Nikola Tesla STEM High School

Redmond, Washington

Research Journal

2021-2022

Introduction

We are proud to present the seventh edition of our annual student research publication. The Nikola Tesla STEM Research Journal enables STEM 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 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: Debi Welcome
Journal Promotion: Kate Allender, Rebecca Townsend
Journal Support: Debi Welcome , Karen Schaeffer, Andy Christensen, Kate Allender, James Boyd

To view current and past Research Journals online go to 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: Cindy Duenas
Associate Principal: Cynthia Burt

Table of Contents

Biomedical

3D Printing Bio-Inspired Heart Valves to Increase Efficiency of Perioperative Care. 1

Bioinformatic Analysis to Find a Universal Sequence From SARS-CoV-2 Genomes for the Potential Development of a Pancoronavirus Vaccine. 7

A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease 16

Parkinson's Disease: A Risk Prediction Model Using Inflammatory Bowel Disease 28

Engineering/Machine Learning

Dampening Phantom Shockwave Jams In Mixed-Autonomy Traffic Via Deep-Rl Control Of Autonomous Vehicles 34

ScGAN: A Generative Adversarial Network to Predict Hypothetical Superconductors. 42

Machine Learning Guided Selection of Bio inks for 3D Bioprinting of Tissue Scaffold 54

Analysis of Ring Galaxies Detected Using Deep Learning with Real and Simulated Data 61

Environmental Science

Rubber Tire Leachate Effect in Lake Washington and Lake Sammamish 77

Predicting Waterfall Forms Through Feature Engineering and Model Selection Machine Learning. 83

It's Flaming Out: Using Artificial Intelligence to Emulate Critical Aspects of Wildfire Growth. 99

Engineering a Surface Flow Constructed Wetland to Evaluate Efficiency for Combating Water Scarcity. 114

Cellular Biology/Chemistry

Creating a Structural Model of a saSrtA-Inhibitor Complex to Predict Efficiency of Class A Sortase Enzyme Inhibitor Molecules. 123

Overexposure to maternal cortisol during gestation and its influence on the regulation of HPA axis in female adults 128

Oxytocin as a Strategy to Enhance the Effectiveness of the Satiation Signal, Cholecystokinin, to Reduce Food Intake in High Fat Diet-Fed Rats 132

Curing Cancer by Cancelling Cell Growth. 141

Behavioral and Social Sciences

Evaluating the Relationship Between Stress and Impulsivity in a Mathematical Contex 144

Combatting Depression using the Gut-Brain Axis via Probiotic Treatments. 152

Development of the Educator Implicit Association Test. 157

Assessment of Yoga Therapy or Mindfulness-Based Therapy Interventions on Symptom Severity and Quality of Life in a Cohort of Adults with Inflammatory Bowel Disease. 165

The Effect of a Cognitive Behavioral Therapy App (Happify) on Adolescent Anxiety 177

A Scientific Look into Mental Visualization and Improved Performance in Taekwondo. 183

Stand up to Stigma: ADHD in Women. 188

The Impact of Credibility When Hitting "Purchase Now". 193

Using the Art of Kuchipudi to Treat Symptoms of Attention Deficit Hyperactivity Disorder. 200

Awards and Achievements 209

Biomedical

3D Printing Bio-Inspired Heart Valves to Increase Efficiency of Perioperative Care

Rhea Kuppa

Nikola Tesla STEM High School

Abstract:

The efficient use of 3D printed heart valves for the use of preoperative practice has hit an all-time low, leading to an influx of failed surgeries in the cardiothoracic field. To improve preoperative practice on heart valves there needs to be a resourceful and cheap way to do so. This project aims to design a product that uses the most bio-realistic and cost-effective material possible to make 3D printed heart valves that will be used for preoperative practice. The valves were printed using three affordable bio-realistic materials - Carbon Fiber, TPU 90a, PLA (independent) - for desktop 3D printers, then were tested using multiple breakage tests - tensile testing, polymerized density, and shore hardness A. Using predetermined efficiency scales, the effectiveness of the prototypes was found (dependent). The efficiency scale values were averaged across all tests, and the best prototype material was established (TPU 90a). The engineering goal stated that the ideal prototype would have a score of 7 or above (1-10) on the scale once averaged and that there would be a difference in efficiency for each material. The experimental results supported the alternative hypothesis (the prototype materials would have a differing efficiency scores), with the TPU 90a having an efficiency score of 8.33 with a p-value of 0.00014 and F-statistic number of 54.17014 proving the results were statistically significant. This experiment has found an ideal material for preoperative testing that is both optimally efficient and cost effective.

Introduction

This project aims to design a product that uses the best material possible to make these bio-printed heart valves, the valves should closely resemble an actual heart valve and make the efficiency of preoperative planning much higher. This project prompts these two questions: “How will the use of different 3D PolyJet materials increase the efficiency of preoperative care of patients with heart valve diseases and disorders?” and “What materials and combinations of materials provide the best model that is the most like realistic heart valves?” Heart valve disease wreaks havoc across the world – every year over 5 million Americans are diagnosed (ACC) causing them to live a life of constant surgeries and procedures. In most cases, one or more of the four heart valves don’t develop properly leading them too not close or open properly (ACC). This causes the blood flowing through the heart to be disrupted. This can be prevented through doctors, surgeons and manufacturers who just need the right tools. Utilizing preoperative planning is the solution to this problem. Bioprosthetic heart valves can be used for surgeons to practice performing on, manufacturers to test their products and more. This will increase the efficiency of recovery from the treacherous heart diseases, and in some cases has proven to have a 30% increase in the success of the later surgery (CCF).

Most current bioprosthetic pieces are accompanied with a multitude of problems and are notorious for not acting like a real heart valve. This means the doctor is required to “guess and check” their work as they preoperatively plan, and heart valve prosthetic manufacturers need to estimate the effectiveness of their product. Research on this subject is incredibly important, it is necessary to build and design a product that can easily be used to help with preoperative care, while sustaining an affordable price. Approximately 75 hospitals in the US and 200 around the world have access to advanced 3D printers (Storrs 2015), with these limited resources it is important to make 3D printing as simple and commercial as possible.

For the effects of efficient preoperative care to be felt around the world it is vital to have a resourceful and cheap way to do so. A recent study, conducted by Dr. Rajat Vashishtha, aims to find an efficient way to make and utilize 3D printed heart valves in hospitals around the world. Specifically, he researched how different bio-ink materials could change the way a surgeon may be able to practice and test devices on. A similar experiment was studied by Dr. Kaiyan Qiu, where his team and him noticed the various complications that comes from doctors limited knowledge of the set they are operating on. To combat this, they printed bio-inspired heart valves with a flexible materials and targeted ways they could reduce postoperative complications. Similarly, Dr. Jinhua Li and his colleagues tested multiple hydrogels materials to find the most compatible and like real heart valve. This project aims to design a product that uses the best material possible to make these bio-printed heart valves, the valves should closely resemble an actual heart valve and make the efficiency of preoperative planning much higher. To find the best solution for the lack of effective and bio-related heart valve prototypes for surgeon, doctors and manufactures, the prototype should best resemble a real biological heart valve to increase the effectiveness of the practices done on it. This can be done by testing multiple materials that resemble the texture and tensile strength of a real biological heart valve. The ideal prototype will have a raw score of 7+ when all scores are averaged. The scores range from (1-10), from poor performance to ideal performance and the highest score will be 7.

Methodology

This project aims to analyze the effectiveness and usability of various heart valve manufactured with multiple materials. All these materials are biocompatible, and to an extent resemble the texture and malleability of a bio-realistic heart valve. All the materials with collected through online resources, all of which are cheap and readily available commercially. A CAD (Computer Aided Design) model was collected from an online resource as well (https://3dprint.nih.gov/discover/3dpx-000452) which is used as the model all materials will print. Using the FreeCad software, the first material is used and printed, TPU 90 A. Once printed, a model is acquired. When finished, a silicone coating is added if applicable to the materials (reference data table). Then the same steps are repeated with all other materials.

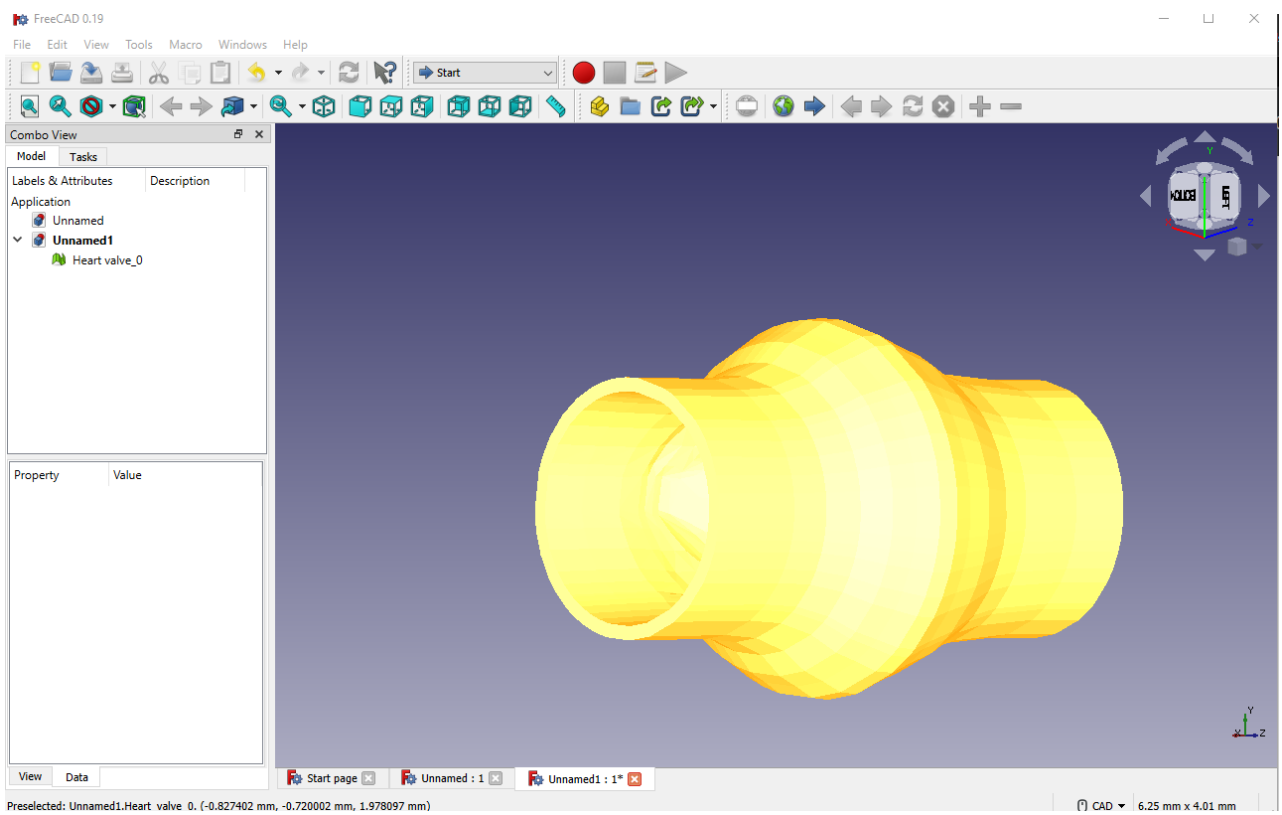


Figure 1 Downloaded CAD design of simple heart valve in FreeCad software, to be printed

To test the prototype, the experimenter must follow the steps illustrated in the flow chart below with each of the materials.

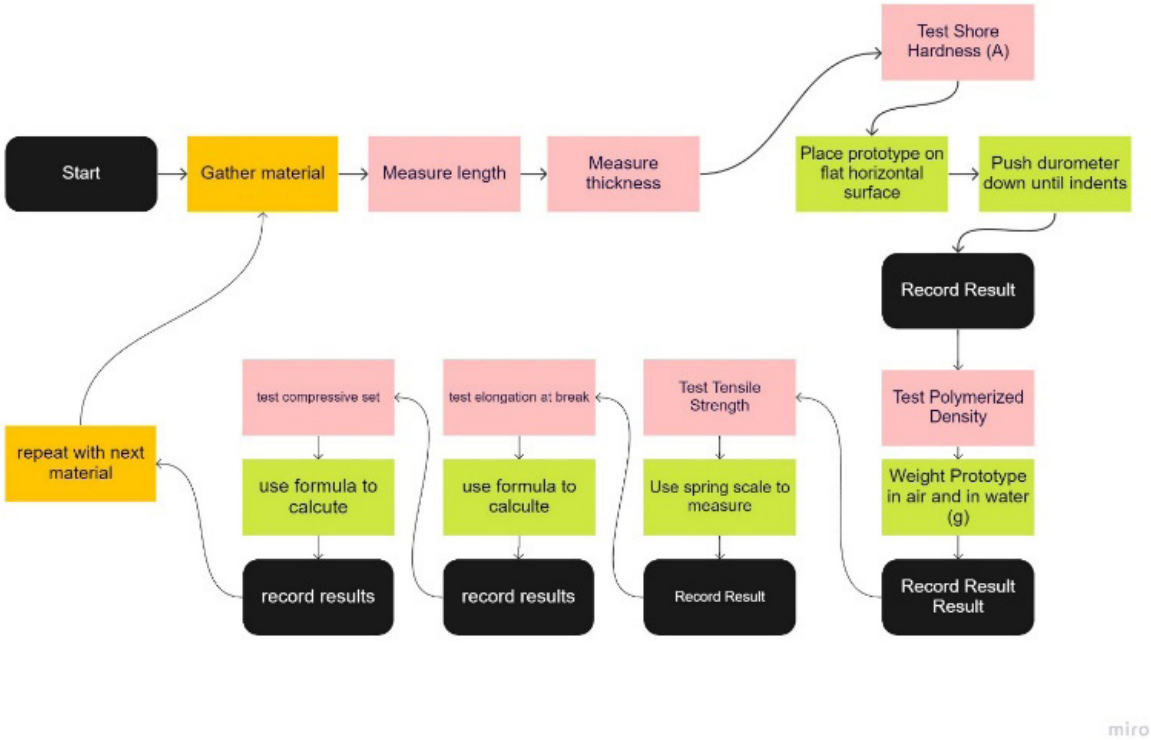


Figure 2 above is a flowchart to be used by experimenter while performing, simplified

The experimenter will start with the first prototype, Tango Plus FullCure 940. The lengths, thickness is measured and recorded in the data table below. Then full testing commences, a shore hardness test (A) is conducted. Shore hardness is a measure of resistance a material has to indentation and is usually used for soft polymers such as the testing materials for these prototypes. The prototype is placed in a flat, hard horizontal surface, like a table. A durometer, that is held between two hands is gently placed against the materials until an indentation is made. The number the durometer produces is recorded (Table 2). After this, the polymerized density test is commenced. Polymerized density is the sum of the mass fractional density of the polymer. First the prototype is weighed on scale in the air, then once immersed in water. The water weight is then dived by the air weight to get the polymerized density. This is then recorded. Next, the tensile strength of the prototype is measured in PSI. Tensile strength is the resistance of the prototype under tension. This break the protype and must be done after all prior tests have been conducted. A short string, the length of the prototype is secured to the prototypes and the hook a spring scale. The top of the spring scale is secured to a wall, and the porotype is pulled until it breaks. The force is the materials measured by the spring scale is recorded. The lengths and the thickness (measured by use of a caliper) before breakage are also recorded. These steps are repeated 3 times with the same materials and prototype, and then done again rotating through each material.

This marks the end of the testing, and now statistical procedures are conducted. First, the elongation at break, the ratio between the changed initial length and the initial length after the prototype, is calculated and recorded using this formula:

$$\text{Elongation} = \left(Delta\frac{L}{L}\right) * 100$$

Where L is initial length and Delta L is final length. Sequentially, the compressive set A is

measured and recorded. Compressive Set A is the percentage of the original prototypes thickness after 30 minutes of the prototype of being printed. This is the formula used:

$$C_A = [t_o - t_1)/t_0] * 100$$

Where t_o is the original prototype thickness and t_1 is the thickness after testing. The statistical analysis is done with every prototype and inserted into the data tables below.

Data Tables

Score Distribution:

Table 1: Score distribution across all materials.

Material:	Score: 10	9	8	7	6	5	4	3	2	1
Tensile Strength (PSI)										
Elongation at Break (%)										
Compressive Set (%)										
Shore Hardness(A)										
Tensile Tear Resistance (lb/in)										

Notes: Score is based off previous testing done in procedure

Table 2: Data table is split into multiple sections, but all data is applicable to the same experiment

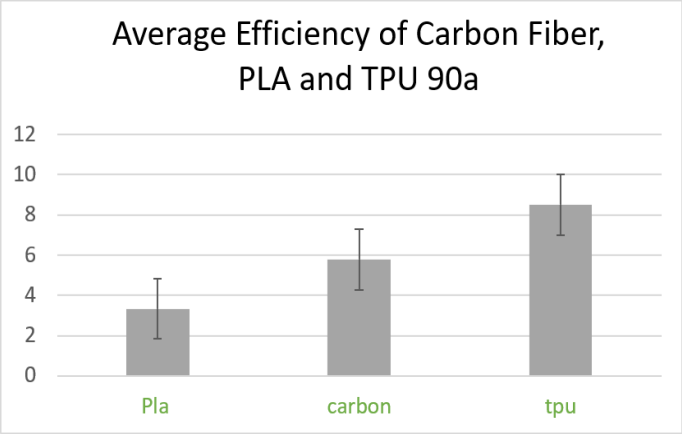
Material	Post Printing Material (if applicable)	Tensile Strength (PSI)	Elongation at Break (%)	Compressive Set (%)	Shore Hardness(A)
Example Material:	Silicone Coating	T1:115PSI T2: T3:	T1: 45% T2: T3:	T1: 1.5% T2: T3:	T1:60-62 T2: T3:
Score:					

Polymerized Density (g/cm)	Qualitative: Effectiveness	Average Score:
T1:1.61 g/cm **T2:**	T1: T2:	
T3:	T3:	

Data Analysis and Results

The averages for each of the materials were averaged one again and graphed as pictured in this bar graph. It's apparent that TPU 90a has the highest efficiency, but it is important to prove that these results are statistically significant To do this, an ANOVA test was conducted between all three materials. This resulted in an f-statistic number of 54.1701, the higher the f-statistic is, the better the model, therefore suggesting high reliability of my results. Additionally, this test produced a p-value of 0.00014, a statistically significant p value would be 0.05 and below. Again, this extremely low p-value shows that the probability of my null hypothesis, which is that all the materials are the same efficiency, being incorrect is incredibly low. Unfortunately, the ANOVA test can only produce a p-value for 2 of the 3 materials tested. It only lets the conductor of the experiment that there is in fact statistical significance but not between which 2 materials, to combat this, 2 more T-tests were conducted. A significant p -value was found between PLA and Carbon fiber, and Carbon fiber and TPU. This means there must be a significance between PLA and TPU as well. This solidifies the results and makes it safe to predict that TPU 90a does indeed have the highest efficiency out of all the materials and there for is the best for pre-operative practices. The expected outcome stated that the most efficient material will have a score of 7 or more. This was proven to be correct when TPU had an efficiency score of 8.5

In all, all engineering goals were achieved. The engineering goals were that this project would be able to build and design a prototype using various commercial materials that can easily be used to help with preoperative care (surgical and manufactural), while sustaining an affordable price. Additionally, the valves should closely resemble an actual heart valve and make the efficiency of preoperative planning higher. All the materials tested were commercially and relatively cheap ranging from 90-120 dollars, therefore, a material was found that is commercially available, cost effective, and efficient. The materials resemble a heart valve proven by the data, based scale value. This project had significant findings, suggesting high reliability in my results. And finally, found extensive testing's result in a replicable model that will inherently increase success rate of surgeries.



References

Li, J., Wu, C., Chu, P. K., & Gelinsky, M. (2020, February 18). *3D printing of Hydrogels: Rational Design Strategies and emerging biomedical applications.* R. Retrieved December 4, 2021, from https://www.sciencedirect.com/science/article/pii/S0927796X20300012.

Maxson, E. L., Young, M. D., Noble, C., Go, J. L., Heidari, B., Khorramirouz, R., Morse, D. W., & Lerman, A. (2019, August 14). *In vivo remodeling of a 3D-bioprinted tissue engineered heart valve scaffold.* Bioprinting. Retrieved November 30, 2021, from https://www.sciencedirect.com/science/article/pii/S2405886619300107.

Tappa, K., & Jammalamadaka, U. (2018, February 7). *Novel biomaterials used in medical 3D printing techniques.* Journal of functional biomaterials. Retrieved December 3, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5872103/.

Tuncay, V., & van Ooijen, P. M. A. (2019, February 15). *3D printing for heart valve disease: A systematic review.* European radiology experimental. Retrieved December 3, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6377684/.

Vashistha, R., Kumar, P., Dangi, A. K., Sharma, N., Chhabra, D., & Shukla, P. (2019, February 6). *Quest for cardiovascular interventions: Precise modeling and 3D printing of Heart Valves*. Journal of Biological Engineering. Retrieved December 3, 2021, from https://jbioleng.biomedcentral.com/articles/10.1186/s13036-018-0132-5.

Bioinformatic Analysis to Find a Universal Sequence From SARS-CoV-2 Genomes for the Potential Development of a Pancoronavirus Vaccine

Anya Vaish

Nikola Tesla STEM High School

Abstract

COVID-19 is caused by the SARS-CoV-2 virus. Present RNA vaccines from Pfizer/BioNTech BNT162b2 and Moderna mRNA-1273 are more than 94% successful in preventing infection. However, the virus is prone to mutation as demonstrated by the emergence of Delta and Omicron variants. The spike protein of the virus is essential for the interaction and internalization of the virus in the host cell and is considered a prime target for vaccine development against the SARS virus. Highly conserved sequences of the spike protein or other sections of the viral genome may be useful targets for the development of a pancoronavirus vaccine, which may be effective against future SARS-CoV-2 strains and other coronavirus families. Bioinformatic analysis of 258,269 full-length SARS-CoV-2 genomic sequences in the NCBI database was carried out using a custom Perl Script. All sequences were compared to the spike protein and full-length viral genome reference to find 100 nucleotide-long segments that were at least 99% conserved across SARS-CoV-2 sequences. The analysis resulted in a >99.5% conserved 114-nucleotide segment on the spike protein and a 99.49% conserved 104-nucleotide segment on the non-spike protein section of the viral genome across 258,269 SARS-CoV-2 genomes. The resulting conserved sequences from this study may be useful in developing an RNA or protein vaccine that could be effective against future SARS-CoV-2 strains or could act as a pancoronavirus vaccine if these sequences are present in other coronavirus families. This method can also be used to find conserved sequences across all coronavirus families.
Keywords: S-protein, spike protein, conservation, pancoronavirus vaccine

Introduction

This study aims to identify highly conserved SARS-CoV-2 RNA sequences that can potentially be used to develop a pancoronavirus vaccine. Coronavirus disease 2019 (COVID-19) is caused by the SARS-CoV-2 virus, which is an RNA virus. As of April 2022, over 500 million people have been infected by SARS-CoV-2 and over 6.1 million have died worldwide (World Health Organization, 2022). Several vaccines for SARS-CoV-2 were quickly developed in response to the outbreak of COVID-19 (Krammer, 2020; Nagy and Alhatlani, 2021). Among these vaccines, two mRNA vaccines from Moderna (BNT162b2) and Pfizer/BioNTech (mRNA-1273), both of which encode the prefusion-stabilized full-length spike glycoprotein of the SARS-CoV-2, are widely used and have shown more than 94% efficacy against symptomatic COVID-19. Other vaccines have shown 60-80% efficacy against symptomatic COVID-19 (Corbett, 2021). However, the virus is prone to mutation, which is demonstrated by the emergence of new variants including Delta and Omicron variants. Current vaccinations may not be as effective against new strains (Pegu et al., 2021; Cao et al., 2022). Therefore, the development of a universal vaccine, which may be effective against future strains of coronaviruses, is highly desirable.

(Dai et al., 2020; Koff & Berkley, 2021) SARS-CoV-2 is a β-coronavirus that is closely related to SARS-CoV-1 and distantly related to MERS-CoV, both of which have previously caused epidemics and are still serious threats to human health, as well as distantly related to common-cold coronaviruses (Premkumar et al., 2020). A pancoronavirus vaccine that covers all strains of SARS-CoV and MERS viruses will be even more desirable to quickly respond to the emergence of new diseases (Cohen, 2021). A pancoronavirus vaccine could potentially be developed by targeting a highly conserved sequence from SARS-CoV-2, as this vaccine would be able to target the same location in multiple coronavirus variants and therefore be at least somewhat effective against them. Conservation analysis can be performed to identify certain conserved sequences for this purpose. Previous studies have performed conservation analysis on SARS-CoV-2 as well as on other coronaviruses. A study conducted conservation and phylogenetic analysis to trace the evolutionary history of SARS-CoV-2 to determine that a bat coronavirus, rather than a pangolin coronavirus, was more likely to be the lowest common ancestor of SARS-CoV-2 (Lei & Zhang, 2020). Another study identified larger regions of genomes of betacoronaviruses lineage B, a group that includes SARS-CoV-2 and SARS-CoV, that were conserved, such as the 3'-UTR and 5'-UTR (Chan et al., 2020). However, since the 3'-UTR and the 5'-UTR are untranslated regions, they cannot be used for the development of a vaccine. Neither of these studies identified specific nucleotide sequences that are highly conserved and can be targeted by a vaccine. Another study performed phylogenetic network analysis using the median joining network algorithm to trace the evolution of SARS-CoV-2, which is useful since tracing the movement of SARS-CoV-2 variants can help researchers predict how future variants will behave (Forster et al., 2020). This study did not relate their findings to the development of a pancoronavirus vaccine. One study analyzed 3,132 viral protein sequences across multiple families of coronavirus using sequence alignment and identified several 9-amino acid epitopes with 89% exact match in the spike protein region that included two epitopes identified from recovered COVID-19 patients (Li et al., 2021).

This project seeks to analyze the more than 250,000 unique SARS-CoV-2 genomes submitted into the National Center for Biotechnology Information (NCBI) database, particularly the RNA sequences for the spike protein (S-protein), to determine highly conserved RNA sequences across the genomes. These conserved sequences may prove to be significant for the development of a pancoronavirus vaccine, a vaccine that targets a conserved sequence and may be effective against multiple variants of coronaviruses. The RNA sequences of the S-protein genomes were the primary focus of this project since the spike protein is an essential part of virus transmission (Xia, 2021), which suggests that the spike protein is likely to be more conserved across SARS-CoV-2 genomes relative to other parts of the genome. Additionally, two approved RNA vaccines for SARS-CoV-2 target the S-protein. Nevertheless, the bioinformatic analysis was carried out on full-length genome sequences as well. This project also aims to perform sequence alignment analysis on common SARS-CoV-2 mutant sequences to gain a better understanding of how closely related they are to each other, as well as where the mutations are commonly occurring in the spike protein sequence and throughout the whole coronavirus genome. This information may be useful to develop a pancoronavirus vaccine and may be used to predict how future variants of SARS-CoV-2 will mutate, and therefore help predict how effective a vaccine may be against them.

Methodology

Existing nucleotide sequences data was obtained from the publicly available, open-source database of SARS-CoV-2 genomes from the National Center for Biotechnology Information (NCBI).

All full-length SARS-CoV-2 genomic sequences from North America in the NCBI database as of February 13, 2022 were downloaded in a FASTA file. A total of 258,269 full-length SARS-CoV-2 genomic sequences were downloaded in the search database set. The RefSeq S-protein RNA sequence was isolated from the full-length SARS-CoV-2 reference sequence and the S-protein RefSeq and the full-length RefSeq were used as reference sequences to parse into nucleotide fragments to use for searching similar sequences in the set of SARS-CoV-2 genomes.

A custom Perl script was written to parse the RefSeqs into smaller fragments and search for the sequences in the set of SARS-CoV-2 genomes. The Perl program was run on the reference sequence to first parse it into smaller segments for analysis. The reference sequence was parsed into segments with a length of 50 nucleotides, 100 nucleotides, and 150 nucleotides, with the start of each segment exactly one nucleotide away from the start of the previous segment. The Perl program was tested on a set of 1000 genomic sequences to determine whether the 50-nucleotide, 100-nucleotide, or 150-nucleotide segments should be used for analysis on a larger set of genomic sequences. The Perl program was then used to analyze the 100-nucleotide segments from S-protein RNA and full-length RNA reference sequences against each corresponding sequence of the SARS-CoV-2 target database, one genomic sequence at a time. If the fragment segments were a perfect match compared to the corresponding segments in the database genomic sequences, the segment was recorded as a match sequence. This process was then repeated for each of the other S-protein sequence fragments and sequences from the rest of the full-length RefSeq. For each match sequence recorded, the percent conservation was calculated using the number of occurrences of the sequence across all the full S-protein sequences. Match sequences that had over 95% conservation (occurred in over 95% of the S-protein sequences) were recorded as highly conserved. If a sequence fragment was not present in at least 95% of sequences in the set, that fragment was removed from analysis. Based on this, sequences with approximately 99.5% conservation were selected as potential candidates for analysis and consideration for a universal vaccine as listed in Table 1.

In addition to this, all match sequences were imported into Geneious Prime bioinformatic software and then translated into the amino acid sequence. The lead nucleotide sequences were compared with the sequence of the RefSeq and prominent mutant SARS-CoV-2 genomes to determine the location of the conserved sequences on the RefSeq and which mutations were significant (which mutations resulted in a change in the amino acid sequence).

Tables and Figures

Table 1. 100-nucleotide (nt) long RNA sequence fragments from S-protein (NC_045512.2_SProt) RNA and the RefSeq genome (NC_045512.2) were measured against a set of 258,269 full-length SARS-CoV-2 genomic sequences for perfect match. Frag. Freq. = frequency of segment across all genomes.

Ref-Seq Position # (NT)	Reference Segment	Frag. Freq.	% Conservation
S-protein			
2945	CACGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGC	257123	99.55
2946	ACGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCU	257124	99.55
2947	CGUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUG	257150	99.56
2948	GUCUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGC	257116	99.55
2949	UCUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCA	257109	99.55
2950	CUUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAG	257118	99.55
2951	UUGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGA	257116	99.55
2952	UGACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAA	257086	99.54
2953	GACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAA	257245	99.60
2954	ACAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAU	257246	99.60
2955	CAAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUC	257134	99.55
2956	AAAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCA	257215	99.59
2957	AAGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAG	257216	99.59
2958	AGUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAGA	257213	99.59
2959	GUUGAGGCUGAAGUGCAAAUUGAUAGGUUGAUCACAGGCAGACUUCAAAGUUUGCAGACAUAUGUGACUCAACAAUUAAUUAGAGCUGCAGAAAUCAGAG	257025	99.51
Full-length			
8409	UUAAAGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACGUAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAG	256963	99.49
8410	UAAAGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACGUAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAGA	256963	99.49
8411	AAAGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACGUAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAGAC	256964	99.49
8412	AAGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACGUAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAGACA	256966	99.49
8413	AGAUUUCAUGUCAUUGUCUGAACAACUACGAAAACAAAUACGUAGUGCUGCUAAAAAGAAUAACUUACCUUUUAAGUUGACAUGUGCAACUACUAGACAA	256956	99.49

Figure 1. Flow chart of the procedure, including data collection and data analysis.

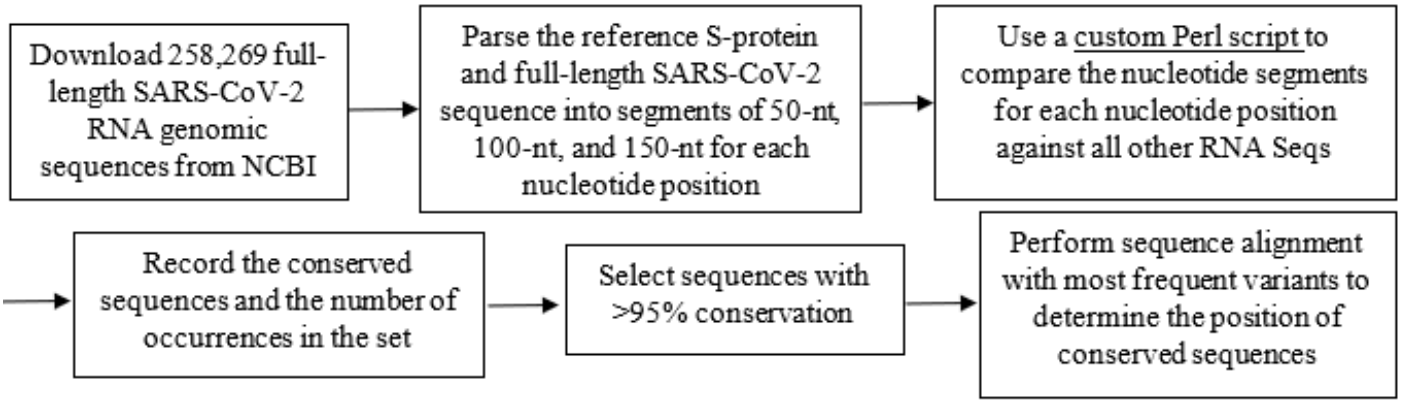


Figure 2. A. Alignment of three lead 100-nucleotide sequences from the analysis of 258,269 full-length SARS-CoV-2 genomic sequences against the spike protein and full-length RefSeq NC_0455512.2, Delta B.1.617.2, Alpha B.1.1.7, Gamma P.1, Beta B.1.351, Iota B.1.526, ETA B.1.525, and Omicron B.1.1.529 (OM570283). Tick marks indicate the positions of mutations in the variant viral genome compared to the RefSeq. B. Three sequences are shown aligned with the full-length genomic sequence (NC_045512.2), spike protein sequence (NC_045512.2Sprot), the receptor binding domain (RBD) sequence, and partial S2 part of the spike protein (OL454754). The three highly conserved sequences align with the S2 domain of the spike protein.

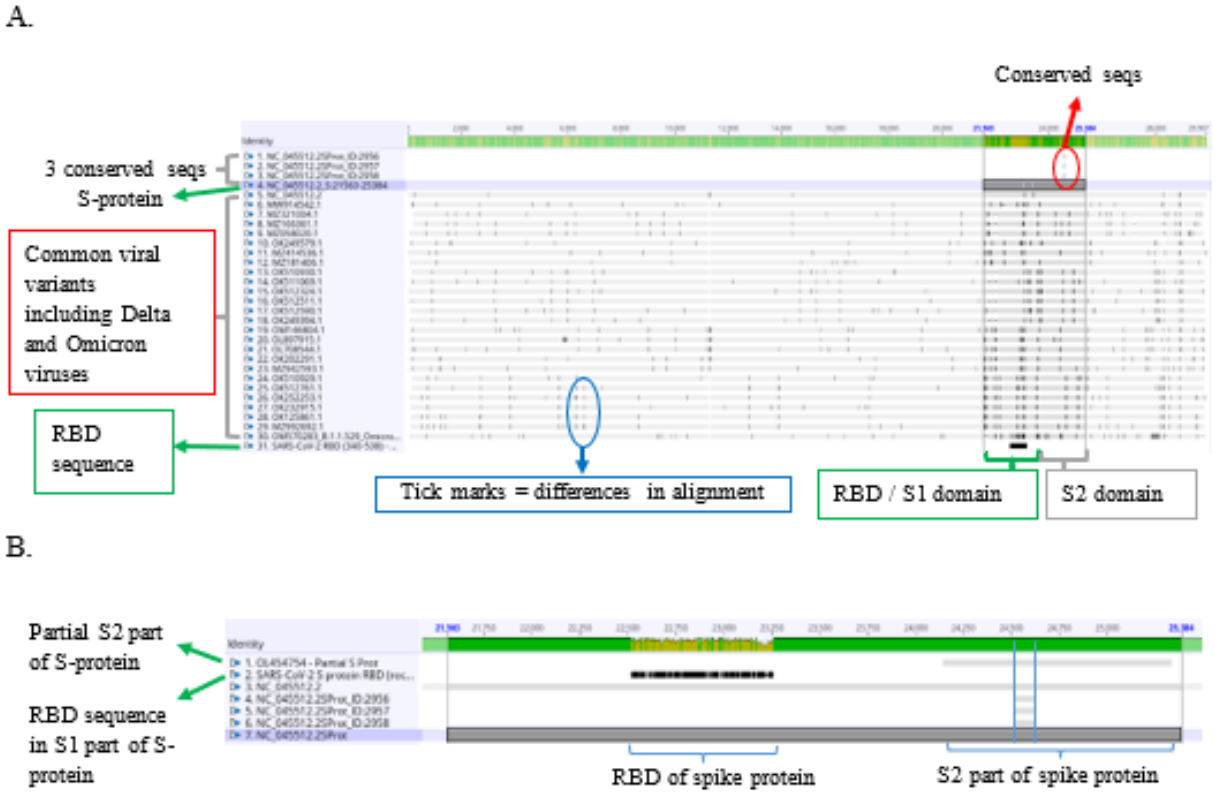


Figure 3. The three 100-nucleotide sequences are shown to align with the RefSeq. The corresponding translated protein sequences are also shown. The sequence ID:2956 is in frame with the RefSeq and depicts the corresponding protein sequence, and can be used for vaccine design. The sequences ID:2957 and ID:2958 are out of frame; no corresponding protein sequences are depicted.

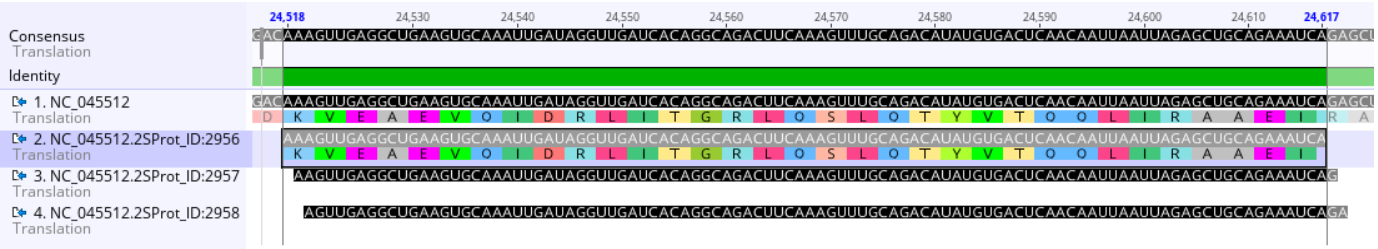
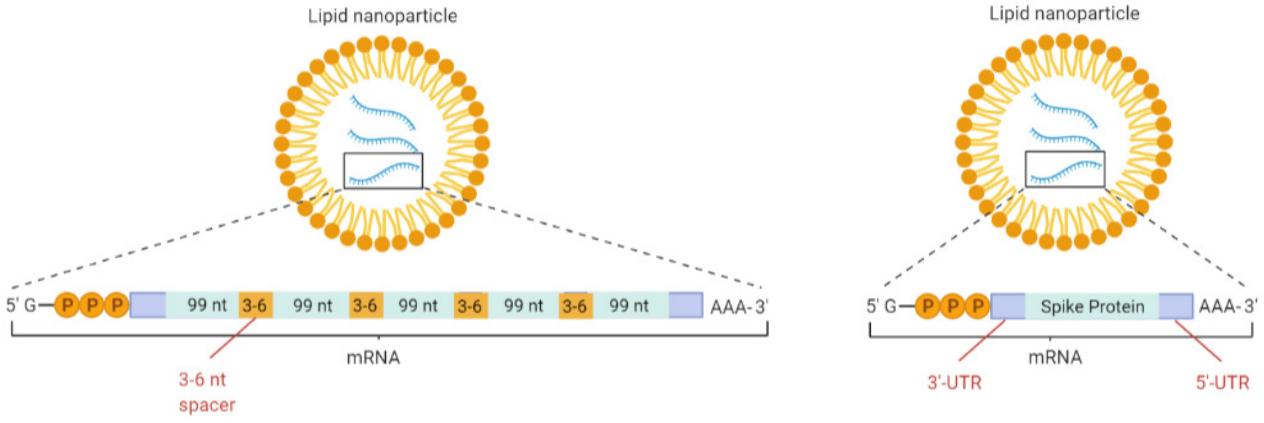


Figure 4. A. RNA vaccine design is shown. A 99-nt fragment from the lead sequence can be used for universal vaccine design by repeating and concatenating multiple fragments separated by 3- or 6-nucleotide sequences. A 3'-UTR and 5'-UTR sequence is appended to the designed sequence to facilitate translation. The RNA contains a 5'G-ppp cap for translation. The RNA is encapsulated in a lipid nanoparticle for delivery into cells. B. Figure B shows the current Moderna or Pfizer/BioNTech vaccines design that use either the whole S-protein or a large S2 fragment of the S-protein.



Results and Discussion

The present study used 258,269 SARS-CoV-2 full-length genomic sequences downloaded from the NCBI database for analysis. The sequences represent all full-length SARS-CoV-2 genomes submitted to the NCBI database corresponding to the North America region as of February 13, 2022. Figure 1 depicts the procedure of bioinformatic analysis. Bioinformatic analysis using a custom Perl script resulted in a 114-nucleotide segment corresponding to the region 2945-3058 on the spike protein (region 24,507-24,620 on the full genome) and a 104-nucleotide segment corresponding to the region 8409-8512 on the non-spike protein section of the viral genome. The 100-nucleotide sequences corresponding to these two highly conserved sequence regions are depicted in Table 1. The sequences in the spike protein region 2945-3058 were most conserved, showing conservation frequency of more than 99.5%. The next most conserved sequence region 8409-8512 was on the non-spike protein region of the genome and had a conservation frequency of 99.49%. Sequences with lower conservation frequencies are not shown.

Figure 2A depicts the alignment of three sequences from the region 2945-3058 of the spike protein with the spike protein (NC_045512.2_S:21563-25384) and full-length RefSeq NC_0455512.2, Delta B.1.617.2, Alpha B.1.1.7, Gamma P.1, Beta B.1.351, Iota B.1.526, ETA B.1.525, and Omicron B.1.1.529 (OM570283). Tick marks indicate the positions of mutations in the variant viral genome compared to the RefSeq. Gaps indicate the missing nucleotide(s). Figure 2A depicts that the RBD is highly mutable and that the S2 domain is more stable as indicated by the observation of a highly conserved sequence segment with >99.5% conservation. The next best conserved sequence with 99.49% conservation occurred in the non-spike protein region of the SARS-CoV-2 genome (Table 1). Figure 2B depicts the alignment of three sequences with the full-length genomic sequence (NC_045512), spike protein sequence (NC_045512.2Sprot), the receptor binding domain (RBD in the S1 domain of the spike protein) sequence, and partial S2 part of the spike protein (OL454754). The shown three highly conserved sequences on the spike protein align with the S2 domain (internalization domain) of the spike protein.

Figure 3 depicts the alignment of three 100-nucleotide sequences with the RefSeq. The corresponding translated protein sequences are also shown. The sequence ID:2956 is in frame with the RefSeq and depicts the corresponding protein sequence. This sequence can be used for vaccine design. The sequences ID:2957 and ID:2958 are out of frame, so corresponding protein sequences are not shown. These sequences cannot be used for vaccine design.

Figure 4A depicts the design of an RNA vaccine. The 99-nt fragment from the lead sequence ID:2956 or any in frame sequence from the region 2945-3058 on the spike protein can be used for universal vaccine design by repeating and concatenating multiple fragments separated by 3- or 6-nucleotide sequences. The translation of 99 nucleotides will result in a 33-amino acid epitope. A 3'-UTR and 5'-UTR sequence is appended to the designed sequence to facilitate translation. The RNA also contains a 5'G-ppp cap to promote translation. The RNA is encapsulated in a lipid nanoparticle for delivery into cells. Figure 4B shows the current Pfizer/BioNTech or Moderna vaccines design that uses full-length S-protein mRNA.

Discussion

Vaccine efficacy can be compromised by the emergence of viral mutations that affect the binding of neutralizing antibodies raised against the vaccine or previous viral infection(s). The emergence of Delta, Omicron, and other variants of the SARS-CoV-2 virus compromising the efficacy of vaccination raised the concern of vaccine efficacy and required booster doses of vaccines to neutralize new viral variants.

Because the spike protein is integral to virus interaction with the cell followed by internalization into cells, the spike protein is considered a prime target for vaccine design. (Li et al., 2021a)

The RBD within the S1 domain of the spike protein is responsible for binding to the ACE2 receptor on the cell surface. The S2 domain of the spike protein is responsible for virus internalization. As the RBD is poorly conserved between SARS-CoVs and other pathogenic human coronaviruses, the RBD represents a promising antigen for detecting coronavirus-specific antibodies in humans (Premkumar et al., 2020). A structure-function and antigenicity study suggested that the S-protein of the Delta variant has evolved to optimize the fusion step to enter cells while the overall structure of the RBD is preserved among all SARS-CoV-2 variants and the reoccurring mutations appear to be limited to a number of sites. Therefore, the study suggested that therapeutic antibodies or universal vaccines should not target the N-terminal domain (NTD) because the escape from anti-NTD antibodies appear to be little cost to the virus; instead, the RBD is a better target for therapeutic antibodies (Zhang et al., 2021). A comparative efficacy study of two mRNA vaccines BNT162b1, which encodes a secreted trimerized SARS-CoV-2 receptor-binding domain, and BNT162b2, which encodes a prefusion stabilized membrane-anchored SARS-CoV-2 full-length spike, revealed that the vaccine containing full-length spike protein mRNA resulted in less systemic reactogenicity compared to the vaccine containing RBD mRNA (Walsh et al., 2020). On the contrary, the current analysis of 258,269 full-length SARS-CoV-2 genomic sequences revealed that the RBD is highly mutable and the S2 domain is more stable as indicated by the observation of a highly conserved sequence segment (Figure 2A), suggesting that the S2 domain of the spike protein is a more suitable target for vaccine development across present and future SARS-CoV-2 strains.

RNA vaccines contain an mRNA code that is translated in cells to produce a peptide/protein antigen. Longer peptides are processed into shorter peptides by antigen presenting cells (APC) that are recognized by CD8+ cytotoxic T lymphocytes (CTL) and helper (CD4+) T-cells. Typically, CD8+ CTL recognizes 8-11-amino acid linear peptides presented in association with Class I MHCs (Major Histocompatibility Class) on APC, and helper (CD4+) T-cells recognize 11-30-amino acid long peptides presented in association with Class II MHC (Malonis, 2019; Slingluff Jr, 2011). Studies have suggested that long 30-amino acid peptides encompassing short minimal epitopes may be more effective immunogens (Slingluff Jr, 2011). The use of full-length spike protein mRNA, as used in Moderna and Pfizer/BioNTech mRNA vaccines, will result in a heterogeneous population of short peptide antigens that will in turn result in a heterogeneous population of antibodies. Some of these antibodies are overrepresented and stored in memory B-cells to neutralize follow on viral infection, while other antibodies are underrepresented and suppressed (Cho et al., 2021). The overrepresented antibody may not be effective against mutant viruses. The current study proposes an mRNA vaccine design with a unique highly conserved 99-114 nucleotides from spike protein RNA. To increase the robust production of corresponding peptides upon vaccination, the nucleotide segment can be repeated 5-30 or more times, thus producing epitopes for production of unique antibodies for a conserved sequence as shown in Figure 4. The other lesser conserved yet still highly conserved RNA sequences can be combined with the most conserved sequence to produce a number of antibodies targeting only highly conserved antigen sequences on the virus.

The conserved sequences selected from this study may be useful in developing an RNA or protein vaccine that could be effective against future SARS-CoV-2 strains or could act as a pancoronavirus vaccine if these sequences are present in other coronavirus families. The method used in this study can also be used to find conserved sequences across all coronavirus families, such as MERS and SARS-CoV-1. A conserved sequence among all families of coronaviruses can be used to design a pancoronavirus vaccine targeting all families of coronaviruses. Designing, manufacturing, and testing the vaccine for effectiveness in laboratory settings is an important future goal of the current study.

Acknowledgments

This research was done with the mentorship of Dr. James McSwiggen, the CEO of McSwiggen Biotech Consulting LLC.

References

Cao, Y., et al., (2022). Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*, *602*(7898), 657-663. doi: 10.1038/s41586-021-04385-3.

Chan, A. P., Choi, Y., & Schork, N. J. (2020). Conserved genomic terminals of SARS-COV-2 as coevolving functional elements and potential therapeutic targets. *MSphere, 5*(6):e00754-20. doi: 10.1128/msphere.00754-20.

Cho, A., et al. (2021). Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination. *Nature, 600*(7889), 517-522. doi: 10.1038/s41586-021-04060-7. Epub 2021 Oct 7.

Cohen, J. (2021). The dream vaccine. *Science, 372*(6539), 227–231. doi: 10.1126/science.372.6539.227.

Corbett, K.S. (2021). Immune correlates of protection by mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. *Science, 373*(6561):eabj0299. doi: 10.1126/science.abj0299.

Dai, L., et al. (2020). A universal design of betacoronavirus vaccines against COVID-19, MERS, and SARS. *Cell, 182*(3), 722–733. doi: 10.1016/j.cell.2020.06.035.

Forster, P., et al. (2020). Phylogenetic network analysis of SARS-COV-2 genomes. *Proceedings of the National Academy of Sciences, 117*(17), 9241–9243. doi: 10.1073/pnas.2004999117.

Koff, W. C., & Berkley, S. F. (2021). A universal coronavirus vaccine. *Science, 371*(6531), 759–759. doi: 10.1126/science.abh0447.

