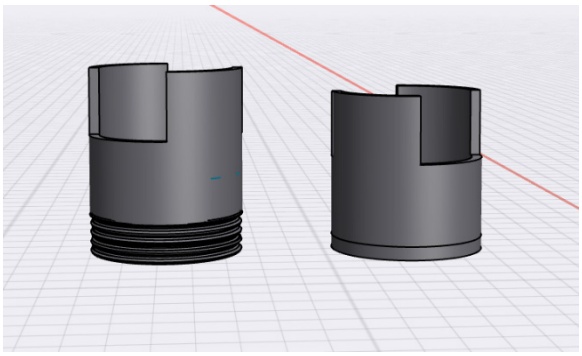


Figure 2: A screenshot of the 3D CAD file of the 3D printed filter attachment.

After the print is complete, the amount of time for this step will vary depending on what 3D printer is used to print the filter attachment, start setting up the mock faucet set up. Place the two 5L buckets next to each other. Stick the 90GPH fountain pump to the side of one of the buckets of your choice.

The fountain pump should be attached to the side of the bucket which is touching the other bucket. Take the 1/2in vinyl tubing and attach it onto the fountain pump, make sure that the tubing falls inside the other bucket. Cut the tubing as shown in the set-up in figure 3.

Now, fill up the bucket, which has the fountain pump, with water all the way to the top. Plug in the fountain pump into an electrical socket to ensure that it is working. It is considered working if water flows over to the other side of the bucket through the tubing. If the set-up is working, unplug the fountain pump and pour all the water back into the bucket with the fountain pump. Grab the microplastics and pour all of the microplastics, 10g, into the bucket with water in it. Using the stirring stick, mix the water continuously until the

microplastics are submerged and mixed in with the water. After all the microplastics are separated, and not in clumps in the water, move on to Testing the Filter Attachment in Part B of the procedure.



Figure 3: the set-up for the testing the prototype in the experimental procedure.

Prototype Testing

Start by testing the membrane filter sheet. Take the membrane filter sheet out of the packaging and rinse the membrane filter with sterilized water and then discard the water.

Weigh the membrane filter sheet on the scale. Make sure to zero the weight of the container that it is held in so that only the weight of the microplastics is accounted for you. Record the weight of the membrane filter in data table 1.1 on the column 'Weight of membrane filter sheet before'. Set the membrane filter to the side. Make microplastics using sandpaper and a PVC pipe. Weigh the microplastics on the scale. Make sure to zero the weight of the container that it is held in so that only the weight of the microplastics is accounted for you. Record the weight of the microplastics in data table 1.1 on the column "Weight of microplastics". Add the microplastics to 100mL of water. Mix the water so the microplastics and the water can mix. Set up the membrane filtration system. Place the filter inside the membrane filtration system. Close the membrane filtration system. Turn on the vacuum. Let the water filter through the system. Close the vacuum and remove the top part of the filtration system. Remove the membrane filter carefully while making sure not to drop any of the microplastics and place it in a petri dish. Weigh the membrane filter and microplastics on the scale. Make sure to zero the weight of the container that they are held in so that only the weight of the membrane filter sheet and microplastics is accounted for you. Record the weight of the microplastics in data table 1.1 on the column "Weight of membrane filter sheet after". After, place the membrane filter sheet under a microscope. Examine and count the number of microplastics that you can find on the filter sheet. Record the number of microplastics counted in table 1.2. Repeat steps 1 – 21 five more times to have a total of 6 trials at the end of testing.

Now test the filter attachment. Take the membrane filter sheet out of the packaging and rinse the membrane filter with sterilized water. Then discard the water. Set the membrane filter to the side.

Take the 3D printed filter and open it. Place a membrane filter sheet inside the filter attachment on the black circle. Then place another black circle stopper on the membrane sheet and close the 3D printed filter attachment. Screw the filter attachment onto the vinyl tubing. Make sure that the filter attachment is completely hanging over the empty 5L bucket. Wrap the filter attachment with a wet tissue, to ensure that all the 3D printing pores are covered. [This step was taken since the 3D printing available which conducting this experiment was not the type which was watertight.]

Add a balloon to the filter attachment bottom to ensure that all the water which is filtered out is caught in the balloon. This allows for measuring the amount of water filtered out and the amount of water leaked separately.

To start, turn on the fountain pump and start the timer for 30 seconds at the same time. Make sure that all the water that leaks is collected in the bucket and all the water passing through the filter is caught in the balloon, none of the water should be lost. After the timer ends, pour the water in the balloon into a graduated cylinder. Record the amount of collected water in the bucket in data table 1.2 on the column "Amount of water leaked from filter".

Pour the water in the balloon into a

graduated cylinder. Record the amount of collected water in the plastic tray in data table 1.2 on the column "Amount of water which passed through the filter". Open the filter and remove the membrane filter sheet.

Place the membrane filter in a labeled sterile bag for future use. Repeat Steps 1-21 three more times to have a total of 4 trials at the end of testing.

Data Collection

1.1: Testing the Membrane Filter Sheet

	Weight of microplastics mixed in water (g)	**Weight of membrane filter sheet before (g)**	**Weight of membrane filter sheet after (g)**
Trial 1	0.0273g	0.1680g	0.1944g
Trial 2	0.0333	0.1682	0.1850
Trial 3	0.0308	0.1816	0.2105
Trial 4	0.0329	0.2368	0.2599
Trial 5	0.0280	0.2945	0.2195
Trial 6	0.0339	0.2636	0.2926

1.2: Outer Shell Testing

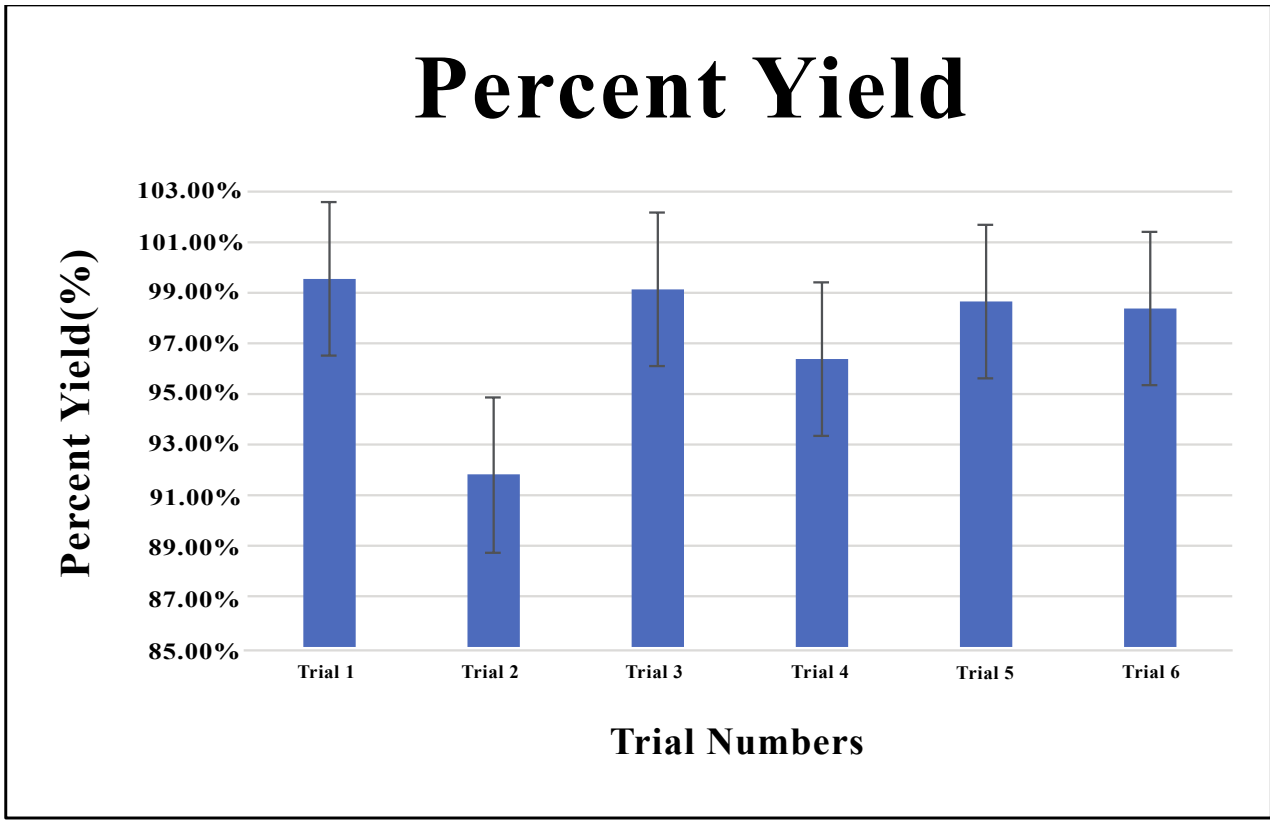
	Amount of water leaked from filter (mL)	**Amount of water which passed through the filter (mL)**
Trial 1	4	20
Trial 2	6	19
Trial 3	5	15
Trial 4	2	30

Data Analysis

5.1.1 Percent Yield with Trial 2 for Testing the Filter Efficiency Data Table

Trial #	Percent Yield
1	99.539
2	91.811
3	99.105
4	96.366
5	98.366
6	98.353

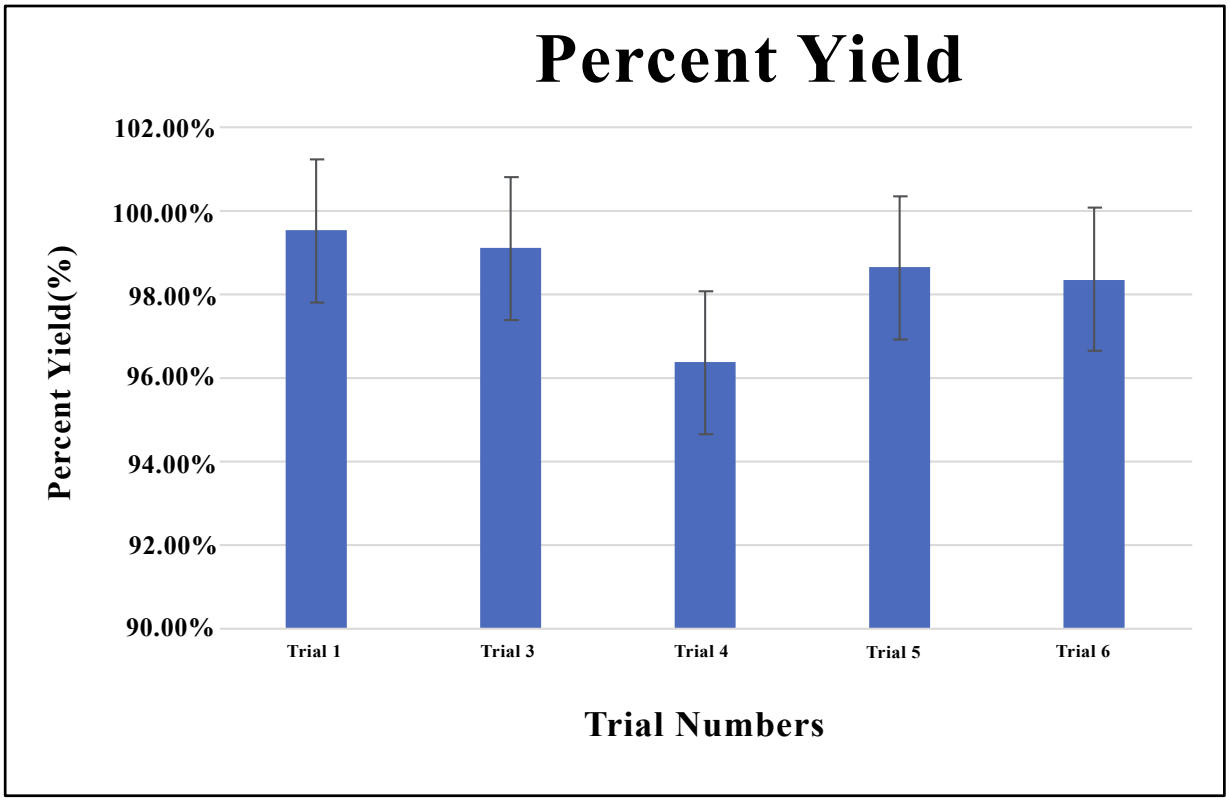
5.1.2 Percent Yield with Trail 2 for Testing the Filter Efficiency Graph



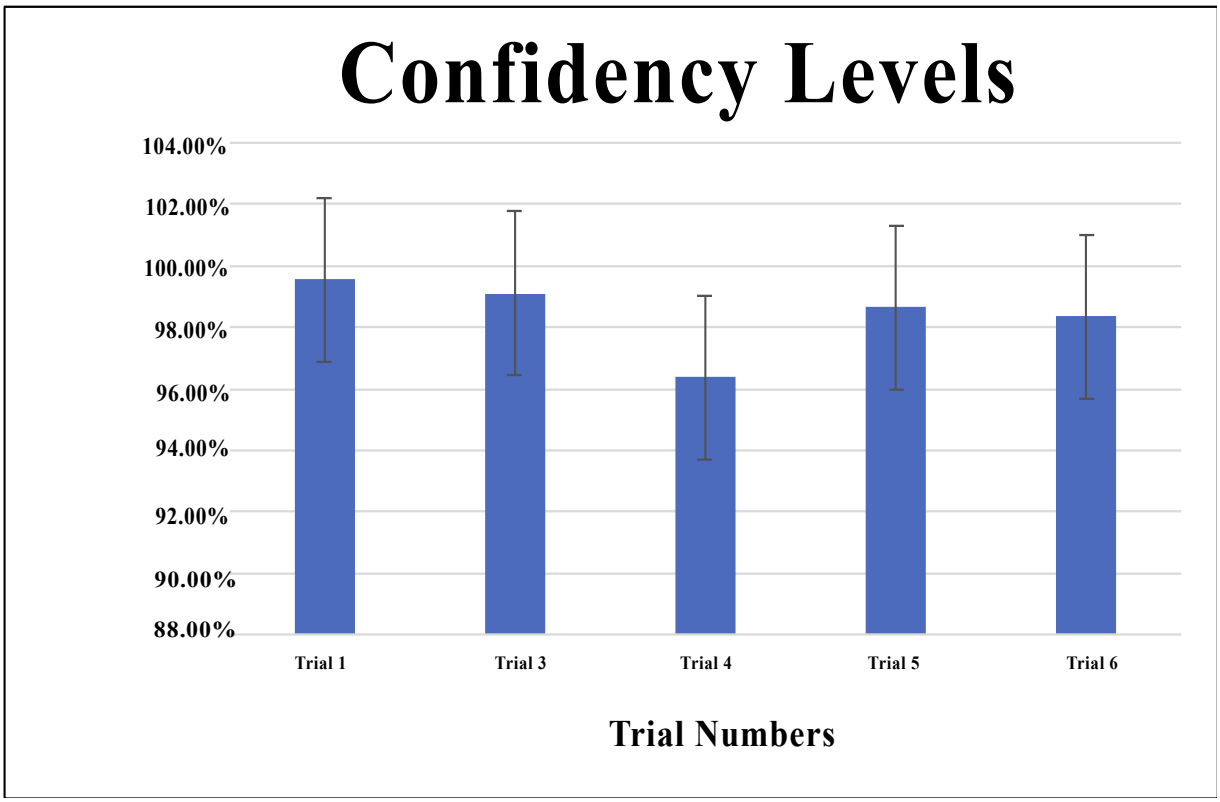
5.2.1 Percent Yield without Trial 2 for Testing the Filter Efficiency

Trial #	Percent Yield
1	99.539
2	91.811
3	99.105
4	96.366
5	98.366
6	98.353

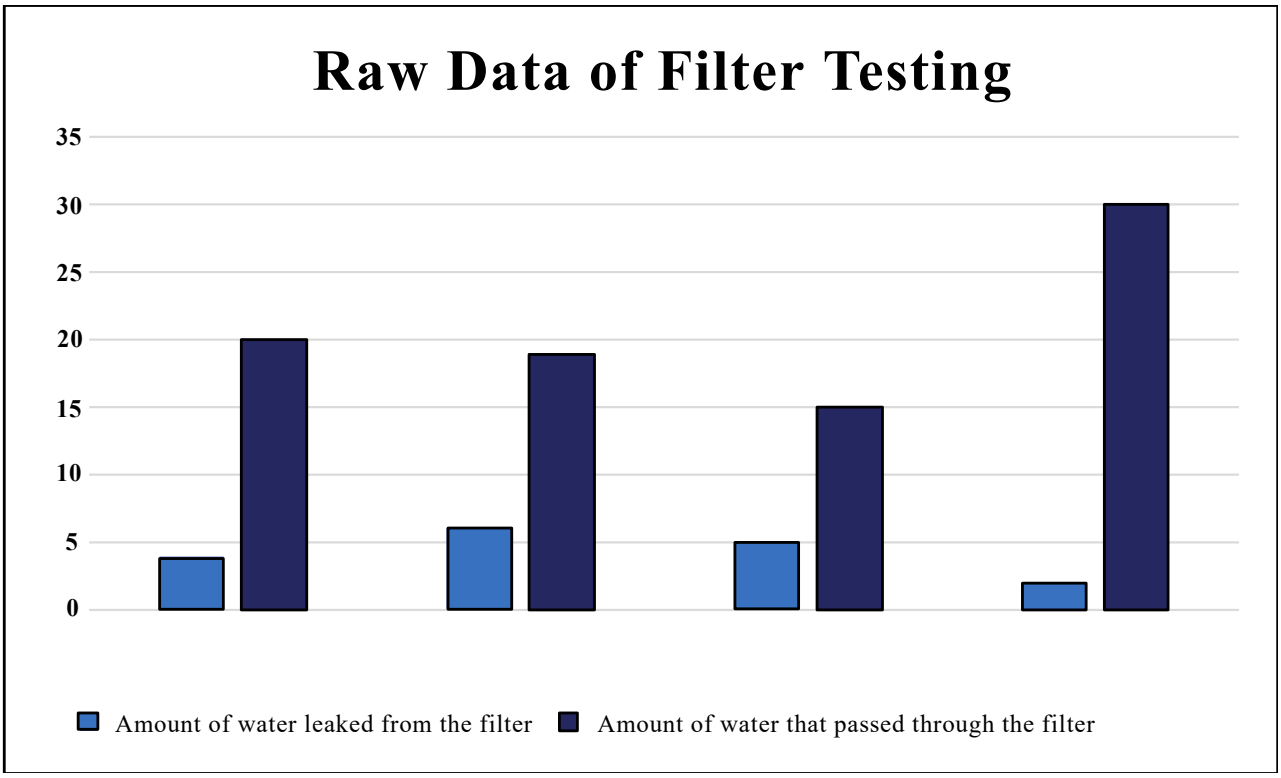
5.2.2 Percent Yield without Trial 2 for Testing the Filter Efficiency Graph



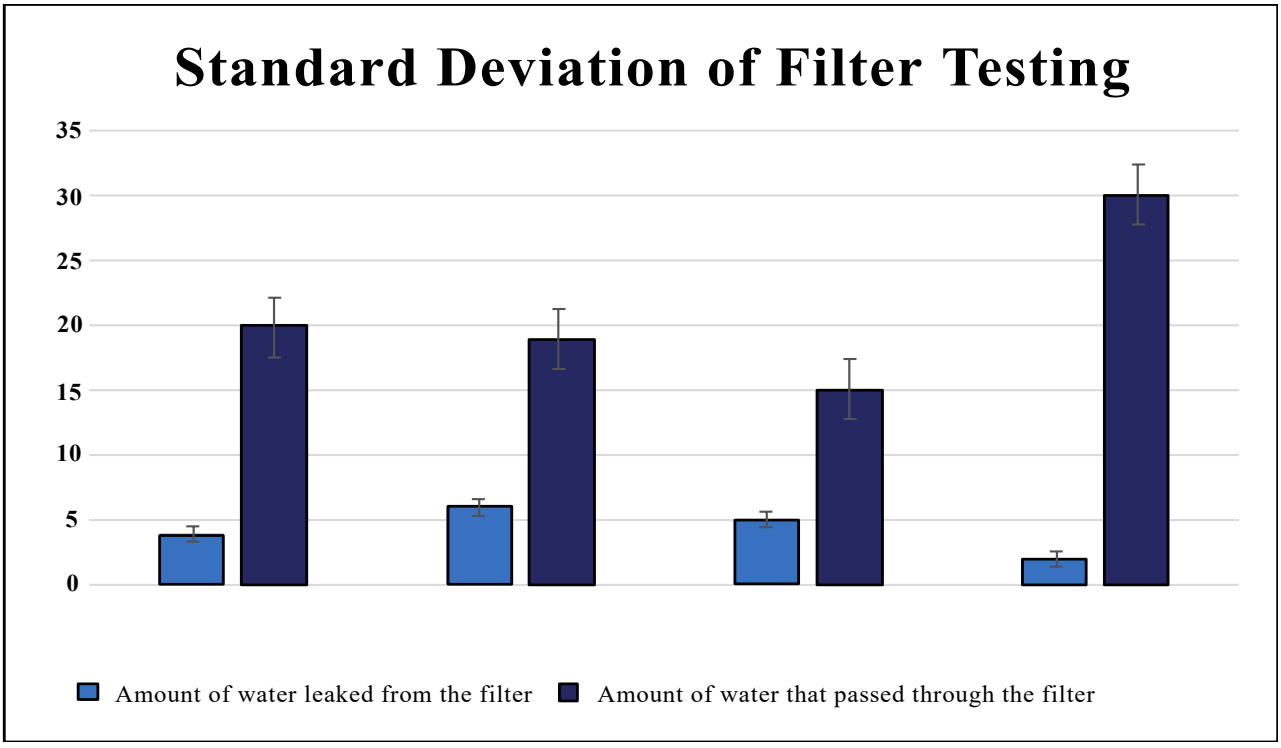
5.2 Confidence Levels without Trial 2 for Testing the Filter Efficiency



5.3 Raw Data Comparison of Leaked vs. Filtered Water During Filter Attachment Testing



5.3 Raw Data Comparison of Leaked vs. Filtered Water During Filter Attachment Testing



Results and Conclusion

At the end of this project, all the engineering goals were accomplished. The 3D microplastic filter attachment created during this project successfully, filtered out the microplastics out of the water with a mean percent efficiency of 97.403%. The filter also showed minimal leaking throughout the testing process with a significant amount of water being filtered out through the filter attachment.

A combination of engineering and experimental procedures was used in the method. The filter attachment was designed using software for 3D modeling and printed using a fusion 3D printer as part of the engineering process. The membrane filter sheet was put through its paces during the experiment, and the filter attachment's capacity to remove microplastics was assessed.

The filter's effectiveness was revealed by the data collection and analysis. The membrane filter sheet's ability to capture and retain microplastics was demonstrated during testing by the weight increase before and after filtration. The effectiveness of the filter was demonstrated by the amount of water that went through the filter and how much water leaked during the outer shell testing.

The results of the percent yield analysis varied from trial to trial, with percent yields ranging from 91.811% to 99.539%. Experimental factors, such as the distribution of microplastics and the efficiency of the filter attachment design, could be the cause of these variations. A greater understanding of the filter's performance would be provided by additional analysis, including confidence level calculations.

Overall, the project produced a working prototype of a filter attachment that appears to be capable of addressing the problem of microplastics in tap water. The filter design can be further improved and refined on the basis of the testing and data analysis outcomes. The filter attachment's efficiency should be improved, water leakage should be reduced, and filtration performance should be consistent and dependable in future iterations.

This engineering project contributes to safeguarding human health and mitigating the potential risks associated with microplastic ingestion by providing a practical solution for filtering microplastics from tap water. This project could serve as a basis for further research and development of microplastic filters for other types of water sources.

Bibliography

Ragusa, A., Svelato, A., Santacroce, C., Catalano, P., Notarstefano, V., Carnevali, O., Papa, F., Rongioletti, M. C. A., Baiocco, F., Draghi, S., D'amore, E., Rinaldo, D., Matta, M., & Giorgini, E. (2021). Plasticenta: First evidence of microplastics in human placenta. *Environment International*, *146*, 106274. https://doi.org/10.1016/j.envint.2020.106274

Guardian News and Media. (2020, December 22). *Microplastics revealed in the Placentas of Unborn Babies*. The Guardian. Retrieved January 25, 2023, from https://www.theguardian.com/environment/2020/dec/22/microplastics-revealed-in-placentas-unborn-babies

Guardian News and Media. (2021, December 8). *Microplastics cause damage to human cells, study shows*. The Guardian. Retrieved January 25, 2023, from https://www.theguardian.com/environment/2021/dec/08/microplastics-damage-human-cells-study-plastic

Joseph, N. (2019, October 1). *Tiny Plastics, big problems. UW College of Arts & Sciences. Retrieved November 16, 2022, from https://artsci.washington.edu/news/2019-10/tiny-plastics-big-problems*

Wagner, M., & Lambert, S. (2017). Freshwater Microplastics: Emerging Environmental Contaminants? Germany: Springer International Publishing.

D2ouvy59p0dg6k.cloudfront.net. (n.d.). Retrieved December 6, 2022, from

https://d2ouvy59p0dg6k.cloudfront.net/downloads/plastic_ingestion_web_spreads.pdf

Zeng, E. Y. (2018). Microplastic contamination in aquatic environments: An emerging matter of environmental urgency. Elsevier.

Yuan, Z., Nag, R., & Cummins, E. (2022, February 7). Human health concerns regarding microplastics in the aquatic environment - from Marine to Food Systems. Science of The Total Environment. Retrieved November 16, 2022, from https://www.sciencedirect.com/science/article/pii/S0048969722008221

Staff, E. D. N. (2022, March 29). The invisible plastic particles in our drinking water. Earth Day. Retrieved December 6, 2022, from https://www.earthday.org/the-invisible-plastic-particles-in-our-drinking-water/

Invisibles Final Report. Orb Media. (n.d.). Retrieved January 25, 2023, from https://orbmedia.org/invisibles-final-report

Lundy, T. (2022, March 24). How to filter microplastics from water. LifeStraw Water Filters & Purifiers. Retrieved December 6, 2022, from https://lifestraw.com/blogs/news/what-you-need-to-know-about-microplastics

Poerio, T., Piacentini, E., & Mazzei, R. (2019, November 15). Membrane processes for microplastic removal. MDPI. Retrieved December 6, 2022, from https://www.mdpi.com/1420-3049/24/22/4148

Lundy, T. (2022, March 24). How to filter microplastics from water. LifeStraw Water Filters & Purifiers. Retrieved January 25, 2023, from https://lifestraw.com/blogs/news/what-you-need-to-know-about-microplastics

An Everting Vine Robot With A Pneumatic Soft Gripper

Ansh Bayappu

Nikola Tesla STEM High School

Abstract

Search and Rescue teams have a hard job on their hands as they explore the dangerous rubble of fallen buildings, so there must be a safer and more efficient way to explore these caverns using robots. This project seeks to investigate a method to more efficiently explore caverns formed during a building collapse to increase the number of survivors found using a novel combination of a vine robot (inflated tubing) with a soft pneumatic gripper (created with a silicone elastomer). The goal was to create a transportable system that can run for 30 minutes of a Search and Rescue while everting (reeled material becoming outside) at a speed of 40 inches every minute in order to transport objects of all shapes into the rubble. A novel internal magnetic mount was designed that could handle larger loads than prior vine robots while lending itself easily to tip-based steering of the newly everted material. Throughout the process, soft pneumatic grippers were adapted to be more resilient to the harsh environment. The vine's movement was easily predictable by time of airflow as shown by the correlation coefficient of 0.99614. The vine traveled at an average velocity of 16.98 in./min. and a maximum velocity of 26.25 in./min. which didn't meet the goal, but future plans include the implementation of a fabric that can handle an increase in pressure and therefore velocity. This prototype shows the prospects of the combination of a vine robot with a soft robotic gripper during a building collapse.

Introduction

To avoid catastrophe, like in the recent Miami Surfside condominium collapse where it took 32 days before everyone was found, we can use everting vine robots with grippers to look through rubble more efficiently and assist in the process of Search and Rescue. Due to the fact that there might not be any clean drinking water under the rubble it is unlikely for survivors to survive for more than 72 hours. This is the time limit for Search and Rescue teams to find people by. There have been cases where people have been found 120 hours later surviving off of rainwater leaking into the rubble. We can try to use vine robots to send water down to anyone trapped. Out of the 72-hour window we really only have 48 hours of time to search as it takes about 24 hours to get crews onsite. It is essential to improve this domain to be more efficient so that we can increase the number of lives saved.

The current process of Search and Rescue begins when a host country's government sends out a response requesting emergency help. Then the collapsed building is inspected to determine the possibility of a further collapse and where people are likely to be (The New Humanitarian, 2009). Common debris found on building collapse sites include broken rebar, glass, concrete, downed power lines, gas leaks, and other sharp materials which the robot could extend over and help spot before people go near them (Occupational Safety and Health Administration). Once a person is found there are special tools that have been developed to drill through wood, metal, and concrete to try and get them out of the rubble.

Vine robots are inflated pieces of tubing, called a vine, which can be controlled. The vine can assist in this process by inspecting voids with minimal removal of debris to decrease the chance of a second collapse. The camera of the vine robot going into the rubble can increase the amount of information that onsite structural engineers have available to them when making decisions on what debris to move and is maneuverable unlike the current method of a camera on a pole.

One current method of searching through rubble are drones for aerial mapping. Drones are helpful for looking over a large area of space in a short time, as well as the surface of the site and thermal imaging in open areas. However, thermal imaging does not work through concrete or walls as they block most infrared radiation from passing. New throwable rovers are another technology currently being used. They have the advantage of reaching areas where humans can't go, especially those with poisonous fumes or if the void has the possibility to collapse. One problem with these throwable rovers is that they tended to get stuck in cracks (Brown, 2021). They also have reliability problems like in one case in New Zealand in 2010 where a robot short circuited and significantly delayed a Search and Rescue operation to find 29 miners trapped in a mine. As these robots drive further away from the operator and possibly into the rubble they tended to lose radio signal and become inoperable. A current vine robot solution was developed in March of 2022 called RoBoa (der Maur et al., 2022). RoBoa is a vine robot that is controlled by pneumatic actuators on the outside at multiple spots along the body. It is able to get into spaces that are under 10.1 cm in diameter. One of its greatest advantages is its ability to decouple the link between steering and eversion. It has a bending angle of up to 120 degrees leading to high maneuverability. It's also able to lift or go under up to a kg in weight. Its main success was that in testing with the Swiss Army rescue troops it showed promise and proved that RoBoa and vine robots have applications for the field of Search and Rescue.

Research shows that vine robots, air filled tubes, can navigate through tight, sticky, and sharp environments, cover gaps below them, and elongate to reach into every corner of a collapsed building (Blumenschein, 2020). With the use of a soft robotic gripper at the end, the system is still flexible to fit through gaps. The gripper has the ability to conform to the object it holds since it is filled with air. By combining a vine robot and a soft robotic gripper, the product can go under doors and unlock them with minimal damage to the doors. With additional attachments it can also help send water or oxygen to located survivors. This combination of the flexibility of the vine robot and the strength of the gripper has never been done before and can be used to advance the field of Search and Rescue. This project seeks to investigate a method to increase the speed of Search and Rescue operations using a vine robot with a soft robotic gripper.

Methodology

The vine robot is being made out of LDPE or Low-Density Polyethylene as a body. This is a cheap plastic and is useful for prototyping, but in real field applications a more expensive ripstop nylon would be better. The LDPE was cut to a length of 9 feet and 3 inches for testing purposes. After marking 1 inch on either side of the robot all the way around the tube we can then begin to seal it. To save money instead of using a plastic sealer this project used an old hair straightener. The initial temperature of the straightener was too high where it would just melt the plastic instead of sealing it, so the addition of an old rag around the tube to decrease heat transfer solves this problem. After sealing one end, 2 holes were poked into the plastic and a string tied between them. The string was then tied to the spool. The purpose of the string is to allow any inverted tubing to go to the tip and allow the robot to go to its true maximum length with the current amount of reeled plastic tubing. Without the string the vine can only reach a maximum length of half of the total reeled length.

In a 44-ounce LocknLock container holes of 1 inch at the top of the container, ½" on the side and 1.25" on the side exactly opposite the ½" hole were drilled. Then 3D parts were designed of a spool, spool cap, motor base, outlet tube, and spool aligner (Fig. 1). These were then printed on a MakerBot replicator 3D-Printer.

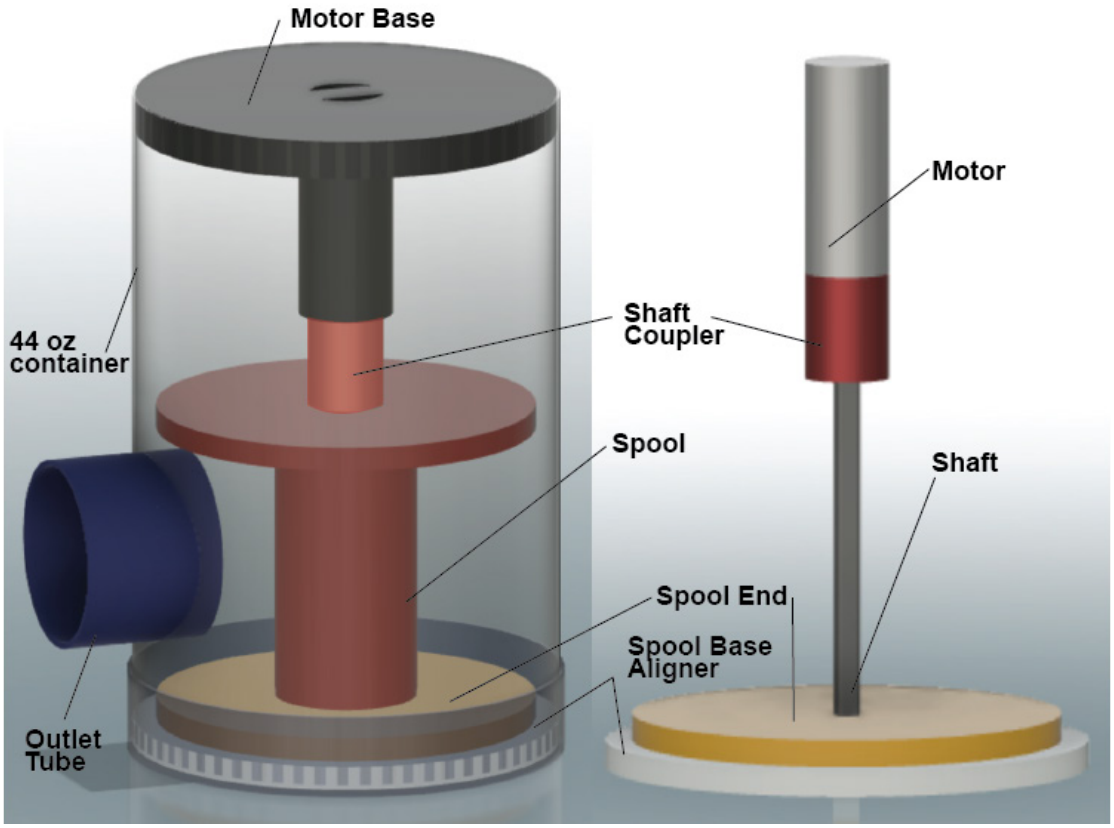
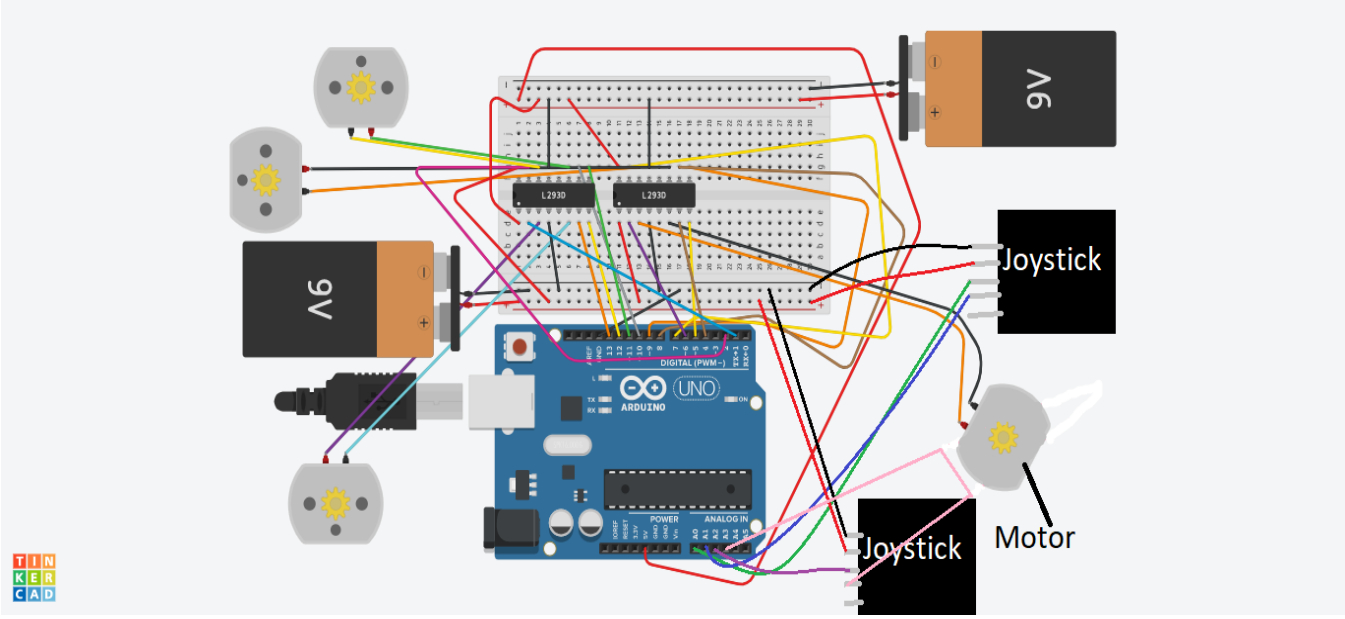


Figure 1. Here is how the frame of the vine robot goes together. The motor base and the spool base aligner are glued to the sides. On the motor base are holes to slide out the motor wires from which they shall be attached to the Arduino.

The interior assembly begins by pushing the d-shaft through the spool. There was a small bit intended leniency to ensure fit. This gap was filled with hot glue at the tip to push against the wall and prevent the d-shaft from coming out. Hot glue was used to hold the spool's basic shape with the spool cap, to ensure the connection epoxy was used. The coupler was then tightened onto the shaft and then the motor. The motor base was then pushed on and I realized that the screw on the coupler blocked the inner cylindrical support. The motor base had to be reprinted with a lower support. Then the motor base was hot glued to the container lid. The motor was pushed into the motor base so that only the leads stick out the back and was subsequently hot glue it in place to the motor base. The bearing was inserted into the spool aligner and kept a rubber band around it to allow for a tight fit at the bottom of the container. The spool aligner was then pushed into the bottom of the container. Keeping the d-shaft in the aligner and screwing the lid back on, two wires were soldered to the motor and the system was tested by keeping both wires to a 9V battery and then wired to the Arduino for 1 motor according to this diagram (Fig. 2).

Figure 2. Here is how the circuit between the Arduino, 2 Joysticks, and 4 motors are. The 9V batteries may be used or an Arduino power module can be substituted. To have more spacing between wires and ports an Arduino Mega 2560 is better.



The cap lid was then unscrewed to take the whole spool out. The string was tied to the spool and winded up all the way and pulled out through a 1.25” hole. The outlet tube was then pushed through the 1.25” hole and required some sanding to fit through. The inlet tube was inserted into the ½” hole and tubing was connected to that. The edges of the outlet tube and insert to the container were hot glued. The tubing that was pulled through to the 1” mark invert it at the tip and hot glue it to the outlet tube. This was adapted in a later design with a tapered rubber seal to provide a better seal and allow for easier fixes of tears. The tubing was then connected between the insert and the air supply with a cone adapter. As an air supply a CAT power station’s air compressor was used.

At this point I tested the system to check if it everted as I kept loose some spool. This did not occur as leaks needed to be sealed after which it everted. Any issues with eversion can be solved by fixing leaks and increasing air pressure. Due to there being no steering system the vine moves in the easiest direction possible. A tube not everting indicates that there is not enough pressure within the tube or that there is not enough force at the tib of the tube that would cause it to evert.

Soft grippers have 4 parts: an inlet hole, expandable silicone, a non-stretch fabric, and a channel. The initial design of a soft gripper consisted of an acetate channel that would split two layers of silicone. A pipe was added into this gap and air would flow inflating this section. The silicone when inflated expands in length. The non-stretch fabric attached on the other side can’t stretch, so the gripper curls towards the fabric side as the silicone in this position is allowed to be of greater length than the fabric. The acetate in the first design (Fig. 3) didn’t split the two layers of silicone well, which caused all the air to flow into one spot causing the gripper to burst. Another problem with this initial design was that each time the pipe was inserted into the hole on the side the hole would expand in size. Air leaks at the edges of the pipe would force this gripper design to become inoperable .

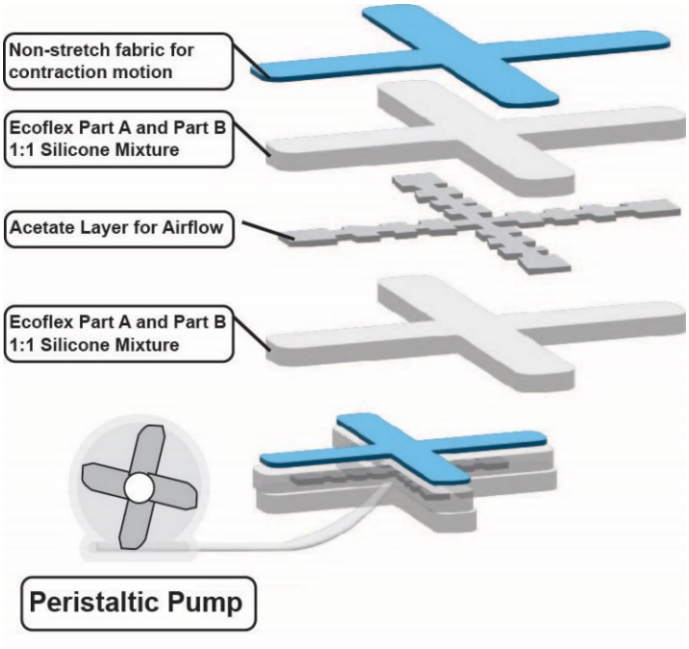


Figure 3. The 4 layers of the soft gripper are a layer of an Ecoflex mixture, an acetate layer, another Ecoflex layer, and a non-stretch fabric layer. The fabric prevents stretching and therefore causes each arm of the gripper to move inward when air is added via the peristaltic pump.

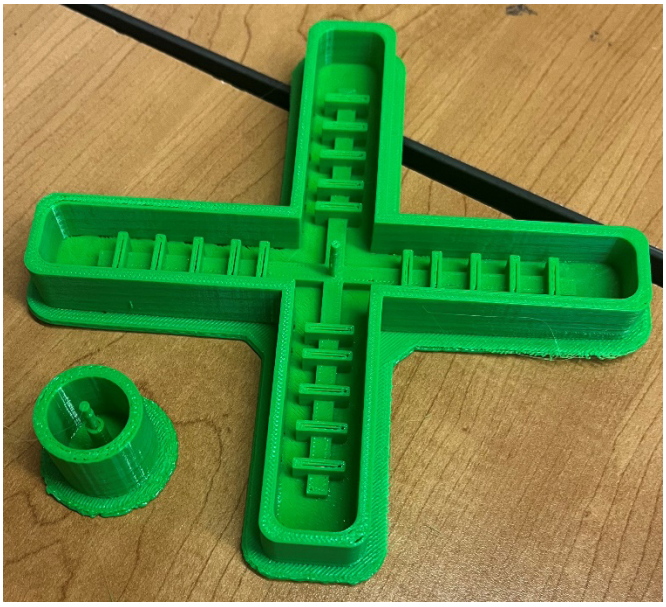


Figure 4. This mold for the gripper creates a channel directly within the silicone. The coupler mold has a hole that tapers towards the gripper and allows us to center the coupler on the gripper.

To solve these aforementioned issues a different type of soft gripper was used: an embedded channel gripper. This gripper has the channel directly embedded into the silicone. The air is pumped directly into this open channel. In this new design a new coupler also made out of silicone was used. It was designed to be smaller than the pipe, so that once stretched and the pipe is inserted it can be held tighter and a semi-airtight seal is created.

The process to make an embedded silicone gripper takes 6 hours if an oven is used or 20 hours if it is let to dry at room temperature. The process to make the gripper involves mixing Ecoflex Silicone Part A and B at a 1:1 ratio. The current size I used (Each leg is 2.5 inches long) required about 78 grams of silicone in total per gripper. With 1 pint of Ecoflex Part A and Part B is enough to make 8 grippers. Out of the 78 grams 66 grams of silicone is used to pour into the initial gripper and coupler molds (Fig. 4).

Once the silicone is poured it goes into the oven for 2 hours or in the open air for 6 hours. The gripper and coupler were then removed from the mold. An additional 12 grams of silicone were used to glue the gripper to the fabric and the coupler to the gripper. The gripper needed to be pushed down on all edges to an airtight seal. Silicone was applied to the bottom of the coupler mold (small hole) and a nail through the center of the coupler mold was used to align it to the center of the gripper and again required to be pushed down to ensure an airtight seal. This must cure for an additional 2 hours in the oven or in the air for 6 hours. Then the fabric was cut around the gripper and slit at the 4 inner corners of the gripper. The gripper was then assessed by inflating via a bike pump (Fig. 5a).

5(a)



5(b)

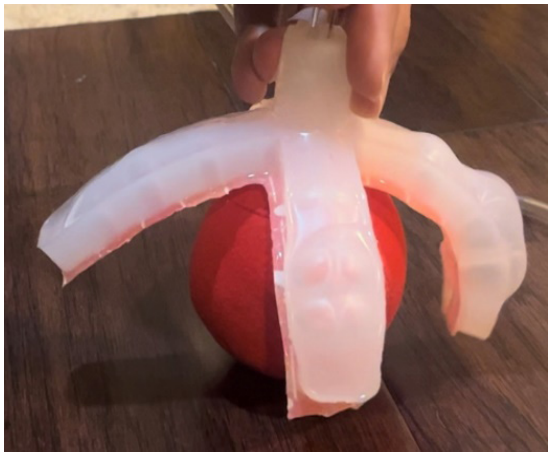
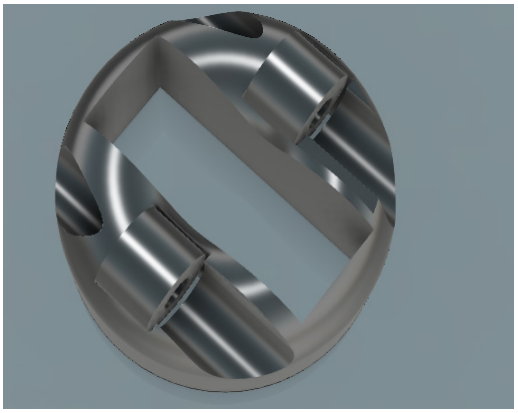


Figure 5. (a) Process to create gripper. (b) Gripper inflation, but uneven movement as left leg doesn't curve as much as other three legs.

One problem noticed during testing was that one leg would inflate before the others. To solve this issue a thin layer of silicone glue is spread over the top of the gripper. This creates a small bit of additional weight which the air needs to push and as air shall travel the easiest path less air will flow in that leg (Fig. 5b).

The inner tip mount's purpose is to constantly remain at the tip of the vine robot even as it extends. The outer tip mount magnetically connects to this carrying the gripper. They are connected via a series of magnetic rollers which allow the vine to smoothly move through as it everts. The initial design for the tip mount was based on Luong's magnetic mount design in 2019. This smooth semi-toroidal design prevents the vine from being torn as it passes over. Due to a scaled down version of design being used the central hole was too small for the uninflated internal tubing to fit through (Fig. 6a). An oval shaped and rectangular hole designs were tested for this and the rectangular design performed better. But, even with this new design the problem of detaching occurred between the inner and outer tip mounts.

To solve the issue of detaching more rollers were added to the tip mount to give a stronger connection between the pieces as well as space between the magnets and the back of the outer mount to give space for the uneven curvature at the tip of the vine (Fig. 6b). The gripper was attached to this with a cylindrical support and hole for the pipe as it goes through the vine robot. This cylindrical piece can be customized to create more modules to fit different tools such as microphones and a GPS (Fig. 6b).



6(a)

6(b)

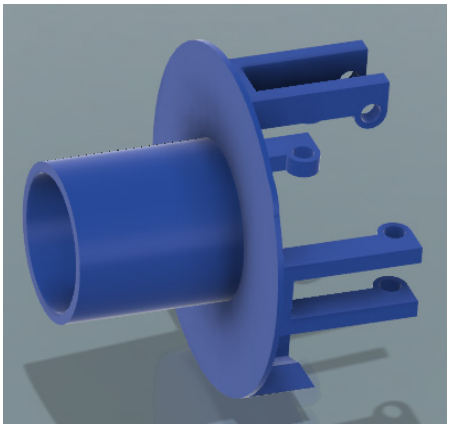


Figure 6. Newly developed (a) inner tip mount and (b) outer tip mount working on the principle of magnetic rollers.

This system was then connected to an air compressor (CAT Power Station) to allow the system to be operable by a single person. A three-way splitter (only two outlets were used) to connect the vine and the gripper. Ball valves were added to allow the vine to be in an airtight environment and seal the gripper in with air once initially filled.

The gripper was attached to the outer mount of the vine. In order to distribute more of the gripper's load on the vine instead of the magnets a new mount was developed where the inner mount sits inside the outer mount and a false roller is used.

When the vine was lengthened with the gripper on it, the sticky silicone gripper dragged on the floor leading to the vine initially having uncontrollable movement and then coming to a complete stop. Initially, to solve this issue a solution was attempted where the gripper did not touch the ground, but this solution would require a base so wide that it would be useless for going into the rubble. I solved this problem by coating the gripper in oil to reduce friction.

Data was then collected on the vine and the gripper. Position data of the vine was collected by laying out a tape measure parallel to the vine. The vine was then continuously pumped with air and allowed to grow. Then afterwards, the video was used to find the position of the tip at different intervals (Table 1). Gripper force was calculated by pulling a ball out of the inflated legs of the gripper using a force gauge or in other words a luggage weight checker, and the maximum value in this process is gripping force (Fig. 7).



Figure 7. Pulling ball out of gripper to test gripping force using force gauge or luggage weight checker.

Trial 1: Vine Length and Velocity					
Time (sec)	**Length(in)**	**Velocity(in/s)**	**Time (sec)**	**Length(in)**	**Velocity(in/s)**
0	0		52	14.5	0.278846154
4	1.75	0.4375	56	16.25	0.290178571
8	3.25	0.40625	60	17.25	0.2875
12	4.75	0.395833333	64	18.375	0.287109375
16	5.125	0.3203125	68	19.25	0.283088235
20	5.875	0.29375	72	20.75	0.288194444
24	7.375	0.307291667	76	21.875	0.287828947
28	9	0.321428571	80	22.25	0.278125
32	9.75	0.3046875	84	23.875	0.28422619
36	10.75	0.298611111	88	24.125	0.274147727
40	11.875	0.296875	92	25	0.27173913
44	13.625	0.309659091	96	27.125	0.282552083
48	13.875	0.2890625	100	27.5	0.275

Table 1. Data from vine's movement as recorded on a tape measure velocity calculated by doing (distance)/(time)

Results and Conclusions

The ability to electronically control the eversion of the vine robot and stay without an outlet for up to 2 hours was achieved. Objects of all shapes were able to be lifted due to the pneumatic properties of the soft gripper. The initial target for eversion speed, nor the ability to lift pounds wasn't reached as an inflatable tube incapable of lifting that heavy of a weight at its tip. This project sought to investigate how an everting vine robot that can carry payloads can assist in Search and Rescue. The prospects are immense and must be further investigated.

Between the length of the vine as it everted and time of airflow into the vine there was a correlation coefficient of 0.99614 (Fig. 8). In a theoretical scenario we want each input of air to cause a change in length. The r value indicates smooth movement of the vine and a near constant amount of airflow into the vine signaling that minimal leaks were being created throughout the process of eversion. The strong positive correlation indicates that the vine had a smooth controllable increase in length.

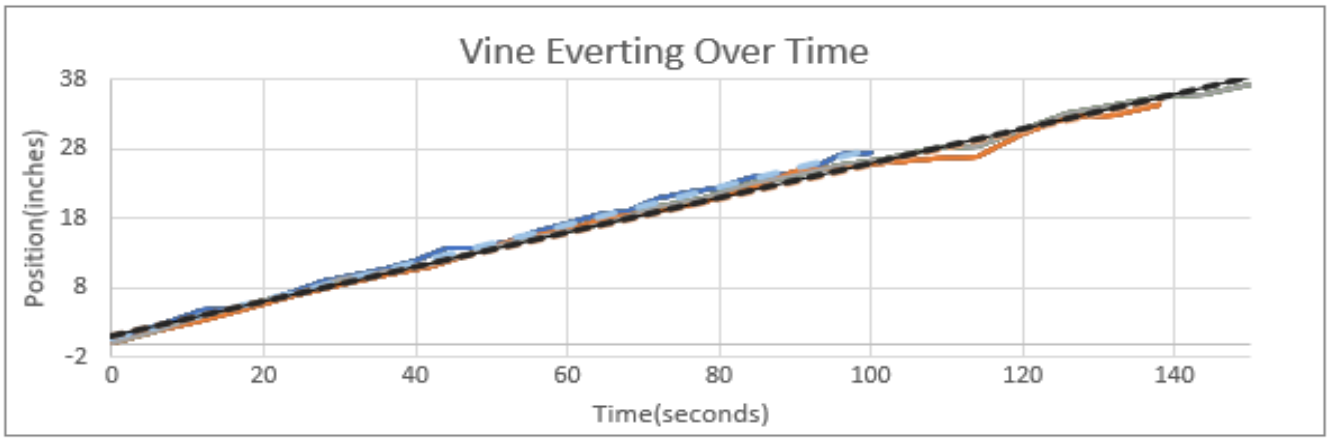


Figure 8. Graph of change in length of vine as it is continuously pumped with air.

The vine reached a maximum velocity of 27 inches per minute (Fig. 9). It had had an average velocity of 17 inches per minute. This speed doesn't meet the engineering goal of a max speed of 40 inches per minute, but a switch to the more durable ripstop nylon will lead to an increase in velocity. Since ripstop nylon can hold more pressure as it is stronger more air can be pumped without it tearing, so speed will increase.

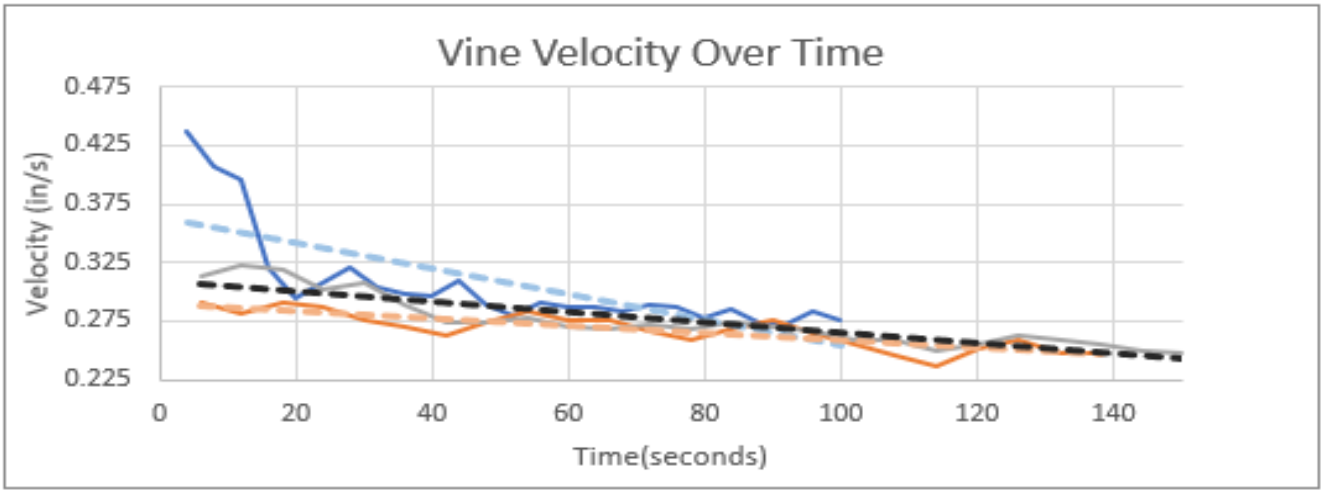


Figure 9. Graph of change in velocity of vine as it is continuously pumped with air and changes in length.

The gripper had a peak gripping force of 2.56 pounds and an average gripping force of approximately 2.32 pounds. This means that it requires an average of 2.32 pounds of force to remove the ball from the grasp of the gripper. This may seem low but can be increased by using a larger payload so less air is required to reach the ball, and more is being used to hold the ball tighter.

In terms of the next steps, this design shall require a steering system. This design hands itself well to a tip based steering system with actuators connected to the inner tip mount. A larger container and vine will be required to fit this. A nearby search and rescue building collapse training ground is being sought to evaluate this design.

References

Blumenschein, L. H., Coad, M. M., Haggerty, D. A., Okamura, A. M., & Hawkes, E. W. (2020). Design, Modeling, Control, and Application of Everting Vine Robots. *Frontiers in robotics and AI*, *7*, 548266. https://doi.org/10.3389/frobt.2020.548266

Brown, D. (2021, June 30). *Throwable military robots sent to assist with Florida condo collapse. The Washington Post. Retrieved January 16, 2023, from https://www.washingtonpost.com/technology/2021/06/30/throwable-robot-florida-condo-collapse/*

der Maur, P. A., Djambazi, B., Haberthür, Y., Hörmann, P., Kübler, A., Lustenberger, M., Sigrist, S., Vigen, O., Förster, J., Achermann, F., Hampp, E., Katzschmann, R. K., & Siegwart, R. (2022, March 28). RoBoa: Construction and evaluation of a steerable vine robot for search and rescue applications. arXiv. Retrieved October 28, 2022, from https://arxiv.org/abs/2203.15145#

Glick, P. E., Adibnazari, I., Drotman, D., Ruffatto Iii, D., & Tolley, M. T. (2022). Branching Vine Robots for Unmapped Environments. *Frontiers in robotics and AI*, *9*, 838913. https://doi.org/10.3389/frobt.2022.838913

Luong, J., Glick, P., Ong, A., deVries, M. S., Sandin, S., Hawkes, E. W., et al. (2019). "Eversion and retraction of a soft robot towards the exploration of coral reefs," in *IEEE International Conference on Soft Robotics* (Seoul), 801–807. doi: 10.1109/ROBOSOFT.2019.8722730

Occupational Safety and Health Administration. (n.d.). *Structural Collapse Guide. Occupational Safety and Health Administration. Retrieved January 19, 2023, from https://www.osha.gov/emergency-preparedness/guides/structural-collapse*

The New Humanitarian. (2009, October 8). How to: Rescue people trapped in a collapsed building. Relief Web. Retrieved January 16, 2023, from https://reliefweb.int/report/indonesia/how-rescue-people-trapped-collapsed-building

Greer, J. D., Morimoto, T. K., Okamura, A. M., & Hawkes, E. W. (2019). A Soft, Steerable Continuum Robot That Grows via Tip Extension. *Soft robotics*, *6*(1), 95–108. https://doi.org/10.1089/soro.2018.0034

Software Applications/ Computational Sciences

A Quantum Optimization Algorithm to Efficiently Route Public Transportation in Cities

Ashwin Kaliyaperumal

Nikola Tesla STEM High School

Abstract

Cars are one of the biggest polluters of greenhouse gases and to reduce their impact on the environment, there needs to be a nation-wide shift towards the use of public transportation - specifically the use of buses. Designing new bus routes can help buses arrive more frequently on time, especially during rush hour, which is an incentive for people to use buses. The Urban Transit Routing Problem (UTRP) is an NP-hard optimization problem, solved with heuristic algorithms that output approximate solutions, that focuses on constructing bus routes from an existing road network. The purpose of this project is to use quantum computing to solve the UTRP by minimizing a road network's average travel time (ATT), a measure of the average travel time for all passengers to travel between all pairs of bus stops. My engineering goal was to use quantum computing, which has a theoretical advantage over heuristics in solving NP-hard problems, to produce routes with a lower ATT than heuristics for Mandl's Swiss benchmark and existing bus routes in downtown Seattle. My algorithm achieved an ATT of 10.11 minutes for the benchmark, which is an improvement over current UTRP solvers. Additionally, my algorithm decreased the ATT in a section of downtown Seattle by 46%. This research supported my engineering goals by showing the feasibility of using quantum computers to develop routes with a lower ATT than both classical heuristic methods for a benchmark and existing bus routes in downtown Seattle.

1. Introduction

1.1. Rationale

With global temperatures on the rise from greenhouse gas (GHG) emissions, cars, which are one of the biggest sources of anthropogenic GHG emissions, need to be replaced by less polluting means of transport. According to the U.S. Department of Transportation, cars were the second largest polluter of carbon dioxide in 2010 (US Department of Transportation, 2010). An increase in the ridership of public transportation, would help reduce these emissions. Public transportation ridership in the U.S., however, is extremely low. According to the United States Census Bureau, in 2019, only about 5% of people took public transportation. Buses were the most used form of public transportation, with about 2.3% of people in the U.S, or about 46.3% of public transportation users, taking the bus. (US Census Bureau, 2019).

Other forms of non car-centric transportation are already being used frequently around the world. The Dutch have become famous for their increased use of bikes thanks to convenient bike parking, an abundance of safe bike lanes, and a traffic system that considers both bikers and drivers in the Netherlands. Europe and China have expanded their high-speed rail networks over the past twenty years. This has happened at such a quick rate in Europe, in fact, many short distance airlines have had to shut down out of being outcompeted by the new high-speed trains. Both of these examples show the feasibility of switching to more renewable modes of transportation, for both short and long-distance trips.

In the U.S, improvements in bus reliability and scope will most likely increase public transportation the most. Adding new bus

routes doesn't require the expensive costs that building new rail networks does, and most of the infrastructure for buses is already built. The existing road network built to serve cars can be used to run buses. Increasing ridership will help buses get more funding to expand their routes and the number of stops they service, reducing the number of cars on the road as more people choose to take the bus.

One reason for a lack of ridership is that buses are regularly delayed from their schedules and are stuck in traffic. According to the Seattle Times, buses from King County Metro arrive on time about 77% of the time and are usually late during rush hour, when people need it most (Gutman, 2018).

A solution to this problem could be the use of a quantum optimization algorithm that would help efficiently route buses, increasing the percentage of buses that arrive on time as well as minimizing the amount of time the buses spend stuck in traffic.

This problem has been introduced and developed as the urban transit routing problem (UTRP), a variation on the more common vehicle routing problem. The UTRP starts with a graph of bus stops, which are connected with edges that each have their own travel time. Each pair of bus stops has their own travel demand that indicates the amount of people that want to travel from a specific bus stop to another. UTRP solutions try to minimize the overall time it takes to get from one bus stop to another throughout the entire set of bus stops, as well as decrease the total number of transfers between stops. It also considers the travel demand and transfer time between routes.

Some classical algorithms that have been used to solve this optimization problem have used metaheuristic algorithms, like genetic algorithms, particle swarm optimization, and simulated annealing.

However, quantum computers are better suited towards solving optimization problems than modern, classical computers. Instead of using bits to store information, which can only be in a binary state, quantum computers use qubits that can be in both states simultaneously. This stems from the quantum mechanical property of superposition and allows for the computational power of a quantum computer to increase exponentially with the number of qubits it has. For example, a quantum computer with fifty qubits matches the processing power of a modern computer with 250 bits. Quantum computers also use other quantum mechanical properties, like interference and entanglement, which allow for additional speedups when computing tasks in parallel and further reduces the computational time.

Modern quantum computers, however, are limited in their size and accuracy. These noisy-intermediate scale quantum computers usually don't work well with large programs and certain qubit operations. The solution to this is to use quantum annealing computers, like the ones developed by D-Wave. These solve optimization problems, formatted in the Quadratic Unconstrained Binary Optimization (QUBO) model, through a heuristic approach called quantum annealing. These computers currently support problems with over 1000 qubits.

Formatting the UTRP into a problem that a quantum computer can optimize will allow for more efficient solutions compared to classical algorithms.

In this paper, I focus on answering two research questions:

1) Can a quantum computing approach to the UTRP produce with a lower average travel time for the Mandl's Swiss Road network benchmark?
2) Can a quantum computing algorithm to solve the UTRP produce routes with a lower average travel time than existing bus routes in cities?

where the average travel time is defined as the average travel time over the entire road network in minutes per transit user.

1.2. UTRP Definition

The UTRP attempts to design routes (in this case, bus routes) with a known set of nodes (bus stops) and a road network that connects the nodes. The graph with the connected road network with the nodes is also called the transit graph. The UTRP is also constrained with a travel time matrix

and a demand cost matrix. The travel time matrix notes the transit time for each pair of nodes that are directly connected. The demand cost matrix describes the number of people that travel from any pair of nodes. Both matrices are symmetric, so that both the travel time and demand between any two nodes are the same in both directions of travel.

The UTRP generally has two parts to it: the operator cost and the passenger cost. In this paper, I only focused on optimizing the passenger cost. The passenger cost, for an initial transit graph, is the average travel time for all sets of stops and for all passengers, with a specific route set. This is defined mathematically by Buba et al. with the formula:

$$ATT = \frac{\sum_{i,j=1}^{n} d_{ij}\,\lambda_{ij}}{\sum_{i,j=1}^{n} d_{ij}} \qquad (1)$$

where n is the total number of nodes, i and j are the node numbers, d_{ij} is the demand between nodes i and j, and λ_{ij} is the minimum total travel time between nodes i and j (Buba et al. 2016). The minimum travel time for a certain route set is calculated by using the Floyd-Warshall algorithm (Floyd, 1962) and the travel time matrix.

Solutions to the UTRP consist of a route set with a certain number of routes. Each route is a path, with a start and end node, that contains a set number of nodes. For example, in the Mandl's Swiss Road Network benchmark, I will be optimizing for four routes with a maximum of eight nodes each. Transfers between routes are allowed, defined to be a switching of routes in a route set and are only allowed if both routes share a common node to transfer from. Each route and the overall route set are subject to some additional constraints as well: 1) all nodes in the original graph must be included, 2) each route must connect to any other route (this constraint guarantees that every node will be able to reach any other node with a certain number of transfers), and 3) each route is free of any repeated nodes.

2. Formulation of the UTRP Objective Function and Conversion to a Quantum Problem

The quantum annealing (QA) algorithm, to solve the UTRP, attempts to optimize and find the candidate routes that, together, meet the highest amount of demand. There are four steps to the algorithm:

Generate candidate route sets from the original transit graph.

Create a QUBO matrix using the candidate route sets and the UTRP cost function.

Optimize the QUBO matrix using D-Wave's hybrid quantum-classical samplers and get the optimized route sets.

If there are any nodes missing from the original transit graph in the optimized route sets, repeat from step 2 and add a weigh factor to the QUBO matrix in step 2.

2.1. Generating Candidate Route Sets

The candidate route sets are generated using a Breadth First Search algorithm to find all possible routes for each pair of nodes. Potential routes that contain exactly the maximum number of nodes per route are added randomly to the list of candidate routes chosen to include about 10% of all routes that had exactly the maximum number of nodes, however, this percentage can be tweaked to allow a greater or smaller sample set of candidate route sets.

2.2. The UTRP Objective Function

The UTRP objective function for the QA attempts to both choose a set number of routes and to make sure the overall route set meets as much demand as possible. To ensure this, the objective function is split up into two parts: a cost function and a constraint function. The objective function is written in terms of binary variables, as it will be easier to convert the function into a form the quantum annealer can solve. This will be gone over in further depth in a later section.

The cost function is mathematically described as:

$$cost(v, w) = \left(\sum_{q_{ij} \in B_{vw}} TT_{vw} d_{vw} q_{ij} \right)^2 \qquad (2)$$

where v and w are nodes on the transit graph, q_{ij} is the binary variable that describes if a certain candidate route is in the optimized route set (i is the route index and j is the candidate route index), B_{vw} is the set of candidate routes that don't contain the nodes v and w, TT_{vw} is the travel time between nodes *v* and *w* on the transit graph, and d_{vw} is the demand between nodes v and w. This cost function

penalizes routes that do not meet the demand of a certain pair of nodes. The overall objective function applies this cost for all pairs of nodes.

The constraint function ensures that, for each route index i, there is only one candidate route j that gets picked. As formulated by Neukart et al., this is described mathematically as:

$$contraint = \lambda \sum_{i} \left(\sum_{j} q_{ij} - 1 \right)^2 \quad (3)$$

where, again, *i* is the route index, *j* is the candidate route index, and q_{ij} is the binary variable that describes if a certain candidate route is in the optimized route set. λ is a scaling factor that guarantees that the constraint is always met (Neukart et al. 2017). It is found by using the value of the maximum cost of any two routes. This ensures that the constraint is always greater than the cost function and only returns *i* number of routes.

Putting the cost and constraint functions together yields the objective function, described mathematically as:

$$Obj = \sum_{v,w \in G} \left(\sum_{q_{ij} \in B_{vw}} TT_{vw} d_{vw} q_{ij} \right)^2 + \lambda \sum_{i} \left(\sum_{j} q_{ij} \right)^2 \quad (4)$$

where the first term is the cost function, applied over every pair of nodes in the transit graph, and the second term is the one-route constraint.

2.3. Mapping of the UTRP Objective Function as a QUBO Problem.

The objective function is mapped as a QUBO problem. Mathematically, a QUBO problem is defined as:

$$Obj(x,Q) = x^{\wedge}T*Q*x \quad (5)$$

where x is a vector of binary variables of length n and Q is an n¬-by-n array that encodes the problem. The goal of optimizing a QUBO problem is to find a vector of binary variables of length n that minimizes the cost of the QUBO objective function.

Quantum annealers can turn a QUBO problem into an Ising Hamiltonian that encodes the problem as an energy minimization problem. This allows both the quantum annealer to natively solve the QUBO and allows for the formatting of an optimization problem as a QUBO problem for the QA to solve.

As stated previously, the UTRP objective function for the QA is written in terms of binary variables. This allows for the reformatting of the UTRP objective function in the format of the QUBO objective function. As a reminder, the vector of binary variables x is a vector of size i * j, where i is the number of routes and j is the number of candidate routes. Each route uses the same list of candidate routes.

To convert the UTRP objective function into the QUBO objective function, all that is needed is to find the parameters for the Q matrix. This can be done explicitly by adding up the cost of the cost function and constraint function separately for Q_{ij}, or, for clarity, Q_{ab}, where a and b are the row index and column index respectively. Both a and b are equal to the length of *x* and correspond to a route number – candidate route pairing.

To explicitly add the cost function to the Q matrix, I use the following procedure:

1) If the candidate route doesn't contain both nodes *v* and *w* and a equals *b*, add the travel time and demand for nodes *v* and w to Q_{ab}
2) If the candidate route doesn't contain both nodes *v* and *w* and *a* does not equal *b*, add the travel time and demand for nodes *v* and *w* to Q_{ab}
3) Repeat steps 3 – 4 for each index *a* and *b*
4) Repeat steps 3 – 5 for each node *v* and *w*

A similar approach is taken to explicitly add the constraint to the Q matrix, which is described in the following procedure:

1) If a equals *b,* subtract λ^2 to Q_{ab}
2) If a does not equal *b*, add $2*\lambda^2$ to Q_{ab}
3) Repeat steps 1 – 2 for each index *a* and *b*

After calculating out a Q matrix for a candidate route set, optimizing it only requires inputting the matrix into D-waves hybrid quantum classical solver. In

this paper, I used the Leap Hybrid Sampler from D-Wave (D-Wave Systems 2017).

3. Comparing the QA UTRP Solver with Classical UTRP Solvers

It isn't possible to directly compare different classical UTRP solvers, as well as the QA UTRP solver, together because objective functions and constraints vary between solvers. To solve this problem, a standardized benchmark, Mandl's Swiss Road Network, as well as different performance parameters have been created in previous research on the UTRP.

Mandl's Swiss Road Network is comprised of three standardized parts: the road network graph, the travel time matrix for each pair of connected nodes, and the demand cost matrix for each pair of nodes. There are five main performance parameters:

1) d_0 -the percentage of demand satisfied without any transfers,

2) d_1 - the percentage of demand satisfied with a single transfer,

3) d_2 - the percentage of demand satisfied with two transfers,

4) d_{un} - the percentage of demand unsatisfied, and

5) Average Travel Time (ATT) - the average travel time over the entire road network in minutes per transit user.

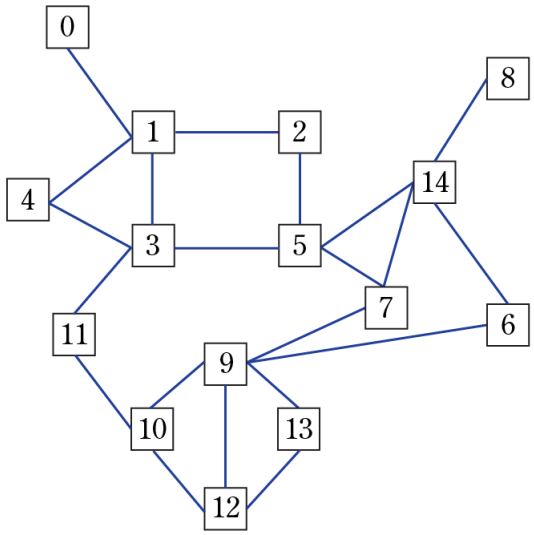


Figure 1: A graph of Mandl's Swiss Road Network benchmark with travel times for each edge labeled

To compare the QA UTRP solver to prior solutions to the UTRP, I obtained the best solution from Mandl's Swiss Road Network using the QA UTRP solver and calculated the performance parameters for the solution route. I compared my solver's best solutions to the different classical UTRP approaches. A summary of the performance parameters are in Table 1.

Ideal ATT, according to Mumford, is 10.01. The Hybrid QA Solver was 0.2% off of the ideal. Compared to the best overall ATT, found by the Differential Evolution Solver, who was at 3.5% off of the ideal ATT for Mandl's Swiss Road Network.

My algorithm was able to decrease ATT from 10.36 minutes from the current best performing classical heuristic, Differential Evolution, to 10.11 minutes. To put this in perspective, in terms of total travel time (TTT), my algorithm decreased TTT from about 161,000 minutes to 157,000 minutes. This is a decrease of 4,000 minutes of travel time for all passengers in Mandl Swiss's Road Network.

Parameters	Initial Heuristic (Fan 2008)	Particle Swarm Optimization (Kechagiopoulos et al. 2014)	Differential Evolution (Buba et al 2016)	QC Solver
d0 (%)	69.94	91.84	94.28	91.65
d1 (%)	29.93	7.64	5.72	8.34
d2 (%)	0.13	0.52	0.00	0.00
dun (%)	0.00	0.00	0.00	0.00
ATT (min)	12.90	10.64	10.36	10.11

Table 1: Comparison of my algorithm (QC Solver) with other classical heuristics

4. Applying the QA UTRP Solver to Optimize Down Seattle Bus Routes

To test the feasibility of UTRP solutions and to compare a real-world application of a quantum annealing solution, I used our algorithm to generate an optimized bus route set for bus stops in downtown Seattle.

4.1. Creating and Formatting Data for Downtown Seattle

I planned on optimizing bus routes that passed through bus stops located in a 1250-meter radius from the center of downtown Seattle, located at the coordinate point (47.608013, -122.335167). I will refer to this area as the "downtown Seattle area" from now on. All of these routes are run by King County Metro, which lists their stops and their route data on Open Transit Data (Open Transit Data).

To locate all of the bus stops in the downtown Seattle area, I created a bounding box defined with four latitude-longitude points and filtered out all latitude-longitude points. I did this by approximating the change in distance from the original coordinate point using the Earth's radius. I got the bounding box coordinates (i.e., the range of coordinates that I used to define the downtown Seattle area.

To match the problem format of the UTRP, I needed to create an undirected graph of with the downtown Seattle bus routes and bus stops, to create the travel time and demand cost matrix.

There are two steps to create the undirected graph of downtown Seattle, 1) creating an OSMnx graph (Boeing 2017) a certain radius around a center point, which I did earlier, and 2) adding in all of the bus stops as singular points.

To calculate the travel time matrix, I used OSMnx get_travel_time() method to calculate the travel time between all pairs of bus stops that were connected by the road network. They were indexed in order of their bus ID number described in the stops.txt file from OpenTransitData.

To calculate the demand cost matrix, I imported Annual Average Weekday Traffic (AAWDT) data from SeattleGeoData (Seattle GeoData) and included it in the attribute data for each street on the OSMnx graph. To approximate the AAWDT for each pair of nodes, I then calculated the shortest path between each node and added up the AAWDT for all of the road segments on that graph.

To import the existing routes, I used the data from trips.txt to approximate the bus routes already taken. This provided a map of downtown Seattle as shown below, where the purple lines are the existing bus routes in downtown Seattle.



Figure 2: Existing route set in downtown Seattle



Figure 3: Optimized route set created by quantum algorithm

4.2. Results

I calculated the ATT for both the existing route set and the optimized route set that my quantum algorithm created, with the results shown in the table below.

Parameters	ATT (min)
Existing downtown Seattle bus routes	7.08
Optimized routes using QC algorithm	3.82

Table 2: Comparison of my algorithm's optimized routes with the existing routes in downtown Seattle. My algorithm was able to decrease average travel time by 46%.

The optimized route set is shown in the figure below, with the lines in blue showing all of the roads covered by the optimized route set.

My routes were able to decrease the average travel time by 46.1%, from a travel time of 7.08 minutes to 3.82 minutes.

However, my algorithm did omit some bus stops on the outer edge of the area I tried to optimize. This shows that my algorithm, along with optimizing the travel time, "chose" certain bus stops to not include in the final route set. This finding shows that the quantum algorithm is actively deciding which bus stops are worth including in the final route set. This can help urban planners to make decisions if certain bus stops should be discontinued.

5. Discussion and Future Applications

The quantum algorithm is still limited in size in both the number of candidate routes and the number of bus stops it can include. This is due to both classical memory constraints (the computer I tested on couldn't hold an array of more than about 500 candidate routes by 50 bus routes of Q matrix data) and quantum restraints (D-Wave only has a limited number of qubits to perform computations with). This means that the algorithm is only able to optimize routes for small cities or sections of big cities. However, classical computers can be upgrades to contain more memory and there is a trend of the size of quantum computers growing quickly, so this limited capability could be fixed in a few years.

My algorithm also only focuses on optimizing existing bus routes using average travel time and demand information; it doesn't consider the number of buses on each line (frequency setting) or the existing bus schedule (timetable setting). My algorithm also doesn't develop new bus stops, which could limit the minimization of average travel time. I would like to improve my algorithm and the UTRP formulation to use demand data about areas, instead of specific bus stops, to create optimized bus routes. In other words, the demand data would contain information about hotspots in the entire city, instead of focusing on specific bus stops, which would allow for the creation of a more informed placement of bus stops.

References

Boeing, G. (2017). Osmnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–139. doi:10.1016/j.compenvurbsys.2017.05.004

Buba, A.T. and Lee, S. (2016) Differential Evolution for Urban Transit Routing Problem. Journal of Computer and Communications, 4, 11-25. http://dx.doi.org/10.4236/jcc.2016.414002

Commuting by public transportation in the United States: 2019 - Census.gov. (n.d.). Retrieved December 15, 2022, from https://www.census.gov/content/dam/Census/library/publications/2021/acs/acs-48.pdf

D-Wave Systems. (2017). D-Wave Initiates Open Quantum Software Environment. Available at: https://www.dwavesys.com/press-releases/d-wave- initiates-open-quantum-software-environment

Fan, Lang, and Mumford, Christine. A metaheuristic approach to the urban transit routing problem. Journal of Heuristics, Vol 16 (3) 2010 pp 353-372. (Also first published online in 2008). A metaheuristic approach to the urban transit routing problem | SpringerLink

Floyd, R.W., 1962. Algorithm 97: Shortest path. Communications of the ACM 5-6, 345.

Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. 2019. Quantum supremacy using a programmable superconducting processor. Nature 574, 7779 (2019), 505–510.

Gutman, D. (2018, February 12). Where's the bus? here's how Metro Transit decides whether your bus is on Time. The Seattle Times. Retrieved December 14, 2022, from https://www.seattletimes.com/seattle-news/transportation/your-metro-bus-is-on-time-more-often-than-you-may-realize-heres-why/

Kechagiopoulos, P.N. and Beligiannis, G.N. (2014) Solving the Urban Transit Routing Problem Using a Particle Swarm Optimization Based Algorithm. Applied Soft Computing, 21, 654-676. http://dx.doi.org/10.1016/j.asoc.2014.04.005

Mumford, C.L. (2013) New Heuristic and Evolutionary Operators for the Multi-Objective Urban Transit Routing Problem. 2013 IEEE Congress on Evolutionary Computation, Cancun, Mexico, 20-23 June 2013, 939-946. http://dx.doi.org/10.1109/CEC.2013.6557668

Neukart F, Compostella G, Seidel C, von Dollen D, Yarkoni S and Parney B (2017) Traffic Flow Optimization Using a Quantum Annealer. Front. ICT 4:29. doi: 10.3389/fict.2017.00029

Open Transit Data (OTD). Open Transit Data (OTD) | Developer resources | Sound Transit. (n.d.). Retrieved December 14, 2022, from https://www.soundtransit.org/help-contacts/business-information/open-transit-data-otd

Public transportation's role in responding to climate change. (n.d.). Retrieved December 15, 2022, from https://www.transit.dot.gov/sites/fta.dot.gov/files/doc/PublicTransportationsRoleInRespondingToClimateChange2010.pdf

Seattle geodata. Seattle GeoData. (n.d.). Retrieved April 30, 2023, from https://data-seattlecitygis.opendata.arcgis.com/

DHEBNet – a Novel Deep Heterogenous Ensemble-Based Learning Model for Deepfake Detection

John Karaca

Tesla STEM High School, Redmond, WA, USA

ABSTRACT

Over the last two decades, the development of deepfakes – a form of hyper-realistic synthetic content generated by artificial intelligence that visually manipulates the content's human subject(s) – has posed a significant threat to the notion of authenticity, credibility, and truth in the modern world. However, despite efforts by the digital forensics community to address the situation, there has been limited research regarding the effectiveness of heterogeneous ensemble learning – a category of machine learning that combines various models (base learners) in producing an output – for accurate and robust deepfake detection. This research project proposes DHEBNet: a novel deep heterogenous ensemble-based learning model that accurately and robustly detects deepfake video content (>90%) by aggregating the predictions of spatial (i.e., methods that analyze frames independently) and temporal (i.e., methods that analyze frames comparatively) deep learning models via a deep neural network (DNN) meta learner. DHEBNet base learners were trained on a diverse pre-processed set of over 1,650 portrait videos – each containing one or more human subjects and categorized as "real" or "fake" – sampled from the FaceForensics++, Facebook DFDC, Google DFD, and Celeb-DF deepfake datasets. Evaluations with a holdout dataset revealed that DHEBNet achieved an accuracy of 95.75% and an F1-Score of 97.49%, representing improvements over the equivalent average base learner performances by 1.80% and 1.09%, respectively. DHEBNet satisfies initial performance requisites, demonstrates the effectiveness of combining models to capture nuances in synthetic content, and provides a comprehensive solution to the evolving problem of deepfakes.

1 INTRODUCTION

Thesis: This project seeks to develop a novel deep heterogenous ensemble-based learning model that aggregates the predictive capabilities of various spatial and temporal deep learning models to robustly discover inconsistencies in video data containing a human subject(s) and reliably detect deepfake content.

Over the last two decades, the development of deepfakes – a form of synthetic media generated by artificial intelligence (AI) that visually manipulates the likeness of a person – has challenged the way human beings fundamentally interpret reality. By merit of advancements in generative algorithms, synthetic content produced by deepfake technology has become increasingly hyper-realistic. The indistinguishability of deepfakes from the real world is unarguable: Korshunov and Marcel observed that individuals, when asked to classify potentially unauthentic content involving one or more human subjects as "real" or "fake", were only able to differentiate state-of-the-art synthetic videos at 24.5% of the time in contrast to an accuracy of 82.2% with authentic videos – an unsettling figure testifying to the capacity of modern deepfakes to reliably deceive human perception (Korshunov and Marcel, 2020). It is further troublesome that over 49,000 deepfake videos were detected to be in circulation online in 2020 – a figure found to be doubling every six months over the two years prior (Patrini, 2019). Combined with the digitally driven nature of information dissemination, the broad availability of computational resources, and

the democratization of AI systems, deepfakes have spawned a growing unease as to the continuity of the concept of the "objective truth". The integrity of information in the 21st century is in question: with deepfakes, seeing is no longer believing.

First coined in 2017 by a user of the same name on the social news website Reddit, the term "deepfake" is a composite of the words "deep learning" – a machine learning approach that imitates the human brain by employing artificial deep neural networks (DNNs) – and "fake", in reference to the unauthentic, albeit believable, nature of synthetic content. Deepfakes function by superimposing a source human identity (x_s) onto the image or video of a target human identity (x_t) through various methods of content manipulation, namely reenactment (where the facial expression(s) of x_t is manipulated by x_s) and replacement (where the face of x_t is completely replaced with x_s) (Figure 1). Since its coining, deepfakes have notoriously made international headlines in their application. For instance, in January of 2019, a supposed deepfake video of Ali Bongo – a Gabonese president who suspiciously hadn't been publicly seen for months due to medical treatment – helped instigate an attempted military coup in the Central African nation of Gabon, plunging the country into temporary violence and volatility with reported casualties.

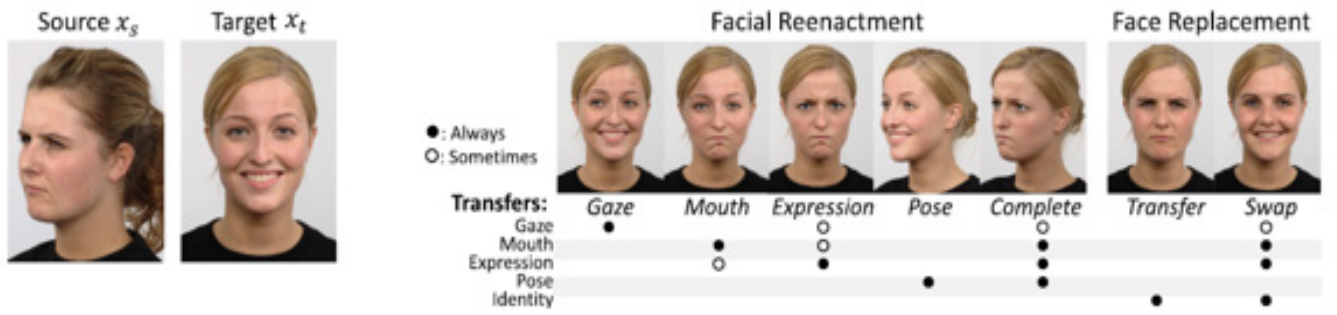


Figure 1: Deepfake Illustration. Examples of facial reenactment and replacement in generating deepfakes between xs and xt. Adapted from "The Creation and Detection of Deepfakes: A Survey" by Y. Mirsky and W. Lee, 2020, ACM Computing Surveys (CSUR), 54(1), p. 3.

However, the groundwork of the technology was laid two decades prior, when Bregler et. al., in 1997, introduced the novel "Video Rewrite" program that enabled facial reanimation (Bregler et. al., 1997). By extrapolating existing video footage through computer vision techniques that (1.) tracked the landmark points on the mouth of a video's subject and (2.) connected the sounds produced by such subject's voice to the extracted mouth landmarks, the software was capable of synthesizing video footage of the same subject enunciating the words of any audio track that themselves may have never spoken in. Despite its novelty, the generated content regularly contained lip movement asynchronies, unnatural articulation of speech, inconsistencies in facial illumination, warping artifacts involving the background environment and the subject, and registration errors in spatial mapping concerning facial landmarks, namely the mouth.

In 2014, Goodfellow et. al. improved on these imperfections by proposing the revolutionary Generative Adversarial Network (GAN): a deep learning-based model that learns from input data in such a way as to generate new input data that could plausibly be interpreted as authentic. To achieve this, GANs are composed of two sub-models: a generative model *G* that captures input data and tries to replicate it, and a discriminate model *D* that tries to successfully classify the output of *G* as "real" or "fake". These models compete to outperform one another until the capacity of *G* to synthesize new input surpasses the predictive capabilities of *D*, thereby indicating the capacity of the GAN to generate believable content (Goodfellow et. al., 2014). In the years following the breakthrough, GAN technology has been improved on in the field of computer vision through the invention of the Style-Based Generator Architecture (StyleGAN) which was primarily designed to facilitate image synthesis. StyleGAN has since been adapted for the task of the high-quality generation of human

faces with remarkable, albeit disconcerting, results.[1] Ultimately, GANs have significantly reduced the artifacts and inconsistencies originally raised with the "Video Rewrite" program and have, therefore, underpinned the trend of deepfakes towards hyperrealism and ever-increasing believability.

The scope, scale, and sophistication of modern generative algorithms have allowed adversaries to increasingly leverage deepfakes to undermine the human world at large. For one, deepfakes threaten the journalism industry, where the proliferation of fake news could hamper consumer trust in general media. The technology can be used as two equally destructive means of attack towards the press: to (a.) spread high-quality, believable unauthentic media whose news story is repercussive or (b.) disseminate low-quality, illegitimate media that compromises the reputation of journalism.

Likewise, deepfakes imperil democracy – a democratic system of government is founded on truth, and the spread of false information has the potential of distorting political discourse and jeopardizing the electorate. If the public were to believe that everything they consume – whether by reading, hearing, or seeing – could have plausibly been faked, polarization would dominate the political landscape at the expense of efficient governance. For instance, in 2018, American internet media company BuzzFeed illustrated how public figures and politicians, through a deepfake of former president Barack Obama, could be easily impersonated and manipulated to spread misinformation via deepfakes.

Additionally, deepfakes endanger domestic and global financial systems, as malicious actors could leverage synthetic media to extort individuals, defraud companies, and inflict harm on markets. For instance, organizations facing internal crises could be damaged by deepfakes seeking to amplify a circulating narrative, and businesses could face financial losses from deepfake voice phishing. Notably, in March of 2019, cybercriminals used deepfakes to impersonate the voice of a Chief Executive Officer of an unidentified German company and to defraud the company's energy subsidiary of €220,000 (~$237,000). A little less than two years later, another set of cybercriminals would steal $35,000,000 in funds from an unidentified international bank by assuming the synthetic identity of an executive at the bank's partnering company.

Furthermore, deepfakes threaten national and international security, as the technology could be used to facilitate espionage and to spread misinformation that could foster domestic and global instability. For instance, over the last several years, the United States has identified a plethora of online profiles on the networking site LinkedIn that list significant connections with the federal government – including positions with members of Congress and jobs within major political think tanks – despite harboring GAN-generated profile pictures. Undoubtedly, global adversaries are employing synthetic content in an effort to bolster their digital espionage initiatives.

In light of the technology's advancements in hyper-realistic content generation as well as its implications, deepfakes have become a priority for the digital forensics community and have led to the proposal of various methods of detection. Initial research exploited the spatial flaws inherent to early generative models (e.g., the first iterations of StyleGAN) through frame-based approaches – Convolutional Neural Networks (CNNs), namely – that leveraged visual artifacts and occlusions left by the image synthesis process, including the distortion of facial attributes and the abnormal distribution of light across the face. Other methods explicitly captured the temporal information of deepfake content in order to identify sequential discrepancies, such as the jittering and flickering of faces, via recurrent methods – including Recurrent Convolutional Neural Networks (RCNNs) and three-dimensional Convolutional Neural Networks (3D CNNs) – that weighed the extent of difference between frames. Despite their successes, recent advancements in generative algorithms (i.e., StyleGAN2 and StyleGAN3) have revealed the vulnerability of early detection methods to new more-convincing distributions of deepfakes, as they inherently assume a static detection environment composed of specific leverageable visual artifacts. For instance, between three benchmark deepfake datasets – Deepfake Detection by Google, Deepfake Detection Challenge by Facebook, and Celeb-DF – released consecutively of one another

[1] https://thispersondoesnotexist.com/

with improvements, spatiotemporal methods of detection attained decreasing AUC scores of 0.86, 0.76, and 0.66, respectively – metrics that show that advancements in generative techniques within a small period of time can significantly damage the performance of spatial and temporal forensic approaches (Mirsky and Lee, 2020).

Despite their pitfalls, all researched methods of detection appear to have their respective strengths and weaknesses against varying techniques of content synthesis. For instance, while spatial methods of detection may yield greater performance on content with distinct frame-based artifacts, they may neglect unrelated temporal artifacts that would have been easily detected by temporal methods. For that reason, researchers have begun studying the potential use of ensemble-based machine learning in deepfake detection by aggregating the predictive capabilities of various forensic approaches. Although recent ensemble techniques have outperformed state-of-the-art detection methods, they remain largely homogenous – curating various models (base learners) of the same type (e.g., exclusively spatial), thereby restricting the scope of their forensic analysis – and are limited in scale.

In this paper, heterogenous ensemble-based learning – a category of machine learning that combines base learners of different types in producing an output – is alternatively proposed as a superior method of digital forensics involving synthetic content. Recognizing the advantages of each aforementioned category of detection, this engineering project seek to develop a novel deep heterogenous ensemble-based learning model (DHEBNet) that aggregates the predictive capabilities of various spatial (i.e., methods that analyze frames independently) and temporal (i.e., methods that analyze frames comparatively) deep learning models via a deep neural network (DNN) meta-learner to robustly discover inconsistencies in video data and reliably detect deepfakes at a high (>90.0%) degree of accuracy.

2 METHODOLOGY

In this section, the collection and processing of data, the architecture of DHEBNet, and the training process of DHEBNet is outlined.

2.1 Data Collection

In order to maximize model performance and generalizability with sufficient and unbiased data, video data was sourced from four independent, public datasets: (1.) FaceForensics++, containing 1,000 authentic videos extracted from YouTube and 5,000 corresponding synthetic videos (Rossler et. al., 2019); (2.) Google Deepfake Detection (DFD), containing 363 crowdsourced authentic videos and 3,065 corresponding synthetic videos (Dufour and Gully, 2019); (3.) Facebook Deepfake Detection Challenge (DFDC) Preview, containing 1,140 crowdsourced authentic videos and 3,065 corresponding synthetic videos (Dolhansky et. al., 2019); and (4.) Celeb-DF, containing 590 authentic videos extracted from YouTube and 5,639 corresponding synthetic videos (Li et. al., 2020).

These four deepfake datasets were compiled into a central dataset that spans over 21,000 portrait videos, with each video containing one or more human subjects and labeled as one of two distinctive groups: *real*, referring to videos that haven't been modified and are authentic; and *fake*, referring to videos that have been modified by a generative algorithm(s). Cumulatively, these portrait videos evenly represent 9 types of deepfake generative techniques (i.e., Celeb-synthesis, DFDC Method A, Deepfake Detection, Face2Face, FaceShifter, Neural Textures, FaceSwap, Deepfakes, and DFDC Method B), 4 classes of video resolutions (i.e., VGA, FHD, HD, and QVGA), 6 different ethnic groups (i.e., Caucasian, Asian, Latino Hispanic, Middle Eastern, African American, and Indian), and all genders (i.e., male and female) (Figure 2). As such, the compiled central dataset contains both a substantial and diverse range of portrait video data for model training, testing, and evaluation purposes.

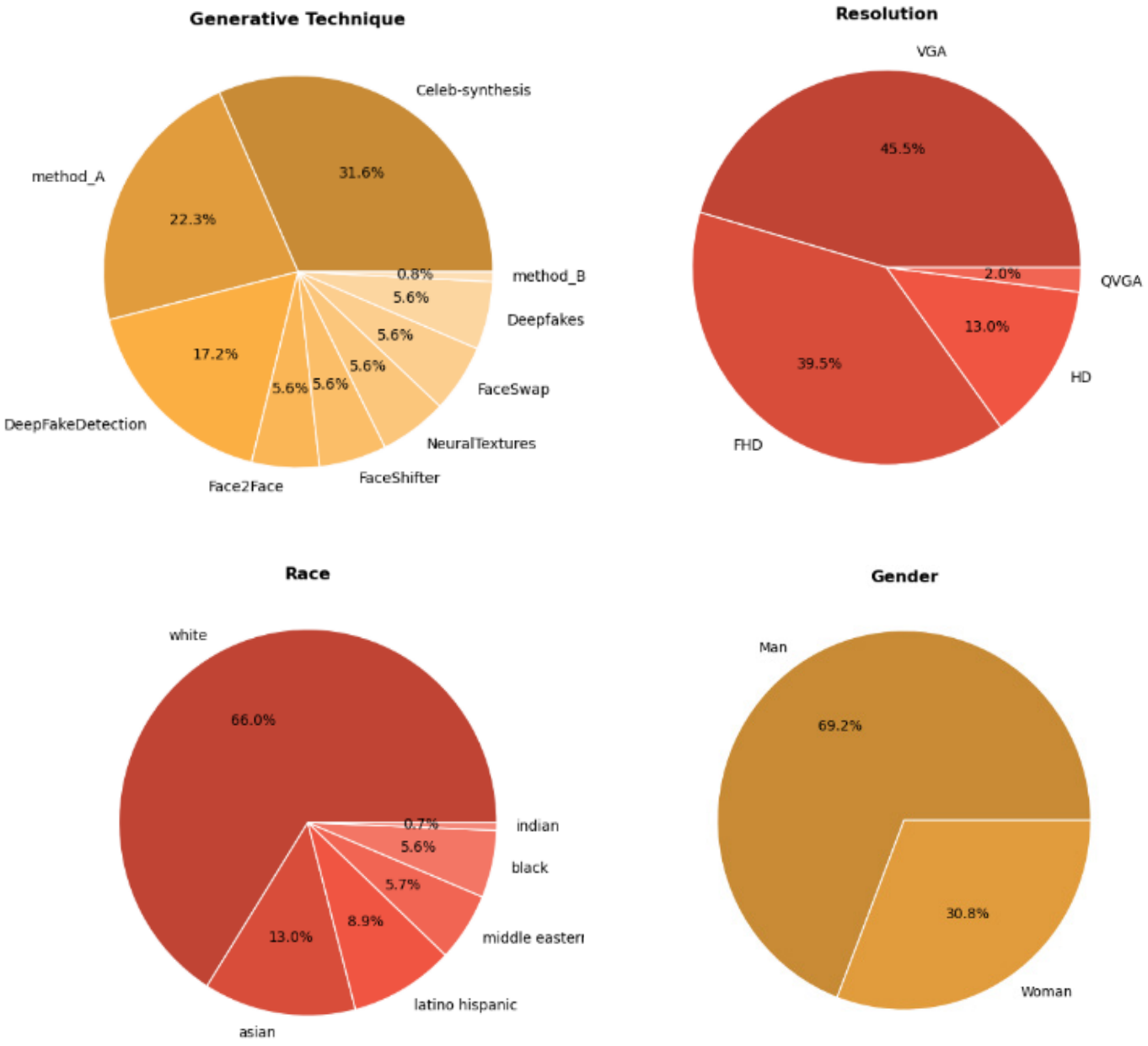


Figure 2: Data Analysis. Analyzing the distributions of generative techniques and resolutions in, as well as the races and genders represented by, the central dataset.

2.2 Data Processing

In the vast majority of deepfakes, visual artifacts and irregularities are predominantly concentrated in the facial area, as generative techniques that synthesize such content most directly manipulate the subject's face. Additionally, the video data within the central dataset varied greatly in resolution and length, and needed to be appropriately resized to a standard, fixed format for model training and evaluation. As such, before splitting and sampling the central dataset, each of its portrait videos was processed through the identification, isolation, resizing, and resequencing of the facial regions of the video's human subject(s) from a select number of frames (Figure 3).

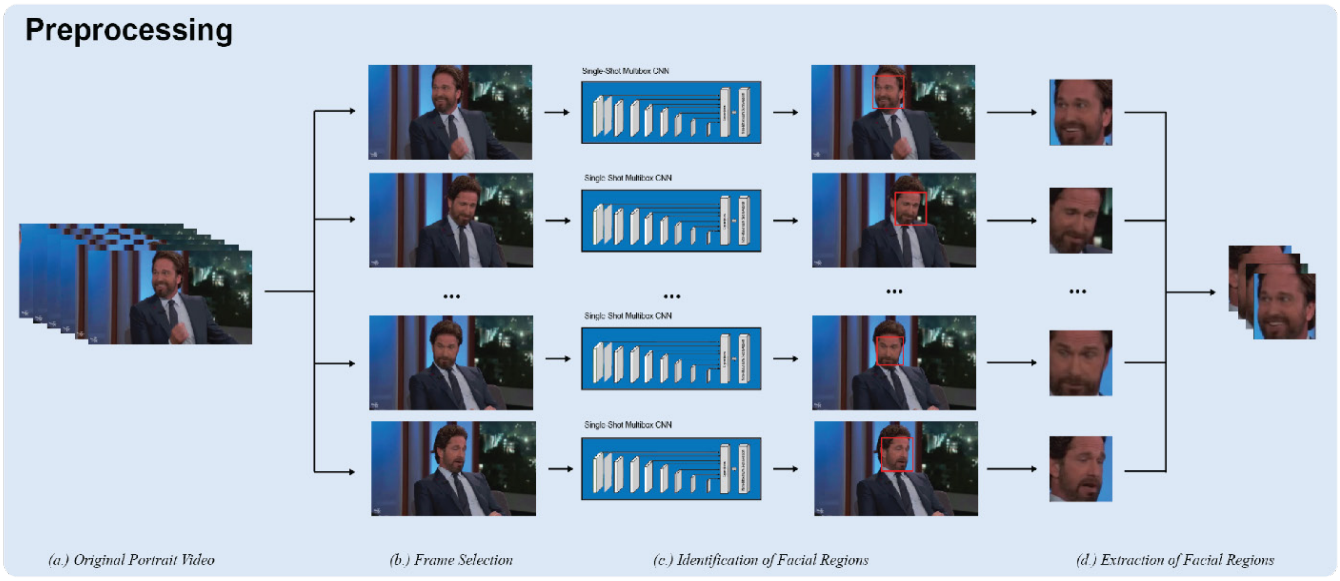


Figure 3: Data Processing Pipeline. Identification, isolation, and sequencing of the facial regions from the selected frames of any given portrait video.

To accomplish this, 5 different face detection algorithms were manually evaluated: Haar Cascade (Viola & Jones, 2001); Histogram of Oriented Gradients (HOG) and Linear Search Vector Machine (SVM) (Dalal & Triggs, 2005); Max-Margin (MMOD) CNN (King, 2015); Multi-Task Cascaded (CT) CNN (Zhang et. al., 2016); and the Single-Shot Multi-Box (SSD) CNN (Liu et. al., 2016). Ultimately, the SSD CNN was chosen as the most viable candidate for the task of video processing, having a low number of false positives, accurate facial bounding boxes, and temporal coherence. The remaining methods lacked these qualities, having been either prone to false positives (i.e., Haar Cascade, MT CNN) or temporal incoherence (i.e., HOG + Linear SVM, MMOD CNN).

The SSD CNN was then employed to appropriately process each portrait video of the central dataset through the identification, isolation, resizing, and resequencing of the facial regions of the video's human subject(s) to produce new videos consisting of 24 frames of a 224x224 resolution. 80% of the central dataset was then reserved for training, with another 20% reserved for validation and testing. Subsequently, a 10% sample (2,100+ portrait videos) of the central dataset was procured due to underlying computational limitations.

2.3 Model

DHEBNet is a deep heterogenous ensemble-based learning model that classifies portrait video content by learning how to best interpret (1.) spatial (i.e., methods that analyze frames independently) and (2.) temporal (i.e., methods that analyze frames comparatively) deep learning models via a deep neural network (DNN) meta-learner (Figure 4)

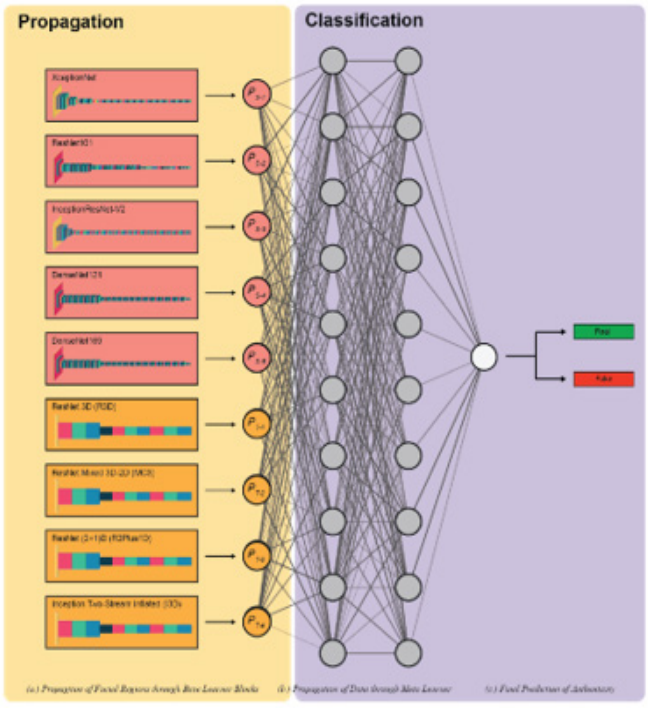


Figure 4: DHEBNet Architecture. Structure of DHEBNet (a. – c.).

2.3.1 Spatial Block

The Spatial Block of DHEBNet – the first of the model types within the ensemble architecture – considers the spatial attributes of input video data and detects inconsistencies within portrait video frames. For any given portrait video, each spatial base learner makes an n number of predictions of authenticity for an n number of frames – classifying the frames individually, allowing it to capture spatial, frame-based artifacts – before averaging the final prediction via a fusion layer (Figure 5).

The block consists of five corresponding state-of-the-art, two-dimension image recognition CNNs specialized in spatial content classification: the (1.) XceptionNet CNN (Chollet, 2016); (2.) ResNet101 CNN (He et. al., 2015); (3.) InceptionResNetV2 CNN (Szegedy et. al., 2016); (4.) DenseNet121 CNN (Huang et. al., 2018); and (5.) DenseNet169 CNN (Huang et. al., 2018). To save resources and improve performance, the transfer learning technique was employed, with each spatial base learner pre-trained on the ImageNet dataset (Russakovsky et. al., 2015).

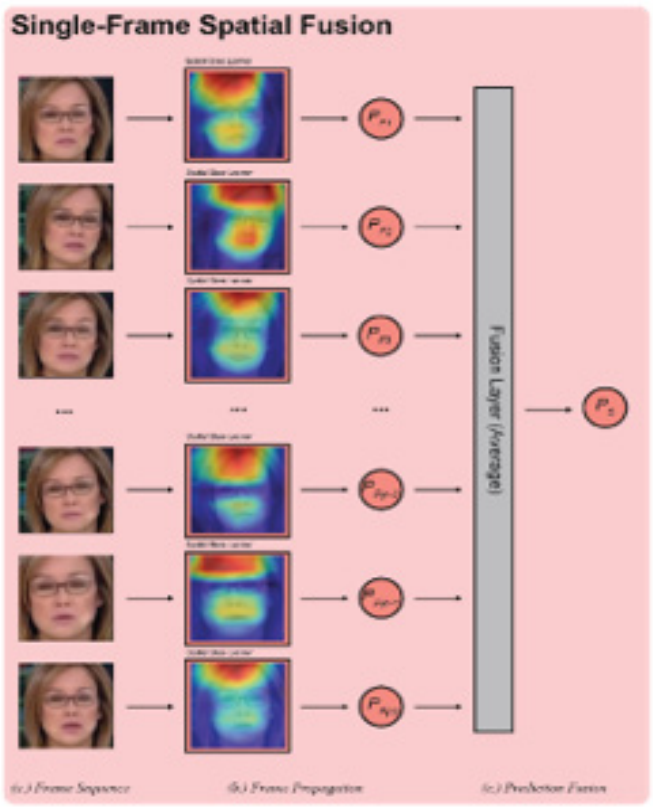


Figure 5: Single-Frame Spatial Fusion. Spatial base learners make individual predictions for each frame of a given portrait video before averaging the predictions together via a fusion layer.

2.3.2 Temporal Block

The Temporal Block of DHEBNet – the second of the model types within the ensemble architecture – considers the temporal attributes of input video data and detects inconsistencies between portrait video frames. In contrast to the Spatial Block, for any given portrait video, each temporal base learner makes a single prediction of authenticity, allowing it to effectively capture temporal artifacts (Figure 6).

The block consists of four corresponding state-of-the-art, three-dimension video recognition CNNs specialized in temporal content classification: the (1.) ResNet 3D CNN (R3D) (Hara et. al., 2017); ResNet Mixed 3D-2D CNN (MC3) (Tran et. al., 2018); ResNet (2+1)D CNN (R2Plus1D) (Tran et. al., 2018); and Inception-V1 Two-Stream Inflated CNN (I3D) (Carreira & Zisserman, 2017). Similar to the Spatial Block, transfer learning was employed, with each temporal base learner pre-trained on the Kinetics 400 dataset (Kay et. al., 2017).

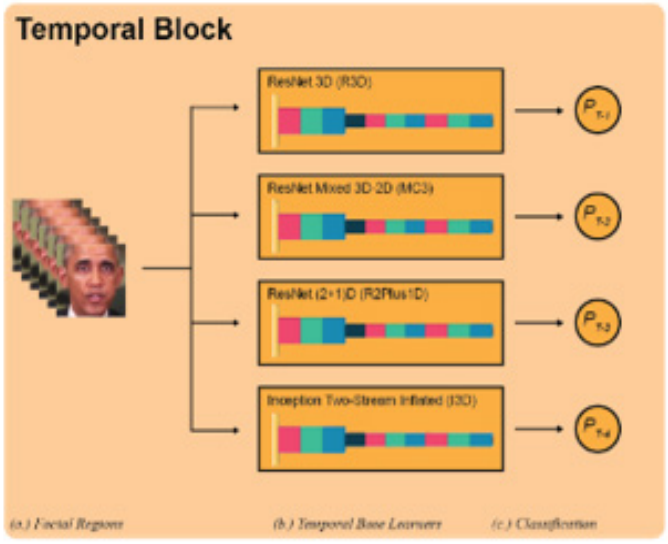


Figure 6: Temporal Classification. Temporal base learners capture temporal, inter-frame inconsistencies by directly making predictions on a given portrait video.

2.4 Training

Each base learner of DHEBNet – independent of which block it resided in – was trained and validated with the training and validation subsets, respectively, of the pre-processed central dataset

sample. More specifically, the spatial base learners utilized an image subset created through the extraction of portrait video frames (49,000+ images), whereas the temporal base learners directly utilized the portrait videos comprising the dataset.

Subsequently, a train-validation loop was created and used to evaluate the broad performance and generalizability of each model in order to monitor the presence of over- and under-fitting. Within the loop, a shallow grid search tuning scheme was employed in order to identify and select the optimal hyperparameters for each base learner. These hyperparameters included, but were not limited to, the number of epochs (i.e., 5, 10, 25), dropout probability (i.e., 0.3, 0.5, 0.8), learning rate (i.e., 1e-3, 1e-4, 1e-5), batch size (i.e., 1, 16, 32), optimization algorithm (i.e., Adam, Stochastic Gradient Descent), and the presence of early stopping.

Once the optimal hyperparameter configuration of each base learner was identified and saved, the DNN meta-learner of DHEBNet was trained and evaluated. Using multiple layers – constituting over 400 trainable parameters – to sequentially extract higher-levels features from base learner classifications, DHEBNet's meta-learning model learns how to best interpret and combine predictions in producing a final prediction of content authenticity.

Additionally, performance metrics were actively calculated and collected following each epoch of each base learner's training process. In specific, the statistical tests of accuracy, precision, recall, and F1-Score were utilized.

3 RESULTS

Once completely trained and optimized, DHEBNet and its base learners were tested using a holdout subset (200+ portrait videos) of the pre-processed central dataset sample. Evaluations revealed that DHEBNet achieved an unprecedented accuracy of 95.75%, a precision of 96.69%, a recall of 98.31%, and an F1-score of 97.49%. Further analysis showed that DHEBNet outperformed its base learners by a significant margin, achieving a 1.80% and 1.09% improvement over the equivalent average base learner accuracy and F1-score, respectively.

Furthermore, DHEBNet achieved a high level of generalizability relative to the data categories – generative technique, resolution, ethnic group, and gender – represented by the portrait videos it was evaluated on. For instance, while the spatial and temporal base learners, on average, achieved an accuracy of 95.34% and 93.86% on portrait videos with an FHD resolution, respectively, DHEBNet exceeded with an equivalent accuracy of 97.26%. Additional performance comparisons can be found in Figure 9, as well as a bar chart visualization in Figure 8.

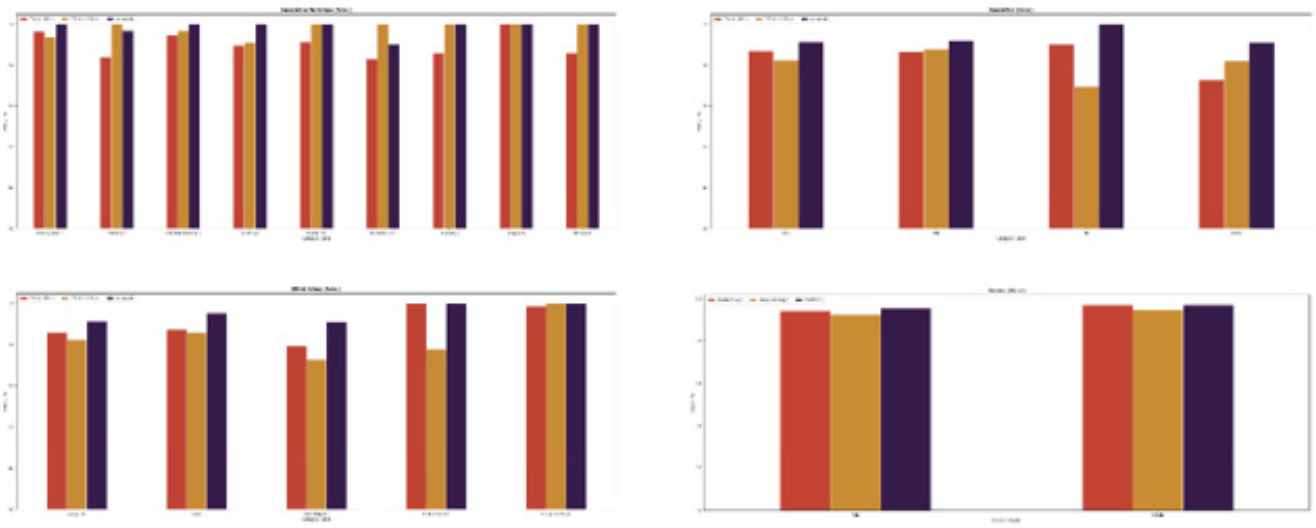


Figure 8: Performance Comparison – Bar Graph. Bar graph of accuracies attained by the spatial (red) and temporal (yellow) base learners, on average, as well as that of DHEBNet (purple) for the generative technique, resolution, ethnic group, and gender data

Data Category	DHEBNet Accu.	Avg. Spatial Accu.	Avg. Temporal Accu.
Generative Method			
Celeb-synthesis	98.39%	**98.71%**	97.97%
Method A	**96.88%**	96.25%	97.68%
DeepFake Detection	**100.00%**	99.35%	93.46%
Face2Face	90.91%	**94.55%**	87.53%
FaceShifter	**100.00%**	100.00%	96.70%
Neural Textures	**100.00%**	98.00%	89.71%
FaceSwap	**100.00%**	96.67%	93.57%
Deepfakes	**100.00%**	100.00%	97.14%
Method B	**100.00%**	100.00%	100.00%
Resolution			
VGA	**94.12%**	94.12%	91.60%
FHD	**97.26%**	95.34%	93.86%
HD	**96.15%**	95.38%	93.85%
QVGA	**100.00%**	98.18%	98.44%
Ethnic Group			
Caucasian	**95.97%**	94.63%	93.10%
Asian	**100.00%**	100.00%	98.64%
Latino Hispanic	86.36%	**87.27%**	86.49%
Middle Eastern	**100.00%**	100.00%	100.00%
African American	**100.00%**	100.00%	88.31%
Gender			
Male	**95.24%**	94.01%	92.34%
Female	**96.92%**	96.92%	94.51%

Figure 9: Performance Comparison – Table. Table of accuracies attained by the spatial (red) and temporal (yellow) base learners, on average, as well as that of DHEBNet (purple) for the generative technique, resolution, ethnic group, and gender data categories.

4 CONCLUSIONS

As deepfakes continue to evolve and proliferate, designing accurate, reliable, and robust detection systems will be essential in preserving global truth, authenticity, and credibility. In this paper, DHEBNet – a deep heterogenous ensemble-based learning model specialized in deepfake detection – is proposed as a novel solution to the growing threat of synthetic video content. By leveraging a DNN meta-learner, DHEBNet learns how to best combine the predictions of distinct spatial and temporal deep learning models in producing precise predictions of authenticity. When evaluated on a diverse slate of real and fake video content, DHEBNet attained an accuracy of 95.75% and an F1-Score of 97.49% – performance metrics that exceeded the equivalent metrics achieved by its single-model base learners. Additionally, DHEBNet likewise exceeded its initial performance requisite of an accuracy of 90% or higher.

Furthermore, DHEBNet significantly outperformed existing state-of-the-art deepfake detection models. When compared with recent research publications that proposed single-model spatial and temporal techniques in synthetic content detection on a variety of deepfake datasets – FaceForensics++, Facebook DFDC, Google DFD, and Celeb-DF, namely – DHEBNet outperformed in accuracy across all scenarios (Figure 10). These unprecedented successes exemplify the capacity of deep heterogenous learning in capturing a broader and richer range of visual artifacts and inconsistencies – a capability that directly translates to improved performance and robustness, as initially hypothesized with DHEBNet.

Methods	FaceForensics++	DFDC	DFD	Celeb-DF
Spatial Methods				
AMTEN [Guo et. al., 2020]	**84.7%**	-	-	-
Face & Context Discrepancy [Nirkin et. al., 2020]	75.0%	-	-	-
Temporal Methods				
Optical Flow CNN [Amerini et. al., 2019]	81.6%	-	-	-
Multi-Task Learning [Nguyen et. al., 2019]	83.7%	-	-	-
CViT [Wodajo & Atnafu, 2021]	71.8%	**91.5%**	-	-
TD-3D CNN [Zhang et. al., 2021]	-	80.0%	-	88.8%
Optical Flow CNN-LSTM [Saikia, et. al., 2022]	**91.2%**	66.3%	-	79.5%
DHEBNet	**96.5%**	**93.2%**	**99.8%**	**94.6%**

Figure 10: Model Comparison. Performance of current detection models with that of DHEBNet. In all instances, DHEBNet exceeds these state-of-the-art deepfake detectors.

Future iterations of, and forthcoming research into, DHEBNet will explore and present solutions to two distinct limitations.

First is the limited size of, and imbalances within, the central dataset. Despite its size, only a slice of the video data – a 10% sample having been procured – was utilized during the training and evaluation process of DHEBNet and its base learners. Furthermore, the dataset contained biases towards some categories of video data, most prominently in favor of Caucasian and Male human subjects. The utilization of comprehensive cloud computing technologies and the incorporation of new deepfake datasets – like DeePhy (Narayan et. al., 2022) and KoDF (Kwon et. al., 2021) – is presented as an approach of better capturing, expanding, and diversifying the central dataset.

Second is DHEBNet's exclusive focus on spatial and temporal features in producing predictions of authenticity on video content. Despite the model's successes, spatial and temporal features are only two of several leverageable attributes in portrait video containing a human subject(s). For instance, biological features – attributes that are exclusive to a video subject's facial physiology, like blinking (Li et. al., 2018) and facial blood oscillation (Ciftci et. al., 2020) – may offer new lenses of deepfake detection. As such, future development of DHEBNet could complement its existing spatial and temporal capabilities with distinct biological base learners.

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The State of Deepfakes: Landscape, Threats, and Impact.

Amerini, I., Galteri, L., Caldelli, R., Bimbo, A.D. (2019). Deepfake Video Detection through Optical Flow Based CNN. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 1205-1207).

Bregler, C., Covell, M., Slaney, M. (1997). Video Rewrite: Driving Visual Speech with Audio. In 1997 Proceedings of the 24th annual conference on Computer graphics and interactive techniques. ACM.

Carreira, J., & Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6299-6308).

Chollet, F. (2017). Xception: Deep learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Rec1ognition (pp. 1251-1258).

Ciftci, U. A., Demir, I., & Yin, L. (2020). Fakecatcher: Detection of Synthetic Portrait Videos using Biological Signals. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 886-893).

Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2019). The Deepfake Detection Challenge (DFDC) Preview Dataset. arXiv preprint arXiv:1910.08854.

Dufour, N., & Gully, A. (2019, September 24). Contributing Data to Deepfake Detection Research [web log]. From https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144

Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. Computer Vision and Image Understanding, 204, 103170.

Hara, K., Kataoka, H., & Satoh, Y. (2017). Learning Spatio-temporal Features with 3D Residual Networks for Action Recognition. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 3154-3160).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4700-4708).

Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., ... & Zisserman, A. (2017). The kinetics human action video dataset. arXiv preprint arXiv:1705.06950.

King, D. E. (2015). Max-Margin Object Detection. arXiv preprint arXiv:1502.00046.

Korshunov, P., & Marcel, S. (2020). Deepfake Detection: Humans vs. Machines. arXiv preprint arXiv:2009.03155.

Kwon, P., You, J., Nam, G., Park, S., & Chae, G. (2021). Kodf: A large-scale korean deepfake detection dataset. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10744-10753).

Li, Y., Chang, M. C., & Lyu, S. (2018, December). In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE.

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A Large-scale Challenging Dataset for Deepfake Forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3207-3216).

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot Multibox Detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. ACM Comput-

ing Surveys (CSUR), 54(1), 1-41.

Narayan, K., Agarwal, H., Thakral, K., Mittal, S., Vatsa, M., & Singh, R. (2022, October). DeePhy: On Deepfake Phylogeny. In 2022 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-10). IEEE.

Nguyen, H. H., Fang, F., Yamagishi, J., & Echizen, I. (2019, September). Multi-task learning for detecting and segmenting manipulated facial images and videos. In 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS) (pp. 1-8). IEEE.

Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2020). Deepfake detection based on the discrepancy between the face and its context. arXiv preprint arXiv:2008.12262.

Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to Detect Manipulated Facial Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1-11).

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115, 211-252.

Saikia, P., Dholaria, D., Yadav, P., Patel, V., & Roy, M. (2022, July). A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-V4, Inception-ResNet and the Impact of Residual Connections on Learning. In Thirty-first AAAI Conference on Artificial Intelligence.

Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., & Paluri, M. (2018). A Closer Look at Spatiotemporal Convolutions for Action Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6450-6459).

Viola P., & Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. I-I).

Wodajo, D., & Atnafu, S. (2021). Deepfake video detection using convolutional vision transformer. arXiv preprint arXiv:2102.11126.

Zhang, D., Li, C., Lin, F., Zeng, D., & Ge, S. (2021, August). Detecting Deepfake Videos with Temporal Dropout 3DCNN. In IJCAI (pp. 1288-1294).

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks. IEEE signal processing letters, 23(10), 1499-1503.

Finding The Source: A School Shooting Prediction Program

Aiden Morrella

Tesla STEM High School

Abstract

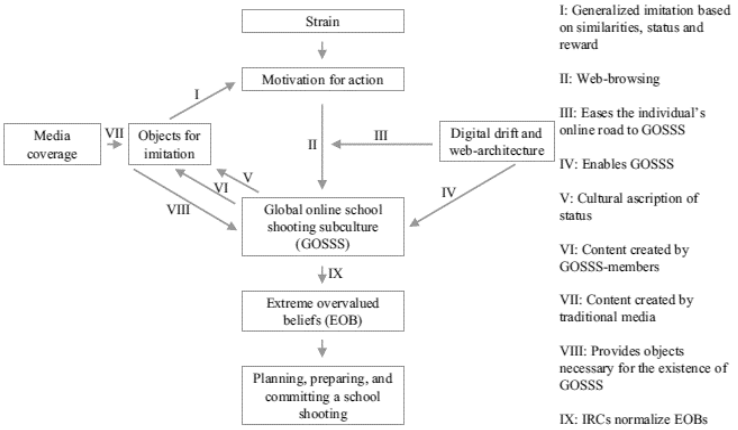
School Shootings have injured over a thousand students, and there has been few impactful efforts that have been put in place to prevent these tragedies from happening. The hypothesis of this project is that a program can used to identify a student's risk of wanting to commit a school shooting based off of only previous research and statistical data so that the students can get the help they need. Using literature and statistics, variables were created such as: Affiliation with being a part of the superior race, a dislike for social norms, fascination with previous shooters, and traumatic childhood experiences. These variables were then implemented into a questionnaire, as it is an easy way to collect the desired information, that uses a"0-5" scoring system that inevitably exports its results into a python program, where the program weights each question's answer and add it up giving an overall score that is between 0-100, though the thing I was testing was changing that total score and seeing the outcome, with that number being the independent variable in this experiment. Some findings that came from this were that Model 1 was significant to 99.967% in comparison to 2022's school shooting statistics. This is important as it means that this questionnaire and program could hypothetically help determine a student at a possible risk for wanting to perform acts of gun violence, and thanks to a scaling system.

Introduction

This project seeks to find and investigate similarities and differences in behavior from the average student and the shooters. School shootings have been a problem in the United States for the past half of a century, this problem has caused too much torment and needs to be stopped. There have been 118 school shootings in 2022 alone and a whopping 1,705 shootings in the United States since 1970. These attacks on our schools have become a problem since the turn of century and have ravaged our school systems with panic. With the United States not implementing crucial gun regulations, such as background checks and psychological tests, the only other way to solve this continuing issue is to find out who actually pulls the trigger before it is pulled. This research is beneficial in the scientific community as it is an attempt at laying a foundation in this new field of psychology. Without an idea of the true identity of the problem, there is no way to find a stop to it. Students are scared to go to school and despite efforts to make the casualties lessen, such as increasing security or implementing mental health classes, things are continuing to get worse. This leads to the big question: Can a program determine a student's threat of potentially becoming a school shooter?

This project seeks to find and investigate similarities and differences in behavior from the average student and the shooters/ Global Online School Shooting Subculture agents (GOSSS). This project used a set of variables that were derived from literature review and data analyses. These variables will then be used to create questions that will spur answers that align with the variable or not. The code was be written in python and will consist of a string of algorithms, databases, a questionnaire, and libraries that are both self-written or from open-

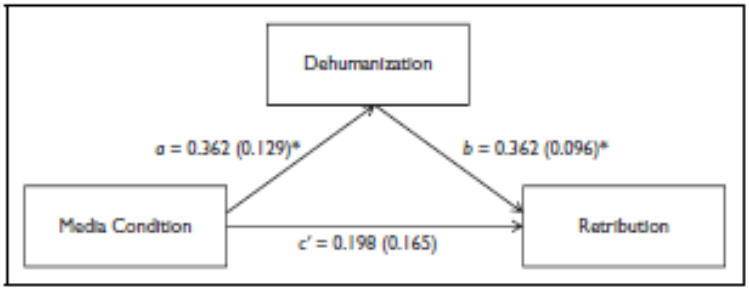
source sites. Research shows that when previous shooters cried for help, it was usually so subtle until it was too late where the person either made a threat or committed the heinous act, which brings the desired goal of this project: the program is to specify who needs help.



When it comes to the beginning stages of finding/ creating variables, research papers supply the best knowledge to be used and applied. Jonatan Werner and Martin Diederichsen have found that psychological problems, such as depression, narcissistic personality traits, lack of empathy, and even racial supremacy beliefs, play a major role in school shootings. They came up with the theory of how shooters come to be and classified these individuals as GOSSS agents (agents used as a general term for a group). This information helps in determining which factors make a shooter. They found two main factors: strain and imitation. Strain would be considered as stresses that are caused by trauma or neurodevelopmental problems, while imitation would be considered as the manipulation of personality to become or be like another person. These GOSSS agents tend to be people who idolize school shooters and often imitate the behaviors, and they inevitably are the individuals' making threats of shooting up school or committing the act.

Joshua D. Freilich from the Department of Justice did an investigation to find obvious factors that could have led to the formation of previous shooters. The team compiled information from every event at a school that involved a gun, then proceeded to upload that information into a database. The team found that 100% of mass school shooters were male, 100% came from middle to low-income households, 100% were white, and 66% experienced family and psychological issues. These are statistically significant as they can provide insight into how to structure the questions and what to look for. These statistics are also important as they align with information that Werner and Diederichsen had found.

Individuals who experience or have experienced psychological trauma or come from economically unstable households tend to have higher chances of becoming potential shooters. Dr. Freilich's goal was to create a program that could classify school shootings, which could help aid in the determination of significant traits that are found in shooters. This differs from my project as this project is about a program that seeks to find similarities in traits before the shooting. These two factors align with the findings of other researchers such as Steven M. Chermak, Ruth Defoster, and Jonatan Mizrahi-Werner, have found through other data, including Dr. Freilich's research. This project again differs from mine as they are pointing out traits from research, while I am using traits to determine a "risk factor" for an individual. Though Dr. Freilich, J. Werner, M. Diederichsen, and I, have one end goal, to reduce and end school shootings. Though thanks to lots of review, a few main traits have been set in stone as red flags: The belief that their race is superior. The end goal of this project is that of the others, but with that, the use of this program is to be trusted with the sole use of helping others, rather than out casting. These terrorist-like acts have destroyed families, ruined lives, and taken away what teachers desire, "a safe learning environment", and if these tragedies continue, they will only get worse, and the average number of victims will continue to rise. This is not a problem that we as a nation can let boil for any longer, one death is too much, and 17 (Nikolas Cruz, 2018) is horrifying (which is more than half of the average classroom). With that said, this project aims to be another support beam in the building that is the study of school shootings and be used to give insight into what societal and personal issues need to be fixed.



Methodology

When it comes to solving this problem, I will be addressing this issue, as said before, computationally. The best way to tackle this computationally would be to use a questionnaire that uses the input from a user to derive a score out of 100 that is based off of probability. To actually get to this point, a group of variables need to be created that are common traits or events that are found in an individual who have committed the shooting. This collection process involves the analysis of many research documents, literature, as well as news articles. These articles are often related to the psychology behind murder, mass murder, suicide, anger psychology, and school shooting studies. After the variables are created, each one should be assigned one to three questions that can achieve a goal of getting an answer that is related to that variable. The number of questions per variable will be determined by the difficulty creating a question that isn't all that intrusive. From that, each variable also equal to a set number of sentences that are related to the variable as answers. If more than one answer is needed, try to create a best to worst answer list that can give a score that is between 0 and the given score. For the overall question that is based on a grade scale for the percentage of that given score that will be added up. After that, create a function that will add all the variable scores up. Create a questionnaire that can be filled out by a user that contains the questions. Have each variable choose a random value in the list of scores. After each variable has chosen its value, have the code then multiply each variable by its weight. Then have all the new variables added together to give total score. Add a tally to score output group and then have the code then repeat steps repeat that 15 million times. Take this function and run it 10 total times and output each group of values into a list. Have the program find the average of the program. Repeat steps that for every time total score is manipulated.

1. Have you ever experienced harassment, verbally or physically, by a second party repeatedly over time?
2. Would you consider yourself to be a sociable person?
3. Has anyone close to you passed away?
4. Have you ever taken any type one drugs before?
5. Do you have friends that you enjoy hanging out with regularly?
6. Are you able to hold strong emotional relationships with anyone?
7. Would you consider yourself to be easily persuaded?
8. Do you enjoy fame?
9. Do you feel comfortable around guns?
10. Have you ever had anyone important abruptly leave your life, emotionally or physically?
11. Do you think about physical resolution when it comes to dealing with emotions?
12. Do you enjoy having control over people?
13. How sociable would you consider yourself?
14. How many times have you experienced or witnessed a physical or sexual assault on yourself or someone else?
15. Do you believe that you are a part of a superior race?
16. How often do you use prejudice in your day to day life?
17. How close are you to your parents?
18. Do you find yourself unable to control your emotions?
19. How strongly does social media and the internet frame the way you see other people?
20. Would you consider yourself to be an intrinsic individual?

```
Q1 = [0.0,0.25,0.5,0.75,1]
Q2 = [0.0,0.25,0.5,0.75,1]
Q3 = [0.0,0.25,0.5,0.75,1]
Q4 = [0.0,0.25,0.5,0.75,1]
Q5 = [0.0,0.25,0.5,0.75,1]
Q6 = [0.0,0.25,0.5,0.75,1]
Q7 = [0.0,0.25,0.5,0.75,1]
Q8 = [0.0,0.25,0.5,0.75,1]
Q9 = [0.0,0.25,0.5,0.75,1]
Q10 = [0.0,0.25,0.5,0.75,1]
Q11 = [0.0,0.25,0.5,0.75,1]
Q12 = [0.0,0.25,0.5,0.75,1]
Q13 = [0.0,0.25,0.5,0.75,1]
Q14 = [0.0,0.25,0.5,0.75,1]
Q15 = [0.0,0.25,0.5,0.75,1]
Q16 = [0.0,0.25,0.5,0.75,1]
Q17 = [0.0,0.25,0.5,0.75,1]
Q18 = [0.0,0.25,0.5,0.75,1]
Q19 = [0.0,0.25,0.5,0.75,1]
Q20 = [0.0,0.25,0.5,0.75,1]

question_values = [5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5,5]

question_answers = [Q1[1],Q2[1],Q3[1],Q4[1],Q5[1],Q6[1],Q7[1],Q8[1],Q9[1],
Q10[1],Q11[1],Q12[1],Q13[1],Q14[1],Q15[1],Q16[1],Q17[1],Q18[1],Q19[1],Q20[1]]

res_list = []
answer_list = []

for i in range(0, len(question_answers)):
    res_list.append(question_answers[i] * question_values[i])


scores = [':D',':)',':|',':C','>:C']

result = (sum(res_list))

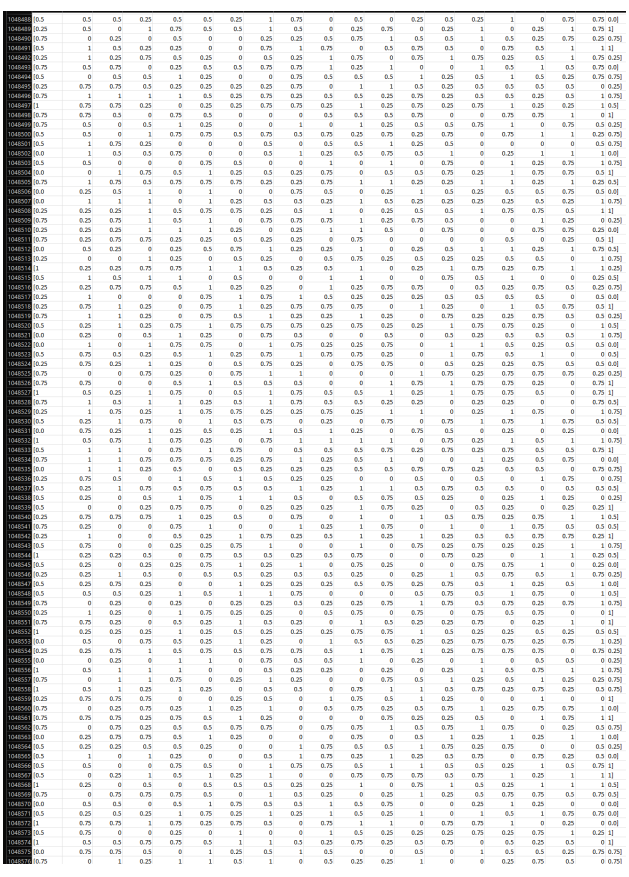
def score_summary(r,l):
    if ((r <= 19) and (r >= 0)):
        print(l[0],r)
    if ((r <= 29) and (r >= 20)):
        print(l[1], r)
    if ((r <= 49) and (r >= 30)):
        print(l[2],r)
    if ((r <= 69) and (r >= 50)):
        print(l[3], r)
    if ((r <= 100) and (r >= 70)):
        print(l[4],r)

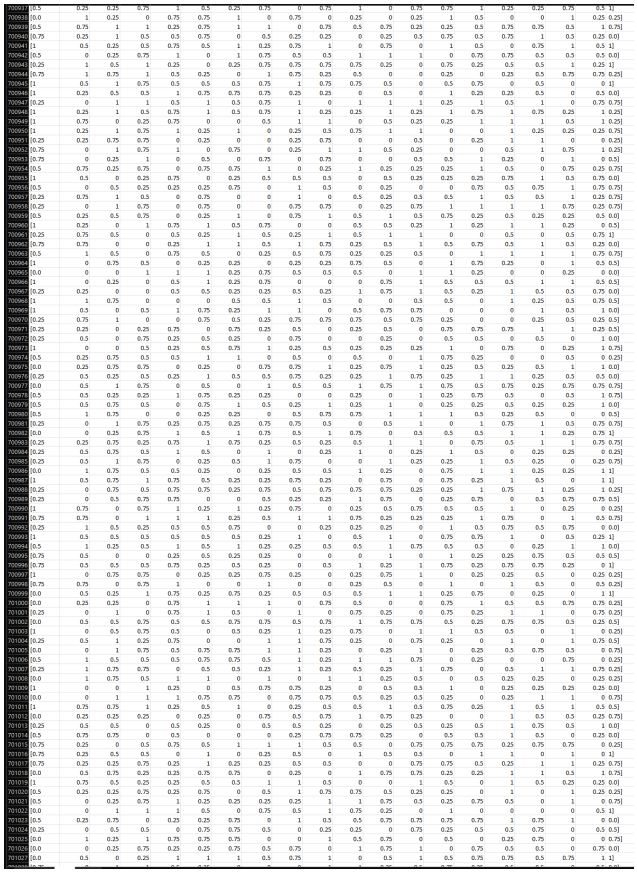
score_summary(result, scores)
```

Data Analysis

Once a rough outline of questions was created, they were uploaded to a Microsoft Form. The form served as a way to collect the data. Though the data used in the project was selected at random & did not directly use the form. The questions' goal is to be specific, yet general enough to be answered by anyone.

The best way to approach this is by finding what creates the motivation & what creates the strain. Also noting what actions were taken by previous shooters & threat makers on social media. "We find that strain motivates the agent to take action, & that this action is informed by objects of limitation that are





generated either through traditional news media or through online subcultures." (J. Mizrahi-Werner, Pathways to School hooting Subculture) Being that there were 15 million students in 2022, & 2,500 threats & shootings, I was hoping to get a severe threat average of 2500. Though this led to nothing as only Data 1 of the 10 had a combination of questions that led to a decent statistical relation of 99.99967% compared to 2022's statistics. Models 1-10 had no correlation with the only model coming even remotely close was Model 2. As the total score increased, the less accurate the model got. There is no consistent rate at which accuracy declines or increases. Model 1 represented a better way to structure not only the weights, but also the number of questions. The score out of 100 had the best results, & with that, the model was run 25 total times. The mean was 2529 individuals with a score of 88 or above. The standard deviation is 80.8, which is amazing being that this is out of 15 million. Though it doesn't prove that the questions have any validity, it does show that random input values can mimic real life scenarios. With the simulation having such strong correlation with what happened in 2022, I believe I achieved one of my research goals as now I know what to base the weights of each question out of.

Model 1	
Threat Level	Number
Little Possible Threat	7631200
No Possible Threat	2570382
Possible Threat	2313
Severe Threat	67
Very little Possible Threat	4801038
	15005000
Model 2	
Threat Level	Number
Little Possible Threat	3653526
No Possible Threat	6233523
Possible Threat	12
Severe Threat	138589
Very little Possible Threat	4974350
	15000000
Model 3	
Threat Level	Number
Little Possible Threat	2085396
No Possible Threat	6504523
Possible Threat	343493
Severe Threat	982354
Very little Possible Threat	5084234
	15000000
Model 10	
Threat Level	Number
Little Possible Threat	3430955
No Possible Threat	2304257
Possible Threat	3916858
Severe Threat	3049385
Very little Possible Threat	2298545
	15000000

Conclusion

Though it doesn't prove that the questions have any validity, it does show that random input values can mimic real life scenarios. With the simulation having such strong correlation with what happened in 2022, this project found and investigated similarities and differences in behavior between the average student and the shooters. It was not only able to find these similarities, but also led to the creation of a model that could be used to accurately test real people.

Bibliography

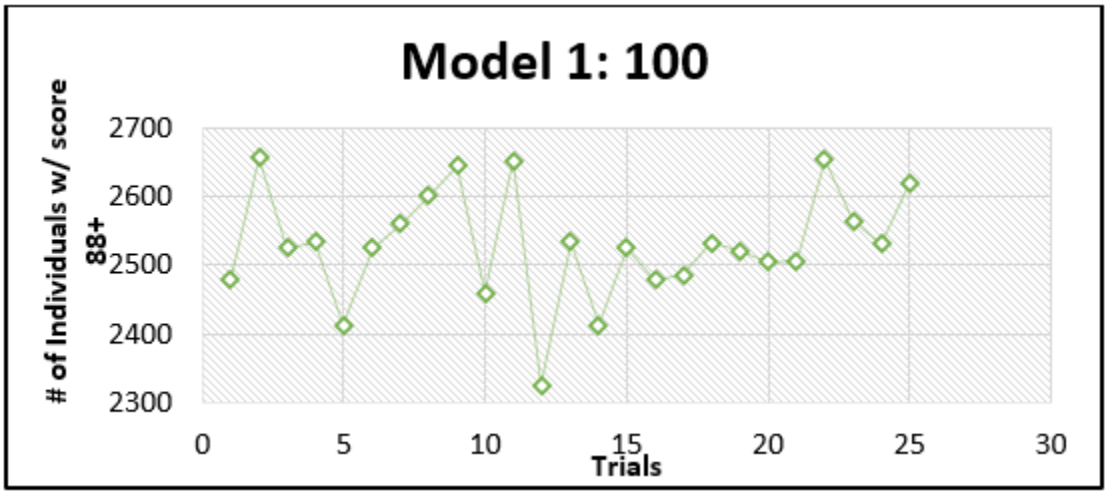
Boerboom, S., Pekel, K. and Dodds, M. (2022) *Mass shooting data & research, The Violence Project. Chief of Police, City of Minnetonka. Available at: https://www.theviolenceproject.org/*

Duzor, S.S.and M. (no date) VOA Special Report: History of mass shooters, VOA Special Report | History of mass shooters. Available at: https://projects.voanews.com/mass-shootings/

GBD compare (no date) Institute for Health Metrics and Evaluation. Available at: https://vizhub.healthdata.org/gbd-compare/

Learning From Error in Violence Prevention: A School Shooting as an Organizational Accident; Authors: Sarah Goodrum, Jessie Slepicka, William Woodward, Beverly Kingston; 2022

Not Another School Shooting: Media, Race, and Gun Violence in K-12 Schools; Authors:



Breahannah Hilaire, Laurie O. Campbell, Viki P. Kelchner, Eric D. Lagaurdia, Cassandra Howard; No date provided

Pathways to School Shooting Subculture: Re-thinking Theory across strain, Imitation and Digital mediation; Authors: Jonatan Mizrahi-Werner, Martine Bech Diederichsen, Benjamin Schou llsoe, Jakob Demant, Atte Oksansen; May 11th, 2022

School shooting data base (no date) *Database of American School shootings, Database of American School Shootings. Available at: https://www.schoolshootingdatabase.com/*

Terrorizing the Masses: Identity, Mass Shooting, and Media Construction of Terror; Author: Ruth Defoster; July 28th, 2017

Understanding the Causes of School Violence using Open-Source Data; Authors: Joshua D. Freilich, Steven M. Chermak, Nadine M. Connell, Brent Klein, Emily Greene-Colozzi; August 2021

When School-Shooting Media Fuels a Retributive Public: An Examination of Psychological Mediators; Authors: Megan J. O'Toole, Mark R. Fondacaro; 2017

Stroke Prediction with Logistic Regression Based on Various Risk Factors

Aparna Iyer

Nikola Tesla STEM High School

Abstract

Strokes are a leading cause of death around the world, with one person dying of a stroke every 3.5 minutes, and with this horrific death rate, predicting strokes is of high importance. This project seeks to investigate stroke prediction based on various potential risk factors. The dependent variable in this project is stroke status, which is classified as yes or no, whereas the independent variables are the risk factors, including age, gender, average glucose levels, hypertension, heart disease, residence type, smoking status, work type, marital status and BMI, all of which are recorded in 5110 entries. The first logistic regression model included all the variables and then the model was reduced to include only variables that were positively correlated to stroke and were statistically significant, as shown by their p-values. The results showed that hypertension, heart disease, age and average glucose level were positively correlated with stroke, with positive values indicating a positive correlation, and a value closer to 1 indicating a stronger correlation. Hypertension had the strongest correlation with a correlational value of 0.539597, and heart disease was second strongest, with a value of 0.404308. The p-values for these variables were all under 0.05, with hypertension having a p-value of 0.001820 and heart disease having a p-value of 0.046889, indicating that these findings are statistically significant. With this information, significant risk factors for stroke can be flagged, and this knowledge can be used in healthcare to minimize the risk of strokes through predictive and preventive care and treatment.

Introduction

The purpose of this study is to determine the change in stroke rates caused by a range of independent variables such as age, BMI, gender, heart disease and hypertension presence, smoking and marital status, work and residence type, and average glucose levels. Strokes have been a prevalent problem in the world for millions of years and occur when blood flow is blocked from reaching the brain. Strokes are separated into two categories: hemorrhagic and ischemic (Guhdar, et al). Hemorrhagic strokes occur when blood vessels found in the brain leak (often because they are ruptured or broken) and enter the brain tissue. This causes pressure on brain cells, which leads to them becoming damaged. On the other hand, ischemic strokes occur when some sort of particle, such as a blood clot, or fatty deposits, known as plaque, prevent blood from reaching the brain. Strokes are often seen in patients who display symptoms such as loss of speech or vision, slurring of speech and numbness or possible weakness on one or both sides of the body as well as on the face. When detected, they must be treated with urgency, as they can have devastating consequences. Strokes can cause problems such as vision damage and/or loss, impairment of sensory and/or motor functions, impairment or damage of bladder control, paralysis and comas (Ghanipour, et al). Strokes are often recurring, and a quarter of people that are affected by strokes have previously had a stroke (Hervella, et al). The immense consequences and occurring frequency lend their hand to high death rates as well, with one person dying of a stroke roughly every 4 seconds (Prathikshana, et al). Alongside the dangerous threats they pose to health, strokes also have enormous financial costs, with billions of dollars being spent yearly on treatment. Between

the years of 2017 and 2018, the U.S. spent over $53 billion on stroke related costs, which included medicine, healthcare treatment and lost productivity (Guhdar, et al). With all of this being said, looking into preventing and predicting strokes is extremely important, as this can help mitigate the risk of strokes in patients and therefore reduce death rates. Interestingly enough, up to 80% of strokes are preventable (Prathikshana, et al), and so looking further into this is essential to help save lives. A lot of research has been done regarding stroke prediction and correlations between various diseases and conditions such as diabetes and blood pressure in relation to stroke. In fact, blood pressure and stroke have shown strong correlations, and it is commonly known that various stressors can cause increased blood pressure (Hervella, et al). Factors such as affluence, family life, and daily habits such as diet and recreational activities (like smoking and drinking) can all play a hand in influencing stress levels. However, the research into these independent variables and their relation to stroke is not something that has been well researched. There are some indications that there may be positive correlations that can support connections between these unique independent variables and stroke, based on some correlational and causational studies. Based on this, exploring the correlations between unique stressors and stroke can be used to potentially help target patient care, and help healthcare providers determine when and how to treat patients appropriately, whether this is through medication, or by teaching patients how to make changes to their lifestyles to reduce their stress and risk of stroke. This project seeks to use logistic regression to create models to determine the correlations between a variety of independent variables and the risk of stroke.

Methodology

For this project, the first step will be downloading the data set (https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset?resource=download) from Kaggle. After downloading this data set, the independent variables are separated into categorical and continuous data. Categorical data encompasses all of the numerical variables like age, BMI, and average glucose level, whereas continuous data is all of the variables that can be divided into categories, like gender, which, in this case, was divided into such as male, female, and other. After this, certain central measures of tendencies were calculated for all of the continuous data, including standard deviation, mean, median, minimum and maximum (see Table 1).

	Std. Deviation	Mean	Median	Minimum	Maximum
Age	22.55613	42.86881	44	23	82
BMI	7.85432	28.89456	28.1	10.3	97.6
Average Glucose Level	44.42555	105.2974	91.68	55.12	271.74

Table 1. *Each of the continuous variables tabulated with information regarding the distribution of their data values (mean, minimum, maximum, standard deviation and median)*

After this, the frequencies for each of the categories within each of the categorical variables was calculated, both with percentages and just raw numbers (see Table 2).

Smoking Status					
	Former Smoker	Never Smoked	Current Smoker	N/A	Total
Numbers of People	885(17.32%)	1892(37.03%)	789(15.44%)	1544(30.22%)	5110
Gender of People					
	Female	Male	Other	Total	
Numbers of People	2994 (58.59%)	2115 (41.39%)	1 (0.02%)	5110	
Number of People with Hypertension					
	People Without Hypertension		People With Hypertension		Total
Numbers of People	4612 (90.25%)		498 (9.75%)		5110
Number of People With Heart Disease					
	People Without Heart Disease		People With Heart Disease		Total
Numbers of People	4834 (94.60%)		276(5.40%)		5110
Marital Status					
	Married At One Point/Still Married		Never Married		Total
Numbers of People	3353 (65.62%)		1757(34.38%)		5110
Type of Work					
	Child-Work	Government Job	Never Worked	Privately Employed	Self-Employed	Total
Numbers of People	687(13.44%)	657(12.86%)	22(0.43%)	2925(57.24%)	819(16.03%)	5110
Type of Residence					
	Rural		Urban		Total
Numbers of People	2514(49.20%)		2596(50.80%)		5110

Table 2. *This table shows the varying categories that fit under each categorical variable, and then the number of people that were registered as fitting under this category as well as the total number of participants. The percentage of people within each sub-category is also recorded in parentheses next to the raw numbers.*

After this, the categorical and continuous data were cross tabulated as well to explore frequencies between the varying types of independent variables (see Fig. 1 and Fig. 2).

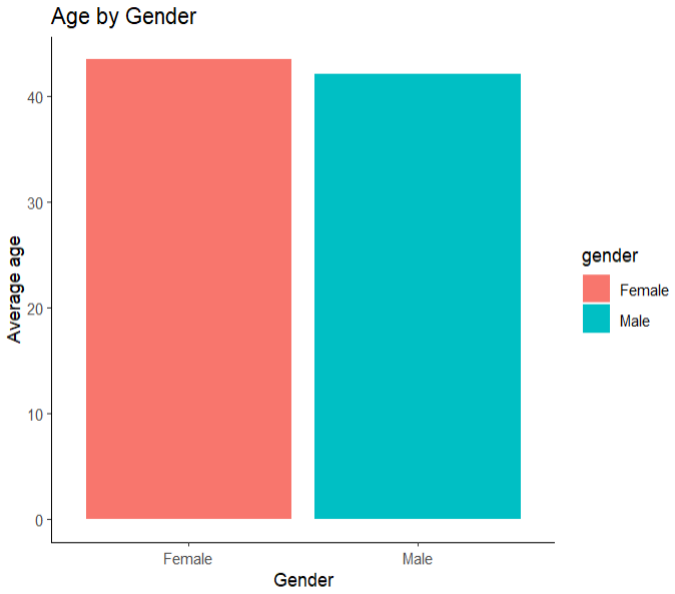


Figure 1. *Bar graph showing relation between the average ages of males and females (other is excluded as only one entry was available).*

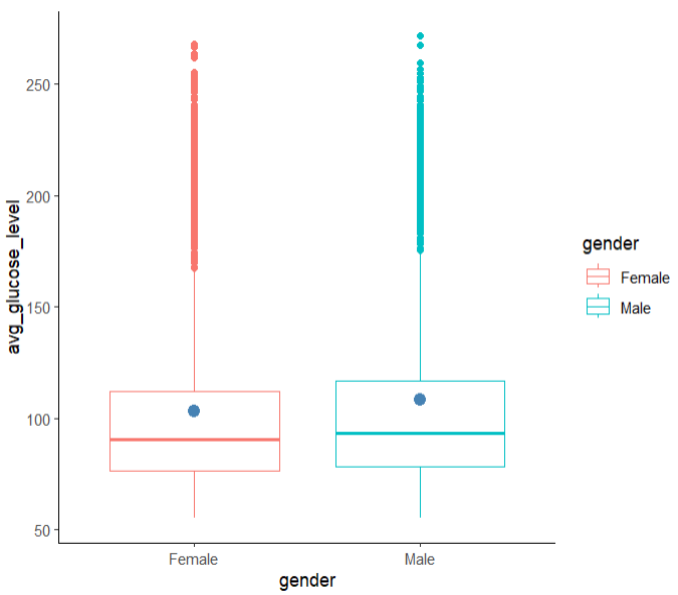


Figure 2. *Box plot showing the distribution of average glucose levels across males and females (other is excluded as only one entry was available).*

After this, code was developed to create training sets with which to test the model. The code developed would randomly run through the data and select 75 percent of the values to create a training set with. The code worked so that no values were replaced to ensure that repetition of values would not interfere with training the model. After this, the initial model was developed, and this was done through logistic regression. This model included all of the independent variables from the data set, and provided correlational values that were input to a table (see Table 3).

	Correlational Value
Gender Male	-0.014625
Hypertension	0.524857
Heart Disease	0.348763
Ever Married (yes)	-0.115175
Work Type – Government Job	-0.681655
Work Type – Never Worked	-0.9823495
Work Type – Private	-0.520849
Work Type – Self Employed	-0.945890
Residence Type – Urban	0.004514
Smoking Status – Former Smoker	0.275333
Smoking Status – Never Smoked	0.208109
Smoking Status – Current Smoker	0.589250
Age	0.073481
Average Glucose Level	0.004652
BMI	0.004062

Table 3. *This table shows the correlational values, in relation to positive stroke status, for all of the variables and many of the individual categories within the variables.*

After the initial model was created, stepwise logistic reduction was done to narrow down the independent variables and determine the most significant correlations. The results from the second model were also tracked in a data table and used to make qualitative conclusions regarding some of the data (see Table 4).

	Correlational Value
Hypertension	0.539597
Heart Disease	0.404308
Age	0.067540
Average Glucose Level	0.004802

Table 4. *This table shows the correlational values, in relation to stroke, for the independent variables hypertension, heart disease, age and average glucose level.*

Conclusion

With this project, the engineering goals were to explore multivariate models and also predict the correlations between stroke based on various independent variables such as age, BMI, gender, work and residence type, smoking and marital stats, hypertension and heart disease presence and average glucose levels. These goals were achieved as the data collected gave correlation values showing the relationship between each independent variable and positive stroke status. The most significant independent variables in relation to stroke were hypertension, with a correlational value of 0.539597 and heart disease, with a correlational value of 0.404308. Although these values are not extremely strong, they do indicate a positive relationship, and the ranked values can be used as a way to determine how to target patient care to then minimize the risk of stroke in the future. The p-values for each of the correlational values were under 0.05, being 0.001820 and 0.046889. As all of these p-values were under 0.05, the results obtained are determined to be statistically significant. So with the goal of determining risk of stroke in relation to a variety of independent variables, it is determined that the variables that are most strongly correlated with stroke risk are hypertension and heart disease.

Future Discussion

This project specifically investigated independent variables that had a positive correlation with stroke, however there were some negative correlations observed during the process that were fascinating, such as males and people who were married having a negative correlation with stroke status. This would be a future alley of investigation, looking into what potential risk factors could influence this negative correlation. Aside from this, coming out of a global pandemic, many people who suffered through COVID-19 expressed having long term damage that affected their health and quality of life. Looking into current stroke rates and how stroke rates fluctuated during the pandemic opens up another pathway for research, as the effects of COVID may have also impacted people's risk for stroke, and this could have negative long-term implications.

Works Cited

Anenberg, S. C., Horowitz, L. W., Tong, D. Q., and West, J. J.: An Estimate of the Global Burden of Anthropogenic Ozone and Fine Particulate Matter on Premature Human Mortality Using Atmospheric Modeling, Environ. Health Perspect., 118, 1189–1195, https://doi.org/10.1289/ehp.0901220, 2010.

Brauer, M., Amann, M., Burnett, R. T., Cohen, A., Dentener, F., Ezzati, M., Henderson, S. B., Krzyzanowski, M., Martin, R. V., Van Dingenen, R., van Donkelaar, A., and Thurston, G. D.: Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution., Environ. Sci. Technol., 46, 652–660, https://doi.org/10.1021/es2025752, 2012.

Fedesoriano, F. (2021, January 26). *Stroke prediction dataset. Kaggle. Retrieved April*

24, 2023, from https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset?resource=download

Fernandez-Lozano, C., Hervella, P., Mato-Abad, V. et al. Random forest-based prediction of stroke outcome. Sci Rep 11, 10071 (2021). https://doi.org/10.1038/s41598-021-89434-7

Ghanipour, Samaneh & Soroush, Yousefzadeh. (2022). Stroke Prediction with Logistic Regression and assessing it using Confusion Matrix.

Guhdar, Mohammed & Melhum, Amera & Ibrahim, Alaa. (2023). Optimizing Accuracy of Stroke Prediction Using Logistic Regression. Journal of Technology and Informatics (JoTI). 4. 10.37802/joti.v4i2.278.

Prathikshana M, Bobby & Muthuvel, Marimuthu & Velvadivu, Palaniswamy. (2021). STROKE PREDICTION. 7.

Wang, W., Rudd, A.G., Wang, Y. et al. Risk prediction of 30-day mortality after stroke using machine learning: a nationwide registry-based cohort study. BMC Neurol 22, 195 (2022). https://doi.org/10.1186/s12883-022-02722-1

Using VR In-vivo Exposure Therapy to Identify Startle Responses in Phobic and Non-Phobic Individuals Through Body Wave Analysis

Minal Pervaiz

Tesla STEM High School

Acknowledgements:

Special thanks to Julia Hall for mentoring this project;
this research would not have been possible without your assistance.
I am also grateful to J. Michael Strick for reviewing this research paper.
I'd also like to recognize Leah and Melanie Cammarano for their incredible praise and encouragement for this project as it came to fruition.
Thank you for your care and support.

Abstract

Arachnophobia is a prevalent phobia impacting between 3 to 6% of the global population, with such a large range due to the varying level of response dependent on each individual, yet the notion of quantifying this fear or identifying the rates of fear in both the phobic and non-phobic population has never been conducted. This study investigated whether visual arachnoid stimuli will elicit a natural fear response from individuals who reported little or no arachnophobia. The study utilized both phobic individuals (control group) and non-phobic individuals (experimental group), to measure their systolic blood pressure and heart rate before viewing the stimulus (independent variable) and after viewing the stimulus (dependent variable). An electroencephalogram (EEG), an electrooculogram (EOG), and sympathetic nervous system (SNS) waves were also utilized to identify the moment at which the individual elicited any response from viewing virtual arachnids, as well as quantifying the rate at which the continued exposure to the stimulus occurred. The heart rate was statistically significant with a p-value of 0.02584, whereas the blood pressure was not, with a p-value of 0.46049 (evaluated from the use of chi-square testing). However, the body waves resulted in the greatest increase in spindle height and frequency between 0.1-1 seconds of viewing the stimulus and about a 30% decrease in spindle height and frequency afterwards, indicating that continued exposure resulted in decreased anxiety. This research supports that exposure to arachnids stimulates a body wave response and effectively declines over a prolonged exposure period for both phobic and nonphobic individuals.

Introduction

Like the many aspects of human life, the essence of fear is constantly evolving; it remains a reliable tool that cautions around potential harm and ultimately keeps us alive. Fear is shaped by our environment and experiences, an accumulation of

knowledge over time from the age of naive infancy to our elderly years. As a population, society has shaped its people and has dampened the influence of natural survival due to extended periods of exposure. Yet external factors within society may also trigger unwanted phobias that are essentially unreasonable and unrealistic – a very different approach to the existential fear humans are born with. This project seeks to understand whether arachnoid exposure elicits a natural and immediate fear response from both individuals with and without arachnophobia – the fear of spiders – taking into account the range of responses produced between the phobic and nonphobic individuals, and determining if such responses are inherent in all people based on the amygdala's response to the stimuli.

As a result of the stimuli that the amygdala associates with danger, the eliciting of fear allows both humans and animals to reach their full muscle capacity, which is definitively the body's greatest strength potential at a given time – a system that allows for the ultimate use of fight or flight skills under the pretense of survival. This system implements the extreme conditions of the human body to be accessed for a potentially short period of time before returning to homeostasis, all activated by the response from the amygdala. The change in muscular and physical abilities during this time are greatly enhanced as compared to their normal capability. These circumstances are only possible in the presence of great fear, as is the case in exposure to major phobias, so hypothetically the body's full muscle capacity could be achieved during such exposure. However, with arachnoid presence a daily occurrence in much of life, non-phobic individuals must have a response elicited from the amygdala that is not significant enough to hinder their ability to continue living as normal – though the amygdala's fear response must provide a brief increase in body function just before the body returns to homeostasis and the exposure's effect has worn off.

Due to the difficulty in measuring the extent of these fear responses, we are limited in investigating the average maximum capacity of muscular function during exposure with a feared stimulus. Instead of measuring this muscle capacity, this project seeks to investigate whether a common stimulus presented frequently within our lives holds any similar – albeit smaller – response of this "fear adrenaline": spiders. Arachnophobia affects around 3 to 6% of individuals globally and remains one of the hardest phobias to treat in a clinical setting due to its high chance of relapse after therapy. It has also been found to affect more females than males – as women tend to hold more specific phobias, close to 80% of the general population (Magee et al, 1996). More research has been presented that suggests a correlation to sex chromosomes and negative reactions to potentially threatening stimuli, even more so towards arachnids. While females are biologically more impacted by arachnophobia, this project will use both male and female participants to conduct the procedure. Additionally, this project will conduct a double-blind procedure, where both the experimenters and participants are unaware of which participants are placed within the experimental or control groups.

To achieve these results, the study will use virtual reality (VR) in-vivo exposure therapy, where the stimulus is presented in front of the participant to simulate automatic exposure. Virtual reality will be utilized in the experimental procedure as the use of VR stimulus exposure still presents an immersive experience to the user and is proven to be an effective treatment for patients (Miloff et al, 2019). The data received from this study will then be used to understand how therapy may decrease the phobic response to a similar extent that a non-phobic individual may experience when presented with arachnoid exposure.

Using behavioral approach testing and exposure therapies, the direct confrontation of an arachnoid stimulus will promote an increase in systolic blood pressure (SBP) and heart rate, and analysis of the electrode signals will identify a significant increase in electroencephalogram signals and body impulses. Under fearful circumstances, our bodies are signaled to involuntarily promote

actions from the sympathetic nervous system under influence of the amygdala, which is responsible for detecting threats. This amygdala response will further allow us to detect these signals via electroencephalogram (EEG), electrooculogram (EOG) and sympathetic nervous system (SNS) electrodes placed around the body.

This study questions whether the exposure of spider stimuli is enough to produce electrical responses in the body from non-phobic individuals. Through the use of VR exposure therapy, direct confrontation of arachnoid stimuli will promote hyperactivity from the amygdala to produce these responses from the sympathetic nervous system of these individuals. As the amygdala has been found to activate threat detection faster in phobic individuals, there is an expected increase in electrical responses towards arachnoid stimuli by way of the participants' EEG, EOG, and SNS waves.

Methodology

This project identifies fear responses through behavioral approach testing and exposure therapy, where patients are monitored and exposed to arachnoid stimuli to gain systematic desensitization.

This project notably involves the use of a VR headset to implement in-vivo virtual reality exposure therapy, as well as a number of forms distributed to participants involved in the study. These forms included a customized Taylor Manifest Anxiety Scale (TMAS – Fig. 1), provided to participants to ensure individuals are not experiencing significantly high anxiety that will hinder their participation in the study. Additionally, a customized Fear of Spiders Questionnaire has been used for this study (FSQ – Fig, 2), of which the original FSQ is found and derived from the Syzmanski and O'Donohue study in 1995 ((1995). Fear of Spiders Questionnaire. Journal of Behavior Therapy and Experimental Psychiatry, 26(1), 31-34. https://doi.org/10.1016/0005-7916(94)00072-T); the FSQ was used to divide participants between the control and experimental groups. The collection of nonintrusive multimodal responses will be collected through the use of the Backyard Brains Heart and Brain SpikerShield equipment, with the SpikerShield's electrodes placed on each participants' head and arms. All individuals were provided an informed consent form that had to be completed before participation in the study, and all participants who were minors were required to include a parent or guardian signature.

Taylor Manifest Anxiety Scale form

This form is provided to you to gather information about your wellbeing leading up to the study. Please answer with the most accurate representation from the answers/numbers provided. All information will remain confidential.

		No		Yes
1.	Do you find it difficult performing any tasks that require more than a minimum effort to complete?	1	2	3
2.	Are you having any difficulty performing physical tasks for long periods of time?	1	2	3
3.	Are you having any difficulty performing physical tasks for short periods of time?	1	2	3
4.	Do you spend most of the day in bed, or feeling tired?	1	2	3
5.	Have you experienced any difficulty sleeping in the past week?	1	2	3
6.	Have you felt a lack of appetite in the past week?	1	2	3
7.	Have you felt physical weakness in the past week?	1	2	3
8.	Have you felt irritable in the past week?	1	2	3
9.	Have you felt worried or anxious in the past week?	1	2	3
a.	Have this interfered with any of your activities or social interactions?	1	2	3
10.	Have you experienced any sickness (nausea, vomiting, weakness, etc.) in the past week?	1	2	3

Figure 1

*Fear of Spiders Questionnaire form**

1. If I encountered a spider now, I would need to get help from someone else to remove it.
2. Currently, I am looking out for spiders.
3. If I were to encounter a spider now, I would think it would harm me.
4. I think a lot about spiders in a negative way.
5. If I encountered a spider in a room before, I would be afraid to enter the room.
6. If possible, I would do anything to avoid a spider.
7. Sometimes I think about being bit by a spider.
8. If I saw a spider now, I wouldn't be able to deal with it efficiently.
9. If I saw a spider now, it would stay in my mind for a long time.
10. If I encountered a spider now, I would need to leave the room.
11. If I encountered a spider now, I may think it will try to jump on me.
12. If I encountered a spider now, I would ask for someone to kill it.
13. If I encountered a spider now, I would feel afraid of it.
14. If I encountered a spider now, I would panic.
15. Spiders are one of my worst fears.
16. If I encountered a spider now, I would feel very nervous.
17. If I encountered a spider now, my heart would likely beat faster.

Figure 2

Participants were gathered by volunteering through an online interest form posted at Tesla STEM High School and was limited to its population of students, teachers and staff. To ensure a representative and random sampling of participants, only one of two persons interested were chosen to participate in the study, decided by a random number generator. These participants were further contacted for questioning their existing fear of spiders, and testing of general anxiety or depression that may potentially be impacted by the study. This study implemented a double-blind procedure, and participants were divided into the control and experimental groups as decided by the results from the FSQ; questions 5-21 will determine these groupings. Participants scoring between 0-3 were sorted into the experimental group, and all other participants were sorted into the control group.

The participants chosen for the study were given instructions for waiting in a separate room from the experiment area. A single participant was chosen at a time to be tested, where the initial SBP and heart rate were taken prior to testing. The SpikerShield electrodes were then placed onto the participant depending on the particular electrode arrangement (from #1-3, where 1 signified the SNS arrangement, 2 signified the EEG arrangement, and 3 signified the EOG arrangement) (EOG electrode arrangement – Fig. 3), and connected to a digital recording device to collect the body waves during the procedure.



Figure 3

Image (cropped) from Backyard Brains at Experiment: Eye Potentials (The EOG) (backyardbrains.com)

The participants were given the VR headset with their eyes closed before being instructed by examiners to initiate the procedure. The participants then opened their eyes and engaged with the arachnoid material for 45 seconds while body waves were collected and examined simultaneously. At the end of the exposure period, the final SBP and heart rate were recorded immediately into the original data sets (see Fig. 4 for data table).

Participants	Initial Systolic Blood Pressure	Final Systolic Blood Pressure	Initial Heart Rate	Final Heart Rate
1	115	104	71	88
2	123	104	77	80
3	107	109	90	94
4	138	141	75	78
5	124	122	83	79
6	137	112	74	83
7	119	114	60	82
8	118	96	88	98
9	125	125	76	71
10	132	130	67	66
11	111	124	79	93
12	128	127	75	75
13	121	127	81	87
14	108	124	66	82
15	130	126	65	66
16	104	113	64	77
17	126	142	82	81
18	117	122	82	83
19	112	117	99	101
21	99	99	89	86
22	110	122	90	107
23	103	105	66	75
24	100	113	76	78
25	133	128	74	87
26	131	135	71	83
27	105	110	80	83
28	94	110	74	100

Figure 4: Phobic and non-phobic participant data after exposure to the arachnoid stimulus, of systolic blood pressure and heart rate.

Data Analysis

After performing chi-square, the data showed that the SBP results were not statistically significant during both trials of the procedure. The first trial had a total of 14 participants conducting the experimental procedure. After chi square testing it was found that the corresponding p-value for the SBP was 0.18925, which was far greater than an expected value of 0.05 or below for the hypothesis to be accepted (seeing an increase in the SBP after stimuli exposure). The second trial held an additional 14 participants in which the SBP data

was again not statistically significant, with a cumulative p-value of 0.46049. The initial and final mean readings of the total SBP data, 116.857 and 117.821 respectively, shows very little difference between the initial and final values. As such, the data indicates that there is very little regard to systolic blood pressure in the startle response of the individuals. However, some considerations must be made in regard to the machine readings of the SBP data: during the first trial, the machine reading output participants' data after the exposure period had been completed, which may have affected the SBP data in regards to the time period in which the actual SBP was calculated; i.e., the blood pressure was not recorded immediately. Fig. 5 shows the first trial results for SBP, showing that the final values for the SBP were generally lower than the initial values. Fig. 6 shows the entirety of the study's participant data results with more instances of final SBP being lower than the initial.

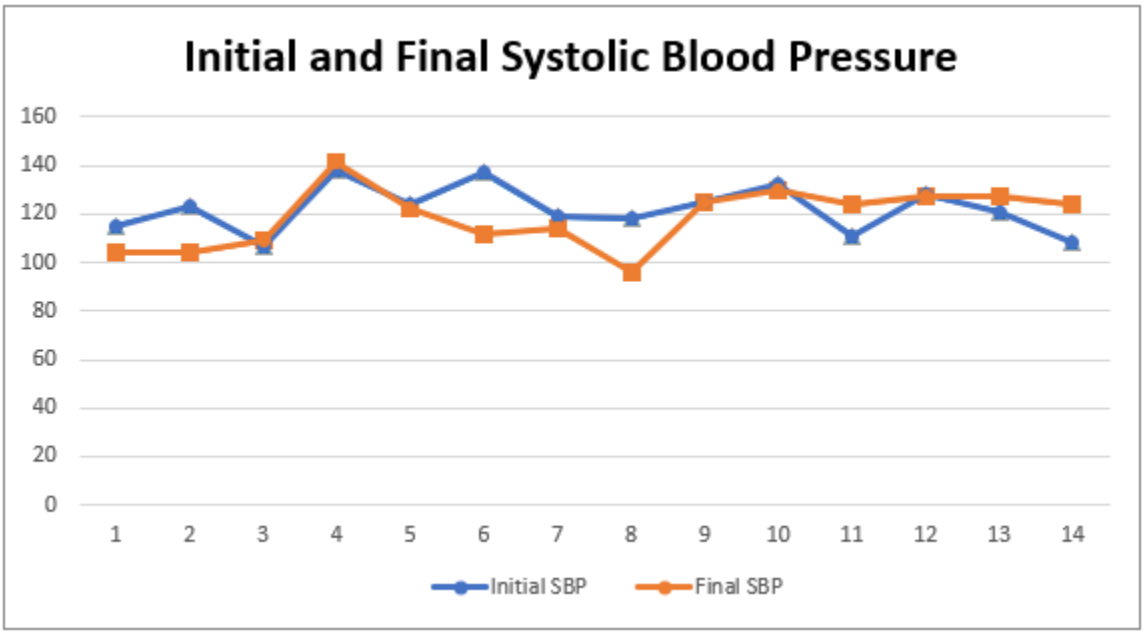


Figure 5: First trial results for systolic blood pressure

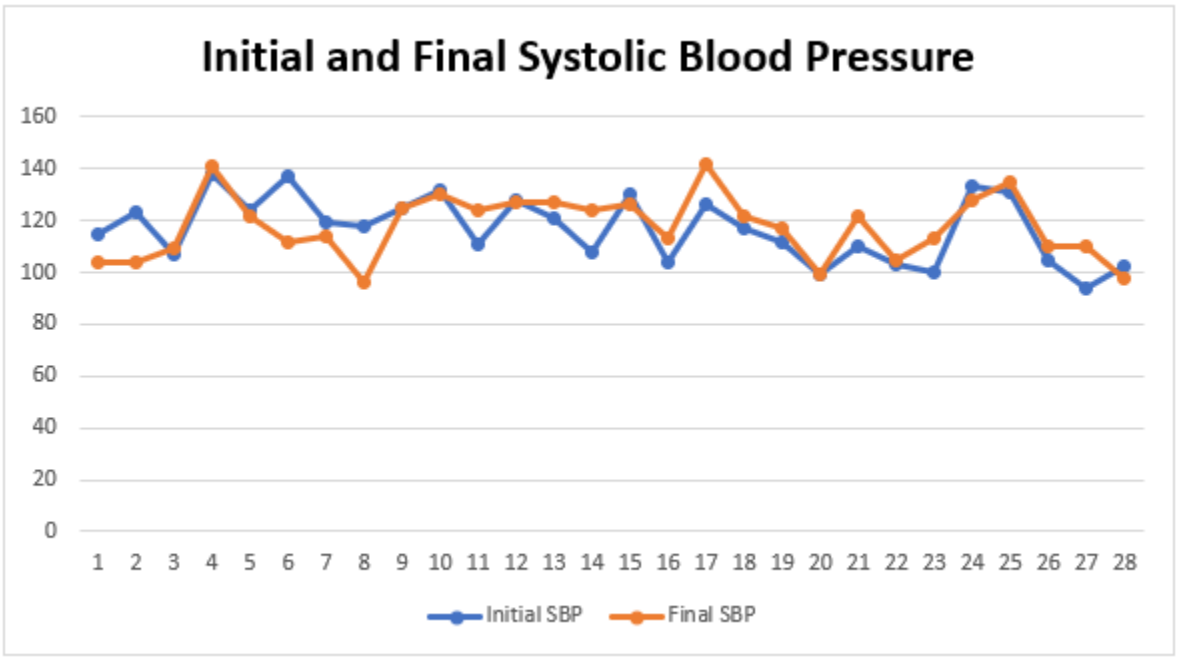


Figure 6: Total results for systolic blood pressure

Chi square testing of heart rate results indicated that the data was statistically significant in both trials of the study. The first trial's 14 participants held a chi square p-value of 0.053476, indicating that the data was correct as it related to the study's hypothesis (seeing an increase in heart rate after stimulus exposure). The second trial's 14 participants held similar results, totaling with a chi square p-value of 0.02584, which proves that the data was statistically significant as the value was found under 0.05. The initial and final mean readings of the heart rate, 76.5714 and 83.3929 respectively, also indicate the significant change in values before and after viewing the stimulus. Fig. 7 shows the first trial results with much of the trial's participant data showing a higher final reading than its initial. Fig. 8 shows the entire study's participant data also relating higher final readings than its initial.