Krammer, F. (2020). SARS-CoV-2 vaccines in development. *Nature, 586*(7830), 516-527. doi: 10.1038/s41586-020-2798-3.

Lei, K. C., & Zhang, X. D. (2020). Conservation analysis of SARS-COV-2 spike suggests complicated viral adaptation history from bat to human. *Evolution, Medicine, and Public Health, 2020*(1), 290–303. doi: 10.1093/emph/eoaa041.

Li, M., et al. (2021). Rational design of a pan-coronavirus vaccine based on conserved CTL epitopes. *Viruses, 13*, 333. doi: 10.3399/v13020333.

Li, X., et al. (2021a). Possible targets of pan-coronavirus antiviral strategies for emerging or re-emerging coronaviruses. *Microorganisms, 9*(7), 1479. doi: 10.3390/microorganisms9071479.

Malonis, R.J. (2020). Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem Rev., 120*(6), 3210–3229. doi: 10.1021/acs.chemrev.9b00472.

Nagy, A. and Alhatlani, B. (2021). An overview of current COVID-19 vaccine platforms. *Comput Struct Biotechnol J., 19*, 2508–2517. doi: 10.1016/j.csbj.2021.04.061.

Pegu, A., et al. (2021). Durability of mRNA-1273 vaccine-induced antibodies against SARS-CoV-2 variants. Science, 373(6561), 1372-1377. doi: 10.1126/science.abj4176.

Premkumar, L., et al. (2020). The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol. 5(*48):eabc8413. doi: 10.1126/sciimmunol.abc8413.

Slingluff Jr, C.L. (2011). The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer J., 17*(5), 343-50. doi: 10.1097/PPO.0b013e318233e5b2.

Walsh, E.E., et al. (2020). RNA-Based COVID-19 Vaccine BNT162b2 Selected for a Pivotal Efficacy Study. *medRxiv.* doi: 10.1101/2020.08.17.20176651.

World Health Organization. (2022, April 20). *Coronavirus disease (COVID-19) pandemic*. Retrieved April 21, 2022, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Xia, X. (2021). Domains and functions of spike protein in SARS-COV-2 in the context of vaccine design. *Viruses, 13*(1), 109. doi: 10.3390/v13010109

Zhang, J., et al. (2021). Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science, 374*(6573), 1353-1360. doi: 10.1126/science.abl9463.

A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease

Anjali Sreenivas

Nikola Tesla STEM High School

Abstract

Alzheimer's Disease, described by several researchers as the public health crisis of the 21st century, is a devastating neurodegenerative disorder of gradual onset, clinically characterized by cognitive deterioration and memory loss. Despite its prevalence, there is yet to be a globally accessible and affordable method by which the disease can be differentially and timely diagnosed. This study seeks to investigate the potential of blood-based molecules as Alzheimer's biomarkers through the interpretation of machine learning models that yield the highest predictive power within metabolomics and transcriptomics datasets aggregated from multiple studies. The Mann-Whitney U test with FDR corrections was used to determine the statistical significance of multiomics feature values in differentiating between the two experimental groups (cognitively healthy controls and Alzheimer's patients), and gene ontology was performed on the top differentially expressed genes identified to determine biological pathways that appear to be altered in Alzheimer's patients. While serum metabolites demonstrated little potential as effective Alzheimer's biomarkers, the Random Forest and Extreme Gradient Boosting machine learning architectures, when applied to the transcriptomics dataset, had AUROC values of 0.713. Further analyses revealed that 78 genes were highly differentially expressed with p-values < 0.01 and $> 90\%$ increases in transcript abundance within the Alzheimer's cohort. The identified genes are robust candidate biomarkers shared across multiple study sites that could one day make diagnosing Alzheimer's with a simple blood test a reality. Furthermore, the biological pathways identified corroborate the blood-brain barrier hypothesis and may be effective therapeutic targets in advancing the quest for a cure.

Keywords: Alzheimer's Disease, blood-based biomarkers, blood-brain barrier, diagnostics, pathogenesis, therapeutic targets

1. INTRODUCTION

1.1 Background and Current Diagnostic Methods

Alzheimer's Disease (AD) is a devastating clinically and patho- physiologically heterogeneous neurodegenerative disorder of insidious onset, characterized by memory loss and cognitive deterioration (Khachaturian, 1985). The vast majority of AD cases have a late onset with clinical symptoms only beginning to manifest at the age of 65 years or older (Guerreiro & Bras, 2015); however, in many cases, pathophysiological changes as- sociated with the disease have been observed to initiate decades before these symptoms first appear (Trushina & Mielke, 2014). While various classes of medications have been approved to relieve symptoms and slow disease progression, there is no cure for the disease today.

The World Health Organization (WHO) reports that AD represents between 60 and 70% of all dementia cases, which amounts to nearly 55 million patients across the globe. With- out the development of effective diagnostic tools and thera- peutics, this figure is expected to triple as early as 2050 as the world's population grows and ages. Furthermore, the 2021 Alzheimer's Facts and Figures Report notes that deaths from AD have increased by 145% over the past two decades, making AD the seventh leading cause of mortality around the world (Alzheimer's Association, 2021).

Despite the prevalence of AD, however, there is yet to be a globally accessible, affordable, and efficient method by which the disease can be differentially and timely diagnosed—one of the most pressing needs in the field today.

Central hallmarks of Alzheimer's Disease are the accumulation of extracellular amyloid-β senile plaques in the brain parenchyma, and the development of intraneuronal neurofib- rillary tangles composed of hyperphosphorylated tau protein (Hampel et al., 2018; Zetterberg, 2019). Apart from conduct- ing post-mortem autopsies, current methods of diagnosis for AD typically involve a combination of performing invasive spinal taps to analyze cerebrospinal fluid (CSF), neuropsycho- logical exams, and PET or MRI imaging to observe potential structural or molecular alterations in the brain. Validated CSF biomarkers of high accuracy are present for both Aβ plaque pathology (characterized by a reduced CSF Aβ42/40 ratio or CSF Aβ42 concentration, owing to the retention of Aβ42 in plaque-laden brain tissue) as well as tangle pathology (char- acterized by the hypersecretion of CSF total tau (t-tau) and phosphorylated tau (p-tau)) (Zetterberg, 2019). However, ob- taining CSF to measure the presence of these biomarkers is not only a time-consuming and painful process, but is also incredibly costly and widely inaccessible around the world (Snyder & Carrillo, 2014), especially in developing regions lacking more advanced infrastructure.

1.2 Blood-Based Biomarkers: What Is Known

The prospect of a simple blood test being able to distinguish AD patients from a cohort of dementia patients would be a ground-breaking milestone in scientific research, in terms of facilitating earlier diagnoses, streamlining the diagnosis process, improving accessibility, and also providing potential therapeutic targets. With infrastructure already in place for blood sampling worldwide, and that too, with no specialists required to conduct blood tests, the discovery of robust blood-based biomarkers would provide a feasible, efficient method to routinely track disease progres- sion in patients, making earlier diagnoses more probable so that AD patients around the world can get access to the proper, timely intervention they need to slow disease progression.

1.2.1 The Blood Brain Barrier: Why Blood-Based Biomarkers are Relevant

While AD is often pathologically characterized by the accu- mulation of Aβ and tau proteins in the brain parenchyma, a growing body of research suggests the existence of additional molecular pathophysiological pathways that interact with one another (Hampel et al., 2018; Yamazaki & Kanekiyo, 2017). At the center of this research is the role which neuroinflam- mation has been found to play in AD pathogenesis (Heneka et al., 2015), and how it affects the blood-brain barrier (BBB).

The BBB, while once believed to function as an imperme- able barrier, is appreciated today for mediating interactions between the immune system and the central nervous system (CNS), such as immune cell trafficking into the CNS and the exchange of cytokines between the CNS and the circula- tion (Erickson et al., 2012). This selective permeability, how- ever, becomes compromised as the BBB is disrupted both with age and the progression of neurodegenerative diseases such as Alzheimer's. Specific consequences of this can include but are not limited to: altered immune cell trafficking resulting in additional interactions between immune cells and the CNS (to combat inflammation), higher amounts of cytokines crossing directly from the CNS into the bloodstream, and the alter- ation or modulation of transporter cells located near the BBB (affecting the concentrations of molecules such as proteins and metabolites in both the bloodstream and the CNS) (Yamazaki & Kanekiyo, 2017). Furthermore, the BBB is composed of brain endothelial cells which themselves become activated in response to chronic neuroinflammation, often inducing cytokine secretions (Erickson et al., 2012).

In essence, the effects of neuroinflammation and the subse- quent disruption of the blood-brain barrier, a key component of AD pathogenesis, extends beyond the CNS and are also present in the bloodstream. Thus, even though AD is a disease affecting the CNS, blood-based biomarkers in the multiomics fields may have potential in monitoring disease progression.

1.2.2 Challenges: Heterogeneity and Parallelism with Other Non-Alzheimer's Dementias

The fact that Alzheimer's Disease is both clinically and patho- physiologically heterogenous (meaning that it has multiple alternative pathways of disease

development and progression) (Yashin et al., 2017) adds an additional layer of complexity when it comes to identifying robust blood-based biomarkers. In addition to this inherent heterogeneity and multifactorial nature of the disease, other proteinopathies and pathologies of- ten coexist in the ageing brain as well as co-morbidities. Such co-morbidities which are often present within AD patients in- clude those with clinical phenotypes that overlap with AD (i.e. Parkinson's Disease) as well as those without (i.e. dementia, stroke, cardiovascular disease) (Hampel et al., 2018; Santiago & Potashkin, 2021). What this ultimately translates to is the fact that AD pathogenesis varies from person to person, which contributes to the difficultly of identifying reliable biomarkers that are generalizable to a greater population. Furthermore, the idea that other non-AD dementias have identical clinical phenotypes and are often co-morbidities of AD helps explain the occurrence of misdiagnoses in Alzheimer's cases as well as the challenge in finding differential biomarkers.

1.2.3 Past Research

Several observational studies have attempted to discover AD- specific biomarkers through a multiomics approach, measuring the concentrations of various biological compounds in blood serum and plasma. However, the results of these studies have lacked reproducibility–a potential result of their relatively small sample sizes, the presence of unaccounted co-morbidities, the imbalance of sex and social factors, and cases of misdiagnosis within the study cohorts (Lin et al., 2019).

At the front of the current blood-based biomarker land- scape today tends to be plasma Aβ1-42 and 1-40 levels, plasma tau, and serum levels of the axonal protein neurofilament light (NFL). While multiple studies have independently identified these biomakers as potentially valuable, there are a number of critical issues: (1) there is still broad disagreement and inconsistency between these studies and several others over the differentiating value of these proteins in the diagnosis of AD; and (2) both plasma tau and serum NFL are not unique to AD. The bottom line is that additional blood-based biomark- ers need to continue to be investigated through independent, large-scale validation with greater sample sizes (Hampel et al., 2018).

1.3 Study Goals

The proposed study seeks to address the lack of reproducibil- ity in previous study results by aggregating metabolomic and transcriptomic blood profiles of AD patients and cognitively healthy controls from several published observational studies to attain a much larger, more generalized data set. Through the interpretation of linear or non-linear machine learning models that yield the highest predictive power, relationships between multiomics feature values (the levels of blood-based metabo- lites and gene expression levels) and the onset of Alzheimer's Disease can be objectively established, and potential therapeu- tic targets can be identified following pathway analysis. This research aligns with a multimillion-dollar request for proposals recently put forth by the National Institutes of Health (2021). Ultimately, the identification of more robust, AD-specific sig- nals across multiple study sites represents a step towards making blood tests a reality for AD diagnosis. If successful, the com- putational methodology used could also be applied to other branches of multiomics beyond metabolomics and transcrip- tomics to identify additional candidate biomarkers or potential therapeutic targets for AD.

2. Methodology

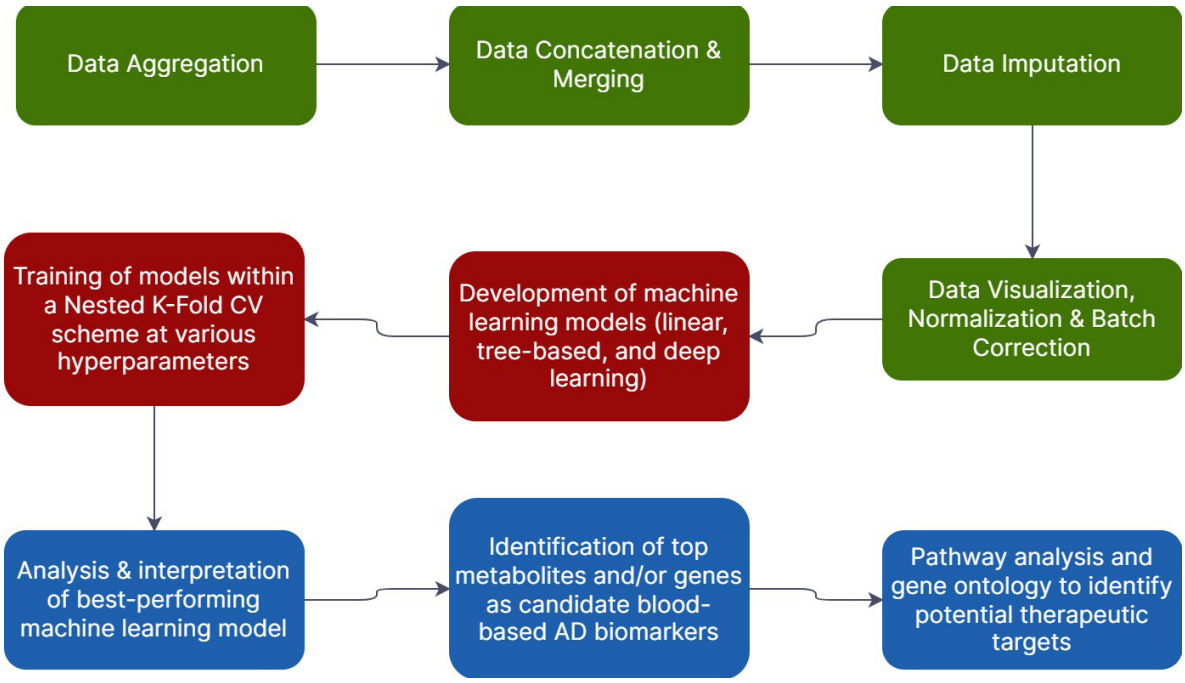


Figure 1. A summary of the entire research procedure broken down into three main phases: (1) aggregating and understanding the data (green), (2) training and testing various machine learning models on the dataset (red), and (3) analysis of the best-performing model to identify robust candidate AD biomarkers and relevant biological pathways.

2.1 Aggregating and Understanding the Data

2.1.1 Metabolomics

Over 2,500 metabolomic blood profiles of humans were aggre- gated from four published observational studies on the Synapse AD Knowledge Portal a. All data utilized were publicly available data upon the filing and approval of Data Access Requests.

Two identical tables were compiled, one for samples taken from the blood serum, and the other for samples taken from the blood plasma. The following data was recorded for each sample: (a) metadata, including the study the sample was taken from, individual ID number, the plate ID number, patient background statistics (age, sex, ethnicity, education level), and final diagnosis (Alzheimer's Disease (AD), Mild Cognitive Im- pairment (MCI), or Cognitively Normal (CN)); (b) targeted p180 metabolite levels quantified via flow injection analysis-mass spectrometry (FIA-MS/MS) and liquid chromatography- mass spectrometry (LC-MS/MS); and (c) targeted bile acid metabolomic levels quantified via LC-MS/MS.

The aggregated plasma and serum metabolomic datasets were pre-processed separately, but in adherence to the same protocol. First, analytes that were not measured across all the metabolomic studies of that given specimen were removed. Subsequently, analytes where >40% of measurements were noted as "<LOD" (less than the lower limit of assay detection) were eliminated. And lastly, all samples and analytes that con- tained >10% non-numeric "NA" measurements among the remaining data values were also eliminated. The result was a condensed set of analytes that were measured across all the integrated studies, and samples with adequate data. All re- maining "LOD" values were substituted with half of the LOD threshold value provided by the study for that given analyte, and any missing or "NA" values in the data set were imputed by computing the average of all other values for the analyte. Finally, transformations of data for all samples in which the the study did not report raw data were accordingly handled.

2.1.2 Transcriptomics

Nearly 1,400 transcriptomic blood profiles of humans were also aggregated from three published observational studies on the Synapse AD Knowledge Portal and from the Alzheimer's Disease Neuroimaging Initiative Database b. All transcriptomic data utilized were also publicly available data upon the filing and approval of Data Access Requests.

Upon identifying and recording the appropriate metadata for each sample using Microsoft Excel's index and match func- tions, the Pandas library in Python was used in order to merge the three dataframes - each containing the transcriptome data and the metadata of samples from a single study. The merge was performed over the gene name column, ouputting a single dataframe of samples from all studies and a concatenated set of genes (comprising of only the genes shared across all three studies).

Next, patient IDs with duplicate entries across various time points (due to patients of some studies being screened multiple times at various checkups) were handled by converting the date column to a pandas date-time object, then sorting the merged data frame by both patient ID and date/time, and lastly, removing all duplicates by only keeping the last (latest) entry. Furthermore, across the various samples, there were several categories of diagnosis: CN, AD, MCI, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), NA, and unknown. All samples with recorded diagnosis of NA, unknown, or EMCI were eliminated, and AD, MCI, and LMCI were all grouped together as the AD category. NA values in any of the transcript abundances within all remaining samples (1,357 total samples) were imputed with the average of the rest of the values for that gene.

2.1.3 Data Visualization and Batch Correction

The t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm c, a dimensionality reduction algorithm, was ap- plied to both the metabolomic datasets and the transcriptomic dataset using the scikit-learn library in Python in order to visualize the data and determine whether batch correction would be necessary. The t-SNE algorithm preserves local structure within a dataset, meaning that the distance between a point and its closest neighbors are maintained in the dimensionality reduction process. Essentially, this allows clusters within the dataset to be easily identified visually when plotted.

For both metabolomics datasets, the t-SNE scatterplot vi- sualizations showed relatively interspersed data points when colored by study site (Figures 2 and 4), meaning that batch variation was likely not significant. Thus, no normalization or batch correction of data points was performed.

[a]https://adknowledgeportal.synapse.org/
[b]http://adni.loni.usc.edu/data-samples/adni-data-inventory/
[c]https://www.mathworks.com/help/stats/t-sne.html

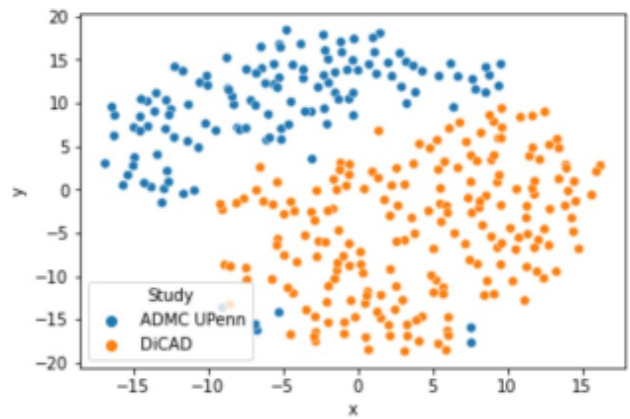


Figure 2. T-SNE dimensionality reduction plot of aggregated plasma metabolomic dataset colored by study site.

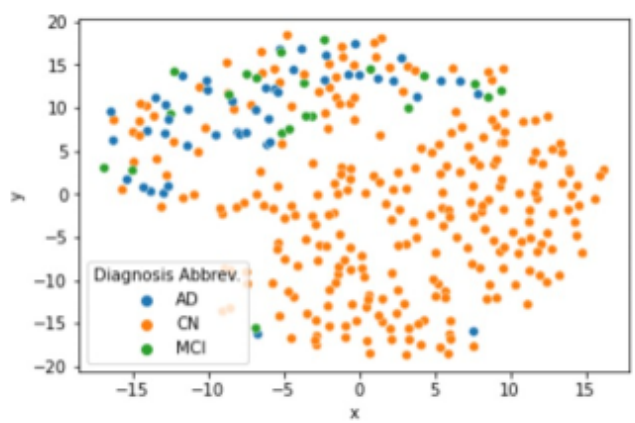


Figure 3. T-SNE dimensionality reduction plot of aggregated plasma metabolomic dataset colored by plate diagnosis.

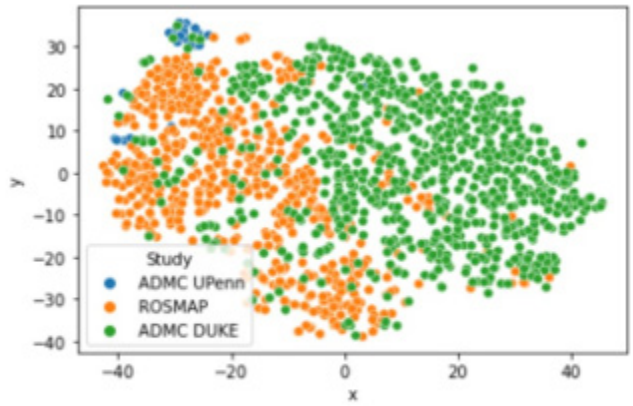


Figure 4. T-SNE dimensionality reduction plot of aggregated serum metabolomic dataset colored by study site.

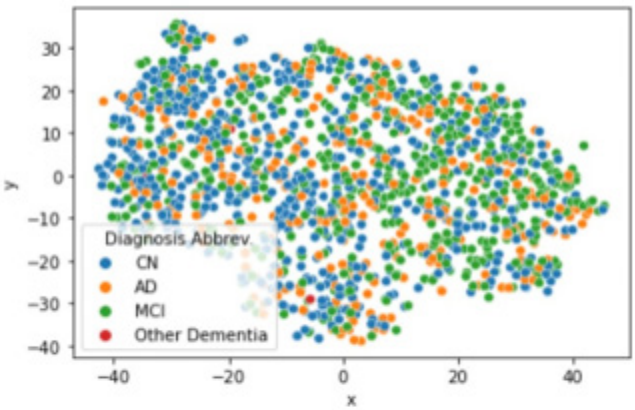


Figure 5. T-SNE dimensionality reduction plot of aggregated serum metabolomic dataset colored by diagnosis.

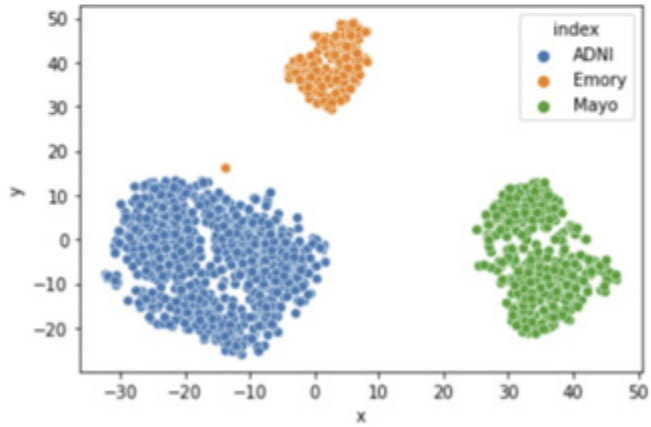


Figure 6. T-SNE dimensionality reduction plot of original merged transcriptomic dataset colored by study/site.

However, critical limitations of the plasma metabolomics dataset, in particular, was revealed by Figure 3, a t-SNE plot for the dataset colored by diagnosis. The scatterplot not only made evident the limited sample size of the plasma dataset (326 samples), but (when considered in tandem with Figure 2) showed that only one of the two studies the dataset comprised of contained Alzheimer's patients. In other words, one of our two experimental cohorts would originate from a single study– a major red flag revealing that any inferences and conclusions we proceed to draw from this plasma dataset are limited in terms of confidence.

On the transcriptomics side, when the t-SNE algorithm was applied to the original merged transciptomics dataset (Fig- ures 6 and 7), there were very clear, distinct clusters visible on the scatterplot by site/study. Within each of these clusters were multiple types of diagnoses, however, suggesting that there were likely no confounding factors affecting both study and diagnosis. These observations demonstrated that statistical transformations such as normalization and batch correction may be beneficial, and also that the dataset may be useful for analytical purposes.

Figures 8 and 9 illustrate the t-SNE visualizations of the transcriptomics dataset following normalization. The separa- tion of the clusters by study site, while still present, visibly decreased, indicating that normalization slightly improved dis- crepancies between study values. Figures 10 and 11 represent the t-SNE scatterplots of the transcriptomic dataset following batch correction on top of normalization. Batch correction was performed using pyComBat d, an open source Python implementation of ComBat, which is a software used for the adjustment of batch effects. The plots demonstrate the relative absence of the clusters that originally separated the studies, and generally intermixed diagnoses. Thus, machine learning architectures will be applied on both the normalized transcrip- tomics dataset as well as the normalized and batch corrected dataset, and the model/dataset combination that yields the best results in the testing cycle will be further analyzed.

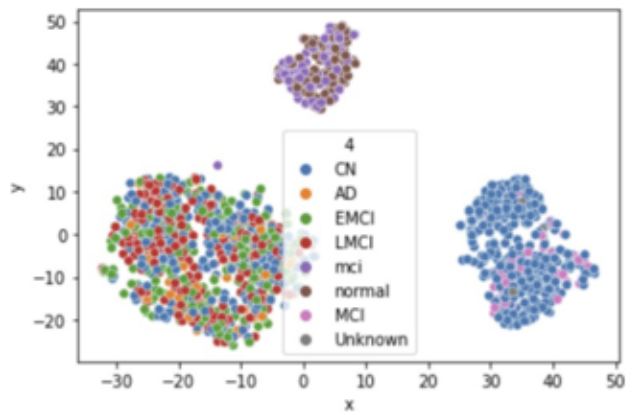


Figure 7. T-SNE dimensionality reduction plot of original merged transcriptomic dataset colored by diagnosis.

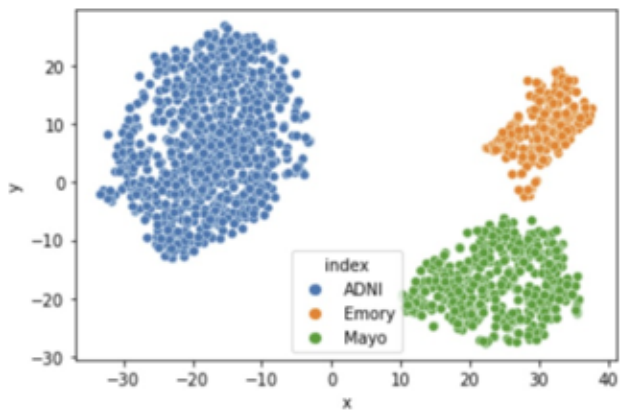


Figure 8. T-SNE dimensionality reduction plot of the normalized merged transcriptomic dataset colored by study/site.

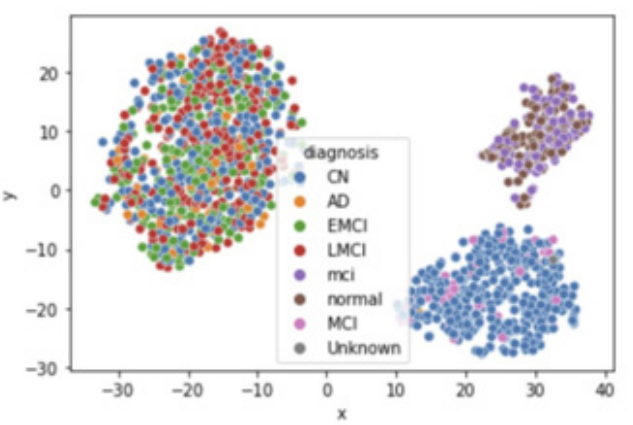


Figure 9. T-SNE dimensionality reduction plot of normalized merged transcriptomic dataset colored by diagnosis.

[d]https://epigenelabs.github.io/pyComBat/

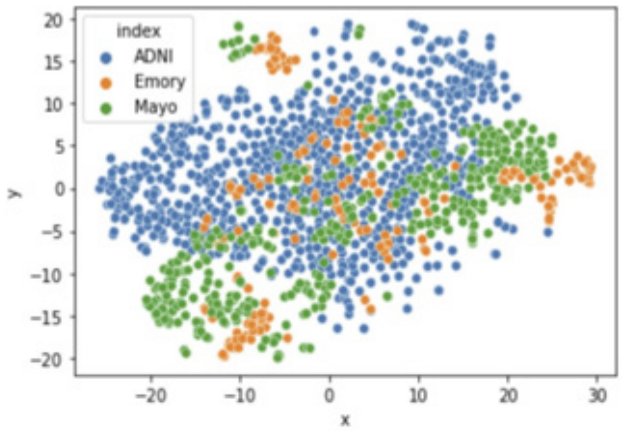


Figure 10. T-SNE dimensionality reduction plot of normalized and batch corrected merged transcriptomic dataset colored by study/site.

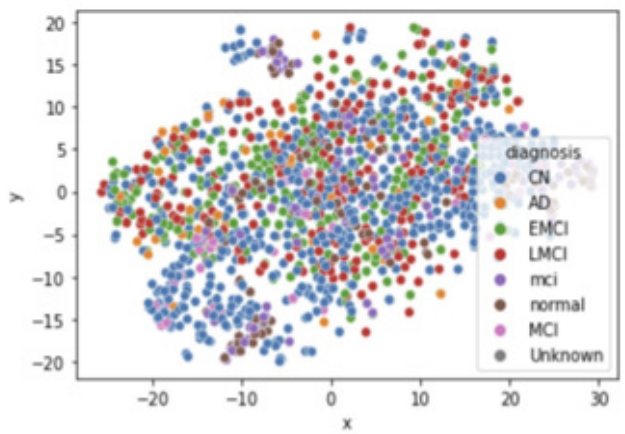


Figure 11. T-SNE dimensionality reduction plot of normalized and batch corrected merged transcriptomic dataset colored by diagnosis.

2.2 Machine Learning

2.2.1 Development, Training, and Testing of Models

Using the sckit-learn machine learning library, four machine learning classification models were developed in Python to evaluate whether there were any significant patterns within the datasets that could be used to differentiate between the AD and cognitively healthy cohorts in the first place. The machine learning architectures explored were Random Forest, Extreme Gradient Boosting (XGB), Support Vector Classifier (SVC), and Logistic Regression. The four models were tested and trained on each of the datasets (serum metabolomics, plasma

metabolomics, normalized transcriptomics, and normalized & batch corrected transcriptomics) within a Stratified K-Fold Cross Validation scheme (where k = 10), in which stratified splitting was first performed to segment the data into multiple testing and training data sets.
Subsequently, a nested cross validation (CV) loop was used to evaluate each model's overall performance and generaliz- ability, in order to ensure that the model was not overfit for the inputted dataset. Within the inner loop of this nested CV scheme (which was passed in the training data set of the kth iteration), grid search was performed to select for the optimal hyperparameters. This best fit model was then applied on the testing data set, and area under the receiver operating charac- teristic curve (AUROC) was measured–representing a ratio of the classifier's true positive rate vs. false positive rate. The AU- ROC score ranges from 0 to 1, where 1 represents the perfect classifier, and 0.5 represents totally random classification.
Figure 12 below illustrates the steps that occur within the nested cross-validation scheme.

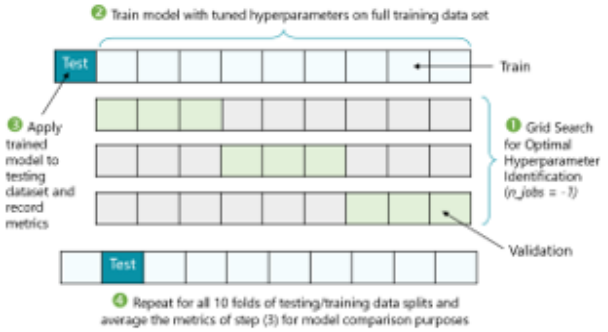


Figure 12. Diagram illustrating the sequential progression of a K-fold nested cross validation scheme.

2.2.2 Results

All aggregated metabolomic samples provided metadata with regards to age, sex, and education, and therefore, the performance of the machine learning models could be compared when applied on just the metabolite blood concentrations versus that of when metadata is also factored in.

Among the four models that were investigated, the best ma- chine learning performance obtained with the serum metabolite concentrations alone was with the Random Forest model, and even those results were very low with an average AUROC of 0.623. Subsequently, the machine learning models were applied on each of the three serum metabolomic study datasets independently to investigate whether there were any corre- lations that could be identified within these individual studies themselves. Even poorer results were obtained with the range of average AUROC scores being from 0.504 and 0.578.

The Random Forest machine learning model was then ap- plied on only the metadata for the serum metabolomics dataset, which resulted in an average AUROC value of 0.671 - a metric higher than what was observed when the model was applied to the metabolite concentrations by themselves. This suggested that AD classification accuracy was significantly better when based solely off of factors like age, ethnicity, gender, etc. than when metabolite concentrations were considered. To validate this, the model was applied to a dataset consisting of both patient metadata and serum metabolite concentrations, and the AUROC score was 0.624–less than that of metadata alone. Thus, factoring in serum metabolite concentrations degraded classification accuracy, demonstrating that serum metabolites likely have very low potential as effective Alzheimer's Disease biomarkers.

More optimistic results were obtained, however, when ma- chine learning methods were applied to the plasma metabolomics dataset. Once again, the best-performing model was the Random Forest architecture for which, when applied to the plasma metabolite values alone, resulted in an AUROC of 0.692 - a significantly higher score than that of the serum dataset. Furthermore, when this model was applied to the plasma patient metadata alone, an AUROC of 0.539 was obtained–indicating that factoring in plasma metabolite concentrations did indeed yield much higher classification accuracy. The results suggest that plasma metabolites may have value in Alzheimer's diag- nosis; however, due to the notable limitations of the plasma dataset revealed by our t-SNE plots, no further analysis was conducted.

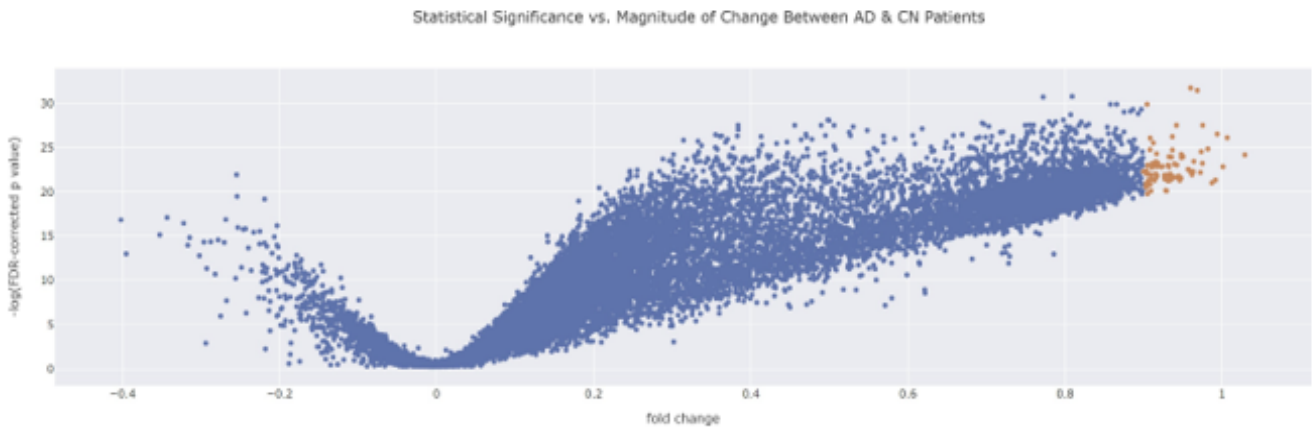


Figure 14. Volcano plot illustrating the distribution of genes' capabilities in differentiating between the AD and CN study cohorts on the basis of p-value and fold change. Orange-colored points indicate highly differentially expressed genes with a | fold change | ≥ 0.9 and a p-value <0.01.

Figure 13 below summarizes the above metabolomics AUROC metrics in a bar chart.

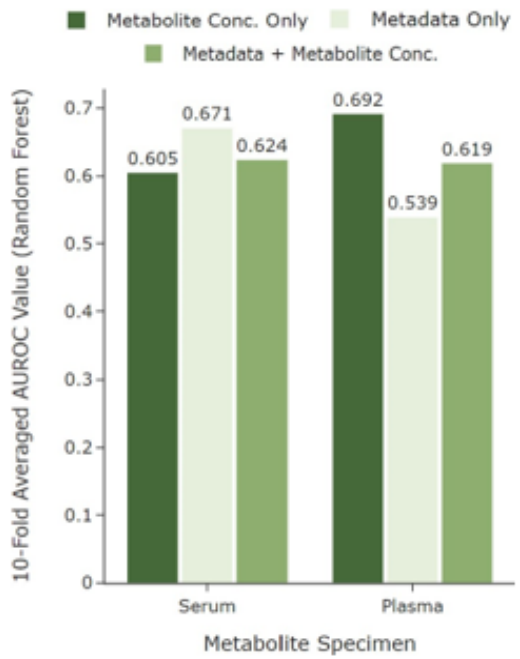


Figure 13. Raw AUROC metrics for the Random Forest classifier (identified as the top-performing model) on the metabolomics datasets.

Transcriptomics data yielded the best machine learning results by far, opening pathways for further analyses and sig- nificant conclusions. The best performance was observed with both the Random Forest and Extreme Gradient Boosting models when applied to the normalized dataset, with an average AUROC score of 0.713. Figure 15 shows the performance of all four machine learning architectures on both the normal- ized and normalized + batch corrected transcriptomics datasets based on average AUROC score.

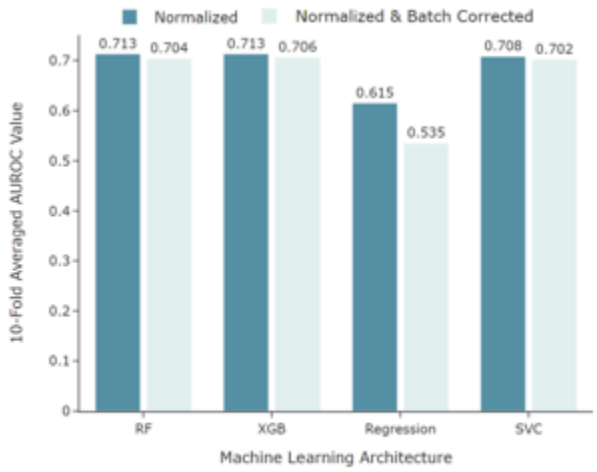


Figure 15. Raw AUROC metrics for all four machine learning architectures on the transcriptomics datasets.

Furthermore, feature reduction was performed on the tran- scriptomics datasets (values on a spectrum from 100 genes to 10,000 genes were tested) to investigate whether the machine learning metrics would further improve if the feature count was truncated. No significant changes in metrics were ob- served.

3. Data Analysis and Results

Based on the machine learning results, further analysis was conducted on the normalized transcriptomics dataset to assess whether specific genes' blood transcript abundance levels could be of notable importance in Alzheimer's Disease diagnosis.

3.1 Statistical Analyses

For each of the 17,729 genes in the transcriptomics dataset, the p-value was computed using the Mann-Whitney U Test to determine the statistical significance of each feature's ability to differentiate between between AD and cognitively normal patients. Additionally, fold change for each gene was calculated as follows: (mean transcript abundance of AD patients – mean of CN patients) / mean of CN patients.

Figure 14 provides a high-level visual representation of these metrics and illustrates that there are indeed genes of likely relevance to Alzheimer's diagnosis based on blood transcript abundance. Points colored in orange represent the 78 genes classified as highly differentially expressed between the two study cohorts, with both a statistically significant p-value <0.01 and a | fold change| 0.9 which exceeded the 90th percentile. All highly differentially expressed genes identified appeared to be upregulated within Alzheimer's patients.

The box plots in Figure 16, illustrating variation in blood transcript abundance between the AD cohort and the CN cohort for five of the highly differentially expressed genes identi- fied (FA4GD, CLCA3P, KRTAP15-1, CD177, and POM121L8P)
show clear upregulation in the Alzheimer's cohort.

3.2 Machine Learning Feature Importance Analysis

In addition to statistical analyses, further analyses were con- ducted by examining the feature importance rankings of the two best-performing machine learning models, RF and XGB. For each of the 10 iterations executed and for both of these machine learning architectures, the quantified importance of each of the genes were recorded and then averaged. The fea- tures were then sorted in descending order according to their average importances, and these values have been plotted for both machine learning models as the scatterplots shown in Figure 17.

On both scatterplots, there is a clear point at which feature importance begins to taper off–this occurs at around index 850 for XGB and 2000 for RF. Taking the lower of those two threshold indices, the fea- ture importance data for the XGB and RF models were then

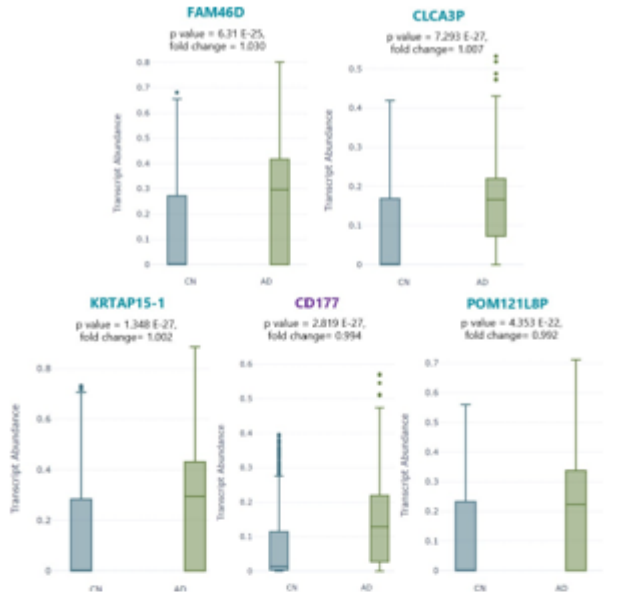


Figure 16. Box plots depicting variation in blood transcript abundance levels for the five highly differentially expressed genes with the greatest fold changes between the Alzheimer's cohort and the cognitively healthy controls.

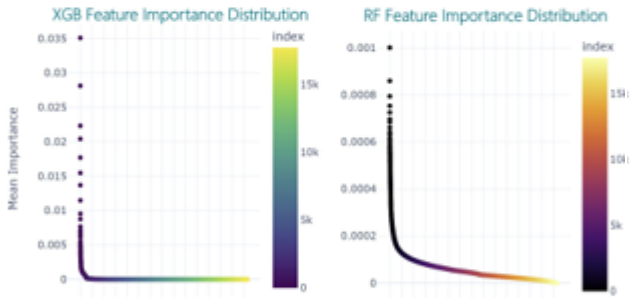


Figure 17. Scatterplots modeling average feature importance rankings for each of the genes within the transcriptomics dataset for the two best-performing machine learning architectures. Elbows are located at an index of 850 for XGB and 2000 for RF.

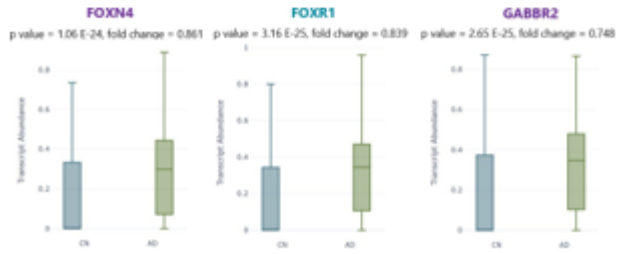


Figure 18. Box plots depicting variation in blood transcript abundance levels for three of the differentially expressed genes ranked important by both top-performing machine learning architectures.

merged over the top 850 features in order to generate a list of genes whose importance in determining AD classification was ranked within the top 850 by both machine learning architectures. 68 genes were identified under this criteria, and further selection parameters were applied to determine which of them have statistically significant

p-values of <0.01 and fold changes of at least 0.5. Nine genes (distinct from those identified through statistical analyses alone) satisfied this cri- teria: FOXN4, FOXR1, GABBR2, FAM83C, ANKRD30A,ANKFN1, NEB, SDC1, EML6. Figure 18 contains box plots showing the variation in blood transcript abundance between the AD and CN cohorts for three of these nine genes.

Table 1 lists the 87 total differentially expressed genes (DEGs) identified via statistical analyses and machine learning feature importance analyses.

Table 1. List of the 87 genes found to be differentially expressed in the blood of Alzheimer's patients versus cognitively healthy controls.

0	FAM46D	OR14J1	FOXB2
1	CLCA3P	EVX1	CCDC105
2	KRTAP15-1	EMX1	CYLC1
3	CD177	ZCCHC16	ROPN1
4	POM121L8P	CCDC54	CETN1
5	APOD	SCP2D1	OR5M9
6	IRS4	HOXD11	SLC26A9
7	GABRB1	LINC00380	LMX1A
8	SIX6	PCDHB8	OR6M1
9	HOXD12	SSX1	UNCX
10	OR10A5	DRD1	C6
11	GPR152	SPATA31D3	DEFB114
12	ATOH1	HIST1H2AA	CSNK1G2-AS1
13	OR52J3	FERD3L	F9
14	SPANXN2	PABPC1L2B	OR8D2
15	LRRC18	SPRR2D	OR10A2
16	ARGFX	OR6X1	KRTAP25-1
17	LCE1C	DSPP	OR2G6
18	OR13G1	DEFB119	KRTAP10-5
19	SUN5	OR2A12	TPD52L3
20	PPY	NEUROG1	FOXN4
21	TNP2	IFNA6	FOXR1
22	SPINK14	DNAJB8	GABBR2
23	OR5AK2	FGF3	FAM83C
24	NPY5R	OTOP3	ANKRD30A
25	OR10A3	ADAM29	ANKFN1
26	SPACA4	PSORS1C2	NEB
27	LINC00906	IMPG1	SDC1
28	APOA5	PAX1	EML6

3.3 Gene Ontology Enrichment Analysis

For these 87 DEGs, gene ontology enrichment analysis was performed to determine which biological pathways appear to be altered within Alzheimer's patients based on over-representation in the list of DEGs. The Gene Ontology Consortium's online resource e was used to conduct the analysis. 17 statistically significant and significantly enriched pathways were observed to be altered in AD patients, as shown in Figure 19.

[e]http://geneontology.org/

In addition to identifying likely compromised purpose of conducting the gene ontology analysis was to verify that the DEGs identified were functionally enriched.

4. Conclusions and Discussion

The 87 DEGs identified may be of incredible importance, representing robust candidate blood-based biomarkers for Alzheimer's Disease shared across a much broader population from multiple study sites. Some of these genes, such as the ones colored in purple in Figures 16 and 18 (CD177, FOXN4, and GABBR2) corroborate the findings of existing smaller-scale studies, and others (such as FOXR1, CLCA3P, and FAM46D) appear to be novel findings that should be further investigated. Collectively, these findings could take us a step closer towards one day diagnosing AD through a simple blood test.

Furthermore, the biological pathways identified, such as alterations to nervous system processes and sensory perception pathways, are characteristic of Alzheimer's pathogenesis, in- dicating that the study results corroborate the Blood-Brain Barrier hypothesis. In other words, the differentially expressed genes identified support the idea that changes that occur up in the brain as part of neurodegenerative disease progression are indeed likely detectable in the blood. Beyond this, some of these biological pathways–such as primary metabolic process and multicellular organismal process–may also represent new targets for drugs and therapeutics to explore to potentially advance the quest for a cure to Alzheimer's Disease.

Further research should study the potential of the identified DEGs in distinguishing AD patients from a cohort of dementia patients, as opposed to AD versus CN patients alone. Addition- ally, research could be conducted to explore whether any of the DEGs found are correlated to other comorbidities. The methodology implemented in this study can also be extended to other branches of multiomics, such as proteomics, to gain additional insights in terms of which blood-based molecules may have the greatest potential as Alzheimer's biomarkers. Along these lines, future studies should further investigate the diagnostic potential of plasma metabolites, for which this study was inconclusive about.

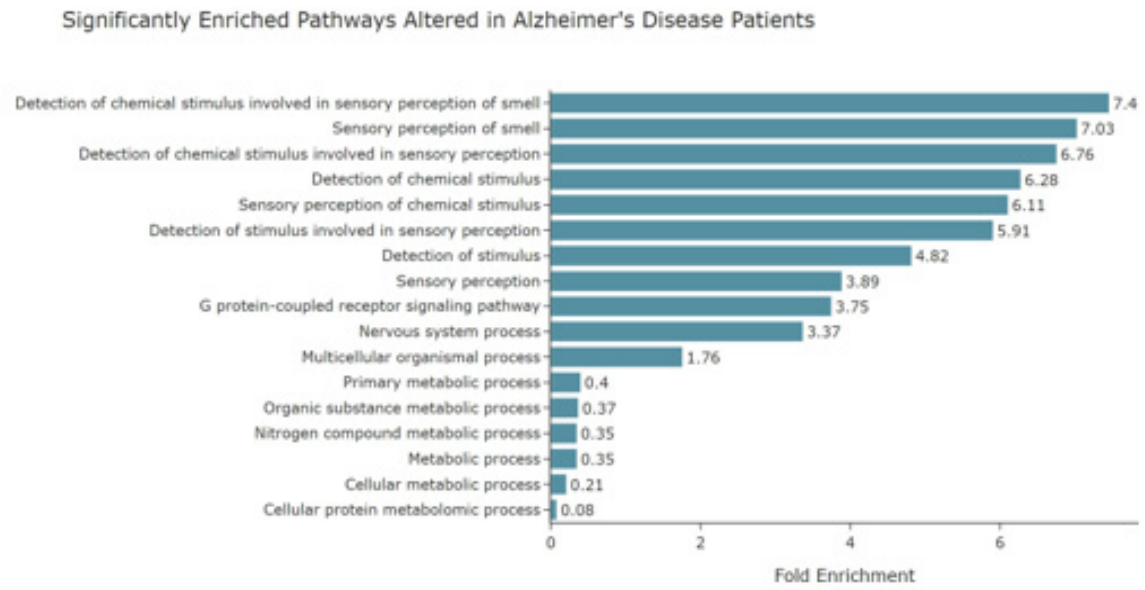


Figure 19. Bar graph illustrating the 17 significantly enriched pathways–determined through gene ontology enrichment analysis–altered in Alzheimer's Disease patients based on over-representation in the identified list of DEGs. Fold enrichment value represents the fold increase in the frequency genes of a given pathway were present in the inputted DEGs list relative to what would be expected in a random gene list. All plotted pathways are statistically significant with p-values < 0.01. In this context, the p-value represents the chance that the observed frequency of genes mapped to the specific term is due to random chance.