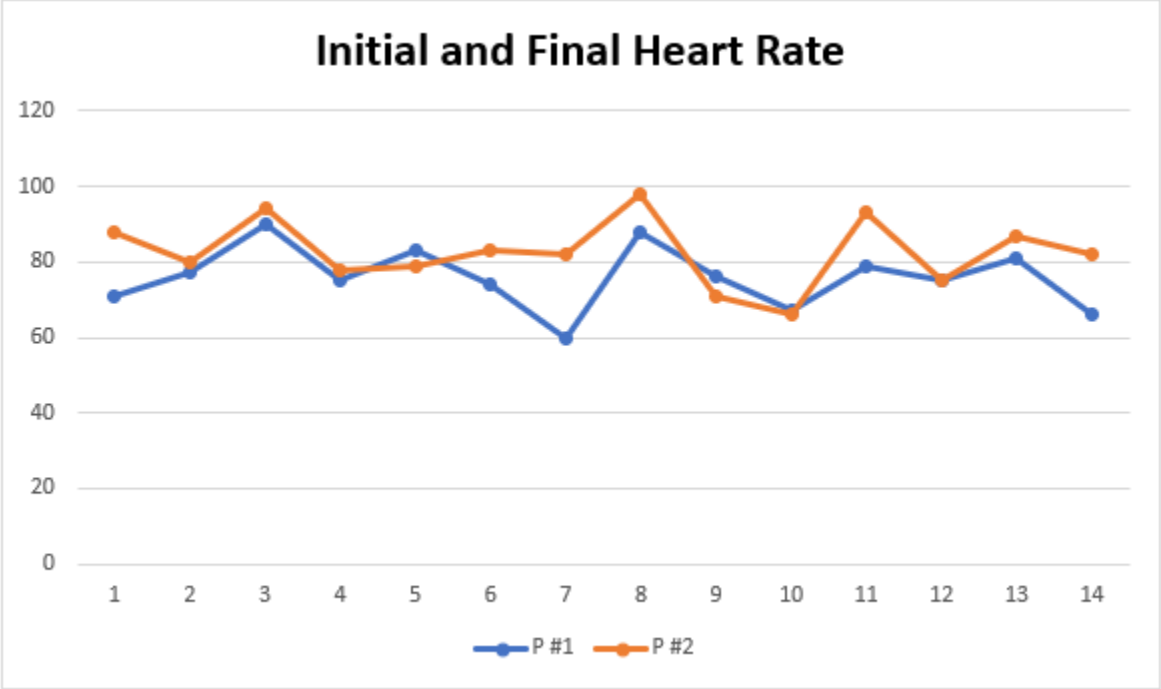


Figure 7: First trial results for heart rate

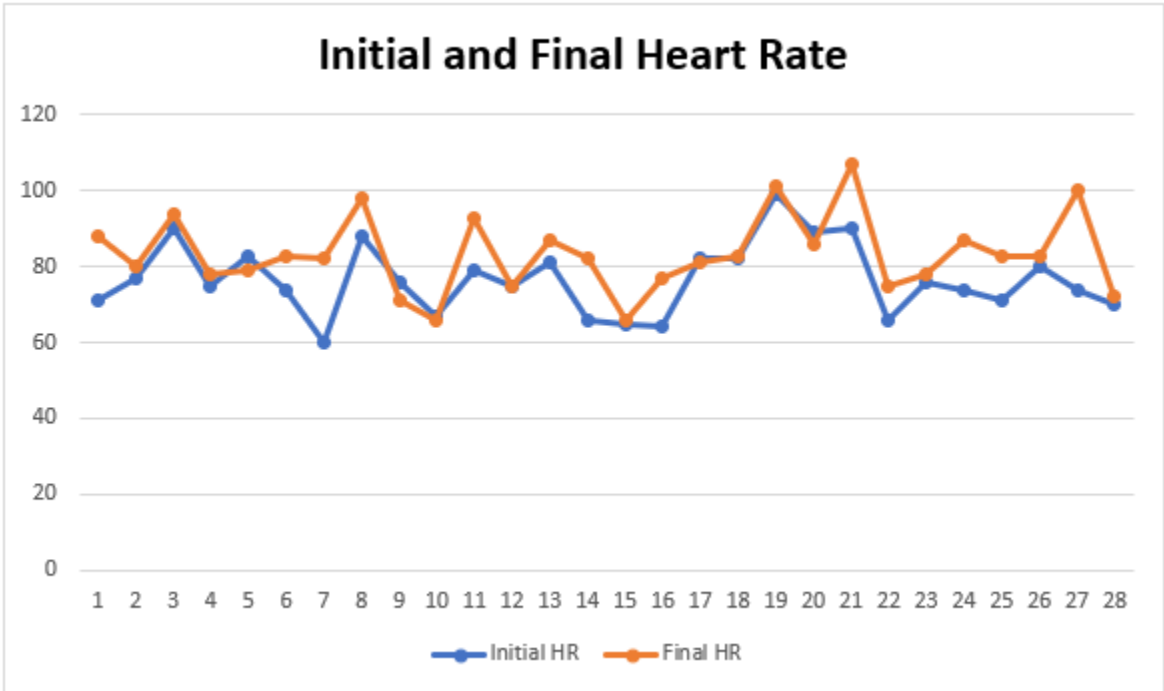


Figure 8: Total results for heart rate

Additionally, under a Fourier transform the corresponding frequencies were compared by each phase of the procedure: pre-exposure, initial exposure, and the prolonged exposure (40 seconds into the exposure period). Through percent difference it was found that the initial exposure was generally about half the frequency as compared to the initial exposure (the identified startle response of the individuals), and the prolonged exposure showed a similar frequency as had been elicited during the pre-exposure period. The startle responses were visually identified as the spindles of the body waves resulted in the greatest increase in spindle height and frequency between 0.1-1 seconds of viewing the stimulus for most participants. Additionally, about a 30% decrease in the spindles' amplitude and frequency was identified around 10 seconds after initial exposure for non-phobic participants (see Figs. 9, 10 and 11).

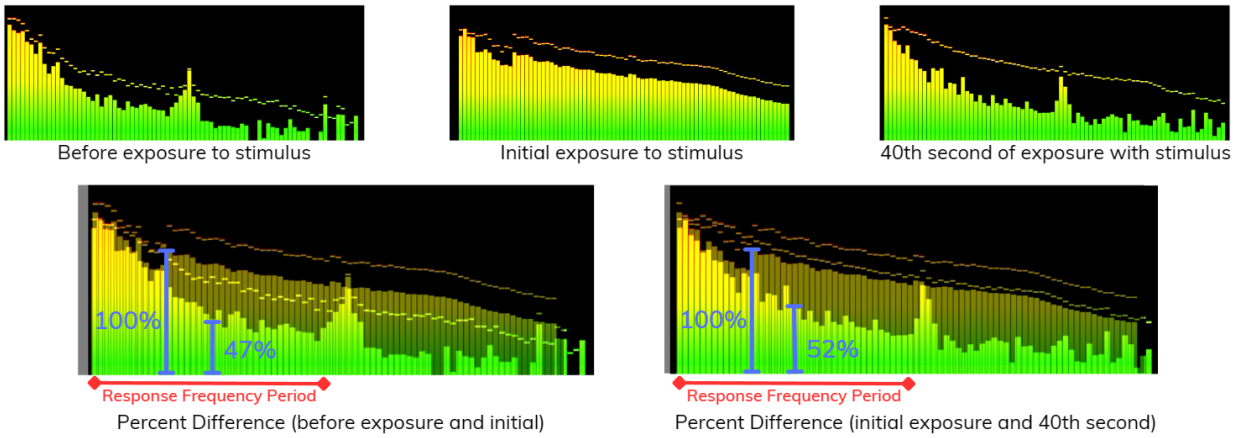
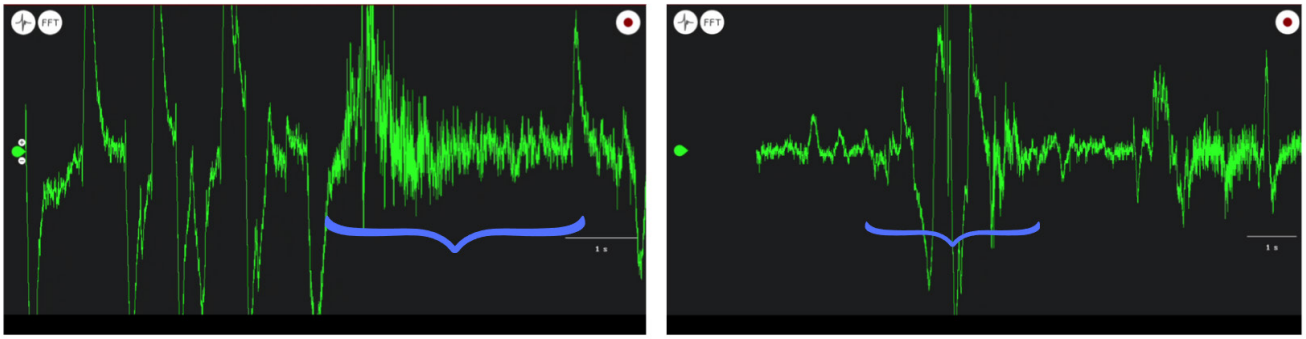


Figure 9: Sample percent difference of non-phobic response each of the three phases of the procedure (pre-exposure, initial exposure, and prolonged exposure)



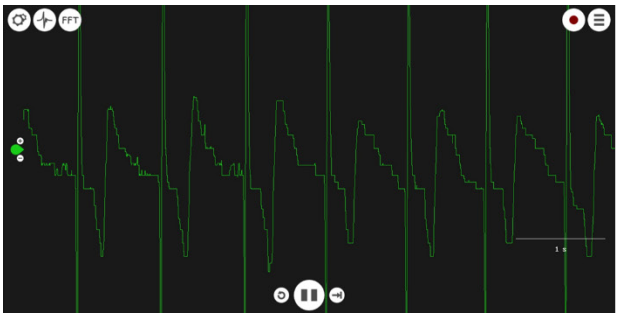
Figures 10 and 11: Identification of the startle response of two non-phobic participants. Figure 10 (left) shows an increase in the spindle frequency following initial exposure to the stimulus, whereas Figure 11 (right) shows an increase in the spindle amplitude during initial exposure.

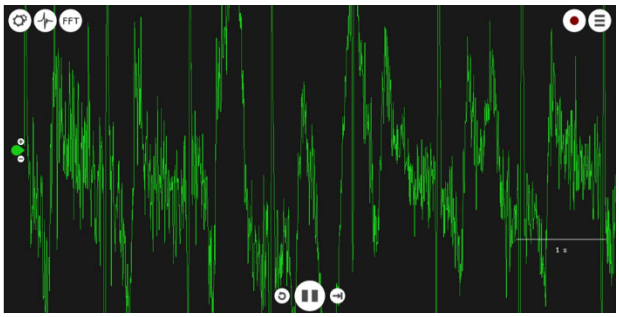
Conclusion:

The SBP data was not statistically significant as it related to the study's hypothesis, as compared to the heart rate data which was statistically significant, with p-values of 0.46049 and 0.02584 respectively. The body waves resulted in the greatest increase of spindle height and frequency between 0.1-1 seconds of viewing the stimulus. There was also about a 30% decrease in the spindles' amplitude and frequency after, indicating that continued exposure resulted in decreased anxiety. Using percent difference, it was found that the initial exposure produced frequencies more than doubled as compared to the pre-exposure period, and 40 seconds into the exposure period displayed frequencies around the same amplitude as the pre-exposure period.

Limitations of the study included the first trial being limited to only auditory data, whereas the second trial handled both auditory and visual data

(see Figs. 12 and 13). Additionally, a delay in the SBP reading was identified for the first trial but was not present during the second trial. There were 6 outliers found from the experimental group (28 individuals, 10 phobic (control group), 18 non-phobic (experimental group)), though imposing frequencies on the electrodes from the surrounding environment must be considered. Moreover, the body wave analysis using the Fourier transform was completed at an instantaneous point in time as compared to the desired analysis over an extended period.





Figures 12 and 13: Pre-exposure and prolonged exposure phases of a phobic participant using the SNS electrode arrangement. Figure 12 (left) displays the initial SNS readings of the individual before stimulus exposure, whereas Figure 13 (right) shows the final SNS readings.

Discussion

This research supports how exposure to arachnids stimulates body wave responses and effectively declines over a prolonged exposure period for both phobic and nonphobic individuals. This research may further be utilized by clinical professionals providing exposure therapy for patients, by understanding the period and extent of the startle response produced by non-phobic individuals as a foundation when reducing the fearful responses of phobic individuals. This study may further be expanded by employing a Fourier analysis of the frequencies over an extended period of time, as well as utilizing a greater number of electrodes placed on the body at once as compared to the study's electrode arrangements (which each included only 3 electrodes at one time). The generalizability of the results may also be expanded upon with a larger and more diverse sample size for future research.

Bibliography

1. Klahn AL;Klinkenberg IA;Notzon S;Arolt V;Pantev C;Zwanzger P;Junghöfer M;. (n.d.). Prepare for scare-impact of threat predictability on affective visual processing in spider phobia. Retrieved November 1, 2022, from https://pubmed.ncbi.nlm.nih.gov/27036648/.
2. Larson CL;Schaefer HS;Siegle GJ;Jackson CA;Anderle MJ;Davidson RJ;. (n.d.). Fear is fast in phobic individuals: Amygdala activation in response to fear-relevant stimuli. Retrieved November 1, 2022, from https://pubmed.ncbi.nlm.nih.gov/16919528/.
3. Peira N;Golkar A;Ohman A;Anders S;Wiens S;. (n.d.). Emotional responses in Spider Fear are closely related to picture awareness. Retrieved November 1, 2022, from https://pubmed.ncbi.nlm.nih.gov/21973031/.
4. Policy on the use of human subjects. (n.d.). Retrieved December 1, 2022, from https://policy.csuci.edu/sp/3/sp-03-029.htm.
5. Use of human subjects in research. (n.d.). Retrieved December 1, 2022, from https://provost.harvard.edu/use-human-subjects-research.
6. Vasey MW;Vilensky MR;Heath JH;Harbaugh CN;Buffington AG;Fazio RH;. (n.d.). It was as big as my head, I swear! biased spider size estimation in spider phobia. Retrieved November 1, 2022, from https://pubmed.ncbi.nlm.nih.gov/21906909/.

Testing the Effects of the Smiling Mind Mobile App on Reducing Anxiety Symptoms

Milana Veera

Nikola Tesla STEM High School

Abstract

Mental health is a continually increasing problem, especially for students who are struggling with school while figuring out their identity. As a result, nearly 1 in 3 teenagers have some type of anxiety disorder. This study investigates the effectivity of the *Smiling Mind* mobile app on reducing anxiety symptoms in teens 16-18 years of age. The *Smiling Mind* app offers meditation and breathing exercises to address anxiety and stress, and there is options for teens. A standard sampling procedure was used to gather a representative sample of teenagers and pre-screen them. Then, the Beck Anxiety Inventory was administered before and after the app was used, giving a score from out of 63 based on the severity of anxiety symptoms. After setting this baseline, the app was distributed and set up informing participants to use the app for 5 minutes every day around the same time, reminded by a notification. This concept is similar to a prescribed dosage approach given for medication, but with none of the harmful side effects. A paired t-test was performed to analyze and compare the values. The mean of the post app group had a 20% reduction in anxiety than without using the app. Using this, a t-statistic value of 4.24 and a p-value of less than 0.001 was calculated, showing that the hypothesis was correct. The *Smiling Mind* app clearly reduced anxiety symptoms by statistically significant amount in these teenagers, and it can be used to help the future of all students.

Introduction

This experiment focuses on testing a meditation based mobile app called *Smiling Mind* on reducing the number of anxiety symptoms they have using the Beck Anxiety Inventory. This inventory is comprised of twenty-one questions to calculate the number of anxiety symptoms based on the participants' ratings. One in five adolescents (12-18 years) are diagnosed with at least one mental health disorder and of these, many do not know or have access to the help they need (Adolescentadmin, 2022). The struggle adolescents have with mental health and discovering their identity is continually getting worse, causing a struggle to get back from the effects of the COVID pandemic when social contact was limited. This specifically is seen in students' performance in school while they try to return to normal routines and keep up with the academics like they used to. Teenagers need a method that is more accessible and readily available for them, using resources around them such as their phones. Therefore, one way to provide teenagers with mental health support is utilizing mental health applications on their own phones.

Using free mental health apps show to have a positive effect in various studies. For example, one study stated that "The current systematic review shows that mobile apps for mental health intervention in college students exists and demonstrates good acceptability and feasibility. They also demonstrate efficacy among students" (Oliveira 2021). This specific study focused on an app called *mHealth* and tested it on college students. The results showed that there were significant differences in students that used the app verses students who did not. Focusing on this

accessible, easy way to get help is the goal. The factor that will be done differently in this study is testing a prescribed dosage approach that none of the apps focus on, and this is where people use it consistently everyday each time. This means that 10 minutes on the app will have similar effects on symptoms as taking an anxiety pill, but without a pill. As a result, any risk of overdose is eliminated but the help will still be provided. Also, teens will not need to go get medication prescribed if they are not comfortable with that, or if the parents do not approve.

Like the previous study displayed improvements, another study researching online apps for anxiety symptoms found that "there was a significant group-by-time effect (the change between groups is different at different times) over 8 weeks ($F_{2,33\ 316} = 15.9$; $P < .001$), meaning those assigned to the mobile platform experienced a faster rate of improvement than without it" (Graham 2020). This shows that apps do provide significant positive effects for the number of symptoms people experience therefore helping improve the problem at hand and teenage mental health.

Other studies have also displayed similar beneficial outcomes. In one study, five apps were tested to see whether those apps improved participants' symptoms before and after the app was presented. It showed that "by post-intervention, all but one participant (out of six) (C3) had achieved a reduction in subjective distress; four to a non-noticeable level (i.e., a rating of < 3)" and similarly other examination showed that, with the exception of one person, everyone displayed apparent discomfort at baseline. However, four subjects had reduced their subjective discomfort by the post-intervention, three of the subjects to an unnoticeable level (Marshall 2022). These were for apps called MoodMission and Destressify which are also mental health apps, and four of the five apps showed similar results that the participants improved from how they were at the baseline (before app use) to where their symptoms were unnoticeable compared to before.

Despite these resources available, about 1/3 of teens in the US have some form of anxiety. Moreover, less than half of these people actually reach out for help. The problem is that many of them do not know about all these mental health resources available right on their phones, so instead of feeling uncomfortable about their problems or having to talk to their parents to make them pay for help, they can try out the online mobile apps that are free for the public and the tap of a screen away.

As of yet, the consistency of how this app affects anxiety symptoms to a similar effect an anxiety pill might provide for teenagers has not been investigated, which is exactly what this study aims to do. Focusing on accessible mobile apps for the public and teens, the testing of this specific app called *Smiling Mind* focuses on a certain amount of mentally helpful activities centered around meditation. It has over 500 rating on the app store, most of which are positive as the app has over 4 stars. This app was particularly chosen because of the prior studies it was in where results were not consistent, and it was only tested on adults. Now instead of this, the app will be used consistently and produce consistent results for teenagers' anxiety. This project seeks to investigate the improvement of anxiety symptoms on teenagers (16-18 years) using *Smiling Mind.* Measurements will be based on the Beck Anxiety Inventory (BAI). This will help the mental health crisis in teens, giving the future the best chance. The study hypothesized that the *Smiling Mind* app reduces the number of anxiety symptoms in teenagers overall, measured using the Beck Anxiety Inventory.

Methodology

In this study, the Beck Anxiety Inventory which is a 21-question test to determine levels of anxiety, a total of seventeen participants that are between the ages of 16 to 18 years old, the *Smiling Mind* free public mobile app, and 10-15 flyers with QR codes to gather participants all were used. For the data, Microsoft Forms will be used to send out

the surveys and collect it, and Jamovi version 2.3.18 to analyze the data.

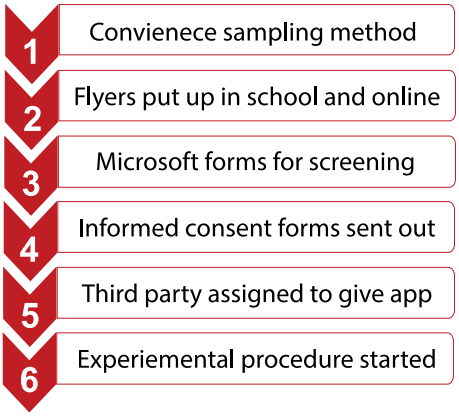


Figure 1: Convenience sampling method and recruitment of participants

First, a convenience sampling method (see Fig. 1 above) was conducted to obtain the participants from the LWSD school district. This means that participants that were the easiest to reach were the ones in my study, so flyers made on Illustrator were put up in public high schools (Tesla STEM, Eastlake, and Redmond High school) and on Instagram to recruit participants. Once they agreed and fit the criteria (16 to 18-year-old high schooler) they were provided a text box to ask any and all questions, confirming their participation through Microsoft forms. It was also checked that they do not already take anxiety medication through prescreening questions in the forms, as that would affect the results. After that was confirmed, informed consent forms were sent out to each participant for signatures with a parent signature for minor participants. To avoid experimenter bias, a third party was picked to distribute, explain app setup, and expected use. This included app installation from the app store, signing in, and setting up notification reminders so that participants maintain the prescribed dosage approach. After the informed consent forms were sent back, they were locked away in a safe box and assigned a number to each of the participants to start the pre-test experimental procedure.

The variables utilized in this study were the time using *Smiling Mind* as the independent variable, and the amount of anxiety symptoms measured through the results of the Beck Anxiety Inventory as the dependent variable.

	Not At All	Mildly but it didn't bother me much.	Moderately - it wasn't pleasant at times	Severely – it bothered me a lot
Numbness or tingling	0	1	2	3
Feeling hot	0	1	2	3
Wobbliness in legs	0	1	2	3
Unable to relax	0	1	2	3
Fear of worst happening	0	1	2	3
Dizzy or lightheaded	0	1	2	3
Heart pounding/racing	0	1	2	3
Unsteady	0	1	2	3
Terrified or afraid	0	1	2	3
Nervous	0	1	2	3
Feeling of choking	0	1	2	3
Hands trembling	0	1	2	3
Shaky / unsteady	0	1	2	3
Fear of losing control	0	1	2	3
Difficulty in breathing	0	1	2	3
Fear of dying	0	1	2	3
Scared	0	1	2	3
Indigestion	0	1	2	3
Faint / lightheaded	0	1	2	3
Face flushed	0	1	2	3
Hot/cold sweats	0	1	2	3
Column Sum				

Scoring - Sum each column. Then sum the column totals to achieve a grand score. Write that score here _____ .

Figure 2: Beck Anxiety Inventory Questionnaire and Scoring

First, have all participants fill out the Beck Anxiety Inventory (see Fig. 2 above). This sets the baseline, record these results (Table 2.a column 1). After this in table 1, the race, age, and gender for participants was recorded for obtaining the demographics. Participants were informed by the third party about the app and when to use the app, as well as to use it for five minutes each day. After getting a notification every day, participants use the app for five minutes, specifically one of the meditation recordings in the 16 to 18 years old category. Each day there is a different recording which focuses on a specific area for their age range, which is 16-18 years old (see Fig 3 below).

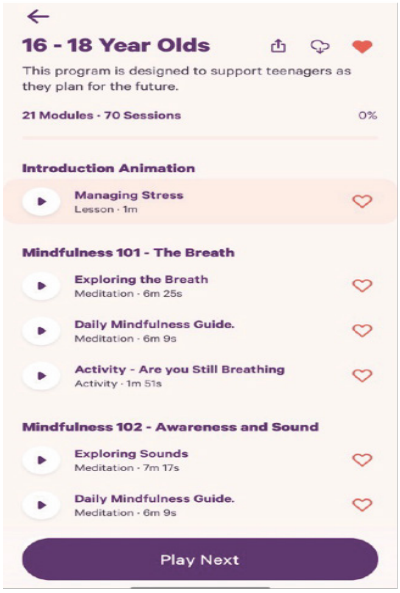


Figure 3: The 16-18 year old category that was used for my participants in the Smiling Mind app

After a continuous week of use, the participants fill out the Beck Anxiety Inventory to check if their ratings have changed post-intervention. The overall group rating was recorded in the data table in the second column (Table 2.a). The ratings were then analyzed using the Jamovi software (Jamovi version 2.3 2022) to perform a paired T-test to see if there was any changes of reduced or increased anxiety symptoms. All the mean anxiety scores before and after intervention were recorded in Data table 1.b to observe changes. At the end of the study, each participants' improvements for individual questions (rating out of 0 to 3) from the BAI was recorded into the table for before and after the app was presented (Table 2.b).

Data Tables

ID#	Race/ethnicity	Age (yrs)	Gender
p1	American	16	M
p2	Indian	18	M
p3	Asian	17	F
p4	Asian	16	M
p5	Indian	16	M
p6	Indian	17	F
p7	Indian	17	F
p8	Caucasian	16	M
p9	Russian	17	F
p10	Mexican	16	M
p11	Bangladeshi	16	M
p12	Latino	16	F
p13	Italian	18	F
p14	Pakistani	17	F
p15	Black	17	M
p16	Australian	18	M
p17	Jewish	18	F

Table 1) Participant Demographics for each participant to collect before intervention

	Pre-app intervention	Post-app Intervention
Mean BAI Score	40.05	32.82
Median BAI Score	42	36

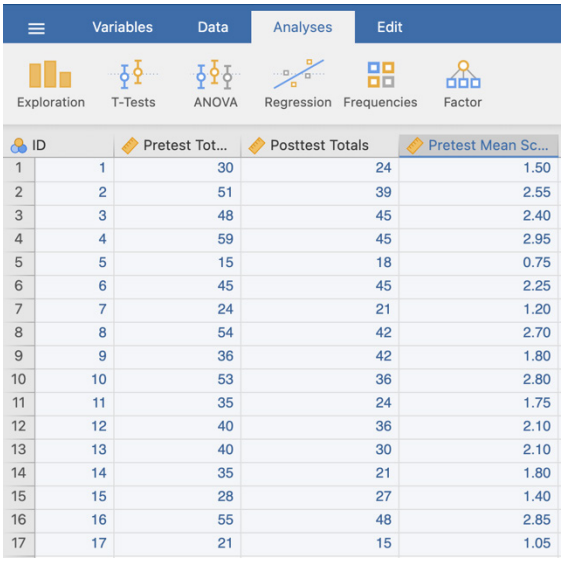
Table 2.a) The overall participant groups mean Beck Anxiety Inventory scores Pre and Post Smiling Mind app intervention for each symptom (0-3)

Pretest mean	Post test mean
1.5	1.2
2.55	1.95
2.4	2.25
2.95	2.25
0.75	0.9
2.25	2.25
1.2	1.05
2.7	2.1
1.8	2.1
2.8	1.8
1.75	1.2
2.1	1.8
2.1	1.5
1.8	1.05
1.4	1.35
2.85	2.4
1.05	0.75

Table 2.b) The overall participant ratings of each question (0-3) on the Beck Anxiety Inventory Pre and Post app use mean

Data Analysis

The data was analyzed using measures of central tendency (mean, median, and mode) using the pre and post Beck Anxiety Inventory Scores. Using the mean, standard deviation of both data sets was calculated for the paired t-test. This information and the t-statistic and p-value was received from Jamovi version 2.3 after the raw data was input (see Fig. 4 below).



	ID	Pretest Tot...	Posttest Totals	Pretest Mean Sc...
1	1	30	24	1.50
2	2	51	39	2.55
3	3	48	45	2.40
4	4	59	45	2.95
5	5	15	18	0.75
6	6	45	45	2.25
7	7	24	21	1.20
8	8	54	42	2.70
9	9	36	42	1.80
10	10	53	36	2.80
11	11	35	24	1.75
12	12	40	36	2.10
13	13	40	30	2.10
14	14	35	21	1.80
15	15	28	27	1.40
16	16	55	48	2.85
17	17	21	15	1.05

Figure 4: Raw data plugged in the Jamovi version 2.3 software for analysis

A paired t-test was used here because it establishes whether two means collected from the same sample differ significantly. In this case the t-test was evaluating the mean of the Anxiety inventory scores before and after the app intervention. This gives me the t-statistic and p-value, which displays the significance of my results. A p-value of 0.05, and a t-statistic of +2 was displayed from the data.

Results

The data from the Beck Anxiety Inventory was input to Jamovi 2.3, giving the central measures of tendency pre and post app intervention. The mean of the Beck Anxiety Inventory score for each symptom before the app was given was 2.00 out of 3.00 which is the rating they gave for each symptom. The mean score after was 1.64. The median of the pre-app score was 2.10 and the post-app score was 1.80 (see Fig. 5). The standard deviation also dropped from 0.669 to 0.547, and the standard error of the mean dropped from 0.162 to 0.133. This means that the data varied less from the mean after the app was given and the improvements in symptoms was consistent across the board.

Descriptives

	N	Mean	Median	SD	SE
Pretest Mean Scores	17	2.00	2.10	0.669	0.162
Posttest Mean Scores	17	1.64	1.80	0.547	0.133

Paired Samples T-Test

			statistic	df	p
Pretest Mean Scores	Posttest Mean Scores	Student's t	4.24	16.0	<.001

Table 3. Jamovi Version 2.3 data tables displaying data analysis results of centreal tendencies, Standard deviation, Standard Error of the Mean, and t and p values from the t-test

These values presented a t statistic of 4.24 and a p value of <.001 (see Table 3 above) which is much greater than .05 which is the basis for statistical tests to accept or reject the null hypothesis. The p-value was calculated based on the t-statistic displaying the differences between the two means. This clearly displays that these results were statistically significant. This means that there is a 99.99% chance that the difference was not out of chance, and the null hypothesis that the app does not improve teenagers' anxiety can be rejected based off my p-value because it does significantly improve their anxiety. These results are also generalizable to a greater population, so if this study were to be replicated by someone else using a much larger population with a representative sample, if would produce similar if not the same results.

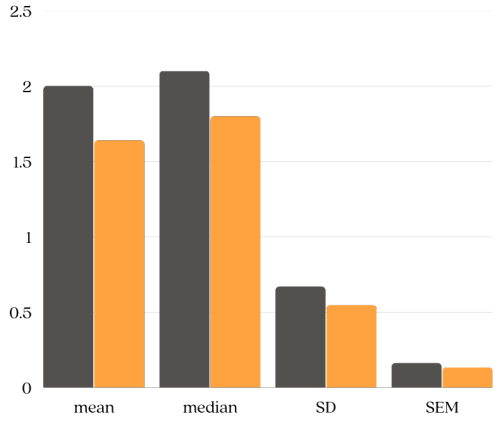


Figure 5: Mean average results for one question before and after app

Study Beck Anxiety Inventory Response for each question (from 0 to 3) Before and After App use

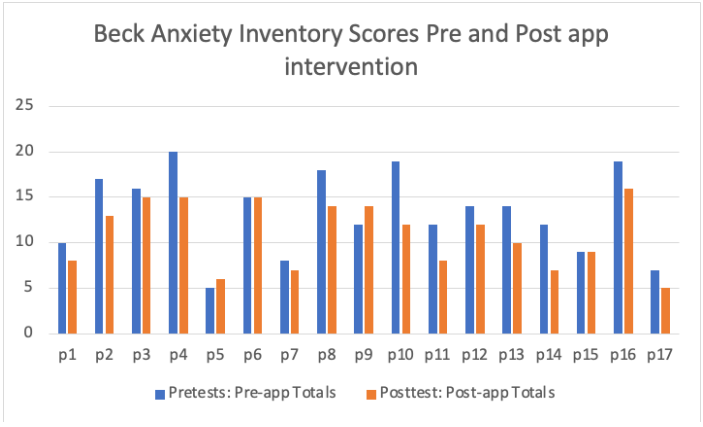


Figure 6: Mean average results for individual participants for observing discrepancies and outliers

Conclusion

The p-value of less than 0.001 that was calculated shows that the study's hypothesis was correct; this data shows that the *Smiling Mind* mobile app reduced anxiety symptoms based on the Beck Anxiety Inventory by a statistically significant amount.

Limitations of this study was the sample size, which was limited to 17 participants. The one-group design also was a limitation because the same group was tested before and after, there is more room for confounding variables such as genetics and personal experiences. Despite this, I did I one-group design because everyone experiences anxiety differently and keeping that standardized to the person also gives a good way to compare results. In further conduction of this study, trying a two-group design can help further confirm that the results are still accurate. The convenience sampling method was also a limitation as the participants that were most easy to access for participation were the ones in the study. This may potentially lead to some bias if the sample is not representative in recreations of this study. In the future versions of this experiment, the goal is to conduct more trials and address these limitations to produce better results and get a greater sample of data.

Discussion

One of the most important steps that can be taken to improve the quality and dependability of these results is to collect more data using a larger sample size with an efficient sampling technique. Utilizing more sophisticated statistical methods to assess the data can lead to more reliable findings. A useful contribution to the subject may be making use of the data to produce a more effective mental health app. This includes adding features and functions that have been shown to be successful in lowering anxiety and tension while building from the prior apps tested. To make sure that the app is in line with mental health practices supported by research, working with mental health specialists will be very important.

Informing kids about the value of mental health applications is also crucial. It's possible that students are unaware of these applications' potential advantages or how to use them. The usage of mental health applications and encouraging adolescents to get assistance when they need it may both be encouraged by school employees. Also adding mental health applications to the school system as a way for kids to cope with their anxiety can be a useful improvement to the selection of resources for mental health offered to teenagers. As stated, it's crucial to make sure that the confidentiality of students is safeguarded, and that usage of these applications is optional. Schools should also offer assistance and resources to students who require more specialized mental health treatment.

Bibliography

Andrea K. Graham, P. D. (2020, September 1). *A coached mobile app platform for the treatment of depression and anxiety among primary care patients*. JAMA Psychiatry. Retrieved November 20, 2022, from https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2766444

Garrido1, S., Cheers2, D., Boydell3, K., Nguyen5, Q. V., Schubert4, E., Dunne2, L., Meade6, T., Brain, 1M. A. R. C. S. I. for, & Garrido, C. A. S. (2019). *Young people's response to six smartphone apps for anxiety and depression: Focus Group Study*. JMIR Mental Health. Retrieved November 2, 2022, from https://mental.jmir.org/2019/10/e14385

Kenney, R., Dooley, B., & Fitzgerald, A. (2016). *Developing mental health mobile apps: Exploring adolescents ...* Health Informatics Journal. Retrieved November 3, 2022, from https://journals.sagepub.com/doi/10.1177/1460458214555041

Marshall, J. M., Dunstan, D. A., & Bartik, W. (1AD, January 1). *Smartphone psychological therapy during COVID-19: A study on the effectiveness of five popular mental health apps for anxiety and Depression*. Frontiers. Retrieved November 10, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2021.775775/full

Oliveira, C., Pereira, A., Vagos, P., Nóbrega, C., Gonçalves, J., & Afonso, B. (2021, April 6). *Effectiveness of mobile app-based psychological interventions for college students: A systematic review of the literature*. Frontiers. Retrieved January 24, 2023, from https://doi.org/10.3389/fpsyg.2021.647606

The jamovi project (2022). jamovi. (version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Virtual Reality for ADHD: Using Technology to Enhance Learning and Attention

Krish Katre

Nikola Tesla STEM High School

Keywords: ADHD, Virtual Reality, Anapanasati Meditation, Attention, Unity, C# Code

Abstract

Almost 10 percent of high school students are diagnosed with ADHD in the United States which causes a loss of concentration, causing them to fall behind academically. Through Virtual Reality technology, students can meditate for a short time period and increase their concentration levels. This project seeks to build a testable VR meditation app and investigate how the app improves concentration in ADHD and non-ADHD high schoolers. This was accomplished through developing a VR app in Unity by utilizing a variety of different texture packages (trees, water, and rocks), and implementing sound effects and interactivity through C# code. This app was then tested on 13 non-ADHD and 13 ADHD participants (variable 1) by providing a concentration test before using the app for 5 minutes and providing it again after the app was used (variable 2). The mean concentration change for ADHD participants was +15.77% and for non-ADHD participants it was +17.85%. The p-value for concentration change is 0.000016 which means the results are highly significant and not due to chance, along with the standard deviation being 2.6, so there are no outliers as all data points are within 2 standard deviations from the mean. These results accomplished the engineering goal by increasing concentration in ADHD students, and through further research this technology can be implemented in classrooms worldwide.

Introduction

The goal of this project is to understand the impact of virtual reality (VR) Anapanasati (breathing based) meditation on ADHD high school students and non-ADHD high school students, specifically the effects on their concentration. 9.6% of high school students nationwide are diagnosed with ADHD, which causes symptoms such as loss of concentration, twitching, fidgeting, and impulsivity. This has a negative correlation in school performance, especially math and reading courses. The paper ADHD in school: prevalence, multi-professional involvements and school training needs in an LEA discusses how teachers lack the training needed to deal with ADHD students and provide them with their basic needs in order to succeed in class (Wheeler 2008). Only 12% of schools in this study received this training, which is not enough to level the playing field between ADHD and non-ADHD students (Wheeler 2008). In order to solve this problem, we need to provide teachers with the right tools to make their lives easier and to help ADHD students succeed at the same time.

Medication has been a proposed solution to ADHD in the past. However, another study called Attention-Deficit/Hyperactivity Disorder, School Performance, and Effect of Medication found out that unless medication was provided at a very young age, it is rendered useless unless it is paired with "educational interventions" (Jangmo 2019). ADHD is rarely diagnosed at this age, so it is difficult to provide medication. So, another solution needs to be discovered. In the study Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool, ADHD symptoms were measured through various

tests in a VR environment. This VR environment was a classroom setting, and the various tests were called Aula Nesplora tests. These tests assessed attention, processing speed, impulsivity, and motor activity, symptoms that have a high correlation with ADHD (Areces 2018). If these tests could be used to assess these symptoms with a high statistical significance, could we also suppress these symptoms with the same VR technology?

The solution for this was to create a VR application that uses Anapanasati meditation in a virtual, calming environment. The advantages of Anapanasati meditation are that it strengthens the pre-frontal cortex, which is responsible for assisting in focus, impulsivity, and emotion control. Anapanasati meditation specifically focuses on mindfulness, and it can be practiced anywhere. The end goal is to implement this technology in classrooms worldwide, and Anapanasati meditation is the most accessible form, as it doesn't require anything but breathing. Two research questions were answered with this study: 1) Can we increase the concentration of ADHD and non-ADHD students after meditating using the VR technology for 5 minutes? 2) Do these results differ between ADHD and non-ADHD students? My goal from this app was to improve concentrations of ADHD and non-ADHD students and test it on a number of high school students. The hypothesis for this study was that five minutes of using the application should improve ADHD concentrations by 5%, and there should be a significant difference between ADHD and non-ADHD students.

Methodology:

Development of Application

To develop this app, the software *Unity 3D* was used. This is a game development software that is a popular tool used by VR game developers. I was assisted in this research process by one of my mentors, Prof. Gaby de Jongh, who specializes in VR technology. I imported various texture packs to set the scene of my VR app (**trees, sky, rocks, grass, soil, etc.**). I then drew a basic terrain using Unity terrain tool, which created mountains and the landscape of the scene (see **Figure 1**). These mountains were carefully drawn to leave space for a lake in the middle. Multiple trials were done in drawing these mountains, until the final terrain was created. Next, I painted the mountains with the **grass, soil, and rock** texture packs installed earlier. This was done using the Unity paint tool and was repeated multiple times to make the scene look as realistic as possible. I then filled the hole with the **water** texture pack to make a lake in the middle and surrounded it with the **flower, rock, and tree** texture packs (see **Figure 2**). After this step, the scenery was created. To create an ambience, I downloaded various .mp4 files such as forest ambience, meditation music, lake sound effects, and a guided meditation voice to regulate breathing. This was done to enhance the effectiveness of Anapanasati meditation and to guide the participant's meditation journey. The final step and perhaps the most important step is the interactive C# code. This code runs a particle system, which triggers based on Oculus Touch Controller input. Holding the controller's grip button will make the particles come in and releasing it will make the particles go away from the user. This provides a visual effect of the breathing and can guide the participant even more, so they are on cadence with the guided meditation voice (see **Figure 3** for a more in-depth analysis of the code).

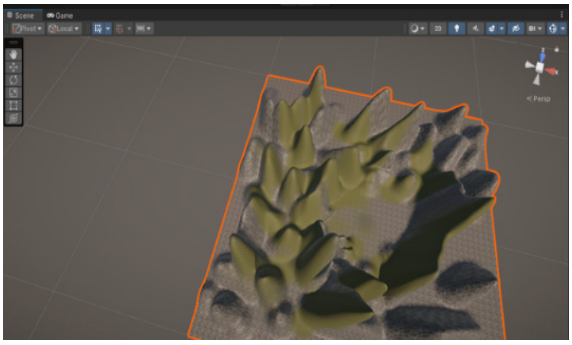


Figure 1: Landscape drawn using terrain tool.



Figure 2: Lake, rocks, and trees surrounding the lake.

Set Breathing States which will be used as cases based on controller input

Move particles in and out based on controller input

```
if (Input.GetButton("XRI_Right_GripButton") && Input.GetButton("XRI_Left_GripButton"))
{
    breathingState = 1;
}
else if (!Input.GetButton("XRI_Right_GripButton") && !Input.GetButton("XRI_Left_GripButton"))
{
    breathingState = 2;
}
else
{
    breathingState = 0;
}
//states
switch (breathingState)
{
    case 0:
        emitIn.enabled = false;
        emitOut.enabled = false;
        break;
    case 1:
        emitIn.enabled = true;
        emitOut.enabled = false;
        break;
    case 2:
        emitIn.enabled = false;
        emitOut.enabled = true;
```

Figure 3: Most important code segment

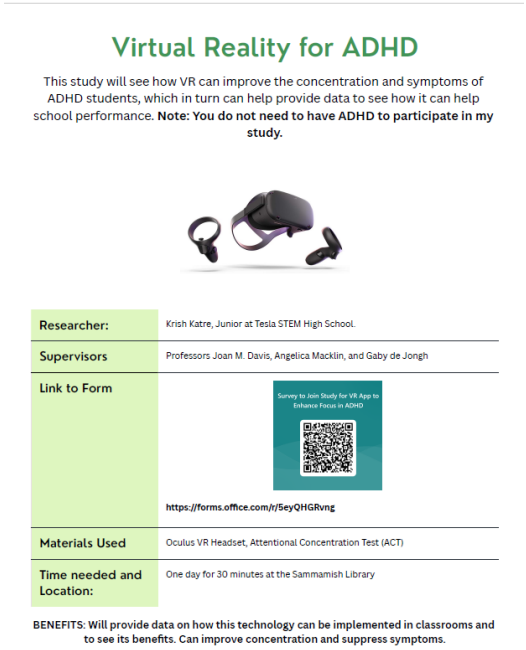


Image of app being used.

App Testing

To test this app, I obtained 26 total participants (n = 26) with an even split of 13 ADHD participants and 13 non-ADHD participants. These participants were all high school students in the grades 9-12, in the Pacific Northwest. This was done through the creation of a flyer which had a Microsoft Forms QR Code attached to it, which upon completion confirms the student's participation in the study (see **Figure 4**). These flyers were posted to local libraries

(Sammamish and Redmond) and schools within a 10-mile radius (Eastlake High School and Redmond High School). Once all responses were collected, they were separated into two groups: ADHD and non-ADHD. They were then assigned random numbers through a random number generator (1-13 inclusive) and sent an informed consent form. This was signed by parents and shredded after they were signed to maintain confidentiality.



Virtual Reality for ADHD

This study will see how VR can improve the concentration and symptoms of ADHD students, which in turn can help provide data to see how it can help school performance. **Note: You do not need to have ADHD to participate in my study.**

Researcher:	Krish Katre, Junior at Tesla STEM High School.
Supervisors	Professors Joan M. Davis, Angelica Macklin, and Gaby de Jongh
Link to Form	Survey to Join Study for VR App to Enhance Focus in ADHD https://forms.office.com/r/5eyQHGRvng
Materials Used	Oculus VR Headset, Attentional Concentration Test (ACT)
Time needed and Location:	One day for 30 minutes at the Sammamish Library

BENEFITS: Will provide data on how this technology can be implemented in classrooms and to see its benefits. Can improve concentration and suppress symptoms.

Figure 4: Recruitment flyer for participation in study

All participants were tested in the local library (Sammamish) to use the app and take the concentration test. They were instructed on how to use the app, controller, and headset. They first took the Total Brain Focus test, which is clinically validated (see **Figure 5** for more details). The first 2 parts of the test are to assess cognitive attention, and they are based off the Stroop Test (used by law enforcement to assess DUIs). In the first part, a word is displayed, and the participant clicks the answer choice that matches the word, regardless of what color the word is. The second part of the test is similar, but the participant clicks the color of the word, which will be different than the word itself. The last part of the test is to assess attention span or attention endurance. This is done through displaying a series of letters, and if two consecutive letters appear, the participant clicks match. This part is the longest in the whole test. Every attempt of the test gives a new random set of colors and letters, so no two tests are identical. Their score is recorded in either Data Table 1 or Data Table 2 (ADHD and non-ADHD, respectively) as a percentage. They then used the app for 5 minutes and followed the guided meditation voice. Once the 5 minutes were completed, they immediately took the focus test again, and the new score was recorded in the same row of either Data Table 1 or Data Table 2. The equipment was cleaned with disinfectant wipes, and the next participant was tested on. This was a single blind study. The data tables below show the results:



Select Actual Meaning of Word

Select Actual Color of Word

2 same letters consecutively, click match

Figure 5: Total Brain Concentration Test Details

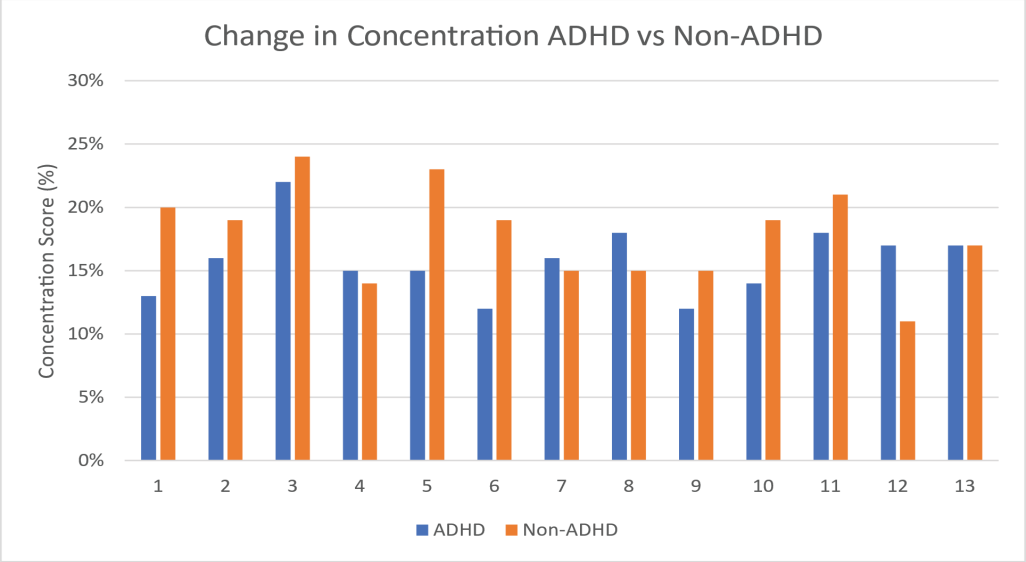
Table 1 – ADHD participant concentration results before and after using VR app

Participant	Grade	Date	Concentration Initial (%)	Concentration Final (%)
1	11th	2/24	58%	71%
2	12th	2/26	63%	79%
3	11th	2/27	50%	72%
4	11th	2/27	39%	54%
5	11th	2/28	43%	58%
6	11th	3/14	47%	59%
7	12th	3/14	45%	61%
8	9th	3/15	42%	60%
9	11th	3/18	57%	69%
10	12th	3/18	45%	59%
11	10th	3/19	47%	65%
12	10th	3/20	51%	68%
13	12th	3/24	41%	58%

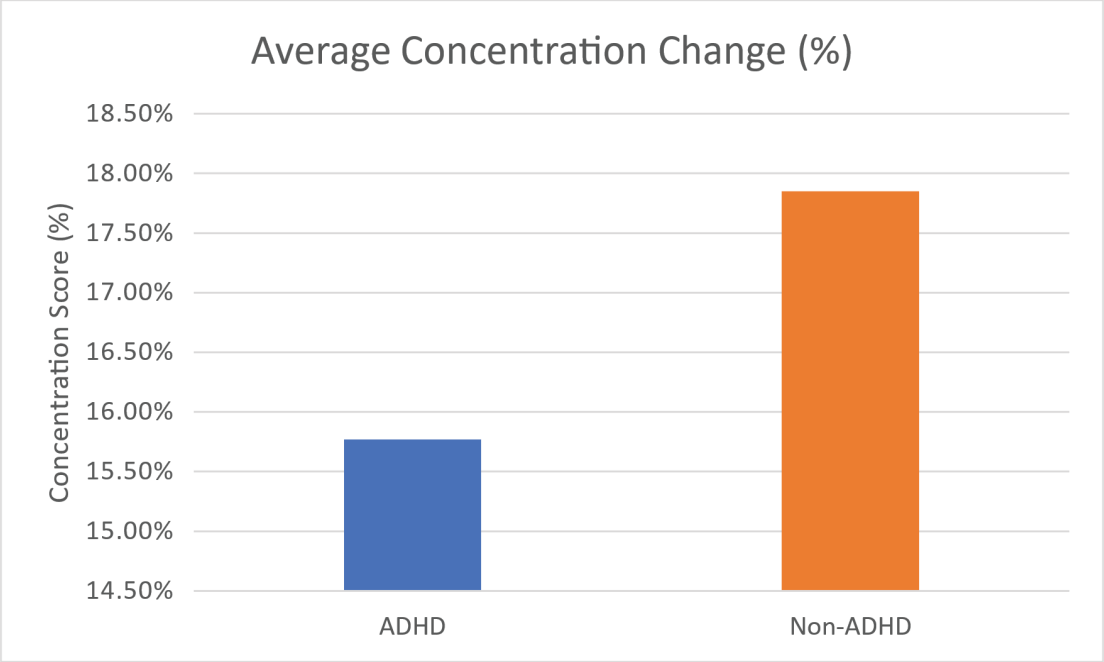
Table 2 – Non-ADHD participant concentration results before and after using VR app

Participant	Grade	Date	Concentration Initial (%)	Concentration Final (%)
1	11th	2/24	67%	87%
2	11th	2/25	56%	75%
3	11th	2/25	50%	74%
4	11th	2/26	56%	70%
5	11th	2/27	54%	77%
6	11th	3/14	51%	70%
7	10th	3/15	55%	70%
8	10th	3/15	61%	76%
9	9th	3/18	58%	73%
10	12th	3/19	41%	60%
11	11th	3/20	50%	71%
12	11th	3/24	59%	70%
13	10th	3/25	52%	69%

The final and initial concentration test results were subtracted to get the change in concentration. These results were displayed as a bar graph in **Graph 1**. These changes in concentration were averaged to calculate the mean change in concentration for each group. These results are displayed in **Graph 2**. The graphs were made in Microsoft Excel, and no participant's data was left out. The mean concentration change for ADHD was +15.77% and for non-ADHD it was +17.85%.



Graph 1: Concentration Change for ADHD and non-ADHD participants



Graph 2: Mean Concentration Change for ADHD and non-ADHD participants (15.77% and 17.85% respectively)

Data Analysis

This data was analyzed using two different methods, Chi-Square Test and Standard Deviation from the mean. The formula for Chi Square Test is $\sum \frac{(Observed-Expected)^2}{Expected}$. This test was done to see if my results were statistically significant or due to random chance. From this test, I got a p-value < 0.05 ($p = 0.000016$). This means we can reject the null hypothesis and my results are very significant.

A standard deviation test was also done to check for outliers in the data. Outliers could skew the data and make it less reliable. The formula for calculating this is $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$. After using this formula, the standard deviation was 2.66 (see **Figure 6** for calculation).

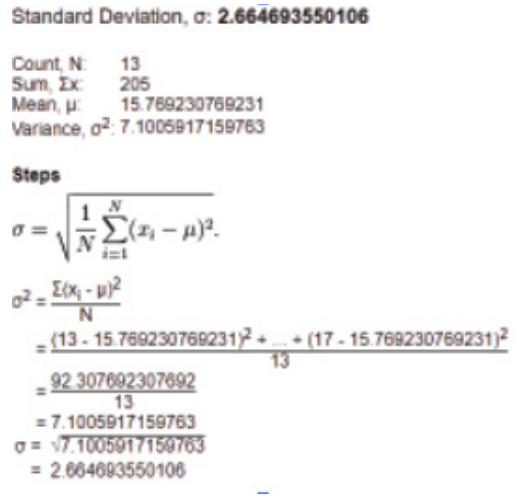


Figure 6: SD Calculation

No data points were outside of 3 standard deviations from the mean, so there were no outliers in my data. This means all my data is reliable and accurate.

Discussion

The hypothesis was partially incorrect. The hypothesis stated that the concentration would only increase by 5%, however it increased by 15 – 18%. There was also not a statistically significant difference between ADHD and non-ADHD participants (*p = 0.5)*. The engineering goals were met as I increased concentration using my app and successfully tested the app. There were some limitations in the study as there were only 13 participants in each category. The sample was also not fully representative as it was slightly male dominated (16 males vs. 10 females). There were also some unintentional distractions, as this study took place in a public library. But this was also a positive, as in a real classroom setting it will not be quiet. It provides a realistic element to the study, so it may have been useful in my results. Some future applications of this study could be to incorporate more ADHD specific tools to lower the p-value of 0.5. This can include fidget toys, stress balls, etc. This would tailor my app to ADHD students more, so we see more of a significant difference. Another application could be to add more scenes to test the effect of the environment on concentration. This could be a beach scene, snow scene, ocean scene, etc. There could also be more interactive games added to the study, such as memory games and Simon Says. These games help ADHD students in the real world. Finally, we can test the effects of normal meditation versus VR meditation. This could tell us if VR meditation is feasible, or will normal meditation do the job. But there are promising results from a basic meditation app, and with more research, this technology can be implemented in classrooms worldwide.

Bibliography

Linda Wheeler, Peter Pumfrey, Peter Wakefield & Wendy Quill(2008) ADHD in schools: prevalence, multi-professional involvements and school training needs in an LEA, Emotional and Behavioral Difficulties, 13:3, 163-177, DOI: 10.1080/13632750802253137

Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool Areces D, Dockrell J, García T, González-Castro P, Rodríguez C (2018) Analysis of cognitive and attentional profiles in children with and without ADHD using an innovative virtual reality tool. PLOS ONE 13(8): e0201039. https://doi.org/10.1371/journal.pone.0201039

Zijlema, Wilma & Kluizenaar, Y. & Kamp, I. & Hartman, C.. (2021). Associations between road traffic noise exposure at home and school and ADHD in school-aged children: the TRAILS study. European Child & Adolescent Psychiatry. 30. 10.1007/s00787-020-01521-8.

Fang Y, Han D, Luo H. A virtual reality application for assessment for attention deficit hyperactivity disorder in school-aged children. Neuropsychiatr Dis Treat. 2019 Jun 4;15:1517-1523. doi: 10.2147/NDT.S206742. PMID: 31239686; PMCID: PMC6559774.

Nguyen MN, Watanabe-Galloway S, Hill JL, Siahpush M, Tibbits MK, Wichman C. Ecological model of school engagement and attention-deficit/hyperactivity disorder in school-aged children. Eur Child Adolesc Psychiatry. 2019 Jun;28(6):795-805. doi: 10.1007/s00787-018-1248-3. Epub 2018 Nov 2. PMID: 30390147.

Behavioral and Social Sciences

Light Therapy as a Biological Intervention for Irregular Circadian Rhythms

Siri Varkala

Nikola Tesla STEM High School

Abstract

Circadian rhythm is the natural, internal process that regulates the sleep-wake cycle and repeats itself every 24 hours. Circadian rhythm disorders can lead to a weakened immune system, cardiovascular disease, and cognitive and behavioral disorders. This project seeks to investigate how well light therapy (IV) can regulate circadian rhythm (DV), and whether a light box or a dawn simulator (DS) is more effective. This experiment was composed of three groups: control and two experimental. All participants recorded their sleep and wake times, the number of times they felt the need to eat or sleep at an unusual time, their percentage of hours of sleep compared to prior to the study over two weeks, and their sleep satisfaction. Participants also indicated confounding variables such as whether they turned down the brightness on screens starting 1 hour before their preferred bedtime, went to bed and woke at the same time every day, avoided strenuous exercise for one hour before bed, and avoided caffeine and alcohol for 4 hours before bed. The results were statistically significant (an average r2 value of 0.8376), with a 5.5% increase in sleep for those in the light box group with a small amount of the confounding factors mentioned. The light box was more effective than the DS with a 5.28% increase in sleep compared to the DS group's percent increase in sleep of 4.9. The outcome of this study determines that the light box is more effective in regulating circadian rhythm.

Introduction

The purpose of this study is to determine the extent to which light therapy can be used to regulate irregular circadian rhythm, and whether a light box or a dawn simulator would be more effective.

Circadian rhythm is the natural, internal process that regulates our sleep-wake cycle, brain wave patterns, hormone production, cell regrowth, temperature, metabolism, and other activities (see Fig. 1) and repeats itself roughly every 24 hours, though with the invention of the light bulb about 150 years ago, it has shifted closer to twenty-five. Circadian rhythm disorders – such as insomnia, sleep-wake phase disorder, night-eating disorder, and jet lag – can lead to a weakened immune system, cardiovascular disease, and cognitive and behavioral disorders (Circadian Rhythm Disorders - What Are Circadian Rhythm Disorders? | NHLBI, NIH, 2022). Though circadian disorders are rare – affecting only 0.17% of the population, adolescents, young adults, and those with mental health conditions are more susceptible – at about 21% (Sleep Science and Sleep Disorders | National Heart, Lung, and Blood Institute (NHLBI), 2013) – likely due to school or work obligations.

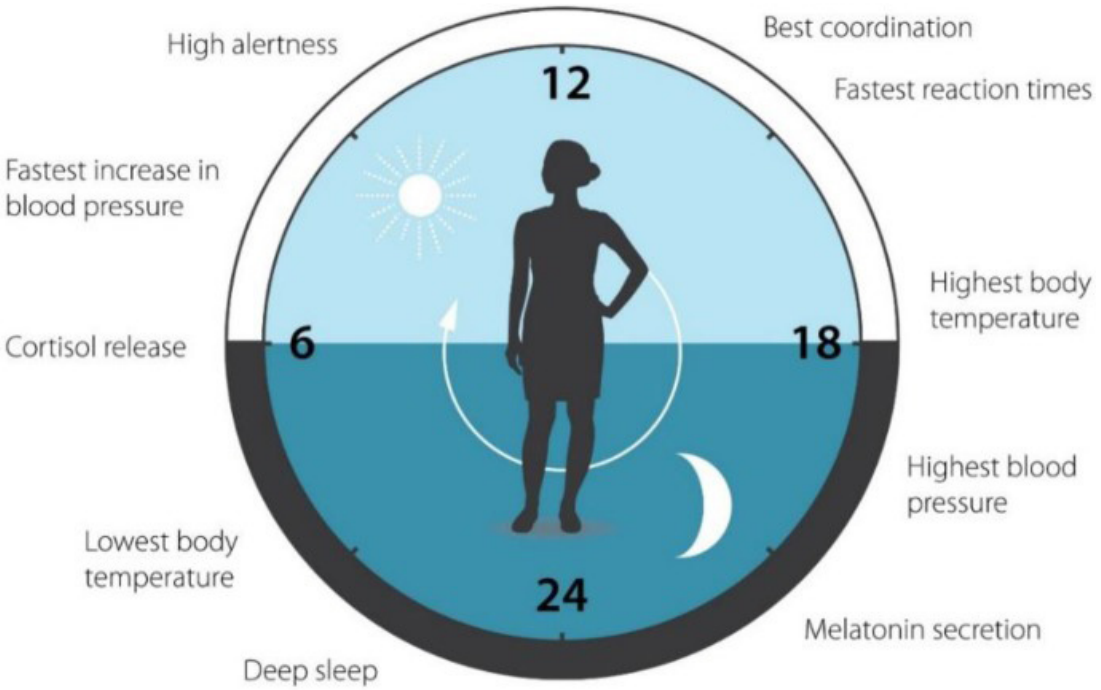
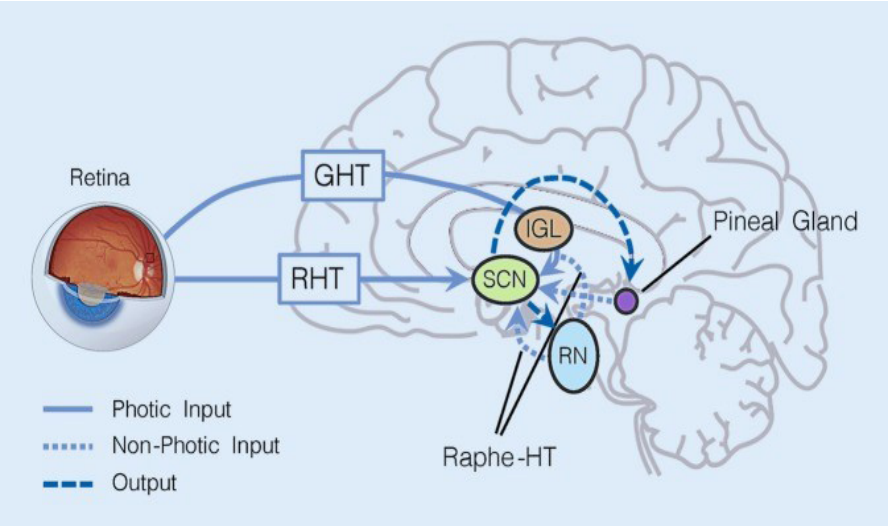


Figure 1. The circadian clock impacts many aspects of our physiology. This clock helps to regulate sleep patterns, feeding behavior, hormone release, blood pressure and body temperature.

The principal factor in regulating circadian rhythm is light. Exposure to bright light during the day can help regulate the sleep cycle especially, which is possible through the suprachiasmatic nucleus (see Fig. 2). The suprachiasmatic nucleus (SCN) is the master clock controlling circadian rhythm, with a near-24-hour rhythm of electrical activity, even without the guidance of environmental cues. This circadian activity reflects the pattern of expression of core genes, called clock genes, regulated by autoregulatory feedback loops. The circadian influences of SCN neurons are distributed to target organs of the body by efferent neural and humoral signals, particularly circulating melatonin. The SCN controls melatonin secretion which regulates one's sleep cycles.



F*igure 2. Input and output pathways to/from the suprachiasmatic nuclei (SCN). The photic input pathways that relay information about the intensity and spectral composition of ambient light are the retinohypothalamic tract (RHT) and the geniculohypothalamic tract (GHT), which connects the retina and SCN via the intergeniculate leaflet (IGL) in the thalamus. Additionally, the SCN also receives non-photic information from the raphe nuclei (RN) via the raphe-hypothalamic tract (raphe-HT) and from the pineal gland. The main output is from the SCN to the serotonergic raphe nuclei (RN, receive information about the phase of the circadian clock and regulate vigilance state of the body) and the pineal gland, where melatonin is produced. Input and output pathways form reciprocal loops.*

Two methods of light therapy commonly used for circadian rhythm disorders are light boxes and dawn simulators. Normal indoor lights are about 1% as bright as the sun; light boxes are brighter, around 10% and therefore more likely to affect circadian rhythm. Dawn simulators on the other hand, slowly shift color and brightness in the same pattern as a natural dawn, set to reach their peak brightness at the time one expects to wake up. Both have been tested in clinical trials and have been proven to be effective, with merely 90 minutes of bright light therapy for five days seeming to achieve a significant improvement in a constellation of circadian, sleep, health, and cognitive factors (Rubiño et al., 2020) with a g value of .41 indicating a strong association. However, no research has ever been done comparing the two in a regulated environment. That was the purpose of this study: To determine the extent to which light therapy could be used to regulate irregular circadian rhythm, and whether a light box or a dawn simulator would be more effective.

Methodology:

This study's participants were recruited through flyers that they saw at their school, workplace, or a business that they frequent. A random number generator was used to select fifteen people – ten female and five male – from volunteers fitting the following criteria:

- Aged between 12 and 50
- Is not severely mentally ill.
- Has an irregular circadian cycle (regardless of whether they have been diagnosed with a circadian disorder)
- Did not have a recent eye surgery or eye problem preventing exposure to bright light.
- Is not taking medication making the skin more sensitive to light (e.g., Tablets against malaria)
- Is not taking medication for their circadian disorder.
- Is not pregnant.
- Can understand English and able to give consent.
- Agrees to participate in the study and signs the informed consent form (along with their guardian if they are a minor)

The fifteen participants were then split evenly into three groups – the control group, experimental group one, and experimental group two – with two females and three males in both experimental groups and one female and four males in the control group to maintain the effectiveness of a representative sample. No names were collected, and the data collection was entirely anonymous, with the participants being referred to and submitting their data as an assigned letter and number through Microsoft Forms every day they participated in the study. See Tables 1, 2, and 3 for the data collected. Participants were informed of the purpose of the study and everything it would entail and gave their informed consent to participate. All informed consent forms were shredded upon completion.

The control group and both experimental groups filled out tables 1, 2, and 3 for every day they participated in the study as well, while undergoing light therapy to help with their moods and regulate their circadian rhythms.

Table 1 (see Table 1) measured the change in participants' circadian rhythm: the percentage of their recommended calorie/carb intake that day, the number of times they ate at an unusual time, and percent of hours of sleep compared to before entering the trial.

Table 1. Changes in circadian rhythm

Participants	Percent of calorie/carb intake compared to beforehand	Number of times you felt the need to eat/ sleep at an unusual time	Percent of hours of sleep compared to before entering the study
Control group:			
Participant 1:	%		%
Participant 2:	%		%
Participant 3:	%		%
Participant 4:	%		%
Participant 5:	%		%
Experimental Group 1:			
Participant 6:	%		%
Participant 7:	%		%
Participant 8:	%		%
Participant 9:	%		%
Participant 10:	%		%
Experimental Group 2:			
Participant 11:	%		%
Participant 12:	%		%
Participant 13:	%		%
Participant 14:	%		%
Participant 15:	%		%
Additional Notes (if applicable):			

Table 2 (see Table 2) measured other factors that might interfere with the attempt to regulate one's circadian rhythm, such as whether or not they dimmed the lights and turned down the brightness on screens starting one hour before their preferred bedtime, went to bed and woke at the same time every day (including weekends), refrained from doing strenuous exercise for one hour before bed, and avoided caffeine and alcohol for 4 hours before bed.

Table 2. Efforts to regulate circadian rhythm.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Time of awakening	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _
Time you went to sleep	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _	_ _ : _ _
Did you…							
Dim the lights and turn down the brightness on screens starting an hour before bed?	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
Avoid doing strenuous exercise an hour before bed?	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
Avoid looking at electronics an hour before bed?	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
Avoid caffeine for 4 hours before bed?	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
Avoid alcohol for 4 hours before bed?	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No	Yes / No
Additional Notes (if applicable):							

Table 3 (see Table 3) measured sleep satisfaction on a scale of one to ten, one being horrible and ten being excellent.

Table 3. Sleep satisfaction

Sleep Satisfaction (1-10)	
Additional Notes (if applicable):	

No one who suffers from an eye disease, has a history of eye diseases in their family, or was experiencing any temporary eye problems participated in the study. Participants were told to cease light therapy immediately and consult their ophthalmologist if they were experiencing any eye problems.

Experimental group one's light therapy consisted of having a dawn simulator on their nightstand about three feet away from their heads. Dawn simulators work by shining increasingly bright light at a distance of about three feet from one's head while sleeping, with a peak brightness of 250 lux – a measure of luminance – and the timing set to hit peak brightness at the time that they expect to wake up.



Figure 3. Miroco light therapy box, the light box used in this study.

Experimental group two's light therapy consisted of sitting in front of a light box (see Fig. 3) emitting 4,000 lux for 45 minutes a day, with the box sitting 16 to 24 inches away from one's face, and not directly into their eyes.

In addition, all three groups tried to dim the lights and turn down the brightness on screens as a part of their wind-down routine starting one to two hours before their preferred bedtime. When this was not possible, they wore blue-light-blocking or red- or orange-tinted glasses, or used candlelight. Additionally, they endeavored to go to bed and wake at the same time every day (including weekends), refrain from doing strenuous exercise for one hour before bed, avoid looking at electronics for at least an hour before bed, and avoid caffeine and alcohol for 4 hours before bed. It is understandable that it is not always possible to follow these rules, so the participants recorded the time they went to sleep, the time they woke up, and whether they did strenuous exercise during the one hour before bed, looked at electronics an hour before bed, avoided caffeine for 4 hours before bed, and/or avoided alcohol for 4 hours before bed.

Both methods of light therapy had been tested before in clinical trials and were deemed safe for human use.

Both methods used broad-spectrum lights that do not produce UV light, so there was no risk of getting a sunburn or any interference with Rx meds that required participants to stay out of the sun.

The LED bulbs have no UV output, so there is also no risk of burning or any type of damage to the skin or eyes. It is still recommended to double-check with a pharmacy or an MD to make sure.

Benefits of completing the light therapy include regulation of sleep cycles, decrease in irregular sleeping and eating, and improvements in mood, skin, and vision.

Conclusions:

In Table 1, the variable "Percent of calorie/carb intake compared to beforehand" was disregarded entirely due to participants not being willing to calculate how much they were eating per day and simply estimating the answer.

	CONTROL		EXPERIMENTAL 1		EXPERIMENTAL 2	
% of sleep compared to before		Need to eat or sleep at an unusual time	% of sleep compared to before	Need to eat or sleep at an unusual time	% of sleep compared to before	Need to eat or sleep at an unusual time
Mean	102.4%	3.88	104.9%	3.57	105.28%	3.26
Median	100%	4	100%	3	100%	3
Mode	100%	4	100%	3	100%	3
Sleep Satisfaction	6.628		7.547		7.824	
Standard Deviation	18.41	1.1582	18.1527	1.1875	13.3487	1.0830
r^2-value			0.8577	0.8523	0.8077	0.8325

Figure 4. Analysis of overall data

As seen in Fig. 4, the mean, median, and mode for the other two variables used to measure regulation in circadian rhythm show decrease in the need to eat or sleep at an unusual time and increase the amount of sleep participants get with the implementation of light therapy, with the light box being the most effective.

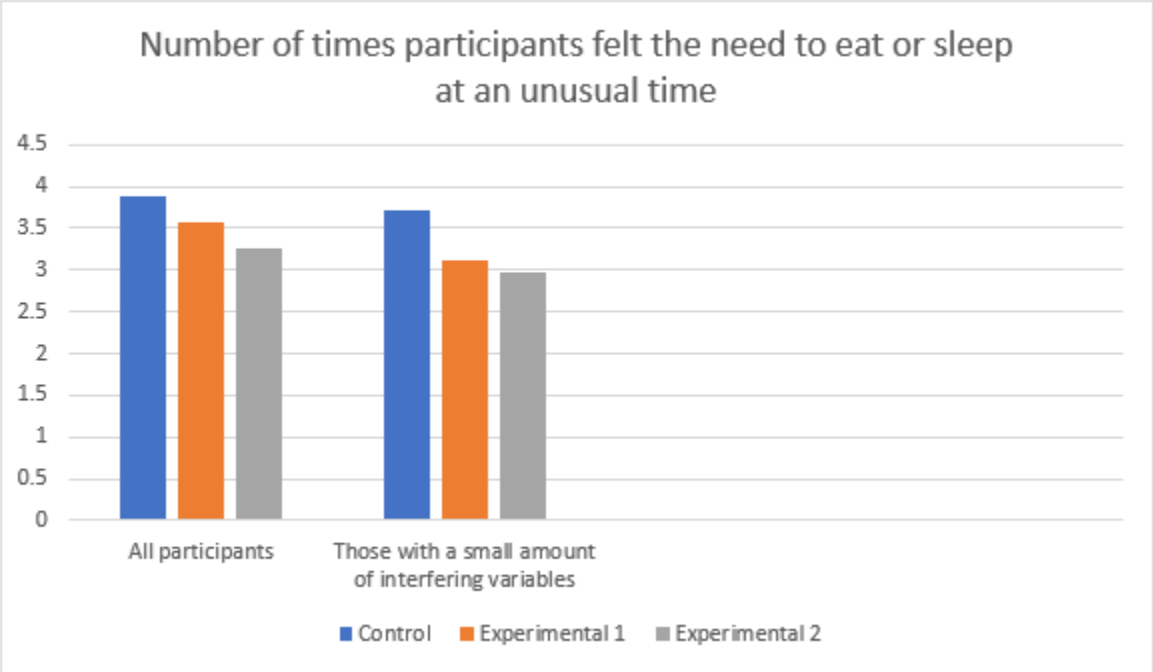


Figure 5. *Comparison of data for the number of times participants felt the need to eat or sleep at an unusual time between all participants and those with two or less of the interfering factors mentioned in Table 2.*

The average number of times experimental group 2 felt the need to eat or sleep at an unusual time was 3.26, compared to experimental group 1's average of 3.57. Those with a smaller number of interfering variables – two or less – had an average of 2.98 in the light box group and 3.12 in the dawn simulator group. This can be compared to the control group's baseline of 3.88 for all participants, and 3.72 for those with a small amount of the interfering variables. (See Fig. 5)

The standard deviation for the number of times a day participants felt the need to eat or sleep at an unusual time was 1.188 for experimental group 1 and 1.08 for experimental group 2. This can be compared to the control group's standard deviation of 1.158 for this variable. (See Fig. 4)

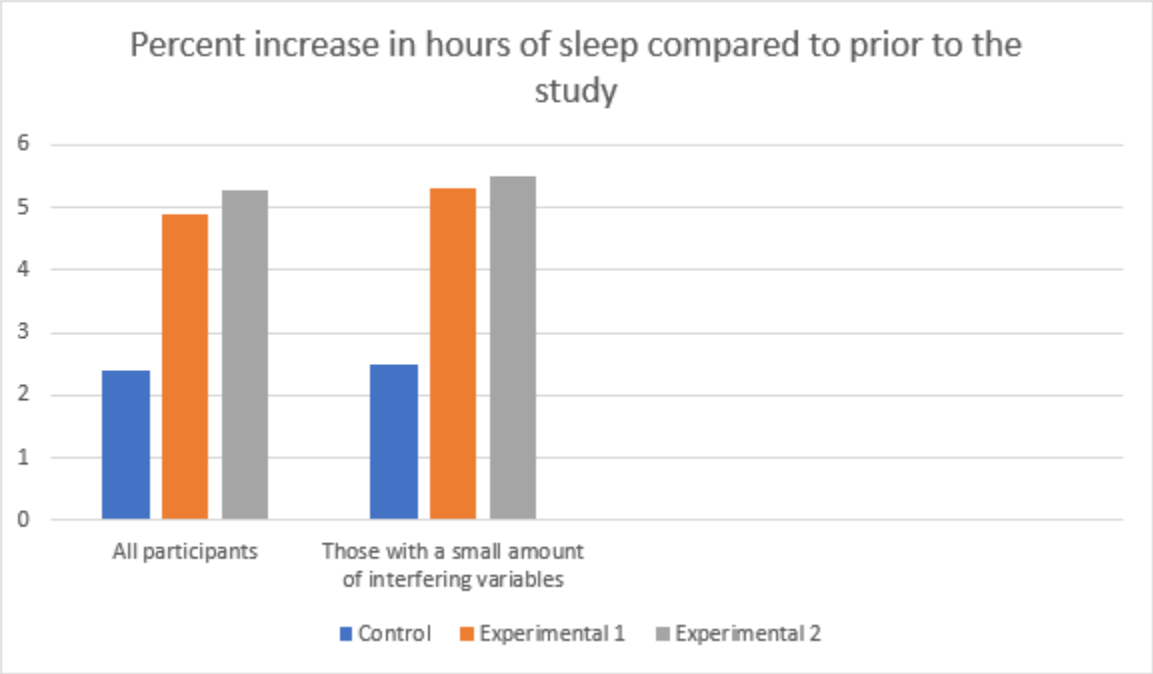


Figure 6. *Comparison of data for the percent increase in hours of sleep compared to prior to the study between all participants and those with two or less of the interfering factors mentioned in Table 2.*

The percentage of hours of sleep increased by 5.28% in the light box group and 4.9% in the dawn simulator group. Among those with the smallest number of interfering variables, the light box group had an average of a 5.5% increase in sleep and the dawn simulator group had an average increase of 5.32%. The control group had a percent increase in sleep of about 2.4%, despite having done no light therapy, likely due to an outlier, and those with a smaller number of interfering variables had a percent increase in sleep of 2.73. This was likely due to a decrease in their workload, influencing the amount of sleep they were able to get. (See Fig. 6)

The standard deviation for the percent of hours of sleep participants got every night compared to prior to entering the study was 18.152 for experimental group 1 and 13.349 for experimental group 2. This can be compared to the control group's standard deviation of 8.41 for this variable. (See Fig. 4)

While the regulation of circadian rhythm was not particularly large, it should be kept in mind that the study only went on for two weeks, and with a light intensity of 4,000 lux for the light box. Most studies regarding light therapy for circadian rhythm went on for a few months and the light intensity was closer to 10,000 lux. A smaller intensity was used for this study because participants usually experience symptoms such as dizziness and nausea with such high light exposure. The standard deviation decreases with the implementation of light therapy (as seen in Fig. 4) – with the exception of experimental group 1's percent of sleep due to an outlier – which indicates a more regular circadian cycle, which was the intended outcome of the study.

The data collected was proved to be statistically significant, with an average r2 value of 0.8376, indicating a fairly strong causation. The r2-values for the decrease in the number of times a day participants felt the need to eat or sleep at an unusual time were 0.852 for experimental group 1 and 0.833 for experimental group 2. The r2-values for the increase in the percent of hours of sleep participants got every night compared to prior to entering the study were 0.858 for experimental group 1 and 0.808 for experimental group 2. (Fig. 4)

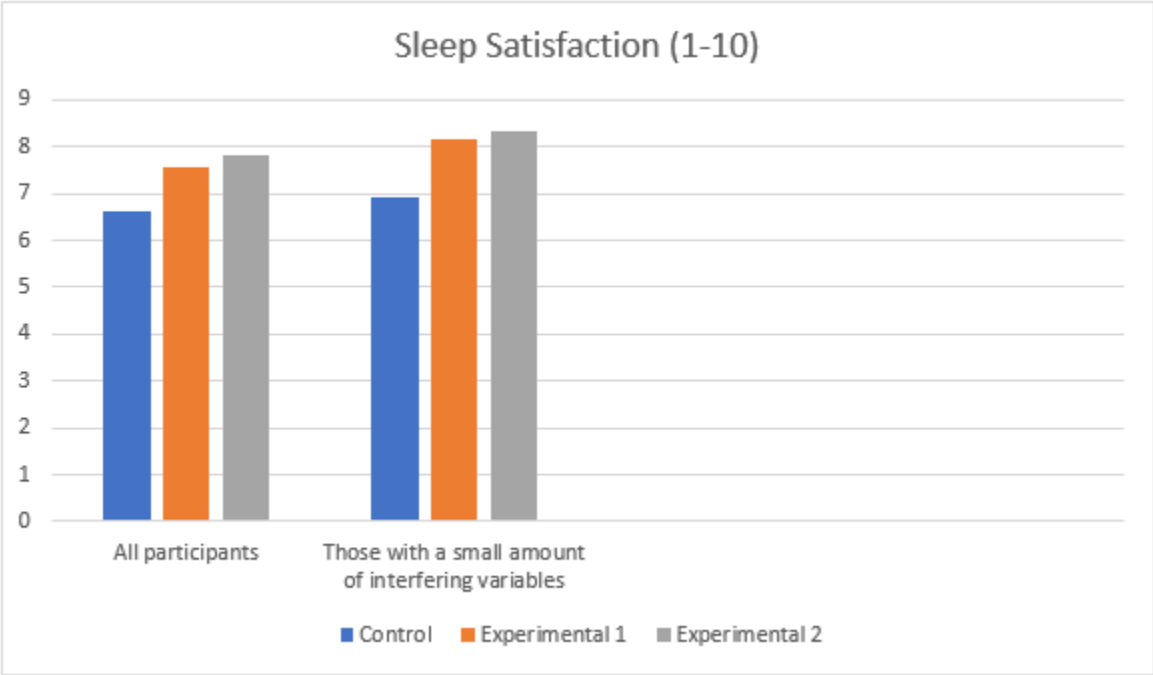


Figure 7. *Comparison of data sleep satisfaction on a scale of one to ten between all participants and those with two or less of the interfering factors mentioned in Table 2.*

In addition to the number of times participants felt the need to eat or sleep at an unusual time and the percent of sleep compared to before entering the study, participants also recorded their sleep satisfaction for every day they participated in the study. Both experimental groups reported an increase in quality of sleep, with an increase of 0.919 for the dawn simulator group and a 1.196 increase for the light box group, from the control group's average sleep satisfaction of 6.628. Among those with two or less of the interfering variables, there was an increase of 1.222 for the dawn simulator group and an increase of 1.405 for the light box group from the control group's mean sleep satisfaction of 6.936. (See Fig. 7)

Limitations of the study include its small sample size – only 15 – and the short duration of the study – only 2 weeks, compared to the several months that light therapy was tested over in prior studies. Had the study gone on longer, there would likely have been an even more significant regulation in participants' circadian rhythms. Another way this study was different than prior ones is that the luminance emitted by the light boxes was only 4,000 lux for 45 minutes a day. Many prior researchers in prior studies used light therapy with as much as 10,000 lux for 45 minutes a day. This was avoided in this study because the participants who were exposed to 10,000 lux for 45 minutes a day experienced symptoms such as dizziness, jitteriness, and nausea.

In the future, the study could be repeated with a larger, more representative group of participants – in terms of age and ethnicity – and for a longer amount of time than in this study. Additionally, the luminance of the light box could have been increased to 6,000 lux, with clear instructions to the participants to withdraw from the study if experiencing any eye problems or other symptoms.

Figure 8.
Epworth Sleep Quality Scale

Situation	**0** Would never nod off	**1** Slight chance of nodding off	**2** Moderate chance of nodding off	**3** High chance of nodding off
Sitting and reading				
Watching TV				
Sitting, inactive, in a public place (e.g., in a meeting, theater, or dinner event)				
As a passenger in a car for an hour or more without stopping for a break				
Lying down to rest when circumstances permit				
Sitting and talking to someone				
Sitting quietly after a meal without alcohol				
In a car, while stopped for a few minutes in traffic or at a light				
Add up the numbers you selected for the eight situations to get your total score. A score of 10 or greater raises concern: you may need to get more sleep, improve your sleep practices, or seek medical attention to determine why you are sleepy.				

1. I had a deep sleep last night

2. I feel that I slept poorly last night

3. It took me more than half an hour to fall asleep last night

4. I woke up several times last night

5. I felt tired after waking up this morning

6. I feel that I didn't get enough sleep last night

7. I got up in the middle of the night

8. I felt rested after waking up this morning

9. I feel that I only had a couple of hours' sleep last night

10. I feel that I slept well last night

11. I didn't sleep a wink last night

12. I didn't have trouble falling asleep last night

13. After I woke up last night, I had trouble falling asleep again

14. I tossed and turned all night last night

15. I didn't get more than 5 hours' sleep last night

All items are scored true / false

The first question does not count for the total score

One point if answer is 'true': questions 2, 3, 4, 5, 6, 7, 9, 11, 13, 14, 15

One point if answer is 'false': questions 8, 10, 12

Maximum score 14 points, indicating poor sleep the night before

Figure 9. *Groningen Sleep Quality Scale*

Furthermore, the variable "Sleep Satisfaction" on a scale of one to ten is highly subjective, as what is an eight for one person could be a six for another. If this variable were replaced with a scale such as the Epworth Sleepiness Scale (Fig. 8) or the Groningen Sleep Quality Scale (Fig. 9), the results gathered from the study would be more reliable.

Further research could include testing how the time of day that participants were exposed to light therapy affected circadian rhythm, as well as different wavelengths of light. It is a known fact that blue light inhibits the release of melatonin and red light stimulates collagen production, giving skin its structure, strength, and elasticity. It would be interesting to test how the wavelengths in between – which have not been researched extensively – affect circadian rhythm.

References:

Auger, R. R., Burgess, H. J., Emens, J. S., Deriy, L. V., Thomas, S. M., & Sharkey, K. M.
(2015). Clinical Practice Guideline for the Treatment of Intrinsic Circadian Rhythm Sleep-Wake Disorders: Advanced Sleep-Wake Phase Disorder (ASWPD), Delayed Sleep-Wake Phase Disorder (DSWPD), Non-24-Hour Sleep-Wake Rhythm Disorder (N24SWD), and Irregular Sleep-Wake Rhythm Disorder (ISWRD). An Update for 2015. Journal of Clinical Sleep Medicine, 11(10), 1199–1236. https://doi.org/10.5664/jcsm.5100

Beauchamp, Marshall T., and Jennifer D. Lundgren. "A Systematic Review of Bright Light Therapy for Eating Disorders." The Primary Care Companion for CNS Disorders, 27 Oct. 2016, 10.4088/pcc.16r02008. Accessed 21 Dec. 2018.

Blume, C., Garbazza, C., & Spitschan, M. (2019). Effects of light on human circadian rhythms, sleep and mood. Somnologie: Schlafforschung Und Schlafmedizin = Somnology: Sleep Research and Sleep Medicine, 23(3), 147–156. https://doi.org/10.1007/s11818-019-00215-x

CDC - Basics About Sleep - Sleep and Sleep Disorders. (2020). Sleep and Sleep Disorders. https://www.cdc.gov/sleep/about_sleep/index.html

Circadian Rhythm Disorders - What Are Circadian Rhythm Disorders? | NHLBI, NIH. (2022, March 24). Www.nhlbi.nih.gov. https://www.nhlbi.nih.gov/health/circadian-rhythm-disorders

Can't Sleep? Try This Illuminating Trick For Instant Results. (n.d.). The Zoe Report. Retrieved November 25, 2022, from https://www.thezoereport.com/wellness/light-therapy-for-sleep

February 25, and 2019 Rebecca Encao Team Lead | Adjunct Faculty Member. "Types of Eating Disorders: Symptoms, Causes and Effects." www.snhu.edu, 25 Feb. 2019, www.snhu.edu/about-us/newsroom/health/types-of-eating-disorders.

Hospices Civils de Lyon. "Interest of Light Therapy in the Treatment of Binge Eating Disorder: Prospective Controlled Randomized Double-Blind Trial." Clinicaltrials.gov, 17 June 2021, clinicaltrials.gov/ct2/show/study/NCT03094260. Accessed 2 Nov. 2022.

Leppämäki, S., Meesters, Y., Haukka, J., Lönnqvist, J., & Partonen, T. (2003). Effect of simulated dawn on quality of sleep--a community-based trial. BMC psychiatry, 3, 14. https://doi.org/10.1186/1471-244X-3-14

National Institute of Mental Health. "NIMH Eating Disorders." www.nimh.nih.gov, Feb. 2016, www.nimh.nih.gov/health/topics/eating-disorders.

Rubiño, J. A., Gamundí, A., Akaarir, M., Canellas, F., Rial, R., & Nicolau, M. C. (2020). Bright Light Therapy and Circadian Cycles in Institutionalized Elders. Frontiers in Neuroscience, 14. https://doi.org/10.3389/fnins.2020.00359

White Light Therapy. (n.d.). Northwestern University Recreation. Retrieved November 25, 2022, from https://nurecreation.com/sports/2018/8/3/lighttherapy.aspx

Correlation between Empathy and Communication with Highly Emotional People

Janelle Chen

Nikola Tesla STEM High School

Abstract

People with empathy are known to be good with others, and people with great communication skills are known for the same thing, but the question still reminds if there is any relationship between the two when strong emotions of everyday life are factored into the equation? This project studies the relationship between empathy and ability to communicate with highly emotional individuals. The hypothesis was that the higher the level of empathy in a person, the better they would be able to communicate with others, specifically those who have strong emotions because empathic people are known to understand others' emotions, which may allow them to communicate better. Empathy was measured using the Independent Reactivity Index through the empathic concern (affective empathy) and perspective taking (cognitive empathy) subscales. Communication was measured through a survey that asked participants to respond to prompts that encompassed one of the six basic emotions (anger, sadness, joy, surprise, disgust, and fear). The results showed an R-Squared value of 0.132, which conveys little to no correlation between the two variables. This suggests that the prompt within the communication aspect of the survey which includes a person who is overly emotional, may have altered the typical relationship between empathy and communication. Overall, this study has shown that no relationship exists between empathy and emotion specific communication due to the addition of strong emotions, telling us that screening for high empathy and communication skills may not always mean someone is effectively able to handle highly emotional individuals.

Introduction

The purpose of this study was to determine the correlation between how much empathy a person has and their ability to communicate with people who are highly emotional. Empathy and communication often go hand in hand. Our ability to interact with one another is an important aspect of being human and can often be hinged on the emotions that the other person feels. Around 90% of the time, humans feel several emotions (Trampe et al. 2015), and that percentage only goes up the more interactions occur. This means that most of the time, people are feeling some sort of emotion while conversing with others. Due to this, it is possible for our emotions to cloud our judgement and make it hard to communicate effectively. The conversation and interactions that people have alter the way they think and how they respond in every one of those situations. Communication itself is the purposeful process of expressing, receiving, and understanding messages containing factual information, feelings, ideas and needs by two or more individuals through common symbols (Sikiti 1998:1). Humans are naturally collaborative creatures, and this collaboration is facilitated by communication, whether that be verbal, non-verbal, or written. These emotions can be broken down into 6 basic groups: joy, anger, disgust, fear, sadness, and surprise. Recently, due to Covid-19, there has been an increase in the number of people who attend virtual meetings by 12.9% (DeFilippis 2022) as many people were forced to turn to online means of contacting one another. The way we come across to one another has started to become even more important. Empathy helps with this as it is made up of three main concepts: understanding why people feel the way they do, being able to put yourself

in their shoes, and understanding how to respond (Davis 1980). This supports the idea that having a high level of ability in all three parts of empathy can help you when communicating with others. It is important to be able to apply understanding and an appropriate response to every situation no matter what emotion you are faced with. Not only is empathy a great facilitator of good communication, but the ability to empathize with others has been a key part of how humans have gotten along. After all, empathy supports a strong role in effective social functioning (Olderbak et al. 2014). Emotions are what people identify with, whether that be extreme sadness or joy. What matters because the level of empathy someone has depends on which of the six basic emotions are being expressed by the other person. People are likely to encounter all these basic emotions throughout the day in some form and it is important to understand if the level of empathy one has is what helps us to navigate these varying emotions in simple conversations.

It was found that there was not any general impairment by anger in empathic accuracy (Weiblen 2021) which implies that perhaps how innately empathic we are is the strongest deciding factor in how well we can communicate with each other. With these factors in mind, the lack of empathy from society could affect our ability to communicate effectively with others. However, the extent to which it could impact the ability to talk with others is still unknown. In another study, the t-values analysis showed (β=.412; $p<.001$) that indeed empathy levels predicted both positively and meaningfully what the communication scores may be (Ogut N et al. 2021). This is significant because it clearly shows how empathy and communication do have some sort of correlation to one another.

Other investigations did not address the varying levels of empathy and the ability to communicate with people who are highly emotional, which is a factor that should be considered as life and conversations can be unpredictable from person to person. A correlation between empathy and communication is something that has been explored as it has been found that teenagers exchange a substantial amount of empathic emotional communication when participating in online communities (Siriaraya et al. 2011). This study encompasses understanding that the amount of empathy that one has could be significant in the ability to communicate with others. Higher empathy in a person could be positively correlated with an increased ability to effectively communicate with highly emotional people.

Methodology

This study used Discord, Reddit, and word of mouth to recruit 50 participants. Final population gender was made up of 80% females, 16% males, and 4% non-binary, all ranging from 18-45 years old and without a severe mental disorder. Participants were given a consent form to sign before any study materials were given to them. Afterwards, they were assigned a number from a random number generator and given the survey. Every odd numbered participant was given the survey which consisted of a questionnaire to measure an individual's empathy level using questions from the Interpersonal Reactivity Index (IRI), a communication level gauging survey, and a demographics survey. The Interpersonal Reactivity Index (IRI) was used for two of its subscales, empathic concern and perspective taking, which helped to measure the cognitive and affective empathy of a person. The communication aspect of the survey consisted of six prompts to cover one of the six basic emotions (anger, fear, joy, sadness, surprise, and disgust) where the participant must convey a message to the individual strongly feeling one of the basic emotions. Confederates who graded the communication prompt responses were unaware of the study premises and hypothesis and scored according to a set rubric to ensure a double-blind study.

Participants took a survey that measured their empathy level on a Likert scale using the subscales, Perspective Taking and Empathic Concern from the Interpersonal Reactivity Index.

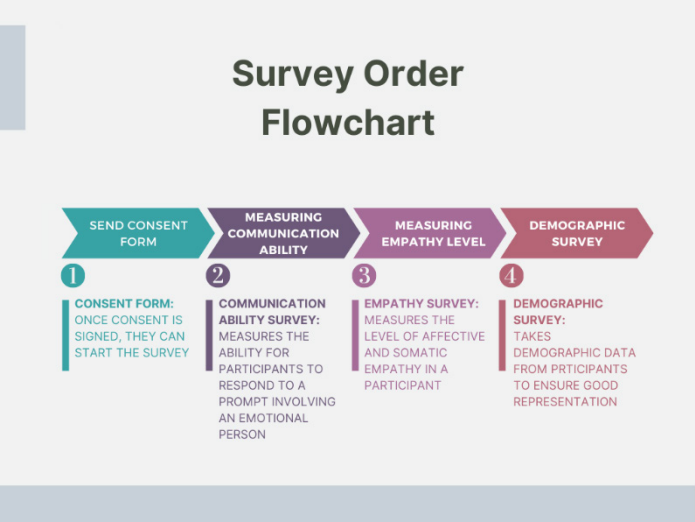


Figure 1. Flowchart showing survey order.

They were given the survey that measures communication level through a series of six prompts (fig. 1). All questions consisted of a fictional prompt that identified with one of the six basic emotions (sadness, joy, disgust, surprise, fear, and anger) within a fictional person and asked the participant to respond to the situation through communicating a set idea despite those emotions. They were then given the empathy level survey to avoid participant bias by preemptively thinking about being empathic. The empathy questionnaire had a total of 14 questions where participants could choose their answer through a 5-point Likert scale from strongly agree to strongly disagree. All participants were instructed to take it in a quiet room and in one sitting. After they completed the survey, participants were given the demographic survey and were debriefed on the entirety of what the study covered as well as thanked for their time.

However, in questions where the inverse scenario was questioned (Ex. “Sometimes, I don’t feel very sorry for other people when they are having problems.”), the point assignment was from 1-5 points respectively. Scores that ranged from 57-70 points were considered ‘highly empathic’, followed by scores that ranged from 29-56 points being considered as ‘somewhat empathic’, and scores that ranged from 14-28 points were considered as ‘Low/ non empathic’ (table 1). R2 was then calculated from all data collected from the empathy survey. Communication level was then measured through confederates using a set rubric to score every question (fig. 2). Each question was scored by three separate graders and the total score was averaged and became the final communication level score.

Affective empathy score (EC)	Cognitive empathy score (PT)	Communication effectiveness score
4, 2, 3, 4, 4, 4, 4, 2,	3, 3, 2, 3, 3, 3	1, -1, 1, 1, 0, 1
4, 4, 4, 4, 2, 2, 4, 4,	4, 4, 4, 5, 2, 3	2, 3, 2, 3, 1, 0
3, 2, 5, 4, 3, 4, 4, 2	3, 3, 1, 3, 2, 2	-1, 0, 0, 1, 0, 2
4, 2, 4, 3, 3, 3, 5, 5	4, 4, 2, 5, 4, 4	1, 1, 2, 0, 2, 1
3, 4, 4, 4, 5, 4, 4, 3	4, 4, 5, 4, 4, 4	1, 3, 1, 0, 1, 1
4, 4, 4, 4, 5, 5, 5, 5	4, 4, 3, 4, 3, 4	1, 2, 3, 1, 1, 3
5, 1, 4, 5, 5, 3, 2, 1	5, 5, 5, 5, 5, 5	1, -1, 1, 1, 0, 0
5, 2, 3, 5, 5, 5, 4, 4	4, 4, 4, 4, 1, 2	1, 2, 3, 0, 1, 3
5, 4, 5, 5, 5, 5, 5, 5	5, 5, 5, 5, 5, 5	1, 1, 3, 1, 1, 2
4, 2, 5, 5, 5, 5, 1, 3	1, 5, 4, 4, 5, 5	2, 1, 1, 1, 1, 1
5, 5, 5, 5, 3, 4, 4, 4	3, 4, 3, 5, 3, 5	1, 2, 1, 0, 0, 0
5, 3, 4, 4, 4, 4, 4, 4	3, 4, 4, 5, 4, 4	1, 1, 1, 0, 1, 1
4, 2, 2, 3, 2, 4, 4, 3	2, 4, 3, 4, 2, 3	1, 0, 0, 0, 0, 1
4, 2, 3, 5, 2, 4, 3, 3	4, 4, 2, 3, 3, 3	2, 2, 1, 0, 1, 1
5, 4, 5, 5, 4, 4, 5, 5	4, 4, 4, 2, 5, 5	1, 0, 1, 0, 1, 1
3, 3, 4, 4, 3, 3, 3, 2	4, 4, 4, 4, 4, 4	2, 2, 0, 1, 0, 1
5, 5, 5, 5, 4, 4, 4, 4	4, 4, 5, 4, 5, 1	0, 0, 0, 0, 1, 0
3, 3, 5, 3, 2, 3, 2, 4	4, 3, 5, 3, 3, 3	2, 1, 2, 2, 2, 2
5, 2, 4, 4, 4, 4, 3, 3	4, 2, 3, 1, 4, 2	0, 1, 0, 0, 0, 1
3, 2, 3, 4, 3, 4, 4, 4	2, 2, 3, 2, 3, 2	1, 1, 0, 1, 0, 1
4, 4, 4, 4, 4, 4, 4, 4	4, 2, 4, 4, 2, 4	0, 1, 0, 0, 2, 2
5, 4, 5, 5, 5, 5, 5, 5,	5, 4, 5, 5, 5, 5	2, 2, 2, 1, 2, 4
3, 2, 5, 5, 4, 1, 4, 4	3, 3, 5, 3, 3, 3	2, 3, 1, 1, 0, 2
4, 4, 4, 4, 2, 3, 4, 3	3, 3, 3, 4, 4, 4	0, 0, 1, 0, 0, 1
5, 3, 4, 5, 4, 5, 5, 4	3, 3, 5, 3, 4, 3	1, 1, 2, 1, 0, 1

Table 1. Data table with results from participant responses.

Rubric:

1. Acknowledged the other person's feelings/paraphrasing the problem — yes or no (1 or 0)
2. Expressing support and encouragement— yes or no (1 or 0)
3. Uses "I" statements when talking to the other person (ex. *I* understand how you feel. *I* get that, something like that has happened to *me* before)— yes or no (1 or 0)
4. Offers (extends) to be there or support/create a safe space for the other to talk — yes or no (1 or 0)
5. Asks open ended questions — yes or no (1 or 0)
6. Does not do the following — yes or no (1 or -1)
    a. Deflect
        i. *"Um. RIP. That, like, sucks. Oh yeah, did you watch the game last night?"*
        ii. Changes the topic
    b. Diminish
        i. *"At least you got 51%. And hey, I heard someone got, like, 20%. You did so well in comparison!"*
        ii. uses "at least" statements or compares situations even if it's comparing them to someone who did worse
    c. Dismiss
        i. *"Calm down. You're overthinking it."*
        ii. Acts/sounds as though they don't care about what the other person is experiencing
        iii. Sound like the situation isn't worth their time or attempt to give them perspective
    d. Directional questioning
        i. *"You're okay, right? I mean, it's been a month...are you feeling better now?"*
        ii. Asking to see if "they're better yet"
    e. Argue or aggravate the issue

Figure 2. Rubric outlining when to award points to participant responses in the communication section of the survey.

Results and Conclusions

The study tested the correlation between levels of empathy and the ability to communicate with highly emotional individuals and hypothesized that there would be a positive correlation, meaning that the higher the level of empathy in a person, the better they would be at communicating with strongly emotional individuals. It was found that there was no correlation or statistical significance between the two variables, indicating that the level of empathy one has is not indicative of how well they are able to communicate with individuals who are extremely emotional.

To break down the study, the survey consisted of two questionnaires, the first one to measure how empathic someone is and the second one on how well someone can communicate with people who are extremely emotional. The empathy level survey yielded a mean score of 52.28 points in total, no score going below 40 points, showing that every participant had been in the "somewhat empathic" (29-55 points) to "highly empathic" (56-70 points) range (table 2). This questionnaire had a median of 52 points and a mode of 56, with a standard deviation of about 7.459 (fig. 3). This indicates that scores were more spread out, creating a wider range of potential scores from participants as well as a slightly lower reliability of those scores being accurate to each participant.

Total Empathy Score	Communication Effectiveness Score
44	3
50	10
41	2
52	7
56	7
58	11
56	2
52	10
69	9
54	7
58	4
56	5
42	2
45	7
61	4
49	6
59	1
46	11
45	2
41	4
52	5
68	13
48	9
49	2
56	6

Table 2. Data table that shows all raw data.

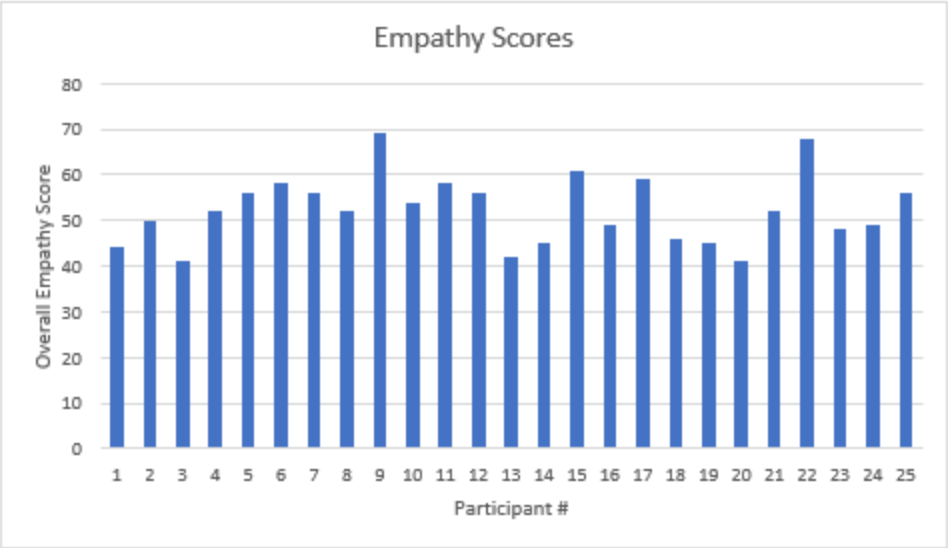


Figure 3. Bar graph showing data points based on empathy score.

Moving on to the communication aspect of the survey, it is important to take into account that less points were able to be allotted in this portion of the questionnaire due to having less criteria needed to get full points. In total, the communication measuring survey yielded an average of 5.96 points, a median of 6, and a mode of 2 to create a standard deviation of 3.352 (fig. 4). This deviation is much larger as it is almost half of the mean and median, telling us that the communication aspect of the survey may not have yielded reliable results.

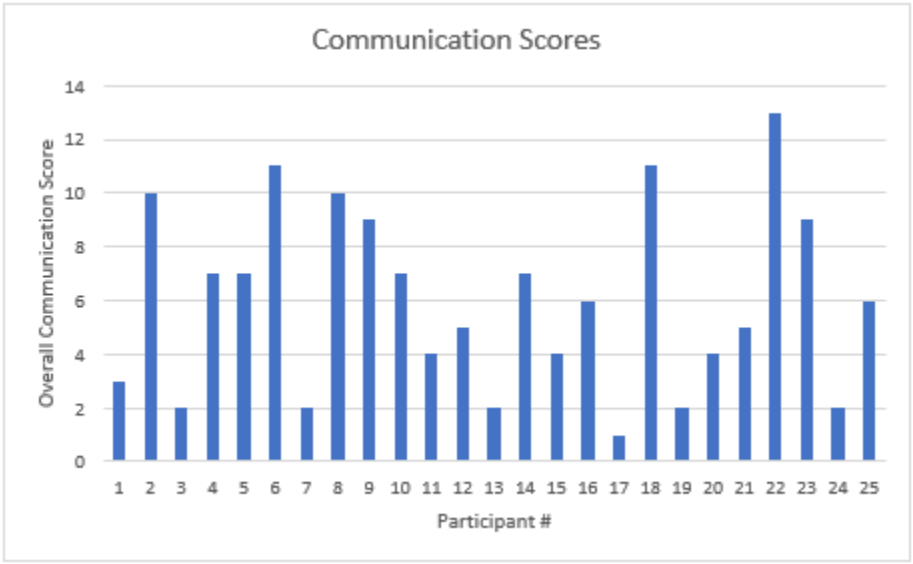


Figure 4. Bar graph showing data points based on communication score.

R^2 analysis was then used to understand the strength and the relationship between empathy level and ability to communicate with highly emotional individuals. Taking the two datasets, Excel was used to calculate the final r-squared value of about 0.132 (fig. 5). This indicates little to no correlation between the two variables, empathy and communication with strongly emotional individuals. It also suggests that despite previous studies, the level of empathy one has is not indicative of how well they will be able to communicate with an individual who is highly emotional.

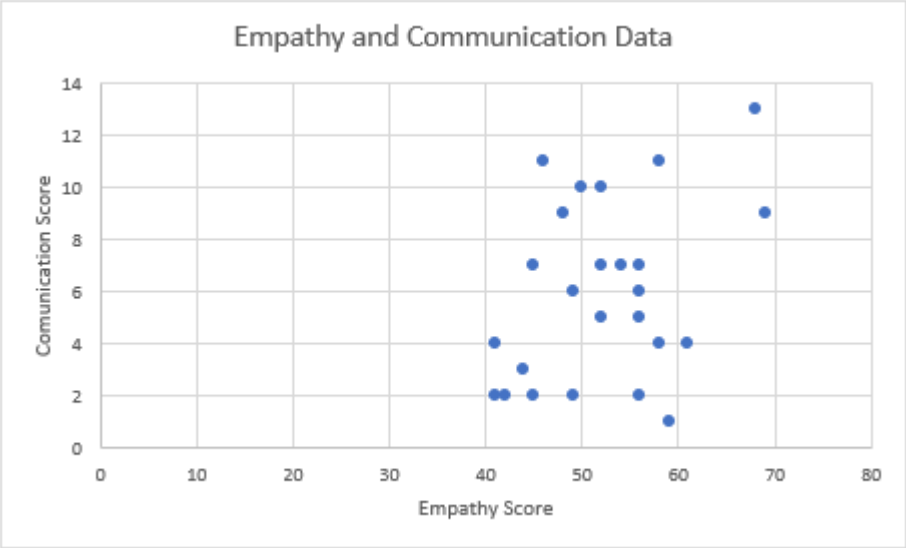


Figure 5. Scatterplot showing data points based on their empathy and communication score

Taking the subscales of the communication aspect of the survey, it can be broken down into the six basic emotions: anger, joy, disgust, fear, surprise, and sadness. If each emotion-specific communication score were to be solely compared against total empathy scores, it is also possible to understand the specific relationship between the ability to communicate with individuals feeling a certain emotion and the participant's empathy level. Starting with joy, participants scored a mean of 1.08 points, a median of 1 point, and a mode of 1 point, altogether creating a standard deviation of about 0.744 (fig. 6). This shows low levels of ability to communicate well with highly happy individuals as well as a higher reliability within this subscale. When considered in comparison to empathy levels, it created a total r-squared value of 0.171, showing a slight correlation between empathy and joy-specific communication. This value is slightly higher than the r-squared value of the overall correlation between empathy and communication.

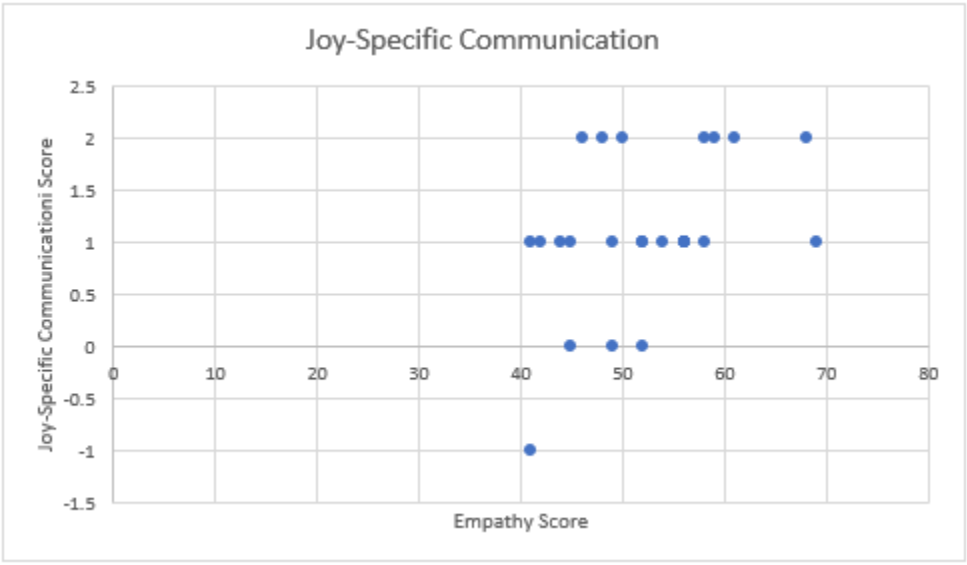


Figure 6. Scatterplot showing data points based on their empathy score and joy-specific communication score.

Anger-specific communication yielded different results as in total, while it may have had a mean of 1.12 points, a median of 1 point, and a mode of 1 point, all values similar to joy-specific communication scores, it yielded an r-squared value of about 0.019 (fig. 7). This implies no correlation between anger-specific communication and empathy, showing that empathy levels cannot be used to predict how well one can communicate with individuals who are highly angry.

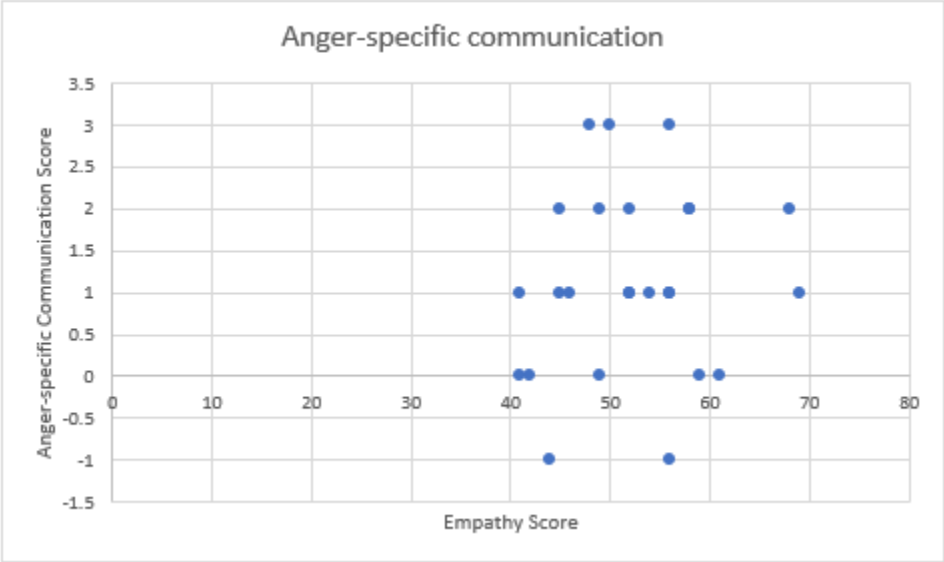


Figure 7. Scatterplot showing data points based on their empathy score and anger-specific communication score.

Fear-specific communication scores had a mean of 1.12 points, a median of 1 point, and a mode of 1 point, altogether creating a standard deviation of about 0.909 (fig. 8). This shows that participants scored slightly higher in fear specific communication compared to other emotion-specific communication questions. This means that fear was an easier emotion for participants to communicate with in comparison to other strong feelings. In total, it yielded an r-squared value of about 0.181 which is higher than the overall r-squared value, implying that there may be a slight correlation between how empathic someone is and how well they are able to communicate with highly fearful individuals.

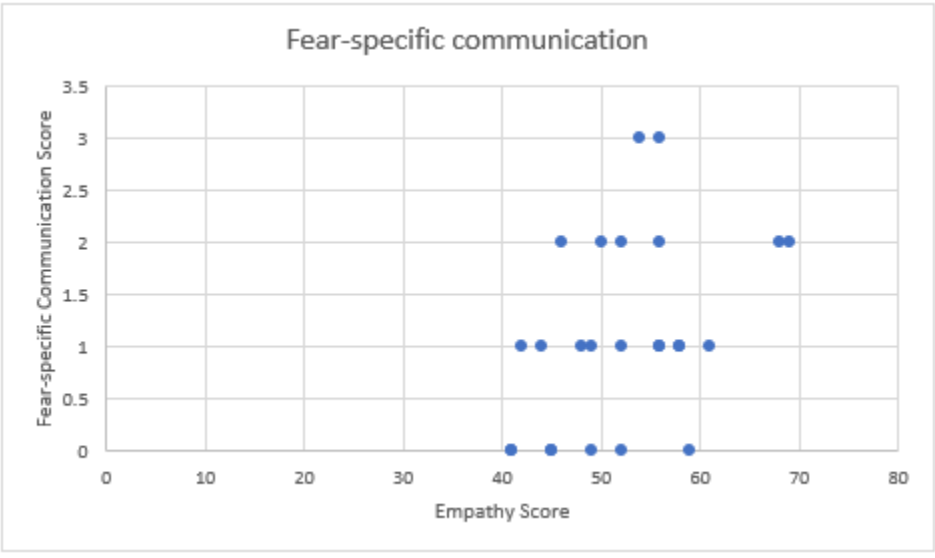


Figure 8. Scatterplot showing data points based on their empathy score and fear-specific communication score.

Disgust-specific communication scores had a mean of 0.64 points, median of 1 point, and a mode of 0 points (fig. 9). These scores show a lack of ability for participants to communicate well with an individual who invokes a feeling disgust in the participant, and the r-squared value of about 0.046 reflects that. The low r-squared value is likely due to the affective empathy aspect of the empathy questionnaire, which measures the ability to feel what another person is feeling, and scoring higher in this portion of the survey can indicate that participants may feel the same disgust as the individual and that may help or hinder their ability to communicate well with them.

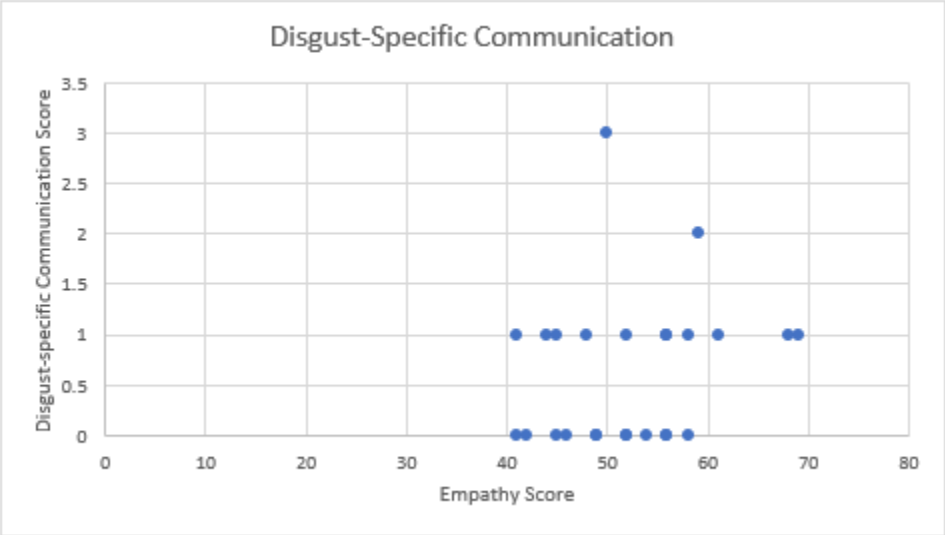


Figure 9. Scatterplot showing data points based on their empathy score and disgust-specific communication score.

Surprise-specific communication showed weak correlation, resulting in an r-squared value of about 0.101. This is emphasized by the mean, median, mode, and standard deviation of 0.8, 1, 0, and about 0.849 points, respectively (fig. 10). Not only does the data indicate the idea that how empathic one is does not tell you how well a person will be able to communicate with a surprised person, but it also implies that people are not able to communicate well with those who are surprised in general.

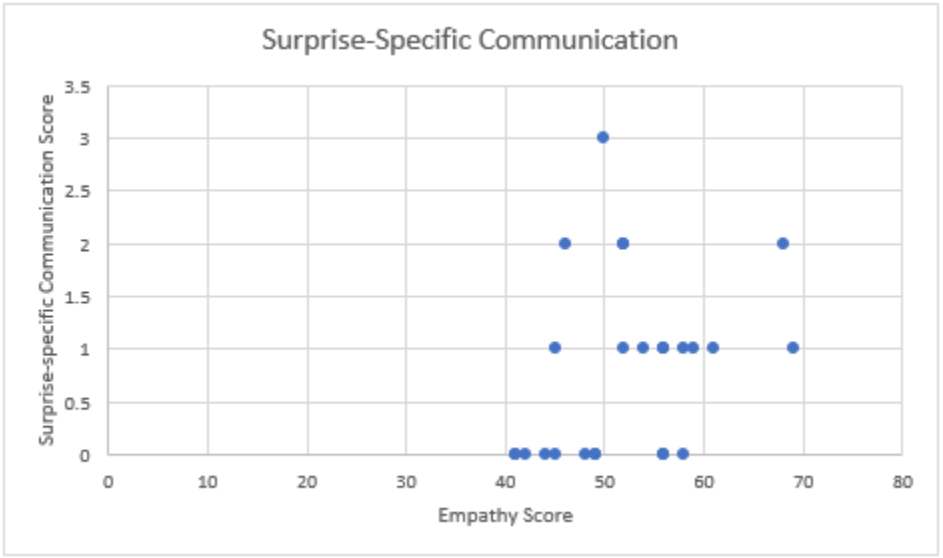


Figure 10. Scatterplot showing data points based on their empathy score and surprise-specific communication score.

Sadness-specific communication scores had a mean of 1.36, a median of 1, and a mode of 1, which created a standard deviation of about 0.933 (fig. 11). This shows more of an ability for the participants to communicate with those who are sad, however that does not seem to correlate with their empathy levels as this emotion-specific communication had an r-squared value of about 0.044. The data shows that there is no correlation between how empathic someone is and how well they are able to communicate with extremely sad individuals.

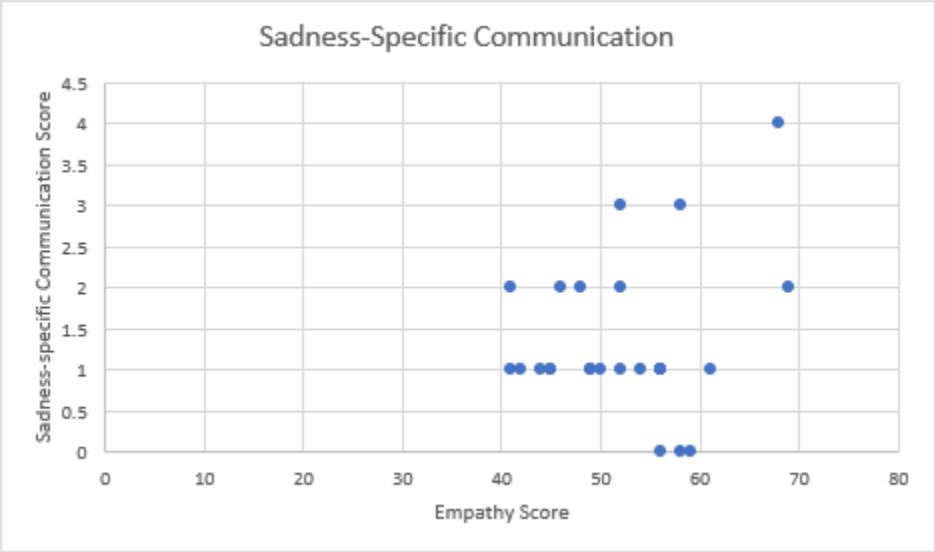


Figure 11. Scatterplot showing data points based on their empathy score and sadness-specific communication score.

Altogether, the data showed that this study's hypothesis was incorrect and that the level of empathy one has is not indicative of how well one will be able to communicate with highly emotional individuals. This implies that regardless of previous studies' findings, how empathic someone is, may be null when you factor in strong emotions of other individuals. This study found that there was a stronger correlation between empathy and communication when participants were asked to respond to those who were extremely happy, fearful, or surprised. This was less prominent when tasked with needing to communicate with people who were extremely sad, disgusted, or angry. Given this information, it allows for a better understanding of how well one can communicate with people with strong emotions.

Discussion

Limitations of this study included the fact that only 25 participants' data was used which may imply that there may not have been a reliable sample size that represented the overall population. This is especially emphasized when looking at the demographics as there was a predominantly female and Asian population amongst my participants. Therefore, more trials should be conducted to verify the results, and in addition to that, the communication survey is made up of qualitative data making the scoring of participants' responses slightly unreliable as confederates may interpret responses differently despite the set rubric. There is also the possibility of participation bias as much of the questionnaire is self-reporting and participants could have given inflated scores on the empathy measuring portion of the survey and reporting an action they may not actually do in the posed situations in the communication questionnaire. However, it is possible to fix the latter two issues through more qualitative, factual research methods such as taking brain scans while asking participants to exercise empathy as well as communicate with highly emotional individuals, something that could be explored in future research. Regardless, the results from this study can be employed in situations from hiring employees to healthcare workers through screening tests, not to search for certain characteristics, but to understand how well an individual will respond to others in certain situations.

Bibliography

Davis, M. H. (1980). Interpersonal reactivity index. PsycTESTS Dataset. https://doi.org/10.1037/t01093-000

DeFilippis, E., Impink, S. M., Singell, M., Polzer, J. T., & Sadun, R. (2022). The impact of COVID-19 on Digital Communication Patterns. Humanities and Social Sciences Communications, 9(1). https://doi.org/10.1057/s41599-022-01190-9

Olderbak, S., Sassenrath, C., Keller, J., & Wilhelm, O. (2014). An emotion-differentiated perspective on empathy with the emotion specific empathy questionnaire. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00653

Rees, C., Sheard, C., & McPherson, A. (2002). Communication skills assessment: The perceptions of medical students at the University of Nottingham. Medical Education, 36(9), 868–878. https://doi.org/10.1046/j.1365-2923.2002.01300.x

Siriaraya, P., Tang, C., Ang, C. S., Pfeil, U., & Zaphiris, P. (2011). A comparison of empathic communication pattern for teenagers and older people in online support communities. Behaviour & Information Technology, 30(5), 617–628. https://doi.org/10.1080/0144929x.2011.582146

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. PLOS ONE, 10(12). https://doi.org/10.1371/journal.pone.0145450

Weiblen, R., Mairon, N., Krach, S., Buades-Rotger, M., Nahum, M., Kanske, P., Perry, A., & Krämer, U. M. (2021). The influence of anger on empathy and theory of mind. PLOS ONE, 16(7). https://doi.org/10.1371/journal.pone.0255068

ÖĞÜT, N., KARAKOÇ, E., & GHİASEE, A. (2021). An empirical study on determining the communication and empathy skill levels of healthcare professionals. Konuralp Tıp Dergisi. https://doi.org/10.18521/ktd.951443

The Effect of Different Modalities and Frequencies of Exercise on Mental Health

Soham Korde

Nikola Tesla STEM High School

Abstract

Over time, Depression is becoming more and more prevalent in our society. As with anxiety and other mood disorders, people are looking for cures to this difficult-to-treat disease. There has been research done on the benefits of exercise in regards to mental health, but these studies have not measured a large enough population or have neglected factors such as time and intensity. For this study, a survey was posted in multiple places that asked about anxiety and depression, exercise habits, and other related questions. The results of the survey were not able to be used due to the lack of participants.

Introduction:

This correlational study aims to find the most effective form/frequency of exercise for improving mental health. While it is widely accepted that exercise is beneficial to one's mental health, it has yet to be determined what kind of exercise is most effective in helping treat poor mental health. Depression continues to be one of the most prominent mental illnesses in modern society. Approximately 8.4% of American adults have experienced a major depressive episode at some point in their lives, while 11.5% and 4.2% of adults experienced mild and moderate depressive episodes respectively (NIMH.gov). There are a variety of treatments, each with its own pros and cons. Antidepressants such as sertraline (Zoloft) are one of the most effective treatments that can be used, usually in major cases of depression. However, the use of drugs has a plethora of issues, including but not limited to dependence and abuse. Medication can also be expensive, which excludes most of the population who believe that it is not worth the thousands of dollars they will have to spend. Even though medication is expensive, trials have shown that there is only a 51% improvement over a placebo (Pigott, 2010). This led people to explore other possibly more effective treatments.

While it is universally accepted that exercise has a countless number of physical benefits, recent studies have shown exercise's benefit on the psychological side. One study compared the effectiveness of HIRT (High Intensity Resistance Training), LIRT (Low Intensity Resistance Training), and GP (General Practitioner) care. The study followed a group of elderly patients with clinical depression as they followed an exercise plan. One group would be doing low intensity resistance training, another would do higher intensity exercise, and the last group would follow general practitioner care. The study found that, while all three showed improvements, HIRT was the most effective form of anti-depressive treatment. The quantification used by this study was the Hamilton Depression index, and a "treated" patient was a patient with a 50% or greater decrease in their score. The HIRT group had a 52+-7% treatment rate ($p = 0.0132$), while the GP group only had a 25+-8% treatment rate. This study, however, was only done on elderly patients, and thus cannot be generalized (Singh, 2005).

Another study found a correlation between obesity and various constituents of mental health such as self-esteem, self-worth, and anxiety and depressive symptoms. This was also a longitudinal study, meaning that the results can reasonably be

generalized to the public. The study found that obesity was correlated to lower feelings of self-esteem, health, and life satisfaction. (Lavallee, et al.) While obesity is not something that will be researched in this paper, the loss of body fat and treatment of obesity is a consequence of exercise, hypothetically leading to improved mental health no matter the modality of exercise. (Lavallee, 2021)

Another study examined the effects of resistance training on obese adults. The exercise interventions ranged from 6-76 weeks, with a median of 12 weeks. 13 groups were aerobic exercise, known as cardio, 8 were a resistance training program like weightlifting, and 6 were a combination of the two. The sessions ranged from 12-90 minutes, which is to be expected due to the variations in modalities. The body image of the people in the exercise groups was significantly improved in comparison to the people in the control group. (Caracca, 2021)

Yet another study followed youths with analogue generalized anxiety disorder. The youths took both a GAD-7 and a Penn state worry questionnaire to quantify anxiety value. They then embarked on a fitness journey of approximately 8 weeks, and 2 days a week. The exercise significantly decreased worry symptoms by an average of 6.49 points ($p=<0.04$), worry engagement by an average of 5.49 points ($p<=0.059$), and reduced anxiety symptoms by an average of 10.50 points ($p<=0.001$). This study shows that even in youth, exercise has a significantly positive effect on mental health. Combined with the studies before, this provides a reliable trend across age groups. (Gordon, 2021)

This study aims not only to investigate the correlation between exercise and mental health, but also to find the most effective exercise modality for helping treat mental health disorders, hopefully leading to a more normalized form of therapy. The hypothesis is that exercise, no matter the modality, will show a consistent increase in overall mental health. The increase will be linear, scaling with time spent weekly exercising, and will show a trend of better mental health as intensity increases.

Methodology:

The materials used will be <=200 participants and a flyer with a QR code that redirects to a form. The flyer will be posted in Redmond downtown park, This QR code leads to a Microsoft Forms document. Before the participants can start filling out the form, they will need to sign a digital informed consent form. After 150 responses to the form, or 1 month since the posting of the flyer, the participants will be randomly selected using a random number generator. This generator will take the original number of participants, select 1/3 of the populations' worth of female and 1/3 of the populations' worth of male participants. The other participants' data will not be used for the experiment. Before the experiment begins, participants will have to fill out an informed consent form that outlines the nature of the experiment. Each participant will fill out a Microsoft form that contains a questionnaire on their exercise habits, such as frequency, modality, and intensity. This data will be input into data table one, alongside the participants' assigned number (assigned when expirement begins to avoid collection of names). The next part of the form will be a standardized Beck Depression Inventory (Fig. 1).

Beck's Depression Inventory

This depression inventory can be self-scored. The scoring scale is at the end of the questionnaire.

1.
- 0 I do not feel sad.
- 1 I feel sad
- 2 I am sad all the time and I can't snap out of it.
- 3 I am so sad and unhappy that I can't stand it.

2.
- 0 I am not particularly discouraged about the future.
- 1 I feel discouraged about the future.
- 2 I feel I have nothing to look forward to.
- 3 I feel the future is hopeless and that things cannot improve.

3.
- 0 I do not feel like a failure.
- 1 I feel I have failed more than the average person.
- 2 As I look back on my life, all I can see is a lot of failures.
- 3 I feel I am a complete failure as a person.

Figure 1: A sample of the Beck Depression inventory, a tool used to quantify an individual's feelings of depression.

GAD-7 Anxiety

Over the last two weeks, how often have you been bothered by the following problems?	Not at all	Several days	More than half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	0	1	2	3
2. Not being able to stop or control worrying	0	1	2	3
3. Worrying too much about different things	0	1	2	3
4. Trouble relaxing	0	1	2	3
5. Being so restless that it is hard to sit still	0	1	2	3
6. Becoming easily annoyed or irritable	0	1	2	3
7. Feeling afraid, as if something awful might happen	0	1	2	3

Column totals _____ + _____ + _____ + _____ =

Total score _____

If you checked any problems, how difficult have they made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all ☐ Somewhat difficult ☐ Very difficult ☐ Extremely difficult ☐

Figure 2: Like the Beck Depression Inventory, the GAD-7 allows scientists to quantify mental illnesses, in this case anxiety.

The participants' score will be input into data table 1. The final part of the form will contain the GAD-7 (Fig. 2). This data will also be input into table 1.

Data Analysis:

The mean of the GAD-7 score and Beck Depression Index score at each intensity level (1-10) and each frequency (in hours/week) will be calculated. These values will then be graphed into 4 different graphs. Table 1: Participant's scores will be entered into this table for data analysis. A linear regression model will then be used, and a standard deviation, standard error, and r-coefficient will be calculated to check if the results are significant.

Participant #	Beck Depression Score	GAD-7 Score	Frequency of exercise	Intensity of exercise
1.	4	10	4	6
2.	7	15	0	1
3.	10	11	3	5
4.	8	11	4	5

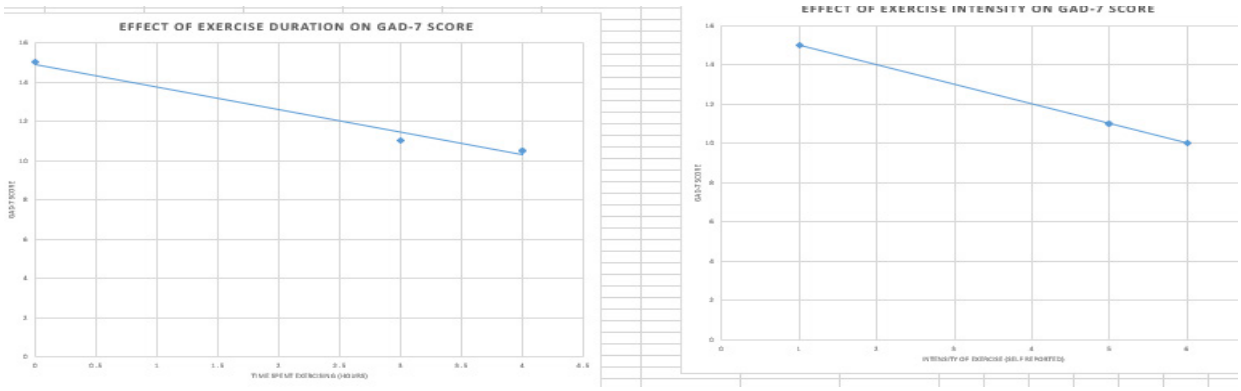


Figure 3: The GAD-7 scores showed a much more linear correlation as compared to the Beck Inventory Graphs

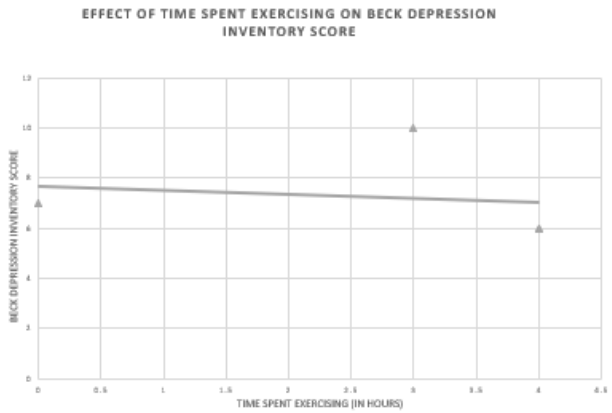


Figure 2: Slight downward trend in graph

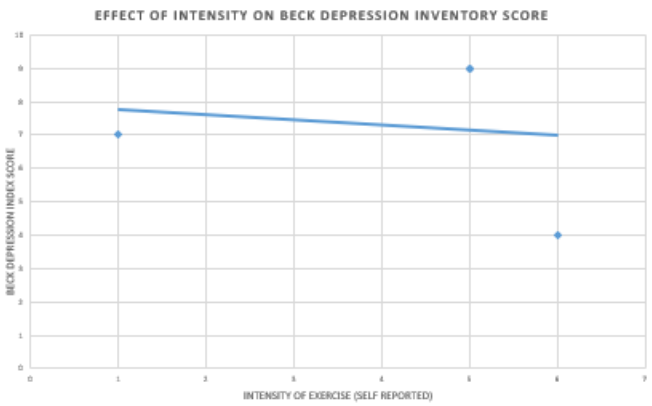


Figure 1: Slight downward trend in graph

Unfortunately, there were not enough data points to calculate an R-Coefficient. The data restriction also led to overlapping SEM lines, which means that the results of the correlational study are inconclusive.

Conclusion:

In the data, there was a slight negative correlation between both frequency and intensity to GAD-7 and Beck Depression Inventory scores. However, the lack of data led to a statistically insignificant correlation, which means that the hypothesis was false. In the future, researchers may be able to devise an experiment to document the change in mental health over a period of exercise. This would not only be more significant than a correlational study, but would also take into account changes in mental health instead of a one-time survey, which is prone to social desirability bias.

Bibliography:

1.) Beck's Depression Inventory. (n.d.). Ismanet.org. Retrieved January 17, 2023, from https://www.ismanet.org/doctoryourspirit/pdfs/Beck-Depression-Inventory-BDI.pdf

2.) Carraça, E. V., Encantado, J., Battista, F., Beaulieu, K., Blundell, J. E., Busetto, L., Baak, M., Dicker, D., Ermolao, A., Farpour-Lambert, N., Pramono, A., Woodward, E., Bellicha, A., & Oppert, J. (2021). Effect of exercise training on psychological outcomes in adults with overweight or obesity: A systematic review and meta-analysis. Obesity Reviews. https://doi.org/10.1111/obr.13261

3.) GAD-7 Anxiety. (n.d.). https://adaa.org/sites/default/files/GAD-7_Anxiety-updated_0.pdf

4.) Gordon, B. R., McDowell, C. P., Lyons, M., & Herring, M. P. (2021). Resistance exercise training among young adults with analogue generalized anxiety disorder. Journal of Affective Disorders, 281, 153–159. https://doi.org/10.1016/j.jad.2020.12.020

5.) Lavallee, K. L., Zhang, X. C., Schneider, S., & Margraf, J. (2021). Obesity and Mental Health: A Longitudinal, Cross-Cultural Examination in Germany and China. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.712567

6.) National Institute of Mental Health. (2022, January). Major Depression. National Institute of Mental Health. https://www.nimh.nih.gov/health/statistics/major-depression

7.) Pigott, H. E., Leventhal, A. M., Alter, G. S., & Boren, J. J. (2010). Efficacy and Effectiveness of Antidepressants: Current Status of Research. Psychotherapy and Psychosomatics, 79(5), 267–279. https://doi.org/10.1159/000318293

8.) Singh, N. A., Stavrinos, T. M., Scarbek, Y., Galambos, G., Liber, C., & Fiatarone Singh, M. A. (2005). A Randomized Controlled Trial of High Versus Low Intensity Weight Training Versus General Practitioner Care for Clinical Depression in Older Adults. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 60(6), 768–776. https://doi.org/10.1093/gerona/60.6.768

Self-esteem and Social Anxiety of South Asian Adolescents with Varying Acne Severity

Prisha Agarwal

Nikola Tesla STEM High School

Abstract

Acne vulgaris is a skin disease causing increased detrimental effects on the lives of adolescents with degrading mental health. This study seeks to investigate the psychological impact of varying acne severity on self-esteem (relating to body image) and social anxiety (relating to social interactions) of South Asian adolescents. The hypothesis proposed was that increased acne severity (independent variable) would result in lower self-esteem and higher social anxiety (dependent variables) in those with greater than 10% coverage of facial acne. The Rosenberg Self-Esteem Scale (RSES) and the Liebowitz Social Anxiety Scale (LSAS) were utilized to quantify self-esteem and social anxiety in this cross-sectional study. The acne severity was graded by using the percentage of surface area coverage of acne through a novel grid system in Adobe Photoshop using edited images provided by participants: less than 10% coverage of acne indicated none to mild acne severity and greater than 10% coverage of acne indicated moderate to severe acne severity. Differences in acne severity between the groups in the study was not statistically significant *(p-value > 0.05* for self-esteem and social anxiety*)*. Moderate to severe acne severity indicated lower average self-esteem (11.7% decrease from control mean) and higher average social anxiety (14.7% increase from control mean). *Acne vulgaris* does not have a significant effect on the self-esteem and social anxiety of South Asian adolescents, but data does indicate a trend like what is proposed in the initial hypothesis.

Introduction

This project seeks to investigate how varying severities of acne vulgaris affect the self-esteem and social anxiety of South Asian adolescents to determine the major consequences of a common dermatological disease. Acne vulgaris is the most common skin disease in the United States, with an 85% prevalence in the adolescent population which can contribute to multiple side effects, both psychologically and physiologically. Acne treatments in total can cost almost $1 billion in culmination, and harsher treatments like Accutane, or isotretinoin, can lead to even harsher symptoms not related to contraction of acne. The disease is a result of the glands and follicles of the skin on either the face, chest, or back becoming blocked by excess dirt, sebum, or oil. This blockage leads to the different types of lesions which can be classified as different types of acne, also indicating varying severities of acne. These different classifications include papules, pustules, cysts, nodules, hyperpigmentation, and acne scarring (Figure 1).

Papules are identifiable since they are smaller bumps, and pustules are a pus-filled developed version of a papule. Higher frequency of papules and pustules indicates none to mild acne severities, and usually causes hyperpigmentation which isn't usually permanent. Cysts are larger, usually redder, lesions which can be quite sensitive to the touch. Nodules are a larger version of a pustule, which usually cause more permanent forms of acne scarring. Whiteheads and blackhead were not identified in any participant during this survey,



Figure 1 *Types of Acne & Treatment*

Note. Images of the varying classifications of acne types with their respective treatment types. *Versed, https://versedskin.com/blogs/learn/types-of-acne*

but they are classified as types of acne which are caused by excess build-up in the pores of the skin.

These varying classifications of acne were applied to a novel Acne Quantification Method utilized in this study. The main goals of this method were to (1) improve existing methods that exist to quantify and grade acne, (2) sort participants into the control and experimental groups for the study, and (3) digitize the acne grading into an existing software rather than up to human biases. Existing methods to grade acne include the Investigator's Global Assessment Scale (IGA) and the commonly used Leeds & Cunliffe Table which is utilized in medical practice as well (Figure 2).

Degree I	Predomiance of comedones, papules and pustules (small and < 10)	Mild
Degree II	10-40 papules and pustules (comedones)	Moderate
Degree III	40-100 papules and pustules, >40 comedones, presence of nodules	Moderate/ Severe
Degree IV	Nodulocystic and conglobata AV with severe, painful l esions, papules, pustules and comedones	Severe

Figure 2 - *Commonly Used Acne Grading Scales*

Note. *The Leeds & Cunliffe Table, details four degrees of acne severity based on presence and frequency of varying types of acne lesions. Research Gate, https://www.researchgate.net/figure/Classification-of-acne-vulgaris-of-Leeds-re-vised-by Cunliffe-2003_tbl1_283834033*

Grade	Description
0	Clear skin with no inflammatory or non-inflammatory lesions
1	Almost clear; rare non-inflammatory lesions with no more than one small inflammatory lesion
2	Mild severity; greater than Grade 1; some non-inflammatory lesions with no more than a few inflammatory lesions (papules/pustules only, no nodular lesions)
3	Moderate severity; greater than Grade 2; up to many non-inflammatory lesions and may have some inflammatory lesions, but no more than one small nodular lesion
4	Severe; greater than Grade 3; up to many non-inflammatory lesions and may have some inflammatory lesions, but no more than a few nodular lesions

Note. *The IGA Scale, utilized in another scientific article using AI to detect acne. Automatic Acne Object Detection and Acne Severity Grading Using Smartphone Images and Artificial Intelligence. doi: 10.3390/diagnostics12081879. PMID: 36010229; PMCID: PMC9406819.*

The common issue with these techniques was the vague language such as "some lesions" or "presence of" which does not thoroughly detail what the varying acne degrees or severities could be. The AQM aims to improve these scales by using a surface area ratio model in a grid overlay system in Adobe Photoshop. The comparison of the acne surface area to the forehead surface area produces a single percentage which can be recorded more easily and quantifies the acne severity using a calculated value. The trapezoidal surface area was utilized to ensure the highest accuracy with the software available, and this shape also matched the shape of participants' foreheads the best (Figure 3).

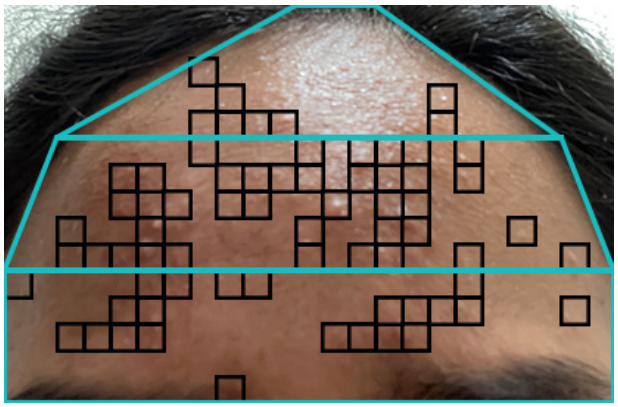


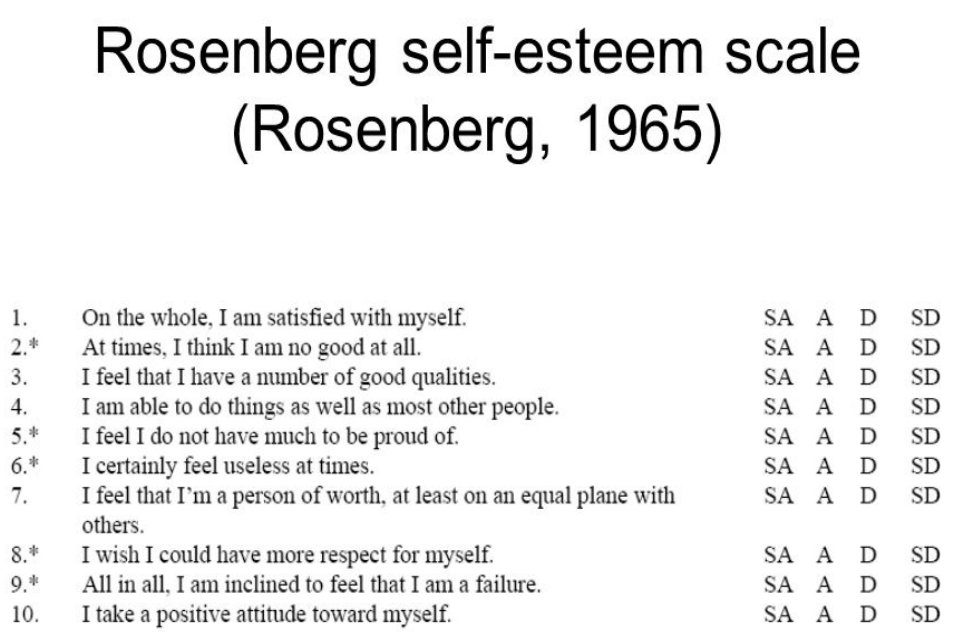
Figure 3
AQM on Participant with Severe Acne Severity

Note. *The two trapezoidal surface areas were added together with a rectangular surface area to ensure the surface area of the participants' forehead did not include areas where acne is not present. Adobe Photoshop by Agarwal, Prisha.*

Self-esteem is found in many aspects of life, and as adolescents are in the stage in their lives where they are still figuring out their identity, self-esteem and confidence in oneself is an important part of figuring out that identity. Maslow's Hierarchy of Needs include self-esteem as one of the stages in order to reach self-actualization or reaching the full potential of oneself. Confidence in one's worth can be affected by a physically damaging disease such as *acne vulgaris*, which can reduce how confident someone is in their worth because they can look and feel different than other people. The Rosenberg Self-esteem Scale evaluates this aspect in this study, using 10-questions which return a score between 1-30. A higher score indicates higher self-esteem, and a lower score indicates a lower self-esteem (Figure 4).

Figure 4

Rosenberg Self-esteem Scale



Rosenberg self-esteem scale
(Rosenberg, 1965)

1.	On the whole, I am satisfied with myself.	SA	A	D	SD
2.*	At times, I think I am no good at all.	SA	A	D	SD
3.	I feel that I have a number of good qualities.	SA	A	D	SD
4.	I am able to do things as well as most other people.	SA	A	D	SD
5.*	I feel I do not have much to be proud of.	SA	A	D	SD
6.*	I certainly feel useless at times.	SA	A	D	SD
7.	I feel that I'm a person of worth, at least on an equal plane with others.	SA	A	D	SD
8.*	I wish I could have more respect for myself.	SA	A	D	SD
9.*	All in all, I am inclined to feel that I am a failure.	SA	A	D	SD
10.	I take a positive attitude toward myself.	SA	A	D	SD

Note. 10 questions in total, and with four answer options including strongly agree, agree, disagree, and strongly disagree. *Slide Player,* https://slideplayer.com/slide/5796808/

Social anxiety, and more specifically Social Anxiety Disorder (SAD) is classified for a patient's apparent fear and/or avoidance of being watched or judged by others. Self-esteem and social anxiety relate in that a person's own confidence in themselves can affect how they feel about the perception of themselves in the eyes of others. SAD can have varying severities as well, which was evaluated in the Liebowitz Social Anxiety Scale. The scale consists of 48 questions divided into two sections which evaluate the fear of certain social situations and the avoidance of those social situations. The scale returns a score between 0-144, with a lower score indicating lower social anxiety and a higher score indicating higher social anxiety (Figure 5).

Figure 5

Liebowitz Social Anxiety Scale

Psychiatry Associates, P.C.
1736 Oxmoor Road. Suite 103. Birmingham, AL 35209
Telephone: 879 2120. Telecopier: 879-2125.

Liebowitz Social Anxiety Scale

Name: _____ Age: _____ Date: _____

Fill out the following questionnaire with the most suitable answer listed below. Base you answers on you experience in the past week and, if you have complete the scale previously, be as consistent as possible in your perception of the situation described. Be sure to answer all items.

Fear or Anxiety:	Avoidance:
0 = None	0 = Never (0%)
1 = Mild	1 = Occasionally (1—33%)
2 = Moderate	2 = Often (33—67%)
3 = Severe	3 = Usually (67—100%)

	Fear or anxiety	Avoidance
1. **Telephoning in public** - speaking on the telephone in a public place		
2. **Participating in small groups** – having a discussion with a few others		
3. **Eating in public places** – do you tremble or feel awkward handling food		
4. **Drinking with others in public places** – refers to any beverage including alcohol		
5. **Talking to people in authority** – for example, a boss or teacher		
6. **Acting, performing or giving a talk in front of an audience** – refers to a large audience		
7. **Going to a party** – an average party to which you may be invited; assume you know some but not all the people at the party		
8. **Working while being observed** – any type of work you might do including school work or housework		
9. **Writing while being observed** – for example, signing a check in a bank		
10. **Calling someone you don't know very well**		
11. **Talking with people you don't know very well**		
12. **Meeting strangers** – assume other are of the average importance to you		
13. **Urinating in a public bathroom** – assume that others are sometimes present, as might normally be expected		
14. **Entering a room when others are already seated** – refers to a small group, and nobody has to move seats for you		
15. **Being the center of attention** – telling a story to a group of people		
16. **Speaking up at a meeting** – speaking from your seat in a small meeting or standing up in place in a large meeting		
17. **Taking a written test**		
18. **Expressing appropriate disagreement or disapproval to people you don't know very well**		
19. **Looking at people you don't know very well in the eyes** – refers to appropriate eye contact		
20. **Giving a report to a group** – refers to an oral report to a small group		
21. **Trying to pick up someone** – refers to a single person attempting to initiate a relationship with a stranger		
22. **Returning goods to a store where returns are normally accepted**		
23. **Giving an average party**		
24. **Resisting a high pressure salesperson** – avoidance refers to listening to a salesperson too long		

Liebowitz MR. Social Phobia. Mod Probl Pharmacopsychiatry 1987;22:141-173

Note. 48 questions in total, and with answers choices for fear as a scale, and choices for avoidance as percentages. *DocHub,* https://www.dochub.com/fillable-form/29059-liebowitz-social-anxiety-scale-scoring-pdf

Altogether, varying acne severities, frequencies of different types of acne lesions, as well as physical effects of the disease push for more research on the dermatological disease and it's psychological implications. With a new method like the AQM which quantifies acne in a new way, further insight can be provided into the effect of varying severities on the mental health of patients. This can further improve dermatological practicing, especially because medical professionals can consider their patients' mental health on top of the already harsh symptoms of oral medications like isotretinoin. More research on this area of the disease could transform patient care and could be one of the first steps to take as we strive to heal not just the body, but the mind as well.

Methodology

First, recruitment of participants was accomplished through posting of the designated flyer on social media and frequented public areas (Figure 6).

Figure 6

Recruitment Flyer



Note. The recruitment flyer was posted with a QR code to the survey to prompt the participants to fill it out. *Canva, from Agarwal, Prisha*

The criterion for the participants includes that they were of a South Asian origin and between the ages of 13-19 to fit the target demographic of the study (Table 1).

Table 1 – Participant Criteria for Study

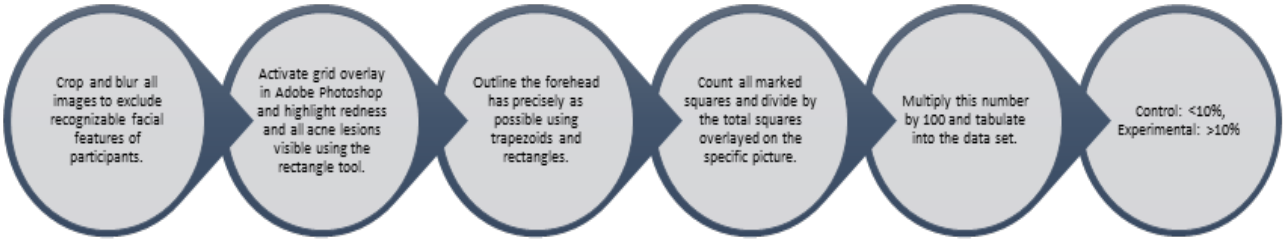
South Asian Nationality	Age
Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka	13-19 years of age

This was checked first in a tabulated data set before continuing with the AQM procedure and creating organized spreadsheets for the independent samples t-test in RStudio. Once this was approved to be in accordance with the criterion, the AQM procedure was performed on the participants' forehead picture. First, the picture is cropped to exclude any identifiable facial features and then a grid is overlayed to use to create the trapezoidal and rectangular shapes surrounding the forehead in order to calculate the surface area. Next, the acne and any other types of scarring or pigmentation were marked in the grid using a square shape tool in Photoshop to provide easy identification and record of the different types of acne marked.

Then the ratio of the number of squares marked of acne was divided by the surface area of the forehead, and then multiplied by 100 to obtain the final percentage for that participant (Figure 7).

Figure 7

AQM Procedure

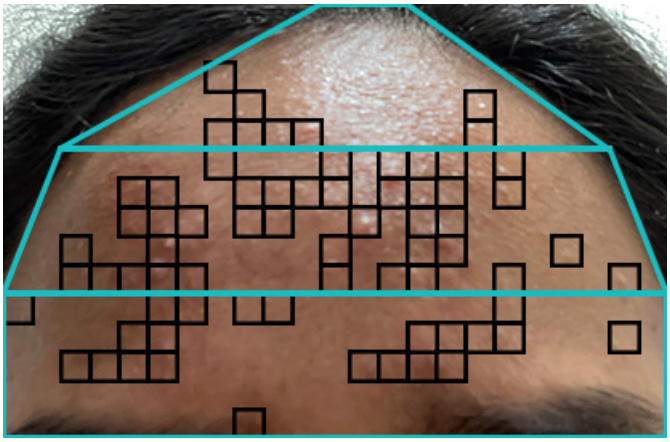


Note. The AQM procedure was performed for each participant, with a total of 19 participants (n = 19) in total. All of this data was recorded in the data sets to be able to sort for the data analysis later on. *Microsoft SmartArt, Agarwal, Prisha.*

If the participant had less than 10% acne coverage, they were recorded as in the control group with none to mild acne coverage. If they had greater than 10% acne coverage, they were recorded as in the experimental group with moderate to severe acne coverage (Figure 8).

Figure 8

AQM Example on Experimental Group Participant



Note. Here, trapezoids and rectangles were used for forehead surface area. This participant had 21% acne coverage, which is quite large and is accurate to the extent that they have many acne lesions. *Photoshop, Agarwal, Prisha.*

Results & Conclusion

An independent samples t-test was used to compare the means between the control and experimental groups for self-esteem and social anxiety. This returns a p-value which indicates whether the difference is statistically significant enough, which is a 95% confidence in the difference between the data since a significant p-value should be below 0.05, or 5%. For both self-esteem and social anxiety, RStudio returned a t-test p-value

greater than 0.05, indicating that the difference between the averages of the control and experimental groups is not statistically significant. This does not indicate 0 significance, but it indicates that the confidence in the difference is not above 95%, which is what is needed to reject the null hypothesis (Figure 9).

Figure 9

Independent Samples T-test Results in RStudio

```
        Two Sample t-test

data:  se by group
t = 0.91177, df = 17, p-value = 0.3746
alternative hypothesis: true difference
in means between group con and group exp
is not equal to 0
95 percent confidence interval:
 -3.090807  7.795352
sample estimates:
mean in group con mean in group exp
         21.72727          19.37500
```

Note. The *p*-value for self-esteem was 0.3746, indicating that there is a 62.54% confidence in the difference between the averages. However, the averages are trending in the direction proposed in the hypothesis. *RStudio, Agarwal, Prisha.*

```
        Two Sample t-test

data:  sa by group
t = -0.70342, df = 17, p-value = 0.4913
alternative hypothesis: true difference
in means between group con and group exp
is not equal to 0
95 percent confidence interval:
 -30.49522  15.24522
sample estimates:
mean in group con mean in group exp
           21.000            28.625
```

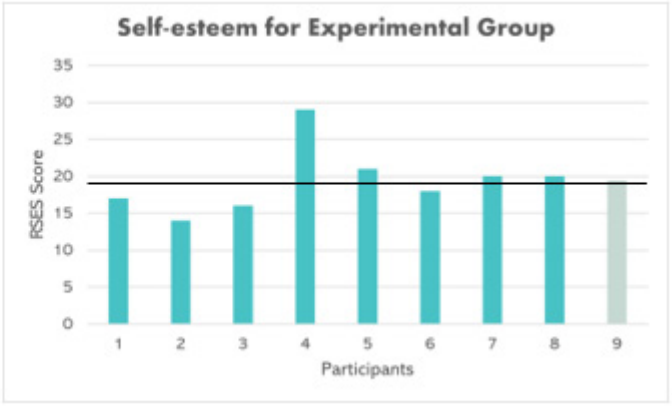
Note. The *p*-value for self-esteem was 0.4913, indicating that there is a 50.87% confidence in the difference between the averages. However, the averages are trending in the direction proposed in the hypothesis. *RStudio, Agarwal, Prisha.*

These *p*-values indicates that the null hypothesis is failed to be rejected, and further steps are required to continue this research. However, the averages for each group are related to the hypothesis proposed, because the experimental group did have a lower self-esteem and higher social anxiety than the control group, but more research and data is needed to confirm the rejection of the null hypothesis (Figure 10).

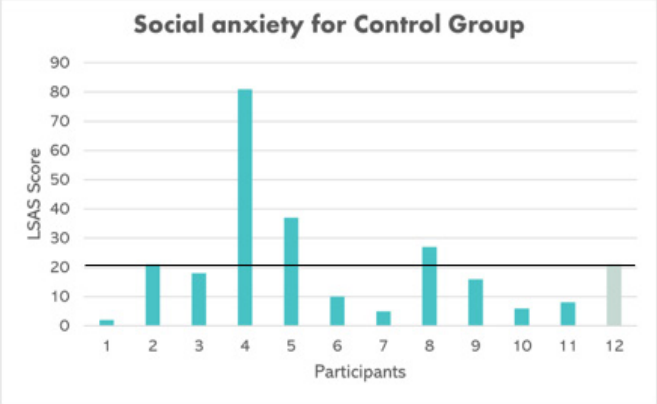
Figure 10

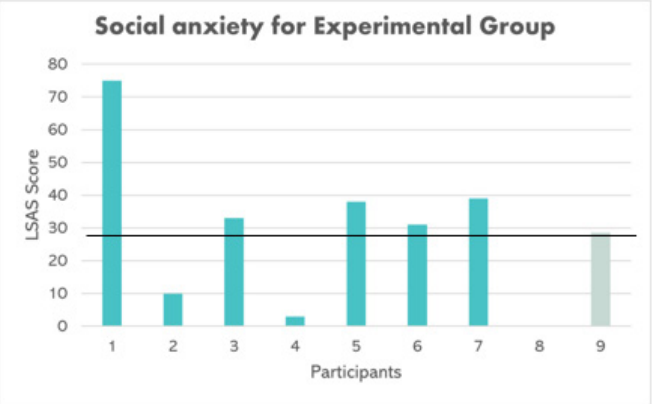
Graphical Representation of All Data





Note. The line helps highlight the different outliers present in the data, however self-esteem scores were more distributed for all participants compared to social anxiety. *Excel, Agarwal, Prisha.*



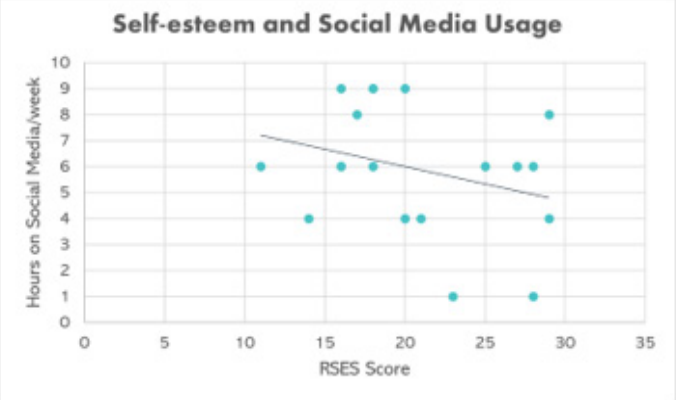


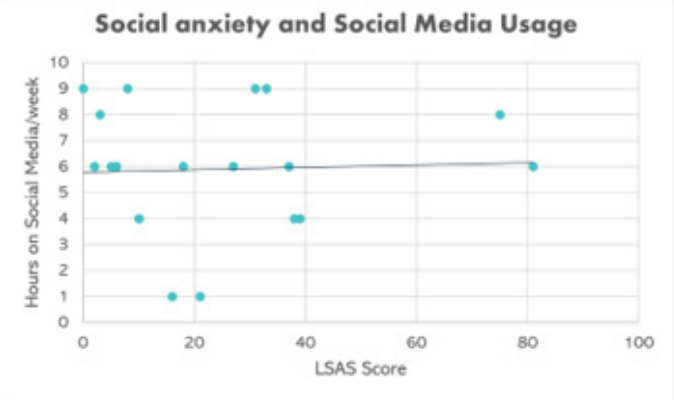
Note. The line is steeper for self-esteem, indicating that as self-esteem increases social media usage decreases. However, the data is very scattered, and the best fit line is not the best trendline. *Excel, Agarwal, Prisha.*

Social media could have also been a confounding variable that contributed to the varying self-esteem and social anxiety of the participants. Confounding variables can end up affecting results, especially in a virtual survey format such as this study. There was a slight correlation between the usage of social media and the dependent variables in this study, however it was not significant enough to be concluded to be a variable affecting the scores of participants (Figure 11).

Figure 11

Graphical Representation of Social Media Usage and DV





Note. The line is steeper for self-esteem, indicating that as self-esteem increases social media usage decreases. However, the data is very scattered, and the best fit line is not the best trendline. *Excel, Agarwal, Prisha.*

Discussion

Although *acne vulgaris* has been shown to have negative consequences and psychological effects, this study found that the effect of *acne vulgaris* on South Asian adolescents was not statistically significant. However, social media's role as a confounding variable could have contributed to the results, especially because social media has been studied before to affect anxiety and depression in adolescents who use it. The Liebowitz Social Anxiety Scale also may have been too focused on the external perceptions, which is straying further from the self-perception associated with self-esteem. The scale may have also asked questions not as related to social situations that are

affected by acne, such as "eating in public" or even "calling someone in public."

A power analysis was also conducted to analyze the results further, using standard deviations and Cohen's *d* in order to figure out whether the results were meaningful enough, or whether a greater sample size is needed. This value indicated the need for a greater sample size from the current 19 participants, and more participants could potentially produce a significant *p*-value (Figure 12).

Figure 12

Power Analysis

```
t test power calculation

          n1 = 11
          n2 = 8
           d = 0.45
   sig.level = 0.05
       power = 0.1499422
 alternative = two.sided
```

Note. A value between 0.1-0.3 indicates the need for a greater sample size. *RStudio, Agarwal, Prisha.*

Bibliography

Acne clinical guideline. American Academy of Dermatology. (n.d.). Retrieved January 23, 2023, from https://www.aad.org/member/clinical-quality/guidelines/acne.

Gorelick, J., Daniels, S. R., Kawata, A. K., Degboe, A., Wilcox, T. K., Burk, C. T., & Douse-Dean, T. (2015, May). *Acne-related quality of life among female adults of different races/ethnicities*. Journal of the Dermatology Nurses' Association. Retrieved October 5, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4456171/.

Hazarika, N., & Archana, M. (2016). The psychosocial impact of acne vulgaris. *Indian Journal of Dermatology*, *61*(5), 515. https://doi.org/10.4103/0019-5154.190102

Huynh, Q. T., Nguyen, P. H., Le, H. X., Ngo, L. T., Trinh, N.-T., Tran, M. T.-T., Nguyen, H. T., Vu, N. T., Nguyen, A. T., Suda, K., Tsuji, K., Ishii, T., Ngo, T. X., & Ngo, H. T. (2022). Automatic acne object detection and acne severity grading using smartphone images and artificial intelligence. *Diagnostics*, *12*(8), 1879. https://doi.org/10.3390/diagnostics12081879

Isotretinoin: Patient safety. American Academy of Dermatology. (n.d.). Retrieved January 16, 2023, from https://www.aad.org/public/diseases/acne/derm-treat/isotretinoin/safety.

from Lish, J. (2016, December 14). *Liebowitz Social Anxiety Scale*. Dr. Jennifer Lish. Retrieved January 23, 2023, from https://worcestercognitivebehaviortherapy.com/forms/liebowitzsocialanxietyscale/.

O'brien, S., Lewis, J., & Cunliffe, W. (1998). The Leeds revised acne grading system. *Journal of Dermatological Treatment*, *9*(4), 215–220. https://doi.org/10.3109/09546639809160698

Rao, Jaggi M. D. (2022, September 16). *Acne vulgaris*. Practice Essentials, Background, Pathophysiology. Retrieved October 5, 2022, from https://emedicine.medscape.com/article/1069804-overview#:~:text=Acne%20vulgaris%20is%20a%20common,also%20the%20back%20and%20chest.

Ritvo, E., Del Rosso, J. Q., Stillman, M. A., & La Riche, C. (2011). Psychosocial judgements and perceptions of adolescents with acne vulgaris: A blinded, controlled comparison of adult and peer evaluations. *BioPsychoSocial Medicine*, *5*(1), 11. https://doi.org/10.1186/1751-0759-5-11

University of Maryland. (2021, June 11). *Using the Rosenberg self-esteem scale*. SOCY Home. Retrieved January 23, 2023, from https://socy.umd.edu/about-us/using-rosenberg-self-esteem-scale.

Zaenglein, A. L., Pathy, A. L., Schlosser, B. J., Alikhan, A., Baldwin, H. E., Berson, D. S., Bowe, W. P., Graber, E. M., Harper, J. C., Kang, S., Keri, J. E., Leyden, J. J., Reynolds, R. V., Silverberg, N. B., Stein Gold, L. F., Tollefson, M. M., Weiss, J. S., Dolan, N. C., Sagan, A. A., … Bhushan, R. (2016). Guidelines of care for the management of Acne Vulgaris. *Journal of the American Academy of Dermatology*, *74*(5). https://doi.org/10.1016/j.jaad.2015.12.037

Examining the Efficacy of Dark Chocolate on Test-Taking Anxiety in High School Students

Zohaib Lughmani

Tesla STEM High School

Abstract

Ingestion of any substance impacts and alters a person's mood and brain. However, can the ingestion of Dark Chocolate positively impact stress, in turn helping with cognition and general mood in humans? Students experiencing test taking anxiety before an academic test may experience discomfort and stress which can cause a decrease in cognitive ability and focus on the test. Ingestion of dark chocolate may decrease stress levels for students to maintain cognitive function in times of stress such as an academic exam. The purpose of this study was to investigate the effects of the acute ingestion of dark chocolate before an exam on test taking anxiety. Screened participants ($n = 19$) were asked to ingest a dark chocolate or a white chocolate control 20 minutes before their test. They were then given the STAI-1 anxiety questionnaire 15 minutes after ingestion to measure their current stress levels. Interventional group STAI scores were significantly lower ($p < 0.034$) than the control scores with a 95% confidence interval. Dark Chocolate as an intervention before a test may be helpful in reducing test taking anxiety.

Introduction

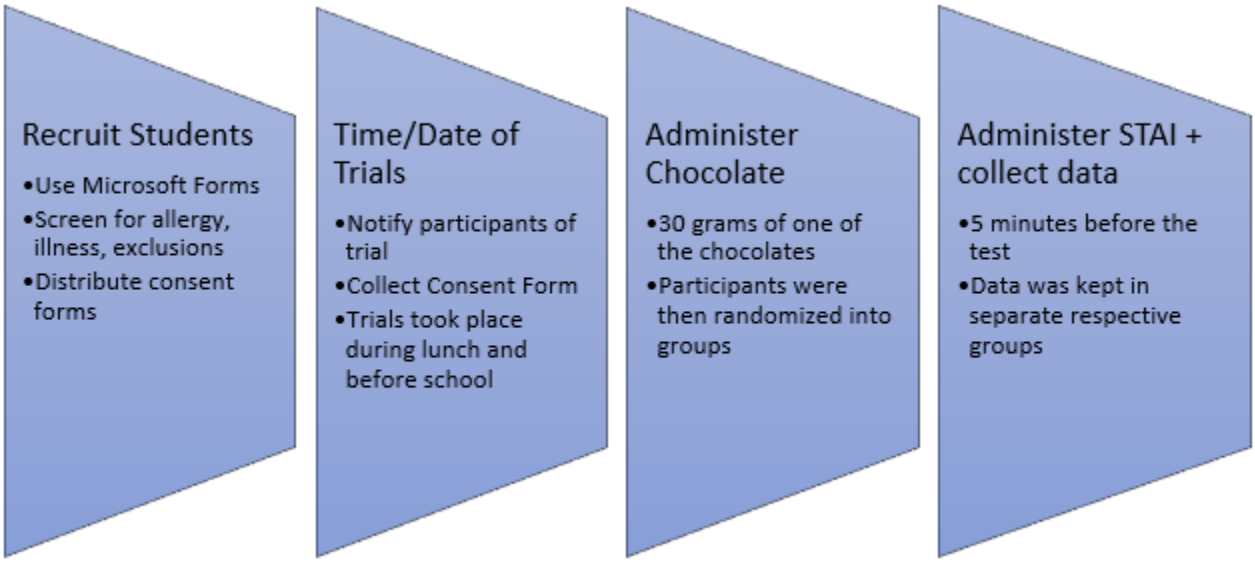
Ingestion of any substance impacts and alters a person's mood and brain; this study seeks to investigate the the ingestion of Dark Chocolate correlated with a positive impact on stress, in turn helping with cognition and general mood in humans. Test taking anxiety is a major problem many students experience in their academic career. Putwain and Daly (2014) reported that 16.4% of British secondary students have test anxiety. Furthermore, Thomas et al. (2018) found that about 25% of undergraduate university students are highly test anxious, suggesting that this issue is an international one. In high stakes environments such as final exams, students can 'crack' under pressure due to test anxiety, leading to harmful habits such as avoiding tests and academic work. In this way, test taking anxiety can be a major obstacle to achieving academic excellence. Prior to test taking, ingesting dark chocolate may reduce test taking anxiety by increasing cognitive function and memory. Research such as an article by Lamport & Christodoulou (2020) found that commercially available dark chocolate increased verbal and episodic memory after ingestion compared to the same dosage of white chocolate. Cocoa products contain chemicals called flavonoids which have been found to increase cognitive function (Tsang 2019). Cocoa contains low amounts of caffeine which have been shown to positively affect mood and cognitive function (Penetar 1994). This project seeks to investigate the efficacy of short-term dark chocolate (85%) ingestion before an academic test for lowering test anxiety in a randomized, controlled setting. Students taking exams face high incidence of anxiety which prevents them from performing up to their potential, despite preparation for the exam. The hypothesis is that taking a dose of 85% dark chocolate will decrease test anxiety compared to the white chocolate group, and that state anxiety scores will be lower for the experimental (dark chocolate) group than the control (white chocolate).

Methods

Poster flyers for sign-ups around the school were advertised for each period of classes and periods (precalculus, calculus, and psychology, as multiple trials were done). QR code on sign up form directed students interested in the study to a Microsoft Forms link. Forms asked if participants are in 1st or 4th period, as time was needed before test for ingestion and STAI questionnaire response (before school or during lunch). If the student was interested, Screening occurred. Any students who were pregnant, planning to get pregnant in the next two weeks, breastfeeding, taking other medications, suffering from long term illnesses, had GI disorders, or allergies to food, were excluded from the study. If a person answers yes to three safety questions they were not enrolled in the trial. A question screening for age was also used. Participants were healthy and not on any other medications. At the bottom of the form students were directed to download the consent form. Participants' parents or guardian must have signed this as well. After the form was filled out, they were asked to turn it in before the trial. After sign ups, participants were emailed the date and time of trial. Consent forms were discarded after the trials were completed and not looked at but only to be kept as proof of consent through the experiment. Email address for the forms account and files were destroyed after trials per fair rules.

Participants were asked to show up to their tests 20 minutes early. Participants were asked to ingest 30 grams of 85% Dark Chocolate or 30 grams of White Chocolate. They were administered the chocolate 20 minutes before their test. 5 minutes before their tests, they were administered the STAI-1 anxiety questionnaire and were asked to fill it out. This process took about 5 minutes. Responses will be collected from each group and kept separate per group. Participants' number and group were written on the paper. Participants entered the classroom and were given the test after forms were collected.

Figure I: Methodology flow chart, left to right.



*breakdown of tasks through trial date

Table I: Data collected including mean, median, and mode.

Participant # (WC)	STAI score	White Chocolate	Participant # (DC)	STAI score	Dark Chocolate
10	56	0	6	40	1
3	58	0	17	42	1
12	48	0	2	54	1
4	48	0	8	29	1
5	66	0	13	33	1
15	38	0	11	30	1
9	25	0	14	40	1
16	31	0	Mean	38.28571429	
7	52	0	Median	40	
18	64	0	Mode	40	
1	49	0	T Test	0.026588396	
19	62	0			
Mean	49.75				
Median	50.5				
Mode	48				

*simple calculations were done in excel, such as measures of central tendency and variance.

Conclusion

A mean STAI y-1 score of 49.75 was found for the white chocolate group. A mean of STAI y-1 of 38.29 was found for the dark chocolate group. Standard deviation from the mean for the white chocolate group was 12.38. Standard deviation from the mean of the dark chocolate group was 8.01. The confidence level of the white chocolate group was 5.97 of the STAI scores and for the dark chocolate 7.41. Upper and lower limit of white chocolate was 55.72 - 43.78. Upper and lower limit of dark chocolate was 46.00 - 30.89. T-test was used to analyze data from both groups with a t-score of 2.31 and a p-value of 0.034 from the subsequent T-Test proving the hypothesis.

This project studied the effects of dark chocolate on test taking anxiety and found a significant difference between the two groups. P value was determined statistically significant for differences in each group (p=0.034). Null hypothesis is accepted, meaning that the hypothesis is correct. Hypothesis stated that the dark chocolate group would show a lower stress level overall compared to the white chocolate. They hypothesis was proven correct and because the p-value is statistically significant, the study is generalizable to a larger population. The STAI-1 questionnaire is a measure of anxiety or stress in an acute, momentary setting, and is measured from a scale of 20-80 (20 being the least amount of stress, 80 being the highest). The mean STAI-1 score for the dark chocolate group was 38.29 points. The mean STAI-1 score for the white chocolate group was 49.75 points. Median score for the white chocolate group was 50.5 points and mean for the dark chocolate group was 40 points. Mode of the white chocolate group was 48 points and the mode for the dark chocolate group was 40 points. All analyzed data, points to a lower stress level of students who ingested dark chocolate before their exam compared to the students who ingested the white chocolate prior to their exam. The margin of error overlaps for both groups due to smaller sample size (n=19). This means that because the margin of error was higher, a larger sample size is needed to solidify the correlation that dark chocolate lowers stress levels in general. This project was limited in aspects which include: a small sample size, no use of double masking, higher sugar levels in the white

Table II: Data analysis and resulting p-value

t-Test: Two-Sample Assuming Unequal Variances

	White Chocolate	*Dark Chocolate*
Mean	49.75	38.28571429
Variance	167.1136364	74.9047619
Observations	12	7
Hypothesized Mean Difference	0	
Degrees freedom (df)	17	
t Stat	2.310164271	
P(T<=t) two-tail	0.03369586	
t Critical two-tail	2.109815578	
Standard Deviation	12.37689379	8.01274495
Confidence Value	5.965471854	7.410553099
Upper Limit	55.71547185	45.69626739
Lower Limit	43.78452815	30.87516119

chocolate than the dark chocolate, and no baseline for anxiety scores.

If these changes were implemented in a new study, the study would be more reliable, but with limited material and time constraints, a small project like this one shows a difference in flavanol rich chocolate in parallel with low amounts of caffeine and theobromine, compared to a confectionary white chocolate control. Because of a higher margin of error, this study does not prove that dark chocolate will lower test anxiety in every case. Due to a low p-value however, the results show that the ingestion of dark chocolate is likely to reduce test taking anxiety in students when taken prior to an exam or test.

Bibliography

Katz, D. L., Doughty, K., & Ali, A. (2011). Cocoa and chocolate in human health and disease. Antioxidants & redox signaling, 15(10), 2779–2811. https://doi.org/10.1089/ars.2010.3697

Lamport, D. J., Christodoulou, E., & Achilleos, C. (2020). Beneficial Effects of Dark Chocolate for Episodic Memory in Healthy Young Adults: A Parallel-Groups Acute Intervention with a White Chocolate Control. Nutrients, 12(2), 483. https://doi.org/10.3390/nu12020483

Letenneur L, Proust-Lima C, Le Gouge A, Dartigues JF, Barberger-Gateau P. Flavonoid intake and cognitive decline over a 10-year period. Am J Epidemiol. 2007 Jun 15;165(12):1364-71. https://doi,org/10.1093/aje/kwm036.