Acknowledgement

I would like to profusely thank my mentor, Dr. Thanaphong Phongpreecha, Instructor at the Stanford University School of Medicine, for offering his advice, supporting me, and having meaningful discussions with me throughout. I also thank Mrs. Kate Allender of Tesla STEM High School for her constant encouragement and support.

References

Alzheimer's Association. 2021, Alzheimer's Disease International

Erickson, M., Dohi, K., & Banks, W. 2012, Pub Med, doi:10.1159/000330247 Guerreiro, R., & Bras, J. 2015, Genome Medicine, doi:10.1186/s13073-015-0232-5

Hampel, H., O'Bryant, S., Molinuevo, J., & Zetterberg, H. 2018, Nature Reviews Neurology, 117, doi:10.1038/s41582-018-0079-7

Heneka, M., Carson, M., & Khoury, J. 2015, Pub Med, doi:10.1016/ S1474-4422(15)70016-5

Khachaturian, Z. S. 1985, Archives of Neurology, 42, 1097

Lin, C., Huang, C., & Lin, K. 2019, Annals of Clinical and Translational Neurology, 117, doi:10.1002/acn3.726

National Institutes of Health. 2021

Santiago, J., & Potashkin, J. 2021, doi:10.3389/fnagi.2021.631770

Snyder, H., & Carrillo, M. 2014, Alzheimer's Dementia, doi:10.1016/j.jalz. 2013.10.007

Trushina, E., & Mielke, M. M. 2014, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1842, 1232, misfolded Proteins, Mitochondrial Dysfunction, and Neurodegenerative Diseases

Yamazaki, Y., & Kanekiyo, T. 2017, Pub Med, doi:10.3390/ijms18091965 Yashin, A., Fang, F., & Kovtun, M. 2017, doi:10.1016\%2Fj.exger.2017.10.020 Zetterberg, H. 2019, Journal of Neuroscience Methods, 319, 2, methods and Models in Alzheimer's Disease Research

Parkinson's Disease: A Risk Prediction Model Using Inflammatory Bowel Disease

Laya Nair

Nikola Tesla STEM High School

Abstract

Currently, Parkinson's Disease (PD) affects over 10 million individuals worldwide, with many patients experiencing gastrointestinal issue years before the onset of motor symptoms of the disease, but what if there was a way to combat it using the gut as an early detector? This project seeks to identify if having a diagnosis of inflammatory bowel disease (IBD) would lead to a higher risk of developing PD by using a clinical prediction model. Using a final data set of 7 different studies, a meta-analysis, which included subgroup analyses and a meta-regression, was conducted after extraction of summary data to determine the risk ratio of developing PD. Overall, a diagnosis of IBD, including Crohn's disease and ulcerative colitis, was directly associated with the risk of developing PD (Risk Ratio: 1.12, 95% Confidence Interval: 1.046-1.17). A higher risk for PD was associated with a diagnosis of IBD, indicating that IBD can be used as an early detector for future PD development and PD risk. This can help high risk groups go through early screening for the disease to take preventative action.

Keywords: Parkinson's Disease, inflammatory bowel disease, systematic review, meta-analysis, inclusion/exclusion criteria

Introduction

Identifying Inflammatory Bowel Disease (IBD) as a risk predictor of Parkinson's disease (PD) will allow for early detection and therefore minimize the debilitating impact of PD. PD is a neurodegenerative disease common in older people, characterized by impaired movement and motor symptoms such as tremors, bradykinesia, and rigid muscles. These symptoms correspond to insufficient levels of dopamine, a neurotransmitter associated with movement of the body. Currently, there are no known causes for PD, but there are several risk factors that have been identified such as heredity and gene mutations in the LRRK2 DJ-1 and glucocerebrosidase genes, along with exposure to harmful environmental factors, like pesticides.

The overall impact of PD treatment is immense, having a $51.9 billion impact in just the United States every year. With the US having around 10% of global PD patients (10 million), the global societal impact may be up to $100 billion. Today, approximately 70% of PD patients have gastrointestinal (GI) issues along with regular symptoms of the disease, like tremors and decreased motility (Petrossian, 2017). Prior research indicates that GI issues are rooted in the enteric nervous system (ENS), a part of the GI tract found to have neurons that acts like a "second brain". This nervous system is connected to the central nervous system (CNS) through the nigrostriatal and mesolimbic pathways and decreased dopamine levels seen in each of these nervous systems create are associated with both PD and IBD, a GI illness that is associated with decreased GI motility and increased inflammation. IBD, which can be manifested as Crohn's disease and ulcerative colitis, is known to be associated with symptoms of abdominal pain, bleeding, and diarrhea (Baumgart et. al., 2007).

In several recent studies, (Brudeck, 2019; Weimers et. al., 2018), a correlation has been found between IBD and PD, suggesting that early detection of IBD may lead to early detection of PD and consequently can be used to develop its prevention measures. For instance, in two groundbreaking studies, researchers found a correlation in approximately 200 gene loci in both IBD and Parkinson's patient data (Liu et. al., 2015; Jostins et. al., 2021). Despite this, much of this correlational research has not been conclusive due to insignificant statistical significance (Sun et. al., 2021).

Understanding the risk factors of PD can lead to reducing the downstream impact of this disease that includes direct costs (treatment, hospitalization) as well as indirect costs (loss of employment, burden to family). This project seeks to investigate the risk of PD and identify predictors/biomarkers such as IBD's significance using a systematic review and meta-analysis. This project seeks to identify that having a diagnosis of IBD would lead to a higher risk of developing PD.

Methods

Existing data on IBD and PD research has been extracted from several open-source studies.

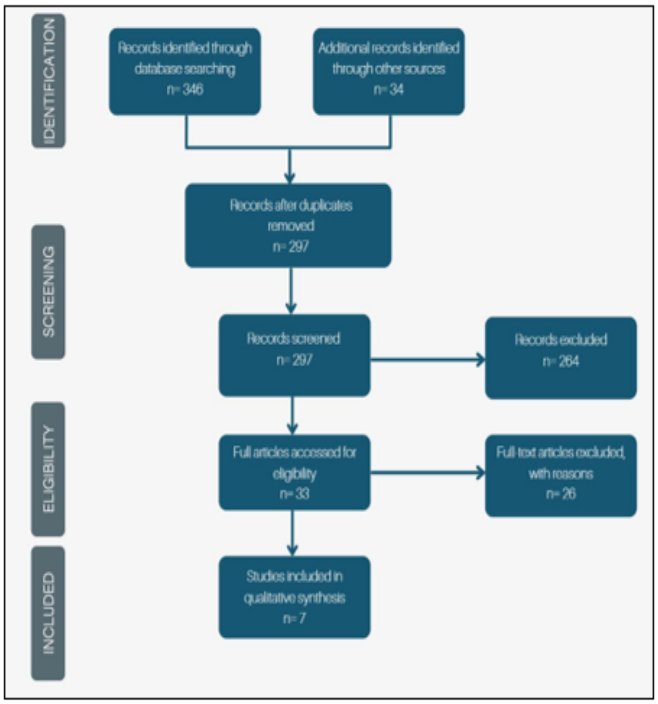
A systematic review of scientifically valid and peer-reviewed studies was used in this project in order to identify risk predictors of PD. A systematic search for pertinent studies using PubMed, MedLine, Cochrane Central, Google Scholar, and Embase for patients with IBD and PD was conducted. Detailed search terms included are listed in *Supplementary Table 1*.

Supplementary Table 1. Search terms used to identify potential studies for screening and inclusion

Search term 1	Search term 2	Search term 3
Parkinson's Disease	Inflammatory Bowel Disease	Humans
Parkinsons Disease	Crohn's Disease	Patients
Parkinson's	Ulcerative Colitis	
Parkinsons		
Parkinsonism		
Parkinsonian dementia		

Before extraction and analysis of the data, each of the studies was manually searched to account for additional studies that could apply to this analysis. Once added to the total number of studies, each study was fully searched through to see if they applied to the inclusion and exclusion criteria of this systematic review, described in the next paragraph such as format of data and type of study. If not applicable to this criteria, the study was removed from the analysis (Figure 1).

Figure 1. Summary of study identification and inclusion



The total number of patients included across all studies to be included in the analysis were tabulated. After careful and methodical screening, there was 7 out of a possible 346 studies that were suited to the inclusion/exclusion criteria were identified. Across these studies, which were conducted in 7 different countries, there were 2.3 million participants, ranging from ages 50-95.

For data extraction, the following values were extracted from each study for further analysis: 95% confidence intervals, hazard ratios, number of participants, and odds ratios. Any data not in this format was either not included or converted into this format. Additionally, only primary studies conducted between January 1970 and January 2022 were used in this review. Any systematic reviews or meta-analyses were to be excluded from this study to prevent overlapping, repetitive data.

These studies were weighted using the Newcastle-Ottawa scale, separately determining weight for cohort and case-control studies. For case control studies, the selection, (representativeness of cases, selection, and definition of controls), comparability (of case and controls based on design),

and exposure (ascertainment, methodology non-response rate) were evaluated. For cohort studies, selection (representativeness of exposed cohort, selection of non-exposed cohort, ascertainment, demonstration outcome of interest was not present), comparability (based on design and analysis), and outcome (assessment, duration and adequacy of follow-up) were evaluated. The Newcastle-Ottawa scale assesses the quality of studies that are used in a meta-analysis. The quality of the studies, in terms of the selection of groups like sampling procedure, comparability of groups, and if exposure or outcome are used in the study, are rated on a 9-star scale. Each study is given a certain weight based on this scale and the higher the weight, the higher the quality of the studies. These weighted values were then input into a generic variance weighted average equation (Figure 2) to be used in the meta-analysis to prevent skew or bias in the data.

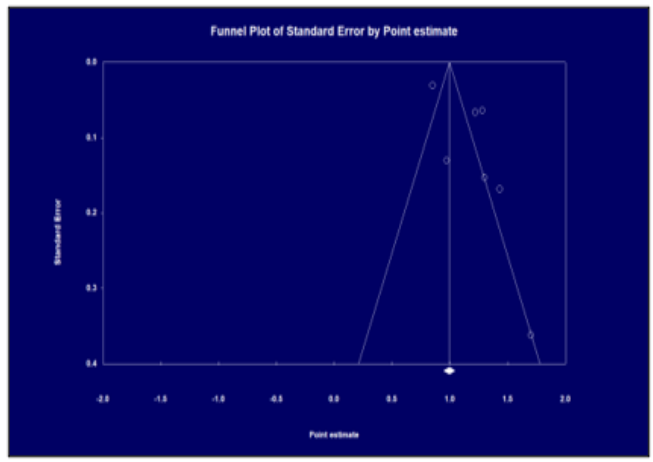
Figure 2. Generic Variance Weighted Average Equation

$$\frac{\sum Y_i(1/\mathrm{SE}_i^2)}{\sum(1/\mathrm{SE}_i^2)},$$

Once studies were weighted, the meta-analysis was conducted. In R-studio and CMA (Comprehensive Meta Analysis) , a fixed ratio meta-analysis was conducted after extracting confidence intervals, odds ratios, hazard ratios, and the number of participants for each study. Cohort and case control studies were grouped with a subgroup analysis. Studies with confidence intervals above one which are above one found a direct association between IBD diagnosis and PD development and below one had an inverse association. Five out of the 7 studies found a direct association as shown in *Figure 3*.

There was a slight publication bias skew to the right (*Figure 4*), since there are more studies that found a direct association between IBD diagnosis and PD development.

Figure 4. Publication Bias Funnel Plot



This effect was mitigated by giving more weight to studies with higher reliability as seen with the usage of the Newcastle-Ottawa Scale.

Results

Overall, the risk ratio was 1.12, meaning there is a direct association between IBD diagnosis and PD development, and this is statistically significant with confidence intervals of 1.046-1.17. The direct association is further reinforced by the low variance of 0.001 indicating that this meta-analysis is representative of the population under study which is patients with PD (*Figure 5*).

Figure 3. Forest plot of included studies

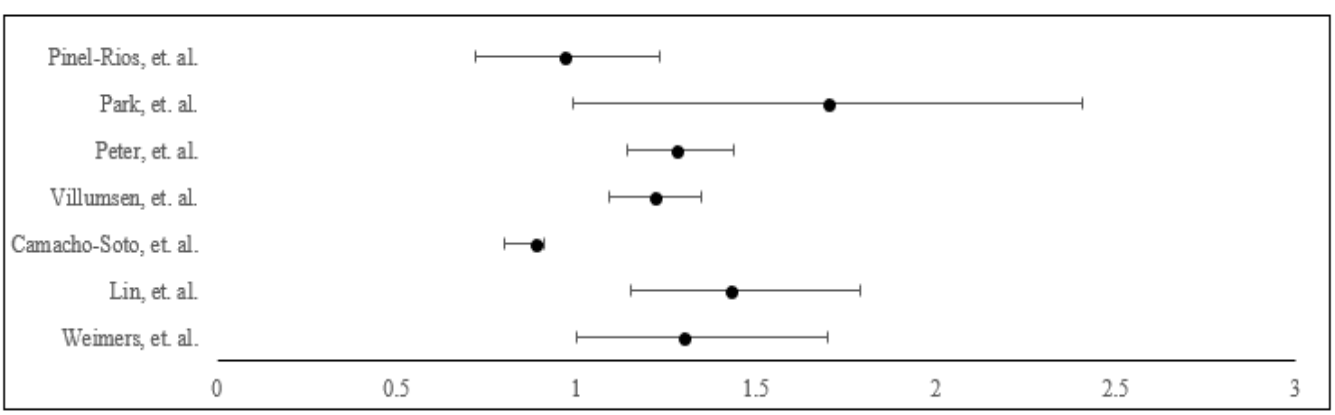
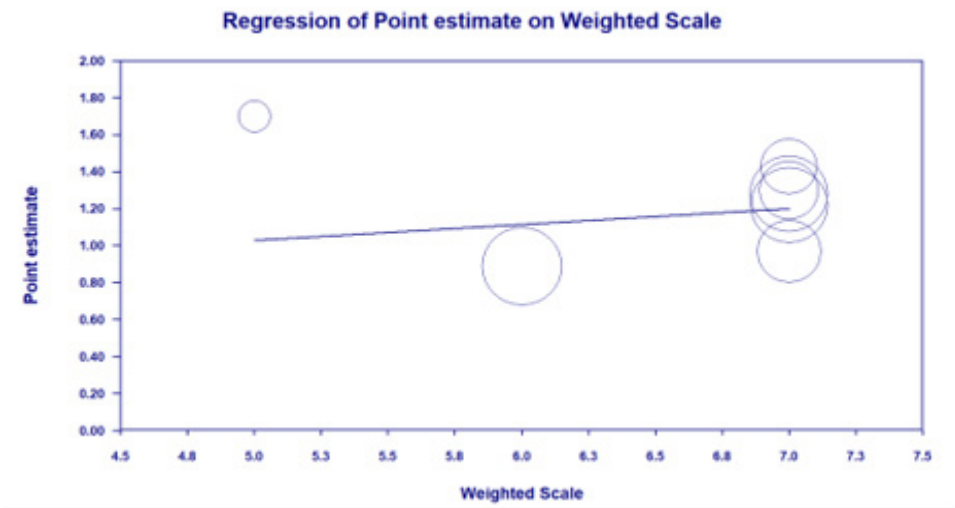


Figure 4. Meta-regression graph using weighted scale



This direct association implies that early and accurate IBD treatment could mitigate development of PD. Additionally, this study had a standard error of 0.032, meaning that the data was highly representative of the population (*Supplementary Table 2*).

Supplementary Table 2. Results

Model	Study name	Statistics for each study							Newcastle	Point estimate and 95% CI
		Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value		-2.00 -1.00 0.00 1.00 2.00
	Weimers	1.300	0.153	0.023	1.000	1.600	8.493	0.000	7	
	Lin	1.430	0.168	0.028	1.100	1.760	8.493	0.000	7	
	Camacho-S	0.890	0.051	0.003	0.790	0.990	17.444	0.000	6	
	Villumsen	1.220	0.066	0.004	1.090	1.350	18.394	0.000	7	
	Peter	1.280	0.064	0.004	1.155	1.405	20.070	0.000	7	
	Park	1.700	0.362	0.131	0.990	2.410	4.693	0.000	5	
	Pinel Rios	0.970	0.130	0.017	0.715	1.225	7.456	0.000	7	
Fixed		1.108	0.032	0.001	1.046	1.170	35.070	0.000		

Conclusions

There was a direct association found between the development of PD and the diagnosis of IBD, meaning that those with diagnoses of IBD are at a higher risk of developing PD. By being able to predict development of PD sooner, with symptoms and risk predictors being recognized earlier, approaches to treatment can be more proactive and preventative rather than reactive. This can reduce downstream impacts of the disease including its financial and material costs. Additionally, this provides tremendous scope for future applications of research by predicting other risk factors of PD. IBD and other gastrointestinal disorders could be used as predictive biomarkers for forms of dementia, including but not limited to PD, such as Lewy body dementia or Alzheimer's disease. This research could be furthered with larger data sets and longitudinal data as well. Different statistical approaches such as the use of the Cox proportional hazard model to predict survival rates of patients with IBD could be used.

This study does have some limitations. This study, although larger than most, did have stringent inclusion criteria and therefore had a small number of studies conducted in few countries, meaning the data cannot be directly applied to a global scale. Additionally, it is acknowledged that there was possible subjectivity in weighting during the use of the Newcastle-Ottawa scale.

There is tremendous scope for future application by predicting risk factors of PD. IBD and other gut disorders could be used as predictive biomarkers for different forms of dementia. There is the possibility of developing individualized data clouds for the management of diseases. The earlier

that symptoms are caught, approaches to treatment can be more proactive and preventative than reactive. Furthermore, anti-inflammatory diets can be utilized to prevent gut inflammation to reduce symptoms of gastrointestinal illness, like IBD, that could lead to risk of developing PD. Overall, there is immense value in understanding the connection between the gut and the brain.

Acknowledgements:

Thank you to Dr. Amy Shepherd from Harvard University for her mentorship & guidance.

Bibliography:

Baumgart, D. C., & Sandborn, W. J. (2007, May 10). I*nflammatory bowel disease: Clinical aspects and established and evolving therapies.* The Lancet. Retrieved January 25, 2022, from https://www. sciencedirect. com/science/article/pii /S014067360760751X ?casa_token=fp8- TPI-jVuQAAAAA%3AO4hzUZcGYy3jAlAHKpT-VTGj-SUR6pw48JdROUAO-W4avfgrNW-QyPA-7kPXd_A5QdMBTUmtO0Tyg

Brudek, T. (2019). *Inflammatory bowel diseases and parkinson's disease.* Journal of Parkinson's disease. Retrieved December 16, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6839501/

Camacho-Soto

Chen, Q.-Q., Haikal, C., Li, W., & Li, J.-Y. (1AD, January 1). G*ut inflammation in association with pathogenesis of parkinson's disease.* Frontiers. Retrieved December 16, 2021, from https://www.frontiersin.org/articles/10.3389/fnmol.2019.00218/full

Inserro, A. (2021, December 1). *Review suggests having IBD may increase risk of PD.* AJMC. Retrieved December 16, 2021, from https://www.ajmc.com/view/review-suggests-having-ibd-may-increase-risk-of-pd

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Philip Schumm, L., Sharma, Y., Anderson, C. A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z., … Cho, J. H. (2012, October 31). *Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature News. Retrieved January 25, 2022, from https://www.nature.com/articles/nature11582

Lin, J., Lin, C., & Hsu, C. (2016). *Association Between Parkinson's Disease and Inflammatory Bowel Disease: A Nationwide Taiwanese Retrospective Cohort Study*. Academic.oup.com. Retrieved April 21, 2022, from https://academic.oup.com/ibdjournal/article/22/5/1049/4561726

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., Ripke, S., Lee, J. C., Jostins, L., Shah, T., Abedian, S., Cheon, J. H., Cho, J., Daryani, N. E., Franke, L., Fuyuno, Y., Hart, A., Juyal, R. C., Juyal, G., … Weersma, R. K. (2015, July 20). *Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations.* Nature News. Retrieved January 25, 2022, from https://www.nature.com/articles/ng.3359

Park , S., Kim, J., & Chun, J. (2019). *Patients with inflammatory bowel disease are at an increased risk of parkinson's disease: A South Korean nationwide population-based study.* Journal of clinical medicine. Retrieved April 21, 2022, from https://pubmed.ncbi.nlm.nih.gov/31398905/

Peter, I., Dubinsky, M., & Bressman, s. (2018). *Anti-tumor necrosis factor therapy and incidence of parkinson disease among patients with inflammatory bowel disease.* JAMA neurology. Retrieved April 21, 2022, from https://pubmed.ncbi.nlm.nih.gov/29710331/

Pinel-Rios, J., Navarro, C., Navarro, M., Tapia, M., Vera, M., Arillo, V., Garcia, M., Castellanos, A., Sevilla, F. (2019, June 19). *Association of Parkinson's disease and treatment with aminosalicylates in inflammatory bowel disease: a cross-sectional study in a Spain drug dispensation records.* OUP Academic. Retrieved January 25, 2022, from https://pubmed.ncbi.nlm.nih.gov/31221869/

Sun, Y., Geng, J., Chen, X., Chen, H., Wang, X., Chen, J., Li, X., & Hesketh, T. (2021, November 29). *Association between inflammatory bowel disease and dementia: A longitudinal cohort study.* OUP Academic. Retrieved January 25, 2022, from https://academic.oup.com/ibdjournal/advance-article/doi/10.1093/ibd/izab300/6445975

Villumsen , M., Aznar, S., & Pakkenberg, B. (2018). *Inflammatory bowel disease increases the risk of parkinson's disease: A Danish nationwide Cohort Study 1977-2014.* Gut. Retrieved April 21, 2022, from https://pubmed.ncbi.nlm.nih.gov/29785965/

Weimers, P., Halfvarson, J., Sachs, M. C., Saunders-Pullman, R., Ludvigsson, J. F., Peter, I., Burisch, J., & Olén, O. (2018, May 16). *Inflammatory bowel disease and parkinson's disease: A nationwide Swedish cohort study.* OUP Academic. Retrieved December 16, 2021, from https://academic.oup.com/ibdjournal/article/25/1/111/4996912

Zhu , F., Li, C., & Gong, J. (2019). *The risk of parkinson's disease in inflammatory bowel disease: A systematic review and meta-analysis.* Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. Retrieved December 16, 2021, from https://pubmed.ncbi.nlm.nih.gov/30309751/

Zhu, Y., Yuan, M., Liu, Y., Yang, F., Chen, W.-Z., Xu, Z.-Z., Xiang, Z.-B., & Xu, R.-S. (2022, February). *Association between inflammatory bowel diseases and parkinson's disease: Systematic review and meta-analysis.* Neural regeneration research. Retrieved December 16, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8463981/

Engineering/ Machine Learning

Dampening Phantom Shockwave Jams In Mixed-Autonomy Traffic Via Deep-Rl Control Of Autonomous Vehicles

Edward Zhou

Nikola Tesla STEM High School

ABSTRACT

Phantom shockwaves, traffic jams that arise in the absence of any bottlenecks, are the core phenomenon behind the 6.9 billion cumulative hours lost to roadway congestion each year. Commonly referred to by drivers as stop-and-go traffic, these shockwaves are local peaks of high vehicular density that emerge in free-flowing traffic when a certain critical density is exceeded. Small perturbations in traffic are observed to grow into traveling waves, spreading downstream for miles even when the average density of the roadway is still low. This project aims to dampen phantom shockwaves by con- trolling a small concentration of connected autonomous vehicles (CAVs), alleviating congestion on simulated roadways through deep reinforcement learning and incorporation of upstream/downstream traffic information. The control policy for the CAVs, which determines a vehicle's next action based on its current state, was developed in the Eclipse SUMO traffic microsimulation software. Simulations were conducted using Markov decision processes to maximize average vehicle velocity and system throughput. For testing purposes, human driving behavior was replicated by the Intelligent Driver Model, providing vehicle acceleration as a function of the vehicle's immediate observations. The introduction of a five percent concentration of trained CAVs into a simulated highway of human drivers resulted in a 14.34% increase in system throughput and a 9.74% increase in average vehicle velocity, achieving near-perfect dampening of phantom shockwaves that were previously visible. This autonomous control policy presents the potential for transfer learning with real-world vehicles, leveraging innovative autonomous technologies to reduce roadway congestion and the vast quantities of energy wasted alongside it.

1 Introduction

This research aims to dampen and dissolve large phantom shock wave jams by computationally training and inserting a small concentration of autonomous vehicles into mixed autonomy traffic systems. Currently, transportation is a major source of American energy consumption, accounting for 28% of the nation's energy usage and 26% of its greenhouse gas emissions. This problem is significantly amplified by traffic congestion, resulting in a cumulative global delay of 6.9 billion hours and economic losses upward of $305 billion in a single year (Cookson et al., 2018). Contemporary traffic control strategies designed to mitigate such congestion mostly consist of fixed actuators, such as traffic lights, ramp metering, and variable speed limits. However, these traditional approaches have limited effects on traffic dynamics that emerge between fixed control points. Most notable of these are phantom shockwave jams, also referred to as jamitons or stop-and-go traffic, which are waves of high vehicle density that emerge in free-flowing traffic without any obvious bottleneck. At present, there are no controllers capable of dampening phantom shockwaves in large roadway systems such as freeways.

But during a recent 2017 breakthrough experiment, Stern et al. were able to achieve an experimental 40% reduction in fuel consumption per desired distance by introducing an autonomous vehicle (AV) into free-flowing ring traffic. Despite the positive results, the AV used in the experiment was not truly scalable, as it was programmed with a controller model that predetermined how it

should react, e.g., maintaining a certain headway distance with the preceding vehicle. Such model-based controllers are limited by the bounds of their model, and it be- comes practically impossible to scale control laws for a complex, connected system like AVs in real-world road networks.

However, recent computational developments in reinforcement learning (RL) have enabled the creation of model- free controllers, using a reinforcement learning trial-and-error method to determine which actions lead to the best outcomes in their environment. These methods are particularly effective in sequential decision-making problems where human-designed models cannot anticipate the range of complexity, as with traffic. RL has already led to significant developments in data-rich stochastic problems, such as with three-dimensional locomotion and video games (Song et al., 2021), pushing the field of transportation science towards possible applications to reduce roadway congestion. This project aims to dampen large phantom traffic waves, particularly in uninterrupted stretches of freeway, using only a small concentration of CAVs and without explicitly changing how other human-driven vehicles operate. This intuitive system of connected, model-free AVs has the potential to alleviate overall congestion and related metrics by way of deep reinforcement learning and incorporation of upstream/downstream traffic information. Such model-free techniques will be developed and trained in the Eclipse SUMO microsimulation software, repeating simulations to improve the policy whereby the average velocity of vehicles will be at its highest and the critical density of traffic at its lowest. Once completed, the system can then be used to engage in policy transfer with real-world vehicles and be used to both reduce both vehicular emissions an individual economic delays.

2 Background

2.1 Related Works

Autonomous Vehicle Control: Autonomous vehicles have been studied in three separate contexts - full autonomy, isolated autonomy, and mixed autonomy. Isolated autonomy refers to a system in which a singular vehicle is operated autonomously with individual control. This means that it relies only on computation and on-board sensors such as camwwweras, radar, LIDAR, behavior prediction, computer vision, and traffic following models. Considerable progress has been made in such automation, and companies like Waymo are already testing such vehicles on public roads. Development of isolated autonomy technologies is solely focused on the performance of the individual vehicle rather than its interactions with other nearby autonomous vehicles. On the other hand, full autonomy is where all vehicles in a road system are operated autonomously with collaborative control, each broadcasting data to help other autonomous vehicles make decisions.

But because the most likely future is one of both human drivers and autonomous vehicles on the same road systems, the most relevant of the three control systems is mixed autonomy. A basic form that's already in use is cooperative cruise control, which uses control theory to simultaneously optimize several adjacent vehicles in a platoon of sorts. This research piece focuses on such mixed-autonomy traffic, the unique difference being that the focus is on an entire system-level optimization of velocity rather than optimizing for specific vehicles that are autonomous.

Traffic Applications: Recently, researchers have begun to apply deep reinforcement learning to traffic systems. CARLA, a three-dimensional autonomous simulator, serves as one of the most promising applications. Yet, it's still in an early stage of development and works to test individual autonomous driving systems rather than interactions between multiple systems. Reinforcement learning is also being tested for traffic control, training agents to more efficiently
control speed limit-based controls and ramp meters. However, both applications are fixed in position and macroscopic, thus too broad to respond to data from individual vehicles.

2.2 Phantom Shockwavesw

When applying widely accepted mathematical equations that model traffic similarly to fluid flow, it is well known that traffic is linearly unstable in large densities. A string of equidistant vehicles traveling at the same velocity cannot maintain the velocity if vehicle density exceeds a certain critical threshold, as small perturbations in traffic easily amplify and disrupt the flow (Flynn et al., 2009).

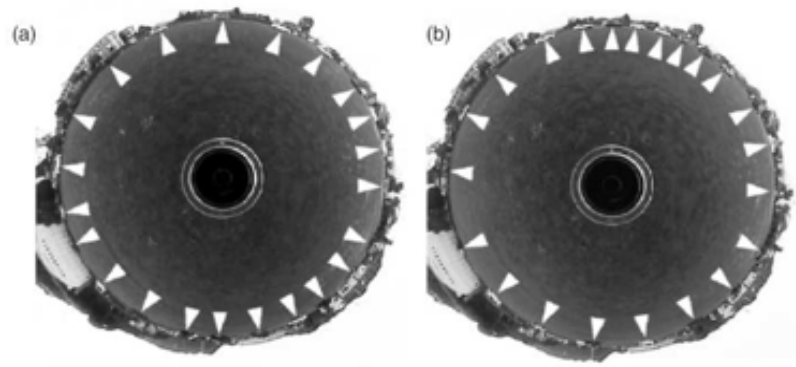


Figure 1: Aerial depiction of the Sugiyama et al. 2008 experiment illustrating the formation of a phantom jam in traffic
a) Positioning of 22 equidistant vehicles at the beginning of the experiment.
b) Positioning of vehicles three minutes into the experiment, displaying a clear phantom jam in the upper right.

Simply put, a small action such as braking can cause the succeeding vehicle to break even harder and so on, creating local peaks of traffic density that propagate back downstream even when the average traffic density of the highway is still only moderate. Many drivers are already aware of this phenomenon, watching the commonly dubbed "stop and go" traffic waves appear and travel downstream even in the absence of any obstacles, bottlenecks, or traffic controllers.

Figure 2: Congestion forming on an open highway, with vehicles being forced to brake upon entering the phantom shockwave and accelerating upon exit.

3 Methodology

3.1 Human Driven Vehicle Dynamics

Because this research focuses on mixed autonomy, the Eclipse SUMO microsimulation software will need a model to replicate human driving behaviors in addition to the autonomous vehicle policy. The longitudinal dynamics for human drivers will be provided by the Intelligent Driver Model (IDM), a microscopic car-following model already built into Eclipse SUMO, which is commonly used to model realistic human driving (Kesting et al. 2010):

$$a_{\mathrm{IDM}} = \frac{dv_\alpha}{dt} = a\left[1 - \left(\frac{v_\alpha}{v_0}\right)^\delta - \left(\frac{s^*(v_\alpha, \Delta v_\alpha)}{s_\alpha}\right)^2\right]$$

$$s^*(v_\alpha, \Delta v_\alpha) = s_0 + \max\left(0, v_\alpha T + \frac{v_\alpha \Delta v_\alpha}{2\sqrt{ab}}\right)$$

aIDM (acceleration of vehicle) is determined by variables va (specific vehicle's velocity), Δva (relative velocity), and sa (distance of preceding vehicle). s* (desired headway for the vehicle) is determined by the set of s0, v0, δ, a, b, which are loaded parameters in the SUMO software specifically calibrated to model open highway traffic. When testing the 5% AV concentration, every 20th vehicle will follow the RL policy while the remaining vehicles will derive their dynamics from the IDM.

3.2 Scenario Setup

This project was conducted under two different testing scenarios, one in a ring-shaped road as seen in previous experiments and another in a multi-lane open freeway network (Figure 3), where vehicles are constantly entering and exiting the environment. For each of these two scenarios, performance metrics of average speed and/or system wide throughout will be gathered. These model-free techniques will be developed and trained in Eclipse SUMO, an open-source traffic simulator commonly used by city planners, integrated with an RL environment that repeatedly runs to improve the policy that dampens phantom shockwaves.

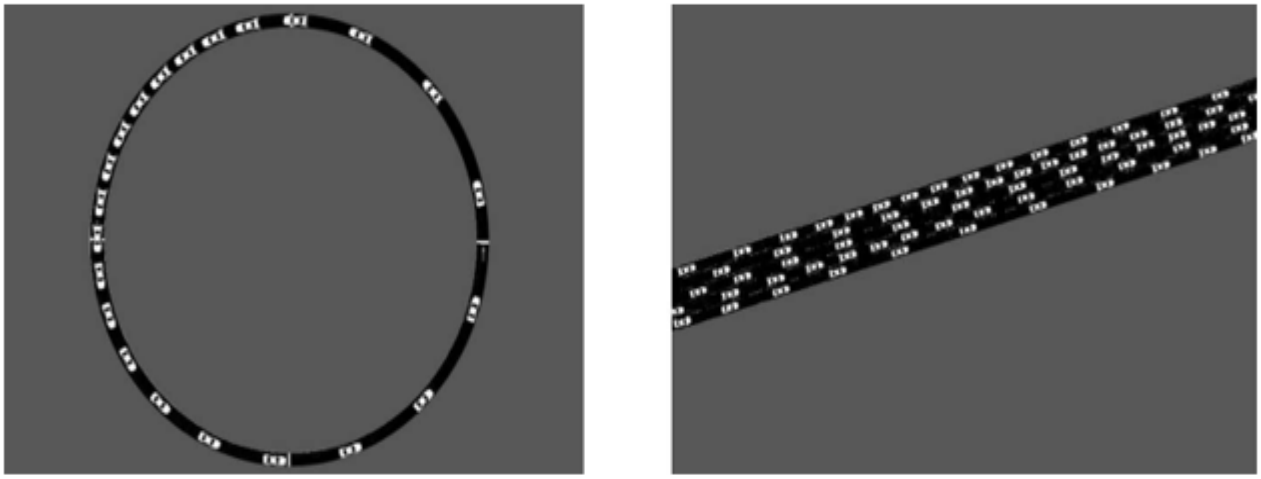


Figure 3: Left: Ring simulation scenario Right: Highway simulation scenario.

The first scenario of a simple traffic ring is designed to study the effects of autonomous vehicles on phantom shockwaves. The circular shape is particularly useful because of its continuous nature, meaning that the exact number of vehicles will be conserved throughout the simulation. As long as the vehicles are distributed at a somewhat realistic distance in regard to real-world traffic conditions, phantom jams will generally form into one single jam and thus only one large shock wave will be observed. This road network will consist of a single lane with a road length of 1,000 meters and 40 equally spaced-out vehicles along the ring.

The second scenario is proposed to test the effects of autonomous vehicles on phantom shock waves in an open highway-like network, where vehicles will continuously enter and exit from the system. This road network will consist of a single lane, horizontal road of length 1,000 meters. The influx of vehicles from the main highway will initially be set to 2000/hour.

3.3 Reward Function - Ring

For the ring-shaped roadway, the reward function seeks to promote high system level speeds in all vehicles, not just those that are autonomous. This trained policy can then be transferred and adapted to the open highway scenario to assess performance under more realistic large-scale testing.

$$R_t = \sum_{i \in CAVs}^{t} max[d_{max} - d(t)]|s_{des} - s(t)|$$

s(t) is equal to a vehicle's speed at timestep t, d(t) is equal to a vehicle's distance with the vehicle in front of it at timestep t, and sdes equal to the desired system-wide velocity. A positive reward value is maintained to penalize any simulation failures caused by vehicle collisions. Congestion will also be penalized, the key indicator being small distances between two successive vehicles.

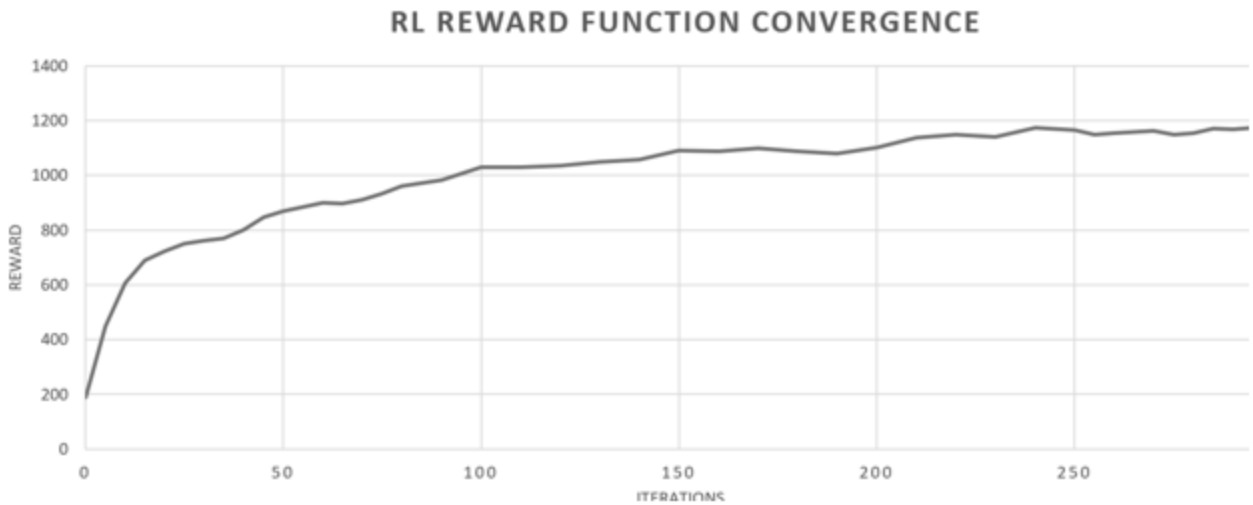


Figure 4: Reward signal convergence of ring scenario.

3.4 Reward Function - Freeway

For the open roadway system, the reward function simply optimizes for vehicle throughput, which is the number of vehicles that are able to exit the simulation space per minute. Because low throughput is a good sole indicator for congestion, optimizing this global reward function should result in the desired improvement. A multiplier of 60 is introduced to change the units from vehicles-per-second to vehicles-per-minute.

$$R_t = 60 \sum_{i \in CAVs}^{t} \frac{n}{t}$$

However, the simple reward function makes it is very difficult to assign credit to which specific vehicle(s)' actions lead to an increase in throughput or system-wide velocity. This problem is amplified by the fact that a vehicle's actions cannot receive immediate feedback from the reward function. Taking a case where one of the CAVs slows down after exiting a phantom jam in ring traffic, it may take another 10 seconds before these actions lead to an actual change in system-wide velocity. Thus, the training period will be a long and intensive process.

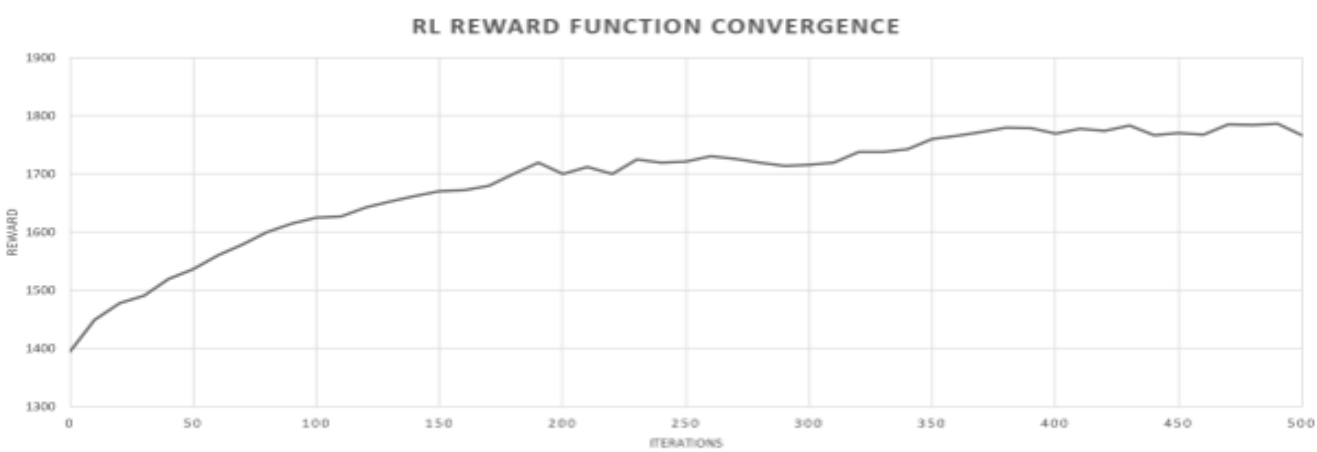


Figure 5: Reward signal convergence of freeway scenario.

4 Experimental Results

For both the ring and freeway scenario, average system-wide speed and system throughput will be gathered under two cases. As a control, all vehicles on the road will be programmed to act as human-operated vehicles. The next test will replace 5% of the human-operated vehicles with autonomous vehicles. Each experiment was conducted over a five-minute time horizon, and final data was averaged from fifty trials after convergence.

4.1 Ring Simulations

Figure 6: Left: Uncontrolled ring simulation (22 Human-Driven, 0 Autonomous)
Right: Controlled ring simulation (21 Human-Driven, 1 Autonomous)

The control experiment for the ring scenario uses 22 human-driven vehicles that are all actively modeled by the Intelligent Driver Model. As expected, one large phantom shockwave forms in the center of the roadway, forcing vehicles to come to a complete stop upon entering. Average system-wide speed for this experiment came out to 22.374 kph. The next experiment replaces one of the 22 human-driven vehicles with an autonomous vehicle, marked by the color red (blue vehicle is observed). As the policy engages, the previously visible phantom shockwave completely disappeared. This resulted in an average system-wide speed of 28.391 kph, a 26.90% increase over the control.

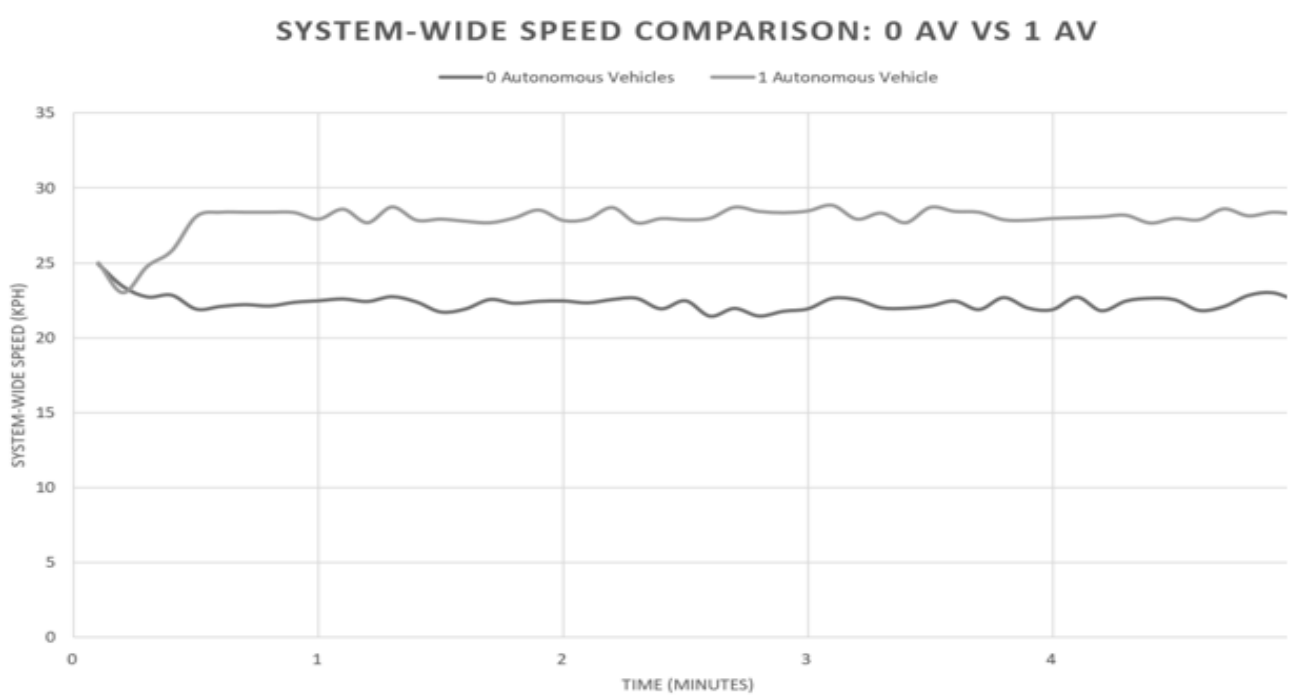


Figure 7: System-wide speed comparison of ring scenario.

4.2 Freeway Simulations

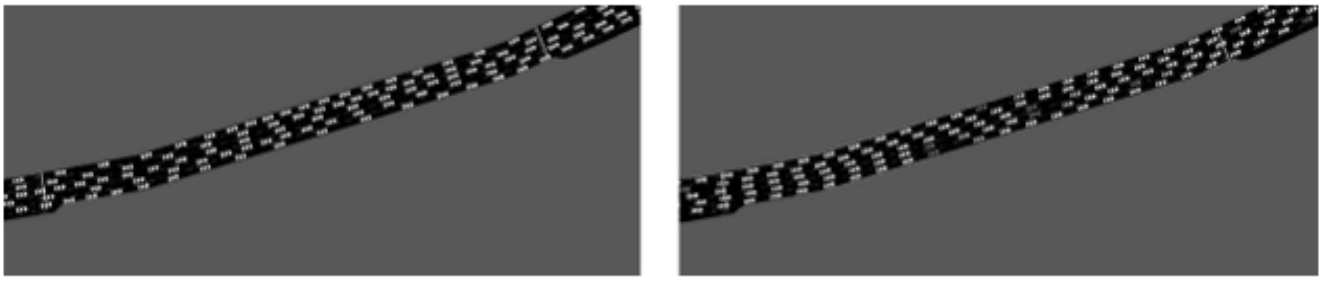


Figure 8: Left: Uncontrolled freeway simulation (100% Human-Driven, 0% Autonomous)
Right: Controlled freeway simulation (95% Human-Driven, 5% Autonomous)

The control experiment for the freeway scenario uses 100% human-driven vehicles modeled by the Intelligent Driver Model. Multiple visible phantom jams form in each lane, causing the flow of traffic to be uneven and slow. The average system-wide speed came out to 47.681 kph and system throughput was 121.310 veh/min. The next experiment replaces 5% of the oncoming human-driven vehicles with autonomous vehicles. Once the first oncoming autonomous vehicles reach the end of the simulation frame, all previously visible phantom shockwaves are dampened, and traffic completely smooths out. This resulted in an average system-wide speed of 52.325 kph, a 9.74% increase over the control. Additionally, the throughput of this new experiment was 138.705 veh/min, a 14.34% increase over the control.

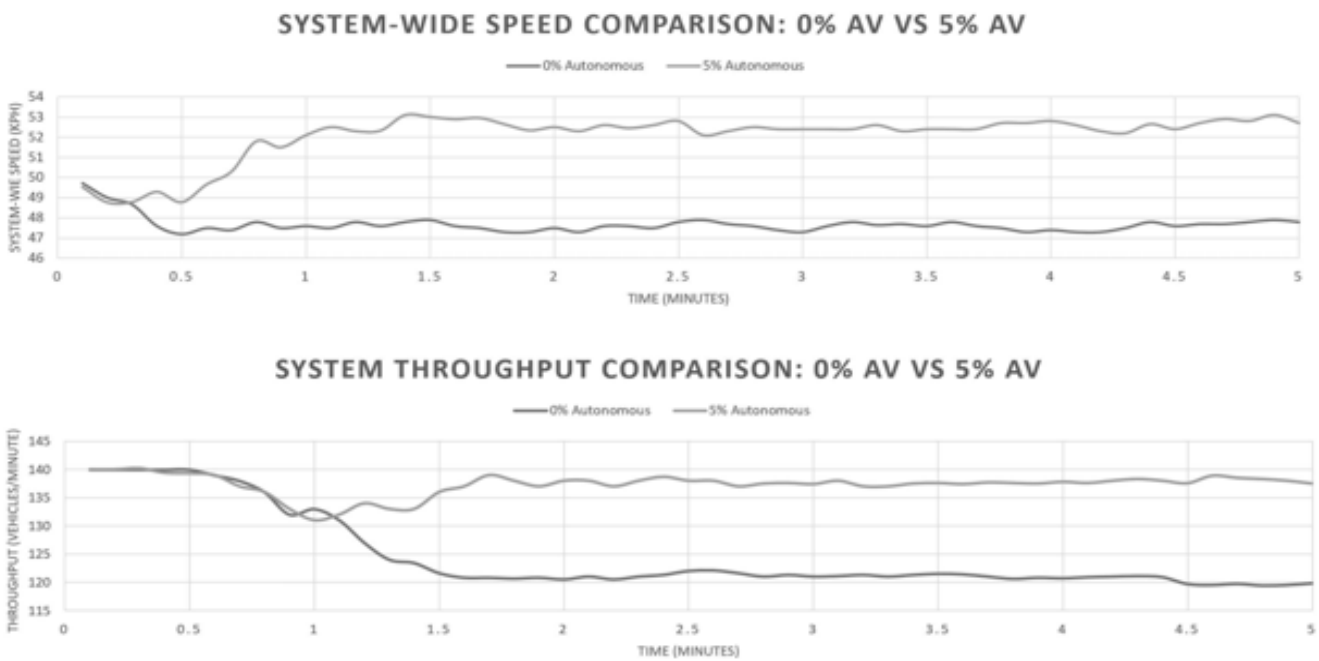


Figure 9: System-wide speed and system throughput comparison of freeway scenario.