Martín, M. A., Goya, L., & de Pascual-Teresa, S. (2020). Effect of Cocoa and Cocoa Products on Cognitive Performance in Young Adults. Nutrients, 12(12), 3691. https://doi.org/10.3390/nu12123691

Montagna, M. T., Diella, G., Triggiano, F., Caponio, G. R., De Giglio, O., Caggiano, G., Di Ciaula, A., & Portincasa, P. (2019). Chocolate, "Food of the Gods": History, Science, and Human Health. International journal of environmental research and public health, 16(24), 4960. https://doi.org/10.3390/ijerph16244960

Tsang, C., Hodgson, L., Bussu, A., Farhat, G., & Al-Dujaili, E. (2019). Effect of Polyphenol-Rich Dark Chocolate on Salivary Cortisol and Mood in Adults. Antioxidants (Basel, Switzerland), 8(6), 149. https://doi.org/10.3390/antiox8060149

The Relationship Between Different Genres of Media Consumed to Burnout Symptoms Reported

Samantha Novitski

Tesla STEM High School

Abstract

Academic burnout is an issue prevalent in almost every educational sector in society, ranging from history majors, dental students, and even teachers; yet does social media play a role in developing academic burnout? This study seeks to examine whether different types of media can play a role in high school students reporting severe academic burnout. It was hypothesized that there will be a positive correlation between consumption of online media to academic burnout symptoms reported, and a negative correlation with consumption of physical media. A sample (n=24) of teenagers aged 15-17 were asked to self-report data on daily consumption of different types of media (in hours) and complete a burnout questionnaire, established by the Maslach Burnout Inventory-Student Survey including sections of cynicism, self-efficacy, and emotional exhaustion. Daily media consumption questions surrounded 6 variables, which included social networks (ex. Facebook), media sharing networks (ex. TikTok), bookmarking/blogging (ex. Pinterest), discussion forums (ex. Reddit), physical tests (ex. books) and online texts (ex. Kindles). Through results and statistical analysis, unfortunately none of the data was concluded to be statistically significant. Although reliability measures like Cronbach Alpha showed values from 0.6734-0.7541 (indicating the methodology were reliable), actual p-values for results of each burnout dimensions were high; 0.05 imply statistically significance, but results showed 0.139 for emotional exhaustion, 0.863 for cynicism and 0.390 for self-efficacy. In the end, the hypothesis cannot be supported, due to a lack of significant data, which could be solved with having more future participants.

Introduction

Academic burnout in current growing youth is an increasing issue that proclaims its roots in the media students are consuming—especially ones that utilize short attention spans and focus on short spanning media. Academic burnout has impacted many sectors of the educational field, with studies showing its prevalence in high schoolers, nursing students, dental students and even history majors. In its definition, academic burnout is a subsection of many different types of career burnouts, though it is often divided into three dimensions, also known as primary symptoms. The first is emotional exhaustion, in direct reference to the schoolwork assigned and given; the next is cynicism, which is one's negative attitude towards school and learning; and the last is efficacy, which is one's low self esteem and evaluation in their school related achievements (Song 2022).

Looking at its very roots, academic burnout is quite prevalent in educational sectors like universities. According to a 2015 study in School of Dentistry at Ege University in Turkey, experimenters examined the on-campus prevalence of burnout across dental majors. Researchers were able to obtain relevant conclusive data—26.59% of students had reported symptoms of emotional exhaustion. In addition, 17.41% reported cynicism, and 19.56% had reported reduced academic efficacy (Atayain 2015). Other studies have shown similar results, such as one conducted in Mashahd

University of Medical Sciences in 2020, which reported that out of 21.60% of students at the campus surveyed, 39.20% had reported experiencing one or more symptoms of academic burnout (Shokrpour 2020). Even so, a 2013 longitudinal study was conducted, which examined academic burnout in Korean middle schoolers, concluding that over time the students slowly worsened in burnout symptoms reported (Kim et. al. 2013). Hence, there is space to assume educational campuses, whether it be in university or lower/higher education, will have some presence of academic burnout. But this is a troubling fact; the percentages presented of students reporting some symptoms of academic are also patterns observed in several other Universities and educational branches as well. There is a level of certainty that there exists an underlying cause in the development of academic burnout exists for these students.

To continue, in the realm of scientific realm of research, many studies have been conducted to investigate potential correlations between academic burnout and a variety of different variables. Such variables examined include marital status, gender differences, types of housing, year of education, and employment status (Shokrpour 2020). Appealingly enough, many of such studies have shown similar patterns between burnout; for example, females and younger/incoming students report more symptoms of academic burnout, respectively. While these studies consider general categories of variables (such as grade or even field of research) not much has been studied on specific, pinpoint habits people may have. In an interesting take of this concept, some researchers have considered social media. Social media is defined as a digital platform designed to allow users to post content onto a online network. Many different genres of social media have been established throughout the modern era; many of which can be uniquely categorized and will be further defined this paper.

However, there are no studies that correlate types of media use, such as physical media or social media, in relation to development of academic burnout symptoms. This study aims at different types of media that are accessible to students grades 9-12 in high school and how they effect the amount/severity of burnout symptoms reported. It can then be suggested that different genres of media that can be consumed, with examples being online texts such as Messenger or WhatsApp, physical texts such as books, media sharing networks such as Instagram, or social networks such as Facebook; could have a positive correlation to academic burnout symptoms reported in high school students in grades 9-12. Evidence to this claim comes from a study published on April 5th, 2022, which set to understand the correlation between abusive internet use with academic burnout and performance in a group of nursing students. Through using the Maslach Burnout Inventory-Student Survey (often abbreviated as MBI-SS), and a sample group of nursing students, results had demonstrated a positive correlation of an R-value of 0.302 within abusive internet usage and academic burnout. In short, as the internet usage increases, the severity and occurrence increase as well (Jafari 2022). But what Jafari's research team did not research was whether different internet platforms consumed (such as *Twitter*, *TikTok, SnapChat*, etc.) play a different role in developing academic burnout symptoms in students. This prompts the question on whether this statistical trend can apply to whether the type of media consumed on the internet or physically, could play a significant role on number burnout symptoms reported. Since research has already observed a positive correlation between abusive internet use and academic burnout symptoms, this should reign true in internet media, but the opposite for physical media, and perhaps different internet platforms could contribute to different rates of burnout symptoms reported. This experiment seeks to answer whether or not there are any correlations between physical versus online media consumption to academic burnout symptoms reported. From this, it can be said that there will be a negative correlation between physical media consumption to academic burnout symptoms reported. Online media that contains short span content, such as media sharing networks will correlate to having more academic burnout symptoms reported, versus long span content like bookmarking, or even online texts.

Operational Definitions of Variables

This study will consider 6 different subsets of media. The first are social networks, which is

online media where users are prompted to talk, interact and network with friends, family, co-workers, or business partners, and the user spends more than 30 minutes to 1 hour on minimum daily. *Examples include Facebook, Twitter, and LinkedIn.* The second is media sharing networks, which is online media where users are prompted to host videos or pictures and can be shared to friends or to garner "likes" and the user spends more than 30 minutes to 1 hour on minimum daily. *Examples include Instagram, TikTok, YouTube and Snapchat.* The third are discussion forums, which is online media where users are prompted to ask questions or engage in thread discussions and the user spends more than 30 minutes to 1 hour on minimum daily. *Examples include Reddit, Quora, and Digg.* The fourth is bookmarking and blogging, or online media where users are prompted to host blogs, pictures, or boards, and the user spends more than 30 minutes to 1 hour minimum daily. *Examples include Pinterest, Tumblr and WordPress.* The fifth are paper texts, or physical media where users consume physical texts for a period of 30 minutes to 1 hour daily. *Examples include books, newspaper, and comic books*. The sixth are online texts, or online media where users consume online text for a period of 30 minutes to 1 hour daily. *Examples include Wattpad, Archive of Our Own, Nook books, and Kindles.* Within this range, it is important to note that participants do not have to use every single platform in the subset, to say they "use a social network".

Another important variable to consider is academic burnout, and how it can be quantified. To do so, this study will utilize the Maslow Burnout Inventory for Students, also commonly abbreviated as MBI-SS. This scale was developed and created by Schaufeli et al, and has been reported to have a Cronbach's Alpha value of 0.740 (hence implying it is a reliable test for measuring academic burnout). In addition, the studies mentioned prior have made usage of this very same burnout measuring scale before, with promising degrees of validity, relevance, and confidence. The scale consists of questions ranging from 6 different answer options, for the 3 different academic burnout symptoms, which include; including exhaustion (*the feeling of exhaust towards one's studies*), cynicism (*the unaffected or distant attitude to one's studies*), and efficacy (*satisfaction towards one's studies, and expectations at school*). Each section will be scored on a point-by-point basis, with MonkeyPuzzle, the distributor of the survey in this study, providing a scoring outline.

Sampling Procedure

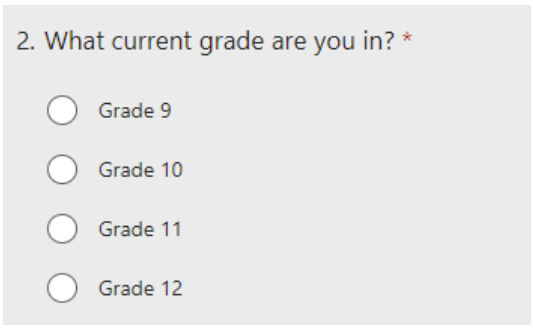
To conduct this study, participants had to be recruited, and specially selected. During a 60-day period, the survey was be distributed via a QR code on physical fliers and was open to responses starting January 5th 2023, ending March 5th 2023. Distribution was done across two high schools-- Tesla STEM High School and Eastlake High School. Once the 60-day period ended, responses for the survey were closed, and all participants (marked specifically by gender) were subsequently transferred to a Microsoft Excel document. Participants were randomly assigned (using a random number generator) either a "1" or a "2". Participants who selected the female option had their names placed into one group, while the remaining participants who selected the male option had their names placed in a separate group. Every name assigned with the number "2" in the female participant group was selected for their data to appear in the final sample. Every name assigned with the number "1" in the male participant group was selected for their data to appear in the final sample. To keep participant gender at an even 50/50 split, participants were leveled off to ensure an even distribution between genders (ex. If there were 55 females and 100 males, the male participants would be leveled off to have 55 males). As an alternative note, this experiment does not utilize race demographics with any correlational variables. Thus, no procedure had been in place for selecting random groups of participants, from different racial backgrounds.

In terms of conducting the study itself, the questionnaire survey was based on Microsoft forms, but did not require an account for participants to proceed with the study. The flyer had contained a QR code, which upon scanning it, participants will had been required to read a debrief of the survey, and indicate they understand the contents of the study, and wish to proceed. Following the debrief, participants were asked to fill out an Informed Human Consent form through Adobe Acrobat and attach it to the embedded area. Survey responses that did not have this form fully filled out and attached were disregarded. On the consent form,

participants under the age of 18 had to have the form filled out by their parents. Participants over the age of 18 were *not* required to do so. There was no distinction between the parental consent versus regular consent, as this is through the same, singular form. Additionally, checkboxes were included to confirm once more, that the participant was willing to be in the study, and that they agree to signing provided forms (which was included on that same page as well). Such forms include both the informed consent form. Participants had then been asked to complete a series of questions, beginning with demographics. Demographic questions included the grade level, specific age, race (if desired) and gender. After demographics, participants were asked questions about their media consumption habits. This included the following variables as indicated in the **operational definitions of variables** section.

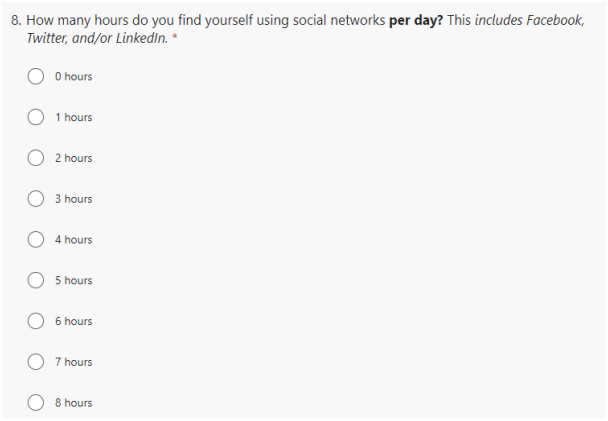
When starting the survey, participants were prompted to begin with the demographics section, with questions ranging from what grade they are a part of, gender, age and race (as pictured in figure 1). Participants were then asked to rate their consumption in hours, which included a range from 0-8, measured via self-reported data and on a daily basis (such is pictured in figure 2). After the media consumption section, participants were asked to answer a series of questions related to burnout, as indicated from the MBI-SS questionnaire, as provided by MonkeyPuzzle. Answers ranged from "extremely true", "once in a while" to "not true at all" (pictured in figure 3). Once the burnout section is completed, participants were able to submit the survey, which sent results over to the researchers. An in-depth debrief of the study was included in the introduction of the forms, as well as providing the informed option of dropping out of the study is desired.

Figure 1- Demographic Based Question



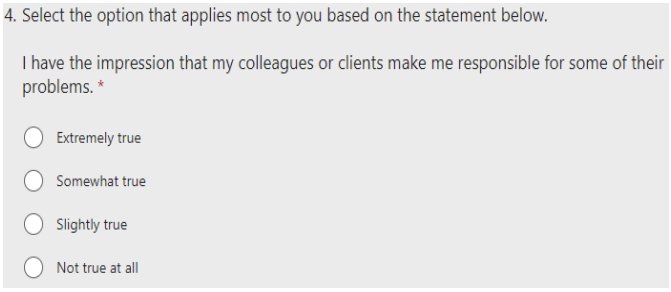
1 - Question prompting the grade of the participant. A follow-up question will be included indicating how old the participant is.

Figure 2- Activity Based Question



2 - Question indicating participant's activity in using social networks, which is one of the variables included in the study. The intensity of the participants usage is determined by the number of hours they spend on one of the listed platforms above.

Figure 3- Burnout Based Question



3 – A burnout-based question which requires a participant to reflect on personal experiences and apply them to a range of answers. These questions have been pre-selected and are apart of the MBI-SS index, used to measure student burnout. Scales and measuring abilities are pre-determined for this as well.

Data Analysis

Once all participants had successfully submitted their survey results, and the survey response time ended (two months), responses were collected, and resultingly analyzed to deduct a conclusion based on the hypothesis. Due to a lack of statistical significance (specifically in p-values and correlational coefficients) in the resulting analysis, the hypothesis that a positive correlation

would result from online media usage and burnout symptoms reported, and that a negative correlation would result between physical media usage and burnout symptoms reported could not be supported.

In the end, the survey was able to amass a total of 24 participants (all of whom had answered the survey in full), 12 of which were female, and 11 of which were male, with 1 prefer not to say (as pictured in table 1). For simplicity sakes, the participants data was not used in the final results, considering the difficulties and unreliability that could come about conducting a linear regression test on only one set of data. Hence, the analysis was confined to the 12 females and 11 males, a total sample size relatively small, but considering the almost even split, is minimally representative of a (small) population in demographics. Continuing, when consider the racial demographics, raw data contained 13 participants who were Asian, 6 White participants, 1 Hispanic participant, and 2's for both Multi-racial and Middle Eastern, which can be observed in table 2. Once more for simplicity sakes, the white participants results would be divided into one group, while the remaining group would be divided into another group, titled "minority", for a linear regression to be conducted on both groups (with some form of insignificance being avoided if only 2 people had their results analyzed). In terms of ages and graduation year, 66.6% of participants were 16 and expected to graduate in 2024, 33.3% were 17 and expected to graduate in 2025.

Table 1- Gender Demographics of Participants

Gender Demographics of Participants

Gender	Freq.	Percent
Female	12	50.00
Male	11	45.83
Prefer not to say	1	4.17
Total	24	100.00

Created with Datawrapper

1-Raw results of the gender demographics question in the questionnaire. The prefer not to say participant will not appear in final data analysis to avoid high chances of statistical insignificance due having only 1 piece of data to analyze.

Table 2- Racial Demographics of Participants

Race Demographics of Participants

Race	Freq.	Percent
Asian (Including Indian Subcontinent)	13	54.17
Hispanic, Latino or Spanish Origin	1	4.17
Middle Eastern or North African	2	8.33
Multi-racial	2	8.33
White	6	25.00
Total	24	100.00

Created with Datawrapper

2-In the raw results for racial demographics, it can be observed that the median results are for Asian participants, with the runner ups being White participants. Remaining racial groups are generally low in number, ex. 2 for multi-racial. 54.17% of participants are Asian, implying a clear majority.

An interesting piece of data came from the results for the raw activity-based questions. As observed in table 4, participants reported not using any type of media for more than an hour a day, for all but 2 media variables—those being 1 hour for social networks, and ~2 hours and 30 minutes for media sharing networks. All other media variables were below 45-30 minutes daily. Additionally, raw burnout-based questions proposed similarly intriguing results, with table 3 showing one set of mean answers to the burnout based questions. Observing table 3**,** the MBI-SS responses were scaled from 0 (never) to 6 (every single day), and the raw results displayed means ranging from 4.912-2.542, indicating no clear trends of burnout being mild, low, or severe without further data analysis.

Table 3- Mean Participant Responses to Burnout Based Questions

Mean Participant Responses to Burnout Based Questions

Variable	Obs	Unique	Mean	Min	Max	Label
burn1	24	4	4.916667	3	6	I feel emotionally drained by my schoolwork.
burn2	24	6	3.875000	1	6	Learning with people all daylong requires a great deal of effort.
burn3	24	6	4.166667	1	6	I feel like my schoolwork is breaking me down.
burn4	24	4	4.833333	3	6	I feel frustrated by my schoolwork.
burn5	24	6	4.041667	1	6	I feel that I work too hard at my schoolwork.
burn6	24	5	2.541667	1	5	It stresses me too much to work in direct contact with people.

3-MBI-SS test results, which is only 1 snippet of 5. Answer choices ranged from 0 meaning never to 6 meaning every single day. The screenshot results imply no clear pattern of burnout within students but would need to be further examined through more data analysis.

Table 4- Number of Hours Used Per Type of Media

Number of Hours Used Per Type of Media

Variable	Obs.	Unique	Mean	Min	Max	Label
burn1	24	6	1.0000000	0	6	Hours: social networks (Facebook
burn2	24	7	2.5833330	1	7	Hours: media sharing (Instagram
burn3	24	4	0.3229167	0	1	Hours: discussion forums (Reddit
burn4	24	3	0.4166667	0	3	Hours: bookmarking/blogging (Pinterest
burn5	24	5	0.5138750	0	6	Hours: online texts (Wattpad..
burn6	24	5	0.7083333	0	4	Hours: physical texts (books..

4-Raw data for activity-based questions related to media consumption. Only 2 types of media are reported to be used for an hour or more a day, those being social networks and media sharing networks at 1 hour and 2.5 hours daily respectively. Remaining media is reported to be less than an hour (30 mins. To 20 mins.) daily.

In terms of further data analysis, a linear regression test was conducted on the program SPSS. Originally, it was presumed that a ANOVA test could have been conducted, but when considering the small sample size of 24, much room for error and statistical insignificance could've been made. Instead, a linear regression was conducted using each of the academic burnout symptom results (e.g. emotional exhaustion, cynicism, and self-efficacy), and done on age, gender, race, employment status, and the actual type of media (ex. media sharing networks, online texts, etc). The tests would be repeated 6 times to ensure all the types of media had a regression test on all 3 symptoms, totaling to 18 sets of regression tests, as for example, table 5-7 all display the regression tests made for physical texts. For all sets of regression tests, the trends were similar—statistical insignificance, whereas P-values must be below 0.05 to remain significant, but the tests P-values were all above (ex. 0.968 for age and emotional exhaustion as pictured in table 5, 0.863 for physical texts and cynicism as pictured in table 6, and 0.962 for female and self-efficacy as pictured in table 7). No less, correlational coefficients were close to 0.00, indicating no correlation. In some cases, such as in table 5, it was observed that the variable "female" had a P-value of 0.104, and a correlation coefficient of 0.486, when regressed with emotional exhaustion. While not entirely significant, this piece of data does support the trend that females are prone to experiencing some more severity over men in academic burnout, but does not present any new results or findings, novel to the scientific realm. Nonetheless, scatterplots of the results P-values show a similar concept, where the visual aid does not show an indication of a positive or negative trend overall between variables, all of which could be observed in charts 1-3. Hence the null hypothesis would be rejected as there is no positive correlation between online media use, and academic burnout symptoms reported. Within this, there would also be no negative correlation between physical media use and academic burnout symptoms reported.

As a final note on data analysis, the decision to conduct a linear regression separately on the variables with the burnout symptoms were made resultingly from correlational matrix tests, such as the Kaiser-Meyer Olkin Test and the Determinant of a Correlational Matric test. The KMO test had a result of 0.417, with the Determinant test having a

result of 0.470, all above 0, or below 1. This factor indicates a lack of an inter-correlational relationship between variables, for example, age and gender, and implies tests should be made separately to avoid statistical insignificance.

Table 5- Linear Regression of Demographic Variables and Emotional Exhaustion Question Results

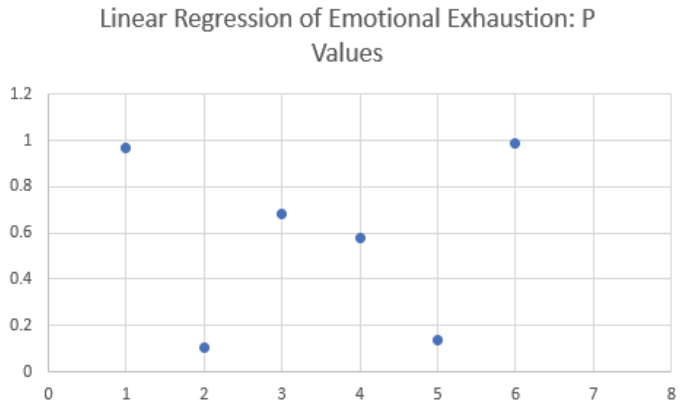
Linear Regression of Demographic Variables and Emotional Exhaustion Question Results

lowpersacc~p	Coefficient	Std. err.	t	P > | t |	95% conf. interval
age	−0.0089635	0.2199950	−0.04	0.968	-.4731123 .4551853
female	0.4860054	0.2825325	1.72	0.104	-.110086 1.082097
minority	−0.1364123	0.3239613	−0.42	0.679	-.8199109 .5470863
employ	0.1717722	0.3024613	0.57	0.578	-.4663653 .8099098
hours6	−0.2081890	0.1340813	−1.55	0.139	-.4910759 .0746979
_cons	0.0555103	3.5454820	0.02	0.988	-7.424802 7.535822

Created with Datawrapper

5-Results for the first linear regression test, with P values all being above 0.05, such as employment being 0.578. Correlational coefficients are also near 0.00, both metrics implying statistical insignificance, and no correlation between the variables.

Chart 1- Linear Regression of Emotional Exhaustion: P Values



1-The observed scatterplot, made from P-values does not indicate a positive or negative trendline, or as well by the values themselves, are not statistically significant to say so as well.

Table 6- Linear Regression of Demographic Variables and Cynicism Question Results

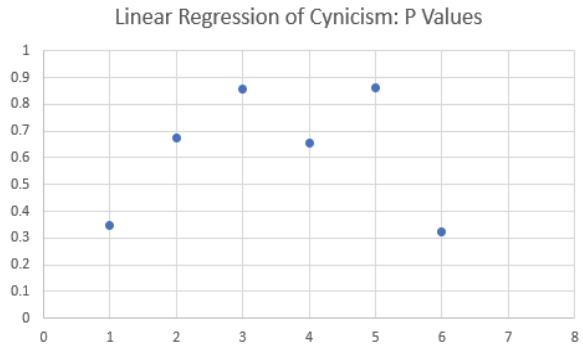
Linear Regression of Demographic Variables and Cynicism Question Results

lowpersacc~p	Coefficient	Std. err.	t	P > | t |	95% conf. interval
age	0.1981459	0.2048113	0.97	0.347	-.2339682 .6302601
female	0.1118781	0.2630326	0.43	0.676	-.4430723 .6668284
minority	0.0548843	0.3016021	0.18	0.858	-.5814406 .6912092
employ	0.1274569	0.2815860	0.45	0.657	-.4666376 .7215515
hours6	−0.0217994	0.1248273	−0.17	0.863	-.285162 .2415632
_cons	−3.3443150	3.3007790	−1.01	0.325	-10.30835 3.61972

Created with Datawrapper

6-Results for the second linear regression test, with P values all being above 0.05, such as female being 0.676. Correlational coefficients are also near 0.00, both metrics implying statistical insignificance, and no correlation between the variables (ex. minority being 0.05).

Chart 2- Linear Regression of Cynicism: P Values



2-The observed scatterplot, made from P-values does not indicate a positive or negative trendline, or as well by the values themselves, are not statistically significant to say so as well.

Table 7- Linear Regression of Demographic Variables and Self-Efficacy Question Results

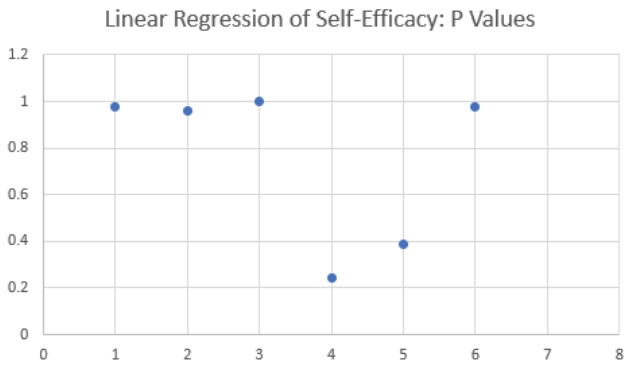
Linear Regression of Demographic Variables and Self-Efficacy Question Results

lowpersacc~p	Coefficient	Std. err.	t	P > | t |	95% conf. interval
age	0.0059565	0.2229503	0.03	0.979	-.4644274 .4763405
female	−0.0138028	0.2863279	−0.05	0.962	-.6179018 .5902962
minority	0.0008482	0.3283132	0.00	0.998	-.6918322 .6935285
employ	0.3692526	0.3065244	1.20	0.245	-.2774574 1.015963
hours6	−0.1197679	0.1358825	−0.88	0.390	-.406455 .1669191
_cons	−0.1149667	3.5931100	−0.03	0.975	-7.695766 7.465832

Created with Datawrapper

7-Results for the third linear regression test, *with P values all being above 0.05, such as age being 0.979. Correlational coefficients are also near 0.00, both metrics implying statistical insignificance, and no correlation between the variables (ex. female being -0.01).*

Chart 3- Linear Regression of Self-Efficacy: P Values



3-The observed scatterplot, made from P-values does not indicate a positive or negative trendline, or as well by the values themselves, are not statistically significant to say so as well.

Conclusions

Even though there was a lack of statistically significant data presented after data analysis, that does not imply there is nothing to conclude, or go forward with. For one, there is no acceptance of the hypothesis, meaning that there was no negative correlation between physical media consumption to academic burnout symptoms reported. Online media that contains short span content, such as media sharing networks will not correlate to having more academic burnout symptoms reported, versus long span content like bookmarking, or even online texts. This comes as a result of data analysis displaying consistent P-value above 0.05 (tables 5-7), implying that there is no significance between the different media variables and academic burnout symptoms reported, and once more, an acceptance of the null hypothesis.

Going further past data analysis, however, there are some claims to be made on the raw data received. Ushering back to table 4, none of the activity-based questions go beyond an hour, for only two media variables—social networks with an hour (ex. Facebook, Twitter, and LinkedIn), and social media networks with two and a half hours (ex. TikTok, Instagram, Snapchat, and Youtube). What this could imply is the prevalence of the social desirability effect. With this psychological phenomenon, participants would not willingly want to paint themselves in a bad or negative light, and would as a result answer questions in an disingenuous way to their actual habits. This for example could be answering a question with 1 hour, instead of 4 (which would be the actual allotted time they spend on a certain media). This could be avoided in the future with the requirement of data to be taken directly from activity trackers on users phones. This would eliminate the self-reported data bias that could have potentially appeared in the raw data, hence causing potential insignificance in data analysis.

On the note of future changes, it could be noted that 24 participants is a low number of data points, which could have contributed to the observed statistical insignificance. In the future, more allotted time to collecting participants should be made, along with a further effort to incentivize participants to participate in the study, hopefully resulting in more than 24 participants. Ideally, the golden number of participants would be over 100,

as this would remove future error of statistical insignificance, allowing for an ANOVA test to be conducted as well as providing a better chance of accepting the hypothesis.

In terms of future applications, there are many ways to interpret this research study. For one, it could be possible to take the idea of different social media genres specifically and compare them to academic burnout symptoms reported. Here, the social media would be grouped by how popular they are, so for example, comparing the monthly users of each app and separating them from "20 million users per month and less", "40-60 million users per month", "60-80 million users per month" etc. Additionally, it could be considered the specific algorithms of the social media platforms considered, such as TikTok, which may have an algorithm designed for prolonged use, versus Tumblr which may not have that algorithm pattern. Such coding differences may play a potential role in social media "addiction" and perhaps a contributing factor to academic burnout. This would all have to be investigated before making any assumptions, however.

Bibliography

Atalayin C, Balkis M, Tezel H, Onal B, Kayrak G. The prevalence and consequences of burnout on a group of preclinical dental students. Eur J Dent. 2015 Jul-Sep;9(3):356-363. doi: 10.4103/1305-7456.163227. PMID: 26430363; PMCID: PMC4569986.

Faranak Jafari, Maryam Janatolmakan, Safura Khubdast, Seyyed Mohsen Azizi, Alireza Khatony, "The Relationship of Internet Abusive Use with Academic Burnout and Academic Performance in Nursing Students", BioMed Research International, vol. 2022, Article ID 2765763, 6 pages, 2022. https://doi.org/10.1155/2022/2765763

Hadžibajramović E, Hansson M, Akerstrom M, Dencker A, Hensing G. Burnout among midwives-the factorial structure of the burnout assessment tool and an assessment of burnout levels in a Swedish national sample. BMC Health Serv Res. 2022 Sep 16;22(1):1167. doi: 10.1186/s12913-022-08552-8. PMID: 36114499; PMCID: PMC9482233.

Kim B, Lee M, Kim K, Choi H, Lee SM. Longitudinal Analysis of Academic Burnout in Korean Middle School Students. Stress Health. 2015 Oct;31(4):281-9. doi: 10.1002/smi.2553. Epub 2014 Jan 23. PMID: 24452964.

Oloidi, F. J., Sewagegn, A. A., Amanambu, O. V., Umeano, B. C., & Ilechukwu, L. C. (2022). Academic burnout among undergraduate history students. Medicine, 101(7), e28886. https://doi.org/10.1097/md.0000000000028886

Shokrpour N, Bazrafcan L, Ardani AR, Nasiraei S. The factors affecting academic burnout in medical students of Mashahd University of Medical Sciences in 2013-2015. J Educ Health Promot. 2020 Sep 28;9:232. doi: 10.4103/jehp.jehp_83_20. PMID: 33209924; PMCID: PMC7652071.

Song, W., Wang, Z., & Zhang, R. (2022). Classroom Digital Teaching and College Students' Academic Burnout in the Post COVID-19 Era: A Cross-Sectional Study. International Journal of Environmental Research and Public Health, 19(20), 13403. https://doi.org/10.3390/ijerph192013403

Quantifying Depression: An art-based approach to reducing the self-report bias in diagnosing depression using Koch's Baum Test and color psychology

Saachi Sharma

Nikola Tesla STEM High School

Abstract

The current ways to diagnose depression are heavily dependent on patients self-reporting their symptoms which is closely related to its high rate of misdiagnoses. The purpose of this investigation was to develop a more accurate screening for major depressive disorder by increasing the ambiguity of the screenings by using Koch's Baum Test – a projective test in which patients draw a tree – in combination with color psychology. Differences between depressed and healthy people that are undetectable with the current self-reported screenings alone are quantifiable with this novel colored version of Koch's Baum Test. Before participants were asked to draw a tree, their group was determined using the Beck Depression Inventory (BDI) – scores 0-14 made up the healthy group, 15-30 made up the mild-to-moderate depression group, and scores exceeding this range were excluded from the study. Color preference was measured in several ways, including the relative luminance and total number of colors used. Additionally, measurements of the tree were taken to validate the accuracy of Koch's Baum Test in relation to depression. Though the tree measurements did not have significant results across all trials, the total number of colors used by depressed participants had a strong positive correlation to BDI scores, $r(15) = 0.72$, $p < .001$. This study was successful in finding a correlation between color preference and depression using Koch's Baum Test and can be furthered to be used in conjunction with the current self-reported screenings for depression to minimize the rate of misdiagnosis of the illness.

Introduction

An alternative to relying on self-report depression indexes that reduces the misdiagnoses of the illness is a colored version of Karl Koch's Baum Test. Depression is a mental illness that affects more than 800,000,000 people worldwide, and the World Health Organization has projected it to be the leading health problem in 2030 (Gu et al., 2020). To diagnose the illness, a symptomatic approach is used in which patients report their symptoms to a professional. Although meant to be reported honestly, this allows for misinformation, whether that be lesser or exaggerated answers, to influence the final diagnosis. In turn, the rate of misdiagnoses of major depressive disorder is nearly 66% (Vermani et al., 2011). An approach to lowering this rate is reducing the patient's power over their final diagnosis by increasing the unconscious actions that are screened for as symptoms of major depressive disorder.

Previous attempts at making such depression screenings include Kock's Baum Test also known as the Tree Test: a mental projection test in which patients draw a tree that is then assessed for quantitative indicators of depression such as the trunk width or canopy area. In a 2020 study on applying the Baum Test to depression, 5 of the 12 tree features analyzed showed significantly different measurements when drawn by someone with major depressive disorder and someone without (Gu et al., 2020). In this sort of screening, patients do

not know which part of the tree they draw is being analyzed to diagnose them, so the ease with which they can manipulate results decreases.

To strengthen the validity of Koch's Baum Test as it relates to major depressive disorder, the present study reanalyzed the significance of 3 of the 5 tree factors that proved conclusive in the earlier study. Additionally, a new factor – color psychology – was added to this projective test to identify more differences between the unconscious thoughts of people with and without depression.

Colors are all integrated throughout the world and are processed both consciously and unconsciously. They have evolved to represent emotions and even the concepts around us, but these associations vary with differing mindsets and populations. Take the example of the color blue, depending on one's exposure to the color, it could be paired with a feeling of sadness or calmness (Fugate et al., 2019). These different pairings can be attributed to how the color is processed unconsciously, which occurs in the early primary visual cortex (V1) of the brain (Breitmeyer, et al., 2004). The present study aimed to show that this unconscious processing is influenced by prolonged emotional states such as that of major depressive disorder; this is partially supported by another study that concluded depressed subjects have a relatively suppressed perception of contrasting luminance than healthy subjects, leading them to prefer colors with the highest contrasts such as black, white, and grey (Sevda et al., 2016). All of these color-emotion studies show that there is a difference in the way different people perceive colors, and the present study applied that concept to depression as a way to help diagnose the illness.

To develop a more accurate major depressive disorder screening, this study aimed to build on the psychological component of color processing in conjunction with the existing Koch's Baum/Tree Drawing Test. The new factor of color provides an extra layer of unconscious differences between healthy and depressed people that are undetectable by the current self-reported screenings.

Methodology

Variables

Beck Depression Inventory (BDI): The BDI is a currently used diagnostic tool for depression in which patients report their own symptoms of the illness on a 4-point scale. The answers are totaled and the higher the number, the higher the indication of depression is. The BDI has 21 questions total and the highest score possible is 63, which would indicate the most extreme depression. Participants of this study filled out the BDI, which was scored by the researcher, and were placed in one of three categories based on their score. A score of 0-14 corresponded to the healthy group, referred to as H, and a score of 15-30 corresponded to the mild-moderate depression group, referred to as D. The last group was participants who scored above 30 on the BDI and were not included in the study for safety purposes due to the vulnerability of this population.

1.		
	0	I do not feel sad.
	1	I feel sad
	2	I am sad all the time and I can't snap out of it.
	3	I am so sad and unhappy that I can't stand it.
2.		
	0	I am not particularly discouraged about the future.
	1	I feel discouraged about the future.
	2	I feel I have nothing to look forward to.
	3	I feel the future is hopeless and that things cannot improve.
3.		
	0	I do not feel like a failure.
	1	I feel I have failed more than the average person.
	2	As I look back on my life, all I can see is a lot of failures.
	3	I feel I am a complete failure as a person.
4.		
	0	I get as much satisfaction out of things as I used to.
	1	I don't enjoy things the way I used to.
	2	I don't get real satisfaction out of anything anymore.
	3	I am dissatisfied or bored with everything.
5.		
	0	I don't feel particularly guilty
	1	I feel guilty a good part of the time.
	2	I feel quite guilty most of the time.
	3	I feel guilty all of the time.

Figure 1: First 5 questions of the 21 question Beck Depression Inventory.

Relative luminance (Y): Relative luminance is the brightness of a point in a color space normalized on a 0 to 1 scale with 0 being closer to the brightness of the darkest black and 1 being closer to the brightest white. The color space used to calculate relative luminance is the CIE color space (Fig. 2), which is the most accurate model of every color the human eye can perceive by way of accounting for the chromatic response of the eye based on the cones in the retina reacting to different colors. The equation for relative luminance (Eqn. 1) is derived using this model, specifically the Adobe RGB interface. In this study, Y values were used in percentage form and used to quantify colors into light and dark groups.

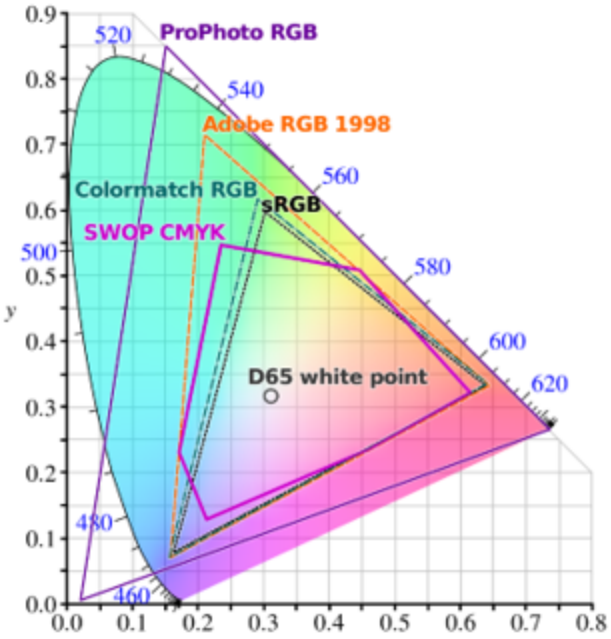


Figure 2: 1931 CIE Color Model - 2D versión.

$$Y = 0.2126R + 0.7152G + 0.0722B$$

where R, G and B are:

$$\text{if } RsRGB \leq 0.03928 \text{ then } R = \frac{RsRGB}{12.92}$$

$$\text{else } R = \frac{RsRGB+0.055}{1.055^{2.4}}$$

$$\text{if } GsRGB \leq 0.03928 \text{ then } G = \frac{GsRGB}{12.92}$$

$$\text{else } G = \frac{GsRGB+0.055}{1.055^{2.4}}$$

$$\text{if } BsRGB \leq 0.03928 \text{ then } B = \frac{BsRGB}{12.92}$$

$$\text{else } B = \frac{BsRGB+0.055}{1.055^{2.4}}$$

and RsRGB, GsRGB, and BsRGB are:

$$RsRGB = \frac{R8bit}{255}$$

$$GsRGB = \frac{G8bit}{255}$$

$$BsRGB = \frac{B8bit}{255}$$

Equation 1: Relative luminance value equation. Multiply answer by 100 to convert to a percent from a 0-1 decimal scale. Coefficients are from the y-coordinates from the points of the Adobe RGB triangle in Figure 2.

Sampling

Participants were recruited through flyers at various high schools and filled out informed consent forms before partaking in the study. They ranged from 13-17 years of age and had to have normal color vision to participate (CIE color model only accounts for normal color vision). Upon completion of forms, they were placed into one of three groups based on their results on the BDI with the three groups being H (Healthy), D (Mild-moderate depression), and illegible (subject was not allowed to proceed with study). Within the H and D groups, participants were assigned to one of two trials, A or B, using a random number generator. Participants did not know what group (H or D) or trial (A or B) they were in, but the researcher did, making this a single-blind study.

Based on their group and trial, participants were referred to as a 3-character index to ensure anonymity. The index started with the letter *H* or *D* followed by a number and a lowercase *a* or *b*. The H and D designated Healthy(H) or Mild to Moderate depression group(D) and the number was a random generation within the range of the group's total number of participants such that each member of the group received their own distinct number. The last letter was which trial they were randomly placed in.

Procedure

Regardless of their group and trial designations, all participants were given the same materials and instructions. A paper with the following was given to each participant:

1. You are to draw a tree on the sheet of paper before you using the colored pens provided – don't worry if it isn't the best drawing, you are not being tested on your art skills.

2. Before you begin, close your eyes, and meditate for around half a minute. Draw the tree that appears in the meditation. If there is no tree in the meditation, open your eyes and draw the tree that is most appealing to you.

3. When done with your drawing, leave it on the table and inform the administrator of the study.

Participants were then provided with an 8.5"x11" sheet of millimetric graph paper and 20 colored pens with the cap side up in no particular order. The pens were deliberately chosen such that they had opaque bodies so the amount of ink left in the pen wasn't visible and in turn would

not influence participants' inclinations to use a certain color. Prior to administering the study, the 20 colored pens were sorted into two groups based on their relative luminance values. Relative luminance was calculated using the RGB values of the pen caps, which were obtained through the Adobe Capture application. 10 of the pens had relative luminance values above 50%, which were categorized as the light colors (Table 2), and the remaining 10 that had relative luminance values below 50% were categorized as the dark colors (Table 1).

Table 1: Dark color pens were categorized by their relative luminance values being under 50.00%. The Staedtler color number refers to the pen brand's distinct label for each pen.

Staedtler Color #	RGB value R	G	B	Relative luminance %
260	162	111	119	20.38
2	208	44	30	15.30
4	235	71	38	22.31
30	265	147	212	26.74
37	41	97	179	12.28
3	34	44	126	3.65
6	96	51	114	6.07
20	234	51	120	21.22
23	213	46	75	16.61
61	181	43	130	13.16

Table 2: Light color pens were categorized by their relative luminance values being under 50.00%. The Staedtler color number refers to the pen brand's distinct label for each pen.

Staedtler Color #	**RGB value** R	G	B	**Relative luminance %**
110	255	208	0	66.37
222	252	179	136	54.71
43	255	177	69	53.13
401	250	173	117	51.50
1	247	205	69	63.87
34	144	222	235	64.17
301	158	192	219	50.08
21	243	174	172	52.31
221	255	168	176	52.50
62	238	187	225	59.15

While each participant drew their tree, the researcher in the room took note of each pen color that was used (a color being used was defined as the ink making contact with the graph paper) by writing down the Staedtler color number for each pen (this was verified again at a later time). Upon completion of the

drawing, the test administrator marked the back of the participant's paper with their 3-character index and when all drawings were collected, they were uploaded to Adobe Illustrator for analysis.

Data Analysis & Results

To validate the use of Koch's Baum Test as it relates to depression, the tree drawings were initially measured without regard to the colors used. Five indexes were measured: trunk height and width, canopy height and width, and total tree height. The measurements were taken from the longest or widest points of the aspect of the tree being measured (Fig. 3). All data (Table 3) was collected on Adobe Illustrator in centimeters.

Table 3: Tree measurement analysis results for all subjects. All values in centimeters.

Participant index	Trunk height	Trunk width	Canopy height	Canopy width	Tree height
H1a	5.5319	4.9468	8.0624	11.3016	12.4003
H2a	9.7747	5.0849	5.769	11.8438	14.8497
H3a	5.2525	4.47	5.5686	4.9448	9.7918
H4a	8.8349	2.9681	5.0705	6.4254	13.6228
H5a	5.254	3.6007	4.9886	7.107	9.5029
H6a	6.6183	3.5896	6.4606	7.7499	12.1085
H7a	9.8429	2.1373	9.5095	5.1751	10.1753
H8a	14.6375	6.2279	13.0971	18.2603	24.2972
H9a	2.7846	1.8551	17.7935	11.5513	18.5165
H10a	8.8237	5.4121	8.5671	12.1007	16.0812
H1b	15.1258	4.2786	16.4323	12.5256	19.4991
H2b	4.2573	2.1737	3.8306	3.1964	8.0778
H3b	2.9143	2.6922	7.4694	5.6269	9.9746
H4b	10.5984	1.9625	7.3516	8.2066	11.3647
H5b	13.0294	11.9611	8.153	11.9104	17.9078
H6b	6.1708	2.9744	5.2766	9.0628	8.7647
H7b	2.4809	1.7746	2.8235	5.4387	4.3919
H8b	9.8763	7.9417	14.5186	17.6604	21.0446
H9b	4.5065	3.0965	7.7718	8.5111	11.3234
H10b	9.0279	6.1835	9.2601	14.2356	15.8336
D1a	1.6542	1.8392	1.4547	2.2098	2.8465
D2a	14.8669	3.6776	4.6491	6.0646	16.0812
D3a	10.5434	7.304	11.9168	15.2273	16.8361
D4a	2.4906	1.5982	3.1889	3.477	5.4179
D5a	9.0315	3.7932	8.0677	8.5364	16.0357
D6a	6.1554	5.0062	9.5198	18.3032	14.4865
D7a	9.3833	6.1897	11.7915	14.7406	17.6646
D8a	10.0548	10.3217	12.0198	13.9566	18.6927
D1b	7.4725	7.073	3.7072	10.6166	10.5804
D2b	14.165	5.6841	11.4611	13.1899	19.7755
D3b	2.3444	3.7129	2.6106	4.357	4.917
D4b	10.3864	7.3498	10.7181	14.7009	16.0562
D5b	6.4626	6.1074	7.3478	7.9339	12.5118
D6b	6.9981	7.5641	8.7596	9.6606	15.1393
D7b	14.8975	8.9613	7.6603	13.6459	22.1358
D8b	11.0116	6.2373	16.5963	17.4083	22.485
D9b	3.9955	4.5917	9.1947	9.5033	12.7273

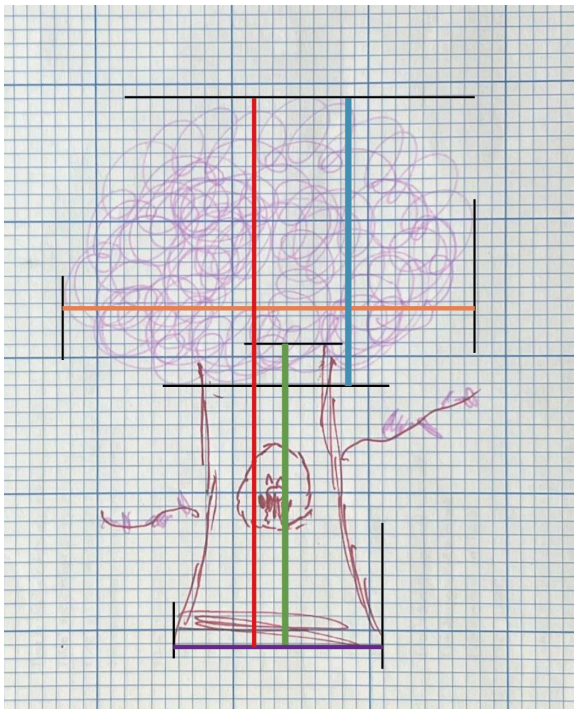


Figure 3 Example of tree measurement analysis - Participant H5a tree drawing. Green line is the trunk height, purple is the trunk width, red is the total tree height, blue is the canopy height, and orange is the canopy width.

Each tree measurement of a participant's drawing was tested for its correlation to the participant's BDI score using the coefficient of determination. These values were averaged among each trial group such that there were five values, one for each tree characteristic, for each of the four groups. Statistical significance was not seen in any of the five tree aspects measured across both trial groups. Possible causes included the subjectivity that comes with measuring without an automated system as used in previous studies that did show significance in similar measurements.

Table 4: Group averages of the R squared values between each participant's tree measurement and BDI score.

	Trunk height	Trunk width	Canopy height	Canopy width	Tree height
Group H					
Trial a	0.1344	0.4739	0.0437	0.4744	0.2931
Trial b	0.0179	0.0827	0.0144	0.0697	0.0184
Group D					
Trial a	0.0017	0.3092	0.2662	0.4105	0.0298
Trial b	0.1821	0.0945	0.0759	0.0299	0.1257

Color data was analyzed in two ways: the average relative luminance of a drawing and the total number of colors used in a drawing. The latter was recorded by the researcher while participants drew their trees and was again verified when drawings were uploaded to Adobe Illustrator for analysis. The researcher also took note of the Staedtler color number of each pen used by a participant while they were drawing. The pens had precalculated relative luminance (Y) values corresponding to these numbers, which were used for analysis of the overall Y value of a drawing. To get the overall Y value, the Y value for each pen used in a drawing was summed and divided by the total number of colors used in the drawing.

Descriptive analysis of the overall relative luminance values of drawings showed indications of a correlation existing between Y values and BDI scores as hypothesized. Both trial groups of the healthy group had higher overall Y values for their drawings compared to their respective mild to moderate depression trial groups and had no overlap in ±1 standard error of the mean of each other (Fig. 4). Inferentially however, no significance was found between the two.

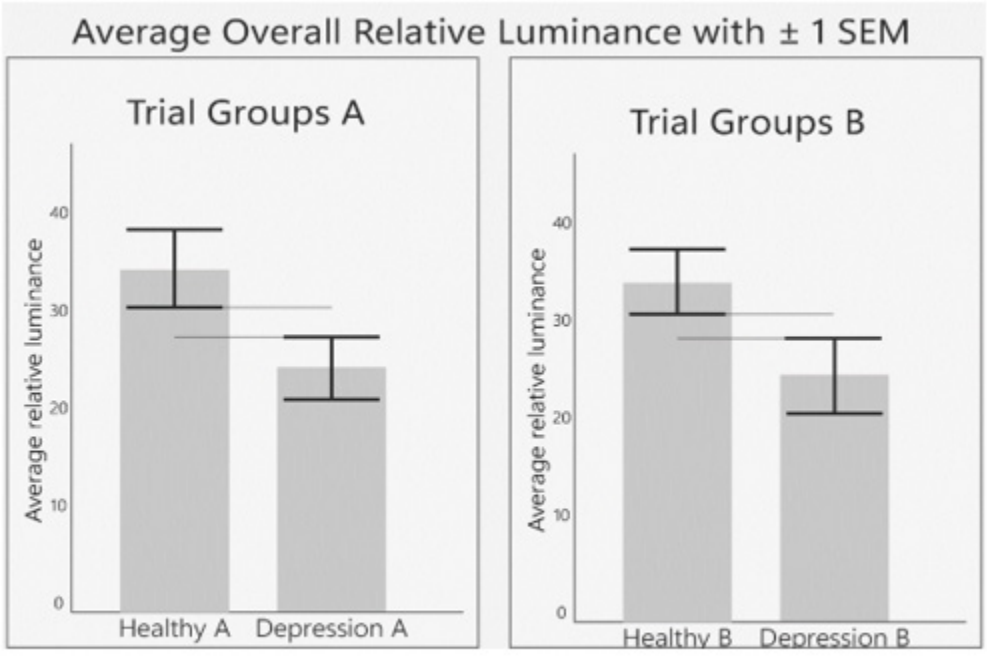


Figure 4: Error bars indicating ±1 SEM for each group's overall average relative luminance values had no overlap.

Conversely, the total number of colors used by each participant did show significance upon analysis, specifically in the mild to moderate depression group. Among the 17 participants in this group, their BDI scores and how many colors they used to draw a tree produced an R-squared value of 0.52 and a p-value less than .001 with Bonferroni correction. This strong positive correlation indicates that the higher the BDI score of a person in the mild to moderate depression group, the more colors they used to draw their tree (Fig. 5).

Table 5: Complete color data including BDI scores and the calculated average overall relative luminance for each drawing.

Index	Color 1 Y%	Color 2 Y%	Color 3 Y%	Color 4 Y%	Color 5 Y%	Color 6 Y%	Color 7 Y%	Color 8 Y%	Color 9 Y%	Color 10 Y%	Color 11 Y%	Color 12 Y%	Color 13 Y%	Color 14 Y%	Color 15 Y%	Average Y%	BDI score	# of colors
H1a	15.3															15.3	6	1
H2a	52.5	53.13														52.815	8	2
H3a	15.3															15.3	5	1
H4a	59.15															59.15	3	1
H5a	59.15	16.61														37.88	6	2
H6a	26.74	12.28	59.15	13.16	16.61	63.87										31.96833	6	6
H7a	53.13	22.31	21.22	52.5	63.87	52.31	26.74	3.65	6.07							33.53333	4	9
H8a	13.16	22.31														17.735	10	2
H9a	52.5															52.5	6	1
H10a	20.38	22.31	13.16	6.07	12.28	16.61	53.13	63.87								25.97625	14	8
H1b	22.31	20.38														21.345	4	2
H2b	22.31	20.38														21.345	1	2
H3b	3.65	64.17	66.37													44.73	14	3
H4b	3.65	66.37														35.01	6	2
H5b	26.74	66.37														46.555	10	2
H6b	21.22	51.5	53.13	22.31	15.3	3.65										27.85167	11	6
H7b	3.65	20.38	21.22													15.08333	8	3
H8b	51.5	16.61	15.3	22.31	66.37	63.87	53.13									41.29857	14	7
H9b	63.87	52.5														58.185	7	2
H10b	15.3	53.13	22.31	52.5	52.31	6.07	3.65									29.32429	4	7
D1a	3.65	21.22														12.435	15	2
D2a	64.17	26.74	3.65													31.52	15	3
D3a	21.22															21.22	15	1
D4a	51.5	21.22	50.08	26.74	20.38	63.87										38.965	18	6
D5a	52.5	15.3	3.65													23.81667	17	3
D6a	53.13	22.31	6.07	51.5	54.71	20.38	16.61	15.3								30.00125	26	8
D7a	3.65	16.61	6.07													8.776667	20	3
D8a	53.13	22.31	13.16	21.22	50.08	15.3	16.61	20.38	26.74	59.15	3.65	6.07	63.87			28.59	26	13
D1b	59.15	3.65	6.07	12.28												20.2875	25	4
D2b	21.22	51.5	63.87													45.53	17	3
D3b	6.07	15.3	3.65													8.34	24	3
D4b	20.38	15.3	22.31	3.65	6.07											13.542	26	5
D5b	26.74	51.5	66.37	3.65												37.065	18	4
D6b	12.28	63.87	50.08	3.65	6.07											27.19	23	5
D7b	53.13	16.61	64.17	3.65	6.07											28.726	26	5
D8b	51.5	6.07	3.65	63.87	52.5	53.13	52.31	54.71	12.28	13.16	15.3	16.61	20.38	21.22	22.31	30.6	30	15
D9b	15.3	3.65	6.07	12.28												9.325	18	4

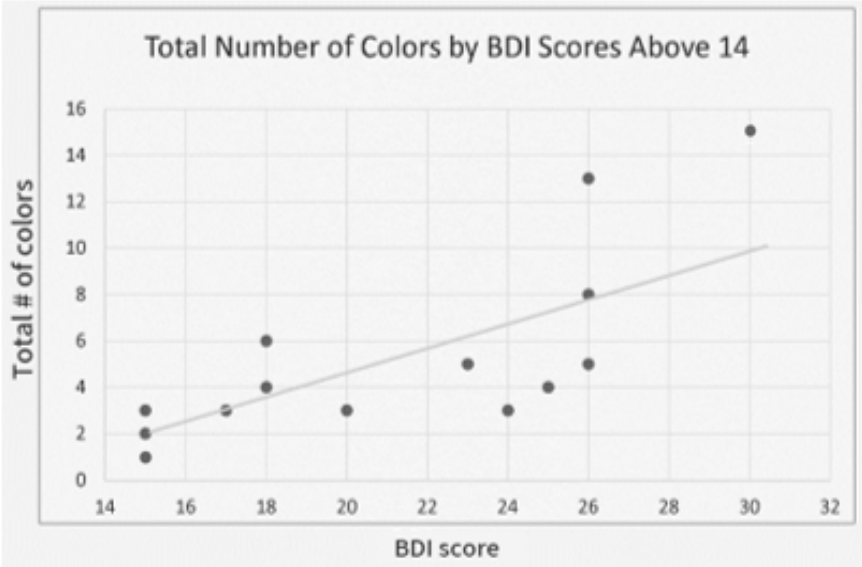


Figure 5: Total number of colors used by mild to moderate depression participants with number of colors on the y-axis and BDI scores on x-axis.

Discussion

This study looked at the unconscious processing of colors in the primary visual cortex and how to utilize it to test to include more ambiguous stimuli in current depression diagnostic screenings. This study found a strong positive correlation between the total number of colors used by people with mild to moderate depression and their reported depression levels from the BDI. However, the hypothesis

of people in this category using darker colors, or those with lower relative luminance values, was only partially supported. In addition, measurements taken solely to validate the correlation between major depressive disorder and Koch's Baum Test were not significant. More data collection across a wider range of ages as well as geographical backgrounds along with automated analysis of the drawings to limit subjectivity and create a more generalizable sample can be pursued to understand these results better.

The correlation found between the total number of colors and BDI scores of people with mild to moderate depression levels can be attributed to the constant high of emotions, lack of calm in depressed people's minds, or differences in their attention spans. Part of using more colors may also have been a way for people of this group to mask their inner feelings of bleaknessand present as more okay, similar to the function of the social desirability effect. Specific biological reasons for this result are unknown but can be studied in the future to understand why more colors may be appealing to people with higher levels of depression.

As the validity of Koch's Baum Test in relation to depression showed no significance, the question of other objects possibly being more accurate projections for people with this illness is brought up. Koch's Baum Test uses a tree because of its neutrality and universality, however, other projective tests use different objects, including houses and people and these too can be studied with color psychology. The lack of significance in this aspect of the study shows why this colored version of Koch's Baum Test may not be enough to be a diagnostic tool on its own, rather it should be used in addition to the currently used standardized depression screenings.

With the goal of reducing the misdiagnoses of the illness, this study combined the concepts of color psychology and projective tests to find correlations between art and depression that are not as easily manipulatable as question-answer screenings can be. Strong correlations were found, and the colored version of Koch's Baum Test that was developed in this study can be applied to analyzing people with depression with less subjectivity and more accuracy.

References

Aragonès E, Piñol JL, Labad A. (2006). The overdiagnosis of depression in non-depressed patients in primary care. Family Practice. 2006 Jun;23(3):363-368. DOI: 10.1093/fampra/cmi120. PMID: 16461446.

Breitmeyer, Bruno & Ro, Tony & Singhal, Neel. (2004). Unconscious Color Priming Occurs at Stimulus- Not Percept-Dependent Levels of Processing. Psychological science. 15. 198-202. 10.1111/j.0956-7976.2004.01503009.x.

Fugate, Jennifer & Franco, Courtny. (2019). What Color Is Your Anger? Assessing Color-Emotion Pairings in English Speakers. Frontiers in Psychology. 10. 10.3389/fpsyg.2019.00206.

Gu, S., Liu, Y., Liang, F., Feng, R., Li, Y., Liu, G., Gao, M., Liu, W., Wang, F., & Huang, J. H. (2020). Screening Depressive Disorders With Tree-Drawing Test. Frontiers in psychology, 11, 1446. https://doi.org/10.3389/fpsyg.2020.01446

Hunt, M., Auriemma, J., & Cashaw, A. C. (2003). Self-report bias and underreporting of depression on the BDI-II. Journal of personality assessment, 80(1), 26–30. https://doi.org/10.1207/S15327752JPA8001_10

Mojtabai, R. J. (2013, April 30). Over-Diagnosis and Over-Treatment of Depression Is Common in the U.S. Retrieved October 5, 2022, from https://publichealth.jhu.edu/2013/mojtabai-depression-over-diagnosis-and-over-treatment

Moller, Arlen & Elliot, Andrew & Maier, Markus. (2009). Basic Hue-Meaning Associations. Emotion (Washington, D.C.). 9. 898-902. 10.1037/a0017811.

Ran, M., Tan, X., Jiang, G., Chen, X., Yang, H., & Yuan, G. (2017, June 25). Probe into Color Preference in Inpatients with Depressive Disorders. Retrieved October 5, 2022, from https://wap.hapres.com/htmls/JPBS_902_Detail.html

Sevda, Korkmaz & Özer, Ö & Kaya, Ş & Kazgan, Aslı & Atmaca, Murad. (2016). The correlation between color choices and impulsivity, anxiety and depression. European Journal of General Medicine. 13. 47-50. 10.15197/ejgm.1592.

Wan, C., Yang, X., & Liu, W. (2014). Reliability and validity test of medical students' application in tree drawing personality test [J]. Chin. J. Behav. Med. Brain Sci, 23, 946-948. https://www.semanticscholar.org/paper/Reliability-and-validity-of-the-tree-drawing-test-Wan-Ran/a5880b99d342c6f7ad8e199f53cc088ba8af3999

Autism and Moral Criteria

Rowena (Ellis) Cai Yin

Nikola Tesla STEM High School

Abstract

Previous studies involving behavior regarding autism spectrum disorder have shown that autistic individuals* are inherently "more moral" than their non-autistic counterparts, however, does autism itself directly result in a positive correlation with ethicality or is personal integrity involved? This study measured the correlation between amount of traits associated with autism and how flexible or strict an individual's moral standard was. The hypothesis was autistic individuals (and persons who have more traits associated with the condition) will have stricter criteria regarding "good" morals. Data was collected through an online self-report survey which measured participants' (23, ages 16+) Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R) test scores and a statement-answer survey which provided a quantitative measure of the strictness which someone adhered to their own moral principles. Results showed a slightly positive correlation between these two variables as the produced R-value or correlation coefficient was 0.1207. However, the hypothesis was still rejected due to a small and ungeneralizable sample. Findings suggest that autism itself and possessing more traits associated with autism (differences in what is considered "normal" social relatedness, sensory-motor, language usage, and circumscribed interests according to the RAADS-R) has little impact on how strict a person is towards flexibility regarding their moral criteria.

*Autism is referred to using identity-first language in order to respect the preferences of *most* autistic individuals (Kapp, Gillespie-Lynch, Sherman & Hutman, 2013).

Introduction

Criteria for "good" morality fluctuates from person to person, however, recent studies regarding this concept in relation to autism often follow a common trend where autistic individuals are observed to be more virtuous than their neurotypical and non-autistic counterparts. However, the concept of "good" morals is not a monolith, primarily having differences based on the culture which someone's morals are derived or influenced from. Perhaps autism and its associated traits positively correlate with ethicality, not through the virtuality of their morals, but rather on how uncompromising someone's adherence to their own principles for "good".

Autism is a neurodevelopmental condition which is often at surface-level presented as repetitive behaviors, aversion to specific stimuli, and difficulty with social interaction. Despite efforts, this condition is heavily underdiagnosed in marginalized groups due to a lack of proper research on such individuals— an issue which provides difficulty when attempting to find generalizable data within this field of research.

In 2014, a study showed that autistic children tend to have more rigid criteria than non-autistic children when it comes to harm, a factor of morality. When presented with stories with morally bad protagonists, approximately 75% of autistic children decided on the harshest option in terms of judging them morally (choice "really naughty")

compared to approximately only 50% of non-autistic children who chose to same option (Li, Zhu & Gummerum, 2014). Unlike autistic children, non-autistic children were more likely to cooperate with those sharing an opposing moral opinion. Overall, the autistic children who participated in this study showed uncompromising adherence to their ethical principles.

But does this result spread across different age groups? While the investigation studied children, the current project aimed to observe adherence to personal values between autistic and non-autistic adolescents and adults.

Interestingly, many previous studies regarding autism rarely factor in the amount of non-diagnosed individuals or individuals who were rejected from a diagnosis for being too "high functioning". This is mainly due to a medical model that both intensely focuses on impairments and has mostly been based off research on a singular type of individuals with this condition— resulting in studies who's sample, and therefore result, does not represent all autistic individuals.

Despite definite variation with behavior between each individual, autistic persons will most likely hold both themselves and others more strictly towards adhering to their own moral criteria because people with this condition are more prone to penalties provided by neurotypical peers from accidental social transgressions.

Methodology

In the present experiment, participants were recruited through advertisements posted in various autism-related forums on social media platforms such as Reddit, Discord, and Twitter. The use of various forums attempts to ensure a diverse and therefore generalizable population. Individuals clinically diagnosed with autism spectrum disorder were especially sought out, however, the autism pre-screening Ritvo Autism Asperger Diagnostic Scale- Revised (abbreviated to RAADS-R) was employed in the case of autistic individuals without the privilege of a diagnosis. Advertisement of the study provided critical information such as allowed age range (16+) and survey contents. Participation was voluntary. All willing participants completed a consent form prior to being given access to the survey and were allowed to opt out of the study at any moment. No compensation was given. See *Figure 1* for more details.

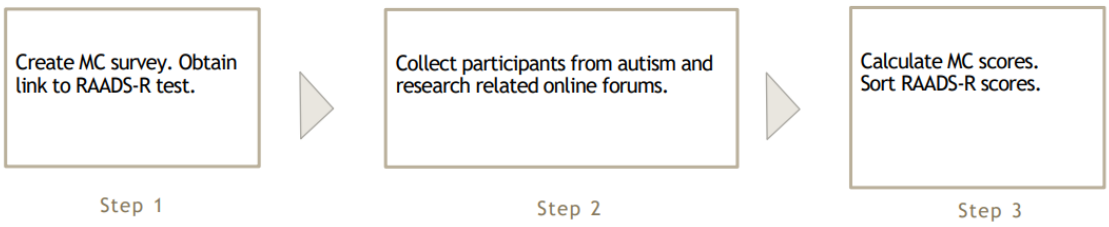


Figure 1. Process of collecting data.

The given self-report contains two major parts. The first half presented the RAADS-R test provided through an attached link while the latter was a statement-answer survey (MC survey) that calculated a score regarding how strict each participant holds themselves to their own moral criteria. Demographic information were also collected. In the demographics section, participants were asked to specify their identified gender, age, ethnicity, and whether or not they have been clinically diagnosed with autism spectrum disorder (ASD). For the former three, participants could choose not to say.

RAADS-R is a self-report instrument which measures traits of autism, commonly used as a prescreening; it is not equivalent to a clinical diagnosis. Scores per question range from 0 (low indication of a particular trait) to 3 (high indication of a particular trait) while total scores range from 0 to 240, with few non-autistic persons having scored over the threshold of 65

(Engelbrecht, 2022). Although higher scores indicate more traits associated with autism, lower scores in no way means someone is "less autistic" unless a score is located below the threshold. All possible scores are then categorized into four range options (0-64; 65-129; 129-194; 195-240) assigned the single-digit numbers 1, 2, 3, and 4 respectively (see *Figure 2*).

RAADS-R score	0 - 64	65 - 129	130 - 194	195 - 240
Numerical value	1	2	3	4

Figure 2. Assigning single digit numerical values to sectioned RAADS-R scores.

The latter half of the survey, or MC survey (see *Figure* 3) contains ten scenario statements which participants are required select a response (never; infrequent; sometimes; often; always) between never (scenario presented rarely occurs in their own life) and always (scenario presented occurs frequently in their own life). Answers are associated with a quantitative number from 1 to 5 (see *Figure 4*), when totaled, produce an MC score (see *Figure 5*). Higher scores indicate stricter adherence towards a participant's own moral criteria while lower scores indicate more flexible adherence towards their own moral criteria. Participant's MC scores are then compared to the range which their RAADS-R score falls into (see **Table 1**). In order to ensure the MC survey was standardized, measures of central tendency (mean, median, and mode) of all participant answers was calculated. A quantitative correlation between the variables was calculated using Pearson's R-value formula (see *Figure 6*).

Statements a. through e.

Ans. Type	never	infrequent	sometimes	often	always
Num. Value	5	4	3	2	1

Statements f. through j.

Ans. Type	never	infrequent	sometimes	often	always
Num. Value	1	2	3	4	5

Figure 4. MC Scoring Guide. Figure above shows possible answers to all scenario statements and their corresponding numerical value.

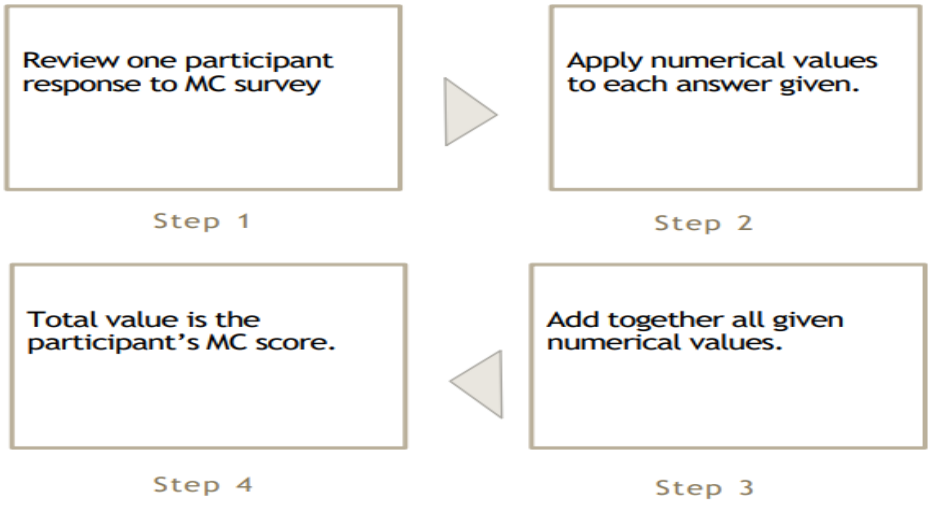


Figure 5. Procedure for calculating participants' individual MC scores.



7. Please consider the following statements **a.** through **j.** *

Choose the option which describes how often each statement applies to yourself.

	never	infrequent	sometimes	often	always
a. I stay quiet when presented with a morally "bad" opinion.	○	○	○	○	○
b. I am willing to cooperate with someone who I deem immoral if the situation calls for it.	○	○	○	○	○
c. I make "good" moral decisions solely because of a possible reward.	○	○	○	○	○
d. I can easily tolerate problems that have morally ambiguous solutions.	○	○	○	○	○
e. My perspective of someone does not automatically become negative if I find out they have once committed an immoral act.	○	○	○	○	○
f. I still feel personally responsible for the consequences of my own immoral actions, even if I was pressured into committing those actions.	○	○	○	○	○
g. I have the urge to correct someone if their opinion or action is one I deem morally wrong.	○	○	○	○	○
h. I dislike having to find a middle ground with a person whose actions I believe are immoral.	○	○	○	○	○
i. I believe that individuals hold full responsibility for their "bad" actions, even if they were enabled by their environment.	○	○	○	○	○
j. I become extremely uncomfortable when a person I look up to commits an immoral act.	○	○	○	○	○

Figure 3. Question 7. Figure above shows all scenario statements and their corresponding answers in parts 1 (statements a. through e.) and 2 (statements f. through j.).

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Figure 6. Pearson R-value formula.

Table 1. How strict a participant holds themselves to their own moral criteria, compared to RAADS-R score and possible clinical ASD diagnosis.

Participant number	RAADS-R score range number	Clinical ASD diagnosis?	MC score (10- 50)
1	1	No	24
2	1	No	33
3	1	No	34
4	1	No	37
5	1	No	38
6	1	No	43
7	1	No	47
8	2	No	34
9	2	No	34
10	2	Yes	35
11	2	No	39
12	3	No*	27
13	3	No*	31
14	3	No	33
15	3	No	34
16	3	No	36
17	3	No	39
18	3	No	39
19	3	Yes	39
20	3	No	41
21	3	No	43
22	3	Yes	47
23	4	No	42

*Individual has been clinically diagnosed with ADHD but not autism spectrum disorder.

Results

From the data collected, a slight positive correlation (R-value: 0.1207) was found between autism and stronger adherence to one's moral values, meaning that autism may result in stricter adherence to a person's own ethical standard. However, the hypothesis was rejected due to the low correlation coefficient and a non-generalizable participant group. *Figure 7* shows a trend between likelihood for autism and adherence to one's own moral values. Overall, the MC survey created solely for this study was standardized due to the small range between the calculated main measures of central tendency (mode= 34, median= 37, mean≈ 37.044).

Figure 8 shows the race of all participants, with the majority of Asian participants (highlighted in *Figure 8* with orange) having RAADS-R scores that lean or reside in the non-autistic range and all White participants' (highlighted in *Figure 8* with white) RAADS-R scores residing above the threshold. This is the cause of the lack of generalizability within the sample and is most likely due to how knowledge on autism stems primarily from research on White individuals.

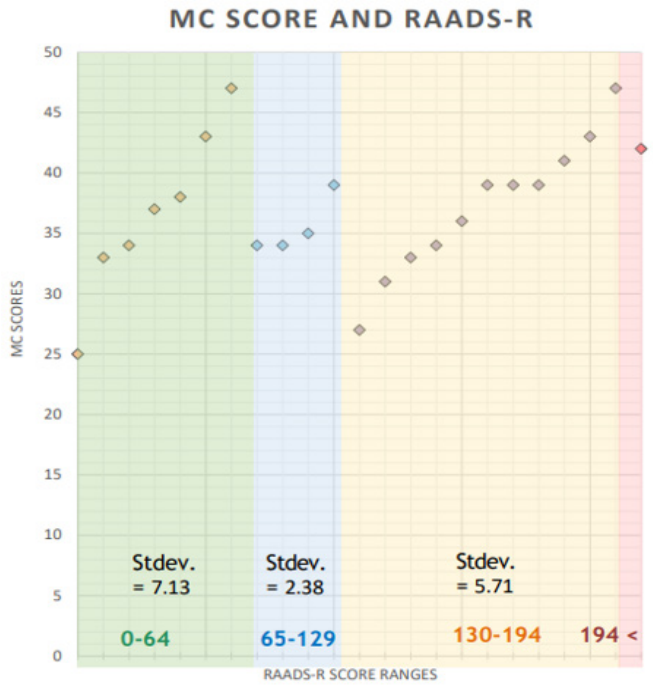


Figure 7. Graph of all participant data in terms of likelihood to have autistic (x-axis) and strictness towards adhering to one's own moral standard (y-axis).

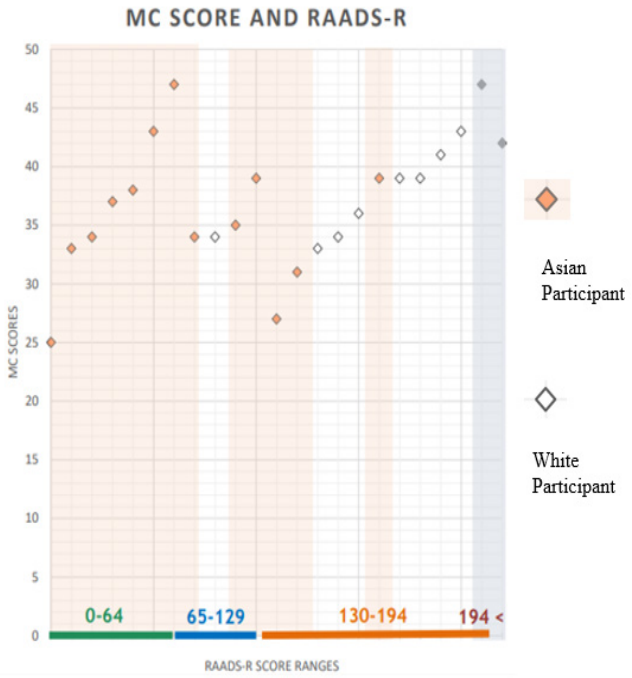


Figure 8. Graph of all participant data in regard to likelihood of autism and adherence to one's moral values— categorized by race.

Interestingly, the participants who had notified that examiner that they were clinically diagnosed with ADHD but not autism had some of the lowest MC score (showing extreme flexibility towards adherence to their morals), only second and third to the lowest, which was scored by a non-autistic individual (see *Figure 9*). However, this does not mean ADHD individuals possibly lack integrity as there are too little data points regarding persons with this condition.

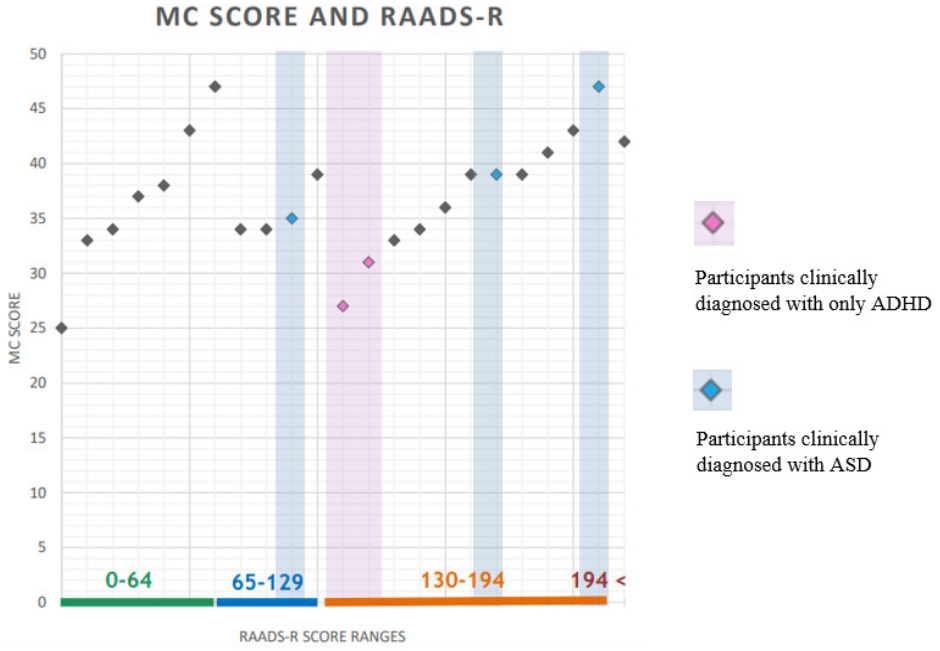


Figure 9. Graph of all participant data in regard to likelihood of autism and adherence to one's moral values— participants with clinically diagnosed ASD and ADHD highlighted.

Discussion

It should be noted that the answers to MC survey statements for autistic participants were more varied and often shifted between extreme ends, while non-autistic participants' answers were more likely to stay neutral or relatively similar throughout. This could be due to polarized thinking patterns, or better known as black and white thinking, which is commonly a trait associated with autistic persons. A lack of heavy correlation in the result could also be due to the aforementioned phenomenon, resulting in where the contrasting answers of autistic participants mirrored that of neurotypical participants when totaled.

The little correlation could also be due to a small sample of participants, attributed to a limitation of collecting people who are willing to take a half-hour long survey with no given compensation. Another major limitation that may have affected results was the understanding of questions administered in both the RAADS-R and MC Survey. As autistic individuals may have literal thinking, or reasoning which relies on actual meanings rather than what us implied, questions on both self-reports, especially the RAADS-R (as the website which hosts this self-report has warned), could have been misinterpreted. Non-autistic individuals could have also very likely faced the same issue.

This study's main finding can be applied as a base for future investigations and understanding of social behavior regarding autistic persons and how it differs from that of non-autistic individuals. The slight positive correlation found between autism and uncompromising adherence to a person's own moral principles should be further expanded upon through research that obtains qualitative data and a more generalizable sample.

Sources

Dempsey, E. E., Moore, C., Richard, A. E., & Smith, I. M. (2020). Moral foundations theory in autism spectrum disorder: A qualitative investigation. *Autism : the International Journal of Research and Practice*, *24*(8), 2202–2212. https://doi.org/10.1177/1362361320939331

Dempsey, E. E., Moore, C., Johnson, S. A., Stewart, S. H., & Smith, I. M. (2021). Moral Foundations Theory Among Autistic and Neurotypical Children. *Frontiers in Psychology*, *12*, 782610–782610. https://doi.org/10.3389/fpsyg.2021.782610

Engelbrecht, N. (2022, August 18). *RAADS–R | embrace-autism.com. RAADS-R. Retrieved April 5, 2023, from https://embrace-autism.com/raads-r/?s=09*

Gillespie-Lynch, K., Kapp, S. K., Brooks, P. J., Pickens, J., & Schwartzman, B. (2017). Whose Expertise Is It? Evidence for Autistic Adults as Critical Autism Experts. *Frontiers in psychology*, *8*, 438. https://doi.org/10.3389/fpsyg.2017.00438

Hu, Y., Pereira, A. M., Gao, X., Campos, B. M., Derrington, E., Corgnet, B., Zhou, X., Cendes, F., & Dreher, J. C. (2021). Right Temporoparietal Junction Underlies Avoidance of Moral Transgression in Autism Spectrum Disorder. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *41*(8), 1699–1715. https://doi.org/10.1523/JNEUROSCI.1237-20.2020

Kapp SK, Gillespie-Lynch K, Sherman LE, Hutman T. Deficit, difference, or both? Autism and neurodiversity. Dev Psychol. 2013 Jan;49(1):59-71. doi: 10.1037/a0028353. Epub 2012 Apr 30. PMID: 22545843.

Li, J., Zhu, L. & Gummerum, M. (2014). The relationship between moral judgment and cooperation in children with high-functioning autism. *Sci Rep* **4**, 4314. https://doi.org/10.1038/srep04314

Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental cognitive neuroscience*, *23*, 45–56. https://doi.org/10.1016/j.dcn.2016.12.006

Ritvo, R. A., Ritvo, E. R., Guthrie, D., Ritvo, M. J., Hufnagel, D. H., McMahon, W., Tonge, B., Mataix-Cols, D., Jassi, A., Attwood, T., & Eloff, J. (2011). The Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R): a scale to assist the diagnosis of Autism Spectrum Disorder in adults: an international validation study. *Journal of autism and developmental disorders*, *41*(8), 1076–1089. https://doi.org/10.1007/s10803-010-1133-5

Adolescents vs. Adults: Upward Social Comparison in a Standardized Test Setting

Varsha Mantravadi

Nikola Tesla STEM High School

Abstract:

The predictive validity of standardized tests has undergone increasing scrutiny in recent years, and science is unable to agree on the topic. This investigation utilizes upward social comparison (comparing to one's superiors), which may introduce subjectivity during test-taking, in order to analyze predictive validity by assessing adolescents' and adults' actual and perceived SAT performances, along with their self-efficacy (the belief in their ability to succeed). A representative sample of 50 adolescents and 50 adults was randomly selected; each participant completed two shortened SATs and received a fake score, and completed a survey on perceived competency and self-efficacy. A negative correlation between upward social comparison and competency, perceived competency, and self-efficacy level in adolescents was hypothesized. No correlation between any of these variables in adults was hypothesized. The results partly support these hypotheses: actual and perceived competencies correlated negatively in adolescents ($R2 = -0.585, -0.557$), suggesting that they responded poorly to a low score, and neither actual competency nor self-efficacy correlated in adults. However, self-efficacy did not correlate in adolescents, and perceived competency correlated positively ($R2 = 0.747$) in adults, suggesting that adults felt better about their performance after the low score. Yet, only the correlation between age groups for perceived competency was significant ($z = -3.357$), indicating that adults sought improvement while adolescents became discouraged due to the importance of the SAT to them. However, the insignificance of the rest of the results suggest that social comparison does not introduce any level of subjectivity, instead boosting the SAT's predictive validity.

Introduction:

This research analyzes the difference between social comparison tendency in adolescents and in adults by administering fake standardized test results and analyzing the subsequent difference in scores, as well as responses to survey questions inquiring into test anxiety, competence, and self-efficacy for both age groups.

Social comparison is defined as the tendency to compare oneself to others, especially on topics such as body image, abilities, and skills, in order to evaluate personal worth. Upward social comparison involves comparing oneself to someone perceived as superior, while downward social comparison involves comparing oneself to someone perceived as inferior (Wheeler, 1966; Wills, 1981). As much as 10% of people's thoughts involve comparing themselves – their achievements, abilities, and appearance – with others, in one way or another (Buunk, et al., 2007). The relationship between social comparison and its impact on our daily lives is an ongoing cycle – the influence on our everyday lives increases our tendency to engage in comparison, which consequently alters how we live our everyday lives. The implications of increased comparison include a harsher attitude concerning one's abilities, and a rosier image of others' abilities brought on by either a feeling of inadequacy or of motivation (Festinger, 1954).

This can create healthy competition for some but intense anxiety for others. Often, social comparison is one of the only ways we can confirm our thoughts, opinions, and actions, and this causes us to not only compare, but surround ourselves with people we would like to emulate. This is especially seen in adolescents: around 75% of teenagers

aged 13-18 identify their peer group as one that challenges them, intellectually as well as socially (Anderson et al., 2018). Comparing themselves to peers, especially those with qualities they would like, allows them to strive for more out of their lives, helps them form identities, and validates unknown thoughts and emotions. Moreover, people who are unsure of their situation – social, educational, occupational, etc. – tend to engage in twice as much social comparison as their self-assured peers (Buunk, 1995). Assuredness of one's situation is much more commonly seen in older individuals than in younger individuals, as many young adults are plagued by worries about their future while older adults focus more on the present and their current situations.

The tendency to socially compare thus decreases across adulthood: older adults almost never deeply consider themselves in relation to others, while younger adults still rely on such habits to validate their successes and failures (Callan et al., 2015). Younger adults also place extreme value on their sense of identity: something that sets them apart from their peers, and something that comparison may serve to form. Younger adults are much more conscious of their peers' behaviors than older adults, and tend to mold themselves in a way that helps them to stand out rather than to fit in. While as much as 71% of adults aged 50-64 would rather blend in, only 42% of adults under 35 would do the same. (Lashbrook, 2012). Following such an inclination to stand out can prove a strong incentive to socially compare, and the behavior of these young adults and their peers' influences on their daily lives showcase yet another effect of doing so.

However, research has shown that adolescents – teenagers between 14 and 19 years of age – experience much more drastic decreases in self-esteem and increases in additional stress than those at any stage of adulthood. Because of the biological and environmental factors governing such a behavioral change, it may be worthwhile to consider that those in this age group will harbor different incentives to socially compare and thus experience significantly altered effects. These factors include a more emotional outlook of oneself due to an increase in amygdala activity, and a still-developing prefrontal cortex which decreases rational thinking and decision-making (Morawetz et al., 2016). Environmental factors include the instability created by constant peer and parent pressure that adolescents often face regarding major decisions for their future, such as higher education. Though adolescents may be pushed to socially compare at an extremely high level by external factors, is this truly what their behavior represents? In stressful scenarios many adolescents face like standardized testing, will they bother comparing their performance to their peers' if their parents may already have done so?

Such questions have yet to be answered by the investigations on situational comparison and the roles of neurobiological factors. Specifically, these investigations failed to address the relationship between age (specifically adolescents versus adults) and intensity of social comparison in a standardized test setting, as the neurobiological differences in adolescents and the environment of a standardized test may contribute to a significant difference in each age group's response to comparisons with peers. After all, adolescents are much more acclimatized to standardized tests than adults, and may even be desensitized to any effects of social comparison. Such effects may instead plague an adult population to incessantly focus on their peers' performance and behavior during the standardized test, a sharp contrast from the studies on the differences between younger and older adults.

Commonly investigated along with academic success, self-efficacy is a significant measure of an individual's confidence through their belief in their abilities to achieve certain behaviors and skills. These skills will typically allow them to accomplish a specific task or succeed at a higher level, and a higher self-efficacy indicates a higher confidence that one will accomplish said success. Research investigating the relationship between self-efficacy and students' academic performance has suggested a strong correlation between the two (Domenech-Betoret et al., 2017; Yusuf, 2011). But such studies have not yet discussed the relationship between social comparison and self-efficacy, as the two in an academic setting could harbor a significant relationship. Not only can self-efficacy be used as a second variable, but it can also measure the effects of social comparison within two age groups, as a lower self-efficacy can indicate which of them felt a higher pressure to perform well in the standardized test environment.

Standardized tests themselves have

undergone much scrutiny in recent years, especially during and shortly after the COVID-19 pandemic. Many studies claim invalidity and unreliability, as they typically do not account for factors such as socioeconomic status, mental health, and access to preparatory materials (Rask, et al., 2009). However, other studies remain steadfast in their claims that such tests have high predictive validity and control for such factors effectively (Shaw, 2015). A scientific consensus has yet to be reached, so a novel approach such as one assessing social comparison between age groups may solidify one or more of these claims, in which one age group holds these tests in high importance while the other does not. Such factors may – or may not – introduce a level of subjectivity that points in a certain direction regarding the predictive validity of standardized tests and whether they should continue to be standard measures of a student's future success.

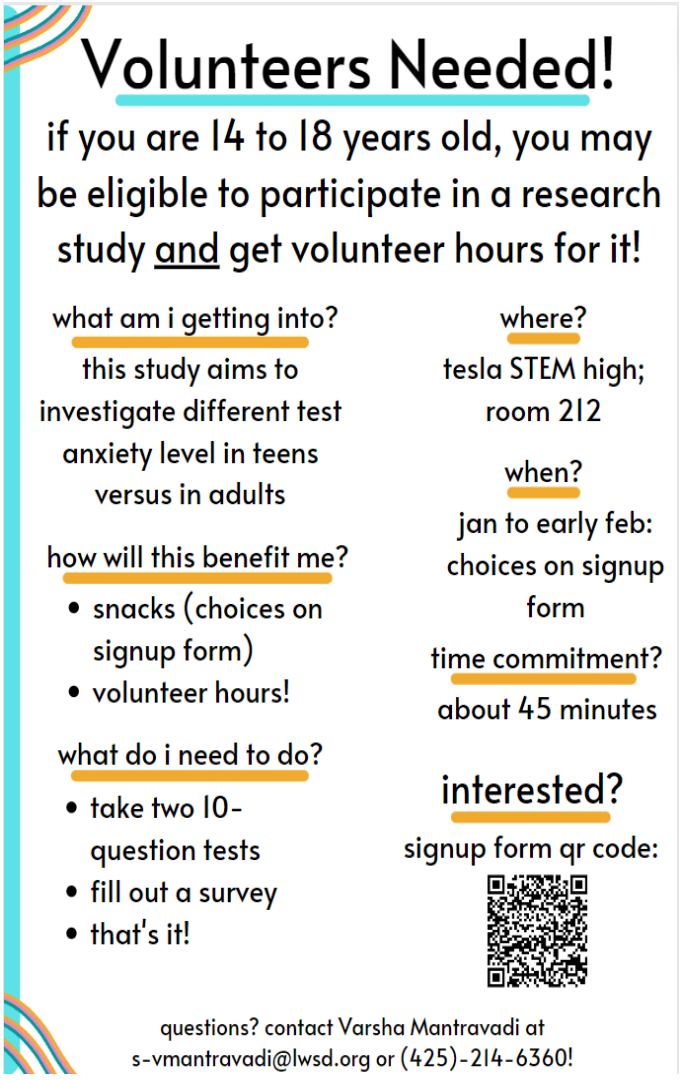
This study seeks to investigate the correlation between adolescence, competence (perceived and actual) at standardized tests, and the resulting self-efficacy in response to upward social comparison in a standardized test setting. The expected outcomes are for the younger participants to engage in a higher level of comparison with their peers, reflected in a lower competence level (both perceived and actual) and a lower self-efficacy than seen in adults, who are expected not to engage in such a high level of comparison with their peers.

Methodology:

This hypothesis was tested through a double-blind procedure in which a standardized test, modeled after the SAT, was administered to a randomly selected group of adolescents and a randomly selected group of adults. After one test, each individual in each of the groups received either a control result – an average score on the test – or an experimental result – a low score on the test. The test was then administered again and the difference in test scores was recorded, as well as responses to a survey inquiring into perceived competency, self-efficacy, and test anxiety.

To obtain a sample population of adolescents, a random number was assigned to every student at Nikola Tesla STEM High School, Redmond High school, Eastlake High School, and Lake Washington High School. A random number generator was used to choose 100 of these students anonymously. These students were then emailed about participation in the study – they were asked if they were in the 14-18 age range, if they would like to participate, and were informed of test and survey administration and how they would receive their "results," as well as an attached signup form and an informed consent form. Fifty copies of a flyer, modeled in Fig. 1, with the same information about the study and a QR code with a link to the signup form with informed consent attached were distributed around Tesla STEM High School, Redmond High School, Eastlake High School, Lake Washington High School, local cafés, and the Redmond Public Library. Once fifty responses were received, each participant was randomly assigned a separate number. Twenty-five of these numbers were randomly selected using a random number generator: those who were assigned these numbers were assigned to the experimental group, and the remaining participants were assigned to the control group.

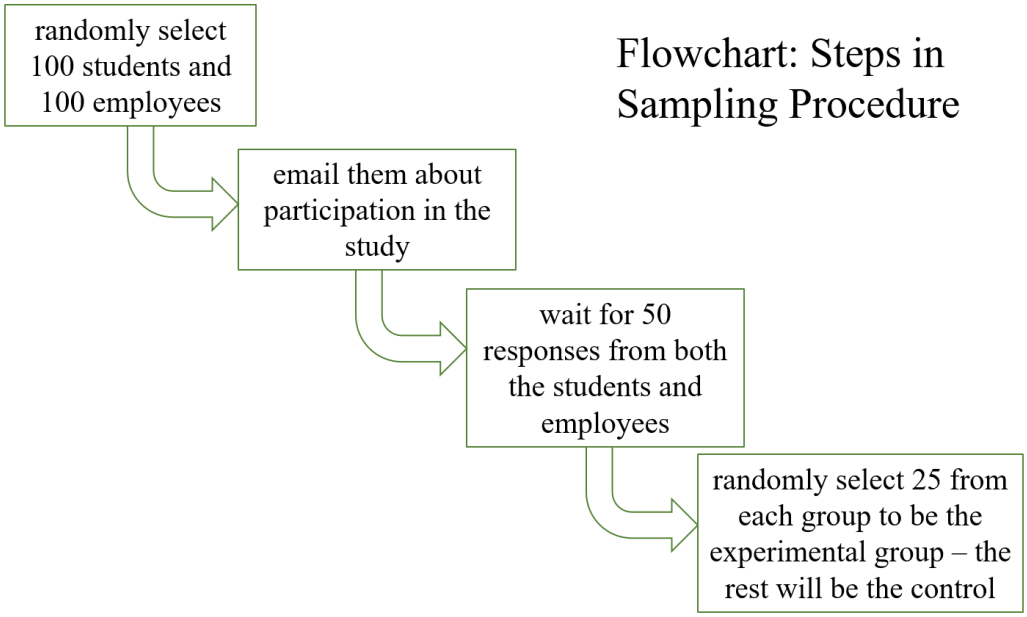
Figure 1. An example flyer for adolescent participation



Note. An example flyer for adolescent participation, presenting a fake purpose, a summary of the procedure, incentives, time and place, the signup form QR code, and contact information. This was posted in the halls of three high schools, local cafés, and the Redmond Public Library.

A sample population of adults was obtained through publicization of the study using flyers. Fifty copies of a flyer with the same information about the study and a QR code with a link to the signup form with informed consent attached were distributed around local cafés and the Redmond Public Library. Once fifty responses were received, each participant was randomly assigned a separate number. Twenty-five of these numbers were randomly selected using a random number generator: those who were assigned these numbers were assigned to the experimental group, and the remaining participants were assigned to the control group.

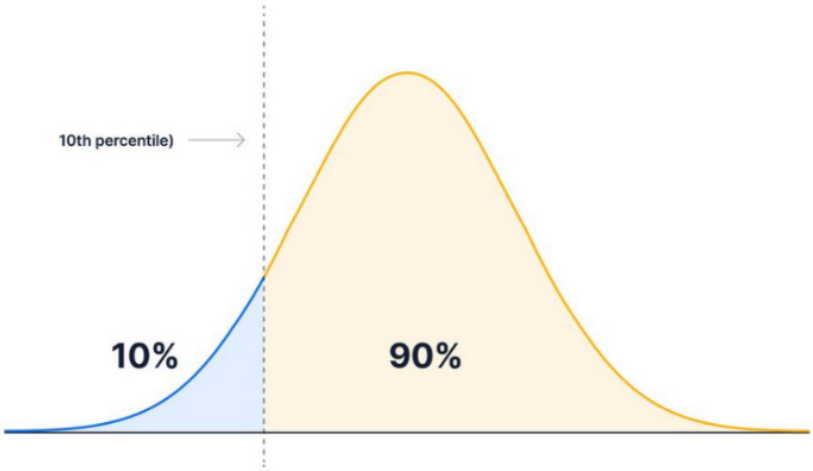
Figure 2. Flowchart modeling the steps of the sampling procedure.



Note. A flowchart modeling the steps of the sampling procedure for adolescents and adults. This flowchart helps quantify each of the steps and present it in a simple visual format, since there are multiple steps where participants are randomly selected.

To create the two standardized tests, two ten-question multiple-choice exams were written, including one reading passage with five comprehension questions and five arithmetic/algebraic math questions per exam. These questions were adapted from SATs from previous years released on the CollegeBoard website. Since the SAT progresses from easier questions to more difficult questions, the material for the adapted tests were taken from the beginnings of each section to control difficulty. The same two tests were administered to the experimental and control groups of both the adolescent and adult age groups. Standard bell curves were created in Excel to illustrate the fake results: one highlighted a result in the 10th percentile of test takers, as depicted in Fig. 3, and one highlighted a result in the 50th percentile. These were then saved as electronic files.

Figure 3. An example diagram of a standard bell curve highlighting a result in the 10th percentile.



Note. Including a diagram such as this allows for more credibility regarding the fake results administered to the participants, allowing them to view their score in a format similar to the SAT score representation.

To administer the tests to the adolescent experimental and control groups, the fifty participants were emailed with a time and place to take the multiple-choice test, as well as a list of items to bring (calculator, water, etc.). A confederate, whose behavior was rehearsed prior to the experiment, was employed to administer the test and the percentiles. The confederate provided each of the participants with a number two pencil and a copy of one test. The participants sat side-by-side in a classroom at Tesla STEM High School and were given fifteen minutes to complete the test. They were reminded when half the time was remaining and again when there were two minutes left. The confederate then collected the test forms and scored them using the scantron system, recording the scores of each experimental group participant in [Table I] and the scores of each control group participant in [Table II]. Then, the distribution curves representing a result in the 10th percentile of test takers were emailed to the experimental group participants, and the distribution curves representing a result in the 50th percentile of test takers were emailed to the control group participants. The confederate then administered the second test, giving the participants fifteen minutes to complete and reminding them when half the time was remaining and when there were two minutes left. The scores were once again recorded in the same manner, and the two distribution curves were emailed to the experimental and control groups accordingly. This process was repeated for the adult experimental and control groups, administering the test at Redmond Public Library and recording the experimental group scores in [Table III] and the control group scores in [Table IV].

To create the surveys inquiring into the participants' self-efficacy and perceived competency, a Google Forms was created in which a text box appeared at the beginning informing the participants that they scored in the 10th percentile of test takers. After this text box, participants were asked to rank the declarative statements provided by the Westside Test Anxiety Scale – inclusion of this measure in the survey was only to ensure the participants would stay unaware of the study's true purpose. The participants' scores were recorded in each group in the corresponding data tables. After the Westside scale, the declarative statements "I feel I am competent at the material on the test administered today," "I feel I can complete this test again and receive a better score," and "I feel my skills and capabilities align with my capacity to earn a high score on this test" were presented. A scale from zero to ten was presented under each statement on which the participants ranked how much they agreed with the statements (modeled after the Likert scale). The participants' Likert scale competence indexes for each statement (1-10) were recorded in each group in the corresponding data table. After this, the participants were presented with a modified version of the General Self-Efficacy Scale (GSE), which involves ranking ten declarative statements according to their resonance with an individual. The second, fourth, fifth, ninth,

and tenth declarative statements were eliminated for the purpose of this study. The participants' rankings of each statement were summed to calculate their GSE score. The participants' GSE scores were recorded in each group in the corresponding data table. A copy of this survey was made, in which the text box at the beginning informed the participants that they had instead scored in the 50th percentile of test takers. The first survey (10th percentile text box) was distributed to the experimental group and the copy of the survey (50th percentile text box) was distributed to the control group right after they took the test.

The participants were debriefed one day after the experiment through email: the purpose of the study and that the social statistics were fake was explained, as well as how each of the tasks they were asked to perform aided in forming a conclusion about the effects of upward social comparison – i.e., how their test scores, perceived competence, and self-efficacy changed when they received a low social statistic versus a higher one. An optional form with questions about which social statistic they received in response to their test (10th percentile or 50th percentile), their opinion on the effectiveness of the survey, and the authenticity of the feedback was included as well in order to improve future study.

Results:

Table I. Adolescent test scores, Westside test anxiety scores, Likert perceived competency indices, and GSE score after receiving a ranking in the 10th percentile of test takers.

Adolescents in experimental group	Test score 1	Test score 2	Test score difference	Westside test anxiety score	Likert perceived competency index – #1	Likert perceived competency index – #2	Likert perceived competency index – #3	GSE score (5-20)
#1	4	6	+2	2.7	8	9	9	19
#2	6	7	+1	2.9	7	10	9	14
#3	3	1	-2	3.7	2	7	3	14
#4	5	6	+1	3.8	5	4	8	16
#5	6	4	-2	3.3	6	3	7	15
#6	5	2	-3	2.9	3	6	6	17
#7	6	3	-3	3.0	7	8	9	16
#8	9	6	-3	4.0	7	7	9	15
#9	7	5	-2	3/6	7	7	8	15
#10	6	7	+1	3.1	7	9	8	14
#11	6	4	-2	2.8	4	10	10	13
#12	9	3	-6	3.5	2	5	2	15
#13	5	6	+1	3.4	2	7	9	18
#14	3	5	+2	3.3	6	7	7	14
#15	3	6	+3	3.1	4	8	8	17
#16	5	4	-1	2.9	6	7	4	15
#17	7	5	-2	2.9	8	10	3	16
#18	6	3	-3	4.8	1	10	3	10
#19	5	2	-3	4.1	2	6	3	12
#20	4	3	-1	3.7	7	10	8	15
#21	5	3	-2	3.8	8	10	8	16
#22	4	5	+1	3.6	7	5	7	14
#23	4	5	+1	3.1	4	8	5	13
#24	7	2	-5	3.4	6	7	4	15
#25	7	6	-1	3.3	0	6	0	14

Table II. Adolescent test scores, Westside test anxiety scores, Likert perceived competency indices, and GSE score after receiving a ranking in the 50th percentile of test takers.

Adolescents in control group	Test score 1	Test score 2	Test score difference	Westside test anxiety score	Likert perceived competency index – #1	Likert perceived competency index – #2	Likert perceived competency index – #3	GSE score (5-20)
#1	4	7	+3	3.0	7	7	8	18
#2	6	7	+1	2.6	9	10	9	16
#3	7	7	0	3.6	5	7	7	16
#4	5	7	+2	2.7	10	10	8	14
#5	7	6	-1	3.5	7	9	7	13
#6	5	6	+1	3.1	3	7	7	14
#7	4	5	+1	3.4	5	6	7	12
#8	6	8	+2	3.7	3	10	10	16
#9	5	4	-1	3.6	4	7	7	13
#10	6	4	-2	3.4	6	8	8	14
#11	6	6	0	3.2	7	9	7	16
#12	5	5	0	3.8	3	10	3	19
#13	5	6	+1	3.7	5	8	6	14
#14	6	7	+1	3.5	4	5	2	13
#15	5	5	0	3.3	6	6	6	15
#16	3	6	+3	3.1	7	3	8	17
#17	6	7	+1	2.9	7	10	8	15
#18	5	5	0	4.1	0	6	3	15
#19	8	5	-3	3.5	8	4	6	13
#20	7	7	0	3.4	7	8	2	18
#21	2	4	+2	3.8	5	9	7	13
#22	8	8	0	2.7	4	4	5	15
#23	5	5	0	3.9	3	9	1	15
#24	4	3	-1	2.0	9	9	9	15
#25	5	5	0	2.8	7	6	5	14

Table III. Adult test scores, Westside test anxiety scores, Likert perceived competency indices, and GSE score after receiving a ranking in the 10th percentile of test takers.

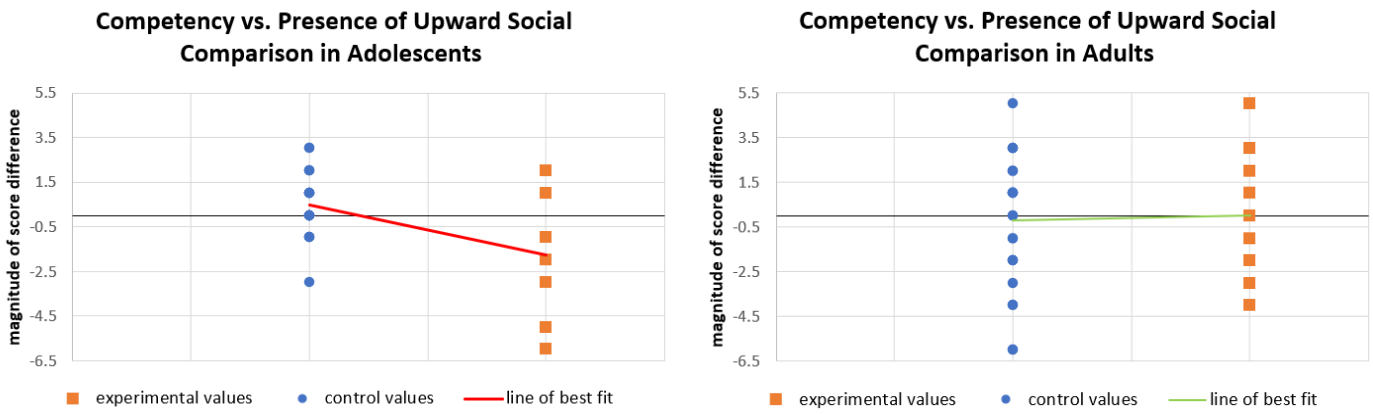
Adults in experimental group	Test score 1	Test score 2	Test score difference	Westside test anxiety score	Likert perceived competency index – #1	Likert perceived competency index – #2	Likert perceived competency index – #3	GSE score (5-20)
#1	3	4	+1	2.5	6	8	8	16
#2	6	2	-4	2.1	5	5	5	19
#3	7	3	-4	2.8	1	5	0	18
#4	5	2	-3	2.9	3	3	6	15
#5	6	2	-4	2.8	10	10	6	16
#6	4	9	+5	2.4	8	5	8	17
#7	3	6	+3	2.9	8	10	10	16
#8	6	6	0	2.0	6	4	6	19
#9	6	6	0	1.3	3	4	3	17
#10	2	4	+2	2.7	5	7	7	16
#11	4	5	+1	4.1	5	9	3	17
#12	5	5	0	2.7	5	5	6	15
#13	3	4	+1	2.8	3	3	3	15
#14	4	3	-1	3.2	8	9	7	18
#15	6	9	+3	2.8	6	8	7	19
#16	4	1	-3	2.3	3	7	8	14
#17	6	7	+1	2.9	10	10	6	18
#18	7	5	-2	2.0	7	8	9	17
#19	8	6	-2	2.6	7	7	7	14
#20	5	6	+1	2.5	1	5	5	16
#21	2	4	+2	2.9	6	7	7	16
#22	5	4	-1	2.8	6	6	9	15
#23	1	2	+1	2.0	3	8	8	16
#24	5	7	+2	2.6	7	8	7	16
#25	4	5	+1	2.5	2	7	2	15

Table IV. Adult test scores, Westside test anxiety scores, Likert perceived competency indices, and GSE score after receiving a ranking in the 50th percentile of test takers.

Adults in control group	Test score 1	Test score 2	Test score difference	Westside test anxiety score	Likert perceived competency index – #1	Likert perceived competency index – #2	Likert perceived competency index – #3	GSE score (5-20)
#1	7	1	-6	2.8	5	7	7	15
#2	4	9	+5	2.0	7	10	4	17
#3	2	5	+3	2.4	5	8	7	16
#4	6	7	+1	3.6	5	8	8	17
#5	4	7	+3	2.1	2	6	0	19
#6	3	6	+3	3.6	5	7	3	16
#7	7	1	-6	2.4	3	5	4	15
#8	5	1	-4	2.7	3	4	5	16
#9	5	2	-3	2.7	3	4	5	16
#10	5	3	-2	2.4	5	4	6	19
#11	5	6	+1	3.8	1	7	7	18
#12	7	5	-2	2.2	5	5	10	19
#13	7	6	-1	2.7	5	5	7	18
#14	6	2	-4	2.4	5	8	6	15
#15	2	4	+2	4.1	3	3	2	19
#16	2	1	-1	2.4	5	8	8	17
#17	3	3	0	1.4	10	10	10	19
#18	4	6	+2	4.4	5	10	0	14
#19	7	7	0	3.0	8	7	8	19
#20	5	4	-1	3.5	5	7	5	15
#21	3	4	+1	2.8	5	6	5	14
#22	2	3	+1	2.9	4	6	5	15
#23	5	3	-2	2.5	1	5	2	16
#24	6	7	+1	1.9	6	10	7	17
#25	2	4	+2	2.8	3	7	7	17

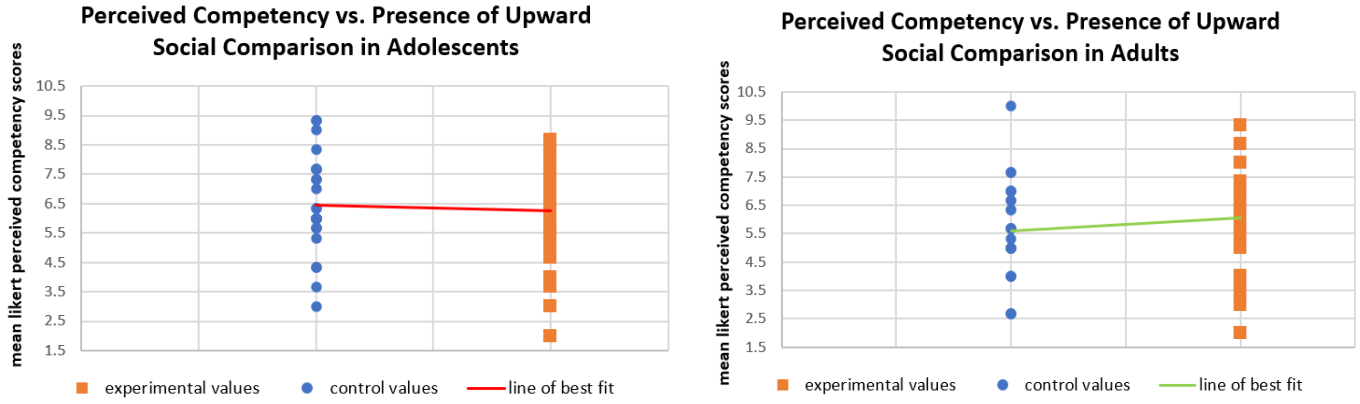
There was a significant negative correlation between competency (score differences) and the presence of upward social comparison (i.e., receiving the low score) in adolescents (R2 = -0.585), but no significant correlation between the two in adults (displayed in Fig. 4). This indicates a decline in adolescent test scores after receiving a low score percentile because of a possibly higher tendency to socially compare than adults, whose test scores did not change significantly due to a low score percentile; this supports the first hypothesis. There was also a significant negative correlation between perceived competency (mean Likert scores) and the presence of upward social comparison in adolescents (R2 = -0.557), and a significant positive correlation between the two in adults (R2 = 0.747), displayed in Fig. 5. This indicates a decline in adolescents' perceived competency at the standardized test after receiving a low score percentile, which supports the second hypothesis. However, the trend in adults does not support this: a positive correlation indicates a higher perceived competency at the test after receiving a low score, as opposed to adolescents who regard their competency with less confidence. There was no significant correlation between self-efficacy (GSE scores) and the presence of upward social comparison in adolescents, but a negative correlation in adults – however, this correlation coefficient was not significant; these results are displayed in Fig. 6. This data does not support the expected outcome for self-efficacy at all, as no correlation between the two in adolescents indicates that their self-efficacy levels did not change after receiving a low score percentile, and the adults' levels only decreased due to chance.

Figure 4. Graphs for competency (score differences) in adolescents and in adults.



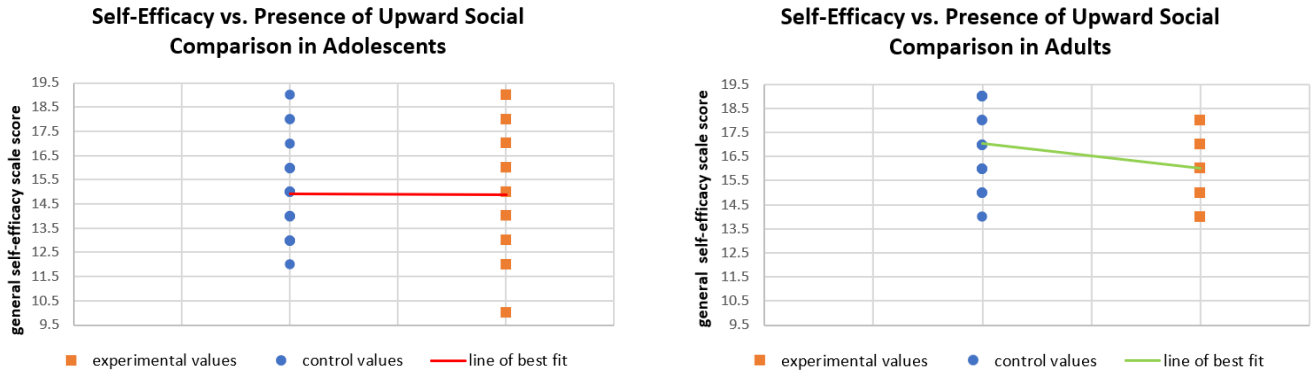
Note. Graphs showing competency as a function of the presence of upward social comparison in both age groups, with the experimental and control values plotted on the x-axis according to binary predictors (2 and 1, respectively) and a trendline displaying the correlations.

Figure 5. Graphs for perceived competency (mean Likert scores) in adolescents and in adults.



Note. Graphs showing perceived competency as a function of the presence of upward social comparison in both age groups, with the experimental and control values plotted on the x-axis according to binary predictors (2 and 1, respectively) and a trendline displaying the correlations.

Figure 6. Graphs for self-efficacy (GSE scores) in adolescents and in adults.



Note. Graphs showing self-efficacy level as a function of the presence of upward social comparison in both age groups, with the experimental and control values plotted on the x-axis according to binary predictors (2 and 1, respectively) and a trendline displaying the correlations.

To analyze these results, correlation coefficients, p-values, and z-values were calculated using Excel and XLSTAT. The data above was inputted into Excel as shown in Fig. 7, and XLSTAT was run on the data for each variable within both age groups (i.e., the data for the score differences in adults was run, then the data for the score differences in teens was run separately, then this was repeated for perceived competency and self-efficacy scores). After XLSTAT was run, it outputted the maximum and minimum, the mean, the standard deviation, the Pearson correlation coefficient and the p-value for statistical significance (shown in Fig. 8). The p-value was used to determine significant positive or negative correlation between each of the variables and the presence of upward social comparison.

Figure 7. Tables inputted into Excel in order to run XLSTAT.

adults

score difference				mean likert perceived competency score								general self-efficacy scale score			
group	experimental values	group	control values	group			experimental values	group			control values	group	experimental values	group	control values
2	1	1	-6	2	6	8	8 7.33333	1	5	7	7 6.33333	2	16	1	15
2	-4	1	5	2	5	5	5 5	1	7	10	4 7	2	19	1	17
2	-4	1	3	2	1	5	0 2	1	5	8	7 6.66667	2	18	1	16
2	-3	1	1	2	3	3	6 4	1	5	8	8 7	2	15	1	17
2	-4	1	3	2	10	10	6 8.66667	1	2	6	0 2.66667	2	16	1	19
2	5	1	3	2	8	5	8 7	1	5	7	3 5	2	17	1	16
2	3	1	-6	2	8	10	10 9.33333	1	3	5	4 4	2	16	1	15
2	0	1	-4	2	6	4	6 5.33333	1	3	4	5 4	2	19	1	16
2	0	1	-3	2	3	4	3 3.33333	1	3	4	5 4	2	17	1	16
2	2	1	-2	2	5	7	7 6.33333	1	5	4	6 5	2	16	1	19
2	1	1	1	2	5	9	3 5.66667	1	1	7	7 5	2	17	1	18
2	0	1	-2	2	5	5	6 5.33333	1	5	5	10 6.66667	2	15	1	19
2	1	1	1	2	3	3	3 3	1	5	5	7 5.66667	2	15	1	18
2	-1	1	-4	2	8	9	7 8	1	5	8	6 6.33333	2	18	1	15
2	3	1	2	2	6	8	7 7	1	3	3	2 2.66667	2	19	1	19
2	-3	1	-1	2	3	7	8 6	1	5	8	8 7	2	14	1	17
2	1	1	0	2	10	10	6 8.66667	1	10	10	10 10	2	18	1	19
2	-2	1	2	2	7	8	9 8	1	5	10	0 5	2	17	1	14
2	-2	1	0	2	7	7	7 7	1	8	7	8 7.66667	2	14	1	19
2	1	1	-1	2	1	5	5 3.66667	1	5	7	5 5.66667	2	16	1	15
2	2	1	1	2	6	7	7 6.66667	1	5	6	5 5.33333	2	16	1	14
2	-1	1	1	2	6	6	9 7	1	4	6	5 5	2	15	1	15
2	1	1	-2	2	3	8	8 6.33333	1	1	5	2 2.66667	2	16	1	16
2	2	1	1	2	7	8	7 7.33333	1	6	10	7 7.66667	2	16	1	17
2	1	1	2	2	2	7	2 3.66667	1	3	7	7 5.66667	2	15	1	17

adolescents

score difference				mean likert perceived competency score								general self-efficacy scale score			
group	experimental values	group	control values	group			experimental values	group			control values	group	experimental values	group	control values
2	2	1	3	2	8	9	9 8.66667	1	7	7	8 7.33333	2	19	1	18
2	1	1	1	2	7	10	9 8.66667	1	9	10	9 9.33333	2	14	1	16
2	-2	1	0	2	2	7	3 4	1	5	7	7 6.33333	2	14	1	16
2	1	1	2	2	5	4	8 5.66667	1	10	10	8 9.33333	2	16	1	14
2	-2	1	-1	2	6	3	7 5.33333	1	7	9	7 7.66667	2	15	1	13
2	-3	1	1	2	3	6	6 5	1	3	7	7 5.66667	2	17	1	14
2	-3	1	1	2	7	8	9 8	1	5	6	7 6	2	16	1	12
2	-3	1	2	2	7	7	9 7.66667	1	3	10	10 7.66667	2	15	1	16
2	-2	1	-1	2	7	7	8 7.33333	1	4	7	7 6	2	15	1	13
2	1	1	-2	2	7	9	8 8	1	6	8	8 7.33333	2	14	1	14
2	-2	1	0	2	4	10	10 8	1	7	9	7 7.66667	2	13	1	16
2	-6	1	0	2	2	5	2 3	1	3	10	3 5.33333	2	15	1	19
2	1	1	1	2	2	7	9 6	1	5	8	6 6.33333	2	18	1	14
2	2	1	1	2	6	7	7 6.66667	1	4	5	2 3.66667	2	14	1	13
2	3	1	0	2	4	8	8 6.66667	1	6	6	6 6	2	17	1	15
2	-1	1	3	2	6	7	4 5.66667	1	7	3	8 6	2	15	1	17
2	-2	1	1	2	8	10	3 7	1	7	10	8 8.33333	2	16	1	15
2	-3	1	0	2	1	10	3 4.66667	1	0	6	3 3	2	10	1	15
2	-3	1	-3	2	2	6	3 3.66667	1	8	4	6 6	2	12	1	13
2	-1	1	0	2	7	10	8 8.33333	1	7	8	2 5.66667	2	15	1	18
2	-2	1	2	2	8	10	8 8.66667	1	5	9	7 7	2	16	1	13
2	1	1	-1	2	7	5	7 6.33333	1	4	4	5 4.33333	2	14	1	15
2	1	1	0	2	4	8	5 5.66667	1	3	9	1 4.33333	2	13	1	15
2	-5	1	0	2	6	7	4 5.66667	1	9	9	9 9	2	15	1	15
2	-1	1	0	2	0	6	0 2	1	7	6	5 6	2	14	1	14

Note. Tables for the score difference values, the Likert scores and each participants' mean Likert score, and the GSE scores for each experimental condition in both age groups. The experimental condition was denoted by a binary predictor (2 for experimental group, 1 for control group).

Figure 8. XLSTAT output for adolescents' score differences and Likert perceived competency scores.

Summary statistics:

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
group	50	0	50	1.000	2.000	1.500	0.505
score diff	50	0	50	-6.000	3.000	-0.640	1.935

Correlation matrix (Pearson):

Variables	group	score diff
group	**1**	**-0.585**
score diff	**-0.585**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

p-values (Pearson):

Variables	group	score diff
group	**0**	**<0.0001**
score diff	**<0.0001**	**0**

Summary statistics:

Variable	Observations	Obs. with missing	Obs. without	Minimum	Maximum	Mean	Std. deviation
group	50	0	50	1.000	2.000	1.500	0.505
likert	50	0	50	2.000	9.333	6.133	1.884

Correlation matrix (Pearson):

Variables	group	likert
group	**1**	**-0.557**
likert	**-0.557**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

p-values (Pearson):

Variables	group	likert
group	**0**	**<0.0001**
likert	**<0.0001**	**0**

Note. XLSTAT output with summary statistics (minimum, maximum, mean, and standard deviation), Pearson correlation coefficient (R2 value, -0.585 for score difference and -0.557 for Likert perceived competency score), and p-value (<0.0001 for both correlations).

Two-tailed z-scores were also calculated to determine the significance of the relationship between the correlations in adults and adolescents for each of the variables. The scores were calculated using the standard z-score formula, which calculates the two scores and compares them:

$$Z_1 = 0.5 * \log\frac{(1+R1)}{(1-R1)}$$

$$Z_2 = 0.5 * \log\frac{(1+R2)}{(1-R2)}$$

$$z = \frac{(Z1-Z2)}{\sqrt{\frac{1}{(n1-3)}+\frac{1}{(n2-3)}}}$$

where:
R_1 = correlation coefficient for first group (adolescents)
R_2 = correlation coefficient for second group (adults)
n_1 = sample size of first group (50)
n_2 = sample size of second group (50)

The z-score for competency was -1.492, which indicates an insignificant relationship between the correlations in adults and adolescents. Though the correlation in adolescents is significant, it cannot be said that the correlation between the age groups is necessarily different due to anything but coincidence – more participants may make this score significant as its absolute value is quite close to the critical value of 1.96. The z-score for perceived competency was -3.357, which indicates a highly significant relationship between the correlations in adults and adolescents. This indicates that the correlation between age groups is significantly different. The z-score for self-efficacy level was -1.413, which indicates an insignificant relationship between the correlations in adults and adolescents.

Discussion:

The purpose of this study was to determine the relationship between social comparison and the competency, perceived competency, and self-efficacy in both adolescents and adults in a standardized test setting, in order to gain an understanding of the differences between the two age groups in that setting and to form a conclusion about these tests' predictive validity. After obtaining data from 100 participants – 50 adolescents and 50 adults – it was found that competency and self-efficacy level did not have a significantly different correlation between age groups, though their individual correlations were significant. However, the difference in the correlation between age groups for perceived competency was highly statistically significant, indicating that the two correlations for this variable can be compared and analyzed in further detail.

These results do not indicate any real difference in social comparison tendency between adolescents and adults, except for when the two

age groups perceived their competency at the standardized test. Adolescents felt they performed worse after receiving the lower score, which is supported by the fact that they hold the SAT in much higher importance than adults, and would thus feel discouraged after getting such a low score and worry about doing well when it matters. This is also supported by the fact that adolescent neurobiology is much more unstable, so they are more prone to quick conclusions and emotions, especially regarding themselves. The results also point to a lack of influence by upward social comparison on self-efficacy level, which is a positive – social comparison in high-stress standardized test settings are not as harmful as some studies that call for removal of the SAT claim. Therefore, this study points to a neutral perspective on the SAT's predictive validity – the results, when considered altogether, do not introduce a significant level of subjectivity that could undermine the quality of standardized tests.

From the competency results, it is safe to state that adolescents are affected much more by a lower score, as they try less and subsequently perform worse on the second test. Adults, on the other hand, perform relatively the same, because the SAT means less to them and their perception of success so they try equally as much after any news of their score, good or bad. From the perceived competency results, it is clear that adolescents perceive their performance as worse after receiving the lower score because their SAT scores are much more important to them. Adults actually perceive their performance as better after the lower score, because they are now motivated to do better and tend to develop strategies instead of specifically focusing on the outcome. From the self-efficacy results, it is clear that adolescents feel relatively the same about their capacity to succeed despite their scores on the test, which is highly productive – they are not attaching their score to their personal worth as much as predicted. Adults' self-efficacy actually decreased after receiving the lower score, but this cannot be attributed to anything but chance since the correlation coefficient was insignificant. From the results of these three variables, this study was unable to promote or demote the predictive validity of standardized tests, as there was no observable and significant difference between the social comparison tendencies of adolescents and adults.

Certain limitations that were not accounted for in the design of the study include students with higher or lower standards for themselves, who may not have perceived a score in the 50th percentile of test takers neutrally. This may also include students with more preparation or confidence, who may not have found the 10th percentile score believable. Other limitations include that adults may attach importance to the SAT in some circumstances and that the adolescent population was possibly biased because the participants were only students in the public school system. Next steps for the study include controlling for some of these limitations such as SAT experience and confidence factors, and generalizing the study to include other standardized tests (ACT, MCAT, GRE, etc.).

References:

(1980). Statistical Methods. Snedecor, G.W. and Cochran, W.G. (n.d.). Statistical Methods. 7th Edition, Iowa State University Press, Ames.

(2010). 5-Point Likert Scale. In: Preedy, V.R., Watson, R.R. (eds) Handbook of Disease Burdens and Quality of Life Measures. Springer, New York, NY. https://doi.org/10.1007/978-0-387-78665-0_6363

Ainscough, L., Foulis, E., Colthorpe, K., Zimbardi, K., Robertson-Dean, M., Chundury, P., & Lluka, L. (n.d.). Changes in biology self-efficacy during a first-year university course. In *Biomedical Sciences.* https://doi.org/10.11187/cbe.15-04-0092 (Reprinted from *Life Sciences Education*, 2016, *15*)

Anderson, M., & Jiang, J. (2018, November 28). *Teens, friendships and online groups*. Pew Research Center. Retrieved January 15, 2023, from https://www.pewresearch.org/internet/2018/11/28/teens-friendships-and-online-groups/

Bukalo, O., Pinard, C. R., Silverstein, S., Brehm, C., Hartley, N. D., Whittle, N., Colaciccio, G., Busch, E., Patel, S., Singewald, N., & Holmes, A. (n.d.). Prefrontal inputs to the amygdala instruct fear extinction memory formation. In *Neuropsychology*. https://doi.org/10.1126/sciadv.1500251 (Reprinted from *Science Advances*, 2015, *1*[6])

Butzer, B., & Kuiper, N. A. (n.d.).

Relationships between the frequency of social comparisons and self-concept clarity, intolerance of uncertainty, anxiety, and depression. In *Science Direct.* https://doi.org/10.1016 (Reprinted from *Personality and Individual Differences*, 2006, *41*)

Domenech-Betoret, F., Abellan-Rosello, L., & Gomez-Artiga, A. (n.d.). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. In *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.01193 (Reprinted from *Educational Psychology*, 2017)

Driscoll, R. (n.d.). Westside Test Anxiety Scale Validation. In *Self Evaliation (Individuals).* (Reprinted from *Test Validity*).

Fam, J. y., Murugan, S. b., & Yap, C. Y. L. (n.d.). Envy in social comparison-behaviour relationship: Is social comparison always bad? In *Research in Progress*. https://doi.org/10.1007/s12646-020-00575-7 (Reprinted from *Psychological Studies*)

Garcia, S. M., Tor, A., & Schiff, T. M. (n.d.). The psychology of competition: A social comparison perspective. In *Reprints and Permissions*. https://doi.org/10.1177/1745691613504114 (Reprinted from *Perspective on Psychological Science*, 2013, *8*[6])

Huguet, P., Marsh, H., Dumas, F., & Wheeler, L. (n.d.). Clarifying the role of social comparison in the big-fish-little-pond effect (BLFPE): An integrative study. In *PubMed*. https://doi.org/10.1037/a0015558 (Reprinted from *Journal of Personality and Social Psychology*, 2009, *97*[1])

Jansen, M., Boda, Z., & Lorenz, G. (n.d.). Social comparison effects on academic self-concepts – which peers matter most? In *Research in Progress*. https://doi.org/10.1037/dev0001368 (Reprinted from *Developmental Psychology*, 2022, *58*[8])

Krayer, A., Ingledew, D. K., & Iphofen, R. (n.d.). Social comparison and body image in adolescence: A grounded theory approach. In *Original Articles*. https://doi.org/10.1093/her/cym076 (Reprinted from *Health Education Research*, 2008, *23*[5])

Lee, J. K. (n.d.). The effects of social comparison orientation on psychological well-being in social networking sites: Serial mediation of perceived social support and self-esteem. In *Original Articles.* https://doi.org/10.1007/s12144-020-01114-3 (Reprinted from *Current Psychology*, 2020, *41*)

Morawetz, C., Bode, S., Baudewig, J., & Heekeren, H. R., (n.d.). Effective amygdala-prefrontal connectivity predicts individual differences in successful emotion regulation. In *National Library of Medicine*. https://doi.org/10.1093/scan/nsw169 (Reprinted from *Social Cognitive & Affective Neuroscience,* 2016, *12[4]*)

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35-37). Windsor, UK: NFER-NELSON.

Yusuf, M. (n.d.). The impact of self-efficacy, achievement motivation, and self-regulated learning strategies on students' academic achievement. In *Research in Progress*. (Reprinted from *Social and Behavioral Sciences*, 2011, *15*)

Comparison of Lung Capacity in Competitive Swimmers vs Non-Competitive Swimmers

Pearly Agarwal

Tesla STEM High School

Abstract:

In the total population, nearly 37 million Americans, 11.4%, including children, live with a chronic lung disease which leads to shortness of breath and difficulties with catching one's breath. A therapeutic form of lung capacity is swimming. The hypothesis is that over time consistent amounts of swimming increase a competitive swimmers lung capacity because swimming leads to the muscles in our body working harder to supply oxygen thus making the lungs stronger. Using a spirometry machine, the lung capacity (dependent variable) was measured as two groups, competitive swimmers (independent variable) and non-competitive swimmers (control group) were required to swim a 100 freestyle. Spirometry values of forced vital capacity (FVC), the total amount of air exhaled, and peak expiratory flow (PEF), measurement of air flowing out of the lungs, showed the most statistically significant values. The results were statistically significant with a p-vale of 0.01, in forced vital capacity, for the competitive males vs non-competitive males and p-value of 0.003, in peak expiratory flow, for competitive females vs non-competitive females. The hypothesis was general to all spirometry values and was correct as shown by the significance between the groups of swimmers and their spirometry results.

Introduction:

This project seeks to investigate how the effect of swimming in competitive swimmers vs noncompetitive swimmers could increase lung capacity and thereby help improve respiratory and cardiovascular problems. The maximum amount of air our lungs can hold, also known as our total lung capacity (TLC), is about six liters. As humans, our lungs mature up until the time we are 20-25 years old. After around the age of 35, the function of our lungs declines and as a result can cause breathing to become more difficult over time. Since 2019, human health around the world has been heavily declining, primarily through increased respiratory problems in the lungs (American Lung Association, 2022). The volume of adult female lungs is typically 10-12% smaller than that of males who are of the same height and age (Bellemare, Jeanerette, and Couture, 2003). Due to this, women tend to have a smaller maximal expiratory flow-volume (MEVF). Maximum expiratory flow volume is a curve that represents the inter-relationship between flow and volume during a maximal expiration. In young healthy subjects, the MEFV curve is highly reproducible within an individual, but shows between-subject flexibility. As a result, their ability to generate increased ventilation during exercise is typically less than that of men. Nearly 37 million Americans, 11.4% of the total population, including around 6 million children, live with a chronic lung disease (Bajaj, 2022). A potential solution for challenges involving the respiratory function of lung capacity is the practice of exercise, specifically competitive or non-competitive swimming. Athletes who participate or compete in sports that involve physical strength, endurance, and speed typically undergo positive effects, including improved mental health, physical health, staying fit, etc. Research evidence suggests that exercise and physical activities have many benefits and positive effects for the health and well-being of individuals (Biddle, Gurley, & Mutrie, 2022). A major factor contributing to an athlete's

performance is their health as it enhances their routine. Practice and training can be exhausting for athletes who may not fuel their bodies properly or be completely healthy, in which healthy exercise is needed to stay fit. Although alternate exercise may also enhance their health in aspects such as mental health and body strength, intense swimming workouts have specifically been shown to significantly improve athletes' cardiorespiratory capacity.

Swimming differs significantly from other sports in that the horizontal position of the swimmer in water causes different gravitational and resistive forces to operate, as well as a different pattern of respiration (Zarzeczny, Kuberski, Suliga, 2022). As a result, its health outcomes for children and adolescents, such as cardio-respiratory fitness (CRF) or physical development, may differ from those offered by other types of exercise. Swimming has a particularly beneficial effect on lung, blood, and heart volumes, as well as maximal oxygen uptake and aerobic capacity. During submersion while swimming, the hydrostatic pressure of water causes compression into the chest wall, which affects the volume of breathing by an approximately 6-11% decrease in vital capacity (Keskinen, Paivinen, and Tikkanen, 2021). While swimming, the thorax is subjected to uncommon conditions. Water immersion places an external weight on the thorax, increasing airway resistance, and the horizontal body position causes central venous engorgement and a reduction in lung compliance. Regular swimmers have a 28% lower risk of early death and a 41% lower risk of death due to heart diseases and stroke (England Health Commission, 2017). Swimmers have about half the risk of death compared to inactive people. Swimming is the fourth most popular sport activity in the United States as Competitive swimming is a sport that provides maintenance of high level of activity and can make your lungs bigger and more efficient in processing oxygen. Larger lungs would allow greater vital capacity which increases the amount of oxygen in the body, allowing competitive swimmers to train longer and without the feeling of extensive effort. On the other hand, decreased vital lung capacity causes shortness of breath and difficulties with catching your breath. Over time, consistent swimming will increase an individual's lung capacity because physical activity makes the muscles, heart, and lungs stronger. As it improves circulation and strengthens tissue around the lungs, helping them function, strengthening the respiratory system. Therefore, competitive swimmers show results of greater lung capacity compared to those who are non-competitive swimmers.

Methodology:

In this study, the comparison of lung capacity in competitive swimmer's vs non-competitive swimmers was tested. This experiment was conducted using a spirometry machine, a 25-yard swimming pool, and a set of strokes and distance for participants to swim. Participants were recruited through flyers providing a brief description of the requirements and experiment, followed by a QR code to the participant questionnaire. These flyers were posted around local schools and stores. The participant questionnaire included questions collecting information on anthropometric data such as age, height, weight, and BMI, along with an informed consent form. Refer to data **table 1**.

Table 1- Anthropometric characteristics of female competitive swimmers

	Age (years)	Body Weight(lbs.)	Body Height (cm)	BMI ((703 x lbs)/ in^2)
Participant #3	**16**	**205**	**172.72**	**31.6**
Participant #5	**15**	**98**	**152.4**	**19.1**
Participant #6	**16**	**113**	**152.4**	**22.1**
Participant #8	**17**	**98**	**165.1**	**16.3**
Participant #12	**16**	**125**	**157.48**	**22.9**

BMI = body mass index

Table 2- Anthropometric characteristics of male competitive swimmers

	Age (years)	Body Weight(lbs.)	Body Height (cm)	BMI ((703 x lbs)/ in^2)
Participant #4	**16**	**150**	**182.88**	**20.3**
Participant #9	**17**	**125**	**167.64**	**20.8**
Participant #11	**16**	**175**	**180.34**	**24.4**
Participant #13	**17**	**170**	**180.34**	**23.7**
Participant #18	**16**	**175**	**167.64**	**28.2**

BMI = body mass index

Table 3- Anthropometric characteristics of female non-competitive swimmers

	Age (years)	Body Weight(lbs.)	Body Height (cm)	BMI ((703 x lbs)/ in^2)
Participant #2	**17**	**130**	**170.18**	**20.4**
Participant #7	**16**	**120**	**162.56**	**20.6**
Participant #14	**16**	**N/A**	**N/A**	**N/A**
Participant #15	**17**	**125**	**170.18**	**19.6**
Participant #20	**17**	**118**	**162.56**	**20.3**

BMI = body mass index

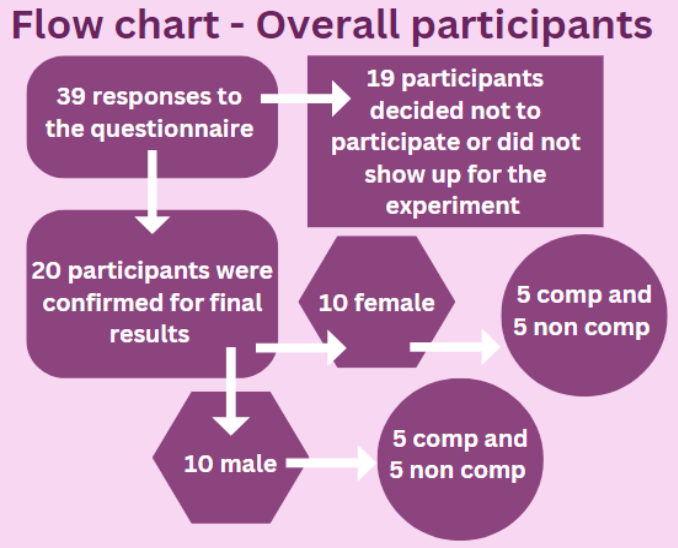
Table 4- Anthropometric characteristics of male non-competitive swimmers

	Age (years)	Body Weight(lbs.)	Body Height (cm)	BMI ((703 x lbs)/ in^2)
Participant #1	**18**	**140**	**187.96**	**18.5**
Participant #10	**16**	**140**	**182.88**	**19.0**
Participant #16	**16**	**113**	**180.34**	**15.8**
Participant #17	**17**	**155**	**162.56**	**26.6**
Participant #19	**17**	**132**	**180.34**	**18.4**

BMI = body mass index

Further questions were asked inquiring about information about participant qualifications such as whether they are high school students, between the ages of 14 to 18 years old, and if they are competitive swimmers or not. Once potential participants had filled out the questionnaire, participants that meet the requirements were contacted back with proceeding information. Final participants consisted of 20 competitive swimmers split as 10 female and 10 male, and 20 non-competitive swimmers, split as 10 female and 10 males. A random number generator was used to assign each person a number as their identifier followed by an informed consent form to be filled in by each participant and their guardian, if the participant is under 18 years old.

Figure 1. Flow Char - Distribution of participants



This is a flowchart representing the distribution of participants from original responses to finalized groups.

Participants were tested at the pool as they were asked to swim a 100 freestyle. After each person had swum the 100 freestyles, the spirometer was used to test lung capacity. To use the spirometer, participants took a deep breath and placed their mouths tightly around the mouthpiece as they exhaled as fast as they could. An image of the machine used in this experiment can be referred to in Diagram 3.

Figure 2. Visual use of the spirometry machine.



This image is of the spirometry machine used in this experiment with a sample display of results.

After each test with the spirometry machine, detach the mouthpiece from the base of the device and clean it with clean running water and antibacterial soap for about 20 seconds, before testing the next participant. Results of the spirometer machine were measured for value in vital capacity, forced vital capacity, forced expiratory volume in first second, peak expiratory flow, the measurement of air flowing out of the lungs, and maximum voluntary ventilation, the largest volume of air in and out of the lungs in a 10 to 15 second interval, in both groups. (Refer to Table 5).

Table 5- Pulmonary function values for female competitive swimmers

	FEV1/ FVC	FVC	FEV1	PEF	MVV
Participant #3	**86%**	**3.34**	**2.95**	**257**	**2544.890**
Participant #5	**97%**	**2.92**	**2.84**	**374**	**2901.289**
Participant #6	**93%**	**3.06**	**2.85**	**336**	**2611.854**
Participant #8	**58%**	**4.15**	**2.21**	**327**	**2532.860**
Participant #12	**87%**	**4.12**	**3.31**	**310**	**2421.186**

VC= vital capacity, FVC= forced vital capacity, FEV1= forced expiratory volume in first second, PEF= peak expiratory flow, MVV = maximum voluntary ventilation

Table 6- Pulmonary function values for male competitive swimmers

	FEV1/ FVC	FVC	FEV1	PEF	MVV
Participant #4	**55%**	**5.52**	**3.06**	**235**	**1845.546**
Participant #9	**83%**	**4.78**	**3.96**	**478**	**3712.081**
Participant #11	**73%**	**4.39**	**3.19**	**236**	**1855.281**
Participant #13	**95%**	**5.08**	**4.83**	**557**	**4328.283**
Participant #18	**100%**	**4.03**	**4.03**	**506**	**3926.607**

VC= vital capacity, FVC= forced vital capacity, FEV1= forced expiratory volume in first second, PEF= peak expiratory flow, MVV = maximum voluntary ventilation

Table 7- Pulmonary function values for female non-competitive swimmers

	FEV1/ FVC	FVC	FEV1	PEF	MVV
Participant #2	**42%**	**4.53**	**1.91**	**136**	**1072.372**
Participant #7	**97%**	**2.42**	**2.35**	**235**	**1834.00**
Participant #14	**95%**	**2.91**	**2.76**	**233**	**1825.427**
Participant #15	**68%**	**3.07**	**2.53**	**230**	**1798.825**
Participant #20	**56%**	**4.43**	**2.50**	**184**	**1447.772**

VC= vital capacity, FVC= forced vital capacity, FEV1= forced expiratory volume in first second, PEF= peak expiratory flow, MVV = maximum voluntary ventilation

Table 8- Pulmonary function values for male non-competitive swimmers

	FEV1/ FVC	FVC	FEV1	PEF	MVV
Participant #1	**85%**	**4.20**	**3.57**	**296**	**2318.719**
Participant #10	**100%**	**4.47**	**4.47**	**683**	**5282.677**
Participant #16	**100%**	**2.99**	**2.99**	**406**	**3147.600**
Participant #17	**86%**	**3.22**	**2.77**	**251**	**1962.768**
Participant #19	**70%**	**3.47**	**2.50**	**230**	**1798.338**

VC= vital capacity, FVC= forced vital capacity, FEV1= forced expiratory volume in first second, PEF= peak expiratory flow, MVV = maximum voluntary ventilation

Results:

Using a TI-84 CE calculator and the 2-sample t-test, the data was analyzed. In doing so, results were provided for standard deviation, mean, and p-value. Data was compared between groups of female competitive and non-competitive swimmers along with male competitive vs non-competitive swimmers. Level of significance is $p<0.05$ and two values resulted to be statistically significant. From the female groups a p-value of 0.0032 for the PEF, measurement of air flowing out of the lungs, and from the male groups a p-value of 0.011 for the FVC, the total amount of air exhaled during an FEV test. When looking at the volume of lungs, females usually have 10-12% smaller lung volume than males, even if they were the same age and height. Due to this the female groups may have showed significance in PEF as they would have a faster breathing rate, and males showed significance for FVC as they have larger lungs.

Table 9 – T-test results between female competitive swimmers and female non-competitive swimmers

	P-value	SD 1	SD 2	ME 1	ME 2
MVV	**.0859**	**180.559**	**334.132**	**2606.416**	**1595.679**
PEF	**.0032**	**42.681**	**96.158**	**320.800**	**119.600**
FEV1	**.159**	**0.344**	**0.316**	**2.632**	**2.41**
FVC	**.465**	**0.583**	**0.951**	**3.518**	**3.472**
FEV1/FVC	**.179**	**0.153**	**0.241**	**0.842**	**0.716**

Female competitive swimmer = group 1, female non-competitive swimmers = group 2, SD = standard deviation, ME = mean, level of significance: $p < 0.05$

Table 10 – T-test results between male competitive swimmers and male non-competitive swimmers

	P-value	SD 1	SD 2	ME 1	ME 2
MVV	**.394**	**1192.052**	**1429.019**	**3133.559**	**2902.020**
PEF	**.397**	**154.969**	**186.044**	**402.400**	**373.200**
FEV1	**.138**	**0.717**	**0.783**	**3.814**	**3.26**
FVC	**.011**	**0.581**	**0.638**	**4.760**	**3.670**
FEV1/FVC	**.322**	**0.332**	**0.390**	**0.632**	**0.522**

Male competitive swimmer = group 1, male non-competitive swimmers = group 2, SD = standard deviation, ME = mean, level of significance: $p < 0.05$

Conclusion:

The hypothesis, over time consistent amounts of swimming will increase a competitive swimmers lung capacity because swimming leads to the muscles in the body working harder to supply oxygen thus making the lungs stronger, was supported. Through spirometry lung capacity was tested in groups of male and female competitive and non-competitive swimmers. When observing the male groups, their results had greater differences than the female groups whose results were closer in value. Physical activity makes the muscles, heart, and lungs stronger. Exercise improves circulation and strengthens the tissue around the lungs, helping them function, strengthening the respiratory system. This is vital because the primary function of the respiratory system is to introduce new air into your body while expelling waste gases.

Limitations:

In evaluating lung capacity there are external factors that can weigh into the results. As mentioned previously, one male participant was an

outlier with the highest result for MVV, this might be because they may not be a competitive swimmer however, they could be a competitive track runner. Data on the history of other sports or years spent on a sport was not collected due to a time constraint. Further information on participant health would have included history of asthma, although due to discomfort this data was not collected or analyzed.

Future Application:

In the future, this could be conducted as a long-term experiment, collecting data on lung capacity prior to beginning a sport and continuing to collect data through the years of participation. Analyzing data across gender groups of male and female to look for results withing a gender gap which could provide further insight. Spirometry is often used amongst asthma patients, so collecting data on a participant's history of asthma, if applicable, could also present variations in the results. Lung capacity has been shown within sports however it is also visible withing instruments. Testing lung capacity in instrument players such as the flute, tuba, bag pipe, etc. could show a new aspect of results and spirometry, beyond that of health through athletics.

Bibliography:

Anastasio, F., Barbuto, S., Scarnecchia, E., Cosma, P., Fugagnoli, A., Rossi, G., Parravicini, M., & Parravicini, P. (2021). Medium-term impact of COVID-19 on pulmonary function, functional capacity and quality of life. *The European respiratory journal*, *58*(3), 2004015. https://doi.org/10.1183/13993003.04015-2020

Berge, H. M., Isern, C. B., & Berge, E. (2015, June 1). *Blood pressure and hypertension in athletes: A systematic review. British Journal of Sports Medicine. Retrieved November 19, 2022, from* https://bjsm.bmj.com/content/49/11/716

Lazovic-Popovic, B., Zlatkovic-Svenda, M., Durmic, T., Djelic, M., Saranovic, S. D., & Zugic, V. (2016, February 22). *Superior lung capacity in swimmers: Some questions, more answers! Revista Portuguesa de Pneumologia (English Edition). Retrieved November 16, 2022, from* https://www.sciencedirect.com/science/article/pii/S2173511515001979

Päivinen, M., Keskinen, K., & Tikkanen, H. (2021, May 20). *Swimming-induced changes in pulmonary function: Special observations for clinical testing - BMC sports science, medicine and Rehabilitation. BioMed Central. Retrieved November 20, 2022, from* https://bmcsportsscimedrehabil.biomedcentral.com/articles/10.1186/s13102-021-00277-1

Schnabel, E., Nowak, D., Brasche, S., Wichmann, H. E., & Heinrich, J. (2011). Association between lung function, hypertension and blood pressure medication. *Respiratory medicine*, *105*(5), 727–733. https://doi.org/10.1016/j.rmed.2010.12.023

Elaboration on the Effect Maternal Education has on Fertility, Age at First Birth, and Contraceptive Use

Moomal Siddiqui

Nikola Tesla STEM High School

Abstract

Women's reproductive rights are a contentious topic with Roe v. Wade overturned and it is important to understand what is influencing women's fertility choices and if these factors can be altered or eliminated. The purpose of this study is to investigate if lower levels of education correlates with higher rates of pregnancy and conception at a young age due to a lack of education on contraceptives. Participants completed a ten-question survey enquiring on their level of education (number of years spent pursuing education), fertility (number of times a woman conceives), age at first birth, and contraceptive use (use of a service to prevent pregnancy). Volunteers participated in a focus group over zoom where additional questions were asked to help understand alternate factors that women feel effect reproductive choices. This project used Pearson correlation to analyze the association between level of education and number of children (R=-0.73), level of education and age at first birth (R=0.59), and used Polychoric correlation to calculate the correlation between level of education and contraceptive use (R=0.67). The project employed a K-Means algorithm to provide a cluster graph showcasing the relationship between years of education and number of children. The clusters showed that women with more years of education (16-20 years) had fewer children than women with fewer years of education (8-16 years) who had more children. The results of this investigation supports the hypothesis that women's education heavily impacts reproductive decisions. This research demonstrates the importance of female education and its influence over fertility.

Introduction

This project aims to investigate how a current lack of female education is leading to prevalent social issues such as adolescent fertility, a lack of knowledge regarding contraceptives, and an increased frequency in prenatal complications. A 2005 study in Kenya found that the median age of marriage was twenty years for women with a secondary education or more, eighteen years for women with a primary education, and sixteen years for women with no education (Ikamari, 2005). As seen in this study, women with higher levels of education had a higher median age of marriage. When pursuing high levels of education, women often end up postponing marriage and children in favor of focusing on completing their education. An increased number of women pursuing a tertiary education will cause a decline in adolescent pregnancy because there will be a strong focus on completing education and obtaining a high-skilled and high paying job (LeMoyne, 2020). Women with higher levels of education will not feel pressured to marry in order to secure financial and social stability. Higher levels of education will lead to the postponement of first births and therefore motivate women to have fewer children of a higher quality (McCrary & Royer, 2011). The term higher quality children refer to children that receive the benefit of a highly educated mother who can better care for her children. As women get older, they recover less quickly after birth and often have fewer children as a result. Children with educated parents often grow up to pursue higher level education as a result of the influence their home environment has had

on them. Providing more women with an extensive education will lead to their children receiving the same and becoming a much more productive part of society.

Low levels of education (especially education that does not cover adolescent years) results in women being less knowledgeable about contraceptives – devices or medication designed to prevent pregnancy. Commonly used contraceptives include plan B (morning after pill), birth control pills, and sterilization. Women who either do not have access to contraceptives or are unaware of them, will have higher rates of fertility because they are unable to safely prevent conception. A study done in Bangladesh using nationally representative data showed a positive association between a woman's level of education and use of contraception. (Waiz, 2000). Providing women with a complete and high-quality education that teaches both academics and life skills through qualified teachers, will give them the opportunity to learn about ways to safely prevent or terminate pregnancies. Most adolescents who conceive do not desire to have children at such a young age and are often driven to unsafe contraceptive or abortion practices. This is highly risky especially for those who are uneducated or living in an un-hygienic environment, as their risk of disease and infection resulting from unsafe procedures is much higher. A study on family planning found that around 32% of maternal deaths and 90% of abortion-related deaths could be prevented through the use of contraceptives (Cleland et al., 2006). The correct use of contraception can save lives if women are educated on how to use them effectively. Currently, millions of women do not have access to contraceptives and millions more are uneducated on how to use them effectively and safely. In 2019, 270 million women had a need for contraceptives (WHO, 2020). Educating women on how to successfully and safely use and access contraceptives and abortion would cause a significant decrease in the number of maternal deaths.

Adolescent pregnancy can also result in prenatal health issues as the adolescent body may not be fully developed. In a study done by the Center for Disease Control from 2007 to 2014, researchers determined that the preterm birth rate for women less than 20 years old in 2014 was 9.82% compared to 8.87% for women ages 25-29 (Ferré et al., 2017). The preterm birth rates for adolescent mothers (under twenty years old) were 0.95% higher than that of women in their twenties. Adolescent pregnancies are more dangerous for both the mother and the infant due to the body's development and adolescent knowledge regarding pregnancy and pre-natal health. Uneducated teenagers often do not know how to prevent unwanted pregnancies or how to care for themselves, and this often results in complications that are harmful for the mother and infant. A longitudinal study that compared the risk of adolescent maternal mortality to that of older women estimated "the risk of death per birth for adolescents aged 15–19 years is only 28% higher than for women aged 20–24 years." (Nove et al, 2014). Adolescents are in greater danger than older women because they are less educated on prenatal health and are much less prepared for pregnancy and children compared to older women who have been educated on how to stay healthy. Adolescent bodies aged 15-19 years are also much less developed and development can vary from person to person, making their bodies less prepared to carry children. I hypothesize that women with lower levels of education will experience higher rates of pregnancy due to a lack of education on fertility choices and contraceptives (birth control, plan B). This will also lead to women having children at a younger age since they will not be pursuing higher level education and will not be working high skilled or high paying jobs.

Methodology:

The participants selected for this correlational study were all female, over sixteen years of age, and had conceived a child at least once in their life. Vulnerable populations included pregnant women and minors who qualify as participants in this correlational experiment. Physical flyers, as seen in Figure 1, containing a QR code leading to the survey were posted in obstetrician-gynecologists offices, and in prenatal classes in order to reach the intended audience. An online eleven-question survey, as shown in Figure 2, was distributed online via Facebook and Instagram accounts. The posts and flyers clearly informed participants that the study requires them to fill out an informed consent form and a two-to-five-minute questionnaire. The top of the flyer and posts indicates that the purpose of the study is to observe how maternal education is affecting fertility, age at first birth, and contraceptive

use, and determine what factors of education are necessary to reduce the effect. In addition, it is stated twice that participants can withdraw from the study at any point in time. Both the survey and the flyer contain an attached informed consent form which outlines the risks (no physical risks to any participants, but the questionnaire may elicit sensitive memories) and benefits: provide data on the effect education has on fertility, age at first birth, and contraceptive use; therefore, helping to bring attention the importance of female education.

The first step in the gaining participants was to distribute the survey on an Instagram account made specifically for this project. A logo developed in canvas was distributed alongside the survey to represent the project (Figure 3). Physical flyers will also be displayed in obstetrician-gynecologists offices and in prenatal classes. The participants were asked to fill out an informed consent form before participating in the survey and if participants were under eighteen, they needed a parent/guardian signature. Once responses and informed consent forms are received a random number generator will be used to gain a representative sample. A minimum of twenty responses will be used and each participant will be identified through a letter throughout the study.

For part one of the correlational study the survey asks participants nine questions, and the responses were recorded in Table 1.

Part two of the correlational study was optional for participants who agreed to take part in a focus group. The focus group was a zoom meeting with no more than five people where we discussed participants thoughts on the survey questions. This projects mentor (Dr. Jodi O'Brien) was also present during the focus group. The focus group discussed ten additional questions that elaborated on the survey questions:

1.) How do you think your level of education impacted the number of children you had or are planning to have?
2.) Do you think the quality of education you received has impacted your fertility choices in any way?
3.) If you have ever used contraceptives, why and when did you make that choice?
4.) When were you educated about contraceptives and different types?
5.) Why did you choose to have children at a specific age?
6.) How many years did you go to school?
7.) Are you currently employed? If so, where?
8.) Were you in a relationship when you conceived?
9.) Was contraception widely talked about when you were growing up?
10.) At what age did your parents first have children? Was it at a much different age then you?

These questions were based on the hypothesis and the responses will help to support or not support the hypothesis regarding the effect maternal education has on fertility, age at first birth, and contraceptive use.

Results:

The hypothesis that women with lower levels of education will experience higher rates of pregnancy due to a lack of education on fertility choices and contraceptives; subsequently, leading women having children at a younger age since they will not be pursuing higher levels education and will not be working high skilled or high paying jobs, was supported by this study. For data analysis this study used both qualitative and quantitative analysis.

Qualitative data analysis found that religion has a very large impact on fertility. Women's belief about sex and children before or after marriage often stems from religion and influences their likelihood of conceiving early. Women who believed in waiting to have sex or children until after marriage often got married and had children later in life because they were waiting for a suitable partner. Women who found it culturally acceptable to have sex or children before marriage tended to have children at a younger age because there was a higher risk of pregnancy. One of the most recognized impacts of religion is belief in the use of contraceptives/abortion which varies among cultures and influences the likelihood of pregnancy. Those women who favored the use of contraceptives/abortion typically had children at specifically planned time – when they wanted children. In terms of the workforce, current literature references that high levels of education in women leads to their immediate return to the workforce

post birth (Knop, 2021). The study's focus group conversations found that this is not always true, and women are likely to stay home for the first 6-12 months post-birth regardless of level of education. There is a common misconception that women with high levels of education are less maternal than other women and this is a societal expectation women need to break out of. This study also found that a younger age at first birth is associated with more children overall and an older age at first birth is associated with fewer children overall. This is likely due to physiological differences since women who choose to have children later in life cannot have as many as quickly for it to be healthy for them or their baby. While women who had children in life had fewer children, these children were of a higher quality, meaning they had the benefit of an educated mother. When children are planned by a couple it is mainly for financial reasons. Families looking for financial security before children often have them later in life to ensure quality of life and steady financial environment.

For quantitative data analysis the relationship between level of education and number of children was analyzed using the Pearson Correlation Coefficient and found a negative correlation with an R value of -0.73 as shown in Figure 5. This negative correlation showcases that women who spend more years pursuing an education tended to have fewer children because they were focused on pursuing an education. This study analyzed the relationship between level of education and age at first birth and found a positive correlation with an R value of 0.59 as shown in Figure 6. The positive correlation here demonstrates that more years of education in women lead to a later age at first birth. Lastly, the study analyzed the relationship between level of education and contraceptive use using a polychoric correlation coefficient using library(polycor) in language R in RStudio, where the x values in the code were years of education and the y values were contraceptive use (as shown in the code in Figure 7). A polychoric correlation was used because the question of whether women have used contraceptives requires a yes or no response which is an ordinal variable not a numerical one. A positive correlation and a R value of 0.67 was found, meaning that more years of education in women is associated with more use of contraceptives to prevent pregnancy.

This study used a type of unsupervised machine learning called a K-Means clustering algorithm to cluster the data because there were multiple features. Each bracket holds data for one participant in the order of years of education, number of children, age at first birth, and contraceptive use (1=yes, 2=no). The algorithm calculated the mean for each person, compared the means, and clustered the values in a graph demonstrating the relationship between level of education and number of children (Figure 4). The red cluster indicated more children due to less education, the blue is fewer children due to more education, and the green cluster is an outlier which is fully expected since this is an observational study dealing with subjective variables.

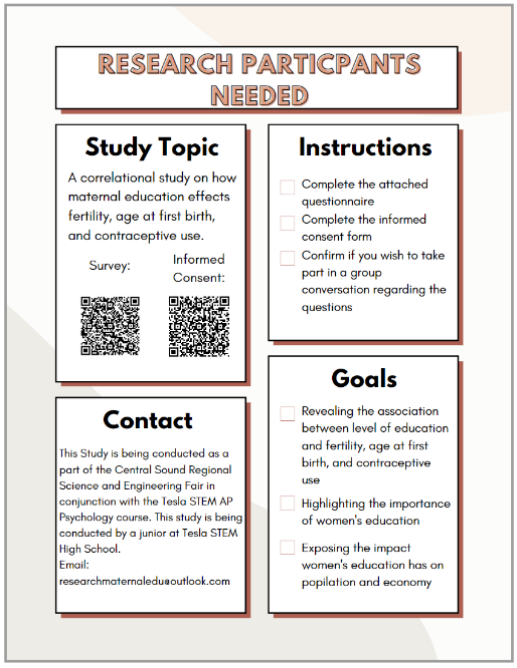
Conclusion:

The hypothesis that women with lower levels of education will experience higher rates of pregnancy due to a lack of education on fertility choices and contraceptives; therefore, leading women having children at a younger age since they will not be pursuing higher levels education and will not be working high skilled or high paying jobs was supported by this study. One of the limitations of this study was the that it only included thirty-five participants. In the future a larger sample size will be obtained from a wider variety of people in order to account for confounding variables. In the future this study could evolve to research the impact religion has on reproductive choices and how that compares or contrasts to education. Another avenue this study could take would be to explore the impact of social norms on women's marriage and reproductive choices and the impact of these complex social patterns on women in modern society.

Diagrams:

Table 1. Survey questions and responses from 35 anonymous participants identified through a letter

What is your highes	What is your relatic	What level of educ	How many pregnan	What was your age	What was the year	Have you ever used	Are you currently w	Would you be willi
16	In a relationship	Masters degree	1	34	2018	Yes	Yes	Yes
16	In a relationship	High School Diploma	1	20	1989	Yes	Yes	Yes
16	In a relationship	Bachelor	2	28	2020	Yes	Yes	Yes
16	In a relationship	Master's Degree	1	28	2017	Yes	No	Yes
16	In a relationship	Undergraduate	2	35	2019	Yes	No	Yes
18	In a relationship	Graduate degree	2	29	2017	Yes	Yes	Yes
18	In a relationship	College	1	36	2018	Yes	No	No
18	In a relationship	Masters, undergradua	2	24	2020	Yes	No	Yes
18	Married		1	26	2005	No	Yes	No
16	In a relationship	Masters	2	25	2005	Yes	Yes	No
16	In a relationship		2	27	2019	Yes	Yes	Yes
18	In a relationship	K-12	1	32	2018	Yes	Yes	No
16	In a relationship	College degree	1	35	2020	Yes	No	No
16	Married	Masters	2	25	2006	No	Yes	Yes
18	In a relationship	Masters	1	21	2000	Yes	Yes	Yes
18	In a relationship	Postgraduate	1	26	2005	No	Yes	No
18	In a relationship	master's degree	1	31	2006	Yes	Yes	Yes
16	In a relationship	Undergrad	2	26	2015	yes	No	No
12	Not in a relationship		4	20	2017	No	No	No
16	In a relationship	Masters	2	26	2005	No	Yes	No
12	In a relationship	Undergrad	3	22	1970	No	No	No
8	dead husband	high school	3	18	1973	No	No	No
18	In a relationship	Masters	1	28	2015	Yes	Yes	No
12	Not in a relationship		3	25	2000	No	Yes	No
12	In a relationship	Undergrad	2	21	2010	No	No	No
18	In a relationship	Masters	1	23	2006	No	No	No
18	In a relationship	Undergrad	1	30	2016	Yes	No	No
16	In a relationship	Masters	2	28	2013	Yes	No	No
16	Not in a relationship		2	24	2009	No	Yes	No
20	In a relationship	Masters	1	35	2017	Yes	Yes	No
20	In a relationship	Masters	2	33	2014	Yes	No	No
12	In a relationship	High School	3	19	2018	No	No	No
16	Not in a relationship		1	21	2015	No	Yes	No
16	Not in a relationship		2	21	2015	No	Yes	No



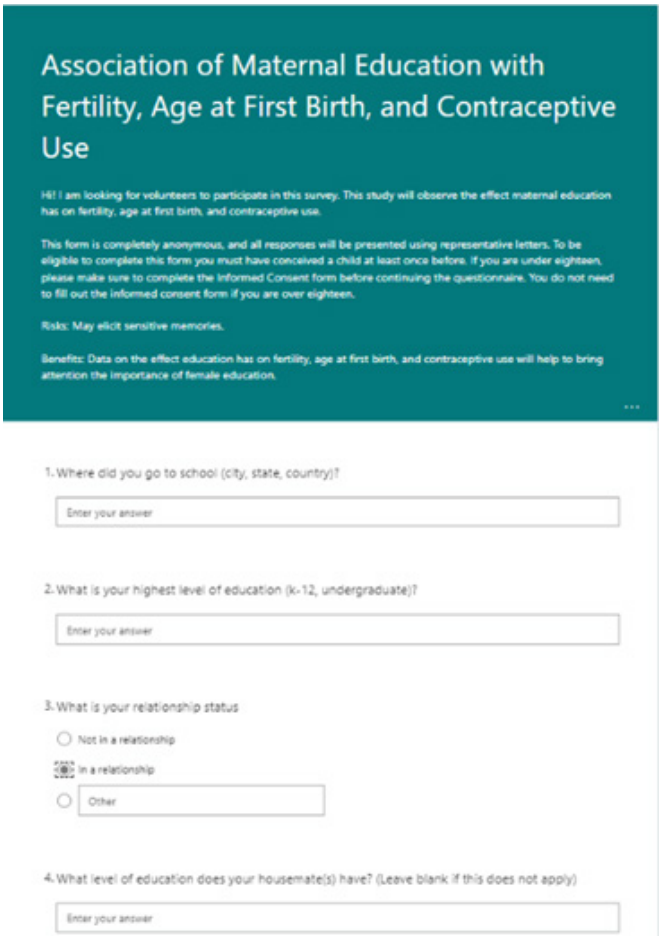


Figure 1. Flyer to be displayed outside prenatal classes and obstetrician-gynecologists offices. It includes the topic and QR code links to the informed consent form and the survey. The flyer also clearly outlines the instructions, goals of the study, and information on why the study is being performed.

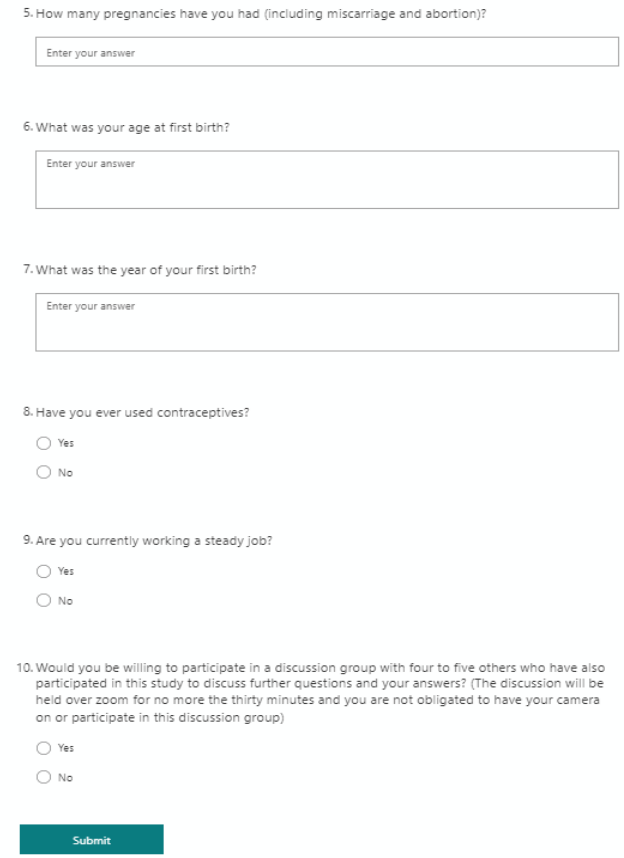


Figure 2. Questionnaire participants will respond to. There are a total of ten questions but participants who select that they are not in a relationship on question three will only have nine questions since they will not be asked to provide the level of education of their partner(s). It is made clear in the description that participants may leave the study at any time.



Simple Circle Logo Template Design With Initials - Logo (canva.com)Figure

Figure 3. Logo developed in Canva to represent this study that was distributed alongside the questionnaire on Instagram.

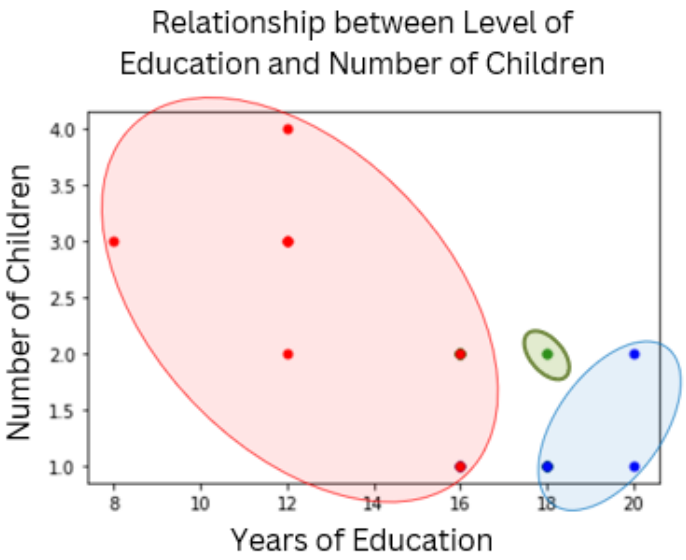


Figure 4. Unsupervised machine learning K-Means clustering algorithm graph on Relationship between Level of Education and Number of Children. K-Means clustering was used to cluster the data because there were multiple features (Years of education, number of children, age at first birth, contraceptive use).

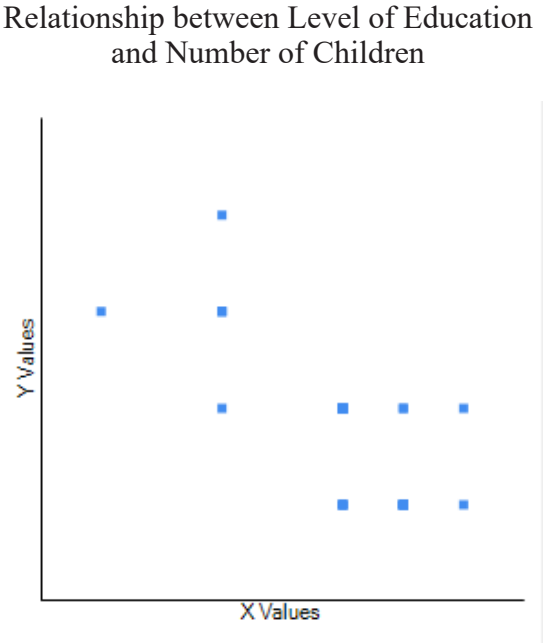


Figure 5. Graph representing the Relationship between Level of Education and Number of Children. X-axis represents years of education. Y-axis represents number of children. Only 11 data points despite 35 participants because of overlap.

Relationship between Level of Education and Age at First Birth

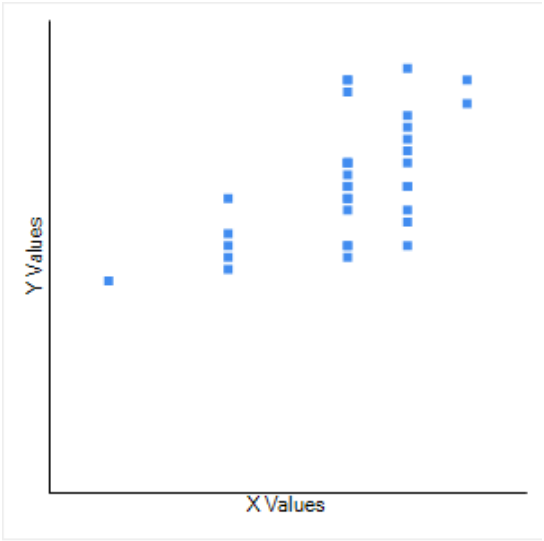


Figure 6. Graph on the Relationship between Level of Education and Age at First Birth. X-axis represents years of education. Y-axis represents age at first birth. Only 28 data points despite 35 participants because of overlap.

Code used to derive polychoric correlation coefficient (x=level of education, y=contraceptive use)

UntitledR.R

```
1  library(polycor)
2  x <- c(16, 16, 16, 16, 16, 18, 18, 18, 18, 16,
3
4  y <- c(10, 10, 10, 10, 10, 10, 10, 10, 20, 10,
5
6  #polychoric correlation for contraceptive use
7  ans <- polychor(x, y, std.err=TRUE)
8  print(ans)
```

Figure 7. Code from library(polycor) coded in language R in R studio that was used to derive the polychoric correlation coefficient. The x values are years of education, and the y values are contraceptive use (10=yes, 20=no).

Bibliography:

Aira Virtala, Sirpa Vilska, Teppo Huttunen & Kristina Kunttu (2011) Childbearing, the desire to have children, and awareness about the impact of age on female fertility among Finnish university students, The European Journal of Contraception & Reproductive Health Care, 16:2, 108-115, DOI: 10.3109/13625187.2011.553295

Cleland, J., Bernstein, S., Ezeh, A., Faundes, A., Glasier, A., & Innis, J. (2006, November 1). *Family planning: The unfinished agenda. The Lancet. Retrieved January 5, 2023, from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2806%2969480-4/fulltext*

Ferré, C., Callaghan, W., Olson, C., Sharma, A., & Barfield, W. (2017, August 17). Effects of maternal age and age-specific preterm birth rates on overall preterm birth rates - United States, 2007 and 2014. Centers for Disease Control and Prevention. Retrieved January 6, 2023, from https://www.cdc.gov/mmwr/volumes/65/wr/mm6543a1.htm#T1_down

Güneş, Pınar 2013. "The Impact of Female Education on Fertility: Evidence from Turkey." Grand Challenges Canada Economic Returns to Mitigating Early Life Risks Project Working Paper Series, 2013-1. https://repository.upenn.edu/gcc_economic_returns/1.

Ikamari, L. (2005). The effect of education on the timing of marriage in kenya - JSTOR. JSTOR. Retrieved January 20, 2023, from https://www.jstor.org/stable/26348131

Japaridze,I.(2019).Femalelaborforceparticipation and fertility differentials. *Journal of Demographic Economics, 85*(2), 123-164. doi:10.1017/dem.2019.1

Karlsen, S., Say, L., Souza, JP. *et al.* The relationship between maternal education and mortality among women giving birth in health care institutions: Analysis of the cross sectional WHO Global Survey on Maternal and Perinatal Health. *BMC Public Health* **11**, 606 (2011). https://doi.org/10.1186/1471-2458-11-606

Knop, B. (2021, October 28). *Among recent moms, more educated most likely to work. Census.gov. Retrieved April 4, 2023, from* https://www.census.gov/library/stories/2019/08/are-women-really-opting-out-of-work-after-they-have-babies.html

LeMoyne. (2020, January 19). Girls' education. UNICEF. Retrieved February 14, 2023, from https://www.unicef.org/education/girls-education.

Liliana Andriano, Christiaan W. S. Monden; The Causal Effect of Maternal Education on Child Mortality: Evidence From a Quasi-Experiment in Malawi and Uganda. *Demography* 1 October 2019; 56 (5): 1765–1790. doi: https://doi.org/10.1007/s13524-019-00812-3

Maeda, E. (2016). *Fertility knowledge and the timing of first childbearing: A cross-sectional study in Japan. Taylor & Francis. Retrieved October 20, 2022, from https://www.tandfonline.com/doi/full/10.1080/14647273.2016.1239033*

Marteleto, L., & Dondero, M. (2013). Maternal age at first birth and adolescent education in Brazil. Demographic Research. Retrieved November 15, 2022, from https://www.demographic-research.org/volumes/vol28/28/28-28.pdf

McCrary, J., & Royer, H. (2011, February 1). The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth. The American economic review. Retrieved January 5, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3073853/

Roustaei Z, Räisänen S, Gissler M, et al. Fertility rates and the postponement of first births: a descriptive study with Finnish population data. BMJ Open 2019;9:e026336. doi:10.11

Waiz, N. (2000). Role of education in the use of contraception - the lancet. THE LANCET. Retrieved April 21, 2023, from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(00)92037-3/fulltext

WHO. (2020). Family planning/contraception methods. World Health Organization. Retrieved February 1, 2023, from https://www.who.int/news-room/fact-sheets/detail/family-planning-contraception

Impact of Stigma on Indian Parental Detection of Generalized Anxiety Disorder in Adolescents

Kenisha Jaiswal

Tesla STEM High School

Abstract

For most teenagers their parents are the people they spend the most amount of their time with, meaning that the parents should be the first people to notice when something is wrong with their child, right? But how is a parent supposed to know when something is wrong with their child if mental illness is not acknowledged? This correlational project explores how parents' negative stigma against specifically anxiety disorders impacts their ability to see generalized anxiety disorder in their teenagers. The hypothesis set for this project was that there would be a negative correlation between the participants' Generalized Anxiety Stigma Scale test score and how well they are able to detect Generalized Anxiety Disorder in adolescents. A Generalized Anxiety Stigma Scale test (GASS) was used as a way to standardize the amount of stigma a parent has against anxiety disorders, the higher the GASS score was the more stigma a participant had. To measure how well parents were able to pick up on symptoms of Generalized Anxiety Disorder, each parent had to read six short stories about a teenager who possibly had an anxiety disorder and were asked to identify the warning signs for Generalized Anxiety Disorder. The results of this project were statistically significant with the value being -0.72 and the null hypothesis was rejected. The results also showed a negative correlation between stigma against anxiety disorders and the parents' ability to correctly identify the warning signs for Generalized Anxiety Disorder, meaning the hypothesis was correct.

Introduction

In 2020, in the midst of COVID-19, the CDC reported a 31% increase in teens mental health-related ER visits. The pandemic brought upon a surge of suicides and mental illness diagnoses, driven by sudden isolation, toxic home lives, and increased screen and social media usage. All of these factors are helping cause a rise in mental health services and it doesn't help that gaining access to these services isn't as easy as it sounds.