5 Conclusion

The engineering goal of dampening phantom shockwaves in mixed autonomy traffic systems has been achieved, increasing average system-wide vehicle speeds in both ring and open freeway roads by over 9%. Previously visible phantom shockwaves from control experiments were completely dampened once the five percent concentration of autonomous vehicles were introduced into their respective simulations. If even a fraction of these increases can be transferred to real freeways, it would save massive amounts of energy and prevent millions of metric tons of carbon dioxide from entering the atmosphere each year. To achieve this policy transfer, the policy developed in this research would need to be trained on additional edge cases, including but not limited to intersections, pedestrians, traffic signals, accidents, and merges.

References

Chan, S. C., Fishman, S., Canny, J., Korattikara, A., & Guadarrama, S. (2019). Measuring the reliability of reinforcement learning algorithms. arXiv preprint arXiv:1912.05663.

Cookson, G., & Pishue, B. (2017). Inrix global traffic scorecard–appendices. INRIX research.

Flynn, M. R., Kasimov, A. R., Nave, J.-C., Rosales, R. R., & Seibold, B. (2009). Self-sustained nonlinear waves in traffic flow. Physical Review E, 79(5), 056113.

Keskin, M. F., Peng, B., Kulcsar, B., & Wymeersch, H. (2020). Altruistic control of connected automated vehicles in mixed-autonomy multilane highway traffic. IFAC-PapersOnLine, 53(2), 14966–14971.

Kesting, A., Treiber, M., & Helbing, D. (2010). Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368(1928), 4585–4605.

Lazar, D. A., Coogan, S., & Pedarsani, R. (2020). Routing for traffic networks with mixed autonomy. IEEE Transactions on Automatic Control, 66(6), 2664–2676.

Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., Goldberg, K., . . . Stoica, I. (2018). Rllib: Abstractions for distributed reinforcement learning. In International conference on machine learning (pp. 3053–3062).

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., . . . Wießner, E. (2018). Microscopic traffic simulation using sumo. In 2018 21st international conference on intelligent transportation systems (itsc) (pp. 2575–2582).

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. arXiv preprint arXiv:1706.02275.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. In International conference on machine learning (pp. 1889–1897).

Song, S., Kidzin´ski, Ł., Peng, X. B., Ong, C., Hicks, J., Levine, S., . . . Delp, S. L. (2021). Deep reinforcement learning for modeling human locomotion control in neuromechanical simulation. Journal of neuroengineering and rehabilitation, 18(1), 1–17.

Sprung, M. J., Chambers, M., et al. (2015). Two decades of change in transportation: Reflections from transportation statistics annual reports, 1994-2014.

Stern, R. E., Cui, S., Delle Monache, M. L., Bhadani, R., Bunting, M., Churchill, M., . . . others (2018). Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. Transportation Research Part C: Emerging Technologies, 89, 205–221.

Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., . . . Yukawa, S. (2008). Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. New journal of physics, 10(3), 033001.

Tajalli, M., & Hajbabaie, A. (2021). Traffic signal timing and trajectory optimization in a mixed autonomy traffic stream. IEEE Transactions on Intelligent Transportation Systems.

Treiber, M., & Kesting, A. (2017). The intelligent driver model with stochasticity-new insights into traffic flow oscillations. Transportation research procedia, 23, 174–187.

Wu, Y., Tan, H., & Ran, B. (2018). Differential variable speed limits control for freeway recurrent bottlenecks via deep reinforcement learning. arXiv preprint arXiv:1810.10952.

ScGAN: A Generative Adversarial Network to Predict Hypothetical Superconductors

Evan Kim

Nikola Tesla STEM High School

Abstract

Superconductors are materials that exhibit zero resistance and expel magnetic fields once cooled below a certain critical temperature, allowing for their application in various technologies such as quantum computers and efficient transmission lines. Of particular interest are High Temperature Superconductors (HTSs), superconductors with a high critical temperature, which lack both an explanation for their mechanisms and a systematic way to search for them, despite having been first discovered over three decades ago. This project proposes ScGAN, a Generative Adversarial Network (GAN) to efficiently predict new superconductors. ScGAN was trained on compounds in the Open Quantum Materials Database and then transfer learned onto four different datasets of superconductors: all of the SuperCon database and three subsets of it chosen by superconducting class. Once trained, 30,000 samples were generated using the version trained on all of SuperCon, and approximately 70.42% of them were determined to be superconducting–a 23-fold increase in discovery rate compared to manual search methods. Furthermore, 99.69% of the samples were novel, demonstrating that ScGAN was able to find new superconductors, including several promising HTS candidates. The other versions of ScGAN also displayed high superconductivity and novelty percentages and predicted superconductors that were similar to their training data, indicating that ScGAN learned the different features of the different classes of superconductors. This project shows a novel, efficient way to search for superconductors, which may be used to find suitable superconductors that can be utilized in technological applications or provide insight into the unsolved problem of high temperature superconductivity.

I. INTRODUCTION

Superconductors, which were first discovered in 1911 by Kamerlingh Onnes, have several defining properties such as zero resistance and being able to expel magnetic fields via the meissner effect. These phenomena allow for a variety of technological applications such as efficient transmission lines, maglev trains, and quantum computers. Quantum computers are especially important as they are believed to be able to solve problems that are too computationally expensive for current computers. However, to be used in these applications, the superconductor must be cooled below a specific critical temperature (Tc) at which point it transitions into a superconducting state. For instance, the components in Google's quantum computer must be maintained at an unreasonably low temperature, around 0.02 K, which is not at all practical for general use (Arute et al., 2019). Thus, superconductors with a high Tc–High Temperature Superconductors (HTSs)–are needed, which has made them a hot research topic for the last couple of decades.

The first HTS was found in 1986, and it did not fit the existing BCS theory of superconductivity (Bednorz & Muller, 1986). And to this day, the mechanisms through which they arise remains an unsolved problem, which means that there is no systematic way to find high temperature superconductors nor predict their critical temperatures (Hirsch et al., 2015). Thus, the current process for finding HTSs is essentially a blind guess and check, which is extremely inefficient. This was exemplified in one study implementing this strategy, which found that only ~ 3% of the approximately 1000 materials they surveyed over 4 years were superconducting (Hosono et al., 2015). Further-

more, they failed to find any superconductor with a Tc > 60 K (i.e. they could not find a very high temperature superconductor). This extreme inefficiency means that the likelihood of manually finding suitable superconductors for applications is extremely low.

To address this difficulty, the use of computational tools has become quite popular recently. In particular, there have been several studies utilizing machine learning to predict whether a given chemical compound will be superconducting or not based on only its composition. The earliest such study was done by Stanev et al. in which two random forest-based models were built–a classification model for predicting superconductivity and a regression model for predicting superconducting transition temperature (Stanev et al., 2018). They were quite successful, reporting a 92% accuracy in classification and a R2 value of 0.88 for their regression model. They did run across one limitation of their machine learning model, however. When trained on a certain class of superconductors (e.g., cuprates), the model was unable to make good predictions on other classes of superconductors. Afterwards, there were several other studies applying similar methods, with a K-nearest neighbors algorithm from Roter and Dordevic and a deep learning model, which was the first of its kind, from Konno et al., just to name a few. The K-nearest neighbors algorithm reported improvements on the previous study from Stanev et al. in terms of overall performance: an R2 of 0.93 and a classification accuracy of 96.5%. The deep learning model, on the other hand, showed that it could be possible to overcome the limitation that Stanev et al. faced, as they were able to make predictions about iron-based superconductors from training data not including them.

The way each of these previous studies predicted superconductors was by running their trained models on a database of known existing chemical compounds and finding which compounds the model indicated could be superconducting. Therefore, they miss out on the vast chemical composition sample space that is not already contained in existing databases. Furthermore, the discovery of superconductors in the lab does not happen in this manner; they are discovered by constructing new materials and testing them (rather than checking known ones), as the researchers who first discovered iron-based superconductors did (Kamihara et al., 2008). A generative model would address both of these issues as it generates compounds, acting as humans would in a lab, rather than checking known compounds.

Generative models refer to a general class of machine learning models which are able to generate things that resemble datasets you feed into them. While generative models for finding superconductors do not yet exist, there exist some for finding hypothetical compounds in general (not superconductors). For example, Dan et al. reported a generative adversarial network (GAN) model which generated compounds with 92.53% novelty and 84.5% validity. They reported that a GAN worked better than other generative models, such as a variational autoencoder, because it was able to learn more hidden rules. Thus, combining the ideas from this generative model and the previous superconductor models, we propose a GAN for superconductors based on chemical composition.

II. METHODOLOGY

As stated in the introduction, we chose the GAN (structure shown in fig. 4) as our type of generative model. For training the model (process depicted in fig. 1), the main idea was to first have the model learn general chemical composition rules by training off of a larger dataset of regular chemical compounds and then transfer learning it onto a dataset of known superconducting materials, which is significantly smaller. The idea of transfer learning is to allow a model to learn even with a limited amount of data, which is the case here as the general compound dataset has on the order of ~106 data points while the superconductor dataset only has on the order of ~104 data points.

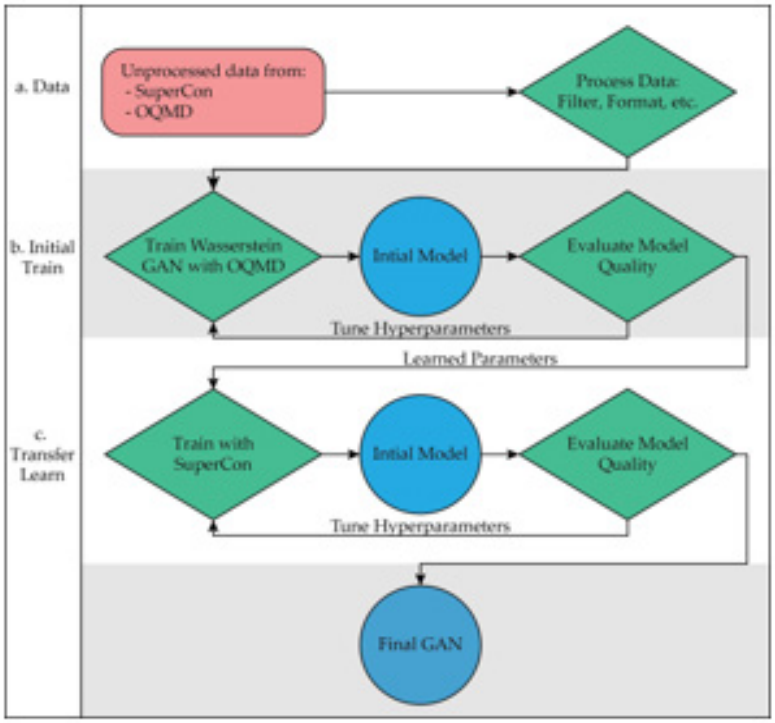


FIG. 1. Depicts the overall training / testing process from the data to the final GAN model. It is essentially composed of three main stages, (a) data processing, (b) training on the OQMD dataset, and then (c) transfer learning onto the superconductor dataset.

A. Data Collection

Data were sourced from two open-source databases: SuperCon (National Institute for Materials Science, n.d.) and the Open Quantum Materials Database (OQMD) (Kirklin et al., 2015)(Saal et al., 2013). SuperCon is the largest database for superconducting materials with around 30, 000 superconductors before filtering. For this project, only the actual chemical compositions of the materials were extracted from SuperCon, though there are other pieces of data such as critical temperature in there. Unfortunately, other potentially useful data such as crystal structure is not well-documented in SuperCon, which in a way forced the use of chemical composition only. OQMD is a slightly larger database with around 106 DFTcalculated compounds, and its RESTful API was queried for around 5, 000 compounds at a time for chemical compositions.

B. Data Processing

Before filtering the data, we had to be put the compounds in a common format so that unwanted datapoints would be caught regardless of small differences in formatting across the databases, such as the ordering of elements. First, each of the datapoints were formatted as

$$\begin{array}{cc} & \text{Data Points} \\ \text{Features} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0.5 & \cdots \\ 0 & 0.2 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0.5 & 0 & 0 & 3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{array}{c} \text{H} \\ \text{He} \\ \text{Li} \\ \text{Be} \\ \vdots \end{array} \end{array}$$

FIG. 2. Chemical composition data represented as a matrix in R^(96×m), where m is the number of datapoints. Each column is a single compound, with each entry representing the number of each element present in the compound. Note that that the numbers in the matrix above are random; they're not real compounds or superconductors.

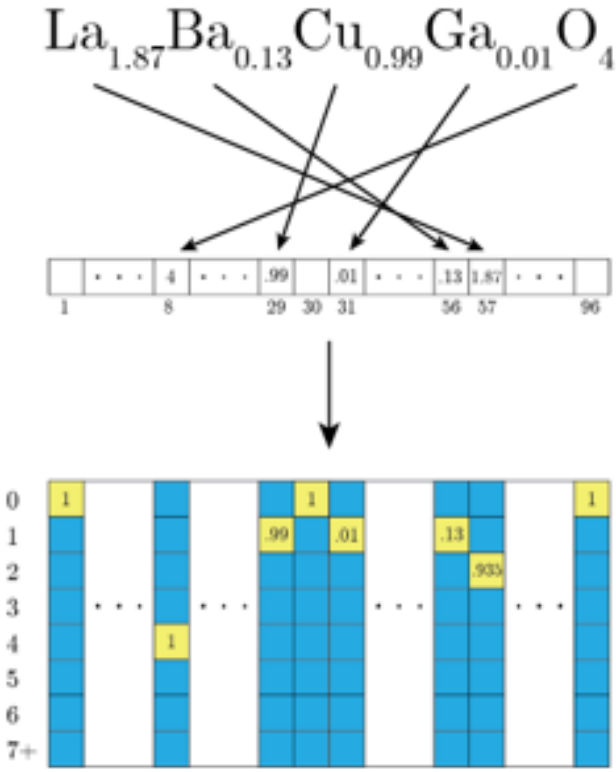


FIG. 3. A "one-hot" encoding of the chemical composition as a matrix in R^(96×8). First, each coefficient is mapped to the corresponding atomic number in the linear vector. Then, each nonzero component x of the linear vector is expanded into an 8 dimensional vector by finding the nearest positive whole number k to x in 1 through 7 and placing a x/k at the index k. For zero components, a 1 is simply placed in the 0 index. The idea here is that the amount of each compound is encoded through both the vertical location of the yellow box and the value in it, which allows the GAN to learn the chemical compositions better.

1 × 96 matrices (the 96 is the maximum atomic number present across all the compounds) and then combined so that each dataset was a matrix in R^(96×m), where m is the number of data points (see fig. 2). Then, both datasets were filtered for duplicates, which reduced each of them to around 800,000 datapoints and 30,000 datapoints. The SuperCon dataset required further processing as several entries were incomplete, and so these were either corrected or removed. Lastly, the SuperCon dataset was split into several different groups. It was split was into different classes– iron-based (IBSCs), cuprates, and others (anything else that isn't iron-based or a cuprate). Table I has the quantitative counts of these different groups. The purpose of this was to test the model's ability to learn the different classes of superconductors, especially the high temperature classes of cuprates and IBSCs. Splitting the IBSCs or pnictides off also allows us to simulate the discovery process that actually happened, since IBSCs were discovered last out of the different groups. (the idea is if the model trained without IBSCs generates an IBSC, then it "discovered" them).

Once the data was filtered, the chemical compositions were then transformed into a form for the GAN to train on. An adaptation of the encoding by Dan et al. was used, a 96 by 8 matrix of real numbers (fig. 3). Different formats were tested, such as the one that Konno et al. used, but these were susceptible to mode collapse.

Class	Quantity	Percentage
Cuprates	7,304	44.4%
Pnictides	1,436	8.7%
Others	7,749	47.0%
Everything	16,489	100%

TABLE I. Distribution of the superconductors in our filtered version of SuperCon by class. Each of these served as different training sets.

C. Model

A GAN is a type of generative model that has two competing neural networks, a generator and a discriminator, as shown in figure 4. Traditional GANs, however, can suffer from issues such as mode collapse and gradient vanishing, so a slightly modified architecture was used

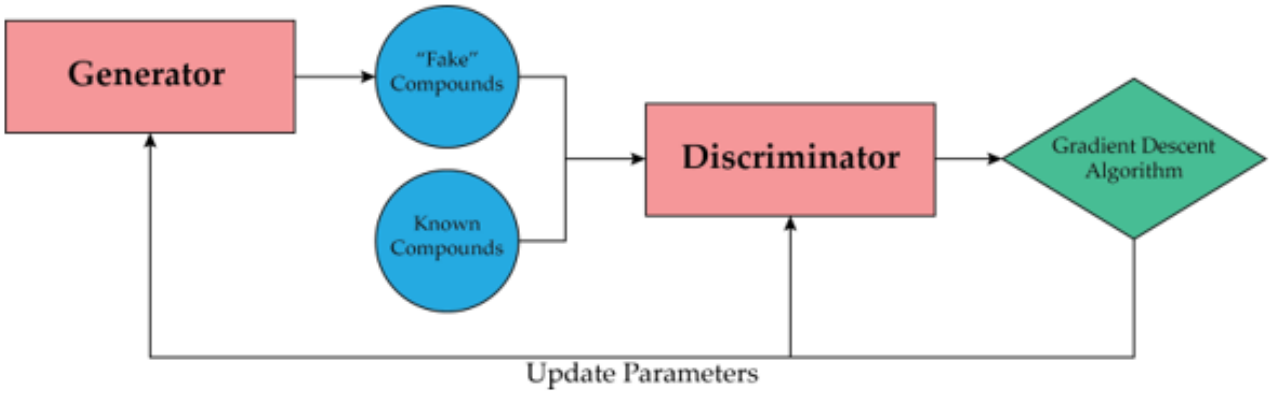


FIG. 4. This is the architecture of the GAN model used. It is composed of two competing neural networks, the generator and discriminator. The generator takes in random noise and generates a "fake" compound, which the discriminator attempts to determine if it is real or not using existing data of real compounds. Then each of them updates their parameters based on the performance of the discriminator.

for the gradient descent: the Wasserstein GAN with Gradient Penalty (Gulrajani et al., 2017). In the Wasserstein GAN with gradient penalty, the loss functions are

$$\text{Loss}_D = \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}[D(\tilde{\boldsymbol{x}})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathbb{P}_r}[D(\boldsymbol{x})] + \lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{\boldsymbol{x}}}}[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\| - 1)^2]$$

$$\text{Loss}_G = -\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}[D(\tilde{\boldsymbol{x}})]$$

Then the parameters are updated with an optimizer,

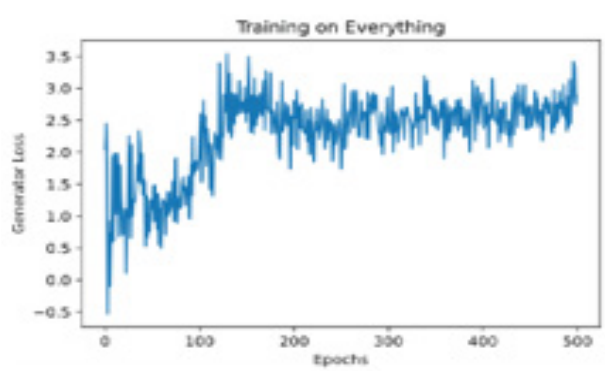
$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \alpha \cdot \textbf{Optimizer}(\boldsymbol{w}, \nabla_{\boldsymbol{w}} \textbf{Loss}_D),$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha \cdot \textbf{Optimizer}(\boldsymbol{\theta}, \nabla_{\boldsymbol{\theta}} \textbf{Loss}_G),$$

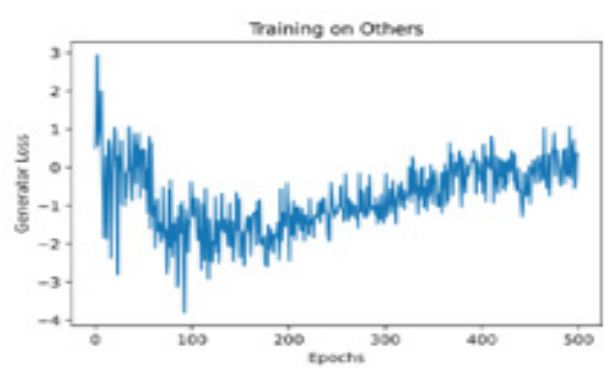
which was chosen to be RMSProp after testing out a few different optimizers such as Adam (Hinton, 2012).

1. Training

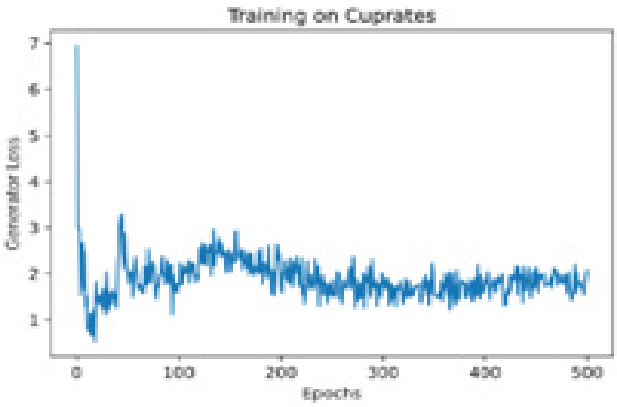
The model was first trained on the OQMD dataset for 400 epochs. Then, it was transfer learned onto the SuperCon dataset or a subset of it, on which it would train for another 500 epochs. The four different datasets (cuprates, pnictides, others, and everything together) resulted in four different versions of the GAN each time. Afterwards, the testing procedure and the data analysis was conducted, and then the hyperparameters were updated based on the results. The training curves for the final model on each of these sets is displayed in figure 5. Notably, they are all able to converge and stabilize over the 500 epochs.



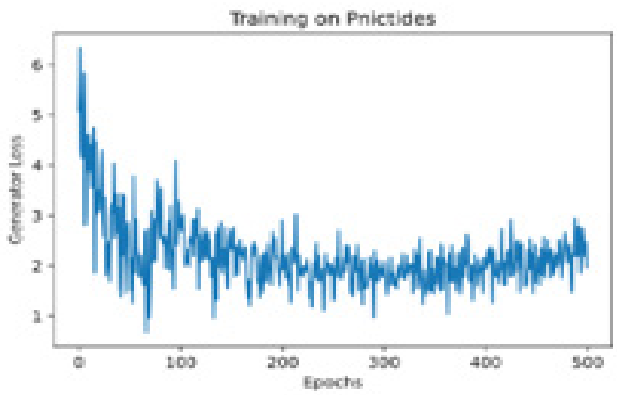
(a)



(b)



(c)



(d)

FIG. 5. The generator loss against training epoch for each of the four datasets the GAN trained on: (a) All of SuperCon; (b) Others (i.e. not cuprates or pnictides); (c) Cuprates; and (d) Pnictides.

2. Testing

After each training process, 5, 000 hypothetical compounds were generated from the versions of the GAN trained on the smaller sets and 30, 000 from the version trained on everything. Then, each of the generated predictions were inspected for various quality checks. Each compound was first tested for validity using the charge and electronegativity check feature of the SMACT package (Davies et al., 2016). The package tests the compound for electronegativity and charge balance to determine whether it is a valid compound or not. Then, each compound was checked for uniqueness–whether the compound showed up earlier in the generated list–and novelty–whether the compound was in the training dataset. These three checks looked at the general quality of the model. Then, more specific to superconductivity, each compound was run through the model from Roter and Dordevic to predict its critical temperature and whether or not it's a superconductor. Lastly, the formation energy of each generated compound was calculwated using the model from Jha et al., which indicates the stability of the compound. These tests will be discussed further in the results section along with the actual results of these tests.

D. Data Analysis

First, to evaluate the overfitting and discriminability of the model, the Fréchet Inception Distance was used, which is defined as

$$FID = \left\|\mu_r - \mu_g\right\|^2 + T_r\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right),$$

where Σ and μ are the covariance and mean of the real (r) and generated (g) datasets. The lower the score, the more closely the generated set resembles the real dataset and the more diverse it is.

1. General Statistics

After collecting the raw data, basic statistical measures were calculated. The mean and standard deviation for each quantity X were calculated according to the following:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\sigma_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - X)^2}\,.$$

The percent superconducting, novel, and unique were of course simply calculated according to

$$\frac{\#\ \text{superconducting / unique / novel}}{\#\ \text{samples}} \times 100\%.$$

2. Analyzing the Ability to Learn the Different Classes

The generated list was then inspected closely. Each superconducting compound was classified into pnictide, cuprate, or other. Then, for each training set, the number of compounds generated

that were part of one of the classes in the training set were counted, and a percentage was calculated off of that. A higher percentage meant that that version of the GAN learned the features in its training data better.

3. t-SNE visulization

The datapoints in this project are all 96-dimensional because each compound can be represented as a 1 by 96 matrix as discussed in II B. Thus, to visualize this higher dimensional sample space, t-SNE is used (van der Maaten & Hinton, 2008). The t-SNE visualizations of both the training set and generated set were created (fig. 8) to see how the GAN filled in the sample space.

III. RESULTS

After training, from the 4 different versions of the GAN, we generated 4 superconductor candidate lists of around 5 to 30 thousand chemical compositions. We then ran the predicted superconductors through a series of tests to evaluate their quality. The first few were general tests and then the rest were in the context of superconductivity using existing computational models (as physically testing all of them is infeasible).

A. Duplicates and Novelty

We first screened the output for duplicates within the generated sets themselves and the duplicates between the generated set and the dataset of known superconductors that it trained on. The results are tabulated in table II. As seen in the table, the number of repeats within the generated samples were relatively low (high uniqueness), with the exception of pnictides, which had quite a bit more duplicates than the rest but still had a relatively high uniqueness percentage. This is likely due to fact that the dataset of pnictides was significantly smaller than the rest (see table I). This low rate of duplicates overal most likely stems from the fact that the model is able to handle decimals, which means there is a massive sample space of superconductor compositions for it to explore. The percent of predicted compounds that were novel, i.e. not already known to exist, listed in table II, is also very high across all 4 GANs. So, these two results demonstrate that all versions of ScGAN can generate a diverse array of new compositions.

GAN Version	Novel %	Unique %
Entirety of SuperCon	99.69%	96.78%
Cuprate	99.74%	92.98%
Pnictides	99.32%	58.74%
Others	98.89%	91.58%

TABLE II. The percentage of generated samples that were novel (not in the training set) and the percentage of generated samples that were unique (distinct from others in the given generated set) for each of the versions of ScGAN trained on the given training datasets on the left.

B. Formation Energy

As mentioned earlier, we found the formation energies of the predicted compounds from the GANs using Elemnet. We see that in the distributions in figure 6, most of the compounds have predicted formation energies less than 0. And while this isn't a perfect indicator of stability, it still is a general indication that most of the generated compounds are stable and able to be created.

C. Superconductivity

Then, we ran the predicted superconductors through the K-Nearest Neighbors (KNN) classification model from Roter and Dordevic for predicting superconductivity based on elemental composition, to see how many were really superconducting. However, the probabilistic nature of the machine learning model had to be taken into account. If psc is the proportion of predictions that came up superconducting according to the model, then we can estimate ρsc, the true proportion of superconducting elements, using Bayesian statistics. Denoting tp and fp as the true positive and false positive rates of the classification model, respectively, the following estimate can be made:

$$\rho_{sc} \cdot tp + (1 - \rho) \cdot fp \approx p_{sc}.$$

Solving for ρ_sc gives

$$\rho_{sc} \approx \frac{p_{sc} - fp}{tp - fp}.$$

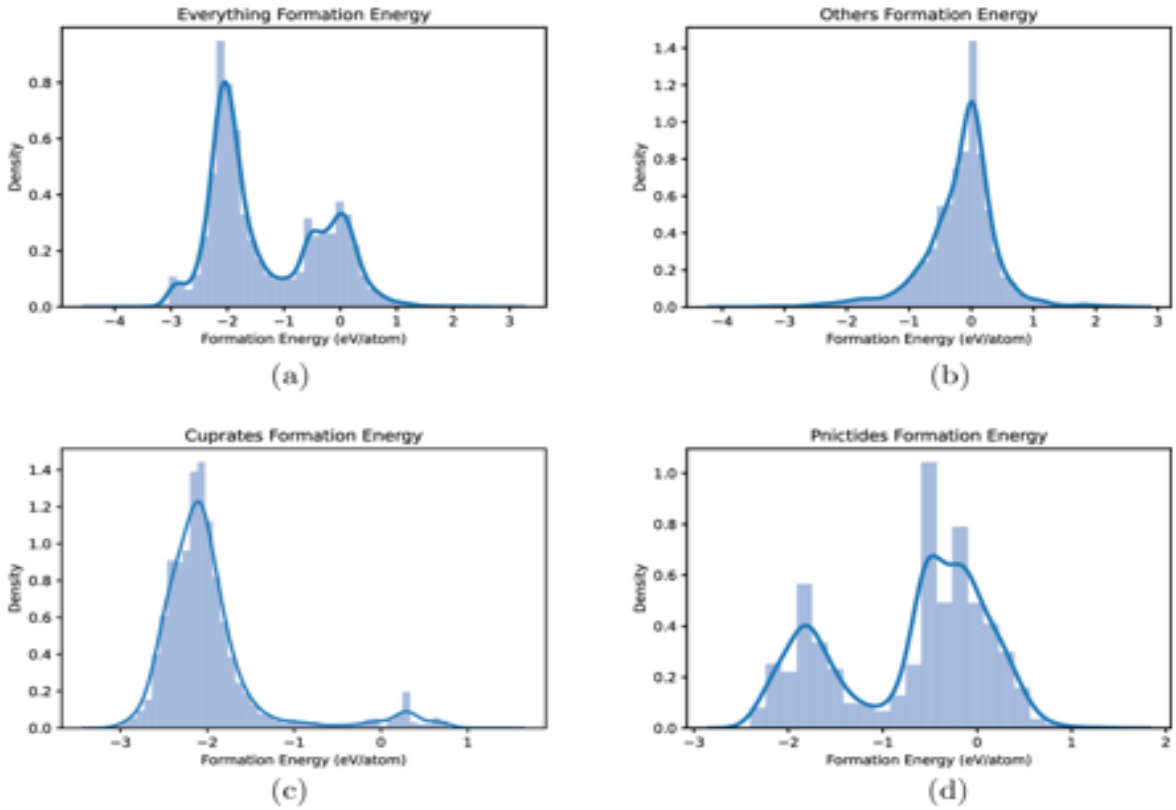


FIG. 6. Distribution of the formation energies of the predicted compounds from the four versions of the GAN: (a) everything, (b) others, (c) cuprates, (d) pnictides.

The true positive and false positive rates here are reported from the paper: tp = 98.69% and fp = 16.94% (Roter & Dordevic, 2020). The output percentages along with the estimates of the true proportions calculated from equation 10 are tabulated in table III.

All 4 GANs achieved very high percentages of superconducting material, especially when compared to the 3% figure from the manual search done by Hosono et al. However, since the evaluation was done with a computational model, there is still some amount of uncertainty, so we should eventually physically test these predictions for superconductivity.

D. Critical Temperature Estimates

We also looked at the critical temperatures of these predictions using the regression model from Roter and Dordevic. Again, just like the superconductivity tests, these values cannot be taken entirely truthfully because of the reported RMSE of approximately 8.91 K. However, since that margin of error is not that large, the regression model still provides us with a general understanding of the capabilities of ScGAN.

GAN Version	Novel %	Unique %
Entirety of SuperCon	74.50%	70.42%
Cuprate	75.76%	71.95%
Pnictides	72.44%	67.89%
Others	69.58%	64.39%

TABLE III. The percentages of generated predictions that were determined by the KNN model to be superconducting for different training sets along with the estimated real percentage of the predictions that were superconducting. The true percentages were estimated according to equation 10.

The critical temperature outputs from the model are summarized in table IV and the distributions are in figure 7. While the distributions are quite spread, the GANs were still able to find several superconductors with predicted critical temperatures of upwards of 100 K, which exceeds the manual search 58 K and several of the previous machine learning approaches. This makes sense as they were limited to those databases, which meant that these forward design approaches would only come up with a handful of candidates with low critical temperatures.

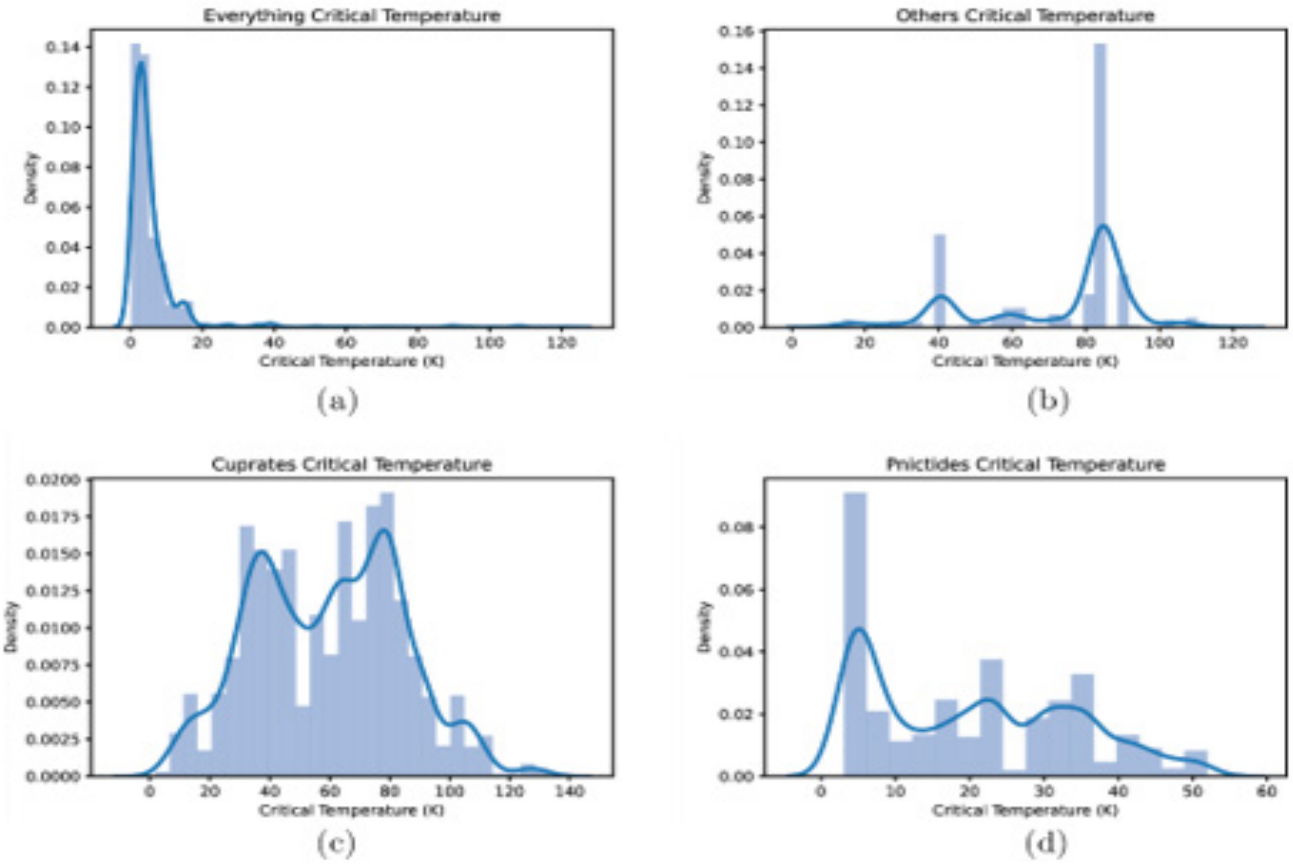


FIG. 7. Distributions of the critical temperatures of the predictions of the four different versions of Sc-GAN: trained on (a) everything, (b) others, (c) cuprates, (d) pnictides.

Training Data	Average T_c	Standard Dev.	Max T_c
Entirety of SuperCon	6.53 K	11.76 K	123.25 K
Cuprate	59.34 K	24.78 K	133 K
Pnictides	20.41 K	13.69 K	51.98 K
Others	72.68 K	21.24 K	116.55 K

TABLE IV. Summary statistics of the predicted critical temperatures of the generated predictions that were determined by the regression model for different training sets.

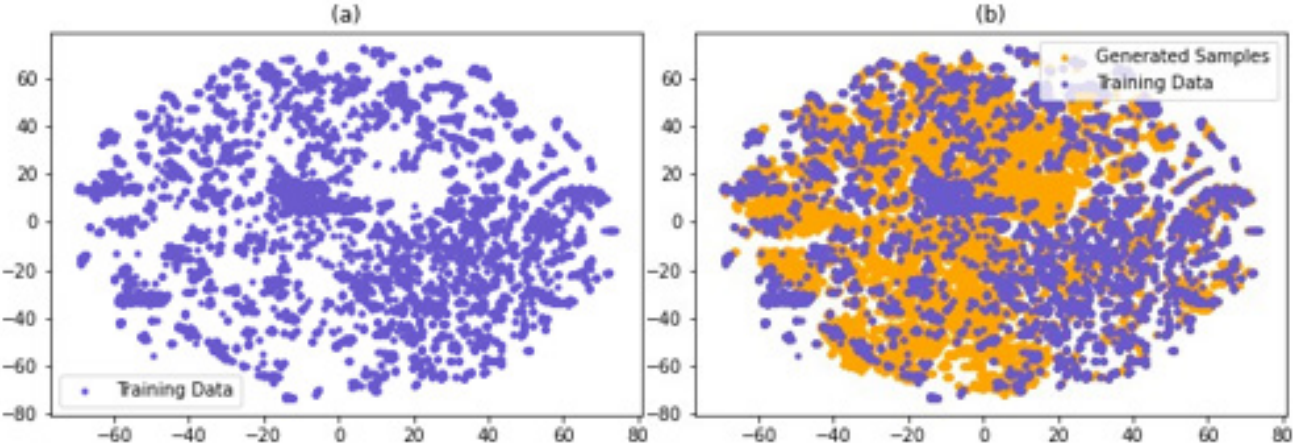


FIG. 8. The t-SNE visualizations of (a) the training set and (b) the training set along with the generated set.

E. Sample Space

The sample space of the compounds is 96-dimensional, so to be able to visualize it, we used t-SNE (van der Maaten & Hinton, 2008). The visualizations for the training data along with the generated samples are in figure 8. We can see that the generated samples fill in the gaps in the sample space with only the training set, indicating that ScGAN is generating predictions that fall in line with the current existing superconductors and is filling in the missing pieces.

Training Data	Cuprate %	Pnictide %	Other %
Cuprate	92.76%	0.06%	7.18%
Pnictides	0.02%	99.84%	0.14%
Others	0.14%	0.6%	99.26%

TABLE V. The distribution of the predicted superconductors across the different classes of superconductors for the different versions of the GAN.

F. Ability to Learn Features

We then looked at the types of superconductors that were generated by the different versions of ScGAN. The distributions are given in table V, and we can see that each the versions generated nearly only superconductors that matched their training data. This indicates the GAN was able to pick up on the different underlying features behind these different major classes of superconductors.

G. Promising Candidates

After running through the Tc prediction model, we identified the ones that looked the most reasonable. Some of them even had very high critical temperatures, greater than the boiling point of liquid nitrogen (77 K), which means they can be cooled with it. It turns out that most of these were cuprates, which makes sense, considering the superconductors with highest critical temperatures are currently mostly cuprates. We then checked the Crystallography Open Database (COD) to see if those compounds were in the database. The ones listed in table VI could not be found in COD or SuperCon, showing that this model overcomes the limitations of the previous forward design models and finds completely new superconductors.

IV. CONCLUSION

The search for High Temperature Superconductors has gone on for several decades now, and higher and higher critical temperatures have been found. However, the process by which these new materials were found were not efficient at all, because, at their core, they are guess and check. Now, aided with computational tools, this search can be made more efficient. This paper introduced a novel method to search for superconductors—discovering candidates with a generative adversarial network. In contrast to previous computational methods which attempted to predict new superconductors off of existing datasets, this model predicted compounds directly. This "inverse design" approach proved to be far more powerful than manual search methods and previous computational methods, with the model being able to generate thousands of candidates with a wide-range of critical temperatures that lied outside of our existing databases.

ScGAN was trained on the chemical formulas of the existing superconductors in SuperCon and no other properties. But even with that one piece of data, the model was able to successfully learn the compositional rules of superconductors and generate many samples that were cross-checked with a separate model to be superconducting. The version of the GAN trained on all of SuperCon achieved success rates of 70.42%, exceeding the manual search success rate by 23-fold. Furthermore, they were also new superconductors, as checking against SuperCon, 99.69% of the generated predictions were novel. Several excellent HTS candidates that were not in COD were found, with critical temperatures of up to 133 K. Previous models would have been unable to find such candidates as they were outside of current databases. Furthermore, the critical temperature of 133 K exceeds the maximum that many of the forward design approaches found, demonstrating the powerful capabilities of this kind of model.

Compound	Predicted T_c	Class
$PrCaBiSr_2Cu_2O_{7.46}$	104.6 K	Cuprates
$YTiSr_2Cu_{2.74}O_{6.76}$	91.7 K	Cuprates
$TeYSr_2Cu_2O_{7.75}$	89.8 k	Cuprates
$C_{2.52}Ni_{0.92}Y_{0.71}Th_{1.0}$	85.3 K	Others
$Si_{0.62}V_{0.91}Zr_{0.83}$	84.7 K	Others
$Al_{2.34}Te_{0.64}Ir_{1.07}$	84.7 K	Others
$TlCaASr_2Cu_2O_{7.82}$	73.9 K	Cuprates
$YCaBa_2ZnCu_{2.36}O_{7.54}$	71.5 K	Cuprates
$HgCsSrCa_2Cu_{2.56}O_{8.66}$	69.8 K	Cuprates
$Be_{0.16}Si_{1.09}V_{2.67}Y_{1.72}$	62.4 K	Others
$Cu_{1.13}Nb_{3.0}Sb_{0.72}Ir_{1.05}$	59.4 K	Others
$GdCaRuSr_{1.83}Cu_2O_{8.71}$	40.8 K	Cuprates
$Ga_{0.62}Nb_{2.88}Sn_{0.65}Te_{0.79}$	40.8 K	Others
$B_{1.73}C_{1.03}Ni_{1.12}Y_{0.66}Pt_{0.64}$	40.8 K	Others
RuTeSeFe	35.6 K	Pnictides
$TeSSeFe_{1.05}$	31.0 K	Pnictides
$CeCoAs_{2.15}Fe_{1.39}$	23.3 K	Pnictides
$CeThPAsFe_{1.59}$	12.2 K	Pnictides
$GaPrCa_{2.58}As_{12.44}Fe_{6.34}$	11.9 K	Pnictides
NdOAsFe	4.5 K	Pnictides

TABLE VI. Promising HTS candidates that do not exist in current databases.

We are currently looking to begin physically testing the candidates (an excerpt of them shown in table VI), so that the computational methods can be completely confirmed. There are also some improvements that can be made with the computational method itself, such as being able to account for charge and crystal structure in the compound encodings. Chemical checks on the predicted compounds (using SMACT as detailed earlier) revealed that while there was electronegativity balance, charge balance was not always exact for the superconductor candidates. Furthermore, crystal structure plays a significant role in superconductivity (at least at low temperatures), so accounting for that may result in improvements. however, this may prove to be a difficult endeavor as crystal structure is not well-documented. Active transfer learning could also be attempted to narrow down the predictions to high temperature superconductors only. Even without these possible improvements though, the model at its current state is still quite promising and can be applied to search for better materials in the rapidly growing field of quantum computers or search for new superconductors that may better explain high temperature superconductivity.

REFERENCES

(n.d.). http://www.crystallography.net/

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. https://arxiv.org/abs/1701.07875

Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., Biswas, R., Boixo, S., Brandao, F. G. S. L., Buell, D. A., Burkett, B., Chen, Y., Chen, Z., Chiaro, B., Collins, R., Courtney, W., Dunsworth, A., Farhi, E., Foxen, B., & Martinis, J. M. (2019). Quantum supremacy using a programmable superconducting processor. Nature, 574 (7779), 505–510. https://doi.org/10.1038/s41586-019-1666-5

Bardeen, J., Cooper, L. N., & Schrieffer, J. R. (1957). Microscopic theory of superconductivity. Phys. Rev., 106, 162–164. https://doi.org/10.1103/PhysRev.106.162
Bednorz, J. G., & Mu¨ller, K. A. (1986). Possible hightc superconductivity in the ba–la–cu–o system. Zeitschrift fu¨r Physik B Condensed Matter, 64 (2), 189–193. https://doi.org/10.1007/ BF01303701
Dan, Y., Zhao, Y., Li, X., Li, S., Hu, M., & Hu, J. (2020). Generative adversarial networks (gan) based efficient sampling of chemical composition space for inverse design of inorganic mate- rials. npj Computational Materials, 6 (1), 84. https://doi.org/10.1038/s41524-020-00352-0
Davies, D. W., Butler, K. T., Jackson, A. J., Morris, A., Frost, J. M., Skelton, J. M., & Walsh, A. (2016). Computational screening of all stoichiometric inorganic materials. Chem, 1 (4), 617–627. https://doi.org/10.1016/j.chempr.2016.09.010
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. CoRR, abs/1704.00028. http://arxiv.org/abs/1704.00028
Hinton, G. (2012). Neural networks for machine learning: Lecture 6a overview of min-batch gradient descent. http://www.cs.toronto.edu/□tijmen/csc321/slides/lecture slides lec6.pdf
Hirsch, J., Maple, M., & Marsiglio, F. (2015). Superconducting materials classes: Introduction and overview [Superconducting Materials: Conventional, Unconventional and Undetermined]. Physica C: Superconductivity and its Applications, 514, 1–8. https://doi.org/https://doi. org/10.1016/j.physc.2015.03.002
Hosono, H., Tanabe, K., Takayama-Muromachi, E., Kageyama, H., Yamanaka, S., Kumakura, H., Nohara, M., Hiramatsu, H., & Fujitsu, S. (2015). Exploration of new superconductors and functional materials, and fabrication of superconducting tapes and wires of iron pnictides. Science and Technology of Advanced Materials, 16 (3), 033503. https://doi.org/10.1088/ 1468-6996/16/3/033503
Jha, D., Ward, L., Paul, A., Liao, W.-k., Choudhary, A., Wolverton, C., & Agrawal, A. (2018). El- emnet: Deep learning the chemistry of materials from only elemental composition. Scientific Reports, 8 (1), 17593. https://doi.org/10.1038/s41598-018-35934-y
Kamerlingh Onnes, H. (1911). The superconductivity of mercury. Comm. Phys. Lab. Univ. Leiden, 122, 122–124. https://ci.nii.ac.jp/naid/10027961803/en/
Kamihara, Y., Watanabe, T., Hirano, M., & Hosono, H. (2008). Iron-based layered superconductor la[o1-xfx]feas (x = 0.050.12) with tc = 26 k. Journal of the American Chemical Society, 130 (11), 3296–3297. https://doi.org/10.1021/ja800073m
Kirklin, S., Saal, J. E., Meredig, B., Thompson, A., Doak, J. W., Aykol, M., Ru¨hl, S., & Wolverton, C. (2015). The open quantum materials database (oqmd): Assessing the accuracy of dft formation energies. npj Computational Materials, 1 (1), 15010. https://doi.org/10.1038/ npjcompumats.2015.10
Konno, T., Kurokawa, H., Nabeshima, F., Sakishita, Y., Ogawa, R., Hosako, I., & Maeda, A. (2021). Deep learning model for finding new superconductors. Phys. Rev. B, 103, 014509. https://doi.org/10.1103/PhysRevB.103.014509
National Institute for Materials Science. (n.d.). https://supercon.nims.go.jp/en/
Roter, B., & Dordevic, S. (2020). Predicting new superconductors and their critical temperatures using machine learning. Physica C: Superconductivity and its Applications, 575, 1353689. https://doi.org/doi.org/10.1016/j.physc.2020.1353689
Saal, J. E., Kirklin, S., Aykol, M., Meredig, B., & Wolverton, C. (2013). Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). JOM, 65 (11), 1501–1509. https://doi.org/10.1007/s11837-013-0755-4
Stanev, V., Oses, C., Kusne, A. G., Rodriguez, E., Paglione, J., Curtarolo, S., & Takeuchi, I. (2018). Machine learning modeling of superconducting critical temperature. npj Computational Materials, 4 (1), 29. https://doi.org/10.1038/s41524-018-0085-8
van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html
Wang, Y., Wu, C., Herranz, L., van de Weijer, J., Gonzalez-Garcia, A., & Raducanu, B. (2018). Transferring gans: Generating images from limited data.

Machine Learning Guided Selection of Bio inks for 3D Bioprinting of Tissue Scaffolds

Rithani Saravanakumar

Nikola Tesla STEM High School

Introduction

The purpose of this study was to use a machine learning approach to train existing data sets of different combinations of bio ink parameters to accelerate the process of synthetic tissue classification and biocompatibility identification, ensuring that the time, resources, and safety are not compromised.

Tissues are an essential part of the human body; they perform an innumerable number of functions, from providing structural support and protection to intracellular communication and carrying out metabolic reactions. In the case of an accident, where the internal organ tissue or underlying skin tissue is damaged, it becomes imperative that the lost tissues are replaced, either with naturally produced tissue lines, or synthetically printed tissues. However, finding readily available tissue donors is a challenging feat. According to NDLM Donation and Transplantation Statistics, annually, "there are approximately 30,000 tissue donors" signed up but "more than 1.75 million tissue transplants" taking place, which means the need for more donated tissues is increasing. An alternative approach is to synthetically print tissues that function in the same way natural body tissues do. These printed tissues, similar to 3D printed objects, are reproduced using a special type of ink called a "bio ink". Bio inks are primarily composed of live cells, polymers, and cellular hydrogels, which all work together to ensure that the newly printed tissue functions properly in the patient's body. However, depending on the region that requires tissue replacement, this composition may vary.

3D printing synthetic tissues is not an easy process. According to the Wyss Institute, it can take up to 6-8 weeks to fully print one functional tissue scaffold that can be used to replace damaged tissue in the body. Taking into consideration the patient's medical history, organ/tissue shape and size, and tissue functionality, it is crucial for doctors to be able to identify the bio inks that hold the most potential in treating the patient. One way to accomplish this is by manually printing tissues from each bio ink and then testing its efficacy by integrating it into an artificial synthetic cadaver. However, printing tissue after tissue to find the one "perfect" bio ink is a time and resource demanding trial and error process. Furthermore, multiple iterations of printing using the same bio ink are required in order to ensure accurate results. Even aside from the printing labor, these scaffolds take a long to be implemented in the specific body region so that they can communicate and interact with the existing tissues. In order for the body to accept and integrate an unknown cellular structure, its biocompatibility with the original tissues must be verified so that the body doesn't reject the tissue because of its foreign nature (Ma et al., 2020).

Testing the biocompatibility of the tissue with the body tissues requires an extensive amount of time and materials, as different combinations of these printing patterns must be tested. In an attempt to do this, scientists are currently experimenting with all of the parameter combinations, tweaking the type of bio ink used, printing settings, and printer used. This is a huge waste of money, as it can take up to $5,000 to print just one tissue (Kirillova et al., 2020). Constant use and unreliable discarding of such materials also poses a hazardous risk to the environment.

Currently, scientists are trying to optimize the time it takes to print these scaffolds by finding patterns in printing parameters (time, speed, spacing, etc.) and identifying the combination of parameter values that will yield the highest quality tissue scaffold in the lowest amount of time. Recent studies show that machine learning (ML) and artificial intelligence (AI) have been used to

improve the 3D bioprinting of materials. Many researchers have developed and tested specific machine learning algorithms, particularly Random Forests, with their training data to accurately predict the quality of the tissue scaffolds. One research article in particular uses two approaches – classification-based and regression-based – both of which are built upon Random Forest classification models (Conev et al., 2020), to predict the quality of the tissue by using the printing error percentage. Other studies have also used ML to determine material printability. Inductive logic programming methodology, an AI-based subfield of machine learning that uses first-order logic, was used to predict the quality of differing polymer mixtures. Establishing a relationship between bio ink properties and scaffold consistency allowed the algorithm to identify promising bio ink combinations (Gu et al., 2018).

However, these investigations did not take into consideration the actual biological composition of the ink, in addition to important physical characteristics such as viscosity. This project seeks to investigate the classification of tissue scaffolds printed from different bio ink factor combinations (PPF composition, printing pressure, printing spacing, and printing speed) using four different machine learning classification algorithms – Logistic Regression, Random Forests, Support Vector Model, and Extreme Gradient Boosting – to design a more efficient tissue classifying model. As predicted before, the biological composition and viscosity features should play an integral role in determining the tissue's quality.

Methodology

The dataset used in this investigation was obtained from a study (Trachtenberg, J. E et. Al., 2016) where they investigated the impact of different bio ink processing variables on scaffold printing consistency. Their variables included material-related (PPF composition in the printing solution), printing-related (printing pressure and speed), and design-related (programmed fiber spacing) factors. The other study (Conev et al., 2020) used these variables and calculated the percentage error for machine precision (Eq. 1) and material accuracy (Eq. 2) for each printing condition. These two metrics reflected the variations from the model's programmed values. As a result, larger MP and MA values indicated larger errors or a "lower" printing quality, and smaller MP and MA values indicated smaller errors, or a "higher" printing quality.

The first step was to use the material precision and material accuracy formulas and the parameters from the data table to calculate the material precision and material accuracy values for each row in the training data. This gave an approximation of the error for each combination of bio ink parameters. Then, a target label called "quality" was created and added to the training data to classify each row of data as "high" or "low" quality tissues. In one research article that employed a very similar methodology, they used threshold values of 50% for material accuracy and 6% for machine precision to classify each row into their respective category. However, for the purposes of this study, the threshold values were 1% higher to

Equation.1

$$Material\ precision \left| \frac{Experimental\ fiber\ spacing - Programmed\ fiber\ spacing}{Programmed\ fiber\ spacing} \right| * 100$$

Equation.2

$$Material\ accuracy = \left| \frac{Mean\ fiber\ diameter - Model's\ fiber\ diameter}{Model's\ fiber\ diamter} \right| * 100$$

ensure more accurate results. Therefore, material accuracy percentages that were less than or equal to 51% were labelled "high" quality and percentages greater than 51% were labelled as "low quality." Similarly, machine precision percentages less than or equal to 7% were considered "high" quality, and those that are greater than 7% were labelled as "low" quality. The obtained values were then recorded into a data table that looks similar to this:

Table 1.1 Classifier label for Material Accuracy Classification Metric

Classification Metric	Wt% PPF	Pressure (bar)	Speed (mm/s)	Spacing (mm)	Tissue Classification (high or low)
Material Accuracy	85	2.0	5.0	1.2	High
Material Accuracy	85	2.0	7.5	1.2	High
Material Accuracy	85	2.0	10.0	1.2	High
Material Accuracy	85	2.0	15.0	1.2	High
Material Accuracy	85	2.5	5.0	1.2	High

Table 1. 2 Classifier label for Machine Precision Classification Metric

Classification Metric	Wt% PPF	Pressure (bar)	Speed (mm/s)	Spacing (mm)	Tissue Classification (high or low)
Machine Precision	85	2.0	5.0	1.2	High
Machine Precision	85	2.0	7.5	1.2	High
Machine Precision	85	2.0	10.0	1.2	High
Machine Precision	85	2.0	15.0	1.2	High
Machine Precision	85	2.5	5.0	1.2	High

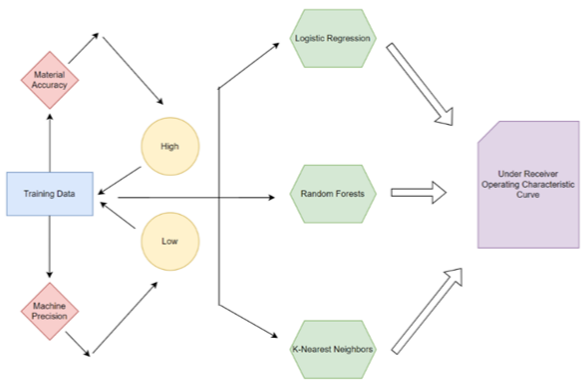
The final columns in both of the tables were added to the training data as a classifier label, so that the result of the machine learning models was able to output a classification.

This project used four different machine learning based models to accurately predict the quality of tissues printed using different printing configurations. These configurations obtained values from a combination of five different parameters: material composition (PPF), printing speed (mm/s), printing pressure (bar), viscosity (Ns/m2), and programmed fiber spacing (mm). These parameters were the input features for all four of the models. Because the objective was to classify the data using two labels, the models were built upon a classification-based approach, where each classification model classified each input printing configuration as "low" or "high" quality. As mentioned previously, the data was labelled with binary values of "low" and "high" using the given

threshold values. Then, each classifier algorithm was run twice, one for each metric (material accuracy and material precision), utilized for accurate prediction.

The Random Forest, Support Vector Model, Extreme Gradient Boosting, and Logistic Regression models were imported from the Scikit-learn library from Python. Additional data preprocessing steps were taken place to resolve any data quality issues to make sure that the data was suitable for model training. Two of the most important were balancing the skewness and clustering of the data and then replacing all non-numerical values with its most commonly occurring value (mode).

Figure 1. Visual Representation of Procedure A and Procedure B.



Caption: Training data is first classified as high- or low-quality tissue with material accuracy and machine precision. Labelled data is fed into four different ML models, which then graph a relationship to evaluate accuracy

For each of the machine learning models, a function called Grid Search was employed to identify the optimal hyperparameters. The purpose of Grid Search is to speed up the long parameter tweaking trial and error process, where instead of inputting arbitrary values for each model to find the combination that yields the highest accuracy, the function automatically returns that sequence without having to manually change the parameters.

After identifying the optimal hyperparameters for each of the machine learning models, each fold of the training data was trained at a time using a Leave One Out Cross Validation technique, where the number of folds is equal to the size of the data set. With approximately 2100 rows of bio-ink feature data, the data set was passed through the model with the selected instance being the one test case and all other instances being the training data. This performance was heavily preferred because it offers a less biased result as opposed to the standard 80% - 20% split, where the data is only compared once. In the LOOCV technique, the data was repeatedly fit to the data set N-1 times, and this also helps to prevent overfitting of the model

Once the hyperparameters had been identified for the Logistic Regression model, they were recorded with its accompanying metrics. The parameters included the learning rate, penalty L1, and penalty L2, and the final metrics of the model were accuracy, precision, recall, and F1 score. For every test case that was being compared with the training data, its respective parameter values and accompanying accuracy results were recorded in Table 2.1.

Table 2.1 Logistic Regression model output measurements

Learning Rate	Penalty L1	Penalty L2	Accuracy

Figure 2.1 Example of Logistic Regression Classification model

Caption: Classification of high- and low-quality bio printed tissue data points and using the determined curve to classify new data points

Once the hyperparameters had been identified for the Random Forests model, they were recorded with its accompanying metrics. The parameters included the depth of the generated trees, number of trees created, and number of variables for decision, and the final metrics of the model were again accuracy, precision, recall, and F1 score. For every test case that was being compared with the training data, its respective parameter values and accompanying accuracy results were recorded in Table 2.2. Because Extreme Gradient Boosting is a slightly modified version of random forests, the similar parameters were used during its model evaluation.

Table 2.2 Random Forests model output measurements

Depth of trees	Number of trees	Number of variables for decision	Accuracy

Figure 2.2: Example of Random Forest model and parameter labelling

Caption: Fine tuning depth of trees and number of trees to achieve highest accuracy classification for Random Forest model

Once the hyperparameters had been identified for the Support Vector model, they were recorded with its accompanying metrics. The parameters included the tradeoff parameter and the insensitivity parameter, and the final metrics of the model were accuracy, precision, recall, and F1 score. For every test case that was being compared with the training data, its respective parameter values and accompanying accuracy results were recorded in Table 2.3.

Table 2.3 Support Vector model output measurements

Tradeoff	Insensitivity	Accuracy

Table 4. Highest output measurement values of each machine learning model

ML Algorithm	Accuracy	Precision	F1 Score	Recall
Logistic Regression				
Random Forests				
Support Vector Model				

After the model output for all four machine learning algorithms had been measured, the results were analyzed using an Under Receiver Operating Characteristic Curve (ROC Curve) and a AUC Score. The ROC Curve is a graphical plot that illustrates the relationship between the true positive rate (Eq.3) and the false positive rate (Eq.4). These values were calculated from the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) of the model's performance. The curve used them to test how well the relationship between the bio ink parameters and tissue quality was, and to select optimal models. A flat graph will indicate a high accuracy, while a non-flat graph signifies that there are deviations. The resulting area underneath this curve is called the AUC score, which represents the how well the model is capable of distinguishing the two classifications.

Equation.3

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR$$

Equation.4

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - TNR$$

Results and Data Analysis

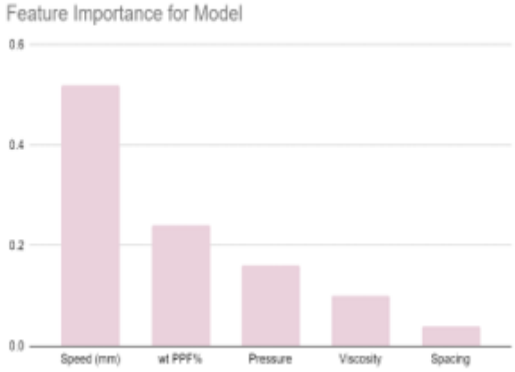
After training the four models and obtaining the final metric values, Random Forests overall had the best performance with the highest values in accuracy (85.9%), precision (0.86), F1 score (0.88), and recall (0.90).

Table 5. Highest output measurement values of each machine learning model

ML Algorithm	Accuracy	Precision	F1 Score	Recall
Logistic Regression	63%	0.62	0.71	0.82
Random Forests	**85.9%**	**0.86**	**0.88**	**0.90**
Support Vector Model	59%	0.72	0.78	0.85
Extreme Gradient Boosting	59%	0.60	0.68	0.79

Overall, comparing the current results to that of the previous tissue classifying model, the inclusion of the biological composition and the viscosity features had an 11% increase in accuracy from 75% to approximately 86%. This significantly impacted the tissue classification accuracy and was able to generate a more efficient selection model, as predicted earlier. Additionally, after plotting the AUROC curve, the resulting AUC score for Random Forests was 0.86, which meant that the model was good at distinguishing between the "low" quality and "high" quality tissue prints. To further analyze the importance of each feature in the model, the model was also coded to rank the five features in order of their importance. The feature that is the most important, or contributes most to the outcome of the model, would be ranked first and have the highest importance value. Likewise, the least important feature, the column that doesn't significantly impact the results, would be ranked last and have the smallest values. The results were statistically significant.

Figure 2.3 Feature Importance for Random Forests model



Caption: Biological Composition was the second most important feature while viscosity was the fourth most important feature, supporting the hypothesis that bio ink properties play a key role in a tissue's structure

In the future, more than five input features should be included in the model to ensure a much higher accuracy and representative data set. Possible experimentation with more biological and rheological parameters such as immune response and shear stress could lead to prediction rates as high as 95%. Furthermore, using an emerging concept called auto machine learning that takes into consideration the feature importance would be a much better alternative to the standard Leave One Out Cross Validation technique because the results would mirror a human's results.

Bibliography

1. Abueidda, D. W., Almasri, M., Ammourah, R., Ravaioli, U., Jasiuk, I. M., & Sobh, N. A. (2019). Prediction and optimization of mechanical properties of composites using convolutional neural networks. Composite Structures, 227, 111264. https://doi.org/10.1016/j.compstruct.2019.111264

2. Algahtani, M. S., Mohammed, A. A., & Ahmad, J. (2019). Extrusion-Based 3D Printing for Pharmaceuticals: Contemporary Research and Applications. Current Pharmaceutical Design, 24(42), 4991–5008. https://doi.org/10.2174/1381612825666190110155931

3. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324

4. Conev, A., Litsa, E. E., Perez, M. R., Diba, M., Mikos, A. G., & Kavraki, L. E. (2020b). Machine Learning-Guided Three-Dimensional Printing of Tissue Engineering Scaffolds. Tissue Engineering Part A, 26(23–24), 1359–1368. https://doi.org/10.1089/ten.tea.2020.0191

5. Grierson, D., Rennie, A. E. W., & Quayle, S. D. (2021). Machine Learning for Additive Manufacturing. Encyclopedia, 1(3), 576–588. https://doi.org/10.3390/encyclopedia1030048

6. Groll, J., Burdick, J. A., Cho, D. W., Derby, B., Gelinsky, M., Heilshorn, S. C., Jüngst, T., Malda, J., Mironov, V. A., Nakayama, K., Ovsianikov, A., Sun, W., Takeuchi, S., Yoo, J. J., & Woodfield, T. B. F. (2018). A definition of bioinks and their distinction from biomaterial inks. Biofabrication, 11(1), 013001. https://doi.org/10.1088/1758-5090/aaec52

7. Gupta, S. (2020, March 1). How to tackle any classification problem end to end & choose the right classification ML algorithm. Medium. https://towardsdatascience.com/how-to-tackle-any-classification-problem-end-to-end-choose-the-right-classification-ml-algorithm-4d0becc6a295

8. Kirillova, A., Bushev, S., Abubakirov, A., & Sukikh, G. (2020). Bioethical and Legal Issues in 3D Bioprinting. International Journal of Bioprinting, 6(3). https://doi.org/10.18063/ijb.v6i3.272

9. Mao, H., Yang, L., Zhu, H., Wu, L., Ji, P., Yang, J., & Gu, Z. (2020). Recent advances and challenges in materials for 3D bioprinting. Progress in Natural Science: Materials International, 30(5), 618–634. https://doi.org/10.1016/j.pnsc.2020.09.015

10. Schwab, A., Levato, R., D'Este, M., Piluso, S., Eglin, D., & Malda, J. (2020b). Printability and Shape Fidelity of Bioinks in 3D Bioprinting. Chemical Reviews, 120(19), 11028–11055. https://doi.org/10.1021/acs.chemrev.0c00084

11. Sperandei, S. (2014). Understanding logistic regression analysis. Biochemia Medica, 12–18. https://doi.org/10.11613/bm.2014.003

12. Tian, S., Stevens, R., McInnes, B., & Lewinski, N. (2021). Machine Assisted Experimentation of Extrusion-Based Bioprinting Systems. Micromachines, 12(7), 780. https://doi.org/10.3390/mi12070780

13. Trachtenberg, J. E., Placone, J. K., Smith, B. T., Piard, C. M., Santoro, M., Scott, D. W., Fisher, J. P., & Mikos, A. G. (2016). Extrusion-Based 3D Printing of Poly (propylene fumarate) in a Full-Factorial Design. ACS Biomaterials Science & Engineering, 2(10), 1771–1780. https://doi.org/10.1021/acsbiomaterials.6b00026

Analysis of Ring Galaxies Detected Using Deep Learning with Real and Simulated Data

Harish Krishnakumar

Nikola Tesla STEM High School

ABSTRACT

Understanding the formation and evolution of ring galaxies, galaxies with an atypical ring-like structure, will dramatically improve understanding of black holes and galaxy dynamics as a whole. Current catalogs of rings are extremely limited: manual analysis takes months to accumulate an apprecia- ble sample of rings and existing computational methods are vastly limited in terms of accuracy and detection rate. Without a sizable sample of rings, further research into their properties is severely restricted. This project investigates the usage of a convolutional neural network (CNN) to identify rings from unclassified samples of galaxies. A CNN was trained on a sample of 100,000 simulated galaxies, transfer learned to a sample of real galaxies and applied to a previously unclassified dataset to generate a catalog of rings which was then manually verified. The properties of these galaxies were then estimated from their photometries. A catalog of 580 ring galaxies, the current largest set of computationally identified rings, was extracted with 6.6 times the accuracy, 7.8 times the detection rate, and 53.7 times the time efficiency of conventional computational methods. The null hypothesis that the specific star formation rates of the ringed and non-ringed identifications were from the same distribution was rejected with a p-value of .015, agreeing with prior studies. With upcoming surveys such as the Vera Rubin Observatory Legacy Survey of Space and Time obtaining images of billions of galaxies, this model could be crucial in classifying large populations of rings to better understand the peculiar mechanisms by which they form and evolve.

Keywords: Ring galaxies (1400), Galaxies (573), Neural networks (1933), Convolutional neural net- works (1938), Galaxy evolution (594), Astronomy data analysis (1858), Spectral energy distribution (2129)

1. INTRODUCTION

Ring galaxies, galaxies with a luminous core sur- rounded by a disk of matter, are thought to have the potential to lead to a greater understanding of galaxy evolution as a whole. However, much still remains un- known about these galaxies. For instance, rings such as Hoag's Object (Hoag 1950) have not shown conclu- sive evidence as to how they have formed (Brosch 1985; Schweizer et al. 1987; Finkelman et al. 2011). Addition- ally, although we can determine that certain ring galax- ies form via galaxy-galaxy collisions (Appleton 1999; Appleton & Struck-Marcell 1996), one of these galaxies, known as the Cartwheel Galaxy, hosts the most ultra- luminous X-Ray sources ever observed in a single galaxy, one of which is possibly an intermediate-mass black hole (Trinchieri 2010). Intermediate-mass black holes are too massive to form through the collapse of a single star, and we have not yet determined how they form (Rose et al. 2021), making studying peculiar galaxies such as the Cartwheel Galaxy crucial.

For these reasons, ring galaxies are of particular inter- est to study. However, only a limited number of rings have been identified. The Catalog of Southern Ring Galaxies (Buta 1995) identified 3692 galaxies south of declination -17°. Most prominent sky surveys, includ- ing the Sloan Digital Sky Survey (SDSS. York et al. 2000) and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS. Chambers et al. 2019), only cover the sky north of around -30°, making this data inconvenient for further analysis (Buta 2017).

The Galaxy Zoo 2 project (Willett et al.

2013) allevi- ated this problem through the usage of data from DR7 of the SDSS (Abazajian et al. 2009). Morphological classifications for 304,122 galaxies were crowd-sourced from more than 50,000 contributors, and compiled into a database. 3962 of these galaxies were identified to have rings (Buta 2017), though many were poorly resolved, and some were misclassified as a re- sult. Less than 1% of the data, in addition, had tendencies of a collisional ring galaxy (Appleton & Struck-Marcell 1996), while most of the data involved a spiral galaxy as a primary com- ponent. Thus, it is clear that more data needs to be obtained to properly understand the formation and evo- lution of ring galaxies.

Over 14 months, however, were required to collect and compile the data inputted by volun- teers for the 304,122 images of the Galaxy Zoo 2 project. With millions of images still left to be analyzed and upcoming digital sky surveys such as the Vera Rubin Legacy Survey of Space and Time (Ivezi´c et al. 2019) collecting images of billions of more galaxies, manual classification of mor- phologies clearly does not suffice for the dramati- cally increasing volume of data.

This problem inspired previous attempts to classify ring galaxies with computational analy- sis (e.g. Shamir 2019; Timmis & Shamir 2017). These methods, how- ever, while simple to utilize, result in a low detection rate and a high misclas- sification rate. The flood fill al- gorithm was used, which looks for a path to the edge of the image from the center pixel, and if so, classifies a galaxy as non-ringed. As a result, objects which are not centered can result in potential problems for the model, resulting in misclassifications of potential ring candidates as non-ringed. Additionally, noise and arti- facts are often misclassified as rings, as these algorithms cannot adapt to edge cases like these. Therefore, these methods do not have a large impact on existing catalogs of ring galaxies.

The analysis of large samples of galaxies with machine learning, on the contrary, can lead to a greater overall classification accuracy (Shamir 2011), with less true pos- itives incorrectly identi- fied, thus leading to the creation of more compre- hensive catalogs of ring galaxies. Partic- ularly, deep learning, and the continuous advancement of convolutional neural networks (CNNs. LeCun et al. 2015) in specific, has propelled efficient and accurate image classification.

Recently, CNNs have increasingly been used to clas- sify galaxies based on their proper- ties. For instance, Wu & Peek 2020 used a CNN to predict spectroscopic data from images of galaxies and Kim & Brunner 2016 used a CNN to distin- guish between stars and galaxies. Specifi- cally, the morphology of galaxies can also be accurately classified by these networks. One of the first applica- tions of this was Huertas-Company et al. 2015, where a CNN was trained to classify the morphol- ogy of galaxies in the CANDELS fields with less than a 1% error rate. Lanusse et al. 2017 described the creation of a CNN to identify gravitational lenses, another unique morphol- ogy, trained on a sample of 20,000 simulated galaxies. Building on this concept, Ghosh et al. 2020 trained a CNN to classify galaxies morphologically on a sample of 100,000 simulated galaxies and used a technique known as transfer learning to accurately retrain this model on a small sample of real galaxies, achieving less than a 5% error rate, thus reducing the need for a larger sample size to train an accu- rate model.

Despite these rapid advancements in auto- mated mor- phological classification, no attempts have been made to detect ring galaxies with CNNs. If we want to further understand the formation and evolution of ring galax- ies, it is clear that obtain- ing an expanded catalog of them is essential. As a CNN can achieve both a high false positive rate while minimizing the misclassification of true positives, they show potential to compile large catalogs of accurately classified ring galaxies. Classify- ing galaxies with CNNs is additionally significantly more time efficient than manual anal- ysis (Ghosh et al. 2020), thus showing promise for the analysis of large datasets in hours, compared to the years it would take with manual classification.

In this paper, I present a CNN with the Inception- ResNet-V2 architecture (Szegedy et al. 2016) to classify ring galaxies 1. This model was first trained on a sample of 100,000 simulated ringed and non-ringed galaxies and then transfer learned onto a sample of galaxies classified by Galaxy Zoo 2. Then, the model was applied on a

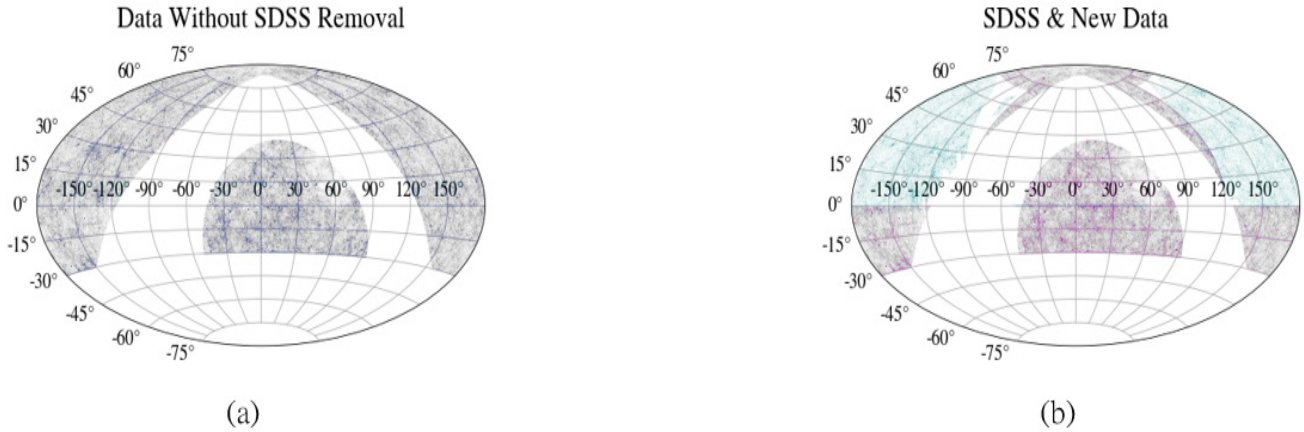


Figure 1. Sky coverage of the application data. Left (a): The sky coverage before the subtraction of the sky covered by the Galaxy Zoo 2 project. Right (b): The blue represents the sky coverage of the Galaxy Zoo 2 project while the magenta represents the sky coverage of the modified catalog after removal.

sample of □ 960, 000 galaxies obtained from the cata- log compiled in Goddard & Shamir 2020a. The galaxies identified as rings were manually verified, and compiled into a new catalog. I then examined the specific star formation rate (SSFR), redshift, color and color-mass diagrams of these galaxies, and compared it to a sample of random non-ringed galaxies to mwake more concrete conclusions about the distinct formation and evolution of ring galaxies. As spectroscopic data is not yet ob- tainable for many galaxies, a technique known as SED fitting (Walcher et al. 2010) was used to derive the prop- erties of these galaxies from their photometric data. I expected the expanded sample of ring galaxies to have a lower SSFR and to be redder than their non-ringed counterparts, as measured in Fernandez et al. 2021 with a different sample of galaxies.

I describe the methods used to obtain the data the model is trained, tested and applied on in § 2. In § 3, I describe the simulation of the initial training data, the selection of a model architecture, the training of the model, and the application of transfer learning. In § 4, I describe the usage of SED fitting to infer properties of the newly obtained catalog. In § 5, I present the expanded catalog of ring galaxies and their properties. Finally, in § 6 and § 7 I discuss the results and their potential applications.

[1]The code used can be found at https://github.com/harishk30/ RingGalaxiesCNNAnalysis/

2. OBTAINING DATA

This paper primarily discusses the usage of a CNN to detect ring galaxies in real data sets. A catalog of ring galaxies was first obtained from the Galaxy Zoo 2 project (Willett et al. 2013) to train the model. 3962 galaxies were classified as ”clean” samples of rings. As Galaxy Zoo 2 participants were shown images from the Dark Energy Camera Legacy Survey (DECaLS), images to train the model were obtained from the DESI Legacy Imaging Surveys (Dey et al. 2019) as it includes DECaLS as a primary component.

Cutouts of the training images were 256 pixels on a side. As the galaxies were arranged in descending order by angular diameter, the resolution of the images was iterated from 0.65” to 0.09” per pixel to ensure that the galaxies occupied equal areas in the frame.

The training sample of galaxies was then manually verified. Only about 3117, or 78.6%, of the initial sam- ple of ring galaxies was found to visually exhibit char- acteristics of a ring. As described in § 3, 80% of this data was reserved for training while 20% was reserved for testing.

The model was applied to detect rings in a catalog of galaxies detected in the Pan-STARRS survey (God- dard & Shamir 2020b). This catalog initially contained 1, 685, 349 galaxies. However, the sky coverage of this catalog overlapped with that of Galaxy Zoo 2, which was used to train the model. Thus, as shown in Fig- ure 1, the sky

covered by Galaxy Zoo 2 was subtracted from that of the Pan-STARRS catalog to obtain a final catalog of 966, 329 galaxies to apply the model to. Images for these galaxies were downloaded from the DESI Legacy Imaging Surveys as previously described for the training data.

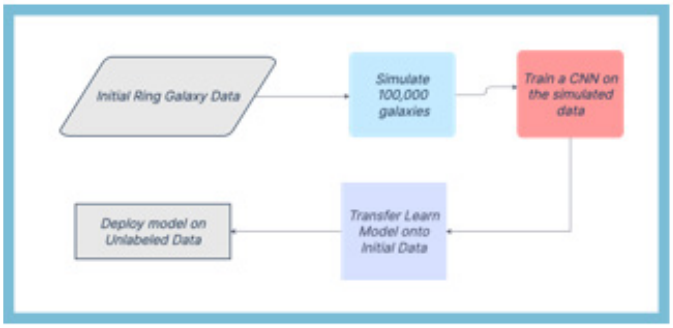


Figure 2. Workflow starting from the initial data to the final CNN. Obtaining the initial data is described in § 2. The simulation of 100,000 galaxies is described in § 3.1. The CNN trained on the simulated data is described in § 3.2. Finally, the process of transfer learning with the CNN is described in § 3.3.

3. CREATING & TRAINING THE CONVOLUTIONAL NEURAL NETWORK

To train an accurate CNN, large data sets are typically required. However, with existing catalogs of ring galaxies being limited to only a few thousand galaxies, a different approach is required to obtain the most ac- curate model. A technique known as transfer learning (Zhuang et al. 2020) is used to first train the model on a sample of simulated galaxies, and the information learnt is used to retrain the model on the smaller sample of real galaxies. The steps used to train the model are shown in in Figure 3.

3.1. Simulating Galaxies

To simulate the initial training sample of ring galaxies, the GALFIT (Peng et al. 2002) program was used on an Ubuntu virtual machine. GALFIT is typically used to fit light profiles to astronomical objects, but in this context, as in Ghosh et al. 2020, it is used to create light profiles for randomized galaxies to be used as training data.

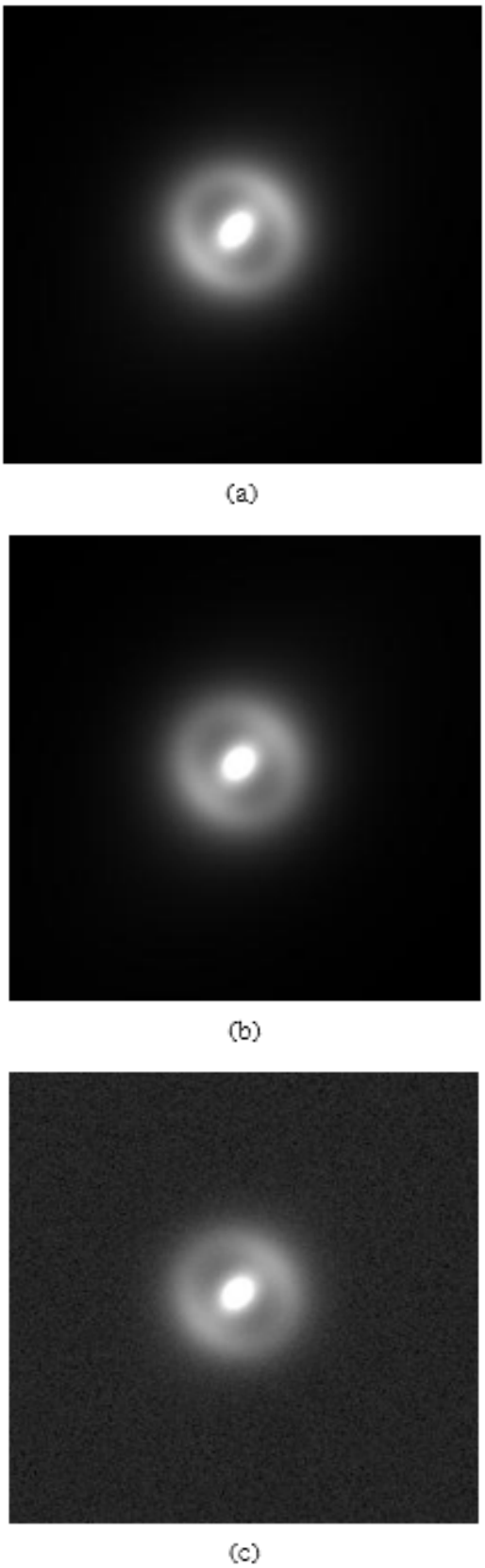


Figure 3. A simulated ring galaxy and modifications to make it further resemble real data. Left (a): The simulated light profile initially created by GALFIT. Middle (b): The simulated light profile after convolution with a point spread function (PSF). Right (c): The simulated light profile after convolution with a PSF and the addition of randomized noise.

The training data set contained 100,000 simulated galaxies, with an even split between ringed and non- ringed galaxies. The galaxies were simulated using a S´ersic profile (Peng et al. 2010), which describes how the intensity of a galaxy varies with distance from its center. The functional form of the S´ersic profile is described by

$$\Sigma(r) = \Sigma_e \exp\left[-\kappa\left(\left(\frac{r}{r_e}\right)^{1/n} - 1\right)\right]. \qquad (1)$$

Σe represents the pixel surface brightness at the radius re, which represents the radius at which half of the galaxy's flux is contained. The S´ersic index of the galaxy, n, controls where the light of the galaxy is con- centrated, and κ is dependent on this parameter to en- sure that half of the flux stays within radius re from the center.

However, a ringed structure cannot accurately be rep- resented by one standard S´ersic function. As ring galax- ies have a pronounced gap in their center, the initial S´ersic function needs to be multiplied by a hyperbolic tangent truncation function. This is given by:

$$0.5\left(\tanh\left[(2-B)\frac{r}{r_{break}} + B\right] + 1\right) \qquad (2)$$

where

$$B = 2.65 - 4.98\left(\frac{r_{break}}{r_{break} - r_{soft}}\right). \qquad (3)$$

In the function, rbreak represents the radius at which 99% of the galaxy's flux is enclosed while rsoft represents the radius at which 1% of the galaxy's flux is enclosed. However, the inner core of the galaxy still needed to be represented, thus requiring a second S´ersic profile without a truncation function applied. When simulating the sample of non-ringed galaxies, only a single S´ersic profile was needed.

To determine the general parameters to simulate a ring galaxy, a light profile was first fit to a galaxy from the sorted Galaxy Zoo 2 catalog. This galaxy was ran- domly selected, and was located at R.A.: 126.77484, Decl.: 21.64528. The sky background and center for this galaxy were determined using SourceExtractor (Bertin & Arnouts 1996), and the other parameters were estimated based on visual inspection of the image.

The parameters of the fit light profile were then used as a baseline to create a randomized distribution of pa- rameters for simulating ring galaxies. Additionally, pa- rameters for simulating non-ringed galaxies were also randomly generated. These parameters are described in Table 1. Finally, the position angle of the galaxies was iterated between 90◦ and 90◦.

The GalaxySim2 code (Ghosh et al. 2020) was used with modifications customized to the ringed galaxies pa- rameters to create the input files for GALFIT, run GALFIT on the input files, and modify the generated images to be more realistic.

After creating the input files, GALFIT was run to create simulated light profiles. The images were then convolved with a point spread function (PSF), representing the dis- persion of light caused by atmospheric disturbances and optical limitations, to be more realistic. This PSF was obtained from the galaxy used as a prototype for the simulated rings. A randomly generated array of noise pixels was then added to the simulated image to be rep- resentative of the noise that would be present in a real image. An example of these modifications is shown in Figure 3.

Following this, a Python script was created to first scale the images, keeping 99.5% of the initial pixels, and then convert them from FITS to JPG files to train the model on, as this was the file format the cutouts of the real galaxies were in.

3.2. Training the Network

A CNN is a type of artificial neural network (ANN) typically applied to analyze images. The simplest type of ANN is a multilayer perceptron (MLP), shown in Fig- ure 4. It consists of three layers - the input layer, the hidden layers, and the output layer. These layers are consisted of structures called neurons which are inter- connected with those of the adjacent layers. These neu- rons each hold a value called their activation. The activations of the input neurons determine the activation of the subsequent layer, and this occurs until an activation for the output neurons are determined.

[2]https://github.com/aritraghsh09/GalaxySim

Table 1. Parameters to Simulate Ringed and Non-Ringed Galaxies

Component Name	Sērsic Index	Half-Light Radius (Pixels)	Axis Ratio	Integrated Magnitude (AB)	Break Radius (Pixels)	Softening Radius (Pixels)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Ring Galaxy			
Outer Ring	1.7	2.0 - 5.0	0.7 - 1.0	15.0 - 25.0	N/A	N/A
Truncation Function	N/A	N/A	0.7 - 0.9	N/A	40.0 - 60.0	24.0 - 34.0
Inner Core	3.0	8.0 - 12.0	0.5 - 0.75	-6.2 - -3.2	N/A	N/A
			Non-Ringed Galaxy			
Sērsic Profile	0.0 - 8.0	55.0 - 88.0	0.1 - 1.0	17.0 - 25.0	N/A	N/A

To determine the activations of the neurons in the subsequent layer, each connection between the previous layer and the next is assigned a weight and each neuron of the next is assigned a bias. Taking first neuron in the second layer as an example, the weights of the connec- tions are stored in a vector w0 = (w0,0, w0,1, w0,2), and the activations of the previous layer are stored in a vec- tor a0 = (a0,0, a0,1, a0,2). The activation of this neuron is then given by

$$a_{1,0} = \sigma\,(w_0 \cdot a_0 + b_{1,0}), \qquad (4)$$

where b1,0 represents the bias of the neuron, and σ rep- resents the activation function, which is rectified linear unit (ReLU) in the CNN used to detect ring galaxies.

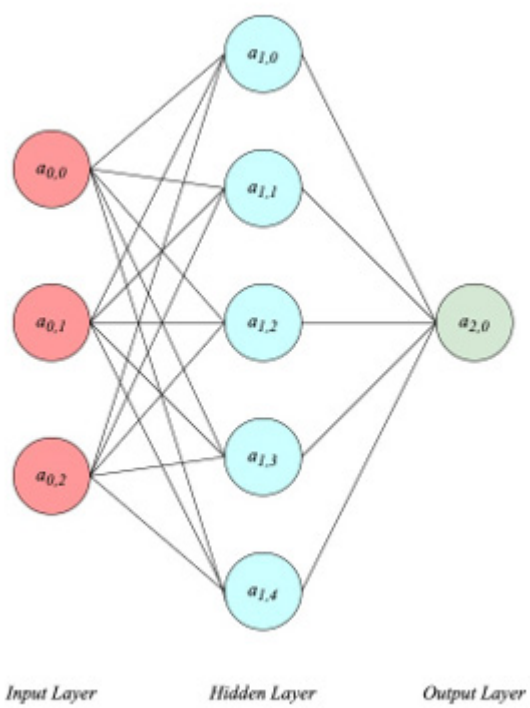


Figure 4. An illustration of a simple MLP. This contains an input layer, only one hidden layer, and an output layer.

The weights and biases are first initialized with ran- domized values. However, this creates a network which performs extremely poorly on an input example. Thus, these values need to be modified through a process called backpropagation. For this, a cost function is needed, which describes how poorly a given network is perform- ing. This is typically referred to as the loss, and in the CNN, binary cross-entropy loss is used, which is given by

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i), \qquad (5)$$

where N represents the number of outputs, yi represents the given class (1 or 0 in a binary case), and y^i repre- sents the outputted probability. To compress the output neuron between 0 and 1, the sigmoid activation function has to be used.
To get the best possible network, L needs to be minimized. In the CNN, this is done through stochastic gradient descent (SGD; Ruder 2017), which uses several iterations in mini-batches to find a relative minimum for the loss function through modifying the weights and biases. Simple SGD is given by

$$w_{t+1} = w_t - \alpha \cdot dW, \qquad (6)$$

where

$$dW = \frac{\partial L}{\partial w_t}. \qquad (7)$$

In the equation, t is the time step, w is the weight of a certain connection and α is the learning rate, which controls how quickly the weights are changed. The same concept is also applied to modify bias. However, SGD can further be improved upon with the Adam optimizer (Kingma

Table 2. Metrics for Various Models with Simulated Data

Model Name	Accuracy	Precision	Recall	F_1 Score	MCC	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
ResNet-50	95.69%	0.912	1.0	0.9545	0.917	0.981
ResNet-101	95.14%	0.912	0.997	0.953	0.906	0.975
Inception-ResNet-V2	99.94%	1.0	0.998	0.999	0.998	1.0

Figure 5. An illustration of the residual connections of ResNet. The identity of the initial function, x, is added to the feature map after convolution, f (x), to produce the output.

& Ba 2017), which uses an adaptive learning rate, meaning that different parameters have different learning rates. This allows for faster convergence on a relative minimum of the loss function. Adam combines two other optimizers, Momentum and RMSProp. Adam is given by

$$w_{t+1} = w_t - \alpha \frac{m_t}{\sqrt{s_t + \epsilon}}, \qquad (8)$$

where

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot dW \qquad (9)$$

and

$$s_t = \beta_2 \cdot s_{t-1} + (1 - \beta_2) \cdot dW^2. \qquad (10)$$

β1, β2 and E are constants which typically have values of 0.9, 0.999 and 10−8 respectively. The same concept can also be applied to find the correct value for bias.

A CNN (Fukushima 1980; LeCun et al. 2015) is a type of ANN which is typically used to classify images. A CNN differs slightly from the basic MLP in that there are stages of pre-processing to the input where key fea- tures are extracted. First, a convolutional layer takes a 3D block of data as an input. For images, this is a 3D matrix of their pixel values. A kernel, a matrix smaller than the input, is initialized with certain values. These values can be learnt through several iter- ations with SGD. This kernel moves across the initial image, and the dot product between the kernel and the smaller patch of the initial image is stored in a new ma- trix referred to as a feature map. An activation function, ReLU in this case, can be applied along with the convo- lution to modify the feature map. A convolutional layer is then followed by a max-pooling layer, which slides a kernel across the feature map and stores the largest value of each patch into a newly created feature map. These layers are alternated until a sufficent abstraction of the initial image is acheived. The outputted feature maps are then inputted into a series of perceptron lay- ers, similar to the MLP, and are reduced to an output containing a prediction as to what class the input im- age belongs to. The parameters for these layers are also learnt through SGD as seen with the ANN before.

Table 3. Metrics for Various Transfer Learned Models

Layers Unfrozen	Accuracy	Precision	Recall	F_1 Score	MCC	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Output	78.9%	0.680	0.699	0.689	0.530	0.841
Block C, Output	92.0%	0.894	0.866	0.879	0.821	0.950
All (No Transfer Learning)	85.0%	0.823	0.702	0.758	0.655	0.884

Typically, more layers can be added to a network to increase its accuracy. For extremely deep networks, the feature maps cannot be further downsized. Then, iden- tity mapping is required to add more layers, where con- volutions are performed with a kernel which will output a feature map with the same dimensions. However, for a standard CNN, this is difficult to learn, and results in the degradation problem, where the accuracy resulting with the addition of more layers initially increases, but falls after reaching a maximum.

To combat this problem, deep residual networks (ResNets; He et al. 2015) were created. As illustrated in Figure 5, these networks take the initial feature map, and store its identity in a variable, x. Then, a convo- lution is applied to the initial feature map to produce f (x). To produce the final feature map, H(x), f (x) is combined with x. These networks also help allevi- ate the vanishing gradient problem, where dW becomes too small to make significant changes to the weights. ResNets are created by stacking many of these types of layers.

To make deeper networks, inception modules (Szegedy et al. 2014) can be used as well. These modules use a concatenation of 3x3 max pooling and 1x1, 3x3 and 5x5 convolutions within a single layer to create a feature map. Inception modules offer the advantage of allowing for the creation of a larger network and the extraction of features at various scales, thus possibly learning more features.

However, with an extremely deep network comes the risk of being too attached to the training data, known as overfitting. Thus, several methods were implemented to ensure that the risk of overfitting was very minimal on the network. The first of these methods was regu- larization, which adds an extra penalty term to the loss function to ensure that it does not assume extreme val- ues. In this network, L2 regularization (Cortes et al. 2012) was used. The modified loss function with L2 regularization is given by

$$L_{new} = L + \frac{1}{N}\frac{\lambda}{2} \quad w_i^2, \qquad (11)$$

where the sum of the squared differences between the real and target weights is multiplied by the reciprocal of the number of outputs, N , and λ , where λ is the L2 regularization constant, which is set at 0.0001 in the network. Additionally, data augmentation, where the initial data is transformed via several methods and ex- panded, was used. A scale of the images by a factor of 0.1, a shear by a factor of 1.2, a zoom by a factor of 0.25 and rotations in the range 0◦ to 360◦ were used.

Using the Keras API, which is built on TensorFlow, the model was created. Three models were initially tested out and evaluated. The first had the ResNet- 50 (He et al. 2015) architecture, which has 50 layers and utilizes residual connections. Similarly, the second had the ResNet-101 architecture, which has 101 layers and utilizes residual connections. Finally, the third had the Inception-ResNet-V2 architecture (Szegedy et al. 2016), which has 164 layers and utilizes a combination of both inception modules and residual connections. These net- works were trained on the 100, 000 simulated images for 11 epochs with a batch size of 32 and a learning rate of 0.0001.

To evaluate these networks, 80% of this data was re- served for training, while 20% was reserved for testing. Half of the data reserved for testing was used for vali- dation, while the other half was used as a true test set, which avoids any influence the process of training might have had. The results of evaluation are described in Table 2. Using a

holistic evaluation of the various met- rics, Inception-ResNet-V2 was clearly the best network for the classification of the simulated ring galaxies. As a result, this network was then used for the process of transfer learning onto the real data.

3.3. Transfer Learning

To transfer the information the model learnt from the simulated data to the real data, a technique known as transfer learning (Zhuang et al. 2020) was used. To get better accuracy when limited data is available, transfer learning helps as the model is pre-trained on a different data set before training on the desired data.

Transfer learning involves freezing particular layers of the model, typically all layers excluding the output lay- ers, and retraining the layers which remain unfrozen. The earlier layers most likely do not need retraining, as they are primarily meant to extract trivial features. A common example of transfer learning is when a model trained on the ImageNet data set is applied to a new data set. However, similar to Ghosh et al. 2020, this model is first trained on a sample of simulated galaxies resembling the real data, and then transfer learned onto the real data.

To get the best possible model, different variations of transfer learning were tested. This included retraining just the output layers, retraining the output layers and inception block C, and retraining the entire model with- out using transfer learning.

The training data set contained about 9318 images, 3117 of which were ring galaxies. This data set was intentionally biased towards non-ringed galaxies to en- sure that there was minimal contamination in the fi- nal sample of ringed galaxies obtained when applied on an unclassified data set.

Data augmentation was used to prevent overfitting, along with early stopping, which stopped training once convergence was reached.

Additionally, a differential learning rate was used. The output layers, which were completely being re- trained, were assigned a learning rate of 10^{-3}. How- ever, the layers behind the output layers, if retrained, were assigned a learning rate of 10^{-5}. After the model reached convergence, an extra fine-tuning step was per- formed. This involved unfreezing the entire model, and retraining it at the extremely low learning rate of 10^{-7}, such that it could make minor adjustments.