In previous research different barriers to gaining access to mental health treatments for Asian populations were studied. One major barrier that they explored was the stigma surrounding mental health. While stigma is not uncommon in different ethnic communities, one review stated, "Asian and Asian American communities have been found to have higher than average prejudices toward those with mental illness." (Chin et al., 2020). In many Asian countries, there has been little education given to older generations surrounding mental disorders and due to this many elders of the community try to avoid talking about it or believe in incorrect information surrounding these disorders. This creates a big problem because stigma makes life a lot harder for those with mental illnesses not only due to outside pressure but sometimes people with mental illness will believe these harmful stereotypes. The review mentioned earlier found that of the 18 studies they conducted, 15 of them showed that prejudice was the primary reason to avoid treatment. The main cause of stigma is lack of information surrounding mental illnesses that then gets passed down generations unless otherwise stopped. This causes people to stray away from professional help not only due to their own stigma but their communities. People tend to worry what other people will think about them and their family. Another problem that the lack of education, causes is that people also have a hard time noticing signs of mental illness in themselves.

This review only looked at adults, but they

failed to see how this stigma affected their children, who most often make up the one of the most vulnerable age groups to mental illnesses. Due to social media and the physical changes happening in their bodies teenagers experience an increase in the release of certain hormones that increase their anxiety levels. This makes them more susceptible to mood disorders like depression or bipolar disorder. This paired with the competitive nature of school and college admissions, means that teenagers are under high amounts of stress.

One major population that is overlooked are teenagers and their biggest barrier: their parents. For adolescents to gain access to mental health, their only link is their parents. Most adolescents lack the resources to afford therapy or are unable to take themselves to appointments causing them to be fully dependent on their parents'. This project explores how Indian parents' negative stigma against specifically anxiety disorders impacts their ability to see Generalized Anxiety Disorder (GAD) in their teenagers. The hypothesis set for this project is that Indian parents' stigma will be negatively correlated with their ability to detect Generalized Anxiety Disorder in teenagers.

Methodology

The first step to this correlational project was acquiring participants. So, to recruit participants a flyer was made through Canva, as seen in *Figure 1*, and was posted on different social media platforms like WhatsApp and Instagram. The flyer specified that the participants needed were Indian parents between the ages 40-60, who had teenagers. All participants were required to fill out an informed consent form to participate in this project that specified that they understood what they were required to do and could back out if they chose to. Once 15 participants were acquired, they were each assigned a random number, through a random number generator, to fill out the questionaries with. Both the questionnaires were done online to reduce the pressure of social desirability bias by the participants. The numbers made it possible to analyze each participant's answers and maintain anonymity, which again helped by reducing the chances of social desirability bias. The first questionnaire each participant took was the Generalized Anxiety Stigma Scale (GASS) test (Personal Stigma Subscale only) to assess how many harmful stereotypes surrounding anxiety they believed in.

Volunteers Needed Poster 1



Figure 1: Flyer searching for participants.

The GASS contained 10 statements about certain beliefs about anxiety disorders each highlighting a negative stereotype about anxiety disorders as seen in *Figure 2*.

Personal Anxiety Stigma Subscale

- An anxiety disorder is not a real medical illness.
- An anxiety disorder is a sign of personal weakness.
- People with an anxiety disorder could snap out of it if they wanted to.
- People with an anxiety disorder should be ashamed of themselves.
- People with an anxiety disorder do not make suitable employees.
- People with an anxiety disorder are unstable.
- People with an anxiety disorder are to blame for their problem.
- People with an anxiety disorder are just lazy.
- People with an anxiety disorder are a danger to others.
- People with an anxiety disorder are self-centred.

Figure 2: GASS Statements

It had a scale of one-to-four, from "I strongly disagree" to "I strongly agree" and the participant had to say how much they agreed with the different statements. The average score was taken out of the ten statements for each participant to be compared later. The range for possible scores were between ten and forty, with the lower scores meaning a lower level of anxiety stigma and a higher score meaning a higher level of anxiety stigma. Each participant then read six vignettes about an unknown teenager and tried to assess whether the teenager possessed Generalized Anxiety Disorder. Each vignette contained a short story describing a hypothetical teenager and their behavior.

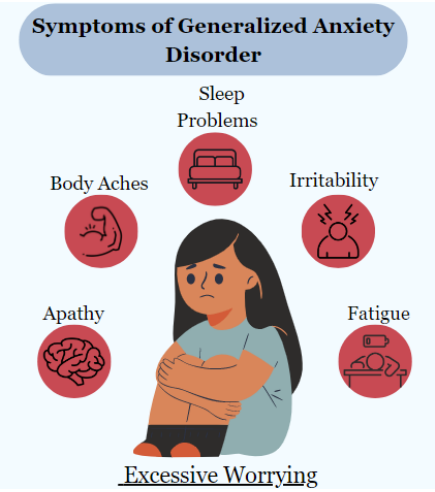


Figure 2: Symptoms of GAD

These vignettes were created with the help of the DSM 5's diagnostic criteria which is what a therapist would use to diagnose a patient with Generalized Anxiety Disorder (GAD). To be correctly diagnosed with GAD the patient needs to have excessive worrying more often than not for six months and needs to be showing three or more of the symptoms highlighted in *Figure 3* for six or more months. These two things were translated into the vignettes and an example of that can be seen in *Figure 4*. For each vignette the parents must try to identify one from the other and try to explain what things they picked up on in the vignette to make sure they weren't random yes/no guesses. The number of accurate answers from this test calculated how well the parents were able to notice the signs of GAD. The average GASS scores and the number of correct answers from the vignettes for each participant was then written down in *Table 1* as well as put on *Graph 1*.

Deepika is a freshman in high school. Her parents have noticed that her grades have been dropping due to Deepika not doing any of her homework for the past 3 months. Her parents have noticed that she has been a lot more tired these past 6 months since school started and have parents ask why she's up so late she always has replied heard her moving around late at night. Whenever her with some variation of "I am worried about something at school tomorrow".

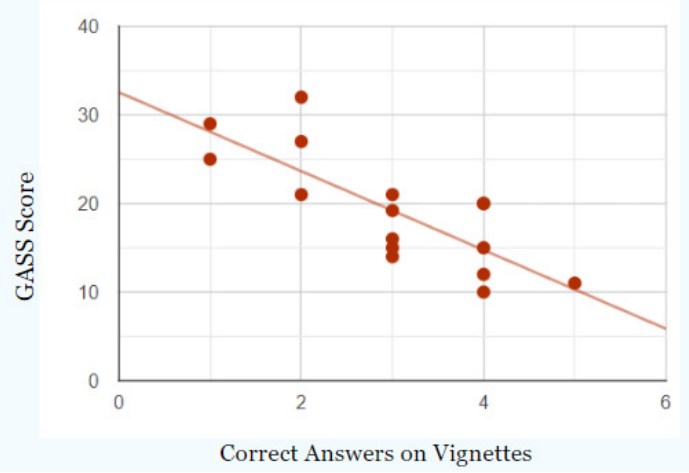
Figure 4: Sample Vignette

Conclusion

The original hypothesis set for this project was that there would be a negative correlation between parent's bias and their ability to correctly pick up on the warning signs for Generalized Anxiety Disorder. This was shown to be true when the calculated r^2 value was -0.72, proving a high negative correlation between the two variables. An r^2 value of -.072 shows a high statistical significance and that there is a relationship between the two variables. The scatterplot also shows a negative relationship with the GASS scores and the number of correct answers on the vignettes. So, the data as a whole shows the original hypothesis to be true but when analyzing the data by sexes this didn't hold true in the same way. The calculated average for men on the GASS was about 17 while for women it was 23. According to the previous data analysis men should have a much higher average for number of correct answers on the vignettes but they didn't. Men had an average of 3.3 while women had an average of 3.1. One reason this could be happening is that there is a stereotype in Indian communities that women care more about their family's public appearance and tend to gossip more so that could be a reason why they have a higher mean GASS score. Women are also better at gauging people's emotion and noticing changes so that may be a reason for the low difference in the number of correct vignette answers. But to truly know why the results came out that way, this project would need to be continued with a larger number of participants and with the Perceived Stigma Subscale of the GASS that measures what a person thinks other people's opinions on how the public see anxiety disorders.

One limitation to this project was its sample size. The project had only 15 participants, which is not enough to be able to completely generalize

an entire population but is a good baseline to start for future research. A small sample size could also mess with the accuracy of the r^2 value by making it higher than it would have been with a larger group of participants. So although with the current sample size the r^2 value does indicate that there most likely is a relationship between the variables, further research should be done with a larger of sample size of Indian parents. Another limitation was the number of vignettes. To improve this project for future research more vignettes should be added as that could help increase test validity. More vignettes would give the participants more time to think about their answers and their reasons as to why they wrote it a certain way and would give more data to analyze later in the project.



Graph 1: Scatter Plot

Next steps to further the depth of this project would be to recreate this with a younger generation and compare results. As time moves on, newer generations are getting more comfortable with the idea of mental illnesses and get help for it. So, one way to explore that is to conduct this study on Indian adults between the ages of 20-40. This is an age group of people about to have their first child or has recently just had their first child, which means that it is necessary that they start thinking about this. This generation also had more education in school compared to the previous about mental illnesses. Since this project did show an inverse relationship between a parent's level of stigma against anxiety disorders and their ability to pick up on the warning signs of Generalized Anxiety Disorder, it would be interesting to see how that relationship holds when people have grown up more educated about mental disorders.

Tables

Table 1

Participant Number	Average score from GASS	Number of correct answers from the vignettes out of six
1	4	12
2	4	15
3	1	25
4	2	32
5	3	15
6	2	21
7	1	29
8	4	20
9	3	21
10	4	20
11	4	10
12	3	14
13	5	11
14	3	16
15	2	27

Bibliography

Batterham, P. J., Griffiths, K. M., Barney, L. J., & Parsons, A. (2013). Predictors of generalized anxiety disorder stigma. *Psychiatry Research, 206(2-3), 282–286. https://doi.org/10.1016/j.psychres.2012.11.018*

Cervantes, R. C., Padilla, A. M., Napper, L. E., & Goldbach, J. T. (2013, September 8). Acculturation-Related Stress and Mental Health Outcomes Among Three Generations of Hispanic Adolescents. Sage Journals. Retrieved November 13, 2022, from https://journals.sagepub.com/doi/10.1177/0739986313500924

Griffiths, K.M., Batterham, P.J., Barney, L. et al. The generalised anxiety stigma scale (GASS): psychometric properties in a community sample. BMC Psychiatry 11, 184 (2011). https://nceph.anu.edu.au/research/tools-resources/generalised-anxiety-stigma-scale-gass

Khambaty, M., & Parikh, R. M. (2017). Cultural aspects of anxiety disorders in India. *Dialogues in Clinical Neuroscience, 19(2), 117–126.* https://doi.org/10.31887/dcns.2017.19.2/rparikh

Naito, T., Chin, J., Kim, T. U., Veera, S., Jeannette, M., & Lomiguen, C. M. (2020, November 12). *Further reduction in help-seeking behaviors amidst additional barriers to mental health treatment in Asian populations: A contemporary review. Cureus. Retrieved November 12, 2022, from* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7733772/

Manjuantha, N., Jayansankar, P., Suhas., Rao, G. N., Golpalkrishna, G., Varghese, M., Benegal, V., & NHMS National Collaborators Group (2022). Prevalence and its correlates of anxiety disorders from India's National Mental Health Survey 2016. *Indian Journal of Psychiatry, 65(2), 138-142.* https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_964_21

Identifying the Levels of Susceptibility of Law Students to Racial Bias

Soha Khan

Nikola Tesla STEM High School

Abstract

African Americans and Latinos constitute 29% of the population in the U.S., but make-up 57% of the prison population and 45% of people serving life sentences, representing a lack of disparate racial representation in the penal system. This project studied the relationship between the susceptibility to racial biases and the level of exposure to anti-bias trainings in current law schools in the U.S. The purpose of this project was to identify whether law students have lower susceptibility to racial biases compared to undergraduate students due to current anti-bias trainings in law schools. Six law students from ABA-rated law schools (independent variable) and six undergraduate students (control group) filled out a Racial Perspective Questionnaire and Harvard Race Implicit Association Test (Race IAT) and reported their responses (dependent variable) in a survey. Each student was tested for implicit racial bias, a measurement of unconscious attitudes and stereotypes depending on the type of race that affect understanding, and for level of susceptibility to racial biases, the percent of law student responses compared to undergraduate responses having a Race IAT score indicating moderate to strong preference. Results of the experiment presented a p-value of 0.136525 showing no statistical significance. The Racial Perspective Questionnaire responses conveyed a 0.9% difference in responses from law students and undergraduates, with both means in the range of being considerably susceptible. The initial hypothesis was incorrect as shown by 1% difference in susceptibility to racial biases for law students compared to undergraduates.

Introduction

This study seeks to investigate how susceptible law students are to implicit racial bias and how their biases compare to those of current judges. African Americans and Latinos constitute 29% of the population in the US, but they make up 57% of the prison population and 45% of the 200,000 people serving life and "virtual life" sentences (Sentencing Project 2018). Racial disparities exist within the justice system, as can be seen in a study of bail-setting in Connecticut, in which judges set bail amounts 25% higher for black defendants than for white defendants in similar situations (Rachlinski, Johnson, Wistrich & Guthrie, 3). Although judges are supposed to be fair, implicit bias, which is based on subtle stereotypical associations, is common among the rest of the population. Judges are susceptible to implicit biases, which can influence the rulings they make. However, developing and implementing training to reduce racial bias should be able to reduce its presence within the justice system.

Based on a report to the UN on Racial Disparities in the Justice System, research shows that the effects of implicit bias can be eliminated based on individual decision-making by training individuals to be behaviorally colorblind (The Sentencing Project, 14). This includes, but is not limited to, police officers, public defenders, prosecutors, judges, jury members, and parole boards. Having judges undergo implicit bias training can help them attain more fair judgments with increased awareness of possible cognitive bias, ignoring the race of a defendant or victim within court when making a ruling, and preventing racial bias from seeping into court decisions. An empirical study of judicial stereotypes tested the idea that negative implicit biases can manifest in federal and state judges against minorities. Results

from their testing, which utilized the Harvard Implicit Association Tests, indicated that each of the types of judges tested displayed negative implicit biases towards the minority groups tested for biases: Asians and Jews, with no statistical difference between the IAT score and the type of judge (Levinson, Bennett, & Hioki, 105). Despite the levels at which judges are placed within the judicial system, from state to federal courts, implicit biases can seep into the process even if the law is used as a basis for decision-making. Even more so, a study on judges moving towards unfavorable precedents, previous decided cases that give a basis to determine decisions in similar contexts or issues, when making decisions was carried out and demonstrated how judges of elite pedigree, adding in the idea of social influences, were more likely to vote towards unfavorable precedents versus judges coming from strong social-legal backgrounds who voted for favorable precedents (Wedeking 84). In this study, examples of such unfavorable precedents can be seen with a Latina judge favoring a Latinx defendant, implying that a decision is made based on background and experiences rather than the objective use of the law (Wedeking 69). While it is not uncommon for more explicit bias to be displayed, as in the example stated previously, removing implicit bias is much harder than removing explicit bias, which is visible and recognized. Law students, as future practitioners in law such as judges and lawyers, are not discriminated against when it comes to bias, either, until recently, when the American Bar Association (ABA) started an attempt to change that. As of February 2022, the American Bar Association mandated mandatory anti-bias trainings within law schools around the nation in an attempt to combat bias when the law is in practice. This study will be assessing the effectiveness of current anti-bias trainings within American Bar Association-approved law schools to decrease the presence and impact of bias within the justice system.

This study was conducted with the hypothesis that law students will have a lower susceptibility to racial bias compared to other undergraduate college students, due to the presence of current anti-bias trainings, indicating that the anti-bias trainings are effective.

Methodology

To understand the full extent to which anti-racial bias training impacts law students, the responses of the Harvard Race IAT and a Racial Perspective Questionnaire will be compared to those of undergraduate college students. This process can be seen in **Figure 2.1**. Undergraduate college students have not been exposed to such trainings routinely like the ABA has suggested law schools are required to carry out, making identification for impact on graduate law students easier to do. Participation was voluntary. If the participants have felt uncomfortable in any part of the survey process, they were able to reach out to the student researcher through an anonymous email to inform the researcher that they would be dropping out of the survey process at any time.

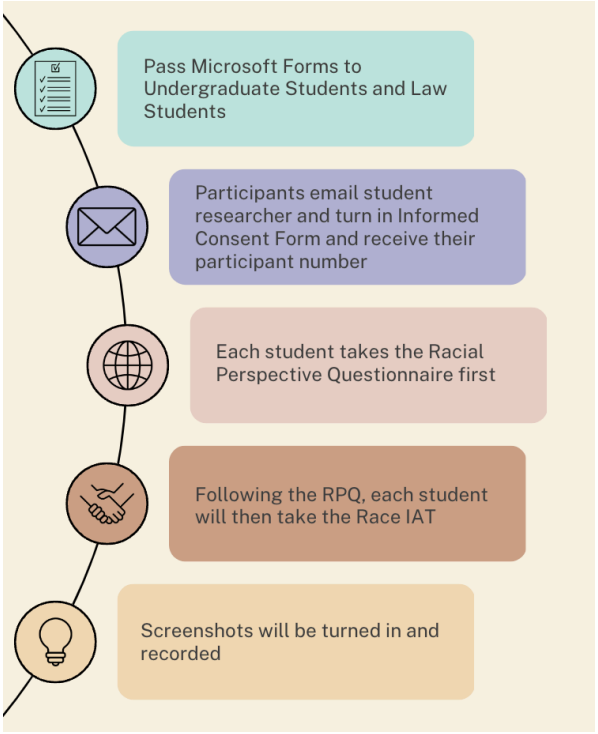


Figure 2.1: The order in which the research will be carried out

Participants were collected from the Seattle University School of Law and the Everett Community College by digitally distributing a poster, as seen in Figure 2.2, with a QR code and links to the Microsoft Forms. 6 undergraduate college students and 6 law students participated, and their responses were collected, totaling 12 participants. When using the Microsoft Forms, participants were first asked to email the student researcher through an anonymous email to be debriefed on the purpose of the research and what the survey process entails, to receive their participant number needed to fill out

the Microsoft Forms, and to turn in their completed Informed Consent Form given at the beginning of the Forms. Participant numbers were assigned using a random number generator to ensure a random sample. Emails between the student researcher and participants were kept minimal, and participants were asked to exclude any details regarding their name and other personal information when emailing the student researcher. The sampling process was kept to a single blind to maintain the anonymity of participants. The anonymous email and all emails, forms, and documents distributed between the student researcher and the participants were destroyed and deleted following the completion of the research.

The Racial Perspective Questionnaire was given to the participants first before asking them to take the Race IAT. This questionnaire was designed based on the questions given to the U.S. Supreme Court judicial nominees as of 2022 and generalized to apply to both law students and undergraduate college students. Microsoft Forms was utilized to make the questionnaire easily accessible to the students and simplify the process of filling out the form. The form was broken down into four sections: the first section was to contact the student researcher to be debriefed; the second section was to turn in their filled-out informed consent form as given in the Microsoft form; and the third section was to receive their participant number to differentiate Race IAT responses from questionnaire responses as they were turned in. The second section was to gain information on the demographics of the students, including race, ethnicity, identifying gender, whether they were undergraduate or law students, whether they were associated with a political party, exposure to racial bias prevention programs as seen **Tables 1.1-1.2**, and an additional question regarding their thoughts on pro bono work. Races of each participant were compared to their Race IAT response as seen in **Tables 1.3-1.4**.



Figure 2.2: Poster made using Canva to post and share across multiple media platforms

Table 1.1: Comparing Law Students' Race IAT Responses to Exposure to Racial Bias programs pulled from the Racial Perspective Questionnaire

Law Student Participant #	Race	Programs attended for addressing racial biases including year, purpose of, application	Race IAT Response
13	Black	Year: 2020 Purpose: Diversity programming requirement; discrimination and biases in general + impact on law career Application: renewed perspective + relevant to Minnesota (school location at time) with history of racial segregation	Moderate preference for European Americans
49	White	Western district of WA racial bias jury video; basics of implicit bias	No preference
8	White	Yearly title IX trainings Purpose: sexual harassment and assault topics involving racial bias as well Application: didn't find helpful	Moderate preference for African Americans
52	White	N/A	No preference
70	White	Year: 2022, unhelpful (not well run) Year: 2016 Purpose: Intrinsic bias based Application: Significant + self-awareness, still applies in daily life	No preference
76	White	Year: 2017 Purpose: police toning, intersections of life, other student experience Application: realize privileges; made sure to make friends with same values for POC, bring up topics regarding stigmas POC face in educational environment	Slight preference for African Americans
47	White	Year: N/A Purpose: bias programs in law + how to not let it affect clients Application: think about going forward and application to future processes	Strong preference for European Americans

Table 1.2: Comparing Undergraduate/High School Students' Race IAT Responses to Exposure to Racial Biases pulled form Racial Perspective Questionnaire

Undergrad/HS Participant #	Race	Programs attended for addressing racial biases including year, purpose of, application	Race IAT Response
18	East Asian, Black	Years: 2021-2022 Purpose: Microaggressions, for POC students to find solidarity Application: forming community, responding to racial bias	Moderate pref. for European Americans
7	Black	N/A	No preference
31	White	N/A	No preference
70	Black	Years: 2017-2021 Purpose: Different ways people can be racist and pin down to characteristic Application: exposure to other POC like participant	Slight pref. to European Americans
37	White, Middle Eastern	N/A	Moderate pref. for European Americans
9	East Asian	N/A	Strong pref. for European Americans
4	South Asian	N/A	Moderate pref. for European Americans

Table 1.3: Comparing Self-Assessment Responses of Each Law Student Participant's Race and their Race IAT Responses

Law Student – Participant #	East Asian	Latinx	South Asian	Pacific Islander	Black	White	Other	Race IAT Response
13	-	-	-	-	X	-	-	Moderate pref. for European Americans
49	-	-	-	-	-	X	-	No preference
8	-	-	-	-	-	X	-	Moderate pref. for African Americans
70	-	-	-	-	-	X	-	No preference
76	-	-	-	-	-	X	-	Slight preference for African Americans
47	-	-	-	-	-	X	-	Strong pref. for European Americans

Table 1.4: Comparing Self-Assessment Responses of Each Undergraduate/High School Student Participant's Race and their Race IAT Responses

Undergrad/HS – Participant #	East Asian	Latinx	South Asian	Pacific Islander	Black	White	Other	Race IAT Response
18	X	-	-	-	X	-	-	Moderate pref. for European Americans
7	-	-	-	-	X	-	-	No preference
31	-	-	-	-	-	X	-	No preference
70	-	-	-	-	X	-	-	Slight pref. to European Americans
37	-	-	-	-	-	X	X	Moderate pref. for European Americans
4	-	-	X	-	-	-	-	Moderate pref. for European Americans

The third section was a self-assessment section, with student responses recorded in **Tables 1.5–1.8** using a 3-point Likert scale and asking for bias for and against East Asian, South Asian, Black, White, Pacific Islander, Mexican, and Latin races using low, moderate, and high standard points to respond with

Table 1.5: Comparing Self-Assessment Responses of Each Law Student Participant Regarding Positive Bias Towards a Specific Race to Race IAT Responses

Law Student – Participant #	East Asian	Latinx	Mexican	South Asian	Pacific Islander	Black	White	Race IAT Response
13	Moderate	Moderate	Moderate	Moderate	Moderate	High	Moderate	Moderate pref. for European Americans
49	Neutral	Neutral	Neutral	Neutral	Moderate	Neutral	Moderate	No preference
8	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate	Neutral	Moderate pref. for African Americans
70	High	Moderate	High	High	Moderate	High	Neutral	No preference
76	Moderate	High	High	High	High	High	Neutral	Slight preference for African Americans
47	High	High	High	High	High	High	High	Strong pref. for European Americans

Table 1.6: Comparing Self-Assessment Responses of Each Law Student Participant Regarding Negative Bias Towards a Specific Race to Race IAT Responses

Law Student – Participant #	East Asian	Latinx	Mexican	South Asian	Pacific Islander	Black	White	Race IAT Response
13	Moderate	Moderate	Moderate	Moderate	Moderate	Neutral	Moderate	Moderate pref. for European Americans
49	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	No preference
8	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	High	Moderate pref. for African Americans
70	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate	No preference
76	Neutral	Moderate	Moderate	Neutral	Moderate	Moderate	Neutral	Slight preference for African Americans
47	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Strong pref. for European Americans

Table 1.7: Comparing Self-Assessment Responses of Each Undergraduate College Student Participants Regarding Positive Bias Towards a Specific Race to Race IAT Responses

Undergrad/HS – Participant #	East Asian	Latinx	Mexican	South Asian	Pacific Islander	Black	White	Race IAT Response
18	High	Moderate	Moderate	Moderate	High	Moderate	Neutral	Moderate pref. for European Americans
7	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	High	No preference
31	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	High	No preference
70	High	High	Moderate	High	High	Moderate	Moderate	Slight pref. to European Americans
37	High	High	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate pref. for European Americans
4	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate pref. for European Americans

Table 1.8: Comparing Self-Assessment Responses of Each Undergraduate College Student Participants Regarding Negative Bias Towards a Specific Race to Race IAT Responses

Undergrad/HS – Participant #	East Asian	Latinx	Mexican	South Asian	Pacific Islander	Black	White	Race IAT Response
18	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate	Moderate pref. for European Americans
7	Moderate	Neutral	Neutral	Neutral	Neutral	Moderate	High	No preference
31	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	No preference
70	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Slight pref. to European Americans
37	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate pref. for European Americans
4	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Moderate pref. for European Americans

Questions asked in this section also included level of familiarity with different races using unfamiliar, moderate, and knowledgeable as standard points to respond with, and regarding exposure to racial bias in personal, professional, and educational environments, using low, moderate, and high as standard points to respond with, with a follow-up fill-in-the-blank as to whether their experience was most in line with being a victim, instigator, or bystander. Section four asked for the participants to fill out the Harvard Race IAT and insert a screenshot of their Race IAT response, labeled with their participant number, in a link to a folder in an anonymous email made specifically for this research. All questionnaires were stored in a secure safe, and following the completion of the research, the documents were deleted and thoroughly destroyed.

The Harvard Implicit Association Test, as seen in **Figure 2.3**, measures attitudes and beliefs that people may have unknowingly or unwillingly and is periodically updated to "reflect our current understanding of the unconscious roots of thought and feeling." Participants are asked to specifically take the Race IAT, which is made to indicate the type of associations people make between European-Americans and African Americans. Both undergraduate and law students were asked to take the Race IAT following the Racial Perspective Questionnaire presented in the Microsoft Forms, their responses can be seen in **Tables 1.8** and **1.9**. The Race IAT and Racial Perspective Questionnaire student responses were compared in relation to their ethnicity and race and how their IAT responses either deviated from or conformed to their self-reported biases and other responses in the Racial Perspective Questionnaire, as seen in **Tables 1.2–1.5**.

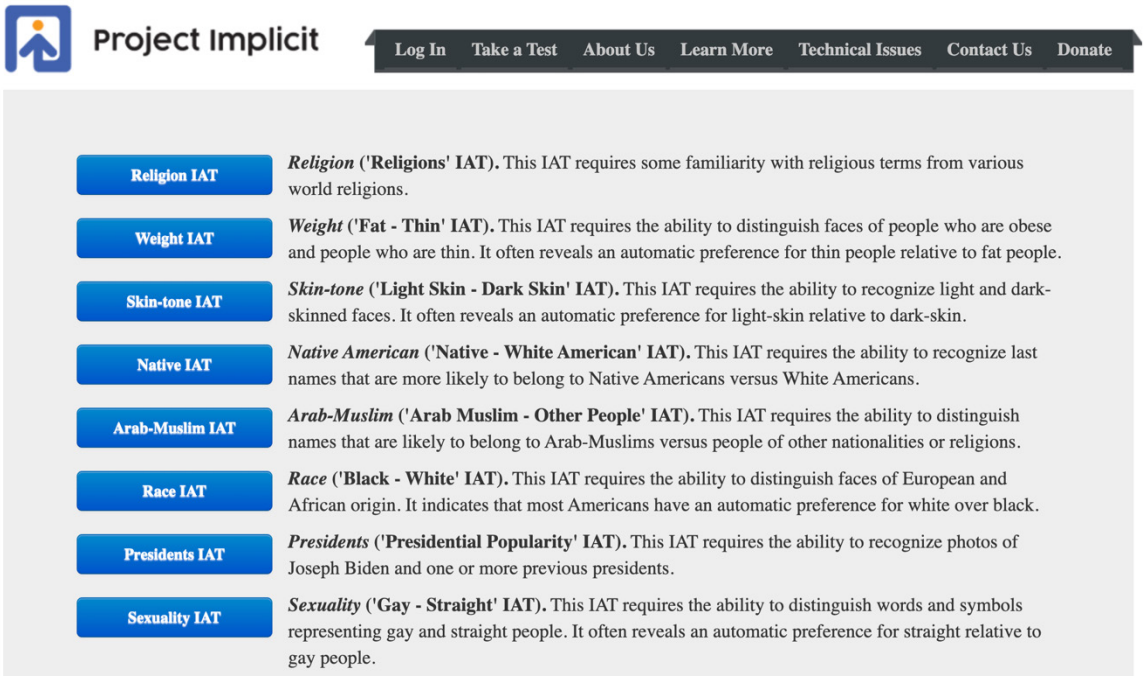


Figure 2.3: Harvard Implicit Association Test website to find the Race IAT

Table 1.8: Race IAT Responses and RPQ Scores of Each Law Student Participant

Law Student – Participant #	Level of Susceptibility (RPQ)	Race IAT Response
13	Considerable	Moderate pref. for European Americans
49	Slightly	No preference
8	Little to No	Moderate pref. for African Americans
70	Considerable	No preference
76	Moderate	Slight preference for African Americans
47	Moderate	Strong pref. for European Americans

Table 1.9: Race IAT Responses and RPQ Scores of Each Undergraduate/High School Student Participant

Undergrad/HS Student – Participant #	Level of Susceptibility (RPQ)	Race IAT Response
18	Considerable	Moderate pref. for European Americans
7	Considerable	No preference
31	Moderate	No preference
70	Considerable	Slight pref. to European Americans
37	Considerable	Moderate pref. for European Americans
4	Slightly	Moderate pref. for European Americans

Responses to both the Racial Perspective Questionnaire and the Race IAT were compiled and calculated based on a set score range. *Any responses that indicated racial bias were given fewer points, and any indication of lacking racial bias was given a greater number of points.*

For the Race IAT, responses indicating a strong preference were given 0 points, those indicating a moderate preference were given 1 point, those indicating a slight preference were given 2 points, and those indicating a neutral preference were given 3 points. This was kept the same regardless of any preference, if any, shown towards African Americans or European Americans.

The Racial Perspective Questionnaire was scored with a similar idea in mind. In the first section of the Racial Perspective Questionnaire, answers to the free response questions were given points based on actions taken by the student that indicate steps towards preventing racial bias. Responses that indicated attending a bias prevention program, whether the program was helpful and applied to the participant's current life, whether they planned on doing pro bono, and if they wanted to do pro-bono with racial bias-related intentions, were all given 1 point. Any responses in the free response section that did not indicate the former were given 0 points accordingly. For the second section of the Racial Perspective Questionnaire, a 3-point Likert scale was used for the following questions: how familiar the participant was with different races based on a scale of unfamiliar to knowledgeable, the average exposure to racial bias within personal, educational, or professional environments on a range from low to high, whether the participant was an instigator, bystander, or victim on average when encountering racial bias; how biased the participants were towards certain races, and how biased the participants were against certain races on a range from neutral to high.

The Racial Perspective Questionnaire score values were broken down into five categories: highly susceptible, moderately susceptible, considerably susceptible, slightly susceptible, and little to no susceptibility to racial bias. Score values ranging from 15 to 20 were set under the highly susceptible category, 21 to 25 under moderately susceptible, 26 to 30 under considerably susceptible, 31 to 35 under slightly susceptible, and 36 to 40 under little to no susceptibility. These categories were directly compared to the Race IAT responses of each participant to determine the difference in the level of racial bias assessed between the IAT and the Racial Perspective Questionnaire. It was found that 33% of undergrad students and 50% of law students indicated greater racial bias in the Race IAT compared to the IAT.

Data Analysis

A Chi-square test was utilized to identify statistical significance. To perform the test, the Race IAT score values of each participant were combined with the Racial Perspective Questionnaire scores, as seen in **Tables 1.9-2.0**. The total scores of each participant were ordered from least to greatest separately for each group of undergrads and law students. The observed values for this project were the law students' data values, and the expected values were the undergrad data values. Each of the expected values was subtracted from the observed values accordingly. Divide the resulting values by the expected value that was subtracted previously and add all of the divided values to get the critical number. Using the chi-squared chart in **Figure 2.4**, based on the degree of freedom of one, the data was found to be statistically insignificant with a p-value of 0.136525.

Table 1.9: Total Scores of Law Student Participants of the Race IAT and RPQ

Law Student Participant #	Racial Perspective Questionnaire Score	Level of Susceptibility (RPQ)	Race IAT Score	Total Score
13	26	Considerable	1	27
49	34	Slightly	3	37
8	38	Little to No	1	39
70	29	Considerable	3	32
76	23	Moderate	2	25
47	24	Moderate	0	24

Table 2.0: Total Scores of Undergraduate/High School Participants of the Race IAT and RPQ

Undergrad/HS Student Participant #	Racial Perspective Questionnaire Score	Level of Susceptibility (RPQ)	Race IAT Score	Total Score
18	30	Considerable	1	31
7	30	Considerable	3	33
31	21	Moderate	3	24
70	29	Considerable	2	31
37	29	Considerable	1	30
4	33	Slightly	1	34

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Figure 2.4: The Chi Squared Chart to identify statistical significance based off degrees of freedom

Conclusion

The initial hypothesis that law students would have a lower susceptibility compared to undergraduate students due to the presence of anti-bias training within law schools was incorrect, as indicated by the statistical insignificance of the p-value of 0.136525. The Race IAT scores conveyed a 1% difference between undergrads and law students, with IAT scores having a slight to strong preference for either race, as represented by the graph in Figure 2.5. About the same percentage of law students and undergraduate students had a slight to strong preference, revealing that law students are equally susceptible to racial bias as undergraduate students. Law school anti-bias trainings were shown to be ineffective as well, since 50% of the law students in the study indicated greater racial bias in the IAT than the Racial Perspective Questionnaire, compared to 33% of undergrads. As shown in Figure 2.6, the graph of the Racial Perspective Questionnaire ranges indicates that a larger number of undergrads indicate susceptibility to racial biases compared to Figure 2.5 of the Race IAT responses, conveying the 1% difference in preferential responses.

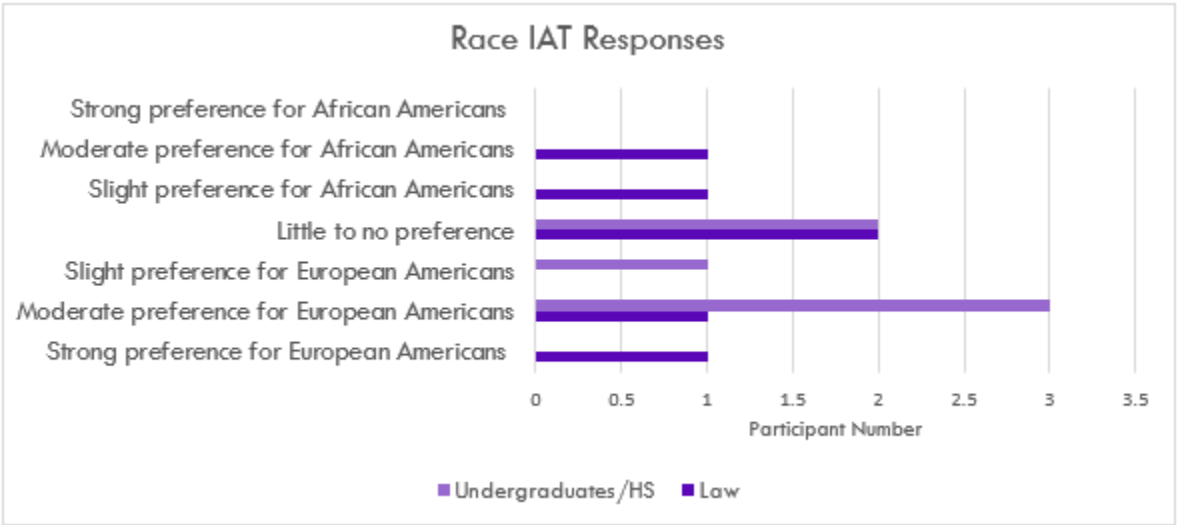


Figure 2.5: A graph of the Race IAT responses of both undergraduate and law students

This conveys that law schools do not have sufficient programs to prevent racial bias, as the American Bar Association claims that they do. In fact, prior to the randomized selection of participant responses used for the study, 54% of the total law student responses stated that they had not gone through any anti-bias trainings by the time they took the study's survey. Anti-bias trainings in law schools are ineffective and poorly run, as mentioned by some law student participants in the study. Trainings lack regulation or organization on a normal basis, and by leaving it up to the law schools to implement properly regulated trainings, they may run loosely without any proper planning to ensure orderly arrangements. Despite the American Bar Association's attempt to create a national standard for law schools in implementing anti-bias training programs, parts of the U.S. make it legally impossible to implement such trainings. Improvement and standardization of anti-bias trainings, where applicable, are essential. Introducing statistics for effective anti-bias trainings will help look at how effective anti-bias trainings are after improvements and illustrate how to make such trainings even better and, if necessary, implement more programs to prevent implicit racial bias. The Anti-Woke Law, which was implemented in multiple states, was created with the idea that applying such anti-racial bias trainings will only highlight the advantages that majority races have over minorities and will encourage further discrimination. With this idea, such trainings were banned in states where the law was enacted. Future research will examine the difference, if any at all, in susceptibility to racial bias of law students from states where the anti-woke law is implicated compared to law students in this study, taken from locations where the anti-woke law is not applied. Looking at how the anti-woke law makes a difference amongst law students in their implicit racial biases as future law practitioners can show the geographical impact on racial biases exhibited in law practices.

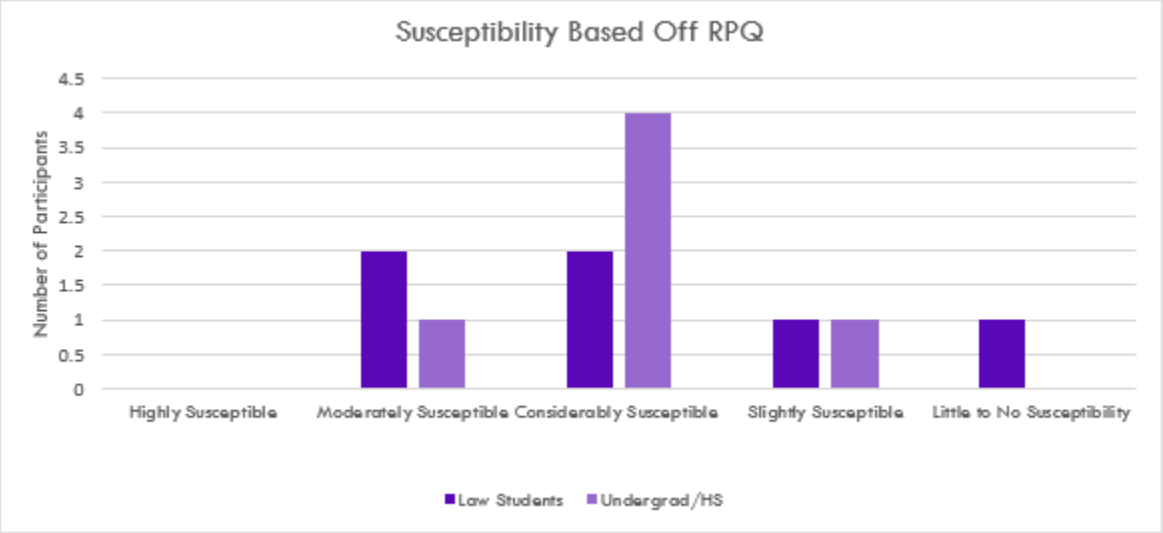


Figure 2.6: A graph of the Racial Perspective Questionnaire Responses in the 5 categorical ranges

Highly Susceptible: 15 – 20

Moderately Susceptible: 21 – 25

Considerably Susceptible: 26 – 30

Slightly Susceptible: 31 – 35

Little to No Susceptibility: 36 – 40

Bibliography

Justin D. Levinson, Mark W. Bennett, Koichi Hioki (2017). *Judging Implicit Bias: A National Empirical Study of Judicial Stereotypes,* Florida Law Review, Volume 69, Issue 1, 63-71.

Justin Wedeking (2012). *Why Do Policy-Motivated Justice Conform to Unfavorable Precedents? The Role of Social-Legal Backgrounds and Precedential Characteristics,* The Justice System Journal, Volume 33, Issue 1, 69-95.

Matthew Clair, Alix S. Winter (2016). *How Judges Think About Racial Disparities: Situational Decision-Making in the Criminal Justice System*, Criminology, Volume 0, Issue 0, 1-28. https://scholar.harvard.edu/files/matthewclair/files/clair_winter_how_judges_think_about_racial_disparities.pdf; Accessed October 3, 2022.

Maya Sen (2015). *Is Justice Really Blind? Race and Reversal in US Courts*, Journal of Legal Studies, Volume 44, Issue S1, 187-229. https://scholar.harvard.edu/files/msen/files/sen_reversal.pdf; Accessed November 25, 2022.

Nick Peterson (2016). *Examining the Sources of Racial Bias in Potentially Capital Cases. A Case Study of Police and Prosecutorial Discretion*, Volume 7, Issue 1, 7-34. https://journals.sagepub.com/doi/full/10.1177/2153368716645842; Accessed September 12, 2022.

Porche' A.W. Okafer, Patricia Y. Warren, Eric A. Stewart (2022). *Plea Bargaining and Trial Decisions in Context: An Examination of Disparities in Punishment*, Journal of Contemporary Criminal Justice, Volume 38, Issue 4, 432-455. https://journals.sagepub.com/doi/abs/10.1177/10439862221111013; Accessed November 14, 2022.

Tetsuya Matsubuyashi, & Rocha, R. R. (2012). *Racial Diversity and Public Policy in the States*, Political Research Quarterly, 65(3), 600–614. https://doi.org/10.1177/1065912911401418; Accessed November 27, 2022.

Evaluating the Effect of Beats Per Minute of Techno Music on Reaction Times to Auditory Stimuli in High School Students

Raneet Hans

Nikola Tesla STEM High School

ABSTRACT

Listening to music in the car is a common experience, but can it be made into a safety feature by using the tempo or genre of the music to influence reaction time and potentially reduce car accident fatalities? The goal of this study is to determine whether an increase in beats-per-minutes (BPM) of techno music will affect the listener's reaction time to an auditory stimulus. The hypothesis was that the increasing BPM would significantly shorten (speed up) the listener's reaction time to the auditory stimulus. Participants were given three versions of the same techno music track at varying tempos (independent variable), then asked to complete an online reaction time test (dependent variable) after listening to each track to see if their reaction times changed in comparison with their times from a track of white noise (control condition). Data was analyzed using a Wilcoxon signed-rank test (non-parametric paired t-test). Results show a statistically significant decrease in reaction time between the control condition and all three techno tracks, with the lowest p-value of 0.0000007 between the control condition and the 134 BPM track. There was also a statistically significant decrease in reaction time between the 94 BPM track and the 134 BPM track, which had a p-value of 0.0107. This indicates that the null hypothesis can be rejected, as the increasing tempo of the techno music did significantly shorten reaction time to the auditory stimulus.

INTRODUCTION

The aim of this study is to determine whether an increase in tempo of techno music will increase the listener's reaction time to an auditory stimulus. In 2020, the leading cause of death for US teenagers and young adults aged 15 to 24 was motor vehicle crashes (CDC, 2022). 61,983 teenagers died in motor vehicle accidents from 1999 to 2020. Many teenagers have very little experience driving, and since the death rate is so high, it is imperative that research be done to investigate ways to prevent such fatalities. Car accidents happen for a variety of reasons, among them the inability to react fast enough. One potential way to remedy this is to shorten reaction time. This would allow drivers to act quicker and potentially avoid motor crashes. Reaction time can be defined as the time elapsed between the original sensation of a stimulus and the motor output. Reaction time to an auditory stimulus is measured by playing some type of audio, and measuring the time between the audio playing and when the listener produces the designated motor reaction. Depending on the number of stimuli, reaction time is classified as either simple reaction time or choice reaction time. If there is only one stimulus, it is simple reaction time; if there is more than one stimulus, it is choice reaction time. Many factors influence simple reaction time, including arousal, gender, and age (Brebner, 1980). Listening to music also shortens reaction time to both visual and tactile stimuli (Anwar et al., 2019). More specifically, listening to techno music for a period of about 30 minutes has been identified to shorten visual reaction time (Meško et al., 2009). Alongside this, techno music increases heart rate, which is a notable marker of arousal (Gerra et al., 1998). Techno music is commonly defined as fast paced music with a tempo near or above 100 beats per

minute. Defining characteristics of techno music often include the use of synthesizers, sequencers, and digital audio workstations to produce the sounds in the pieces. Chords are played using arpeggios, or playing each note in a chord in quick succession, rather than all at once. Heavy bass, riffs using sixteenth notes, and themes of urgency and adrenaline are all markers of a techno song. Techno music is also often in common time (4/4 time signature), and the progression of the song comes mainly from changes in timbre rather than lyrical verse or chorus changes. In 4/4 time, beats per minute, a measure of tempo, comes from counting the number of quarter notes within one minute of a musical track. It has been shown that listening to a track with a fast and increasing tempo can also increase heart rate (Agrawal et al., 2013).

The connection between reaction time, heart rate, arousal, and fast paced music such as techno is evident. What remains to be investigated is the effect of techno music on reaction times to auditory stimuli. This project connects the link between tempo, techno music, and reaction time to investigate the effect of the beats per minute of techno music on reaction time to auditory stimuli. This experiment could pose as a preliminary study, and the results from this study could later be applied to improve reaction time via techno music and prevent fatal car crashes. Using the data from this investigation, a playlist could be accumulated on apps like Spotify or Apple Music, containing techno music similar to that used in the study. The playlist could be listened to in a car to keep reaction time alert and potentially reduce motor vehicle crashes, especially among teenagers, who commonly listen to this style of music. The experimental hypothesis for this study is that the increasing tempo of techno music will shorten the auditory reaction time of the listener.

METHODOLOGY

Sampling Procedure

A flyer (see Fig. 1) was printed and posted around various locations in the Redmond-Sammamish area, namely the hallways of Eastlake High School and Nikola Tesla STEM High School, but also in grocery stores such as Albertsons and Starbucks coffee shops in popular areas. The flyer was also advertised online on Reddit and Instagram using forums for regions in King County. It contained a few sentences describing the experiment as well as a QR code that linked to more detailed information and an informed consent form. Minors who participated were required to get parent/guardian consent through the informed consent form. The flyer also included the contact information of the researcher conducting the experimental procedure if applicants had any questions. The applicants filled out a Microsoft Form or a Google Form indicating their race, age, gender, handedness (the mouse being used is designed for right-handed people), and if they had any hearing disabilities or medications that could interfere with the auditory system or arousal levels. Race and gender were only collected to ensure a representative sample, and any applicant who indicated that they had hearing disabilities or interfering medications was not eligible for the study. Each applicant was then randomly assigned a number via a random number generator. Then, when choosing applicants to participate, a random number generator was used to generate 12 numbers, corresponding to 12 applicants. Those 12 applicants were contacted via the email they provided on the Microsoft or Google Form to schedule a time to come in and conduct the experiment.



Figure 1: Advertisement flyer. This flyer was printed and posted in public community locations such as the Redmond King County Library and the bulletin board in Perrigo Park. It was also posted inside Tesla STEM High School and Eastlake High School. The target audience is teenagers ages 14 to 18.

Experimental Procedure

A confederate conducted the entire experimental procedure to reduce experimenter bias. Upon arriving, the confederates brought the participants one at a time to the established experimenting room, where they completed the tests. All measurements were performed at Nikola Tesla STEM High School. The measurement apparatus included a Dell Latitude 3190 2in1 Pentium computer, a computer mouse (Microsoft Sculpt Comfort mouse, Bluetooth enabled, H3S-00003), and a pair of over-the-ear headphones (LORELEI X8 Over-Ear Wired Headphones). The measurements were conducted via the Auditory Reaction Test on cognitivefun.net (see Fig. 2 below). The website plays a monotone beep through the headphones at a random time and the participants must focus and react by clicking the left mouse button as fast as they can when the beep is played. The control data was collected from the same individuals as the experimental data, but under the control condition of white noise. The participants were separated into 3 equal and randomly generated groups of 4 participants each and listened to the 3 versions of the track along with one period of measurements after white noise to serve as a control in different orders. The first group, referred to as the Increasing group, listened to the tracks in an order of increasing tempo. The second group, referred to as the Decreasing group, listened to the tracks in an order of decreasing tempo. The third group, referred to as the Random group, listened to the tracks in an order unrelated to tempo. This grouping allowed for statistical analysis to later reveal if participants' reaction times increased solely due to practice and not the tempo itself. Groups were separated and measured as follows:

Increasing group: 1-2-3-4

Decreasing group: 4-3-2-1

Random group: 2-4-1-3

Note that 1 means white noise, 2 means the original track at 94 BPM, 3 means the sped-up track at 114 BPM, and 4 means the sped-up track at 134 BPM. Increasing group participants will be used as the example for this procedure as all participants completed the same tests, just in different orders.

.Before experimental testing began, each participant in the Increasing group listened to white noise for 5 minutes. After that, they completed 5 trials of the auditory reaction time measurement. Times were recorded in milliseconds (see Table 2.1 below). After the control experiment, the participants were given sterilized over-the-ear headphones and listened to a loop of *Wasted (nightcore)* by Murkish and Huken at 60 dB for 5 minutes. *Wasted* was chosen because it has no lyrics, holds a consistent and predictable tempo (94 BPM), and displays many of the main characteristics of techno music that were mentioned in the introduction.. Immediately after the 5-minute period ended, the music was shut off and the participants completed 5 more reaction time test measurements (Table 2.2). Following the tests was a 5-minute period of silence to allow any of the effects the music may have had on the participants, such as their heart rate, to settle. After the silent period, the participants listened to a sped-up version of *Wasted* (114 BPM) for another 5 minutes and performed another five trials of measurements with the reaction test afterwards (Table 2.3). .Another 5-minute silent period followed, and then the participants listened to the third version of *Wasted*, sped up to 134 BPM, and completed five final trials of measurements with the reaction test (Table 2.4). The order the tracks are presented varied by group (see Table 1), but there was a 5-minute period of silence between the tracks for all groups. The pattern of the track being played and the measurement completed was repeated in the order of tracks assigned to the participant's group. See Fig. 3 below for a diagram of the time schedule for the Increasing group. The edited versions of *Wasted* were sped up via the audio editing program Audacity.

Figure 2

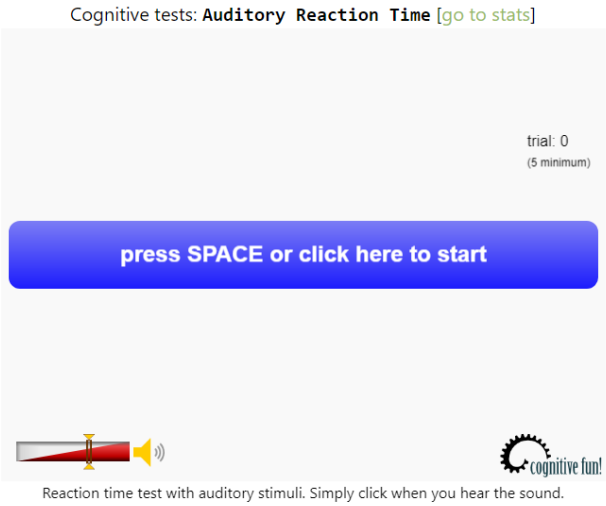


Figure 2: Screenshot of auditory reaction time test on cognitivefun.net. Upon clicking the start button, a monotone beep is played at a random time within a few seconds of starting. The time between the beep and the mouse click is recorded, and after 5 trials of this test, the values are displayed on the screen along with the mean score.

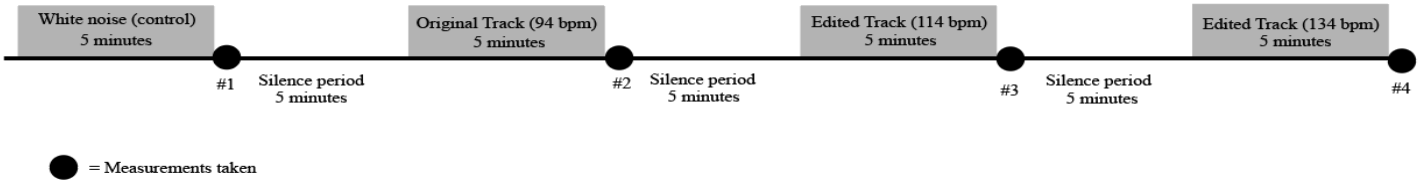


Figure 3: Increasing group experiment timeline. For the Decreasing group, tracks were played in the opposite order, starting with the 134 BPM track and ending with the white noise. The Random group listened to the original track first, then the 134 BPM track, then the white noise, and ended with the 114 BPM track.

RESULTS

Table 1 – Group Participants and Track Order

Group	Order of tracks/measurements	Participant Numbers of group members
Increasing	1,2,3,4	1, 4, 7, 10
Decreasing	4,3,2,1	2, 5, 8, 11
Random	2,4,1,3	3, 6, 9, 12

Table 2.1 – Measurement #1 (0 BPM/white noise)

Participant Number	Trial 1 (ms)	Trial 2 (ms)	Trial 3 (ms)	Trial 4 (ms)	Trial 5 (ms)	Mean
1	305	359	371	451	372	371.6
2	543	357	324	373	358	391
3	495	339	328	435	386	396.6
4	371	350	341	384	321	353.4

5	358	364	366	326	369	356.6
6	313	333	369	468	333	363.2
7	433	270	341	361	245	330
8	327	368	386	369	382	366.4
9	353	377	366	327	345	353.6
10	303	288	305	302	314	302.4
11	309	289	477	357	353	357
12	409	388	417	312	389	383

Table 2.2 – Measurement #2 (94 BPM)

Participant Number	Trial 1 (ms)	Trial 2 (ms)	Trial 3 (ms)	Trial 4 (ms)	Trial 5 (ms)	Mean
1	378	384	370	294	374	360
2	321	360	314	388	372	351
3	400	278	370	291	297	327.2
4	308	352	285	275	314	306.8
5	359	336	304	353	304	331.2
6	310	304	345	383	387	345.8
7	345	263	419	386	400	362.6
8	405	328	292	332	340	339.4
9	317	388	375	315	321	343.2
10	325	399	383	304	252	332.6
11	333	468	304	321	344	354
12	359	394	302	350	273	335.6

Table 2.3 – Measurement #3 (114 BPM)

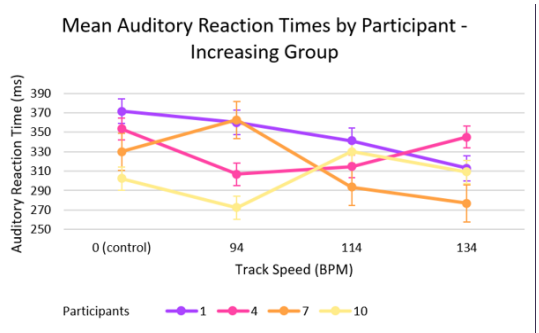
Participant Number	Trial 1 (ms)	Trial 2 (ms)	Trial 3 (ms)	Trial 4 (ms)	Trial 5 (ms)	Mean
1	372	360	314	388	272	341.2
2	393	278	360	375	254	332
3	349	287	232	351	341	312
4	323	324	311	358	258	314.8
5	393	311	279	352	388	344.6
6	324	327	297	321	354	324.6
7	261	273	294	258	382	293.6
8	278	342	354	391	273	327.6
9	333	258	320	339	279	305.8
10	323	341	386	230	370	330
11	292	333	235	334	413	321.4
12	345	293	305	276	277	299.2

Table 2.4 – Measurement #4 (134 BPM)

Participant Number	Trial 1 (ms)	Trial 2 (ms)	Trial 3 (ms)	Trial 4 (ms)	Trial 5 (ms)	Mean
1	268	257	312	359	369	313
2	297	329	307	223	334	298
3	223	225	368	381	347	308.8
4	398	354	335	323	315	345
5	331	297	393	364	342	345.4
6	324	358	337	252	266	307.4
7	227	295	267	320	276	277
8	357	224	308	294	373	311.2
9	352	234	274	304	275	287.8
10	326	303	315	279	323	309.2
11	268	368	349	345	361	338.2
12	302	320	301	292	273	297.6

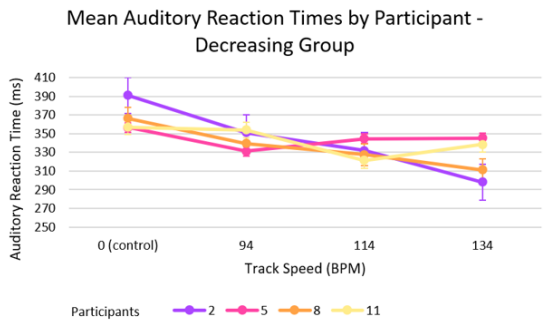
CONCLUSION

The experimental hypothesis for this study is that the increasing tempo of techno music will shorten the auditory reaction time of the listener. This hypothesis was proven to be partially correct. As figures 4, 5, 6, and 7 will show, the mean reaction times of participants in the Increasing, Decreasing, and Random groups all had general downward trendlines, indicating that the reaction times for participants, regardless of order of tracks, shortened as the tempo of the music increased.



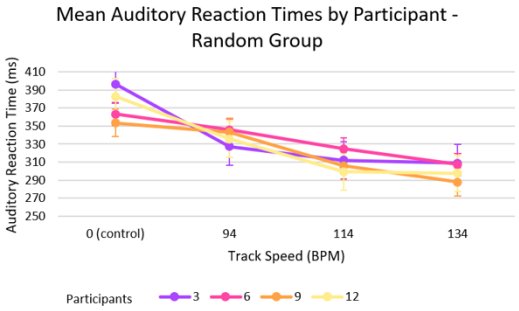
Track Speed (BPM)

Participant	0 (control)	94	114	134
1	371.6	360	341.2	313
4	353.4	306.8	314.8	345
7	330	362.6	293.6	277
10	302.4	272.6	330	309.2



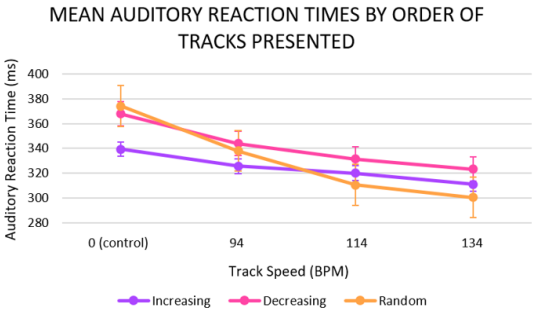
Track Speed (BPM)

Participant	0 (control)	94	114	134
2	391	351	332	298
5	356.6	331.2	344.6	345.4
8	366.4	339.4	327.6	311.2
11	357	354	321.4	338.2



Track Speed (BPM)

Participant	0 (control)	94	114	134
3	396.6	327.2	312	308.8
6	363.2	345.8	324.6	307.4
9	353.6	343.2	305.8	287.8
12	383	335.6	299.2	297.6



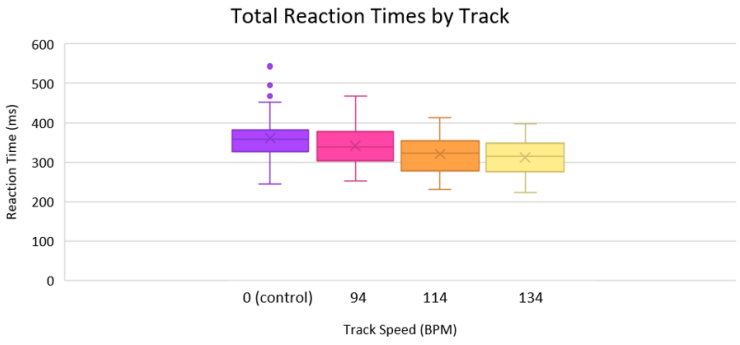
Track Speed (BPM)

	0	94	114	134
Increasing	339.35	325.5	319.9	311.05
Decreasing	367.75	343.9	331.4	323.2
Random	374.1	337.95	310.4	300.4

Above: Figures 4, 5, 6, and 7. Graphical and table representations of each participant per group and their mean track speeds. Figure 7 represents the overall mean values of each group for each track speed.

As figure 4 above shows, there were some outliers. Participants 7 and 10 each had varying mean reaction times. Participant 7 had a longer reaction time at the 94 BPM track than any of the others, and participant 10 had a much shorter reaction time at the 94 BPM track, which then increased at the 114 BPM track. However, when all the values per track per group were consolidated into overall means, the trendlines are still negative, indicating a shortening of reaction time as tempos increased.

A statistical evaluation was conducted after the means were calculated. Using a Wilcoxon Signed Rank test, a version of a non-parametric paired t-test, p-values for significance between reaction times per track were generated. These values are represented below in figure 8.



Statistical comparisons: using a non-parametric paired t-test (Wilcoxon Signed rank test)

condition	P-value	
control vs 94	**0.031**	*
control vs 114	**3.24E-05**	****
control vs 134	**6.98E-07**	****
94 vs 114	0.0595	ns
94 vs 134	**0.0107**	*
114 vs 134	0.2672	ns

Figure 8: Table and graphical representation of the p-values generated for each track comparison.

As Figure 8 shows, there were only 4 significant values out of the 6 values tested. The comparison from the control track to the 94 BPM track, with a p-value of 0.031, indicated a significant decrease in reaction time between the control and the 94 BPM track. Other significant decreases are in the control track versus the 114 BPM track (p-value of 3.24E-05), the control track versus the 134 BPM track (p-value of 6.98E-07) and the 94 BPM track versus the 134 BPM track (p-value of 0.0107). The comparisons of the 94 BPM track to the 114 BPM track and the 114 BPM track to the 134 BPM track were not significant. The only significant decreases were the control track compared to any of the music tracks, indicating that techno music, no matter the tempo, shortens reaction time, and the 94 BPM to 134 BPM track, indicating that there is a possible threshold between tempos where the reaction time shortens significantly. If the listeners' reaction time does not change significantly at a track speed change

of 20 BPM, but the listeners' reaction time changes significantly when the track speed change is 40 BPM, then it is possible that the amount of change in speed needed is anywhere between 20 BPM and 40 BPM. This is cause for future research to further optimize techno tracks to provide the maximum change in reaction time with minimal change to the tracks. Other future research could analyze different genres to see if the change in reaction time is unique to techno music, as prior research suggests, or if it is solely related to the BPM of the track. A playlist could also be created that combines various techno songs of fast tempos to be played in a car to potentially increase alertness and reaction time on the road and help to reduce fatal car crashes.

BIBLIOGRAPHY

Agrawal, A., Makhijani, N. & Valentini, P. (2013). The effect of music on heart rate. Journal of Emerging Investigators, 1.

Anwar U, Fazal A, Mirza F, Ahmed S. Impact of music on reaction time, attention, short term memory and verbal fluency: A gender-based study. APP. 2019; 6(1):1-7.

Bishop, D. T., Karageorghis, C. I., & Kinrade, N. P. (2009). Effects of musically-induced emotions on choice reaction time performance. The Sport Psychologist, 23(1), 59-76.

Brebner, J. T. (1980). Reaction time in personality theory. In A. T. Welford (Ed.), Reaction Times New York: Academic Press, 309–320.

Centers for Disease Control and Prevention. (2022, October 5). Leading causes of death and injury - pdfs|injury Center|CDC. Centers for Disease Control and Prevention. Retrieved January 15, 2023, from https://www.cdc.gov/injury/wisqars/LeadingCauses.html

Gerra, G., Zaimovic, A., Franchini, D., Palladino, M., Giucastro, G., Reali, N., ... & Brambilla, F. (1998). Neuroendocrine responses of healthy volunteers to techno-music': Relationships with personality traits and emotional state. International journal of psychophysiology, 28(1), 99-111.

Meško, M., Strojnik, V., Videmšek, M., & Karpljuk, D. (2009). The effect of listening to techno music on reaction times to visual stimuli. Acta Gymnica, 39(1), 67-73.

Pancar, Z., Özdal, M., Pancar, S., & Biçer, M. (2016). Investigation of visual and auditory simple reaction time of 11-18 aged youth. European Journal of Physical Education and Sport Science.

Red Light for Sleep Quality and Performance in Gymnasts

Sophia Callo

Tesla STEM High School

Abstract

Optimizing your sleep can increase the many mental and physical benefits of it, and something researchers have found that may help with this is red light. The purpose of this study was to determine if there is a correlation between the use of red light and changes in sleep quality and performance in gymnastics. The hypothesis was that there would be a strong correlation between use of red light and improved sleep quality and athletic performance. In this study, participants spent the hour before going to sleep in red light before a gymnastics competition. They took the Pittsburgh Sleep Quality Index questionnaire before and after red light use. After using red light, participant's all-around score in gymnastics increased by an average of 1.3167 points, and their Global Pittsburgh Sleep Quality Index score decreased by an average of 3.6667 points (lower scores indicate better sleep quality). The p-value for the improvements in the global Pittsburgh Sleep Quality Index score was 0.0697, and the most statistically significant component of the Pittsburgh Sleep Quality Index was daytime dysfunction (p= 0.0474). The p-value for improvements in the all-around gymnastics score was 0.0572, and the most statistically significant event was bars (p= 0.0224). The hypothesis was partly correct because every participant had improvements in their Pittsburgh Sleep Quality Index scores and gymnastics scores, though not all of these improvements were statistically significant.

Introduction

The purpose of this study is to investigate the correlation between the use of constant red light in the hour before sleep and changes in sleep quality and performance in gymnasts. Good sleep quality and exercise go hand in hand. Exercise has repeatedly been found to improve sleep quality, which combined with sleep quantity is necessary for optimal athletic performance and recovery. Feelings of daytime sleepiness are a prevalent problem and an indicator of poor sleep quality, and this study looks into a possible solution to that. Red light bulbs are easily accessible, inexpensive, and could possibly decrease sleepiness during the day without requiring more time asleep, by improving sleep quality. This study will also be looking at a possible correlation between use of red light before sleep and performance in gymnasts, because gymnastics is a physically and cognitively demanding sport. Because of this, results could give insights into the physical and cognitive effects of sleep quality.

According to the National Institutes of Health, chronic sleep problems are experienced by about 70 million Americans. This is a problem because of the essential role that sleep plays in health, fitness, cognitive function, mood, and the general functioning of the human body. Although the quantity of sleep is very important, so is the quality. For example, someone who has good quality sleep will feel more rested than someone with poor sleep quality, even if they both slept for the same amount of time. More than just daytime sleepiness, poor quality sleep is associated with frequent waking, taking longer to fall asleep, difficulty concentrating, weight gain, and more feelings of stress, anger, and emotional exhaustion (Suni 2022). Though what

constitutes sleep quality and how to define it has yet to be clearly understood, the effects and solutions have begun to be researched.

Since people with poor sleep quality often experience frequent waking during the night, they experience less deep sleep (stage 3) and REM as waking up disturbs their sleep cycle. Stage 3 is the stage in which you experience the deepest sleep. This stage is thought to be vital for good quality sleep, bodily processes, and cognitive functions such as memory and creativity. During REM sleep, brain activity increases. This stage plays an important role in learning, emotional processing, and brain development. Deep sleep and REM are needed in order to have restful and rejuvenating sleep, in addition to their effects on cognitive functioning. If someone is not experiencing these stages of sleep in high enough quantities, they will not feel as rested, and cognitive and physical performance will likely decline.

Multiple studies have found red light to increase melatonin production. One study found an increase in sleep quality, serum melatonin levels, and endurance performance in response to using red light therapy, though only the increase in serum melatonin was statistically significant (Zhao, 2012). Studies finding red light to improve sleep quality is likely due to the intrinsically photo-sensitive retinal ganglion cells (iPRGCs) in the eye that contain a photopigment called melanopsin. Melanopsin, when activated, sends neural signals to the suprachiasmatic nucleus (SCN), Edinger-Wesphal Nucleus (EWN), and the pineal gland. The SCN controls the central body clock and circadian rhythms, the EWN controls pupillary constriction in response to light, and the pineal gland produces melatonin. The sensitivity curve shows that melanopsin is most sensitive to shorter wavelength light such as blue light. When melanopsin is activated by these shorter wavelength lights, the neural signals to the SCN tell the body that it is daytime, signals to the EWN constricts the pupils, and signals to the pineal gland to suppress melatonin production. Red light has a longer wavelength though, and the wavelengths of red light are at the limit of the melanopsin sensitivity curve. This means that melanopsin will not be activated as much when exposed to red light. Therefore the body will not be told that it is daytime, the pupils will not constrict as much, and there will be less melatonin suppression. This makes red light ideal for use before sleep or when it is dark outside for maintaining regular circadian rhythms and sleep quality, while still being able to have some light so you can go about your regular tasks. However, many of the studies conducted pertaining to red light and sleep have used high intensity red light therapy or long exposure to red light in regular hours of sleep. This study focuses on how effective red lightbulbs are when used before sleep as a more accessible solution than the expensive light therapy tools used in previous studies.

A study by the National Sleep Foundation found that 90% of Americans use devices like phones and televisions in the hour before sleep (Gradisar 2013). These devices expose you to shorter wavelengths, which as discussed before, will negatively affect circadian rhythms and ability to fall asleep. Use of room light before sleep is also a common practice. One study found that exposure to room light delayed melatonin onset and shortened melatonin duration (Gooley 2011). This suggests that chronic exposure to room light before bed can make it harder to fall asleep and put your body into a constant biological summer. So, replacing room light before bed with red light has the potential to prevent these effects so you can experience better quality sleep.

Physical performance relies on feeling rested and energized, but often also relies on cognition such as your ability to focus, especially in gymnastics. One study found that CrossFit athletes with good sleep quality improved their performance in 6 categories of CrossFit exercises faster than those with poor sleep quality, especially in cognitively demanding and technical categories (Klier 2021). These results suggest a correlation between sleep quality and physical and mental performance. Gymnastics is a sport that is both physically and cognitively demanding, so based on prior research, increased sleep quality should improve performance. The previous research around red light, sleep quality, and athletic performance informed the hypothesis of this study that there will be a strong correlation indicating that red light leads

to improved sleep quality and athletic performance.

Methodology

Sampling Procedure

Participants recruited included female gymnasts who were currently competing in levels 5 to 10 in the USAG Junior Olympics program, or levels Gold to Diamond in the USAG Excel program (n=3).

Participants were gathered by sending emails or using a flyer that was distributed to gymnastics gyms in Washington State. The flyer directed them to a Microsoft Form where they provided their email address to receive more information. Then they were sent another Microsoft Form to fill out about their general information including age, weight, height, etc. They were also asked if they were currently taking any medication that could affect their sleep quality, if they had a sleep disorder, and what dates they would be competing through February. Any participants who indicated that they were taking medication that would affect their sleep or had a sleep disorder were excluded from the study. Participants who did not have a competition during the time frame of when the study was taking place were also excluded.

Each participant was assigned a random number to use so their surveys could be linked together for data analysis, but not linked to the individual. They also filled out an informed consent form and were informed of the purpose of the study. Since participants were under 18, their parents had to sign the informed consent form as well. Participation in this study was voluntary, and participants could drop out at any time.

Correlational Procedure

Each participant received a packet with details about the study, instructions, and a list of stretches with pictures and descriptions.

First, participants completed the Pittsburgh Sleep Quality Index questionnaire and reported their scores from their most recent gymnastics competition in a Microsoft Form to establish their sleep quality and gymnastics scores while using room light before sleep. The Pittsburgh Sleep Quality Index (PSQI) questionnaire is a self-reported questionnaire that assesses sleep quality based on seven components (subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, use of sleep medication, and daytime dysfunction) and assigns a number to the sleep quality, with higher scores signifying worse sleep quality. Scores were received during the competition, based off the participants' performance on each event, and the scores of each event were added together for the all-around score. Scores for each event are out of 10 points and were determined by 2 judges who completed the USA Gymnastics Judging Accreditation Program.

Before their next competition, participants were instructed to spend the hour before going to sleep in a room lit with red light by a red light bulb that they were provided. During this hour, they completed the stretches in the packet that they were given. They were also asked to ensure no other light sources were present in the room (alarm clocks, light coming from under the door, etc.) and use of electronic screens was not allowed. Participants were instructed to sleep in darkness.

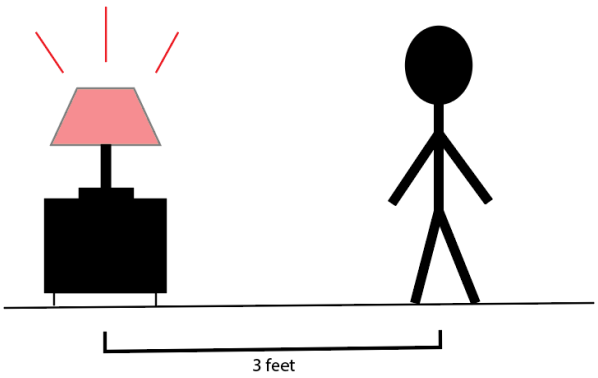


Figure 1. Participants will put their regular or red-light bulb in a lamp and will do their stretches 3 feet away from the lamp.

After the competition, participants were given another Microsoft Form where they filled out the PSQI questionnaire and reported their scores from the competition. Then the data was put into Table 1 and Table 2.

Table 1. Results from the PSQI Questionnaire with and without red light use for each participant.

Component of the Pittsburgh Sleep Quality Index	Without red light	With red light	Difference	P-value
Subjective sleep quality	1.3334	0.3334	1	0.1012
Sleep latency	0.6667	0	0.667	0.1161
Sleep duration	0.6667	0.3334	0.3333	0.6779
Sleep efficiency	0	0	0	n/a
Sleep disturbance	1	0.6667	0.3333	0.3739
Use of sleep medication	0	0	0	n/a
Daytime dysfunction	2.6667	1.3334	1.3333	0.0474
Global PSQI score	6.3334	2.6667	3.6667	0.0697

Table 2. Scores received in the gymnastics competition without red light use and with red light use for each participant.

Average Event Score	Without red light	With red light	Difference	P-value
Vault (out of 10)	8.4167	8.5833	0.1666	0.6263
Bars (out of 10)	8.3333	8.7333	0.4	0.0224
Beam (out of 10)	8.2917	8.875	0.5833	0.0539
Floor (out of 10)	9.2167	9.3833	0.1666	0.2953
All Around (out of 40)	34.2583	35.6417	1.3834	0.0572

Results and Conclusion

The hypothesis that there would be a strong correlation between use of red light before sleep and improved sleep quality and athletic performance in gymnasts was partly correct, as all participants had improvements in their PSQI scores and gymnastics scores, though not all these improvements were statistically significant. Only one participant had a slight decrease of 0.025 points on one gymnastics event, and two components of the PSQI (sleep efficiency and use of sleep medication) stayed at zero with and without red light use.

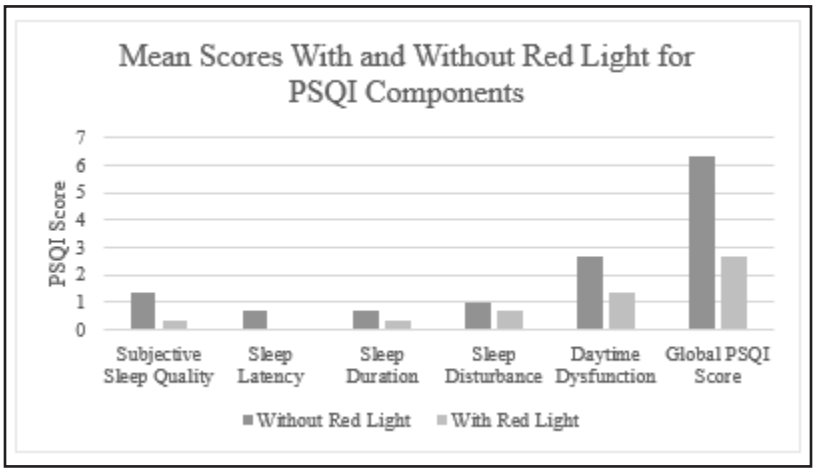


Figure 2. Scores in the components of the PSQI and Global PSQI score before and after red light use. Improvements were seen in subjective sleep quality (p=0.1012), sleep latency (p=0.1161), sleep duration (p=0.6779) sleep disturbance (p=0.3739), daytime dysfunction (p=0.0474), and global PSQI score (p=0.0697), sleep efficiency and use of sleep medication are not shown as the average was 0 with and without red light use.

The components of the PSQI that had the largest improvements were subjective sleep quality, having a mean improvement of 1 point, and daytime dysfunction, having a mean improvement of 1.3333 points (Figure 2). The improvement in daytime dysfunction scores was statistically significant (p=0.0474), meaning there is less than a 5% chance of the null hypothesis being correct, so hypothesis can be accepted for daytime dysfunction, and red light did help improve that aspect of sleep quality. The mean global PSQI score before red light use was 6.3334, which is in the poor sleep quality range, and the mean global PSQI score after red light use was 2.6667, which is in the good sleep quality range. So the red light may have helped participants achieve good sleep quality. Since participants had the larger improvements in subjective sleep quality and daytime dysfunction, they generally felt better during the day after red light use, even if the improvements in other components such as sleep disturbances were not as large or statistically significant.

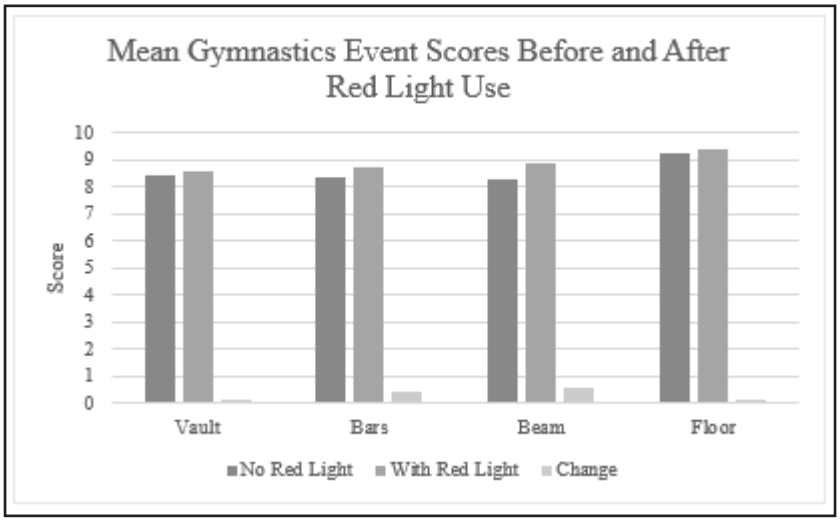


Figure 3. Gymnastics scores on each women's gymnastics event before and after red light use. The average scores improved on each event: vault (p=0.6263), uneven bars (p=0.0224), balance beam (p=0.0539), and floor exercise (p=0.2953). The average increase in all around scores was 1.3176 points (p=0.0572).

The gymnastics events with the largest improvements after red light use were uneven bars, having a mean improvement of 0.4 points, and balance beam, having a mean improvement of 0.5833 points (Figure 3). Bars and beam are typically the more cognitively demanding events, being more technical and requiring more focus, while vault and floor are typically the more physically demanding events, requiring more energy and power. The improvement of bars scores after red light use was statistically significant (p=0.0224), meaning there is less than a 5% chance that the null hypothesis is correct. So the hypothesis can be accepted for this gymnastics event, meaning red light use and improved sleep quality did correlate with improved performance on bars. Since the largest and most statistically significant improvements were on the more cognitively demanding events, red light use and improved sleep quality could have a larger impact on mental performance than physical performance.

Though it cannot be confidently concluded that red light use before sleep improved sleep quality and gymnastics performance, the results of this study indicate a relationship and requires further research.

The limitation of this study was the small sample size. Because of the sample size, a representative sample was not achieved, meaning that it cannot be concluded that the results are generalizable. The small sample size also meant that a control group could not be used, which would have helped account for factors such as improvements in gymnastics performance from getting more practice in between the competitions. A control group would allow for more confidence that the results seen could be attributed to red light use.

A next step could be a continuation of this study with more participants and controls. In further research, use of a control group and having more controls, such as having the same gymnastics judges at each competition and conducting the sleep aspect of the methodology in a laboratory setting, would likely yield more concrete results. Additionally, it could be beneficial to study red light use over a longer period of time, such as a month, and with different sports or measures of cognitive and physical performance. Studying red light use over a longer period of time would likely produce more accurate results, and different measures of sleep quality such as melatonin levels or brain scans could be more accurate than the Pittsburg Sleep Quality Index.

Bibliography:

E;, L. S. T. K. D. Y. K. H. T. L. (2021, January 7). *Sleep quality and athletic performance according to Chronotype. BMC sports science, medicine & rehabilitation. Retrieved November 2, 2022, from https://pubmed.ncbi.nlm.nih.gov/33413572/*

Figueiro, M. G., Sahin, L., Roohan, C., Kalsher, M., Plitnick, B., & Rea, M. S. (2019, May 3). Effects of red light on sleep inertia. Nature and science of sleep. Retrieved November 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6506010/

Gooley, J. J., Chamberlain, K., Smith, K. A., Khalsa, S. B. S., Rajaratnam, S. M. W., Van Reen, E., Zeitzer, J. M., Czeisler, C. A., & Lockley, S. W. (2011, March). Exposure to room light before bedtime suppresses melatonin onset and shortens melatonin duration in humans. The Journal of clinical endocrinology and metabolism. Retrieved November 29, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3047226/?report=printable

Gradisar, M., Wolfson, A. R., Harvey, A. G., Hale, L., Rosenberg, R., & Czeisler, C. A. (2013, December 15). *The sleep and technology use of Americans: Findings from the National Sleep Foundation's 2011 sleep in america poll. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine. Retrieved January 23, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836340/*

Ho Mien I;Chua EC;Lau P;Tan LC;Lee IT;Yeo SC;Tan SS;Gooley JJ; (2014, May 5). Effects of exposure to intermittent versus continuous red light on human circadian rhythms, melatonin suppression, and pupillary constriction. PloS one. Retrieved December 4, 2022, from https://pubmed.ncbi.nlm.nih.gov/24797245/

Klier, K., Dörr, S., & Schmidt, A. (2021, October 28). *High sleep quality can increase the performance of CrossFit® athletes in highly technical- and cognitive-demanding categories - BMC sports science, medicine and Rehabilitation. BioMed Central. Retrieved November 10, 2022, from https://bmcsportsscimedrehabil.biomedcentral.com/articles/10.1186/s13102-021-00365-2*

Mah, C. D., Mah, K. E., Kezirian, E. J., & Dement, W. C. (2011, July 1). The effects of sleep extension on the athletic performance of collegiate basketball players. Sleep. Retrieved November 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3119836/

Suni, E. (2022, March 11). *Sleep quality: How to determine if you're getting poor sleep. Sleep Foundation. Retrieved January 23, 2023, from https://www.sleepfoundation.org/sleep-hygiene/how-to-determine-poor-quality-sleep*

U.S. Department of Health and Human Services. (n.d.). What are sleep deprivation and deficiency? National Heart Lung and Blood Institute. Retrieved January 23, 2023, from https://www.nhlbi.nih.gov/health/sleep-deprivation

Zhao, J., Tian, Y., Nie, J., Xu, J., & Liu, D. (2012, December). *Red Light and the sleep quality and endurance performance of Chinese female basketball players. Journal of athletic training. Retrieved November 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3499892/*

Exploring The Effects of Music Genre on Memorization

Trevor O'Gorman

Nikola Tesla STEM High School

Abstract:

Is there a link between auditory stimuli and memory that can be used to improve student life? 56% of high school students will forget a topic they are taught within one hour of learning it, leading to that student to be 32% less likely to choose to relearn that topic, causing a lack of motivation for learning in many students. If music is played while learning something, could that music be used to recall those same memories easier later? Participants were asked to recite pictures and sentences taught twenty-four hours earlier to model memorization, and they listened to music genre playlists by a popular music app, Spotify. All participants also had a picture and sentence that they had to memorize with no auditory stimuli. This experiment used the Anova test (comparing the no music to each genre) to find whether there was a statistical significance between music listened to and ability to memorize, and all of the genres had P-values for the sentence scores were above 0.56, meaning that the data was statistically insignificant and no cues were made when memorizing the sentences. Despite that, the images did have a statistical significance returning P- values as low as 0.03, this means that the memorization of images was affected, but when looking at the data it is clear that the data was affected negatively. This means that in the real-world when memorizing language music is ok, but not ok when trying to memorize visual stimulus.

Introduction:

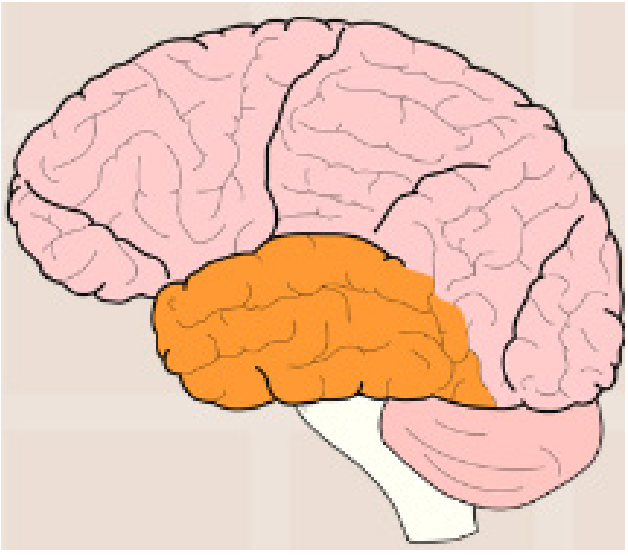
If we are able to find ways to create cues for our memories, we will be able to remember much easier; researchers at Melbourne Graduate School of Education Michael Tesoriero and Nikki Rickard found a consistent congr uence between auditory stimulus and memory recall in 100% of their participants, meaning that auditory stimuli may be the way to make these cues (Tesoriero, Rickard 2017). The researcher Youki Terada discovered that within one hour of getting taught something, 56% of students will forget what they were taught (Terada, 2017). It has also been noted by another researcher, Aia Haruvi that when we know less about a topic, they will be 32% less likely to be motivated to learn more about that topic (Haruvi, 2022). So, if we could find a way for a person to remember something in an easy, cheap, and simple way it could change many students' lives and allow people to learn or memorize faster, and in turn allow for more students to become more interested in those topics because they understand it better.

A potential solution for this problem is having a person associate a sound, or certain auditory stimuli, with a certain topic or subject of knowledge. When we associate that subject of knowledge or topic/ fact with an auditory stimulus, we could use the same stimuli to act as a cue to recall that encoded knowledge. Much like the proven phenomena of how we associate certain knowledge with certain olfactory stimuli. Considering that language is encoded into the temporal lobe, it could be said that speaking and words may be most susceptible to this effect when music is played, because the auditory cortex is also in the temporal lobe, meaning that a cue may be more plausible because of the similar location (see Figure 1). Previous studies by Aia Haruvi have also shown that all types of music can be used to help people have a higher amount of focus and cognitive ability showing that there is a possible relationship (Haruvi, 2022). This relationship has never been tested in depth; especially when you

bring in more than one genre and more than one topic to learn about.

One major resource and inspiration for this study was a study conducted by Michael Tesoriero and Nikki Rickard, where the studied the recall of emotionally encoded memories when music was played, or rather, memories that were encoded when of strong emotions occurred and finding if music had the ability to act as a cue for that pre-existing memory. This study found that there was a correlation, and that there was evidence of music-facilitated recall. Proof of a music-facilitated recall allows us to exploit music to recall certain memories whenever needed. They proved that music can recall already made memories with a similar cue, but this study uses music as a cue rather that the emotion associated with music as the cue like Tesoriero and Rickard.

If a person listens to a type of music while learning, listening to that music again will allow them to recall that memory easier. This phenomenon has been observed with taste (i.e. chewing gum), it is believed that this is because the place where taste is processed in the brain (the insular cortex) is near the Hippocampus, where memories are encoded. Since language memories and spatial awareness are encoded into the temporal lobe, the connection between that type of memory and an auditory stimulus may be made. This could then act as a cue when recalling said memory, allowing the person to memorize things at a much better pace when re-exposed to that music.



(Figure 1) Showing the auditory cortex and where language memories are processed. The highlighted section is the temporal lobe which is where this occurs.