The model was trained with a batch size of 32 and the early stopping had a patience of 10 epochs.

The model was again trained on 80% of the initial data, with 10% being reserved as the validation data set and 10% being reserved as the test data set. The results of the transfer learning are described in Table 3. The best suited model was holistically determined based on the statistics to be where both inception block C and the output layers were retrained. Then, this model was run on the unlabeled data. The galaxies classified as rings are described in § 5.

4. INFERRING THE PROPERTIES OF THE GALAXIES WITH SED FITTING

Using the BAGPIPES package (Carnall et al. 2018), I am able to infer stellar population properties from pho- tometric data. First, the photometric data from 11 bands for the identified ring galaxies is queried from the VizieR database. These bands are GALEX: FUV, NUV; PAN-STARRS: g, r, i, z, y; and WISE: W1, W2, W3 and W4. As data from only the visual part of the spec- trum will likely lead to a poor approximation, UV ob- servations from GALEX, visual observations from Pan- STARRS and infrared observations from WISE are com- bined. Then, along with the filter transmission curves, these observations are inputted into BAGPIPES.

BAGPIPES is able to use the Flexible Stellar Popula- tion Synthesis package (FSPS; Conroy & Gunn 2010) to create a model with multiple stellar parameters which represents a SED. Using Bayesian inference, this model is fit to the photometric data. Bayes' theorem, when applied in the context of SED fitting (Calistro Rivera et al. 2016; Carnall et al. 2018), states that

$$P(\theta|d) = \frac{P(d|\theta)\,P(\theta)}{P(d)}, \qquad (12)$$

where θ is the parameters which are used to define the model, and d is the initial photometric data.

P (θ|d) is also known as the posterior probability distribution, rep- resenting the information that is known with parameters

d. P (d|θ) is known as the likelihood function, or L (d|θ), and it represents the likelihood that the model defined by the parameters θ accurately models

d. P (θ) rep- resents the prior probability, or the information about the parameters already known, and P (d) represents the normalization constant. To minimize the error achieved by the model, BAGPIPES maximizes the natural log of the likelihood function (Acquaviva et al. 2011), known as the log- likelihood. This corresponds to a minimization of the $\chi 2$ value, thus producing the best possible model.

Values for specific parameters, along with error, can be inferred through the posterior probability density function (PDF) of a parameter. This can be found through integrating P (θ|d) over the other parame- ters (Calistro Rivera et al. 2016), a process known as marginalization, given by where θ_1 is the parameter in question in this example. However, calculating this N − 1 dimensional integral can get very computationally expensive, and not fea- sible when applied on the large scale. However, alterna- tive methods exist to estimate this, including MultiNest (Feroz et al. 2009) and Markov Chain Monte Carlo (MCMC; Pandya et al. 2018) sampling. MultiNest is applied in this scenario to explore the parameter space and determine PDFs for each parameter, thus being able to determine properties for these parameters such as their value and error.

$$P(\theta_1|d) = \int P(\theta|d)\, d\theta_2 d\theta_3 \cdots d\theta_N, \qquad (13)$$

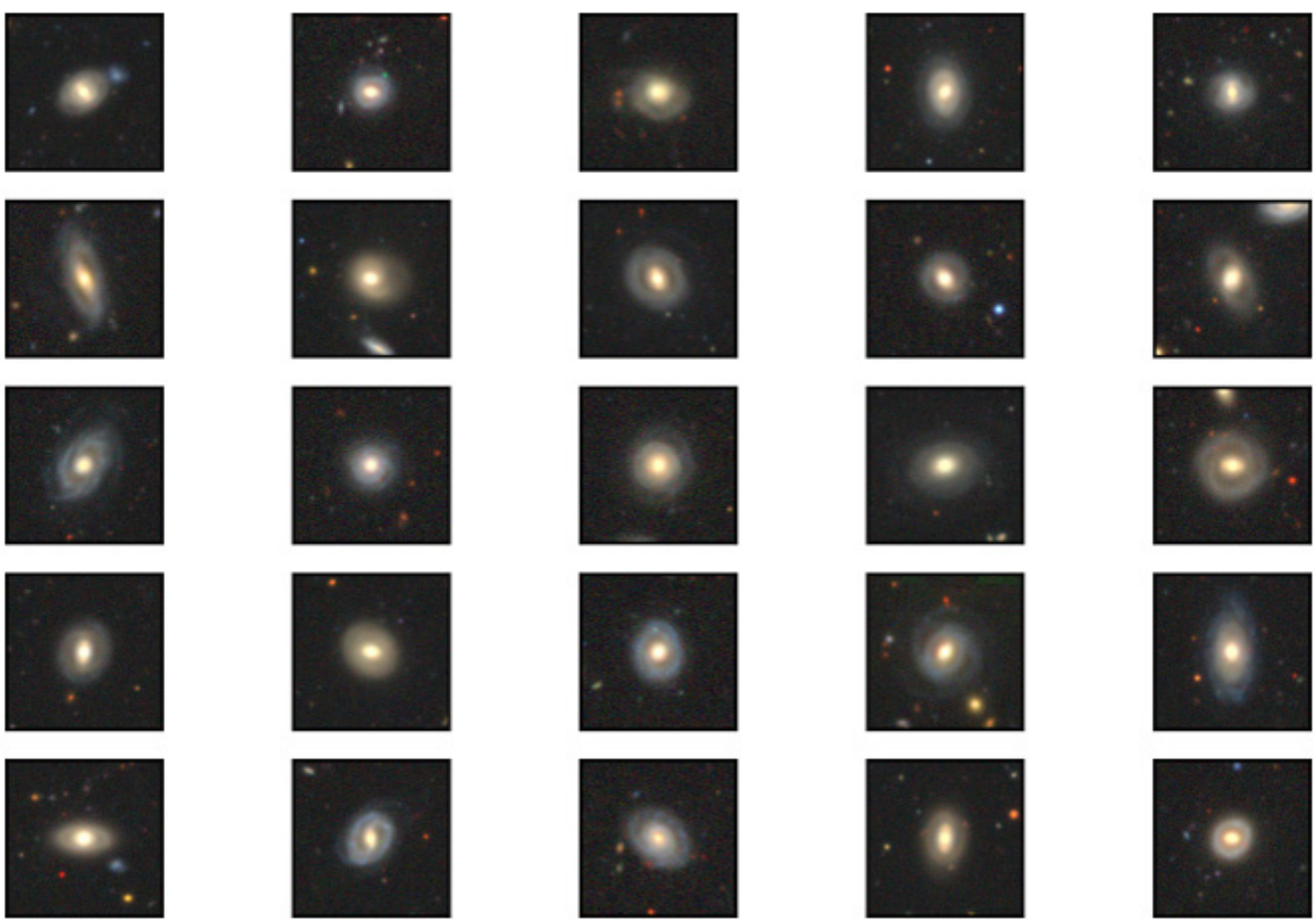


Figure 6. 25 randomly selected galaxies from the catalog of 580 rings.

Several particularly interesting parameters can then be analyzed, including the specific star formation rate (SSFR), stellar mass (M□), and redshift. In § 5 and § 6, I will describe these results and how they can be used to infer the formation and evolution of these galaxies.

5. RESULTS

5.1. Classification of Pan-STARRS Data

The model was applied to 966,329 previously unclassi- fied Pan-STARRS DR1 galaxies after training. A cutoff of a 97% probability to classify a galaxy as a ring was applied on the output from the final sigmoid layer of the model to ensure that the least number of non-ringed galaxies were misclassified as rings.

After the classifications of the model were manually filtered, 580 previously unidentified ring galaxy candidates were found, marking the creation of the largest computationally identified catalog of rings thus far. Additionally, the application of this model to the □950,000 Pan-STARRS galaxies took only around 10 hours, markedly better than the 14 months it took

Galaxy Zoo 2 volunteers to sort through the □300,000 galaxies of SDSS DR7.

Table 4 shows the catalog of identified ring galaxies, along with their coordinates and properties (obtained through SED Fitting). These galaxies have not previ- ously been classified as rings, making this the first time they have been recognized as rings. Figure 6 shows im- ages of 25 randomly chosen galaxies from this catalog.

The unclassified data set differed from the training, validation and confirmation data in that it was not specifically filtered to include a certain proportion of non-ringed and ringed galaxies: in a ”real” data set, less than 0.1% of the galaxies are rings, while the model was trained with a 2:1 ratio of non-ringed galaxies to rings. This led to a larger proportion of non-ringed galaxies being misclassified relative to ringed galaxies: 637 non- ringed galaxies were misclassified. However, the accu- racy is much better than conventional computational methods regardless, and only 0.06% of the entire data set is misclassified. Sorting through the misclassifica- tions takes mere hours for a researcher to perform, while sorting through the entire dataset would take on the or- der of months.

The misclassifications tended to mostly be small el- liptical galaxies, like the one shown in Figure 7. The model likely identified a common feature in these galax- ies which it misidentified as ringed structure, causing it to misclassify them.

Figure 7. A galaxy misclassified as a ring.

5.2. Analysis of Properties

Through SED Fitting, the stellar mass (M_*), specific star formation rate (SSFR) and redshift of the 580 ring galaxies were determined. Additionally, the apparent magnitudes of the galaxies were determined by

$$m = -2.5 \cdot \log_{10}\left(\frac{F}{F_0}\right), \qquad (14)$$

where F is the flux density of the galaxy through a given filter and F0 is the zero point flux density through that filter. Different colors of the galaxies, including FUV - r and g - r color, could be obtained by subtracting the apparent magnitudes of the galaxies in those filters in their respective combinations.

A control sample of 979 non-ringed galaxies was cho- sen from the initial catalog of Pan-STARRS galaxies. These were spiral galaxies, and were required to have r-band magnitudes between 16 and 19, the magnitude region which the ring galaxies were clustered at. This was to make sure that no fainter or brighter galaxies which would potentially have different properties would be chosen.

Using the Kolmogorov-Smirnov (KS) test, the distri- bution of SSFRs in the sample of ring galaxies was com- pared to that of the control sample of non-ringed galax- ies. The null hypothesis that both were from the same distribution was rejected with a p-value of .015, less than the p-value threshold of .05 which I had chosen.

The KS-test compares the empirical distribution func- tion (eCDF) of both samples of galaxies. Looking at the eCDFs of the galaxies in Figure 8, it is clear that the probability of finding a ring at a lower SSFR value is higher than that of finding a spiral at a lower SSFR value.

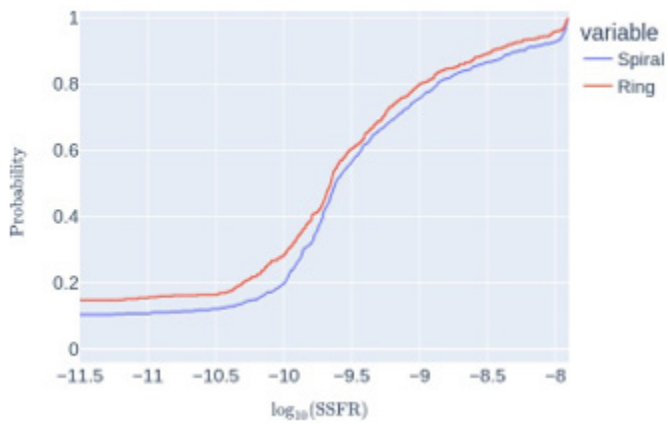


Figure 8. The eCDF of rings compared to the control sample of spirals.

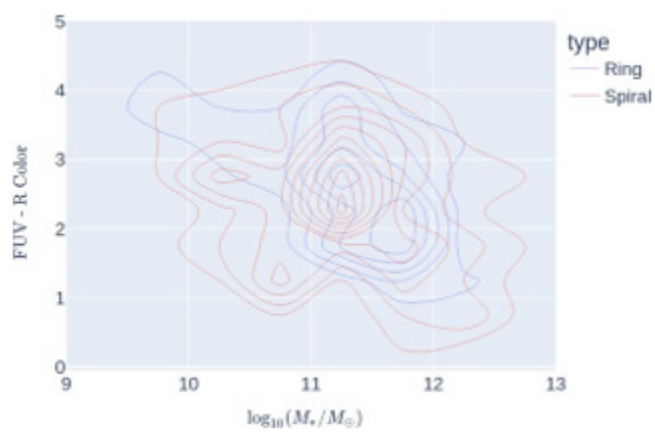


Figure 9. Color-mass diagram for FUV - r color for rings and spirals.

After this, the color-mass diagrams of both samples were compared against each other. On the color-mass diagram, redder galaxies (which have a lower apparent magnitude in the r-band filter) are typically older and less star forming than those which are bluer.

The FUV - r color of both rings and spirals were plot- ted against log10(M□/M0) on a density contour plot in Figure 9. Unlike the rings, the spi- rals were seen to ex- tend into bluer/lower values, this effect being more pro- nounced at lower mass- es.

However, unlike FUV - r color, this effect was not seen in g - r color. As seen in Figure 10, the rings and spirals were seen to peak at around the same areas, and neither the spirals nor the rings extended into redder or bluer values.

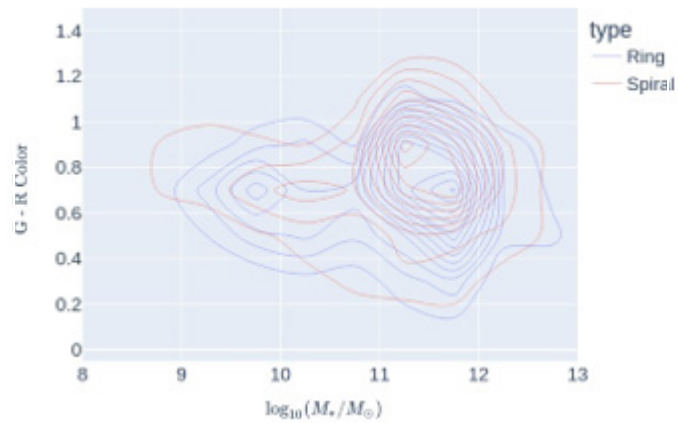


Figure 10. Color-mass diagram for g - r color for rings and spirals.

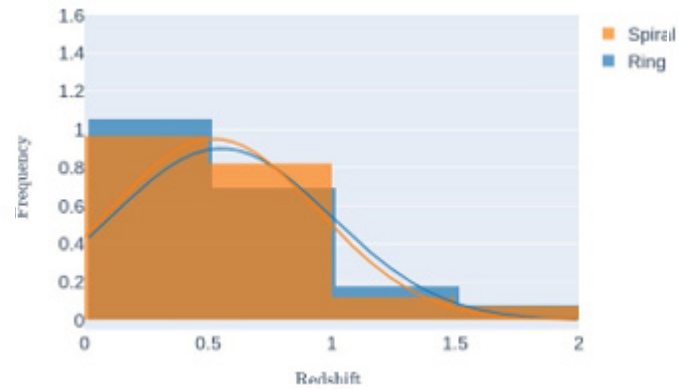


Figure 11. Frequency histogram of redshifts for both rings and spirals.

The FUV filter has a much lower wavelength range than the g filter, with its λmean being 1545.83 AO while the λmean of the g filter is 4900.12 AO.

Finally, the frequency of the redshifts of these galax- ies was plotted as a histogram, as seen in Figure 11. Although the distributions were ex- tremely similar, the redshifts of both samples were found to range from 0.012 to 1.998, which is an extremely large range of possible values.

6. DISCUSSION

6.1. Model & Applications

When applied to the unclassified data set, the model identified 580 ring galaxies candidates after manual verification. Compared to conven- tional computational mod- els for identifying ring galaxies, this model performed considerably

better.

The flood fill algorithm presented in Timmis & Shamir 2017 achieved a 6.9% accuracy when applied to unclas- sified data, and 77.2 ring galaxies were detected per 1 million. On the contrary, this model achieved a 47.6% accuracy when applied to unclassified data and detected 600.2 ring galaxies per 1 million. This was 6.6 times the accuracy and 7.8 times the detection rate of previous al- gorithms, dramatically improving the viability of com- putational methods for the detection of ring galaxies.

Ring galaxies are much harder to detect than the typ- ical galaxy due to the wide range of morphologies they can take: some are face-on and some almost lie hidden on the axis of the galaxy. Additionally, many features resemble rings, but are not actually indicative of them.

This was reflected in the misclassified sample of galax- ies. Most, like that shown in Figure 7, were small ellipti- cal galaxies. Some, however, were mergers of two galax- ies and even fewer were small spiral galaxies. Although this model did not misclassify images with noise and ar- tifacts, unlike previous models, it had trouble determining the classifications of some galaxies which exhibited certain traits of rings.

One potential future step to improve the accuracy of the model could be to increase the variety of the initial simulated sample of galaxies and try to model collisional and hoag-type ring galaxies in addition to just standard rings. Additionally, multiple CNNs could be stacked in an ”ensemble” network ensure that the best, most supported prediction is made for every image.

These results additionally show that a CNN can be built to detect other peculiar morphologies, for exam- ple, that of green pea (GP) galaxies. Even without large datasets, this research shows that methods such as transfer learning paired with synthetic data can al- low for a competitive accuracy to be obtained.

With this model, future sky surveys which produce billions of images, such as the LSST, can easily be probed for ring galaxies. Additionally, existing sky sur- veys which have not yet been explored can easily be searched for thousands of more ring galaxy candidates. Without having to spend as much time to obtain an op- timal sample of ring galaxies, their formation and evo- lution can hence be studied in more depth with larger samples.

6.2. Properties & Analysis

As the sample of ring galaxies identified by the CNN had a lower star formation rate and was redder than its spiral counterparts, it could be deduced to be further in its stages of galactic evolution. These deductions agreed with the findings of Fernandez et al. 2021.

However, they found that these properties were exhib- ited by their sample of rings to a much greater extent than in this study. While the spirals clearly extend be- yond the rings into a bluer range in the FUV - r color- mass diagram, Fernandez et al. 2021 finds that there is a clear distinction between rings and spirals in the g - r color-mass diagram as well.

As previously mentioned, the FUV filter is at a much lower wavelength than the g filter. As the distinction between spiral and ring is clearly visible in FUV - r but not in g - r, my sample of rings can be determined to be less red than that of Fernandez et al. 2021. The most likely explanation resulting from this is that my sample of galaxies is younger than what they had examined.

However, when looking at the redshift distributions of my sample, another potential difference appears. My redshift distribution was spread out from 0.01 to 2, while that of Fernandez et al. 2021 was restricted to be be- tween $0.01 < z < 0.1$. This large range of redshifts could be a potential factor in the variance of the properties of both samples of galaxies as well.

From these results, existing knowledge about ring galaxies was affirmed with new data from a relatively unexplored area of the sky. Additionally, the proper- ties of the galaxies which were obtained could possibly be used to conduct further investigations into ring galax- ies, and researchers could build upon existing knowledge about these galaxies.

The sample of 580 newly classified rings, in addition, could be further investigated. Simulations could possi- bly be conducted to theorize how they formed, and spec- troscopic data can be collected by researchers to further explore the dynamics of these galaxies. Additionally, a subset of collisional ring galaxies can be extracted, and they could potentially be investigated for the presence of inter-

mediate-mass black holes.

7. CONCLUSION

In this work, I investigate the usage of a convolutional neural network to detect large catalogs of ring galaxies from sky surveys. Rings are crucial to understand more about galaxy dynamics, and current catalogs of them are extremely limited.

My network was first trained on a sample of 100,000 synthetic galaxies - simulated using GalFit - and was then transfer learned to a sample of real galaxies con- sisting of 3117 rings. It was found that the Inception- ResNet V2 architecture acheived the best accuracy, and the best variation of transfer learning was when convo- lutional block C and the output layers were retrained. The source code of the model is made public at https: //github.com/harishk30/RingGalaxiesCNNAnalysis.

After training, this network was applied to an unclas-
sified sample of □950,000 galaxies from Pan-STARRS. It detected 580 previously unclassified rings, making this the largest catalog of computationally identified rings. Additionally, the network had 6.6 times the accuracy and 7.8 times the detection rate of conventional algo- rithms, showing promise for the future application of computational methods to identify ring galaxies.

The properties of this sample were then analyzed through SED fitting with photometric data from 11 dif- ferent filters. The null hypothesis that the star forma- tion rates of both the rings and spirals were from the same distribution could be rejected with a p-value of
.015. Additionally, the probability of finding a ring at a lower SSFR was higher than that of a spiral, as shown in Figure 8, aligning with prior studies. The rings were found to extend to bluer colors in the FUV - r color- mass diagram, as seen in Figure 9, but not in the g - r color-mass diagram, as seen in Figure 10, most likely indicating that the sample of ring galaxies was younger than that of previous studies.

In the future, this model could be used to extract large catalogs of rings from future sky surveys which will collect images of billions of more galaxies. Addi- tionally, current sky surveys which have not yet been explored could be classified by this model to compile thousands of more existing rings which have yet to be classified. The sample of 580 rings which were identified by this model could further be explored through con- ducting simulations and collecting spectroscopic data, and insight about the formation and evolution of rings as a whole can be derived.

8. ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Bryce Kalmbach, for his supervision throughout the duration of my project. The astropy, matplotlib, numpy, scipy, pandas, bagpipes and MultiNest libraries were used.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agu¨eros, M. A., et al. 2009, ApJS, 182, 543, doi: 10.1088/0067-0049/182/2/543

Acquaviva, V., Gawiser, E., & Guaita, L. 2011, 7, 42–45, doi: 10.1017/s1743921312008691

Appleton, P. N. 1999, in Galaxy Interactions at Low and High Redshift, ed. J. E. Barnes & D. B. Sanders, Vol. 186, 97

Appleton, P. N., & Struck-Marcell, C. 1996, FCPh, 16, 111 Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393,
doi: 10.1051/aas:1996164 Brosch, N. 1985, A&A, 153, 199

Buta, R. 1995, ApJS, 96, 39, doi: 10.1086/192113

Buta, R. J. 2017, Monthly Notices of the Royal Astronomical Society, 471, 4027–4046, doi: 10.1093/mnras/stx1829

Calistro Rivera, G., Lusso, E., Hennawi, J. F., & Hogg,
D. W. 2016, Astrophysical journal., 833, 98

Carnall, A. C., McLure, R. J., Dunlop, J. S., & Dav´e, R. 2018, Monthly Notices of the Royal Astronomical Society, 480, 4379–4401, doi: 10.1093/mnras/sty2169

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2019, The Pan-STARRS1 Surveys. https://arx-

iv.org/abs/1612.05560
Conroy, C., & Gunn, J. E. 2010, FSPS: Flexible Stellar Population Synthesis. http://ascl.net/1010.043
Cortes, C., Mohri, M., & Rostamizadeh, A. 2012, L2 Regularization for Learning Kernels. https://arxiv.org/abs/1205.2653
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, The Astronomical Journal, 157, 168, doi: 10.3847/1538-3881/ab089d
Fernandez, J., Alonso, S., Mesa, V., Duplancic, F., & Coldwell, G. 2021, Astronomy Astrophysics, 653, A71, doi: 10.1051/0004-6361/202141208
Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601–1614, doi: 10.1111/j.1365-2966.2009.14548.x
Finkelman, I., Moiseev, A., Brosch, N., & Katkov, I. 2011, Monthly Notices of the Royal Astronomical Society, 418, 1834–1849, doi: 10.1111/j.1365-2966.2011.19601.x
Fukushima, K. 1980, Biological Cybernetics, 36, 193, doi: 10.1007/bf00344251
Ghosh, A., Urry, C. M., Wang, Z., et al. 2020, The Astrophysical Journal, 895, 112
Goddard, H., & Shamir, L. 2020a, The Astrophysical Journal Supplement Series, 251, 28, doi: 10.3847/1538-4365/abc0ed
—. 2020b, The Astrophysical Journal Supplement Series, 251, 28, doi: 10.3847/1538-4365/abc0ed
He, K., Zhang, X., Ren, S., & Sun, J. 2015, Deep Residual Learning for Image Recognition. https://arxiv.org/abs/1512.03385
Hoag, A. A. 1950, ApJ, 55, 170

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, The Astrophysical Journal Supplement Series, 221, 8, doi: 10.1088/0067-0049/221/1/8
Ivezi´c, Z˘., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873,
111, doi: 10.3847/1538-4357/ab042c
Kim, E. J., & Brunner, R. J. 2016, Monthly Notices of the Royal Astronomical Society, 464, 4463–4475,
doi: 10.1093/mnras/stw2672
Kingma, D. P., & Ba, J. 2017, Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980
Lanusse, F., Ma, Q., Li, N., et al. 2017, Monthly Notices of the Royal Astronomical Society, 473, 3895–3906,
doi: 10.1093/mnras/stx1665
LeCun, Y., Bengio, Y., & Hinton, G. 2015, nature, 521, 436 Pandya, V., Romanowsky, A. J., Laine, S., et al. 2018, ApJ,
858, 29, doi: 10.3847/1538-4357/aab498
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002,
The Astronomical Journal, 124, 266–293, doi: 10.1086/340952
—. 2010, The Astronomical Journal, 139, 2097–2129, doi: 10.1088/0004-6256/139/6/2097
Rose, S. C., Naoz, S., Sari, R., & Linial, I. 2021, The Formation of Intermediate Mass Black Holes in Galactic Nuclei. https://arxiv.org/abs/2201.00022
Ruder, S. 2017, An overview of gradient descent optimization algorithms. https://arxiv.org/abs/1609.04747
Schweizer, F., Ford, W. Kent, J., Jedrzejewski, R., & Giovanelli, R. 1987, ApJ, 320, 454, doi: 10.1086/165562
Shamir, L. 2011, The Astrophysical Journal, 736, 141, doi: 10.1088/0004-637x/736/2/141
—. 2019, Monthly Notices of the Royal Astronomical Society, 491, 3767–3777, doi: 10.1093/mnras/stz3297
Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. 2016, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. https://arxiv.org/abs/1602.07261
Szegedy, C., Liu, W., Jia, Y., et al. 2014, Going Deeper with Convolutions. https://arxiv.org/abs/1409.4842
Timmis, I., & Shamir, L. 2017, A catalog of automatically detected ring galaxy candidates in PanSTARRS. https://arxiv.org/abs/1706.03873
Trinchieri, F. P. A. W. G. 2010, Chandra observations of the ULX N10 in the Cartwheel galaxy. https://arxiv.org/abs/1003.4671
Walcher, J., Groves, B., Budav´ari, T., & Dale, D. 2010, Astrophysics and Space Science, 331, 1–51, doi: 10.1007/s10509-010-0458-z

Environmental Science

Rubber Tire Leachate Effect in Lake Washington and Lake Sammamish

Hahwon C. Chue

Nikola Tesla STEM High School

Abstract

The effect of rubber pollutants from the playgrounds our children play in has been minimally studied, offering potential unknown risks. The project seeks to investigate the correlational effects that rubber tire leachate from playgrounds has on Lake Washington & Lake Sammamish and interpret rubber leachate pollutants levels in our lakes. The test sites of rubber granulate possessing playgrounds ('source') were tested in set distances of 1, 5, 15, and 25 meters away from the source and pollutant levels were tested. The hypothesis of rubber tire leachate being observed at a higher quantity the closer it was to the source was proven to be statistically significant through the R squared value of 0.828 for Zinc as well as a R squared Value of 0.793 for Nitrate. Both these pollutants having known harmful attributes to the human microbiome as well as marine life. This experiment showed that the playgrounds possessing rubber tire granulate will in fact leach into nearby lakes at a higher quantity the closer to the source.
Keywords: rubber tire leachate(s), source(s), end of life tires, granulate

Introduction

This project aims to find the leachate sources in Lake Washington and Lake Sammamish and interpret rubber leachate pollutant levels to properly understand the effects of rubber debris in water quality. Rubber mulch, a material widely regarded to be safe due to its popularity in recent decades, can be found in many urban and rural landscaping projects throughout the United States. Often derived from recycled rubber tires, rubber mulch is regarded as an affordable and long-lasting loose-fill surfacing material, frequently enduring years of wear. The United States generates just under 300 million waste tires annually (CMRC), with most of these tires going through a recycling process. The effect of rainwater can also be observed on Leachate deposited waste layers when the humidity is over 45% (MDPI). Along the same lines, a study uncovered a detrimental effect that may be produced as a byproduct of the leachate exhibited from the water passing through the rubber mulch. (Gualtieri, 2005). The leachate contains a multitude of heavy metals, pigments, and oils that accumulate during manufacturing runoff into the surface and groundwater (Halsband, 2020). Studying this topic would give proper information to make healthy choices for the environment and mitigate potential contamination done upon the water sources. By assessing this study, this project will determine the severity and effect of the impact that rubber mulch has on the ecosystems and on eco life it resides in Lake Sammamish and Lake Washington.

The rubber mulch leachate would pollute the nearby aquatic ecosystems with observed toxins at a higher rate when closer to the pollutant source due to the higher leachate concentration exhibited from the ELT (End of Life Tire(s)). This will exhibit a positive Pearson's correlation coefficient proving that there is statistical significance between the rubber leachate tire source and the Lake nearby. Instigating an area of research potentially un-resolved by prior research and findings in a setting applicable to both industrial and non-industrial water sources.

Methodology

As each test site is deduced based on prior background research of finding end of life tires and

rubber tire granulate source playgrounds, systematic sampling of various locations in the Lake Sammamish/ Lake Washington area has been conducted as means of this experiment.

Collection of data has been done on-site at each location. Each experiment has been set up with the test kit materials for the pollutants Zinc, Pyrene, and Bisphenol-A. The pH level of the water has been tested at each site to mitigate correlational biases. After the test kits, Zinc Check and Water Test Kit are set up, distances have been recorded using Distance Pinner on the App store. The respective distances: 1 meter, 5 meters, 15 meters, 25 meters, will be used as a control from the leachate water source.

After all the test kits and safety precautions of Face masks, Rubber Gloves, and Eye protection, were set up on-site at the respective distance. Sampling has been done using a plastic pipette to draw water from the source. The pipette will then disperse water into a granulated glass beaker for Test kit analysis. One strip of each test kit has been placed into the granulate beaker until data is able to be interpreted using the product. Once both strips are used, the experiment will had been done three additional times for accuracy of the data.

After all three sets of data were interpreted at each distance, data had been inserted into the tables 1-4 below. This has been done for each site as means of differentiating any correlational influences that may have occurred in one specific rubber tire leachate source.

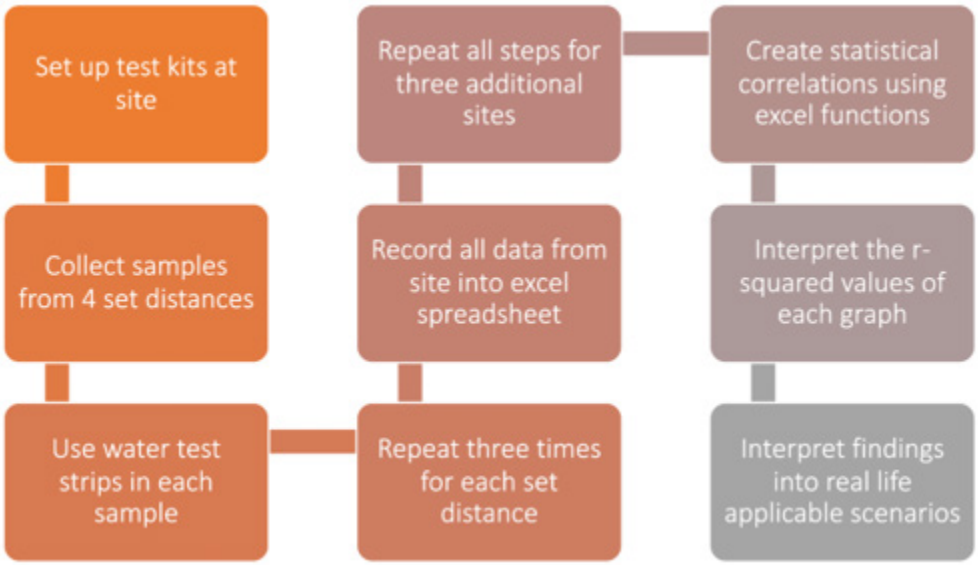


Table 1
Pollutant Concentrations at distance away from Site One(ppm)

	Distances			
Pollutants	1 Meter	5 Meter(s)	15 Meter(s)	25 Meter(s)
Zinc	1.33ppm	0.66ppm	0.66ppm	0ppm
Nitrate	6.66ppm	3.33ppm	0ppm	0ppm
Nitrite	0.33ppm	1ppm	0ppm	0ppm
PH Level	6	6.16	6	6.33

Note. Pollutant's concentrations are the average of three test-stripped experiments

Table 2
Pollutant Concentrations at distance away from Site Two(ppm)

	Distances			
Pollutants	1 Meter	5 Meter(s)	15 Meter(s)	25 Meter(s)
Zinc	1.66ppm	0.33ppm	0.33ppm	0ppm
Nitrate	6.66ppm	1.66ppm	0ppm	0.33ppm
Nitrite	0ppm	0ppm	1ppm	0ppm
PH Level	6	6	6	6

Note. Pollutant's concentrations are the average of three test-stripped experiments

Table 3
Pollutant Concentrations at distance away from Site Three(ppm)

	Distances			
Pollutants	1 Meter	5 Meter(s)	15 Meter(s)	25 Meter(s)
Zinc	3.66ppm	2.33ppm	1ppm	1.33ppm
Nitrate	3.33ppm	0.66ppm	1.66ppm	0ppm
Nitrite	0.33ppm	0.33ppm	0ppm	0ppm
PH Level	6.16	6.16	6.16ppm	6.16ppm

Note. Pollutant's concentrations are the average of three test-stripped experiments

Table 4
Pollutant Concentrations at distance away from Site Four(ppm)

	Distances			
Pollutants	1 Meter	5 Meter(s)	15 Meter(s)	25 Meter(s)
Zinc	1.66ppm	1ppm	1.33ppm	1ppm
Nitrate	5ppm	2.66ppm	0.66ppm	2.33ppm
Nitrite	1.66ppm	0.66ppm	0.33ppm	0.33ppm
PH Level	6.16	6	6.16	6

Note. Pollutant's concentrations are the average of three test-stripped experiments

Data Analysis

EPA Standard Levels. EPA standard levels for each pollutant are derived from levels of industrial activity as well as source pollution into the water streams. Working with other agencies such as "Our Living River", the EPA works to minimize the pollution of water while improving the water quality of the source. The acceptable EPA levels for each pollutant being:

Zinc: 5mg/L (5 ppm)
Nitrite: 1ppm
Nitrate: 10ppm
pH: 6.5-8.5

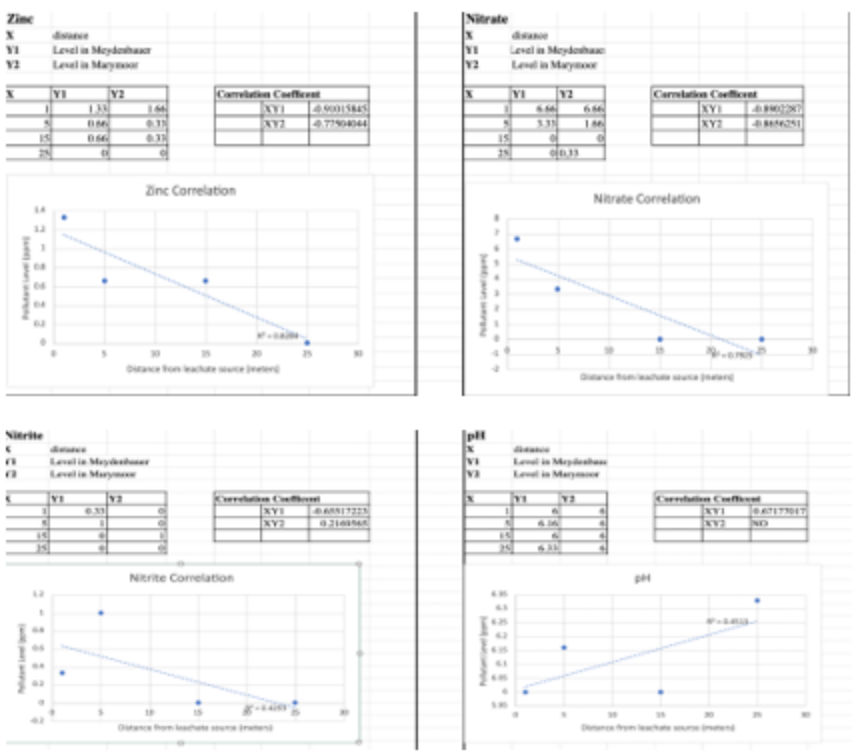
Pearson product moment correlation. To identify the strength of the correlation between the Rubber tire leachate source and the distance at which the sampling data was recorded from, Pearson's product moment correlation, or r-score, was calculated using the data of ppb averages in tables 1-4 in relation to the distance recorded from. To calculate the Pearson's product co-efficient, data is manually inserted into the excel document from Microsoft, giving a correlation coefficient and R Squared value that is either statistically significant or non-significant.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Investigative research. After interpretating the statistical significance of the rubber tire leachate source and pollution levels with distance, further analysis of publicly available data regarding the health of marine-life subsiding within the Lake-Washington and Lake-Sammamish regions will be conducted using science journal articles cited below.

Investigative research. After interpretating the statistical significance of the rubber tire leachate source and pollution levels with distance, further analysis of publicly available data regarding the health of marine-life subsiding within the Lake-Washington and Lake-Sammamish regions will be conducted using science journal articles cited below.

Figure.1 Correlational Charts for Zinc, Nitrate, Nitrite, pH levels



Note: This table(s) summarizes the R-Squared value for Zinc, Nitrate, Nitrite, and pH levels

Figure.2 Meydenbauer location

Note: This setup depicts the materials used in accordance with the methodology

Results

After collecting the data across four different sites and comparing the R squared values within a correlational graph, the only statistically significant pollutants were found to be Zinc (0.82 R Squared value) and Nitrate (0.79 R Squared value). Thus, proving a correlation between the source of rubber leachate and the distance away in the water source, an analysis is conducted regarding the potential effects of Zinc and Nitrate due to their respective EPA guideline levels. As the pollutant Zinc was below the EPA guideline level of 5ppm and the pollutant Nitrate was below the EPA guideline level of 10ppm, no immediate cause of harm. Though not proving an immediate cause of harm, this statistical relationship between Zinc and Nitrate to the distances from the water source and the rubber leachate source proves to be a severe effect in assessing the quantity of the correlation between these two variables. This poses potential risks with increased leachate effects due to other variables such as heat (Dillon) and may be a cause of harm for our nearby lakes soon. As global warming reaches unprecedented scales, we must counteract against any potential pollutant sources that may pose a dire effect on our Lake Sammamish and Lake Washington. These include rubber tire granulate possessing playgrounds which have the capacity to raise Zinc and Nitrate levels across our Lakes, posing a harm to human health as well as the health of marine life. Future steps may be to be more conscientious about our usages of recycled rubber tire products and their omnipresent detrimental effects once past their means of original use. This conscientiousness may be implemented in forms of research assessing more in-depth correlational ground water to rubber tire leachate sources such as that from playgrounds.

Acknowledgements

This project was done with the mentorship of Dr. Lyda Harris, a microplastic(s) researcher from the Seattle Aquarium with a PhD in the department of Biology UW.

References

Capolupo, M., Sørensen, L., Jayasena, K. D. R., Booth, A. M., & Fabbri, E. (2020). Chemical composition and ecotoxicity of plastic and car tire rubber leachates to aquatic organisms. Water Research, 169, 115270. https://doi.org/10.1016/j.watres.2019.115270

Correlation Coefficient: Simple Definition, Formula, Easy Steps. (2021, July). Retrieved January

15, 2022, from Statistics How To website: https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/

Defois, Clémence, et al. "Environmental Pollutant Benzo[A]Pyrene Impacts the Volatile Metabolome and
Transcriptome of the Human Gut Microbiota." Frontiers in Microbiology, vol. 8, 15 Aug. 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5559432/, 10.3389/fmicb.2017.01562. Accessed 1 Oct. 2021.
Defois, C., Ratel, J., Denis, S., Batut, B., Beugnot, R., Peyretaillade, E., Engel, E., & Peyret, P. (2017).
Environmental Pollutant Benzo[a]Pyrene Impacts the Volatile Metabolome and Transcriptome of
the Human Gut Microbiota. Frontiers in Microbiology, 8.
https://doi.org/10.3389/fmicb.2017.01562
Dillon, E. (2017). The Effect of Temperature on Zinc Leaching from Rubber Tire Mulch. In Journal of
the South Carolina Academy of Science. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1154&context=jscas
Duda1, A., Kida2, M., Ziembowicz2, S., Koszelnik2, P., DA, B., KL, B., TL, G., M, C., L, S., KDR, J.,
AM, B., E, F., Statistical Office, C., KE, D., KE, H., JL, M.-S., CT, B., BJ, D., AR, D., … PE, J.
(2020, July 20). Application of material from used car tyres in geotechnics-an environmental
impact analysis. PeerJ. Retrieved October 15, 2021, from https://peerj.com/articles/9546/.
Gualtieri, M., Andrioletti, M., Mantecca, P., Vismara, C., & Camatini, M. (2005, March 24). Impact of
tire debris on in vitro and in vivo systems. Particle and fibre toxicology.
https://www.ncbi.nlm.nih.gov Gualtieri M;Andrioletti M;Vismara C;Milani M;Camatini M; (n.d.). Toxicity of tire debris leachates.
/pmc/articles/PMC1079942/.
Halsband, C., Sørensen, L., Booth, A. M., & Herzke, D. (1AD, January 1). Car tire CRUMB Rubber:
Does LEACHING produce a toxic chemical cocktail in coastal marine systems? Frontiers.
https://www.frontiersin.org/articles/10.3389/fenvs.2020.00125/full.
Halsband, C., Sørensen, L., Booth, A. M., & Herzke, D. (2020). Car Tire Crumb Rubber: Does Leaching
Produce a Toxic Chemical Cocktail in Coastal Marine Systems? Frontiers in
Environmental Science, 8. https://doi.org/10.3389/fenvs.2020.00125
K.E. Day, K.E. Holtze, J.L. Metcalfe-Smith, C.T. Bishop, B.J. Dutka, Toxicity of leachate from
automobile tires to aquatic biota, Chemosphere, Volume 27, Issue 4, 1993, Pages 665-675,
ISSN 0045-6535, https://doi.org/10.1016/0045-6535(93)90100-J.
Šourková, M., Adamcová, D., Winkler, J., & Vaverková, M. D. (2021). Phytotoxicity of Tires Evaluated
in Simulated Conditions. Environments, 8(6), 49.
https://doi.org/10.3390/environments8060049
Tamis, J. E., Koelmans, A. A., Dröge, R., Kaag, N. H. B. M., Keur, M. C., Tromp, P. C., & Jongbloed, R.
H. (2021). Environmental risks of car tire microplastic particles and other road runoff
pollutants. Microplastics and Nanoplastics, 1(1). https://doi.org/10.1186/s43591-021-00008w
Treading Water: Tire Wear Particle Leachate Recreates an Urban Runoff Mortality Syndrome in Coho
but Not Chum Salmon. (2021). Environmental Science & Technology.
https://pubs.acs.org/doi/10.1021/acs.est.1c03569

Predicting Waterfall Forms Through Feature Engineering and Model Selection Machine Learning

Siddhant Porwal

Nikola Tesla STEM High School

Abstract

Waterfalls are vital for the renewal of our rivers, eroding mineral-rich rocks that enrich water streams, enhancing river health, and supporting life habitat across 17,100 locations in the USA. Waterfalls change forms over time due to erosion and increased stream flows as a result of climate change – thus, analyzing waterfalls can efficiently enhance river erosion and flood management, ensuring these crucial habitats remain uncompromised. Therefore, this unique project was undertaken to predict waterfall form (Plunge, Horsetail, Steep-Cascade, Shallow-Cascade, Rapid) based on pertinent geological features, such as dimension, angle-of-decline (AOD), and seasonal streamflow. Each geological feature was determined relevant and validated for machine learning applicability through calculated r2 values (AOD with highest r2=0.66), then collated and curated in a database through robotic process automation. Utilizing Feature Engineering, data of 10,590 waterfalls was cleansed and statistically engineered to improve data quality, and as waterfalls could not be classified by conventional methods, machine learning models best suited for supervised classification (Random-Forest (RF), Support-Vector-Machine (SVM), Naïve-Bayes, and k-Nearest-Neighbors) were leveraged and trained on the geological features and cross validated for performance. The SVM exhibited the highest AUC accuracy of 94.16%, and RF displayed the highest F-Measure accuracy of 86.73%. This novel methodology was validated and compared to traditional ML without feature engineering and showed statistically significant improvement in AUC and F-Measure accuracy by +22.03% and +31.66% respectively. While waterfall forms take a long time to change, in the event that a change occurs, there is very little time left to react. Thus, this project makes it possible for enhanced flood and disaster preparation, discovering waterfall angle-of-decline, perpendicular run, and height as critical waterfall features in need of proactive monitoring to remediate any drastic ecological changes. Alongside, this project presents a modular approach that can be applied to other understudied, ecologically critical scenarios such as glacial melting and landslides, both of which will frequent the globe as time goes on.

Key Words

Waterfalls, Machine Learning, Naive-Bayes, Support Vector Machine, Random Forest, k-Nearest Neighbors, Feature Engineering, Model-Selection, Erosion, Fluvial Knickpoints, Data Extraction, Collation, Robotic Process Automation (RPA)

Goal

This project (1) collates a dataset that hosts waterfall features, (2) develops a diverse set of candidate machine learning models to predict waterfall forms using a novel methodology, and (3) identifies critical ecological thresholds to monitor for addressing erosion management/river health as our climate continues to change. This project provides modular framework for further research and opens newer frontiers on environmental engineering.

Introduction

Waterfalls — known as fluvial knickpoints — are crucial landforms, interconnected with river systems, erosion, stream flows, and climate changes. Maintaining Earth's ecological balance, waterfalls add to rivers' dissolved oxygen contents, ultimately purifying water. Naturally dividing rivers, waterfalls enrich flora and fauna, providing nutrient-rich deposits for plants and animals alike.

Even replenishing current demands, waterfalls provide for human dependencies, for example, up to 16% of global energy demands through hydropower (USGS 2018). In fact, according to data from the Worldwide Waterfall Database, over 17,000 waterfalls tower rivers in the United States of America, sustaining over 17,000 diverse environments, all critical to humans, wildlife, and the Earth's balance in these such ways. How so? Changes brought by waterfalls impact not only topology, but also changes in ecology, which varies as the climate changes.

However, as important as waterfalls are, they currently lack proactive monitoring. Only sparse data exists, even in our governmental agencies, regarding these crucial landscapes, hence it's impending that a broader, more proactive approach to understanding waterfalls be taken. After all, waterfalls are not static landforms, continuously eroding upstream by the sheer force of a rushing river, thus playing a central role in geological problems. Adding on, knowing that many waterfalls surround current human infrastructure as well (whether tourist locations or hydropower plant), it is equally important to have continuous data tracking and analysis to monitor and manage hazards as flash floods increase in prominence by 72% every decade, according to data from the National Weather Service.

Hayakawa et al. proposed waterfall classifications by measuring erosive strength through a Force-Resistance equation that utilized streamflow, precipitation, width, height, and rock strength, although it was only applicable to the 17 waterfalls of their study area; they could not generalize their equation to the overall population of waterfalls.

Goudie et al. analyzed common metrics of classifying waterfalls, including the 1939 Lobek Classification System and the 2018 National Geographic Classification System (Goudie et al. 2020). Goudie concluded that every system had its benefits and drawbacks, later suggesting the metric employed by Hayakawa et al. (2005) as among the most holistic. This revealed gaps in addressing classification, the crucial precursor to understanding impacts on ecology under climate changes.

This project investigated and classified 10,590 waterfalls based on their geological parameters, including for example waterfall dimensionality, angle of decline, perpendicular run, streamflow levels, and rock type. By making use of currently limited data, the project employed d*ata extraction, collation,* and *feature engineering* to expand the amount of usable data available. This data obtained was then run against four machine learning algorithms—k-Nearest, Support Vector Machine, Naive-Bayes, and Random Forest—to not only show the applicability of machine learning as a tool for furthering waterfall research, but also identify critical parameters to monitor for thresholds in erosion and hazard management.

Within the scope of the project were the following experimental subgoals,

1. To engineer available features through normalization and standardization—to meet this end, data was collated from the World Waterfall Database and statistically placed relative to all known data points on a common scale.
2. To develop candidate machine learning models capable of recalibration and feature addition. With this, waterfall form can be iteratively tested, with the end goal to enhance prediction accuracy (ROC AUC and F-Measure).

The model goal of this project was to achieve an F1-Metric/F-Measure accuracy percentage above 75% and AUC Score above 90% by developing ML-model configurations with a high specificity and sensitivity.

Waterfall Classifications and Parameters:

Figure 1: Waterfall Forms/Classification Types. These are the major waterfall classification types; there will always be a few outliers. A majority of the recorded waterfalls are of the plunge and horsetail types; only a few rapids have been recorded.

The waterfall forms predicted by this project can be summarized by Figure 1.

This system originates from the classification developed by the World Waterfall Database. Accordingly, a Plunge waterfall consists of a steep decline in altitude, with typically a pitch angle around 90 degrees. Next, a Horsetail waterfall is one with a fanning drop of water: this is often due to the occasional contact with bedrock changing the course of water as it falls. Steep Cascade waterfalls are waterfalls in which water remains in full contact with the bedrock, with pitch angles comparable to Plunge and Horsetail waterfalls; Shallow Cascades, similar to Steep Cascades, have pitch angles within 10 to 40 degrees. Lastly, Rapids are all waterfalls with pitch angles less than 10 degrees. Rapids are usually not considered as waterfalls, which is why they have the least amount of recorded data. Along with that, many waterfalls exist as outlier cases, not fully resembling any one of these five types.

The geological parameters utilized in the machine learning models are summarized by Figure 2. All waterfalls have a width, a height, a run, a pitch angle/AOD, drops, a source, streamflow volumes, rock types, a perpendicular run, and the height of the tallest drop. These features were used to train Machine Learning.

Figure 2: Waterfall Features used in Models. This diagram highlights the definitions of all features mentioned, with the Snoqualmie Waterfall as an example. Note: θ represents Pitch.

All waterfalls have a width, a height, a run, a pitch angle/AOD, drops, a source, streamflow volumes, rock types, a perpendicular run, and the height of the tallest drop. These features were used to train Machine Learning.

Machine Learning Model Types:

Figure 3: Candidate Machine Learning Models. (1) Support Vector Machine bounds instances of different classes through linear or curved *vectors* to predict data classification. (2) Random Forest creates a series of tests in the form of a *tree* to determine data classification. (3) Naive Bayes utilizes Bayes Theorem of probability to classify data. (4) k-Nearest Neighbors utilizes the classifications of similar data instances to predict classification.

The four machine learning models employed are summarized with Figure 3. These four machine learning models are classification models, meaning they draw trends by grouping like data. Each model has a different architecture and is unique from the other three:

Model Technicalities in Depth:

The most complex of the models is the Support Vector Machine (SVM). Visualizing data graphically in a "feature space," the Support Vector Machine generates margins between instances of data, shown in two dimensions on Figure 3). The soft margins extend for a fixed space on either side of the true dividing line, allowing for any visible outliers.

As this project investigates 10 waterfall features, the dividing line exists instead as a hyperplane of n–1 (9th) dimension within a feature space of the nth (10th) dimension, where n=10 is the number of features.

To account for a margin in any dimension, SVM utilizes the kernel trick, which is shown below:

$$\hat{y}(x_i) = \text{sign}\left(\sum_{j=1}^{N_s}[\hat{a}_i\, t_i\, \kappa(x_j, x_i)] + b_0\right)$$

Figure 4: Kernel Trick Decision in Support Vector Machines. The Support Vector Machine Kernel Trick creates a hyperplane boundary by optimizing for the best nth dimension feature space. The classification decision is resulted from the location of the data instance to classify relative to all margins.

The next model, Random Forest (RF), creates multiple decision trees with conditions where if the feature condition is true, the decision moves to a leaf "node" consisting of the classification. If the feature condition is false, the decision moves to the next decision "node," which operates with a different feature condition. Eventually, each tree results to a certain decision.

What makes Random Forest unique is its resolution to a classification, which is the majority output of all its decision trees. For instance, a Random Forest made up of three trees can have tree decisions end up at "Plunge," "Steep Cascade," and "Plunge," but the Random Forest will output the waterfall as of the Plunge majority decision.

The third machine learning algorithm used was Naive Bayes. Utilizing Bayes Theorem, Naive Bayes assumes (hence naïve) all classifications are possible, thus calculating the probability of waterfalls fitting within each waterfall form type, given its features:

$$P(y \mid x_1 \cap x_2 \cap \ldots x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

Figure 5: Bayes Theorem in Naive Bayes. Bayes Theorem determines waterfall form classification likelihood by the probabilities of the data instance's features occurring.

The last machine learning algorithm used was k-Nearest-Neighbors (KNN). As was the case in SVM, KNN visualizes data instances in nth (10th dimension) feature space, but instead of generating margins, finds the k nearest waterfalls in feature space to determine waterfall form classification through the Euclidean Distance Formula.

$$E(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

Figure 6: Euclidean Distance Formula. Given two waterfall data instances a and b in feature space, the Euclidean Distance Formula calculates their proximity.

Training Data Extraction:
Using Robotic Process Automation on Power Automate, a script was built to automate the waterfall data extraction flow. The script accessed webpage HTML elements on the Worldwide Waterfall Database (WWD) and USGS State Geological Compilation, public domain datasets the project was given permission to use. The end result of this process was that waterfall feature data, including classification, was transferred to usable comma delimited text. A total of 11,530 waterfalls had been logged to the dataset using this process; however, around 10,590 waterfall instances were fully collected, including duplicates and instances with little data.

```
Loop Fall from 0 to n:
     Navigate to worldwaterfalldatabase.com/
        waterfall + Fall
     Extract Data from Pages
     Write to Excel Workbook
```

Figure 7: A simple script written using abstracted tools on Power Automate to obtain waterfall feature and classification data from WWD.

Figure 8: Example Waterfall Data obtained through Robotic Process Automation (RPA):

Waterfall Identification		*h, w,* run (ft)	Class	Pitch (deg.)	Flow (ft^3/s)[a]	Drops	Source
Iguazu Falls	BRA	269 x 8800, 50	Plunge	90°	61.660 cfs	1	Headwaters
Niagara Falls	USA	167 x 3950, 50	Plunge	90°	85,000 cfs	1	Lake
Nevada Falls	USA	594 x 80, 350	Horsetail	75°	585 cfs	1	Lake
Yosemite Falls	USA	2425 x 100, 1800	Plunge	85°	307 cfs	6	Lake
Angel Falls	VEN	2648 x 350, 150	Plunge	90°	150 cfs	1	Headwaters
Strondsfossen	NOR	1245 x 100, 2200	Steep C.	70°	75 cfs	1	Glacier
Lehamite Falls	USA	1180 x 30, 1200	Steep C.	60°	5 cfs	1	Runoff
Bridal Veil Falls	USA	1291 x 75, 2080	Horsetail	75°	40 cfs	7	Lake
Fishtail Falls	USA	25 x 25, 75	Shallow C.	59°	600 cfs	2	Glacier
Little Falls	USA	6 x 240, 60	Rapids	15°	3,500 cfs	1	Headwaters
Shomyo-daki	JPN	1148 x 50, 800	Horsetail	80°	150 cfs	4	Headwaters

[a] the flow depicted above is the maximum high streamflow volume, not the average streamflow.
* not all features collected are shown above.

Preprocessing: Once acquiring the raw data, manual preprocessing began on Microsoft Excel. Incomplete and duplicate data instances were filtered out, resulting in only 941 data instances, and categorical data, such as waterfall source, was mapped to a number, such that it could be interpreted by machine learning architecture. Data Features with both number and word, such as height extracted as "13 meters" or "62 meters" from the database, were also reorganized using commands such as =LEFT(LEN(cell)–7) and converted to number from text. (The reason 7 is used above is because "meters" is a total of 7 characters, space included).

The feature engineering techniques of Normalization and Standardization were employed to preprocess the fully-numeric feature data. Additional columns were created in the Excel dataset for each feature. Each feature was transformed using the equations in Fig 9, giving additional feature data the machine learning models could train on. With this statistical process, the dataset was less impacted by outliers and data gaps, hence improving model accuracy and robustness.

Normalizing Formula: $x' = \frac{x - x_{min}}{x_{max} - x_{min}}$

Standardizing Formula: $z = \frac{x - \mu\ (average)}{\sigma\ (standard\ deviation)}$

Figure 9: Feature Engineering Methodologies: Normalization and Standardization. Normalization is a statistical transformation that compares features on the same reference scale, from 0 to 1. Standardization is a statistical transformation that compares features by z-score (number of standard deviations from the mean), as defined above.

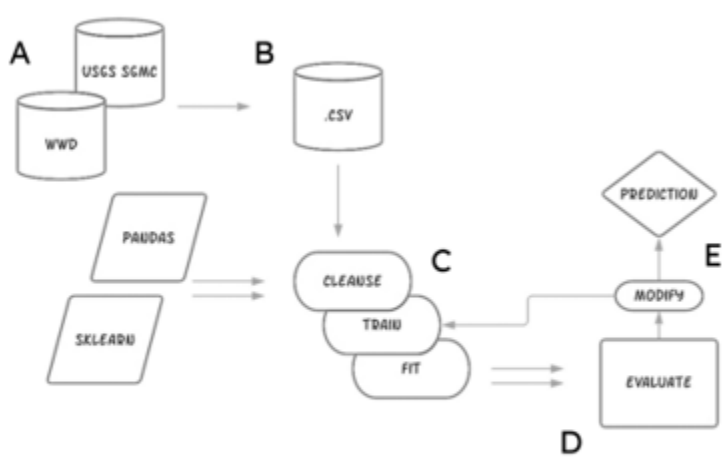


Figure 10: Machine Learning High Level Workflow. The architecture above consists of: (A) Feature Extraction, (B) Feature Engineering, (C) Model Training, (D) Model Validation and Iterative Accuracy Testing, and (E) Final Model Accuracy Testing.

Model Programming – In Depth: Using Google Colab, a browser-based .ipnyb notebook editor, the machine learning models for waterfall classification were trained. The foundations of the models were set by importing SciKitLearn, Pandas, Plotly, YellowBrick, and MatPlotLib. SciKitLearn included each of the machine learning base algorithms, and modules for preprocessing, train_test_split, cross_val, confusion_matrix, f1_ score, and roc_ auc_score.

Once attaching the preprocessed training data into Colab program files, an exploratory data analysis was conducted by visualizing the feature space through MatPlotLib. This was to take note of which features best correlated with waterfall classification, to help identify the right signal from the noise. In this process, viable features were identified by their r2 values, calculated through a least squares regression line. The significance was to validate data relevance and preemptively resolve data inaccuracies.

With Pandas, the waterfall training data was stored for the models. A partition of 60:40 train to test was utilized consistently. Before each of the 4 classifiers was fit and trained on this data, Pandas and SciKitLearn-based preprocessing began. Reading the stored feature data, SciKitLearn was utilized to verify the completed feature engineering through .StandardScaler() and .Normalizer(). This also enabled for MatPlotLib-based feature space visualization through pyplot.scatter(), which visualized the features best aligned with waterfall types to denoise the feature space. The purpose of doing so was to provide the relevant information for model testing iterations – seeing a visual representation of the data guided the iterative nature of refining the ML models. For example, certain iterations used only normalized data to classify waterfalls. If it was found from these visualizations that features correlated weakly with waterfall classification, certain standardized features were substituted to improve model robustness.

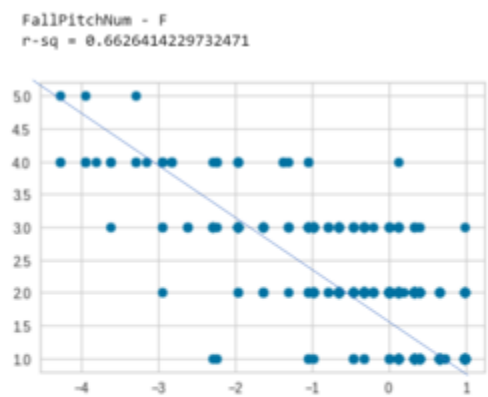
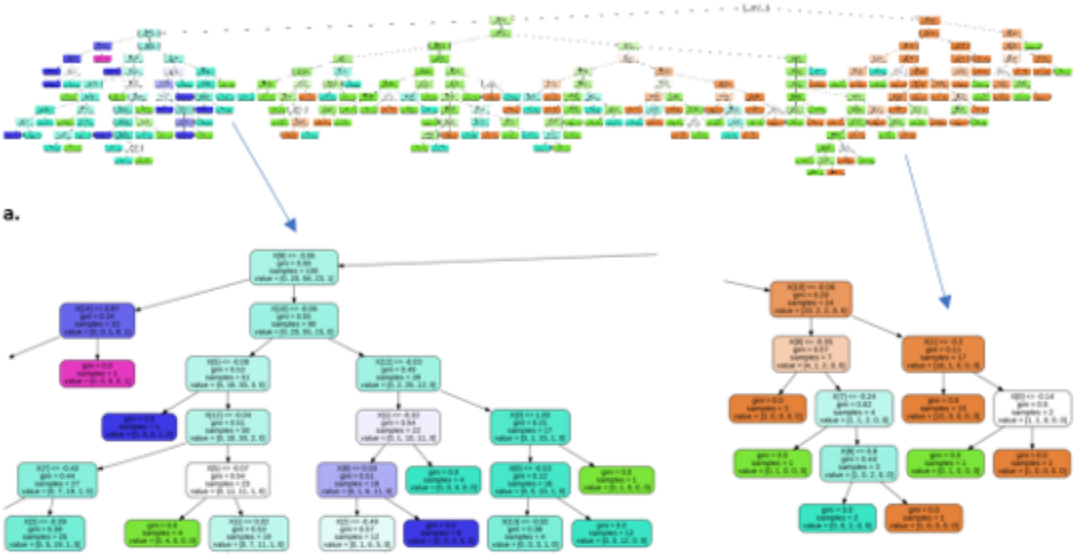
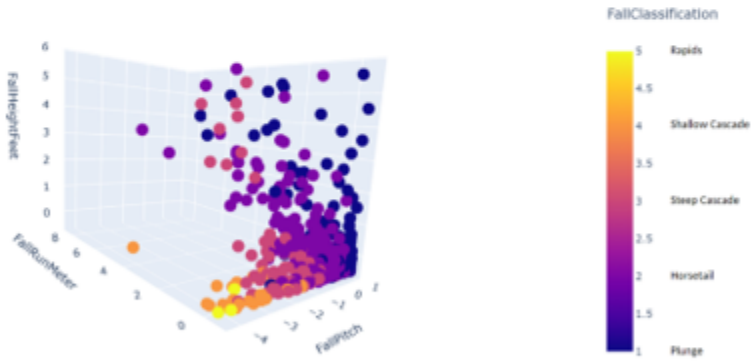


Figure 11: Explanatory Data Analysis through r^2 value. The pitch/angle of decline feature (AOD) exhibited a high r^2 value, revealing an inverse linear correlation between pitch/AOD and waterfall type. This graph shows each waterfall form as 1.0, 2.0, ..., 5.0 on the y-axis (plunge, horsetail, steep cascade, shallow cascade, rapid respectively). On the x-axis is the standardized pitch from feature engineering. The pitch yielded an r^2 value of 0.66; similarly, waterfall run yielded $r^2 = 0.58$ & height $r^2 = 0.43$, all three later determined as critical parameters to monitor for thresholds.

In the model development section was the methodology for allocating a sample of the dataset as withheld testing instances. Through SciKitLearn train_test_split(), the project saved 40 waterfalls from every 60 for validating the machine learning results. Utilizing the GaussianNB(), KNeighborsClassifier(), SVC(), and RandomForestClassifier(), models were then iteratively developed, each utilizing its unique algorithmic processes, as seen in Figures 3-6, to classify the testing instances.



a.



b.

Figures 12a, 12b: Machine Learning Process Visualization. Zoomed out to its full length, the top image showcases one of the calculated decision trees of 81 in the final `Random Forest`. Each color represents one of the 5 waterfall form predicted. (a) showcases fractions of this decision tree zoomed in; (b) visualizes the feature space (restricted to 3D with pitch, run, and height), classified by radial `SVM()` hyperplane. Both RF and SVM are shown above as they result with the highest accuracies (*see* Results).

```
[ ]  # feature engineering: using StandardScaler() to standardize features
     post_df[features] = preprocessing.StandardScaler().fit_transform(pre_df[features])


     # utilizing appropriate conventions to split data into training & testing set
     # 60:40 train:test -> 0.4 test_size
     X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.4)


     # training support vector on training data set to predict waterfall forms
     form_pred = support_predict.fit(X_train, y_train).predict(X_test)


     # finetuning machine learning models, example: random forest
     rf_grid = {
           'criterion' : ['gini', 'entropy'],
           'max_depth' : [10, 12],
           'bootstrap': [True],
           'max_features': ['auto'],
           'min_samples_leaf': [1],
           'min_samples_split': [5, 7, 9],
           'n_estimators': [81, 83, 85],
           'class_weight' : ['balanced', 'balanced_subsample']
     }
     form_pred = GridSearchCV(estimator = RandomForestClassifier(), param_grid = rf_grid)
```

Figure 13: Code Summary. This screenshot summarizes the most crucial techniques used to arrive at waterfall predictions with the most important code snippets. Shown as # comments, each line is explained at a high level.

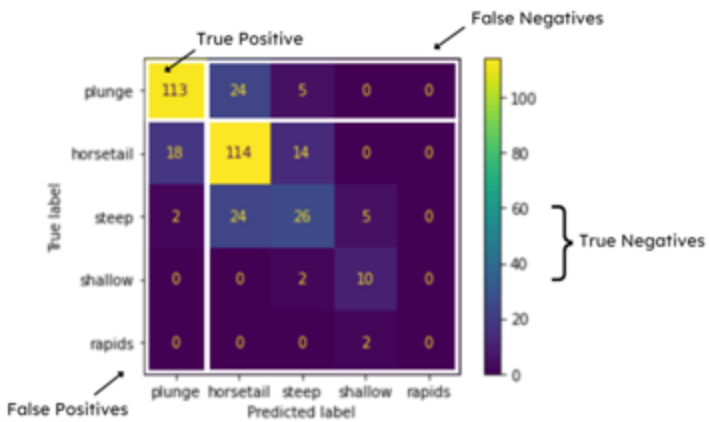


Figure 14: Iteration Confusion Matrix Example. The confusion matrix above exhibits a diagonal line, hence resembling a near-accurate machine learning model (this example iteration had an F-Measure accuracy of 73.25% (later improved, *see* Results)). Labeled are the true/false positives/negatives for the plunge classification. Each classification label has its own true/false positives/negatives that are found using the same definitions shown in the picture. This confusion matrix exhibits the diagonal line, showing high accuracy.

With classification predictions made, the project displayed model accuracy through the Confusion Matrix, F-Measure Score, and Receiver Operating Curve Score. However, before displaying accuracy, SciKitLearn cross validation through cross_val_score() was used to ensure replicability of results. Through 5 cross validations, each algorithm was retested to ensure non-outlier results. This was done due to the random nature of data allocation in train_test_split() potentially yielding biased training and/or testing data subsets.

Model Selection Programming – Accuracies:

Among the many means of obtaining accuracy, the first utilized was the Confusion Matrix (CM). The confusion_matrix(), a grid, visualized the predicted and expected classifications for each individual model. In the case of the project, the matrix was of size 5 x 5 for 5 classification types. With the CM, the true positive, false positive, true negative, and false negatives of each type were obtained, as shown in Figure 14. CMs that exhibit a diagonal line are near accurate as they show waterfall form classification's results map as expected.

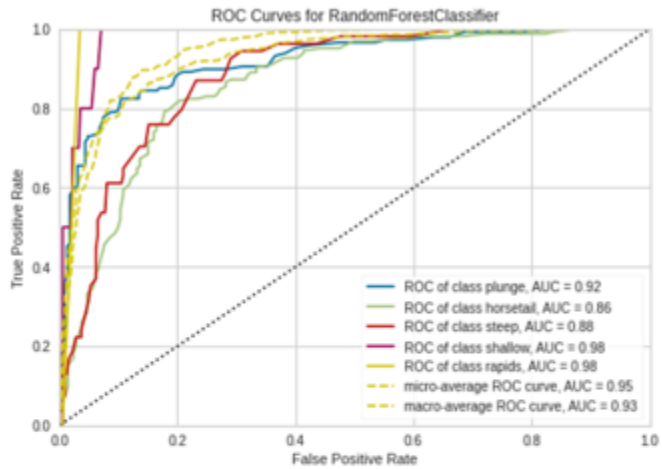
The F-Measure, also known as F1 or F-Score, was the next accuracy measure obtained. The purpose of obtaining the F-Measure was to understand accuracy in terms of classification precision and recall of each waterfall form (Figures 14, 15). The model F-Measure, f1_score(), was calculated by solving for the harmonic mean of precision and recall (Figure 15), where precision is how variable the algorithm was in its classifications and recall is how predictions compared to reality.

$$\text{F-Measure} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} = \frac{2\left(\frac{TP}{TP+FP}\right)\left(\frac{TP}{TP+FN}\right)}{\left(\frac{TP}{TP+FP}\right)+\left(\frac{TP}{TP+FN}\right)}$$

where precision: $\frac{TP}{TP+FP}$ and recall: $\frac{TP}{TP+FN}$

Figure 15: Macro F-Measure Formula. The F-Measure is the harmonic mean of precision and recall. The closer to 1 (or 100%) the F-Measure is, the more accurate the model is.

The Receiver Operator Curve was the next accuracy metric used to evaluate the machine learning models. The ROC curve visualized the False Positive Rate against the True Positive Rate, which in essence showcased the ability of the machine learning model to correctly classify each data instance over different False Positive allowance thresholds, which is different than the F-Measure. Macro ROCs closest to 1 reflected a higher ability.




```
macro roc-auc is 92.50374747837311%
<matplotlib.axes._subplots.AxesSubplot>
```

Figure 16: Example ROC Area Under Curve for Iteration of Random Forest. The ROC is the curve between the graph of TP vs. FP. The AUC of the ROC curve is the area under the curve, or the integral $\int_0^1 \text{ROC}\, dFP$. The linear dotted line is the benchmark set by random guessing: models with an AUC of 0.5 do as well as a random guess, and models above 0.5 do better than random guessing.

Iterative Testing and Model Selection: Each of the four candidate models were iteratively tested to obtain a higher accuracy. Each model was recalibrated with respect to (1) its features selected, (2) feature engineering, and (3) model parameters. Certain features were denoised (removed) from the feature space if they decreased F-Measure model accuracy. Specifically, for example, certain features were normalized over being standardized, few were standardized over being normalized, and others weren't statistically modified. All of these changes ultimately improved waterfall classification accuracy. The other incremental changes were made to the models themselves: by changing model-specific hyperparameters, such as changing the value of k in KNN to take more neighborly data instances into account or changing the depth of the trees in the random forest, accuracy was improved, and models were made more robust. The highest performing models of each algorithm were selected as the best models for waterfall classification.

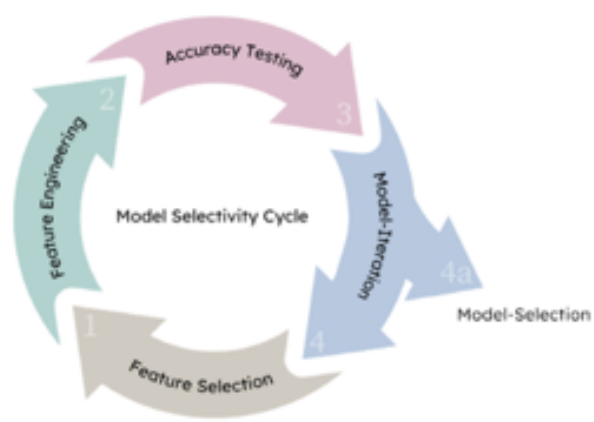


Figure 17: Iterative Accuracy Testing and Model Selectivity Cycle for candidate ML Models, inspiration from Scrum Methodologies. As each step of this cycle is passed, the candidate model is improved in terms of its accuracy.

Conclusion - Results: Table 1 exhibits the calibration of the best candidate models of each algorithm investigated and their accuracies. As described with Figure 17, the candidate models were recalibrated to achieve these high-scoring accuracies. For example, the highest performing model in terms of F-Measure was the candidate Random Forest Model. To achieve its high F-Measure of 86.73%, RF was tuned to have a maximum tree depth of 10, a forest of size 81, and more (Table 1). Likewise, the highest performing model in terms of ROC AUC was SVM. For it to attain its high ROC AUC of 94.16 ± 2.38%, SVM was tuned with an RBF kernel, meaning the Radial Basis Function was used in the Kernel Trick (Figure 4).

The results show that the F-Measure and ROC AUC for each model was an improvement from traditional classification (traditional model refers to initial models without feature engineering and finetuning techniques; these traditional models were the initial iterations). This validates the methodology used to classify waterfalls improved accuracy. Additionally, the results do show that Support Vector Machine, Naive-Bayes, and KNN had statistically significant (95% Confidence Interval) ROC AUC scores and F-Measures from said alternate/traditional models. SVM led with the highest AUC Score of 94.16 ± 2.38%, which was significant compared to the alternate's 72.13 ± 4.55%. K-Nearest Neighbors enjoyed an AUC of 90.45 ± 2.98%, which was significant compared to the alternate's 78.12 ± 4.19%. Naive-Bayes also had an AUC of 87.53 ± 3.35%, significant compared to its alternate which had an AUC of 79.31 ± 4.11%. (Figure 18).

In terms of F-Measures (Table 1), the highest belonged to Random Forest, at a max of 86.73%, with Support Vector Machine following at 84.04%, KNN at 79.78%, and NB at 76.56%.

The sheer strength of the results verified Machine Learning applicability in predicting waterfall forms. Already, with its AUC scores and maximum F-Measures, the project is shown to have met the model goals to achieve a 90%+ ROC AUC and 75%+ F-Measure across all four models investigated. With these impressive achievements in accuracy, river erosion and flood management can be better addressed.

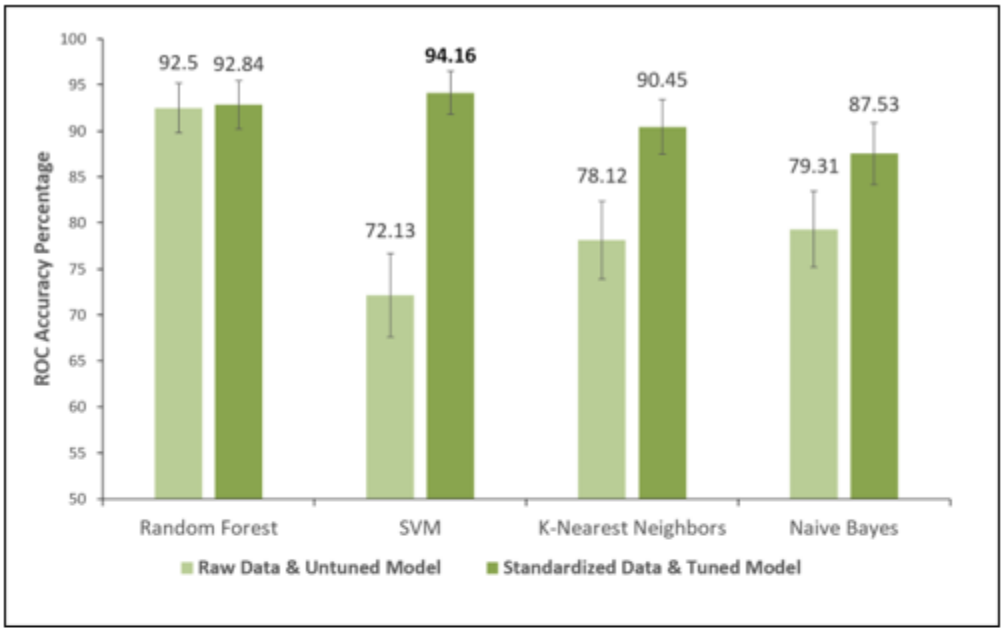
Table 1: Summarized Results of Machine Learning Models attaining best accuracies.

Models	Model-Specific Hyperparameters*	Features Necessary	F-Measure*	Max F-Measure	AUC Score*
Candidate Support Vector	kernel='rbf', + Standardization & 60:40	Num Drops and Source excluded.	75.13 ± 4.38 %	84.04 %	94.16 ± 2.38 %
Candidate Random Forest	max_depth=10 max_features='auto' min_samples_leaf=1 min_samples_split=5 n_estimators=81 class_weight='balanced' + 60:40 only	Num Drops and Source excluded.	79.36 ± 4.11 %	86.73 %	92.84 ± 2.61 %
Candidate Naive Bayes	var_smoothing=1.15 + Standardization & 60:40	Num Drops and Source excluded.	65.87 ± 4.78%	76.56 %	87.53 ± 3.35 %
Candidate k-Nearest Neighbors	n_neighbors=14 + Standardization & 60:40	Num Drops and Source excluded.	72.95 ± 4.50 %	79.78 %	90.45 ± 2.98 %

* These metrics vary due to random train-test partitioning; ± accounts for probable values within 95% confidence interval

* Hyperparameters vary per each candidate model because, as discussed, each model functions with different logic.

Table Caption: Shown above are the fine tuned hyperparameters for achieving the corresponding F-Measure and ROC AUC. The highest F-Measure comes from Random Forest, at 86.73%, and the highest Point Estimate AUC score comes from Support Vector Machine, at 94.16%. Random Forest's high F-Measure signifies its high precision and recall; Support Vector Machine's high AUC signifies its ability to accurately classify is relatively consistent, regardless of the number of false positives that may occur.



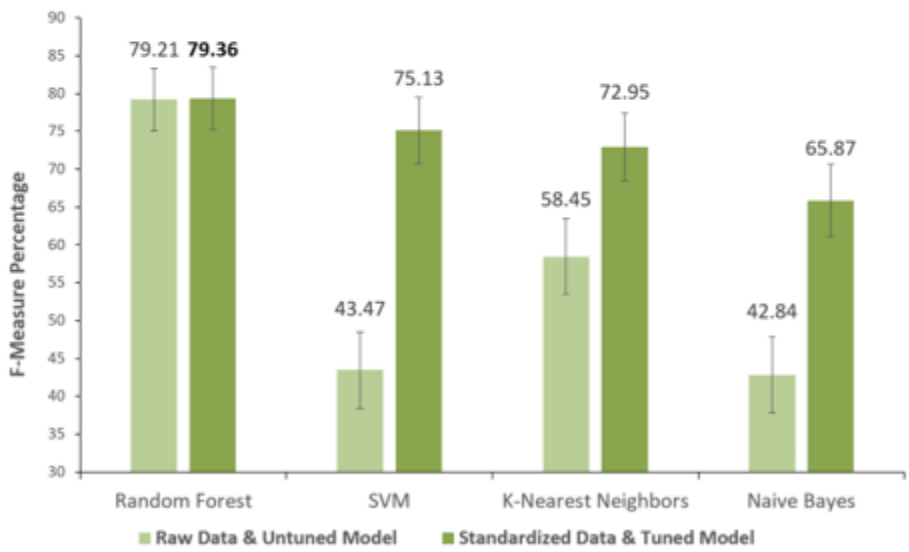
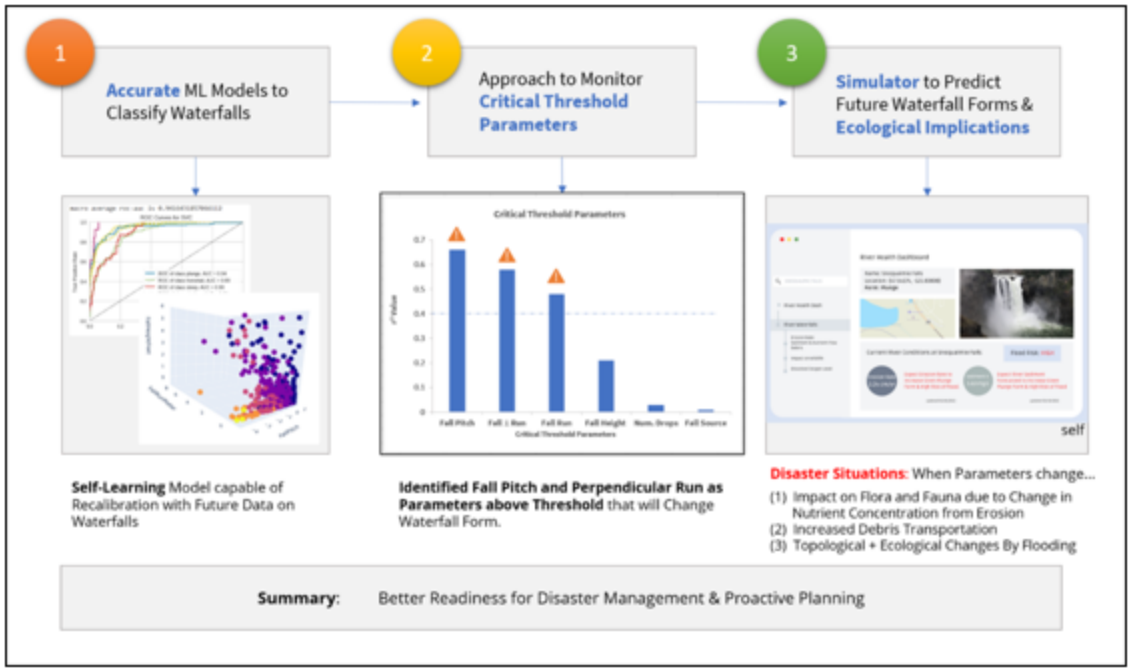


Figure 18: In-Depth Graph of Candidate Machine Learning Models' accuracies. This graph displays information from Table 1 in the context of traditional models trained on the same waterfall dataset (no feature engineering and no model hyperparameter finetuning). As seen above, the increase in F-Measure and AUC was found to be significant for SVM, KNN, and NB relative to their traditional models. In bold are the highest AUC and F-Measures. The highest F-Measure was from Random Forest at 86.73%.



Conclusion - Discussion: *(Point #1 - ML)*

Predicting waterfall forms with Machine Learning means revealing not only river topology, but also pinpointing ecological impacts of waterfalls, crucial given today's increasing floods. As floods become more extreme due to climate changes, waterfalls erode at more unprecedented rates: what this has in store for ecology depends largely on waterfall form.

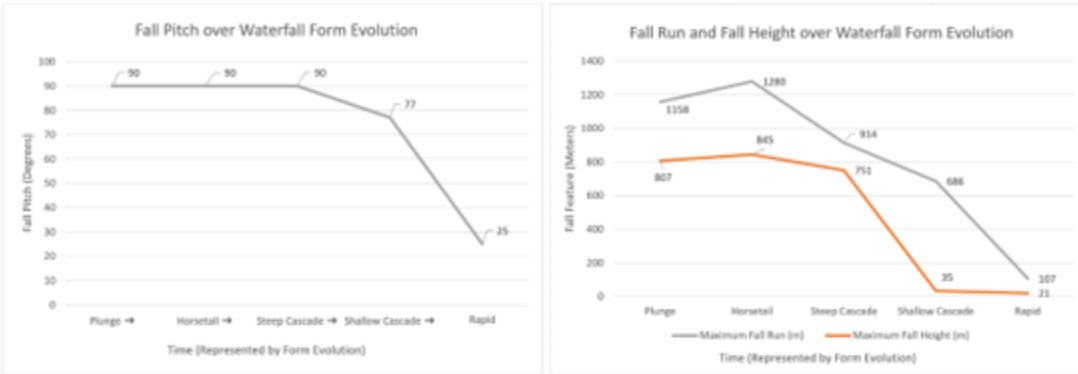
Figure 19: Snoqualmie Falls, WA, under (1) Normal Conditions and (2) Flood-Induced Stress.

For example, Snoqualmie Falls in Washington State has an average streamflow of 285 cubic feet per second. Over floods, it reports a streamflow up to 20,000 cubic feet per second, a 70-fold increase from its original river state. Due to its plunge waterfall form (with a 90-degree angle of decline), we now know its erosion rate is higher than that of a Shallow Cascade with the same streamflow.

So, what does this high erosion rate mean? This means the Snoqualmie River experiences increased debris transportation, potentially eroding unstable bedrock into riverside settlements, while also decreasing critical dissolved oxygen due to decomposition. This means the Snoqualmie River increases in heavy metal concentration, potentially poisoning fish and accumulating upon riverside vegetation. These possible geophysical disasters are limitless—due to the waterfall's form.

Over time, if floods continue to increase,

there will be a point in time where Snoqualmie Falls no longer sustains its plunge form due to erosion and evolves into a horsetail or steep cascade waterfall form, until it becomes a rapid. While this form change may take many years to occur, when it does occur it leaves the environment no time to adapt to new waterfall conditions. The time to prepare and react to the new geophysical disasters posed by the new waterfall form is, thus, severely constricted. Figure 20 showcases these novel discoveries.



Waterfall Form	Max. Pitch	Max Run	Max Height	Pitch S_x	Run S_x	Height S_x
Plunge	90°	1158 m	807 m	6.46°	183.2 m	171.6 m
Horsetail	90°	1280 m	845 m	7.70°	147.4 m	127.2 m
Steep Cascade	90°	914 m	751 m	12.93°	161.0 m	144.3 m
Shallow Cascade	77°	686 m	35 m	16.18°	118.9 m	9.23 m
Rapids	25°	107 m	21 m	7.63°	48.29 m	10.96 m

Figure 20: Novel Discoveries: Critical Waterfall Features to monitor for thresholds. As time progresses, waterfalls evolve in their lifecycles, from Plunge to Rapid. Hence, monitoring said features for changes in waterfall form can provide crucial data for topological and ecological disaster management.

This project is the start, ultimately addressing flash flood and erosion disaster management by predicting waterfall form, a necessary first step. This project makes it possible to prepare for ecological disasters with its four continuously learning machine learning models predicting waterfall forms with high accuracy in real-time, making it a modular approach that is applicable to glacial erosion, lava flows/lahars, and landslides. With this, all associated risks to topology, ecology, and civilian life are mitigated.

(Point #2 – Novel Discoveries):

This research determined three critical waterfall features to monitor:

1. Waterfall pitch angle,
2. Waterfall run,
3. Waterfall height.

From the charts above, these features were found to vary the most among the 5 waterfall form types (also shown with Fig 11). As forms evolve over time, these three features are critical in that they are extremely sensitive, completely bringing direct change to waterfall topology (and ecology).

Figure 20 also shows how waterfall forms evolve over time, from the Plunge to the Rap-

id (left to right on the x-axis). We have already witnessed evolving waterfall forms: for example, the canyon downstream Niagara Falls was once a plunge waterfall; now it has evolved into the remanent rapids visible today.

The graphs show that initially, at Plunge and Horsetail forms, waterfalls change slightly over time (this is shown with the slight increases and decreases in pitch angle, run, and height).

But then, as seen with the steep decrease in slope over time (from Steep Cascade to Rapid) for Fall Pitch, the shape of the waterfall changes drastically at this last step. Likewise with the steep decrease in slope over time (from Steep Cascade to Rapid) for Fall Run and Fall Height – the waterfall drastically changes shape. (This is also verified with the standard deviations from Figure 20.) It is this last form change that poses no time for environments to adapt to new waterfall conditions – it is this last form that brings the most extreme geophysical disasters.

(Point #3 - Simulator): Therefore, given that knowing waterfall forms can help pinpoint disasters in (1) flood conditions and (2) waterfall evolutionary lifecycle, this project ultimately brings attention to the need for flood and disaster management. This project makes it possible to prepare such disasters, with its four continuously learning, living machine learning models predicting waterfall forms with high accuracy through its real-time data pipeline. This project also demonstrates a modular approach that is applicable to glacial erosion, lava flows, lahars, and landslides, ultimately mitigating ecological, topological, and civilian risks.

With this look into machine learning models suited for waterfall classification, future research spurred from this project can now be machine-learning oriented, placing added focus on large-scale flood management and erosion/hazard mapping. Future applicability of waterfall research would then be limitless, extending beyond classification, to applications in satellite imagery/remote sensing, flood management, designing infrastructure, modeling climate change, and much more.

4. References

BaelDung CS. (2020). F-1 Score for Multi-Class Classification. BaelDung CS. https://www.baeldung.com/cs/multi-class-f1-score.

Berrar, D. (2018). Bayes' Theorem and Naïve Bayes Classifier. Encyclopedia of Bioinformatics and Computational Biology, 1(1), 403-412. https://www.researchgate.net/publication/333199040_Encyclopedia_of_Bioinformatics_and_Computational_Biology_ABC_of_Bioinformatics.

Cheng, J. (2021). Waterfall Classification. World of Waterfalls, https://www.world-of-waterfalls.com/what-types-of-waterfalls-are-there/.

Dewan, F., Zhang, L., & Chowdhury, R. (2014). Hybrid decision tree and naïve bayes classifiers for multiclass classification tasks. Expert Systems with Applications, 41(4), 1937-1945. https://doi.org/10.1016/j.eswa.2013.08.089

Garosi, Y., Sheklabadi, M., & Conoscenti, C. (2019). Assessing the performance of GIS based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Science of the Total Environment, 664(1) 1117-1132, https://doi.org/10.1016/j.scitotenv.2019.02.093.

Garreta, R., & Moncecchi, G. (2018). The Fundamentals of Machine Learning. In Packt Publishing (Eds.), Sci-Kit Learn: Machine Learning Simplified (pp. 7-16). Packt.

Goudie, A. S. (2020). Waterfalls: Forms, Distribution, Processes, and Rates of Recession. Quaestiones Geographicae, 39(1), 59-77. https://doi.org/10.2478/quageo-2020-0005.

Hakayawa, Y., & Matsukura, Y. (2010). Stability analysis of waterfall cliff face at Niagara Falls: An implication to erosional mechanism of waterfall. Engineering Geology, 116(2), 178-183. https://doi.org/10.1016/j.enggeo.2010.08.004.

Hayakawa, Y. (2005). Reexamination of a Predictive Equation of Waterfall Recession Rates in Boso Peninsula, Chiba Prefecture, Japan. Geographical Review of Japan, 78(5), 265-275. https://doi.org/10.4157/grj.78.265.

Hayakawa, Y., Matsukura, Y. (2009). Factors Influencing the Recession Rate of Niagara Falls since the 19th Century. Geomorphology, 110(4), 212-216. https://doi.org/10.1016/j.geomorph.2009.04.011.

Ramzai, J. (2020). Clearly Explained: What, Why, and How of Feature Scaling-Normalization & Standardization. Towards Data Science, https://towardsdatascience.com/clearly-explained-what-

why-and-how-of-feature-scaling-normalization-standardization-e9207042d971.

Shelef, E., Haviv, I., & Goren, L. (2018). A Potential Link Between Waterfall Recession Rate and Bedrock Channel Concavity. Journal of Geophysical Research: Earth Surface, 123(1), 905-923. https://doi.org/10.1002/2016JF004138.

StatQuest with Josh Starmer. (2020). Machine Learning Fundamentals: The confusion matrix. YouTube. https://www.youtube.com/watch?v=Kdsp6soqA7o&ab_channel=StatQuestwithJoshStarmer.

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: Applications, Variations and Vulnerabilities – A Review of Literature with Code Snippets for Implementation. Soft-Computing, 25(1), 2277-2293. https://doi.org/10.1007/s00500-020-05297-6.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical Implementations. NESUG 2010, http://www.nesug.org/Proceedings/nesug10/hl/hl07.pdf.
Algorithms and Libraries:
Pedegrosa, F., Varoquaux, G., et. al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12(85), 2825-2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

Jeff Reback, jbrockmendel, Wes McKinney, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, Simon Hawkins, gfyoung, Matthew Roeschke, Sinhrks, Adam Klein, Terji Petersen, Jeff Tratner, Chang She, William Ayd, Patrick Hoefler, Shahar Naveh, Marc Garcia, Jeremy Schendel, … Skipper Seabold. (2021). pandas-dev/pandas: Pandas 1.3.4 (v1.3.4). Zenodo. https://doi.org/10.5281/zenodo.5574486.
Plotly Technologies, Inc. (2015). Collaborative Data Science. Plotly Technologies, Inc. https://plot.ly.

Software:
Google Colaboratory. https://colab.research.google.com/.

Microsoft Power Automate. https://powerautomate.microsoft.com/.

Visual Studio Code. https://code.visualstudio.com/.
World Waterfall Database:

Swan, B., Goss, D. (2002). World waterfall database: a comprehensive, complete record of waterfalls found throughout the world [Data set]. World Waterfall Database. https://www.worldwaterfalldatabase.com.

United States Geological Survey: State Geologic Map Compilation:
Horton, J.D., San Juan, C.A., and Stoeser, D.B. (2017). The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1, August 2017): U.S. Geological Survey Data Series 1052, 46 p., https://doi.org/10.3133/ds1052.

Equation ROC and F-Measure:
Zhu, W., Zeng, N., Wang, N. (2010). Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations. North East SAS Users Group. https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf.

Equation Force-Resistance:
Hayakawa, Y., Matsukura, Y. (2003). Recession rates of waterfalls in Boso Peninsula, Japan, and a predictive equation. Earth Surface Processes and Landforms. 28. 675-684. doi:10.1002/esp.519.

All self-created diagrams were made using Adobe Illustrator and/or online flowchart tools. All other diagrams were obtained with requisite consent from their respective sources. All Data obtained from the WWD is Open Source, Public Domain. © Siddhant Porwal, 2021-2022.

Acknowledgements

This project could not have been completed without the support of (1) Ms. Kate Allender of Tesla STEM High School, who was my research advisor at school and assisted related to independent study and research; (2) Mr. Amit Sharma of DocuSign, who was my mentor and verified my methodology and provided pointers; (3) Mr. Swan of the World Waterfall Database, who generously lent access to his database and provided support for any questions; and (4) Mr. Lipovsky of Harvard University/University of Washington, who validated my research concept.

It's Flaming Out: Using Artificial Intelligence to Emulate Critical Aspects of Wildfire Growth

Druhin Bhowal, Arihant Singh

Nikola Tesla STEM High School

Abstract

Wildfires have caused devastation across the American West during the summer months, eradicating entire belts of natural plant growth and resulting in immense loss of life and property. Further, these disastrous incidents are becoming more commonplace as anthropogenic climate change raises global temperatures and decreases precipitation levels. To help address these growing environmental challenges, this project seeks to model the propagation of wildfires given factors relating to local meteorological and topographical data, with the ultimate goal of being able to project the spatial extent of a specific wildfire with a minimum F1-Score over a 24-hour period of 0.7. Requisite data for this project, such as wind speed and direction, elevation, aspect, and fuel type, were collected from nationwide datasets, along with historical fire perimeters and propagation vectors. In order to achieve the target, a novel neural network was designed and tested. This machine learning model received an initial fire perimeter with corresponding environmental characteristics, producing an estimate of the same fire's perimeter a predetermined number of days into the future. In all, the model achieved an unprecedented F1-score of 0.928 and an AUC score of 0.9730 across 2,688 withheld fire instances, reflecting a high degree of precision and recall across a diverse range of burns with a low computational runtime of 67 ms per prediction. The product of this research project, a machine learning algorithm capable of modeling changes in wildfire perimeters, satisfied design requirements for model efficiency and accuracy, and provides a new modeling tool for firefighting agencies to better plan fire containment efforts.

Keywords: AUC, Neural Network, F1-Score, Wildfire Spread, Fire Radiative Power, OPTICS

1 Introduction

Thesis: The goal of this project is to use a deep learning model to forecast the progression and propagation of a fire over its lifetime.

According to the National Interagency Fire Center (NIFC), there were nearly 60,000 wildfires in the territorial bounds of the United States in 2020 alone, together burning over 10 million acres of land and threatening over 4.5 million U.S. homes. This statis- tic, although alarming, is but a singular sample in a phenomenon which is continuing to manifest in ever greater intensity across the world. Williams et al. observed an eight-fold increase in Californian forest fire final extent (or spread) during summer months since 1972, directly correlated with the five-fold increase in annual burned area within the state during the same period. The planet's changing climate has been the primary driver behind the observed surge in wildfire occurrences, leading to widespread forest mortality, heightened levels of air pollutants in population centers, and excess carbon emissions (Abatzoglou & Williams, 2016). Consequently, it has become increasingly important for government agencies and first responders to be able to forecast the ignition and propagation of wildfires, a prediction that has, until now, required intense physical examination (i.e, on-the-ground fuel mapping) and screening of potential burn sites, followed by deterministic estimation of impact from ignition at such sites.

However, the advent of machine learning in wildfire modeling has provided new methods to forecast fire spread using pre-existing datasets

regarding various fire properties. Deep learning models such as FireCast have recently been developed to model the risk of wildfire spread in specific locations using such datasets, achieving predictive accuracy and recall scores far beyond those of traditional physics-based models (Radke et. al, 2019). When tested, FireCast was able to make predictions of high-risk fire regions up to 14 days in advance, although prediction accuracy fell to values between 6 and 8 percent after 6 days. Further studies have utilized ensemble modeling techniques, most notably gradient-boosted and random forest architecture, in order to forecast ignition and spread probabilities given sufficient biotic and abiotic ignition factors (Coffield et. al, 2019). The forecasts produced by ensemble models could also be applied in categorizing fires as small, medium, or large, mainly based on ignition characteristics and external conditions relating to burn area (e.g. wind speed, fuel type, temperature, and humidity). However, these ensemble models suffer critically from a limited scope, in that they are often trained on and built for fires originating in a certain region, and, indeed, are not practical for future forecasting as they rely on accurate meteorological observations rather than model forecasts, which would be the only available sources of meteorological data in a prediction scenario. Additionally, though past models have reached high general accuracy, it is at the cost of temporal resolution - the models are typically able only to output a coarse prediction of a fire's ultimate spread rather than its evolution throughout its life, the latter, of course, being far more invaluable to accompanying firefighting efforts.

While there are differences between existing machine learning algorithms conceived to model wildfires, all seek to extrapolate characteristics of fire behavior, most important being the fire rate of spread (RoS) over diverse terrain. To do so, it becomes necessary to analyze the factors affecting the intensity and overall behavior of a wildfire, such as: fuel loading, or the mass of dry-fuel per cubic meter of fuel bedding, heat of preignition, or the heat required to ignite a single kilogram of fuel, wind factor, which is a dimensionless constant accounting for the effect of wind on the heating of adjacent fuel particles, and aspect factor, or the direction a plot of land faces. Rothermel's surface fire spread equation, which takes into account these variables, has been employed by fire management services for the last five decades (Andrews, 2018).

$$R = \frac{I_R \xi (1 + \phi_w + \phi_s)}{p_b \in Q_{ig}}$$

Equation 1: Rothermel's Fire Spread Equation (1972), which describes the fire rate of spread (R) in meters/min in terms of the fire's propagating flux constant ($I_R\xi$) in $J/kg^2/min$, the wind factor (ϕ_w), the slope factor (ϕ_s), bulk density (p_b) in kg/m^3 and heat of preignition (Qig) in J/kg.