Methodology:

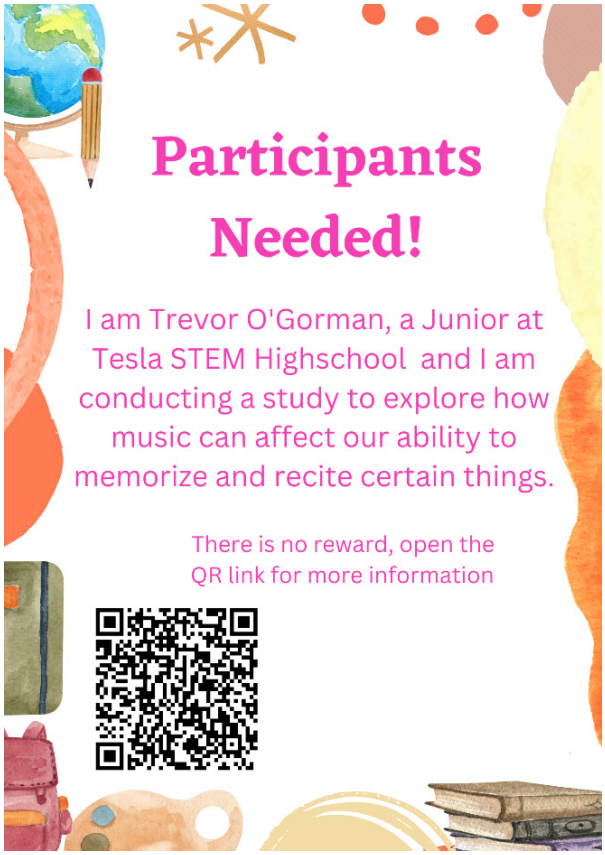
The materials needed for this experiment were 6 (Selected Short) Sentences Created from "Random Sentence Generator" (https://www.thewordfinder.com/random-sentence-generator/), 6 ai generated images, make the image the prompt will be "random objects" (https://deepai.org/machine-learning-model/text2img), Access to the following Spotify Made Spotify Playlists (filtered to be non-explicit): Country – https://open.spotify.com/playlist/37i9dQZF1DWZBCPUIUs2iR?si=69286e97896a42d1, Pop – https://open.spotify.com/playlist/37i9dQZF1DX11otjJ7crqp?si=54885c49d0d54ddf, Rap/Hip-Hop – https://open.spotify.com/playlist/37i9dQZF1DX9oh43oAzkyx?si=03e359a959c64f63, Rock – https://open.spotify.com/playlist/37i9dQZF1DWXRqgorJj26U?si=9ec3ef2f27364e8d, Classical – https://open.spotify.com/playlist/37i9dQZF1DWWEJlAGA9gs0?si=00de9b683a3440d1, Access to a Timer (10 minutes), An item that can play the music to all participants evenly, whether it be headphones, speaker, etc., and A process that displayed sentences and images (Power point or google slides)

To begin, a flyer that had a QR code (see Figure 2) on it that links to a Google form where people can learn about the experiment and give contact information like a phone number or email was created. This Google form also had the informed consent form attached so the person that was interested could fill it out. Then the flyers were posted in the 6 local high schools, because the age range was highschoolers, 13-18. After waiting for 20 participants to fill out the form, each participant was assigned a number 1-20. Then10 random numbers between 1 and 20 were generated, for each repeat one extra was generated. The 10 numbers generated corresponded to 10 participants with the same number that they were assigned. Then the participants were contacted to see if they would truly participate, if any said no or backed out prior to the actual beginning of the experiment more numbers were generated to fill that spot.

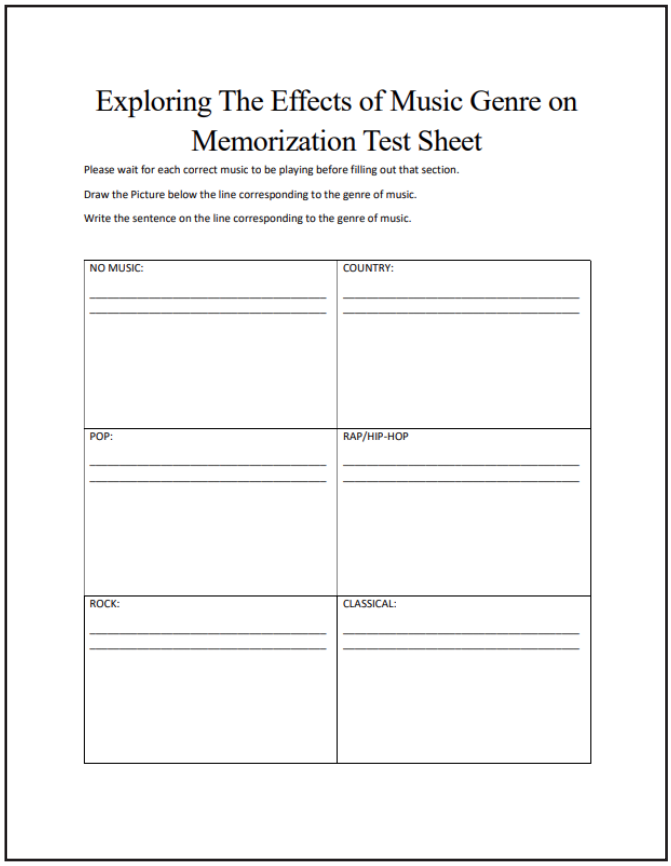
To test the participants, a capable confederate needs to be found to conduct the teaching, testing, and grading. This was done to create a double-blind experiment to reduce or completely get rid of bias. The confederate convened with the 10 selected participants at Tesla Stem High School then, they

showed the first random sentence and picture generated to the participants while no music was playing for 3 minutes. Then they waited 3 minutes with no music or learning. The confederate would then play one of the six albums and show the next random sentence and picture generated for 3 minutes of silence, then they would repeat that again with the rest of the music, pictures, and sentences.

When testing the confederate met again with the participants after 24 hours, this time they were first given a test sheet (see Figure 3). They had 3 minutes to fill out each section of the test sheet where during those 3 minutes the type of music noted on the sheet would be playing. After testing the confederate would also grade the responses, where the number of mistakes on each sentence and each photo was tallied (see table 1.1 and 1.2). This means that the confederate graded their responses, by seeing how close they are to the original sentences and pictures, counting the number of mistakes made. A mistake in the sentences is any miss worded parts or any words missing; to do this in the pictures, the confederate will count any incorrect key features of the photo or any incorrect key features.



(Figure 2.) – Recruitment Poster


Exploring The Effects of Music Genre on Memorization Test Sheet

Please wait for each correct music to be playing before filling out that section.

Draw the Picture below the line corresponding to the genre of music.

Write the sentence on the line corresponding to the genre of music.

NO MUSIC:	COUNTRY:
POP:	RAP/HIP-HOP
ROCK:	CLASSICAL:

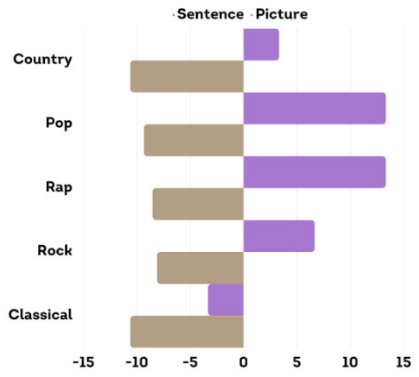
(Figure 3) – Testing Sheet

Table 1.1

Genre of music	No Music	Country	Pop	Rap/Hip-Hop	Rock	Classical
Avg % Correct of sentences After 24 Hours	20	23.333	33.333	33.333	26.667	16.667

Table 1.2

Genre of music	No Music	Country	Pop	Rap/Hip-Hop	Rock	Classical
Avg % Correct of Images After 24 Hours	10.606	0	1.282	2.083	2.5	0



(Figure 4) – Testing Sheet graph of averages minus the control

P-values	Sentence	Picture
Country	0.858313	0.034581
Pop	0.59947	0.076108
Rap	0.563385	0.182791
Rock	0.758332	0.034581
Classical	0.858313	0.034581

(Figure 5) – Table of P- values

Conclusion:

The null hypothesis cannot be rejected as all p-values supporting that the cues were made are much too high. The clearest conclusion from this experiment is that no cues were made when listening to music, but there were no negative effects to the sentences when they were exposed to music. This is because images are not processed in the temporal lobe meaning it is harder to recall them when exposed to a conflicting stimulus that is processed in the temporal lobe. This means that in the future we can listen to music if it is a language related concept, and we should avoid it if we are trying to memorize images.

Discussion:

Throughout this project, there were numerous adjustments that could have been made to improve accuracy, most apparent of which being that there was not much data to work with, the experiment only had 6 participants, meaning that the data, while correct, could be more accurate if there were more participants. Another adjustment that would be great if this experiment was re-done would be to survey the participants on what their favorite genre of music is because looking at Figure 4 it is clear that rap and pop music performed the best, which brings the question that they may have just been more motivated to memorize that because they liked the music more, as rap and pop are popular amongst this demographic. A final thing, it is apparent that cues may have not been made due to the inconsistency of genres, this experiment used genres of music as the cues which can be inconsistent in the way music sounds, even if they share similar ideas. Songs by themselves may have been more effective as they are extremely recognizable.These factors may have contributed to

the unfortunate P-values of the experiment and if this is redone, these should be taken into account.

Bibliography:

Haruvi, Aia. "Measuring and Modeling the Effect of Audio on Human Focus in Everyday Environments Using Brain-Computer Interface Technology." *Frontiers*, 27 Jan. 2022, www.frontiersin.or/articles/10.3389/fncom.2021.760561/full

HosseinShahabia1SaharMoghimiabcPersonEnvelope, A. links open overlay, HosseinShahabia1, a, 1, SaharMoghimiabcPersonEnvelope, b, c, AbstractThe purpose of this study was to investigate the effective brain networks associated with joyful, Wu, J., Vuilleumier, P., Shen, L. L., Peretz, I., Miskovic, V., Milde, T., Koelsch, S., Kay, B. P., Greenblatt, R., Flores-Gutiérrez, E. O., Davidson, R. J., … Cristianini, N. (2016, January 14). *Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity*. Computers in Human Behavior. Retrieved December 4, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S074756321630005X?via%3Dihub

LiqinZhanga1JingjingWangb1HuiyingSunaGuodongFengaZhiqiangGaoaPersonEnvelope, A. links open overlay, LiqinZhanga1, a, 1, JingjingWangb1, b, HuiyingSuna, GuodongFenga, ZhiqiangGaoaPersonEnvelope, Highlights•Auditory stimulation is essential for hippocampal function development.•Auditory overstimulation induces hippocampal changes, & AbstractIncreasing evidence has shown that noise overexposure could lead to impaired hippocampal function. Hippocampal alteration is also observed in several auditory deficits. (2022, February 3). *Interactions between the hippocampus and the auditory pathway*. Neurobiology of Learning and Memory. Retrieved December 4, 2022, from https://www.sciencedirect.com/science/article/pii/S1074742722000132

Music-enhanced recall:an effect of mood congruence, emotion arousal or ... (n.d.). Retrieved January 24, 2023, from https://findanexpert.unimelb.edu.au/scholarly-work/1189119-music-enhanced-recall-an-effect-of-mood-congruence--emotion-arousal-or-emotion-function%3F

Proverbio, A. M., Lozano Nasi, V., Alessandra Arcari, L., De Benedetto, F., Guardamagna, M., Gazzola, M., & Zani, A. (2015, October 15). *The effect of background music on episodic memory and autonomic responses: Listening to emotionally touching music enhances facial memory capacity*. Scientific reports. Retrieved October 4, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4606564/

Terada, Y. (2017, September 21). *Why students forget-and what you can do about it*. Edutopia. Retrieved November 10, 2022, from https://www.edutopia.org/article/why-students-forget and-what-you-can-do-about-it

Evaluation of Horse Comfort in English and Western Riding Style Saddle

Eshaal Omar

Nikola Tesla STEM High School

Abstract

An estimated thirty million people participate in the sport of horseback riding in the United States alone, which means that it is important to investigate how riding affects the horse. This project investigates how the saddles, which are the most distinct differences in the two most popular riding styles, English and Western riding style, affect the comfort of the horse while being ridden. The hypothesis was that horses ridden in English riding style would display more signs of discomfort than horses ridden in Western riding style. Participants started off by tacking up their horse in their respective riding style (IV) and were videoed riding around an arena, following a specific course that included the most common activities a horse and rider would perform in a walking gait. This video was analyzed and the measures of comfort (DV), eye blink rate and behaviors on the horse grimace pain scale, were added up. The experimental results did not support the hypothesis as a chi-squared test conducted for eye blink rate yielded a p-value of 0.662, while the horse grimace pain scale yielded a p-value of 0.2708, which meant that there was not a statistically significant difference in comfort behaviors in English versus Western riding style. The results of this study, if supported by other future studies, can help researchers in the equine field know that when investigating the riding styles and comfort of the horse, their focus can be directed to tack other than the saddle.

Introduction

This project seeks to investigate how the type of saddles typically used in each of the riding styles, English and Western style riding, affects the comfort of the horse while being ridden. As of 2022, an estimated thirty million people participate in the sport of horseback riding in the United States alone. This means that while horses aren't commonly used for transportation and labor anymore, they are still being ridden for recreation, sport, and even therapy. Therefore, a better understanding of the comfort or lack of comfort a horse feels while being ridden in the two most popular riding styles, English riding style and Western riding style, could help riders and their horses develop a better relationship with each other.

There has not been much research done regarding comparisons between the English and Western riding styles. However, one study that has touched upon this topic focused specifically on Dressage as an English riding discipline and trail riding as a Western riding discipline and found that overall, there was no statistically significant difference in discomfort behaviors displayed between Western and English horses (von Borstel et al., 2013). As measures of discomfort, they focused on several things, like certain behaviors of the horses and their heart rates, but mainly on rein tension. Dressage and trail riding are quite different, and these differences go beyond one of them being an English riding discipline and the other being a Western riding discipline. Reining would be considered more of the Western equivalent of dressage, and trail riding generally stays the same between the two riding styles, except for differences in tack, something that will always be there when comparing any of the riding discipline of English and Western riding style. It should be noted that it is Western riders that generally do trail riding,

which could be a reason why the study focused specifically on trail riding when they generalized it to Western riding style in the title. Due to these differences, certain measures of discomfort the study used, especially rein tension and heart rates of the horse, likely vary due to the differences in what the horse would be doing in dressage versus trail riding. This means that the results of the study are more applicable to just dressage and trail riding rather than English or Western riding styles as a whole. Even when making comparisons of the overall riding styles, it is something to consider as it compares aspects of English and Western riding styles that will likely be present regardless of whether specifically dressage or specifically trail riding is present.

So when measuring the comfort of the horses and making comparisons between English and Western riding styles as a whole rather than specific riding disciplines within them, it is important to consider other factors that might affect the measures of comfort in the horse. An example is heart rate, which is commonly used as a measure of comfort in horses. While it can be effective for this purpose, it is a variable that change depending on the gait or the different riding disciplines within English and Western riding styles, which is why it is often used along with other measures to measure the comfort of a horse. Some other ways of measuring comfort in a horse can be more consistent over riding styles and gaits, an example is oral stereotypies which horses display as a sign of discomfort, like crib biting and wood chewing. One study conducted that used oral stereotypies as a measure of comfort in the horse investigated whether management factors of riding styles can be significant in behavioral problems in horses and found that when it came to riding style, oral stereotypies occurred most often among horses ridden in the English riding style (Normando et al., 2011). They concluded that while restrictive stabling was also most common in English riders, it wasn't linked to the oral stereotypies, but the way a horse is typically ridden in English riding style (more physical contact with horse than Western riding style, which might not translate correctly with inexperienced riders) combined with the restrictive stabling could have contributed to overall results of English riding style horses displaying the most signs of discomfort.

There have been studies regarding saddle fit in specifically English riding style or specifically Western riding style and the effects on the rider and horse (including discomfort behaviors), but there is a lack of studies comparing anything relating to saddles between each of the riding styles. For example, a study compared the pressure distribution of two types of dressage saddle (meaning an English riding style saddle) and while discomfort of the horse wasn't one of the main variables being observed, heart rate of the horse was observed. However, no comparisons were made dressage saddles and saddles used in Western riding style. Similarly, as mentioned earlier, studies comparing the English and Western riding styles have been made but none of them focus specifically on the saddle. This study will focus specifically on the type of saddle used in each of the riding style and how that will affect the comfort or lack of comfort in each of the styles. In previous studies where English and Western riding styles are compared and the comfort of the horse is evaluated, English horses usually displayed more signs of comfort. Factors other than the saddle or type of tack used in this type of riding style could have contributed to that, but it is likely that when focusing specifically on the saddles, horses being ridden in English riding style will still display more signs of comfort than horses being ridden in Western riding style.

Materials and Methods

Participant Recruitment

To avoid sampling bias, random sampling was done when recruiting participants. Flyers that provided a detailed explanation of the requirements and what will be done in the experiment were put up around the area where the experiment took place. In the flyer, a QR code was included that could be scanned to show a link of a survey where people indicate if they ride their horse in Western riding style or English riding style, if they are a minor or adult, if they own a horse that they ride, and if their horse has any health conditions that would make it tired easily. When people responded and fit the requirements (own a horse that they ride, adult, ride the horse with English riding style or Western riding style, horse doesn't have health conditions that make it tired easily), ten participants that responded that they ride in Western riding style were randomly chosen and ten participants that responded that they ride in English riding style were randomly chosen.

To avoid bias, someone other than the person conducting the experiment randomly picked the participants and used a random number generator to assign them their number from one to eight. The eight participants were then prompted to fill out an informed consent form. Informed consent forms with participants' names and other information were filed away during the experiment and won't be referenced to avoid any bias. These were discarded immediately after the experiment is over.

Experimental Design

There were two experimental groups throughout this experiment as described previously: Western riding style group and English riding style group. Those in the Western riding style group used a Western style saddle (as seen below in Figure 1). These saddles have a horn, skirt, latigo, gullet, stirrup, cantle, and seat. Those in the English riding style group will use an English style saddle (as seen below Figure 1). These are distinguishable from Western style saddle by the lack of a horn in the front. These saddles have a cantle, pommel, skirt, saddle flap, stirrup and stirrup leather and seat.

Each participant chose to perform their experimental procedure in the arena that they most frequently ride at, as this could reduce chances of a horse displaying signs of discomfort due to unfamiliarity with the environmental rather than anything related to the saddle of the riding style. These arenas often had dimensions of around 20 by 40 meters, although there was a slight variation. Participants started off by tacking up in their respective riding style. When the rider was done with tacking up, they rode their horse around the arena in a walking gait for five minutes for the purposes of warm-up. Once the warm-up was finished, riders had an observer, who was standing a safe distance away from the horse and rider, videotape them completing the procedure described in Figure 2 below. The observer focused on videotaping the horse's head, for the purpose of maintaining participant confidentiality and since the comfort behaviors that were analyzed later were focused on this part of the horse, as the horse and rider rode on the left rein on the track around the area once, then made a circle. They changed direction across the diagonal and rode around the arena on the right rein once and made a 20-meter circle again. A walking gait was maintained throughout this process. This specific procedure was chosen to include as many activities that a rider and horse would normally do in a walking gait. After this process wa over, the rider dismounted their horse, removed tack, and returned their horses back to their homes.

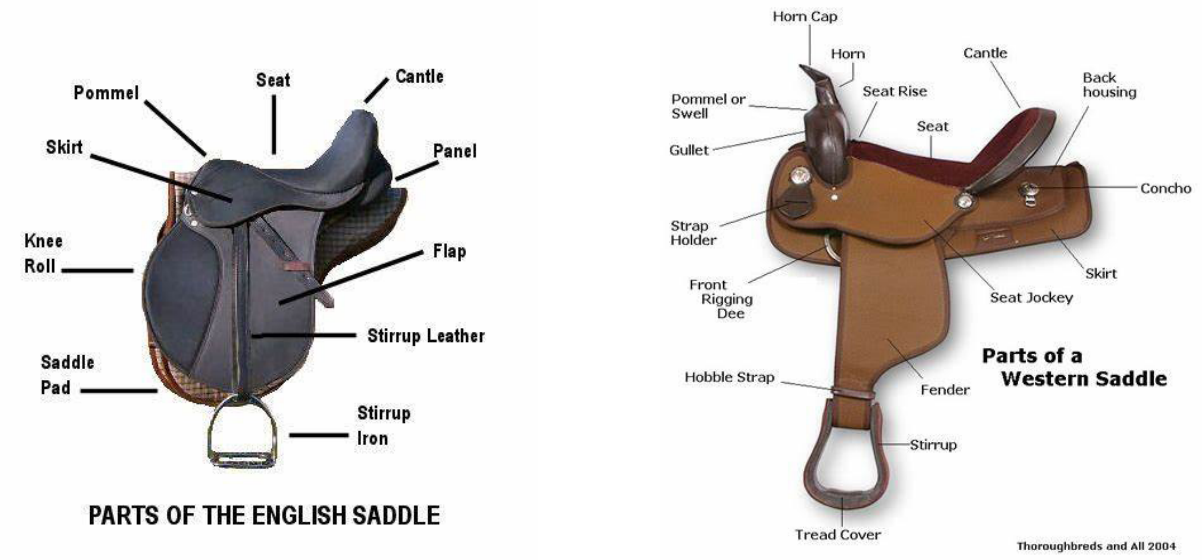


Figure 1. English style saddle is show on the left with the parts labeled. Western style saddle is shown on the right also with the parts labeled. (Robinson, 2022).

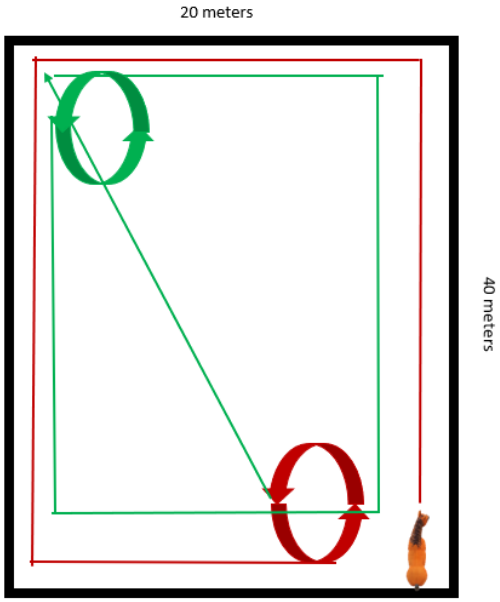


Figure 2. Visual representation of procedure participants followed in experiment

The video of the horse and rider walking around the arena was sent to the experimenter and it was then analyzed retrospectively. Several observations were made, including the amount of eye blinks of the horse, which was observed and added up. One blink is defined by when the horse shuts and opens its eye fully. A higher blinking rate in horses is linked to their comfort. The six behaviors on the HGS (horse grimace pain scale), seen in Figure 3 below, were observed in each video by splitting the video into three equal intervals. A screenshot was taken of the horse's head at each interval and the HGS behaviors were observed and averaged for the three intervals, resulting in one value for each horse that was an average of the HGS behaviors seen throughout the video.



Figure 3. Behaviors on the HGS (Horse Grimace Pain Scale), also known as FAUs or facial action units, are seen above. Each behavior is rated on a scale of 0 to 2, 0 being not present, 1 being moderately present, and 2 being obviously present. The maximum possible score overall would be 12. (Dalla Costa et al., 2014).

Data Tables

Shown below are examples of how data tables looked for a horse in English riding style saddle and a horse in Western riding style saddle. Information about the assigned number given to the horse and participant, the riding style, video time, eye blink rate per minute, and the average score for the HGS scale was recorded for each horse.

Data Table 1.1 – Comfort Behaviors of Horse with English Tack

Horse	Discipline	Video Time	EBR/min	HGS (avg.)
1	English	1:00	27	5.6

Data Table 1.2 – Comfort Behaviors of Horse with Western Tack

Horse	Discipline	Video Time	EBR/min	HGS (avg.)
2	Western	2:16	28	4.3

Data Table 1.3 – Raw Data Collection for All Horses

Horse	Discipline	Video Time	EBR/min	HGS (avg.)
1	English	1:00	27	5.6
2	Western	2:16	28	4.3
3	Western	1:26	24	5
4	Western	1:45	20	4.6
5	English	1:40	19	4
6	English	1:22	29	5.6
7	Western	1:20	24	4.3
8	English	1:34	26	5

Data Analysis

The mean was taken for the values obtained for eye blink rate per minute and the average HGS score for English and Western riding styles. There was only a slight variation in averages for the eye blink rate in English riding style versus Western riding style, as the English riding style average was 25.25 blinks per minute, while the Western riding style average was 24 blinks per minute. Similarly, the average HGS score for English riding style was 5.05, while the average HGS score for Western riding style was 4.55, which meant there was even less of a difference between the two riding style for behaviors on the HGS scale compared the eye blink rates. Nevertheless, the two riding styles showed very little difference in both ways of measuring comfort in the horse. To see if the results were statistically significant, a chi-squared test was performed on eye blink rate and on behaviors on the HGS scale. The chi-squared test yielded a p-value of 0.662 for eye blink rate and a p-value of 0.2708 for behaviors on the horse grimace pain scale. These p-values mean that the results were not statistically significant and there was not a statistically significant difference in the comfort behaviors displayed in English and Western riding style.

Conclusions and Discussion

After data analysis was completed, which included taking the mean and obtaining the p-value through performing the chi-squared test, it can be concluded that there is no statistically significant difference in the comfort behaviors in the two most

popular riding styles, English and Western riding styles, based on this study. These results did not support the hypothesis that horses with English riding style saddles would yield results showing less comfort behaviors.

Possible reasons for these results that were not statistically significant could have been due to several limitations, like the fact that there was only eight participants in the study, four for each riding style. With more participants, it could have been possible that data analysis would show results that a horse in one of the riding style experienced more comfort than the other. It is also difficult to determine whether it was only the saddle in each riding style that caused the results as most riders were not comfortable riding without a bridle and bit, which meant that all riders used a bridle as well as the saddle. These bridles differed slightly between the riding styles and could have been a contributing cause to the comfort the horse experienced as well as the saddle itself. Additionally, there was no control group in this study, although a control group might have caused its own problems. Ideally, this study would have a control group of bareback riders, but it is difficult to find riders experienced enough in bareback riding, meaning that a control group with bareback riders might show results with the horse experiencing discomfort due the rider's lack of experience in bareback riding rather than anything to do with a saddle or lack of saddle.

Despite not yielding statistically significant results, this study can still be helpful to those researching the English and Western riding styles, the differences between the two, and how that would potentially affect the horse. There is already research on rein tension between the two riding styles and how that affects the horse and now that this study showed results supporting the idea that there is not a statistically significant difference between the comfort a horse experiences in one riding style versus the other, those interested in this field of research can focus on tack other than the saddle and reins, for example the girth or cinch. Furthermore, this topic of English and Western riding style saddles and comfort of the horse can still be investigated more thoroughly, perhaps focusing on specific parts of the saddle, as there are not any other published works of research regarding this topic. In the end, the goal is to maximize the amount of comfort a horse experiences while being ridden, so if any research regarding tack in the different riding styles and comfort of the horse do yield statistically significant results where a horse experiences more comfort in one riding style, parts of the tack in that riding style could be implemented into the other riding style, while still maintaining the original purposes of each riding style.

Bibliography:

Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC (2014) Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. PLoS ONE 9(3): e92281. https://doi.org/10.1371/journal.pone.0092281

Kuhnke, Sandra & König von Borstel, Uta. (2016). The symmetry of rein tension in English and western riding and the influence of human and equine laterality. Journal of Veterinary Behavior: Clinical Applications and Research. 15. 91. 10.1016/j.jveb.2016.08.054.

König von Borstel, Uta & Kassebaum, L. & Gauly, Matthias. (2013). Effect of discipline (Western vs Dressage) and skill level on learning and discomfort behavior of riding horses. Journal of Veterinary Behavior: Clinical Applications and Research. 8. e13. 10.1016/j.jveb.2012.12.029.

Merkies, K., Ready, C., Farkas, L., & Hodder, A. (2019). Eye Blink Rates and Eyelid Twitches as a Non-Invasive Measure of Stress in the Domestic Horse. Animals : an open access journal from MDPI, 9(8), 562. https://doi.org/10.3390/ani9080562

Normando, Simona & Meers, Lieve & Samuels, William & Faustini, Massimo & Ödberg, Frank. (2011). Variables affecting the prevalence of behavioural problems in horses. Can riding style and other management factors be significant?. Applied Animal Behaviour Science. 133. 186-198. 10.1016/j.applanim.2011.06.012.

Schleese, J. (2022, March 12). Saddle fit – English vs. western. Horse Journals. https://www.horsejournals.com/riding-training/english/other/saddle-fit-english-vs-western

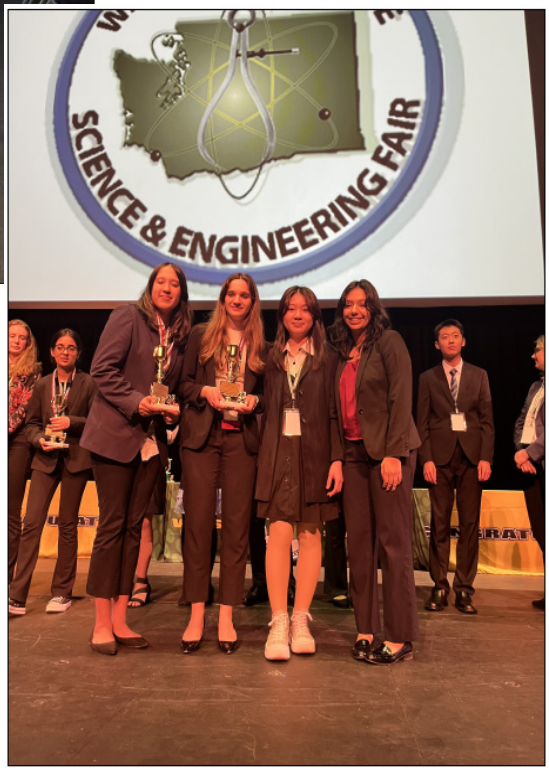
Awards & Achievements

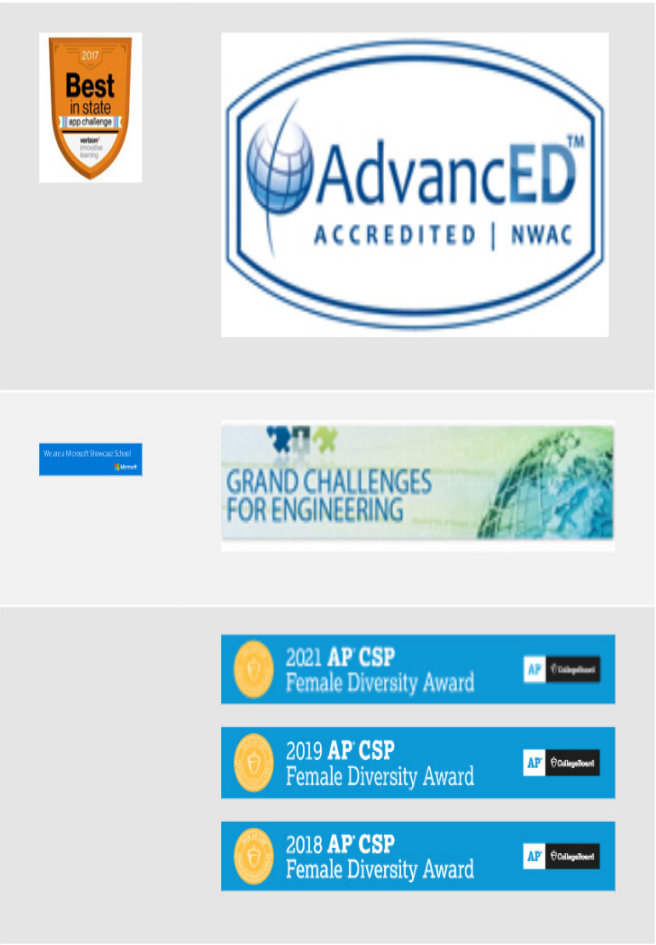














2022-23 College Board AP® School Honor Roll

AP Computer Science Female Diversity Award

2022 International Olympiad in Astronomy and Astrophysics

This past summer, senior Evan Kim brought home a gold medal at the 15th International Olympiad in Astronomy and Astrophysics (IOAA) in Kutaisi, Georgia as part of the US team.

The IOAA is an annual competition for the top astrophysics students from over 40 countries. At the IOAA competitors compete in three rounds—theory, data analysis, and observational—which collectively capture the necessary skills in Astronomy and Astrophysics. In the theory round, contestants must answer a series of long answer questions analyzing astrophysical systems, ranging from eclipsing binaries to relativistic jets. In the data analysis round, contestants apply their theory knowledge and number crunch through real world astronomical data to make conclusions. And in the observational round, the most studied for round, contestants must identify various astronomical objects by sight, on both star charts and in the night sky.

American Mathematics Competition

From Math Club Advisor Mr. Kyle Ostlie:

Each year, the math club participates in the **American Mathematics Competitions(AMC)** which is the first of a series of three challenging competitions in high school mathematics that determine the United States team for the **International Mathematical Olympiad (IMO).** The AMC is one of the largest mathematics competitions in the nation with roughly 160,000 students participating, and we spend most of the year in math club preparing for it.

This year we had 29 students participate and 15 students qualified for the next round, the American Invitational Mathematics Examination (AIME) on February 7th. Only the top 10% of AMC scores in the nation are invited. This is by far the most students we have had qualify. Even more exciting, we had 3 students score in the top 1%!

9th and 10th Grade Division Qualifiers:

Ekaansh Agrawal (Top 5%)

Pranav Kannepalli (Top 1%)

Sarthak Lodha

Reese Long (Top 5%)

Wesley Lu (Top 5%)

Neil Parikh (Top 5%)

Alexander Peev (Top 1%)

Sarah Wen

Winston Zhou (Top 5%)

11th and 12th Grade Division:

Yue (Beth) Ding

Evan Kim (Top 5%)

Sebastian Kumar (Top 5%)

Chris Ma

Rui Meng (Top 1%)

Abhiram Yakkali

Regeneron Science Talent Search Finalist

Congratulations to Tesla STEM senior Evan Kim, who was named one of the 40 finalists among nearly 2000 students who participated in the 82nd Regeneron Science Talent Search. The final competition will take place in Washington, DC, in March 2023

2 Tesla STEM Scholars named to Top 300 in Regeneron Science Talent Search

Congratulations to Ms. Kate Allender and her two science scholars, Evan Kim and Harish Krishnakumar, who have been named as two of the top 300 young scientists of the 82nd Regeneron Science Talent Search (STS). The Regeneron Science Talent Search is the nation's oldest and most prestigious pre-collegiate science competition, providing a prominent forum for original research that is recognized and reviewed by a national jury of professional scientists. Additionally, Evan Kim has advanced to the final level of competition which will take place in Washington D.C. the week of March 9th. All 40 finalists will be evaluated rigorously by judges to determine the top ten winners, competing for the top prize of $250,000. They will present their original science research project to the public and will have the opportunity to meet prominent leaders and scientists through field trips and other special events.

Evan Kim's Project: ScGAN: A Generative Adversarial Network to Predict Hypothetical Superconductors

Harish Krishnakumar's Project: Analysis of Ring Galaxies Detected Using Deep Learning with Real and Simulated Data

American Mathematics Competitions (AMC)

Each year, the math club participates in the **American Mathematics Competitions (AMC)** which is the first of a series of three challenging competitions in high school mathematics that determine the United States team for the **International Mathematical Olympiad (IMO)**. The AMC is one of the largest mathematics competitions in the nation with roughly 160,000 students participating, and we spend most of the year in Math Club preparing for it.

This year Tesla STEM had 29 students participate, and 15 students qualified for the next round, the **American Invitational Mathematics Examination (AIME)** on February 7th. Only the top 10% of AMC scores in the nation are invited. This is by far the most students we have had qualify. Even more exciting, we had 3 students score in the top 1%! Please congratulate these scholars:

9th and 10th Grade Division Qualifiers:
Ekaansh Agrawal
Pranav Kannepalli
Sarthak Lodha
Reese Long
Wesley Lu
Neil Parikh
Alexander Peev
Sarah Wen
Winston Zhou

11th and 12th Grade Division Qualifiers:
Yue (Beth) Ding
Evan Kim
Sebastian Kumar
Chris Ma
Rui Meng
Abhiram Yakkali

Science Olympiad: Results from Four Invitationals

Science Olympiad is a set of 23 events including biological, physical, chemical and earth sciences, as well as engineering, for teams of up to 15 students. They compete in groups or 2 or 3 in every event.

The "season" is divided into two: invitationals, which allow for practice against different levels of competition, and regional/state/national finals, which are more formal events.

This year, Tesla STEM Science Olympiad participated in four invitationals, with excellent results:

- The **Boyceville SO Invitational,** in December, was a remote event hosted by Boyceville HS (of the Wisconsin Boycevilles). Tesla STEM entered 3 teams, and our 1st team finished in 10th place out of nearly 150 teams.
- Also in December, the team traveled south for the Camas SO Invitational. Team STEM won 1st Place, Team Leaf earned 5th Place, Team Root earned 17th Place, and Team Bud earned 24th Place.
- In January, the team spent a day with aircraft at the Raisbeck Aviation SO Invitational. Team STEM won 1st Place, Team Leaf earned 5th Place, and Team Root earned 22nd Place.
- Finally, in February, we sent one team to cold, rainy Berkeley for the Golden Gate Science Olympiad. This is one of two national-level invitationals on the West Coast and, in a field of 50 teams that included recent national champions, Team STEM earned 7th Place (and the 6th Place trophy, because we have two teams from Troy HS to beat!).

We have listed students' results at each event below. It is hard to overstate how proud we are of these teams for much more than the results—it is a joy to see the leaders build a culture and esprit de corps that are just as important as the results they produce.

This is a schoolwide effort--we have support from our admin, our office staff (Ms. O'Brien and Ms. Ferguson), our health staff (Ms. Welcome and Ms. Johnson), and a bunch of teachers that indulge our teams' requests for equipment use, materials, and opportunities to practice. We are grateful.

Specific Event Results:

Boyceville SO Invitational (out of 148 teams):

- Team STEM: 10th Place overall
- Chemistry Lab: Sebi Kumar & Sami Badal Khan, 2nd place
- Detector Building: Divij Aswinkumar & Sami Badal Khan, 5th place
- Dynamic Planet: Aashray & Ashwin Kaliyaperumal, 2nd place
- Environmental Chemistry: Sebi Kumar & Ashwin Kaliyaperumal, 6th place
- Experimental Design: Anjali Sreenivas, Siddhant Porwal, & Sree Sarvepalli, 1st place
- Fermi Questions: Evan Kim & Sebi Kumar, 6th place
- Green Generation: Olivia Wang & Divij Aswinkumar, 4th place
- Wifi Lab: Sami Badal Khan & Evan Kim, 6th place

Camas SO Invitational (out of 25 teams):

- Team STEM: 1st place overall
- Team Leaf: 5th place overall
- Anatomy & Physiology: Rithani Saravanakumar & Shuhui Yang, 3rd place
- Astronomy: Evan Kim & Aashray, 2nd place; Harish Krishnakumar & Rajit Joshi, 3rd place

- Cell Biology: Shuhui Yang & Grace Zhao, 2nd place; Prerana Kota & Angela Wang, 3rd place
- Chemistry Lab: Sebi Kumar & Sami Badal Khan, 1st place; Rajit Joshi, Parik Venkataraghavan, 5th place
- Codebusters: Olivia Wang, Sebi Kumar, & Divij Aswinkumar, 1st place
- Disease Detectives: Anjali Sreenivas & Grace Zhao, 3rd place; Sachita Ghosh & Keerthi Rajesh, 4th place
- Dynamic Planet: Aashray & Ashwin Kaliyaperumal, 2nd place
- Environmental Chemistry: Sebi Kumar & Ashwin Kaliyaperumal, 1st place
- Experimental Design: Anjali Sreenivas, Siddhant Porwal, & Sree Sarvepalli, 1st place
- Fermi Questions: Evan Kim & Sebi Kumar, 1st place
- Flight: Tyler Chen & Shrey Goel, 4th place
- It's About Time: Farzad Hasan, 4th place
- Forensics: Rithani Saravanakumar & Anjali Sreenivas, 1st place; Prerana Kota & Tyler Chen, 2nd place
- Forestry: Olivia Wang & Neha Madanagopal, 2nd place
- Green Generation: Prerana Kota & Brenda Li, 1st place; Divij Aswinkumar & Olivia Wang, 2nd place
- Remote Sensing: Aashray & Evan Kim, 1st place
- Rocks & Minerals: Aashray & Sami Khan, 2nd place; Brenda Li & Saharsh Bhargava, 3rd place
- Bridge: Andy Pham & Neha Madanagopal, 5th place
- Detector Building: Divij Aswinkumar & Sami Badal Khan, 1st place
- It's About Time: Andy Pham & Divij Aswinkumar, 5th place
- Scrambler: Siddhant Porwal & Andy Pham, 4th place
- Solar Power: Sree Sarvepalli & Ashwin Kaliyaperumal, 1st place
- Trajectory: Siddhant Porwal & Sree Sarvepalli, 3rd place
- Wifi Lab: Sami Khan & Evan Kim, 4th place
- Write It Do It: Madeleine Vanderheiden & Mehuli Nath, 4th place

Raisbeck Aviation Invitational (out of 27 teams):

- Team STEM: 1st place overall
- Team Leaf: 5th place overall
- Anatomy & Physiology: Rithani Saravanakumar & Shuhui Yang, 3rd place
- Astronomy: Evan Kim & Aashray, 2nd place; Madeleine Vanderheiden & Rajit Joshi, 5th place
- Bridge: Neha Madanagopal & Andy Pham, 1st place; Sriram Marasanapalle & Mehuli Nath, 3rd place
- Cell Biology: Shuhui Yang & Grace Zhao, 1st place
- Chemistry Lab: Sebi Kumar & Sami Badal Khan, 2nd place
- Codebusters: Olivia Wang, Sebi Kumar, & Divij Aswinkumar, 1st place
- Detector Building: Divij Aswinkumar & Sami Badal Khan, 1st place
- Disease Detectives: Anjali Sreenivas & Grace Zhao, 2nd place; Sachita Ghosh & Keerthi Rajesh, 4th place
- Environmental Chemistry: Sebi Kumar and Sami Badal Khan, 1st place
- Experimental Design: Anjali Sreenivas, Siddhant Porwal, & Sree Sarvepalli, 1st place
- Fermi Questions: Evan Kim & Sebi Kumar, 2nd place
- Flight: Rithani Saravanakumar & Neha Madanagopal, 3rd place
- Forensics: Anjali Sreenivas & Rithani Saravanakumar, 1st place; Prerana Kota & Tyler Chen, 2nd place
- Forestry: Olivia Wang & Neha Madanagopal, 1st place; Naomi Stahl & Katya Rukhlinskaia, 4th place
- Green Generation: Prerana Kota & Brenda Li, 5th place

- It's About Time: Andy Pham & Divij Aswinkumar, 1st place
- Remote Sensing: Aashray & Evan Kim, 1st place
- Rocks & Minerals: Aashray & Sami Khan, 2nd place; Brenda Li & Saharsh Bhargava, 4th place
- Scrambler: Siddhant Porwal & Andy Pham, 2nd place
- Solar Power: Shrey Goel & Madeleine Vanderheiden, 1st place; Sree Sarvepalli & Aashray, 2nd place
- Trajectory: Siddhant Porwal & Sree Sarvepalli, 2nd place
- Wifi Lab: Sami Badal Khan & Evan Kim, 3rd place

Golden Gate Science Olympiad (out of 53 teams):

- Team STEM: 7th Place overall
- Chemistry Lab: Samiullah Badal Khan & Sebastian Kumar, 2nd Place
- Disease Detectives: Anjali Sreenivas & Grace Zhao, 8th Place
- Dynamic Planet: Chegu Vijay Aashray & Ashwin Kaliyaperumal, 7th Place
- Environmental Chemistry: Ashwin Kaliyaperumal & Sebastian Kumar, 6th Place
- Experimental Design: Siddhant Porwal, Sreevatsa Sarvepalli, & Anjali Sreenivas, 3rd Place
- Fermi Questions: Evan Kim & Sebastian Kumar, 1st Place
- Green Generation: Divij Aswinkumar & Olivia Wang, 7th Place
- Remote Sensing: Chegu Vijay Aashray & Evan Kim, 1st Place
- Trajectory: Siddhant Porwal & Sreevatsa Sarvepalli, 4th Place

Washington State High School Chess Championships

From coach Matt Sturtevant:

The Tesla STEM Chess Club/Team finished their regular season Mid-February in the KingCo chess league, finishing in 5th place and earning a bid to the State Team Tournament held this last weekend in Stanwood, Washington.

The Team competed in 2 rounds (60 minutes per player) Friday evening and another 3 rounds (90 minutes per player) on Saturday.

These young adults represented their district and school extremely well, exhibiting great sportsmanship and teamwork throughout the weekend.

I am happy to report that our Team finished 6th in state, earning 16.5 out of a possible 25 points. The top team finished with 18.5. Please congratulate these students:

Board 1: Anthony He

Board 2: Stephanie Velea

Board 3: Aakash Banerji

Board 4: Sean Ge

Board 5: Vignesh Srinivasan

National History Day

From advisor Andrew Ivy:

This past Saturday (March 4) was the Regional National History Day competition, held at Evergreen Middle school, and attended by 23 of our students who produced 12 different projects.

NHD is an annual competition where student conduct primary source research on a topic of their choosing, but in conjunction with a specific annual theme. They present their work in a variety of formats, from historical papers to scripted performances. This year's theme was "Frontiers in History: People, Places, Ideas," and 8 of our entries qualified to go on to the state competition.

Druhin Bhowal, Individual Documentary – Sowing the Seeds of Progress: The Green Revolution and its Unintended Consequences

Arihant Singh, Ethan Yu, Teddy Dong, Ziang Chen, William Gong, Group Documentary – From Atrocities to Accountability: The Nuremberg Trials as a Frontier in International Justice

Cindy Yang, Individual Exhibit – Crossing the Western Frontier: Chinese Labor on the Transcontinental Railroad

Nakul Rajpal, Individual Exhibit – The Frontier of Redlining in Seattle

Brooke North, Prakshi Shukla, Group Exhibit – Words Were Her Weapon: Ida B Wells and the Fight Against Lynching

Saachi Sharma, Samantha Novitski, Group Performance – In the Wasatch Valley: A Historical Fight for Women's Rights

John Bannon, Research Paper – The Seattle Jazz Frontier: The Success of the African American Community in Shaping the American Identity

Anson Chen, Research Paper – Taking on the Titan: the Progressive Era Pioneers who Brought Down Rockefeller's Standard Oil

State competition is April 22.

2023 Washington State National History Day

From advisor Mr. Andrew Ivy:

This past Saturday 14 students from Tesla STEM competed in the State level competition for National History Day. There were 8 different projects, including research papers, documentaries, exhibits and a performance, and all but 2 of those made the "finals" round of judging.

Two of those pieces of work qualified for the National competition, held at the University of Maryland, June 11-15. And so please congratulate:

- Samantha Novitski and Saachi Sharma for their second-place finish in the Group Performance Category. Their piece was titled "Journalist Emmeline B Wells: The Frontier of the Women's Rights Movement in the Wasatch Range."
- Cindy Yang for her second-place Individual Exhibit entitled: "Crossing the Western Frontier: Chinese

Labor on the Transcontinental Railroad."

Also, John Bannon won a cash award for the best use of Regional/Local archives for his project entitled: "The Seattle Jazz Frontier: The Success of the African American Community in Shaping the American Identity."

This overall strong performance by Tesla STEM students (including 2 "just missed" third-place finishes), resulted in Tesla STEM being acknowledged as the 2023 Washington State History Day Outstanding School, Senior Division.

2023 Scholastic Art and Writing Contest -Silver Key Award

Two Tesla STEM students won the **Silver Key Award** at the 2023 Scholastic Art and Writing Contest.

- Junior **Rui Meng** has won The Silver Key Critical Essay Award for his work about "technology, war, and the destruction from war".
- Freshman **Naman Mutalik Desai** has won The Silver Key Award for his memoir entry titled "The Best Apology is Changed Behavior".

"The Region-at-Large West Writing program is administered by the Alliance for Young Artists & Writers in New York City with the support of adjudication partners from around the country. This remote program serves all students located or attending school in the western United States where the regional Awards are not administered by a local Affiliate Partner." (Scholastic Art and Writing Awards)

2023 Pacific Northwest Regional Science Bowl

On Saturday 25 February, Tesla STEM HS sent three teams to the Regional Science Bowl, hosted by the U.S. Department of Energy's Pacific Northwest National Labs.

- Team 3 (Eva Jain, Saharsh Bhargava, Cindy Yang, Krishna Senthilkumar & Jas Singh) finished in **8th place**
- Team 2 (Prerana Kota, Reese Long, Shuhui Yang, Aakarsh Balla, & Anson Chen) finished in **6th place**
- Team 1 (Evan Kim, Aashray, Sami Badal Khan, Sebi Kumar, & Tejas Pakalapati) finished in **1st place!**

Because they won, Team 1 will head to Washington, D.C. for the National Science Bowl in April 2023! They have won this difficult competition in 3 of the past 4 years.

2023 King County Envirothon

On a rainy Thursday in March, a six-student team of Tesla STEM students competed outdoors (and got wet) in the King County Envirothon. They earned the highest score in three of the five events and earned 1st place in the competition!

Please congratulate Aashray, Saharsh Bhargava, Madeleine Vanderheiden, Sami Badal Khan, Johan Karukayil, and Anson Chen, who will travel to beautiful Longbranch, WA, in May to participate in the State Envirothon.

American Invitational Mathematics Examination (AIME)

From Math Club Advisor Mr. Ostlie:

Last month we had 15 students participate in the American Invitational Mathematics Examination (AIME), and we just got our results back. Just participating in the competition was a big accomplishment, but we had 3 students qualify for the next round of testing, the USAMO, from which students will be selected for the national Mathematical Olympiad Program!

Rui Meng
Alexander Peev
Reese Long

This is a huge accomplishment! To put it in perspective, the series of examinations started with roughly 160,000 students taking the AMC, then only 6500 students qualified to take the AIME, and now these 3 are in the rare company of only 558 participants invited to take the USAMO. Our students will be competing against the top mathematical teen minds in the country. We have only had one other student ever make it this far.

The USAMO is a proof-based mathematics examination that will take place on both Tuesday, March 21st and Wednesday, March 22nd during the school day. Each day consists of 3 questions over 4.5 hours.

Central Sound Regional Science & Engineering Fair (CSRSEF)

From instructor and advisor Ms. Kate Allender:

Please congratulate our Tesla STEM scholars on a tremendous showing at CSRSEF over the weekend of March 11-12, which was a hybrid event. The first day was held remotely over zoom, and then 16 projects were selected to move onto finalist judging in person. Students raked in a record-breaking 24 category awards including 1 grand prize winner who won a bid to the Regeneron International Science and Engineering Fair (ISEF). More than 50% of the category awards were awarded to Tesla STEM students.

The following students received the following awards:

Behavioral and Social Sciences

- **1st place:** Prisha Agarwal, Self-esteem and Social Anxiety of South Asian Adolescents with Varying Acne Severity*
- **2nd place:** Pranay Nookala and Tejas Pakalapati, EEG-analysis of Brain Activity in Individuals with Varying Levels of Social Media Usage*
- **3rd place:** Samantha Novitski, The Relationship Between Different Genres of Media Consumed to Burnout Symptoms Reported
- **4th place:** Saachi Sharma, Quantifying Depression: An art-based approach to reducing the self-report bias in diagnosing depression using Koch's Baum Test and color theory

Biomedical Engineering and Translational Health

- **3rd place:** Mitra Junu, Pollen Coated Microneedle Patch to Desensitize Patients to Pollen Allergies Using Immunotherapy
- **4th place:** Anushka (Annie) Adhikary, NEURAID: Identification of Postnatal Presymptomatic

Neuroimaging Biomarkers for Autism Through Recurrent Neural Network-Based Image Analysis of Functional and Structural T1 MRI Data Towards Enhanced Autism Diagnosis

Computational Biology & Bioinformatics (COMP)

- **2nd place:** Shivani Sama*, Reconnect: A Computational Approach to Using DBS to Combat Neurodegenerative Disease Through Induced Plasticity

Biology Sciences

- **2nd place:** Aditi Kalluru*, Impact of Vitamin C on Chemically Induced Mouse Tumor Cells of Squamous Cell Carcinoma
- **3rd place:** Eshaal Omar, Evaluation of horse discomfort in English and Western riding style tack

Environmental Sciences and Technology, Energy, Math & Design

- **2nd place:** Sophia Callo, Red Light for Sleep Quality and Performance in Gymnasts
- **3rd place:** Divij Aswinkumar, Using Machine Learning to Predict the Power Conversion Efficiency of Perovskite Solar Cells
- **4th place:** Aaryan Dhingra, Mycoremediation: Employing Aspergillus Niger to Remove Aluminum Ions From Soil Contaminated With Coal Incineration Fly Ash

Earth and Environmental Science

- **3rd place:** Willem Magee, Predicting PM2.5 Concentrations using Meteorological Data
- **4th place:** Saraswathy Amjith, A Novel Multimodal Approach to Illegal Logging Detection

Engineering Technology and Math Science

- **1st place:** Nikki Taleghani*, Carbon Dot and Cyanoacrylate Fuming Method for Latent Fingerprint Detection
- **3rd place:** Yutaro Urata, Generating Novel Bioplastics Using GAN
- **4th place**: Shriya Tiku, Ham Radio Antenna Communication Set-up for First Responder Communication during Disaster Relief

Physics, Chemistry, and Mathematics

- **2nd place:** Nikhar Khamesra, Bio-flocculants in Microplastic Water Filtration

Robotics & Intelligent Machines

- **2nd place:** Ansh Bayappu*, An Everting Vine Robot with a Pneumatic Soft Gripper
- **3rd place:** John Karaca, Ensemble-Based Learning Model for Deepfake Detection
- **4th place**: Srikrishna Gurumurthy, Development of a Machine Learning Model To Predict the Presence of Advanced Liver Fibrosis in Patients With Type 2 Diabetes Mellitus

Systems Software and Embedded Systems

- **2nd place:** Krish Katre*, Virtual Reality for ADHD: Using Technology to Enhance Learning and Attention
- **4th place:** Anisha Vaish, Automated Pill-Dispenser to Improve Medication Compliance Efficiency and Accuracy

...and congratulations to one of the five **GRAND PRIZE WINNERS** who won the extraordinary honor of competing at ISEF....**Nikki Taleghani** for her project on Carbon Dot and Cyanoacrylate Fuming Method for Latent Fingerprint Detection. Nikki will compete along with approximately 1,800 high school students from more than 75 countries, regions, and territories IN PERSON in Dallas, TX this May. Students at ISEF are awarded the opportunity to showcase their independent research and compete for on average $5 million in prizes.

2023 ExploraVision

From Instructor Rebecca Townsend:

Our anatomy/biomedical engineering classes had three groups earn honorable mentions in the national ExploraVision Competition this year. For ExploraVision, the students predict what a current technology will be like 10+ years in the future. There are approximately 2,000 teams competing each year. (They award one national winner, five regional winners, and honorable mentions per age level.)

- Using Virtual Reality to Treat ICU Psychosis (Bhavya Nandikanti, Tanvi Sapar, and Tanvi Anand)
- OHCT: One Hair Cell at a Time – Using Atoh1 Molecules to Reverse Hearing Loss (Madhu Rasiah, Aubrey Grove, and Nuha Rahman)
- HemoSyn: A Novel Approach to Immune Functionality in Hematology (Manovay Sharma, Ayush Singh, and Arihant Singh)

National Economics Challenge: Washington State Competition

From advisor Ms. Ryfka Schafer:

Four students from the Tesla STEM Economics Club competed in the National Economics Challenge and placed 3rd in WA State!

Vidit Khandelia
Rithvik Koppolu
Vaibhav Attre
Aahil Irfan

They've all won money, and they now move onto the semi-finals, competing against the top teams from every state. If they place in the semi-finals, they advance to the finals in-person in New York City with all travel costs covered by the National Council on Economic Education.

Washington Music Educators Association Honors

From Mr. Matthew Kruse:

The following students participated in the 2023 WMEA All-State and 2023 NAfME All-Northwest honor ensembles:

2023 NAfME All-Northwest:

Manami Chiba - 2023 NAfME All-NW Band - Flute

Chris Ma - 2023 NAfME All-NW Orchestra - Violin

Davis Yang - 2023 NAfME All-NW Orchestra - Bb Soprano Clarinet

2023 WMEA All-State:

Guhan Kumar - 2023 WMEA All-State Orchestra - Violin

Varsha Mantravadi - 2023 WMEA All-State Concert Band - Percussion

Nakul Rajpal - 2023 WMEA All-State Concert Band - Oboe

Sarah Youn - 2023 WMEA All-State Concert Band - Flute

2023 Washington State Technology Student Association (TSA) Competition

From advisor Ms. Bethany Kankelborg:

The Washington State Chapter of the Technology Student Association held it's competition in late March in SeaTac Washington. Tesla STEM High School's chapter had 64 students attend the event where they tested their creativity, communication, professionalism, and technical expertise against their peers from across Washington State. This year's Conference had over 4000 students from both middle school and high school attend, the largest in WTSA's 40 year history. All of our scholars brought their best work to their events and made their school proud. Below I have listed all the finalists and semi-finalists from our school some of whom will attend the National Competition in Louisville, Kentucky at the end of June.

In addition our own Tiana Dumitrescu was elected to serve as Washington TSA's Vice President for the 2023-24 school year.

Forensic Science:
- **2nd Place:** Anjana Punniamoorthy and Divij Aswinkumar
- Top 12: Aaryan Dhingra and Ritu Doshi

Coding:
- Top 12: Saraswathy Amjith and Rithvik Koppolu
- Top 12: Vasudha Narayanan and Emir Akture

Music Production:
- Top 5: Colin Xie

Digital Video Production:
- Top 12: Samyuktha Krishnan, Saahiti Mamidipaka, and Harshita Raja

Audio Podcasting:
- Top 5: Likitha Nanduri and Abhigna Nimmagadda
- Top 5: Sanjana Katti

Webmaster:
- **2nd Place:** Aryaman Rtunjay, Raghav Kalyana Sundaram, Pranav Kannepalli, Achintya Agrawal, and Shruthika Balasubramanian

Biotechnology:
- Top 5: William Bo, Aaryan Dhingra, and Ritu Doshi
- Top 12: Tiana Dumitrescu, Prakshi Shukla, and Lauren Wetzel

Data Science and Analytics:
- **1st Place:** Rishika Gautam

- Top 5: Shivani Sama and Annie Adhikary
- Top 5: Raghav Kalyana Sundaram and Aryaman Rtunjay

Architectural Design:
- Top 12: Vrinda Sankarakumar
- Top 12: Artur Sobol, Denys Davydenko, Willem Magee, and Grigory Shatalin

Board Game Design:
- **1st Place**: Brenda Li and Keerthana Radapuram
- Top 5: Laura Shang, Sanskriti Sinha, Pallavi Thiruveedu, and Shreeyaa Muhunthan

Engineering Design:
- **2nd Place:** Ekansh Arora, William Bo, Aaryan Dhingra, Farzad Hasan, and Saahil Menta
- Top 5: Divya Ramu, Shivani Sama, Shriya Tiku, and Annie Adhikary

Children's Stories:
- Top 12: Iris Dey and Vivian Li

Virtual Reality Visualization (VR):
- **2nd Place:** Vivaan Pradhan and Krishnaprasath Senthil Kumar
- **3rd Place**: Aryaman Rtunjay and Raghav Kalyana Sundaram

VEX Robotics:
- **1st Place:** Pranay Nookala and Farzad Hasan

Geospatial Technology
- **2nd Place:** Siddhant Porwal and Anusha Sharma
- Top 5: Prakshi Shukla, Lauren Wetzel, and Tiana Dumitrescu

Software Dev:
- Top 5: Ryan Chhabra, Rithvik Koppolu, and Saraswathy Amjith

2023 Tesla STEM Estimathon

On the perfect date (April 25), Tesla STEM High School hosted its first Estimathon. The competition was fierce, the answers were approximate and, in the end, the food and fame made it all worthwhile, especially for the winners:

1st Place: Reese Long (competing solo!)

2nd Place: Kshitij Rao, Andy Pham & Steven He

3rd Place: Kedar Venkatachalam, Jai Sainani & Shiven Abhyanker

In all, more than 50 students participated and a good time was had by all. Congratulations also to Pranav Teegavarapu, Evan Kim, and Sami Badal Khan, who organized the event, hosted it, and wrote the questions.

Future Business Leaders of America (FBLA) Regional Competition

Accounting I
- 2nd Anjana Punniamoorthy

Advertising
- 6th Anusha Pandey

Agribusiness
- 4th Sneha Murali
- 3rd Anana Punniamoorthy

Business Calculations
- 4th Johan Karukayil

Business Communication
- 1st Vasudaha Narayanan

Computer Problem Solving
- 4th Anshu Aggarwal
- 1st Roby Matta

Cyber Security
- 3rd Sarah Wen
- 1st Rithvik Koppolu

Economics
- 3rd Aahil Irfan
- 1st Vidit Khandelia

Health Care Administration
- 3rd Felix Hallmann

Human Resource Management
- 2nd Felix Hallmann

Insurance and Risk Management
- 3rd Johan Karukayil

Intro to Business Concepts
- 6th Kaushal Dabbiru
- 3rd Ekaansh Agrawal
- 2nd Winston Zhou

Intro to Business Procedures
- 1st Regan Hendrickson

Intro to Financial Math
- 4th Pranav Kannepalli

Intro to Information Technology
- 4th Regan Hendrickson
- 1st Pranav Kannepalli

Intro to Marketing Concepts
- 3rd Regan Hendrickson

Intro to Parli Pro
- 3rd Alekhya Nidadavolu

Networking Infrastructures
- 4th Rithvik Koppolu
- 2nd Rohit Karthik

Organizational Leadership
- 5th Rithvik Koppolu

Personal Finance
- 1st Vidit Khandelia

Political Science
- 3rd John Karaca

Securities and Investments
- 1st Vidit Khandelia

UX Design
- 6th Madeline Tsai
- 4th Achintya Agrawal
- 2nd Rohit Karthik

Spreadsheet Applications
- 3rd Vasudah Narayanan

Banking and Financial Systems
- 5th Pia Dave and Kaavya Trivedi

Business Management
- 1st Anusha Pandey, Shriya Tiku

Hospitality Management
- 6th Rajiv Arvind, Tejaswi Erattu

Impromptu Speaking
- 6th Kaavya Trivedi

Marketing
- 6th Divij Aswinkumar, Nimai Belur, Pranay Nookala
- 3rd Samyuktha Krishnan, Saranya Pammi

Sports and Entertainment Management
- 2nd Immaz Khan, Pranav Sayal, Chaithu Thandu

Broadcast Journalism
- 6th Divya Singh, Laya Nair
- 2nd Pia Dave, Kaavya Trivedi
- 1st Kshitij Rao, Anirudh Seshadri, Rohan Singh

Business Ethics
- 5th Vidit Khandelia, Rithvik Koppolu, Ayaan Kumar

Data Analysis
- 6th Pearly Agarwal, Gayathri Pillai
- 4th Arpan Agarwal, Roby Matta, Dhruv Naik

Digital Video Production
- 1st Kshitij Rao, Anirudh Seshadri, Rohan Singh

E-Business
- 3rd Divij Aswinkumar, Tejas Pakalapati
- 1st Aahil Irfan, Saathvik Somujayabalan

Electronic Career Portfolio
- 4th Luis Padilla
- 1st Harini Vijeyanandh

Graphic Design
- 1st Ahanya John, Grace Zhao

Intro to Business Presentation
- 6th Rachel Cai, Elaine Jiang, Sarah Youn
- 4th Varshini Hari, Layla Kaim

Intro to Social Media Strategy
- 6th Madison DeSilva, Shriya Kapila, Liya Sebastian

Job Interview
- 3rd Harini Vijeyanandh

Public Service Announcement
- 6th Vaishnavi Rayaprolu, Mehek Sathe
- 2nd Aparna Iyer, Likitha Nanduri, Abhigna Nimmagadda
- 1st Shruthika Balasubramanian, Megan Halim, Aishika Koppolu

Public Speaking

- 1st Cindy Chen

Publication Design
- 5th Likitha Nanduri, Divya Ramu, Shruthika Saravanakumar
- 3rd Karina Kejriwal, Aarav Khanna, Isayiah Lim

Sales Presentation
- 4th Aaryan Bondre, Tanush Contractor, Gyan Dave

Social Media Strategies
- 6th Rajiv Arvind, Abhinav Nandupalli
- 5th Anshu Aggarwal, Varsha Mantravadi, Prakshi Shukla
- 3rd Karina Kejriwa, Aarav Khanna, Bhavya Nandikanti
- 1st Rishi Dilip, Tashfiq Rashid, Sourish Singh

Website Design
- 5th Sarah Wen
- 3rd Anshu Aggarwal, Saathvik Somuhayabalan
- 1st Monisha Krothapalli, Shreya Pandey

Future Business Leaders of America (FBLA) State Leadership Conference

Students in the top 4 automatically advance to the national competition in Atlanta, GA this summer. Those in 5 – 10 have the potential to attend.

Presentations:

Banking and Financial Systems: 1st Pia Dave and Kaavya Trivedi

Broadcast Journalism: 5th Omid Faradahl, Laya Nair and Divya Singh; 3rd place Kshitij Rao, Anirudh Seshardi, Rohan Singh, 1st place Pia Dave and Kaavya Trivedi

Business Ethics: 10th Vidit Khandalia, Rithvik Koppolu and Ayaan Kumar

Business Financial Plan: 10th Abhinav Nadupalli and Alekhya Nidadavolu

Business Management: 4th Shriya Tiku and Anusha Pandey

Business Plan: 7th Aaryan Bondre, Gyan Dave and Karina Kejriwal

Coding and Programming: 6th Ayaan Hussain, Tejas Pakalapati and Siddarth Pallovola; 5th Divij Aswinkumar, Rishi Dilip and Vasudha Narayanan

Data Analysis: 10th Pearly Agarwall and Gayathri Pillai

Digital Video Production: 4th Kshitij Rao, Anirudh Seshardi and Rohan Singh

E-Business: 9th Aahil Irfan and Saathvik Somujayabalan

Electronic Career Portfolio: 4th Harini Vijeyanandh

Future Business Leader: 6th Tarun Mahadevan

Hospitality Management: 6th Rajiv Arvind, Tejaswi Erattu and Pranav Kannepalli

Impromptu Speaking 4th Kaavya Trivedi

Introduction to Business Presentation: 1st Varshini Hari and Layla Kaim

Introduction to Social Media 8th Madison Desilva, Shriya Kapala and Liya Sebastian

Job Interview 2nd Harini Vijeyanandh

Marketing: 8th Achintya Agrawal, Samyuktha Krishnan and Saranya Pammi; 3rd Rishi Dilip, Tashfiq Rashad and Sourish Singh; 1st Divij Aswinkmar, Nimai Belur and Pranay Nookala

Mobile App Development: 7th Krish Katre and Tejas Pakalapati, 2nd place Farzad Hasan, Mitra Junu and Rohit Karthik

Public Speaking 5th Cindy Chen

Publication Design: 5th Karina Kejrwal, Aarav Khanna and Isayiah Lim

Sales Presentation 2nd: Aaryan Bondre, Gyan Dave and Tanush Contractor

Social Media Strategies: 6th Anshu Aggarwal, Varsha Mantravadi and Prakshi Shukla, 2nd Rishi Dilip, Tashfiq Rashid and Sourish Singh, 1st: Aarav Khanna, Karina Kejriwal and Bhavya Nandikanti,

Sports Management: 8th Immaz Khan, Chaithu Thandu and Pranav Sayal

Spreadsheet Applications: 1st Vasudha Narayanan

Website Design: 2nd Anshu Aggarwal and Saathvik Somujayabalan, 1st place Shreya Pandey and Sarah Wen

Objective tests

Advertising 5th Anusha Pandey

Business Communications 4th Vasudha Narayanan

Computer Problem Solving 2nd Roby Matta

CyberSecurity: 8th Sarah Wen

Economics 2nd Aahil Irfan, 1st Vidit Khandalia

Introduction to Business Concepts: 9th Kaushal Dabbiru, 4th Winston Zhou

Introduction to Information Technology: 9th Regan Hendrickson

Introduction to Marketing Concepts 9th Regan Hendrickson

Introduction to Parliamentary Procedure 3rd Alekhya Nidadavolu

Networking Infrastructure 3rd Rithvik Koppolu

Personal Finance 1st Vidit Khandalia

Securities and Ethics 5th Vidit Khandalia

UX Design 6th Achintya Agrawal, 4th Madeline Tsai, 1st Rohit Karthik

2023 United States National Chemistry Olympiad

From Chemistry teacher Mr. Rick Herzog:

Congratulations to Sebi Kumar, Evan Kim, and Aashray (Chegu Vijay Aashray).

In March they took the United States National Chemistry Olympiad local Puget Sound Section test. Sebi and Evan qualified to move on to the next level, while Aashray was chosen as a runner up.

In April Sebi and Evan took the National exam. Sebi scored in the top 20 of about 1000 students from around the country and has been chosen to attend the Summer Study Camp which will be held June 4-17. The top four students will represent the US team in the International Olympiad.

From Jennie Mayer, the Education Committee Chair, American Chemical Society - Puget Sound Section: *"In the many years (a couple decades, I believe!) Prof Bhat has administered this program through the Puget Sound Section, no student from our section has held this honor before. We are very proud of Sebastian and his hard work and dedication."*

2023 ArcGIS Washington State Map Story Contest

This year, our entire 9th grade class created Map Stories for the annual ArcGIS Washington State Map Story Contest.

One team of seniors submitted a project that won 1st place:

- **Druhin Bhowal & Siddhant Porwal**, Fjord Estuaries: Washington's Hidden Gems (Link: https://bit.ly/3KSpD2o)

Four teams of 9th-graders also were recognized:

- **Madeleine Vanderheiden**, Impacts of Hydroelectricity Generation on Salmon (Link: https://bit.ly/41N97Y2)
- **Mina Liu**, Net-Zero (Link: https://bit.ly/3oxsiHf)
- **Ekaansh Agrawal**, Electrify Your Drive: The Future of Transportation (Link: https://bit.ly/41u3dvs)
- **Sree Sarvepalli**, Protect the Puget Sound (Link: https://arcg.is/3USC1DO)

In addition, the following project won 1st place in the Health Science division of the Washington Tracking Network Youth Science Contest:

- **Keerthi Rajesh & Mehuli Nath,** Lung Cancer in Washington State (Link: https://arcg.is/rLi4L)

Washington State Science and Engineering Fair (WSSEF)

(indicates qualification to ISEF):*

WSSEF 2023 Grand Champion Premiere Awards:
Nikki Taleghani, Tesla STEM, Grand Medalist Champion (Gold)
Annie Adhikary, Tesla STEM, Grand Medalist Champion (Silver)

WSSEF First Place Trophy

*Annie Adhikary, Tesla STEM High School (also winner of the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship and the Washington State University - College of Arts and Sciences Scholarship award)

*Yue Ding, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award and the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship award)

*Saachi Sharma, Tesla STEM High School (also winner of the Graphic Design Award, the American Psychological Association - Achievement in Research in Psychological Science Award, the US Army Awards for Outstanding STEM Accomplishments, the Wolfram Mathematica Software Award, and the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship award)

*Nikki Taleghani, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award and the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship award)

Arpan Agarwal, Tesla STEM High School (also winner of the USAID Science Champion Award and the

US Metric Association Award)
Saraswathy Amjith, Tesla STEM High School (also winner of the Ricoh Sustainable Development Award and the USAID Science Champion Award)
Divij Aswinkumar, Tesla STEM High School (also winner of the US Air Force Award, Wolfram Mathematica Software Award, Ohio Wesleyen University - Outstanding Achievement in Science Scholarship and the Voiland College of Engineering and Architecture at Washington State University Scholarship award)
Stephanie Haining, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Raneet Hans, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Farzad Hasan, Tesla STEM High School (also winner of the GENIUS Olympiad - International Project Fair on the Environment award)
Aparna Iyer, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award, the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship and the Voiland College of Engineering and Architecture at Washington State University Scholarship award)
Inaya Kamal, Tesla STEM High School (also winner of the St. Michaels Medical Center Biomedical Engineering and Health Science Award)
John Karaca, Tesla STEM High School (also winner of the Yale Science and Engineering Association Award for Most Outstanding Exhibit, the Sigma Xi Award, the Wolfram Mathematica Software Award and the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship award)
Aiden Morrella, Tesla STEM High School (also winner of the American Psychological Association - Achievement in Research in Psychological Science Award)
Vasudha Narayanan, Tesla STEM High School (also winner of the Olympic College - Excellence in STEM Awards)
Venkatapranay Nookala, Tesla STEM High School (also winner of the American Psychological Association - Achievement in Research in Psychological Science Award)
Eshaal Omar, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Sri Sai Tejas Varma Pakalapati, Tesla STEM High School (also winner of the American Psychological Association - Achievement in Research in Psychological Science Award)
Divya Ramu, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Sophia Rodriguez, Tesla STEM High School (also winner of the US Air Force Award and the GENIUS Olympiad - International Project Fair on the Environment)
Aryaman Rtunjay, Tesla STEM High School (also winner of the GENIUS Olympiad - International Project Fair on the Environment award)
Shivani Sama, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Shriya Tiku, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award and the Ohio Wesleyen University - Outstanding Achievement in Science Scholarship award)
Anisha Vaish, Tesla STEM High School (also winner of the Olympic College - Excellence in STEM Awards the SPEEA IFPTE Local 2001 Excellence in Engineering Award and the Wolfram Mathematica Software Award)
Yuqing Wang, Tesla STEM High School (also winner of the Wolfram Mathematica Software Award)
Minal Pervaiz, Tesla STEM High School
Johan Karukayil, Tesla STEM High School
Kenisha Jaiswal, Tesla STEM High School
Mitra Junu, Tesla STEM High School
Chegu Vijay Aashray, Tesla STEM High School
Prisha Agarwal, Tesla STEM High School
Talah Albadawi, Tesla STEM High School
Ansh Bayappu, Tesla STEM High School
Janelle Chen, Tesla STEM High School

Shravya Daggumelli, Tesla STEM High School
Tiana Dumitrescu, Tesla STEM High School
Eleanor Marso, Tesla STEM High School
Aylin Ozdemir, Tesla STEM High School
Moomal Siddiqui, Tesla STEM High School
Yutaro Urata, Tesla STEM High School
Milana Veera, Tesla STEM High School
Rowena (Ellis) Yin, Tesla STEM High School

WSSEF Second Place Trophy

- Pearly Agarwal, Tesla STEM High School (also winner of the Henrietta Lacks Award)
- Ashwin Kaliyaperumal, Tesla STEM High School (also winner of the Olympic College - Excellence in STEM Award)
- Aditi Kalluru, Tesla STEM High School (Society for in Vitro Biology Outstanding Achievement for Ability and Creativity Award)
- Nikhar Khamesra, Tesla STEM High School (also winner of the Ricoh Sustainable Development Award and the Water Environment Federation - US Stockholm Junior Water Prize)
- Shauryan Deb, Tesla STEM High School
- Jack Ghioni, Tesla STEM High School
- Srikrishna Gurumurthy, Tesla STEM High School
- Nitya Kakulamarri, Tesla STEM High School
- Krish Katre, Tesla STEM High School
- Soha Khan, Tesla STEM High School
- Zohaib Lughmani, Tesla STEM High School
- Varsha Mantravadi, Tesla STEM High School
- Michael Meleshenko, Tesla STEM High School
- Samantha Novitski, Tesla STEM High School
- Trevor O'Gorman, Tesla STEM High School
- Sebastian Predescu, Tesla STEM High School
- Anjana Punniamoorthy, Tesla STEM High School
- Siri Varkala, Tesla STEM High School

WSSEF Third Place Trophy

- Sophia Callo, Tesla STEM High School
- Shruthika Saravanakumar, Tesla STEM High School

WSSEF 2023 Special Awards Grades 9 – 12

American Psychological Association ~ Achievement in Research in Psychological Science Award

- Aiden Morrella
- Venkatapranay Nookala
- Sri Sai Tejas Varma Pakalapati
- Saachi Sharma

GENIUS Olympiad ~ International Project Fair on the Environment

- Farzad Hasan
- Sophia Rodriguez

Graphic Design Award

- Saachi Sharma

Henrietta Lacks Award

- Pearly Agarwal

Ohio Wesleyen University ~ Outstanding Achievement in Science Scholarship
- Annie Adhikary
- Divij Aswinkumar
- Yue Ding
- Aparna Iyer
- John Karaca
- Saachi Sharma
- Nikki Taleghani
- Shriya Tiku

Olympic College ~ Excellence in STEM Awards
- Ashwin Kaliyaperumal
- Vasudha Narayanan
- Anisha Vaish

Ricoh Sustainable Development Award
- Saraswathy Amjith
- Nikhar Khamesra

Sigma Xi Award
- John Karaca

Society for in-Vitro Biology Outstanding Achievement for Ability and Creativity Award
- Aditi Kalluru

SPEEA IFPTE Local 2001 Excellence in Engineering Award
- Anisha Vaish

St. Michaels Medical Center Biomedical Engineering and Health Science Award
- Inaya Kamal

U.S. Air Force Award
- Divij Aswinkumar
- Sophia Rodriguez

U.S. Army Awards for Outstanding STEM Accomplishments
- Saachi Sharma

U.S. Metric Association Award
- Arpan Agarwal

USAID Science Champion Award
- Arpan Agarwal
- Saraswathy Amjith

Voiland College of Engineering and Architecture at Washington State University Scholarship
- Divij Aswinkumar
- Aparna Iyer

Washington State University ~ College of Arts and Sciences Scholarship
- Annie Adhikary

Water Environment Federation ~ U.S. Stockholm Junior Water Prize
- Lakshmi Agarwal
- Nikhar Khamesra

Wolfram Research, Inc. Mathematica Software Award
- Divij Aswinkumar
- Yue Ding
- Stephanie Haining
- Raneet Hans
- Aparna Iyer
- John Karaca
- Eshaal Omar
- Divya Ramu

- Shivani Sama
- Saachi Sharma
- Nikki Taleghani
- Shriya Tiku
- Anisha Vaish
- Yuqing Wang

Yale Science and Engineering Association Award for Most Outstanding Exhibit

- John Karaca

International Science and Engineering Fair (ISF)

Annie Adhikary: Grand Award - 3rd place Award of $1000 in Translational Medical Science
Special Award - $5,000 scholarship to the University of Texas, Dallas

Identification of Novel Diagnostic Neuroimaging Biomarkers for Autism Spectrum Disorder Through Convolutional Neural Network-Based Analysis of Functional, Structural, and Diffusion Tensor Imaging Data Towards Enhanced Autism Diagnosis

Annie Adhikary, 16, Junior, Nikola Tesla STEM High School, Redmond, Washington,
Teacher-Sponsor: Kate Allender

Annie was the WSSEF silver grand champion, representing WSSEF at ISEF. Also at WSSEF she won a first place trophy, a $15,000 scholarship to Ohio Wesleyan University and a $2000 scholarship from the WSU College of Arts & Sciences. At ISEF she won a 3rd Grand Award of $1000 in Translational Medical Science for her project. She also won a $5000 renewable scholarship to the University of Texas at Dallas.A paper based on her project was accepted for the 16th International Conference on Brain Informatics, and she was invited to be a panelist. The paper is in a second round of peer review for the journal Molecular Autism.

Yue (Beth) Ding: Grand Award: 3rd place in computational biology and bioinformatics

Early Detection of Cancer Using Microbial Taxonomic and Functional Information

Yue Ding, 17, Junior, Nikola Tesla STEM High School, Redmond, Washington,
Teacher-Sponsor: Kate Allender

At WSSEF, Yue Ding was named one of seven ISEF finalists. Also at WSSEF, she won a first place trophy and a $15,000 scholarship to Ohio Wesleyan University. At ISEF she won a 3rd grand award of $1000.

Saachi Sharma: Grand Award: 4th place in behavioral and social sciences
Special Award: $10,000 scholarship to the University of Texas, Arlington

Quantifying Depression: An Art-Based Approach to Reducing the Self-Report Bias in Diagnosing Depression Using Koch's Baum Test and Color Psychology

Saachi Sharma, 16, Junior, Nikola Tesla STEM High School, Redmond, Washington,
Teacher-Sponsor: Kate Allender

At WSSEF, Saachi Sharma was one of seven ISEF finalists. Also at WSSEF she won a first place trophy, the American Psychological Association Award, the Graphic Design Award, a US Army Award, and a $15,000 scholarship to Ohio Wesleyan University. At ISEF, she won a 4th Grand Award of $500 and a $10,000 scholarship to the University of Texas, Arlington.

Nikki Taleghani: Special Award- 1st Place from National Security Agency Research Directorate

Carbon Dot and Cyanoacrylate Fuming Method for Latent Fingerprint Detection

Nikki Taleghani, 16, Junior, Nikola Tesla STEM High School, Redmond, Washington,
Teacher-Sponsor: Kate Allender

Nikki was the WSSEF top gold grand champion, winning a trip to ISEF. Also at WSSEF she won a first place trophy, and a $15,000 scholarship to Ohio Wesleyan University. At ISEF she won a first place award from the National Security Agency Research Directorate. At the Central Sound Fair she was one of the 5 grand champion awards, and 1st place in engineering and materials.



International Science and Engineering Fair (ISF)



Naman Mutalik Desai recipient of the **2023-24 Mike Yarrow Peace Fellowship** for improving soil health and reducing fertilizer pollution

2023 Award for Aspirations in Computing (AiC) NCWIT

Varshini Hari: Affiliate Honorable Mention;
Anya Vaish: National Honorable Mention, Affiliate Winner;
Rithani Saravanakumar: National Honorable Mention, Affiliate Winner;
Nitya Kakulamarri: Affiliate Honorable Mention;
Anisha Vaish: National Honorable Men

The Washington Tracking Network (WTN) Youth Science Contest

is an opportunity for high school students in Washington state to develop their science and communication skills by engaging with health and environmental data from their own communities.

Students chose between three tracks: Health Science, Community Engagement, and Science Communication. All projects were required to use WTN data and have an equity analysis.

Together **Keerthi Rajesh** and **Mehuli Nath** won **1st place** in the Health Science track – Group division.

Holocaust Center for Humanity's Annual Art/Film/Writing Contest

Over 50 schools entered. We swept the digital art and film categories and won 3 out of the 4 awards for physical art. Two essays were entered, and one of them won.

Digital Art 9th-12th Grade:
- 1st Place: Brian Yin, Grade 12, Tesla STEM High School
- 2nd Place: Davis Yang, Grade 12, Tesla STEM High School
- 3rd Place: Deepalakshmi Thirumal, Grade 12, Tesla STEM High School
- Honorable Mention: Evan Kim, Grade 12, Tesla STEM High School

Physical Art 9th-12th Grade:
- 1st Place: Shelly Netz, Grade 12, Tesla STEM High School
- 3rd Place: Aaryan Bondre, Grade 12, Tesla STEM High School
- Honorable Mention: Triana Fernando Grade 12, Tesla STEM High School

Argumentative Writing 9th-12th Grade:
- 3rd Place: William Gong, Grade 12, Tesla STEM High School

Film 9th-12th Grade:
- 1st Place: David Gershony, Grade 12, Tesla STEM High School
- 2nd Place: Ananya Unnikrishnan, Grade 12, Tesla STEM High School
- 3rd Place: Ethan Yu, Grade 12, Tesla STEM High School
- Honorable Mention: Arihant Singh, Grade 12, Tesla STEM High School

Technology Student Association (TSA) State National Competition and Conference

Louisville Kentucky June 28th – July 2nd 2023

24 Students Attended the Conference accompanied by Ms. Kankelborg and Mr. Boyd.
The following teams/students made top 10 in their events bringing them up on stage at the award ceremony:

- Biotech Design - William Bo, Aaryan Dhingra, and Ritu Doshi – 10th place
- Board Game Design - Brenda Li and Keerthana Radapuram) – 4th place
- Data Science and Analytics – Rishika Gautam – 1st place
- Engineering Design – William Bo, Aaryan Dhingra, Farzad Hasan, and Saahil Menta – 4th place
- Forensic Science – Vasudha Narayanan and Anjana Punniamoorthy – 6th place
- Software Development – Siddhant Porwal and Anusha Sharma – 4th place
- Webmaster – Achintya Agrawal, Shruthika Balasubramanian, Raghav Kalyana Sundaram, Pranav Kannepalli, and Aryaman Rtunjay – 1st place



Technology Student Association (TSA) State National Competition and Conference

Tesla STEM Expo

Congratulations to the following finalists for advancing to our live event:

Research Track

- The Effect of Vitamin C on the Growth of Mouse Squamous Carcinoma Cells – Aditi Kalluru
- A 3D Printed Scoliosis Brace Designed to Increase User Comfort and Mobility – Inaya Kamal
- DHEBNet – a Novel Deep Heterogeneous Ensemble-Based Learning Model for Deepfake Detection – John Karaca

Startup Track

- Alumina – Ava Hagwell & Cindy Chen
- Budget Bites – Angelina Durbin & Ayshwaryaa Hariharan
- Lingwa – Adrian Brunke, Shelly Netz & Tyler Running
- Piper's Kitchen – Kiki Kalisova & Diya Karthic

These students will be competing for prizes gifted by our generous sponsors, foundry10 & Taskade.

4th Annual STEM Expo Live Event

Congratulations to our winners from this year's STEM Expo:

Research Track
- 1st Place: Aditi Kalluru – The Effect of Vitamin C on the Growth of Mouse Squamous Carcinoma Cells
- 2nd Place: John Karaca – DHEBNet - a Novel Deep Heterogeneous Ensemble-Based Learning Model for Deepfake Detection
- 3rd Place: Inaya Kamal – A 3D Printed Scoliosis Brace Designed to Increase User Comfort and Mobility

Startup Track
- Budget Bites (Angelina Durbin and Ayshwaryaa Hariharan) earned Best Overall Startup for their service of providing cheap, sustainable meals for college students while also minimizing food waste in grocery stores.
- Piper's Kitchen (Kiki Kalisova and Diya Karthic) earned the Entrepreneurship Award for their reusable, eco-friendly, and easy to clean piping bag (used for baking and cooking).
- Lingwa (Adrian Brunke, Shelly Netz, and Tyler Running) earned Audience Favorite for their mobile application that helps you reach fluency in a language.

A special thanks to Kate Allender for your collaboration on this event, Mike Keefer and Martha Perez on helping us transform the West Commons into an event space and back again, Brandon Inoue for helping us out with the sound/audio system, Cindy Duenas and Sumeyye Cardakli for giving these students an opportunity to share their work with the Tesla STEM Community, and one final thanks to the seniors in the STEM Expo Organizing Committee (Hahwon Chue, Saishree Devineni, Triana Fernando, Sidharth Ganesh, Nakul Rajpal, Anirudh Seshadri, and Hrudai Thungathurthi).

Our sponsors were foundry10 (Bronze donor who funded $1000 for the Startup track and $1000 for the Research track) and Taskade who gave different tiers of subscriptions to only the Startup track. Please see below for what each winner of the STEM Expo won:

Startup Track
- Budget Bites (Angelina Durbin and Ayshwaryaa Hariharan) earned Best Overall Startup and received $500 cash and Taskade Lifetime Upgrade
- Piper's Kitchen (Kiki Kalisova and Diya Karthic) earned the Entrepreneurship Award and received $300 cash and Taskade 5-Year Upgrade
- Lingwa (Adrian Brunke, Shelly Netz, and Tyler Running) earned Audience Favorite and received $200 cash and Taskade 5-Year Upgrade

Research Track
- 1st Place: Aditi Kalluru received $500 cash
- 2nd Place: John Karaca received $250 cash
- 3rd Place: Inaya Kamal received $250 cash



STEM Expo-Startup Winners



STEM Expo-Research Winners



National Merit Scholarship Finalists

Chegu Aashray
Anshu Aggarwal
Saraswathy Amjith
Nimai Belur
Shauryan Deb
Aparna Iyer
John Karaca
Rithvik Koppolu
Chris Ma
Varsha Mantravadi
Vasudha Narayanan
Aylin Ozdemir
Anusha Pandey
Divya Ramu
Saathvik Somujayabalan
Colin Xie



2023 WA State AAUW High School Scholar Recognition



Team Stem from Tesla STEM High School wins Science Olympiad NW Regionals



2023 FBLA State Leadership Conference