As Rothermel's model shows, variables impacting fire spread mainly relate to topography, vegetation, fire properties, and meteorology. While research as far back as 1993 has identified wind speed to be the most impactful factor in fire propagation, the same study has found that the presence of other natural characteristics/features, such as uncut grass or a high density of dry vegetation, can play significant roles in the extent to which fires burn (Cheney et al., 1993). These unaccounted-for factors, along with a potentially rapidly changing landscape, can significantly alter the behavior of a fire; this volatility limits the effectiveness of static, physics-based models such as Rothermel's fire spread calculation. Machine learning algorithms are built to recognize associations in data, ensuring adaptability when modeling the propagation of fires. In addition, as a result of field studies, newer models take into account a large array of environmental data sourced from satellite sensors and field estimates, in the hopes of maximizing predictive power.

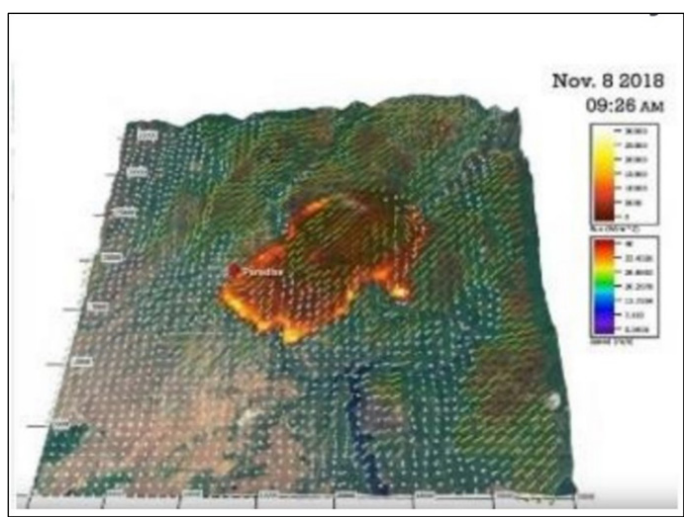


Figure 1

Figure 1: The output from a PYRECAST simulation in Central California. Such mountainous regions can either curb or exacerbate fire propagation, thus making it crucial for models to utilize topographical data in generating forecasts.

Most previous wildfire models, including those relying on physics-based equations, utilize satellite-obtained data regarding various fire properties. Landsat is a satellite used in modeling wildfires, acquiring ground imagery that can be processed to calculate the Normalized Difference Vegetation Index, a graphical indicator of the presence and density of live, green vegetation. Next, the Geostationary Operation Environmental Satellite (GOES), operated by the National Oceanic and Atmospheric Administration (NOAA), collects measurements of fire latitude and longitude, fire size (km2), fire radiative power (FRP) in Watts (W), and the fire temperature (°C). Another relevant tool used by wildfire modelers is the Visible Infrared Imaging Radiometer Suite (VIIRS), a sensor placed on the NOAA-20 weather satellite that detects wildfire "hotspots", or points of intense and potentially harmful burning. Studies done with hotspot data produced by the VIIRS satellite and its precursor, the Moderate Resolution Imaging Spectroradiometer (MODIS), involved the aggregation and mapping of such hotspots into a usable fire perimeter (Briones-Herrera et al., 2020). However, errors in perimeter generation come from omissions in hotspot detection. Both VIIRS and MODIS are sensors placed on polar-orbiting satellites that follow a polar orbit, ensuring the entire Earth revolves beneath them; as a result, every possible location is viewed a series of times in a single 12-hour period. Thus, any fire hotspots occurring at locations during times in which such locations are not viewed go undetected. A new model predicting fire perimeters, one with high (70 percent or more) accuracy for fire perimeter prediction over multiple days and with a wide range of meteorological, topographical, and initial conditions must be developed to this end.

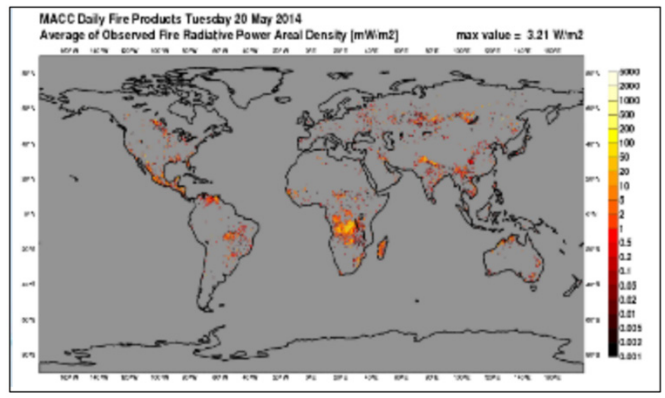


Figure 2: Fire Radiative Power (FRP) across the world as captured by the VIIRS sensor aboard the NOAA-20: the indices are normalized such that areas with fires are brighter than those without.

Because the aforementioned satellites (LANDSAT, VIIRS, and GOES-16) measure different sets of fire characteristics, it becomes necessary to merge their data products over a pre-determined time span. An outstanding issue with this procedure lies in the disparate spatial resolutions of each satellite; while the most advanced GOES-R model, GOES-16, operates at a resolution of 2 km every 5-15 minutes, Landsat operates at 30 meters, and VIIRS operates at both 375 meters and 700 meters. Thus, merging data requires extensive processing of high-resolution data to match the range covered by low- resolution data. Researchers have evaluated the effectiveness of each satellite in detecting wildfire-afflicted regions at any time, identifying fire size and temporal period as having the greatest impact on GOES detection probability (Li et al., 2020). In addition, this evaluation found that errors in fire detection, or omission (where a fire is present but is not detected), decreases across all three satellite/sensors by up to 33% when the temporal range is expanded to 8 hours. As a result, our proposed solution would have to operate over a relatively large span of time in order to use data products made after fire ignition, allowing such an algorithm to better forecast the longer-term spread of the wildfire into neighboring ecosystems.

Thus, as shown by the previous descriptions, wildfire modeling is a complex task, drawing on a variety of data sources. Each source retains certain properties, such as the resolution of collected data and format of the data upon collec-

tion. However, it is necessary to integrate these disparate features into one larger feature-set, including both dynamic wildfire characteristics (such as weather conditions during burns) and more static characteristics (such as elevation and vegetation type). Taking into account all relevant forms of environmental data allows one to produce a balanced, precise model with minimal forms of error and bias.

The primary optimizations in this project are as follows: a more robust and flexible pipeline for rapid, automated data processing, an improved model featuring specialized branches for processing and feature extraction in order to more efficiently predict inter mediate and final fire perimeters, and a display interface that maps out model results (burned vs fire-free cells) in an intuitive and usable way. Our processing pipeline is an amalgamation of three major sources of data: fire perimeter shapefiles, which are resized to proper convolution dimensions, rasterized, and overlain with an RGB (color)-based plot system to better display the fire's progression over a region; topographical data in the form of raster files, which are matched to corresponding fires and compressed; and meteorological data, which is parsed into arrays based on spatiotemporal properties and stored for later usage. Automation and efficient runtime/memory requirements not only allow the model to run on a larger number of input cells, but also ensure representative data (taken across a fire's potentially long lifetime) is utilized when predicting fire spread.

The project culminates into one successful model: a Neural Network harnessing both input tensors describing topographical and meteorological conditions as well as a basal fire perimeter, in all seeking to formulate predictions for spread. As a whole, a reliable model for wildfire spread should be developed that is not as computationally intensive as models currently in use, while also procuring a high (>70%) degree of accuracy. These measures were identified as satisfactory from a preliminary comparison with current accuracy measures of physics-based models such as FlamMap, which peak at ≈65%. (Radke et. al, 2019).

2 Methodology

Whenever Machine Learning is utilized to solve a problem, the major issue encountered is a lack of suitable data for the production of accurate results. In order to alleviate this issue, this study will expand the potential search area for fire spread data from specific regions and vegetation types to a broader scope centered in the Western United States, one of the areas most susceptible to wildfire proliferation. For instance, in the period from 2000 – 2018, the Geospatial Multi-Agency Coordination Center registered nearly 180,000 records of fires in this region (excluding the states of Alaska and Hawaii), a metric more than significant enough to produce training data for a model. These records were formerly hosted on a webserver managed by GeoMAC itself, but have since been moved to the website of the National Interagency Fire Center (https://data- nifc.opendata.arcgis.com/datasets/nifc::wfigs-current-wildland-fire-perimeters/explore), where they are available for public download in several formats. Shapefiles and Geo- JSONs, however, are the only formats which enable rendering of the geometry of any particular fire. To that end, for each unique fire recorded within GeoMAC's databases, there are certain data points of special interest – these include the date of perimeter collection, either by hand or through satellite imagery, a polygon representing the shape of the fire at the given timestamp, the lateral bounds of the fire, and the area of the fire.

All coordinates are provided in NAD83 – EPSG:4269, a geodetic coordinate system for North America. Additionally, the records for each fire have varying temporal density–for certain incidents, there is a breadth of data available for the fire's entire lifespan, sometimes multiple records a day, whereas for others, only singular datapoints are avail- able. Keeping these factors in mind, a cursory analysis of the data was conducted, after which the year 2012 was decided upon as the focus for data collection; by and large, fire incidents recorded in the year were a good mix of very large and very small fires over a variety of vegetation and terrain types, providing an ideal arrangement of data.

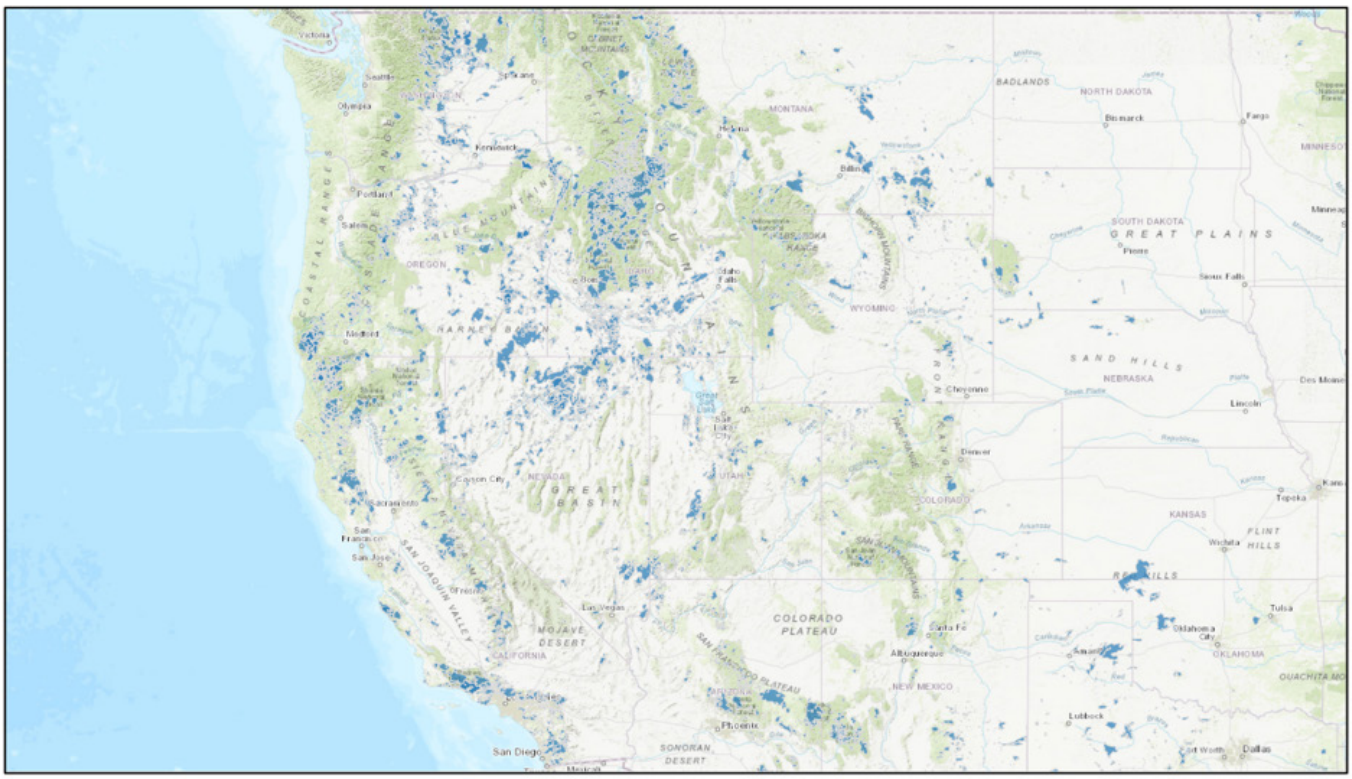


Figure 3: Shapefiles representing recorded fire incidents in the period from 2000-2018, with a notable concentration in center of the image

Several steps were taken to parse the data into the proper format. Firstly, using Python libraries to interpret shapefile data, the complete set of fire perimeters from 2012 was parsed into a data-frame containing tabular data about each individual fire record. A projection was applied onto this data using metadata contained within the shapefile itself – all coordinates were transformed from NAD83 – EPSG:4269 to NAD83 Conus Albers – EPSG:5070. This served to bypass issues with floating point calculations and, more importantly, to bring the data to the same projection system as the GeoTIFF files which were utilized to provide ground-level data – vegetation type, terrain elevation, and so forth – for each fire incident. To perform such transformations the other way around – i.e. to transform the GeoTIFF rasters to the projection system of the fire perimeters – would have been impractical, as each raster contained billions of pixels and would be far too much data to process at once; projection is not a parallelized operation, and would thus require a lump sum of volatile memory at the outset which was simply not available.

After processing the shapefiles into a dataframe, extra steps were taken to "clean" the data into an ideal format. Firstly, certain data columns with confusing or obfuscated names were renamed with more human-friendly names. Additionally, the dataframe as a whole was sorted by fire ID and secondarily by date of fire occurrence in order to arrange fire records for each fire incident temporally. Certain fires were also culled from the dataset in order to account for missing data or to ease processing load – Alaskan fires, for instance, were removed due to a lack of vegetation data, and two choice fires from Eastern Washington and Montana were dropped from the dataset due to their excessive size.

The entire set of fire perimeters was then selectively indexed for fire incidents which had at least seven records each, thereby removing fires with too few records to be of meaning to the model. For each unique fire within this new set, the maximum and minimum bounds were padded such that the length and width were divisible by 7680 (the size of a single convolution in meters), while the area was calculated and stored in a new dataframe. The fire instance was then compressed into a .pkl file and stored for later use.

Each fire perimeter was then rasterized by first plotting the shapefile, adjusting the bounds and resolution of the resulting figure to that of the actual fire, and reading in the RGB values of each pixel as a plot buffer – a binary representation of the color data in the image – which was then parsed into a matrix, with the actual fire occupying white pixels and non-fire areas occupying black pixels.

Figure 4: A plotted raster representing a given fire incident, padded and scaled between 0 and 1 to result in pixels that are either black (no fire) or white (fire).

All of the rasterized fire perimeters were then compressed and exported into individual.pkl files – if two perimeters for a single fire existed for the same day, they were combined using a logical or operator. In all, 3045 unique fires and the temporal evolutions of their perimeters were retrieved for analysis.

2a – Topographical Data

There exist several factors impacting a wildfire's spread. Among the most important of these is the presence of vegetation, as fire cannot spread without fuel. Several efforts have been made to extrapolate this – and the type of vegetation present at a given location – into a metric describing how a fire will spread at a given location. One such metric is the Scott and Burgan Fire Behavior Fuel Model, which describes the rate of spread of a fire across 40 vegetation types. This, combined with the slope elevation, aspect, and vegetation concentration provide a holistic overview of a fire's potential spread factors – while additional factors may be considered, in situations where active fires must be modeled, it is often impractical to collect such data. Therein, considering only the slope elevation, aspect, vegetation concentration, and S&B fuel model index allows one to minimize the amount of data required for successful prediction, enabling maximum flexibility of the model itself.

All data for the aforementioned factors was fetched from the LANDFIRE government website, which stores each as GeoTIFF rasters – essentially, images with each pixel representing a certain value on a scale and with additional metadata regarding the coordinate projection system utilized to generate such pixels. LANDFIRE uses EPSG:5070 as a native coordinate projection system across all rasters. The rasters are too large, however, to practically reproject; therefore, all other data are reprojected to meet LANDFIRE's standards.

In order to minimize the amount of data that must be stored, rather than matching each fire bound with its respective topographical data, the entire fire incident, with its maximum and minimum bounds, is indexed unto the topographical rasters, from where the data is compressed into a unified file for each unique fire. The set of unique fires from previous processing of fire perimeters is loaded as a dataframe, pending which a list of topographical "layers" – with layer codes and file-paths – is defined. No real further processing of the rasters is required; for each fire, the bounds must simply be fetched and the requisite portion of each raster cut out. To this end, for each unique fire within the loaded dataframe, the left bound and bottom bounds, as well as the width and height in 30 x 30-meter pixels, is fetched. For each, a new "tensor" is created, holding the ID of the fire, the "reference point" (the left-bottom coordinate), and the shape of the tensor required (in meters). As the rasters are in geoTIFF format, it is easy to input four EPSG:5070 coordinates as a mask to cut out of the raster as a whole – this is done for each of the four layers, after which they are all appended to thew aforementioned pickle and compressed into a file with a filename based on the fire ID to which they correspond.

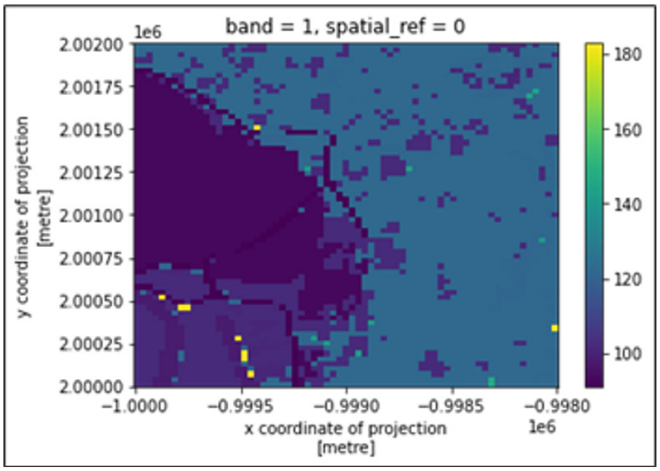


Figure 5: A plotted raster representing a section of raster data from the Scott and Burgan Fuel Loading GeoTIFF file.

2b – Meteorological Data

Meteorological data is the third tenet of fire spread. For the purposes of this study, data regarding wind-speed, wind direction, precipitation rate, and relative humidity was utilized, all measured at an altitude of 10 meters. The data itself was sourced from the National Renewable Energy Laboratory's Wind Toolkit API, wherein a data server can be remotely configured and requests to it made for wind and weather data over the Continental United States. After an API key to access the server was generated, a GET request was generated to the server and a stable connection established. The parameters of interest aforementioned were each parsed into singular arrays – these did not contain the actual data, but a series of hourly timestamps representing the data available for the given parameter. These arrays were then converted into dataframes and parsed so as to retain only data from 2012.

A program was developed for parsing coordinates in EPSG:5070 to longitude-latitude pairs, pending which a request could be sent to the NREL server. A section of data over a number of hours was collected as an array, with the x and y dimensions representing the number of "cells" within the spatial extent of the specified bounds (as the Wind Toolkit consists of data derived from a meteorological model).

For each unique fire, bounding boxes were calculated and converted into longitude/ latitude coordinates, then matched with the total temporal spread of the fire and indexed from the arrays of parameters created formerly, yielding manipulable data in a 2D array describing a meteorological phenomenon at points within the requested space. These arrays, combined with the ID of the fire, its start and end dates, its shape, and its bounds, were stored into a dictionary and compressed to a pickle file for later usage.

2c – Convolution Generation

For each fire perimeter on record, "chunks" of 256 x 256 pixels were extracted into subsections for examination. A traversing convolution was used to index over the perimeters with a step length of 32 pixels, keeping the convolutions with a perimeter and empty space present and discarding those which were entirely empty space or entirely burnt already by the fire. Each chunk was paired with its respective slice of topographical and meteorological data. This was repeated for each recorded perimeter for each unique fire - then, in order to represent the progress of each fire, new tensors were generated, containing the data for each fire as well as its perimeter at time 0, along with the perimeter of the fire at the next recorded timestamp. Topographical data was then normalized in order to make it viable to train on - the Scott- Burgan Fuel Model was vectorized to represent its categorical information in a numerical way, slope and vegetation concentration were normalized between 0 and 1, and the terrain aspect was re-formatted as a series of unit vectors. Numerical meteorological data was normalized in a similar manner, with the wind direction also being vectorized into unit vectors. Pending this, the produced tensors were pickled and stored for further use in training the model itself.

2d – Model Training

The model architecture was designed from scratch to have three sections, each with a specialized role. A structure was decided upon with one convolutional branch to extract important topographical details, a Long-Short-Term-Memory (LSTM) module to summa- rize meteorological occurrences, and a multi-layer perceptron to take

the outputs of these branches, plus the initial fire perimeter, as inputs and produce probabilities for further spread. The inputs for these, were cut from the previously generated convolutions, being, respectively, 256 x 256 x 4 tensors containing the topographical information for a given convolution, a 24 x 6 tensor for the wind data, and the concatenation of the two former branches plus the 256 x 256 raster of the fire perimeter. A sigmoid activation layer - producing outputs constrained between 0 and 1, therein representing model confidence - was applied unto the output neuron of the perceptron. This produced a new 256 x 256 raster representing the spread of a fire in a given convolution, which was then pieced together with other such convolutions for that same time interval in order to produce a finished output prediction.

A base, standard structure for each of the branches (pertaining to their archetypical form) was used as an initial state for the model. For instance, the Convolutional seg- ment consisted initially of an alternating sequence of two convolutional and dense layers resulting in a final output. Pooling layers were later added with experimentation, with a standard downsampling procedure (to the final filter of 32) used in order to extract more general characteristics of topography. The LSTM module was altered to summarize meteorological data into batches of 24 hours, producing finally six outputs representing the characteristics of the six meteorological factors evaluated in the model. Additionally, the number of hours (axis 0 of the input shape for the network) was left unspecified, therein enabling the processing of any number of hours of meteorological data. The perceptron was altered into a funnel-like structure, therein distilling the occurrences due to meteorological and topographical data into meaningful iterations upon the primary fire perimeter. Different optimizers and activation functions were tested, with TanH activation attempted on the final output layer in order to constrain the output more closely towards 0 and 1 - predictions of maximum and minimum confidence - with little success. Kernel-based relu activation (based on a paper by Scardapane et. al) interleaved between dense and convolutional layers and was also attempted successfully - model training time was reduced significantly, and the f-score was raised by 4.7 percent. Dropout percentages were changed as needed, mostly upwards, in order to prevent overfitting and data memorization.

Because the total number of fire samples was too large to fit into computer memory, a data-streaming solution was developed and fitted to the model. An instance of the sequence basal object was created and modified in order to produce such a solution - instructions for the fetching of a singular batch of items and the parsing / deletion of such data pending training were written, and a batch-size of 64 samples was decided upon, therein resulting in a total quantity of 21760 samples per epoch. This solution not only circumvented the issue of having to fit data into memory, but also cut model training time to a mere 49s / epoch, enabling rapid iteration.

The model itself was trained on a train-test-validation split of 80-10-10%, over 32 epochs, with a mean-squared-logarithmic loss function, along with a modified ADAM optimizer with an initial learning rate of 0.001 tied to a polynomial decay function over 10,000 steps to result in an eventual learning rate of 0.00005. As the fire propagation maps are themselves composed of 1s and 0s, the loss function simply sums them and compares the summed value - these values, when squared, can get very large, and are compounded over batches, therein making it necessary to affix them to a logarithmic scale. The ADAM optimizer is ideal for situations in which compute time is limited, or there is an excess of data present and an approximate result must be achieved quickly - this is the case here, for there was approximately 100 gigabytes worth of convolutional data to consider.

```
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
topographyInput (InputLayer)    [(None, 256, 256, 4  0           []
                                )]

conv2d_2 (Conv2D)               (None, 256, 256, 32  6304        ['topographyInput[0][0]']
                                )

activation_8 (Activation)       (None, 256, 256, 32  0           ['conv2d_2[0][0]']
                                )

max_pooling2d_2 (MaxPooling2D)  (None, 64, 64, 32)   0           ['activation_8[0][0]']

conv2d_3 (Conv2D)               (None, 64, 64, 64)   18496       ['max_pooling2d_2[0][0]']

activation_9 (Activation)       (None, 64, 64, 64)   0           ['conv2d_3[0][0]']

max_pooling2d_3 (MaxPooling2D)  (None, 32, 32, 64)   0           ['activation_9[0][0]']

meteorologyInput (InputLayer)   [(None, 24, 6)]      0           []

flatten_1 (Flatten)             (None, 65536)        0           ['max_pooling2d_3[0][0]']

lstm_1 (LSTM)                   (None, 24)           2976        ['meteorologyInput[0][0]']

dropout_4 (Dropout)             (None, 65536)        0           ['flatten_1[0][0]']

dense_7 (Dense)                 (None, 6)            150         ['lstm_1[0][0]']

dense_6 (Dense)                 (None, 1024)         67109888    ['dropout_4[0][0]']

firePerimeter (InputLayer)      [(None, 65536)]      0           []

activation_11 (Activation)      (None, 6)            0           ['dense_7[0][0]']

activation_10 (Activation)      (None, 1024)         0           ['dense_6[0][0]']

concatenate_1 (Concatenate)     (None, 66566)        0           ['firePerimeter[0][0]',
                                                                  'activation_11[0][0]',
                                                                  'activation_10[0][0]']

dense_8 (Dense)                 (None, 1024)         68164608    ['concatenate_1[0][0]']

activation_12 (Activation)      (None, 1024)         0           ['dense_8[0][0]']

dropout_5 (Dropout)             (None, 1024)         0           ['activation_12[0][0]']

dense_9 (Dense)                 (None, 512)          524800      ['dropout_5[0][0]']

activation_13 (Activation)      (None, 512)          0           ['dense_9[0][0]']

dropout_6 (Dropout)             (None, 512)          0           ['activation_13[0][0]']

dense_10 (Dense)                (None, 256)          131328      ['dropout_6[0][0]']

activation_14 (Activation)      (None, 256)          0           ['dense_10[0][0]']

dropout_7 (Dropout)             (None, 256)          0           ['activation_14[0][0]']

dense_11 (Dense)                (None, 65536)        16842752    ['dropout_7[0][0]']

activation_15 (Activation)      (None, 65536)        0           ['dense_11[0][0]']

==================================================================================================
Total params: 152,801,302
Trainable params: 152,801,302
Non-trainable params: 0
```

Figure 6: A diagram representing the architecture of the model created to generate perimeter spread predictions.

2e – Model Evaluation

General statistical tests were used to determine the final accuracy of a model with respect to testing data. In specific, the mean squared logarithmic error, precision / recall score, F1 Score, and AUC (area under receiver operating characteristic curve) will be used:

$$-\,loss = \frac{1}{n}\sum_{i=1}^{n}(\log(y_i + 1) - \log(\hat{y}_i + 1))$$

wherein n is the number of cells, M is the number of unique classes (2), y_i is the actual wnumber of cells on fire, and y^i is the predicted number of cells on fire.

$$-\,precision = \frac{\text{\# of cells classified correctly as on fire}}{\text{\# of cells classified as on fire}} = \frac{TP}{TP + FP}$$

wherein the precision is taken as a proportion of correctly identified positives out of the total number of positive predictions (note that TP is true positive and FP is false positive).

$$-\,recall = \frac{\text{\# of cells classified correctly as on fire}}{\text{\# of total cells actually on fire}} = \frac{TP}{TP + FN}$$

wherein the recall is taken as a proportion of correctly identified positives out of the total number of cells actually on fire (note that TP is true positive and FN is false negative).

$$-f_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

wherein a more representative metric of overall model accuracy is determined as a score from 0 to 1 (the higher the better)

These metrics were registered with the Keras backend and thereafter noted as evaluation functions in the registration and activation of the model, allowing them to be output alongside the standard accuracy and loss measures that Keras naturally provides.

2f – Perimeter Generation Technique

Because high-resolution, hand-drawn perimeters may not always be available for on- going fires, preliminary work into a method allowing for live predictions of the perimeters of existent fires was developed. To achieve this, the infrared ABI spectral channel of the GOES-R satellite (at mid-wave IR bands of around 4 µm) was analyzed for Fire Radia- tive Power (FRP) readings - with this metric, "burn points" within any fire's expanse can be detected due to their individual thermal output. Individual FRP points can be amalgamated into a single perimeter by bounding such points in a concave hull, although only points detected in the same period (i.e, within the same 6–12 hour period) and in the same general area can be subject to such clustering.

To begin clustering, a dataset containing fire instances detected by the GOES-R satellite in September 2019 was obtained. This dataset contained more than 149,000 such fire instances; however, many fire instances contained null values, that is, the value -9999 for quantitative measurements such as FRP (in Watts) and temperature (◦C). As FRP is the primary focus of this program, any instances with null FRP values were filtered out. Further, the resultant FRP points were plotted on a map of the United States to gauge the distribution of recorded values and confirm that most burn points occurred in the arid Western region.

Even after preliminary processing (removing rows with null values), approximately 30,000 valid fire instances remained. To more efficiently initialize and tune a clustering algorithm, the first

4,000 such instances were stored in a coordinate tuple to ensure simultaneous processing of all data points. A preliminary OPTICS clustering algorithm was applied to the coordinate tuple. OPTICS identifies density-based clusters in data with spatiotemporal variation; indeed, the distribution of radiative power hotspots has variation in spatiotemporal density, and thus an OPTICS algorithm can more efficiently detect meaningful clusters of FRP "hotspots" in a given region.

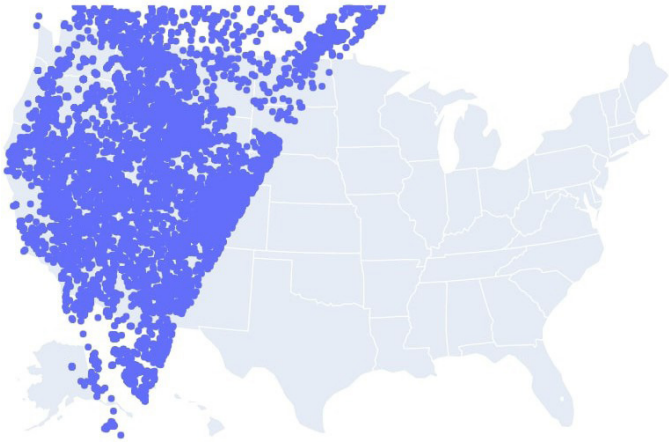


Figure 7: A distribution of FRP hotspots detected by the GOES-R satellites in September 2019

Once initialized, the OPTICS clustering algorithm was fit to the coordinates of the subset of all FRP points, with hyperparameters min samples, which controls the minimum number of samples required to consider a point for a cluster, and max eps (epsilon), which controls the maximum dis tance between two data points such that one can be considered adequately proximal (or in the local "neighborhood") to the other, set to standard values of 15 and 0.2 respectively. In order to refine the clustering algorithm, OPTICS's hyperparameters were tuned to maximize the Calinski-Harabasz index, a measure of the density and separation between clusters, where greater internal density and external separation reflect a higher index. First, a series of epsilon values were tested, from 0.2 to 0.5, where for each an OPTICS algorithm was generated and applied to the coordinate set with the given epsilon value. The epsilon value resulting in the highest Calinski-Harabasz index was stored (as other hyperparameters were kept constant). In addition, a Randomized Search was performed using Sklearn's RandomizedSearchCV class; this technique allows for a range of values to be tested for hyperparametrical optimization. For the OPTICS clustering algorithm, the minimum samples parameter was given any random value between 5 and 25, while the epsilon parameter was given any value from 0.1 to 0.9. Due to constraints with the RandomizedSearchCV class, the metric completeness score was used as a measure of accuracy, as it describes the level at which data points clustered together are indeed part of the same class or section. Because Randomized Search is a more cost-effective/less computationally expensive technique, it randomly samples values for a specified set of parameters, resulting in the identification of an optimal combination of hyperparameters. The optimal epsilon-samples combination was output after continuous running of both the iteration approach and the randomized search assessment.

To conclude the clustering sequence, the values of the hyperparameters for a finalized OPTICS clustering algorithm were set to those identified as optimal by the aforementioned methods. The finalized OPTICS algorithm was then run across all 30,000 valid FRP hotspots to generate a color-coded map of FRP clusters across the continental United States. Only those clusters with over ten internal hotspots were visualized, as such clusters are more likely to represent a single fire and not a disparate set of minor burns. 635 clusters were generated for the 30,000 data points, with an average of 22 points per cluster. Note that not all hotspots were accounted for in the 635 clusters; many fire hotspots were isolated and could not be contained in any such cluster.

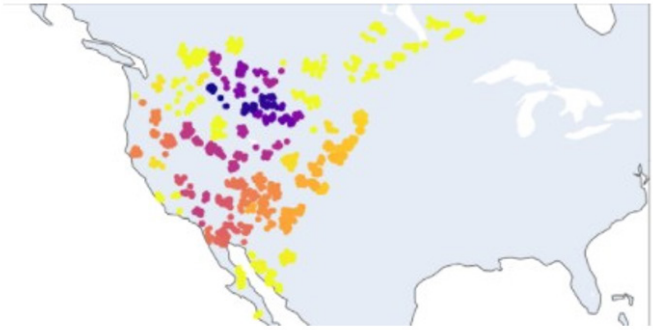


Figure 8: Color coded FRP hotspots generated by the OPTICS algorithm

The next step in the process was iterating through our hotspot clusters and bounding them to actually create the approximate fire perimeters. The concave hull, or bounding perimeter, was set up with an alpha shape object, which essentially surrounded every point in the cluster. These alpha shapes could be converted into fire perimeters given the previously described processing pipeline, enabling the model to harness fire perimeters generated with real-time satellite data at a high temporal resolution. Further evaluation of the accuracy of this method against recorded perimeters is required before assessing its true practicality; however, this method holds promise for a fast, efficient way to supplement the data available through the NIFC database and scale the model itself on a practical level for nationwide applicability.

3. Conclusions

3a – Results

Our model was empirically validated across a large subset of with-held data completely unseen by the model. To verify the general applicability of the model, such data was taken from fires across the United States over a wide variety of biological, topographical, and meteorological conditions. Predictions on this data over 2730 samples yielded an unprecedented F1-Score of 0.92, a loss value of 0.0295, recall value of 0.9262, precision value of 0.925, and A.U.C. score of 0.973. Additionally, pending visual analysis of several specific predictions, the model proved adept at predicting instances where fires spread, stayed stagnant, and recessed, a significant improvement over traditional physical models, such as FIRETEC and FlamMAP, that assume a basal constant of spread, essentially propagating a given fire in perpetuity. Furthermore, because of the spatial and temporal resolution of the data perused in the model - slices of 7680 by 7680 meters with times in between input and output fire perimeters of up to 168 hours - this model is highly capable at understanding emergent properties in fire spread. The model generated predictions for the entire dataset in only 564 milliseconds.

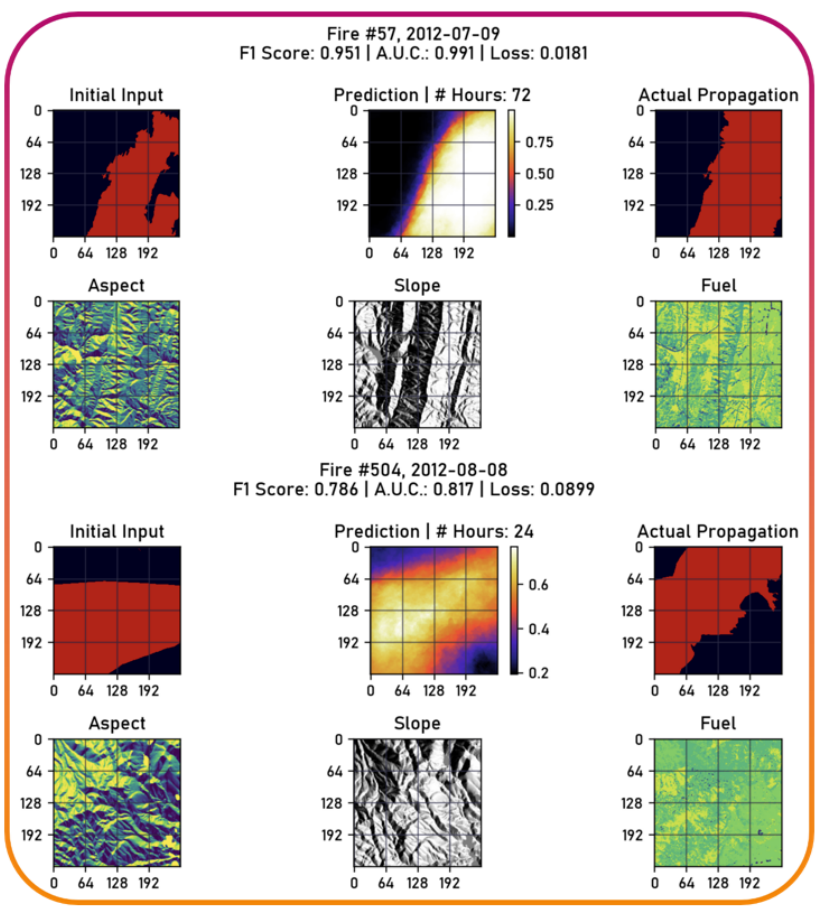


Figure 9: A sample projection generated by the model, together with the supplementary topographical data afforded to it.

Next, our clustering algorithm generated 635 clusters of fire radiative power (FRP) "hotspots", with the purpose of detecting fires in real-time through satellite observations. Our final OPTICS algorithm achieved a metric completeness score of nearly 0.81, displaying a relatively high level of homogeneity and goodness of split (a characteristic more so for unsupervised structures). It demonstrated the ability to differentiate between areas of a fire perimeter to include in a cluster, those surrounding actual hot spots, and extraneous regions that had a low probability of burning at this time.

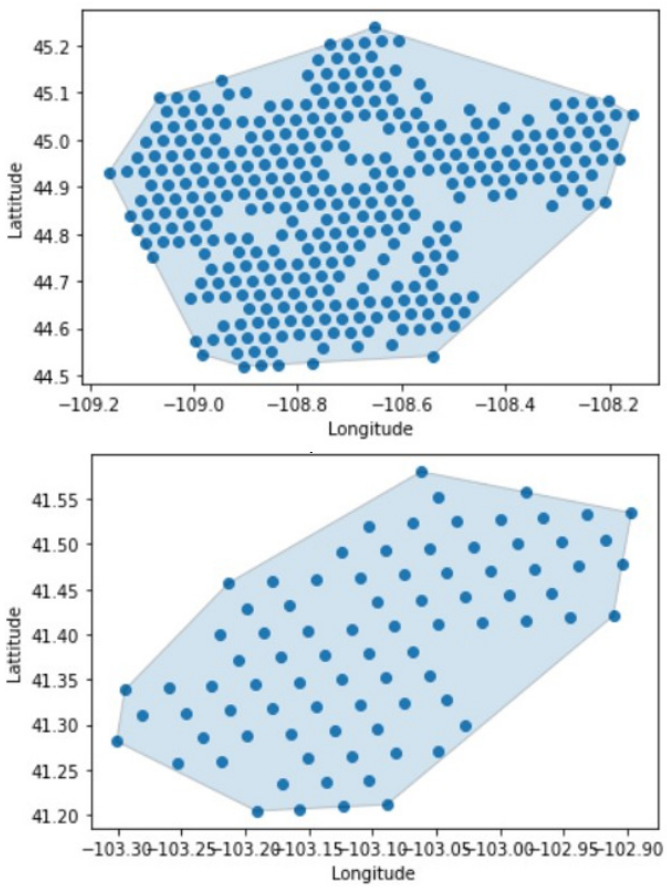


Figure 10: Above are two representative example fire perimeters generated by the OP- TICS algorithm - the top example is that of a more expansive fire, evident from the larger latitude and longitude scale present on the graph, while the bottom example is from a smaller fire, with a correspondingly lower number of internal FRP hotspots. Note that the GOES satellite often detects hotspots in linear "belts".

3b - Discussion and Future Work

Current fire spread models provide statistically significant prediction accuracies only up to a few hours or a day at maximum, and rapidly falter afterwards. Further, they require significant computational power to function, with FIRETEC, despite being run on a supercomputer at the Los Alamos laboratory, running slower than real time, rendering it cumbersome to use in practical applications. A model such as ours will enable firefighting agencies to make more methodical, planned decisions regarding efforts to combat the spread of a wildfire with confidence in its ability and accuracy, completely revolutionizing the process of fighting and controlling wildfires.

Future iterations of the model might focus on its application on a practical basis, moving away from the technical demonstration that it is currently. Possible routes for improvement include the development of an ensemble modeling system, in which slight modifications are made to the model's parameters, predictions are generated, and then averaged for a final result. Additionally, the complete automation of the model's data processing pipeline is necessary for a system that actively processes and predicts potential spread for ongoing wildfires - coupled with live satellite imagery and a perpetual bank of static information - topographical information, for instance - supplemented by local vegetation and meteorological forecasts current to the fire instance can create a robust, efficient, automatic system for fire spread projection. With the implementation of the perimeter clustering algorithm previously mentioned, this automation can be achieved, and with radiative power metrics acquired from such an algorithm, even more accurate projections, accounting for the strength of the fire at certain points, can be made.

The potential of linking this model with a meteorological model to provide, for instance, smoke or atmospheric pollution forecasts is also significant. This has been attempted in the past with the unification of FIRETEC and HIGRAD, but can be made much more practical with the greater accuracy this model offers. The fire spread model can provide accurate estimates of the amount of vegetation burnt, which can then be used as a parameter in the smoke forecasts of advanced meteorological models like HYSPLIT. As large fires generate their own weather, such a model could be used to supplement the predictions of the spread model, therein taking into account the very dynamic nature of wildfires.

To provide a complete end-to-end solution

for wildfire spread prediction, secondary models for a fire's initial spread could be utilized. Currently, our model performs with middling accuracy when the initial perimeter is too small - a model built on the Asynchronous- Advantage-Actor-Critic architecture could be used to circumvent this issue. Such a model would treat each individual pixel on fire as an "agent", working to maximize its collective "reward", or the delta in spatial spread between an input and output fire. This approach would likely result in high accuracy at the outset of a fire, after which the neural network would take over to provide more long-scale and long-range predictions.

References

Abatzoglou, J. T., amp; Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. Proceedings of the National Academy of Sci- ences, 113(42), 11770–11775. https://doi.org/10.1073/pnas.1607171113

Andrews, P. L., Cruz, M. G., Rothermel, R. C. (2013). Examination of the wind speed limit function in the Rothermel surface fire spread model. International Journal of Wildland Fire, 22(7), 959. https://doi.org/10.1071/wf12122

Ballı, C., Acar, M., Caglar, F., Tan, E., Onol, B., Karan, H., Unal, Y. S. (2012). The Figure of Merit in Space (FMS) and Probability Analyses of the Concentrations for Forecasted Transport of Particles using the WRF and HYSPLIT Models over Istanbul for January and July, 2009. NASA ADS, 11910. https://ui.adsabs.harvard.edu/abs/2012EGUGA..1411910B/abstract.

Briones-Herrera, C. I., Vega-Nieva, D. J., Monjar´as-Vega, N. A., Brisen˜o-Reyes, J., L´opez-Serrano, P. M., Corral-Rivas, J. J., Alvarado-Celestino, E., Arellano-P´erez, S., A´lvarez-Gonz´alez, J. G., Ruiz-Gonz´alez, A. D., Jolly, W. M., Parks, S. A. (2020). Near Real-Time Automated Early Mapping of the Perimeter of Large Forest Fires from the Aggregation of VIIRS and MODIS Active Fires in Mexico. Remote Sensing, 12(12), 2061. https://doi.org/10.3390/rs12122061

Cheney, N., Gould, J., Catchpole, W. (1993). The Influence of Fuel, Weather and Fire Shape Variables on Fire-Spread in Grasslands. International Journal of Wildland Fire, 3(1), 31. https://doi.org/10.1071/wf9930031

Coffield, S. R., Graff, C. A., Chen, Y., Smyth, P., Foufoula-Georgiou, E., Randerson, J. T. (2019). Machine learning to predict final fire size at the time of ignition. International Journal of Wildland Fire, 28(11), 861. https://doi.org/10.1071/wf19023.

Hodges, J. L., Lattimer, B. Y. (2019). Wildland Fire Spread Modeling Using Convolu- tional Neural Networks. Fire Technology, 55. https://doi.org/10.1007/s10694-019-00846-4

Jiang, W., Wang, F., Fang, L., Zheng, X., Qiao, X., Li, Z., Meng, Q. (2021). Modelling of wildland-urban interface fire spread with the heterogeneous cellular automata model. Environmental Modelling Software, 135, 104895. https://doi.org/10.1016/j.envsoft.2020.104895

Li, F., Zhang, X., Kondragunta, S., Schmidt, C. C., Holmes, C. D. (2020). A preliminary evaluation of GOES-16 active fire product using Landsat-8 and VIIRS active fire data, and ground-based prescribed fire records. Remote Sensing of Environment, 237, 111600. https://doi.org/10.1016/j.rse.2019.111600

Li, Y., Tong, D. Q., Ngan, F., Cohen, M. D., Stein, A. F., Kondragunta, S., Zhang, X., Ichoku, C., Hyer, E. J., Kahn, R. A. (2020, July 28). Ensemble PM2.5 forecasting during the 2018 campfire event using the HYSPLIT transport and dispersion model. AGU Journals. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JD032768.

Liang, H., Zhang, M., Wang, H. (2019). A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors. IEEE Access, 7, 176746–176755. https://doi.org/10.1109/ACCESS.2019.2957837

Linn, Rodman, Reisner, John, Colman, J. Jonah, Winterkamp, Judith (2002). Studying wildfire behavior using FIRETEC. International Journal of Wildland Fire, 11, 233 - 246. Retrieved December 5th, 2021, from https://www.frames.gov/catalog/4584

Long, X., Tie, X., Cao, J., Huang, R., Feng, T., Li, N., Zhao, S., Tian, J., Li, G., Zhang, Q. (2016, August 2). Impact of crop field burning and mountains on heavy haze in the North China plain: A case study. Atmospheric Chemistry and physics. https://acp.copernicus.org/articles/16/9675/2016/

Maillard, O., Vides-Almonacid, R., Flores-Valencia, M., Coronado, R., Vogt, P., Vicente- Serrano,

S. M., Azurduy, H., An´ıvarro, R., Cuellar, R. L. (2020). Relationship of For- est Cover Fragmentation and Drought with the Occurrence of Forest Fires in the De- partment of Santa Cruz, Bolivia. Forests, 11(9), 910. https://doi.org/10.3390/f11090910

McCarthy, N. F., Tohidi, A., Aziz, Y., Dennie, M., Valero, M. M., Hu, N. (2021). A Deep Learning Approach to Downscale Geostationary Satellite Imagery for Decision Sup- port in High Impact Wildfires. Forests, 12(3), 294. https://doi.org/10.3390/f12030294.

Morvan, D., Dupuy, J. L. (2004). Modeling the propagation of a wildfire through a Mediterranean shrub using a multiphase formulation. Combustion and Flame, 138(3), 199–210. https://doi.org/10.1016/j.combustflame.2004.05.001

Rios, O., Valero, M. M., Pastor, E., Planas, E. (2019). A Data-Driven Fire Spread Simulator: Validation in Vall-llobrega's Fire. Frontiers in Mechanical Engineering, 5. https://doi.org/10.3389/fmech.2019.00008

Pais, C., Miranda, A., Carrasco, J., amp; Shen, Z.-J. M. (2021, July 7). Deep fire topol- ogy: Understanding the role of landscape spatial patterns in wildfire occurrence using artificial intelligence. Environmental Modelling amp; Software. Retrieved October 26, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S1364815221001651.

Pan, L., Kim, H. C., Lee, P., Saylor, R., Tang, Y. H., Tong, D., Baker, B., Kondragunta, S., Xu, C., Ruminski, M. G., Chen, W., Mcqueen, J., / Stajner, I. (2020, May 7). Evaluating a fire smoke simulation algorithm in the National air Qual- ity forecast Capability (NAQFC) by using Multiple observation data sets during the southeast Nexus (SENEX) field campaign. Geoscientific Model Development. https://gmd.copernicus.org/articles/13/2169/2020/.

Partanen, T. M., Sofiev, M. (2021). Forecasting the regional fire radiative power for regularly ignited vegetation fires. Natural Hazards and Earth System Sciences Dis- cussions, 1–18. https://doi.org/10.5194/nhess-2021-262

Patricia, L. Andrews (2018). The Rothermel surface fire spread model and associated developments: A comprehensive explanation. General Tech Report (GTR), U.S Forest Service, 371. https://www.srs.fs.usda.gov/pubs/55928
Radke, D. (n.d.). FireCast: Leveraging deep learning to predict ... - IJCAI. Retrieved October 26, 2021, from https://www.ijcai.org/Proceedings/2019/0636.pdf

Sayad, Y. O., Mousannif, H., Moatassime, H. A. (2019, January 24). Predictive modeling of wildfires: A new dataset and machine learning approach. Fire Safety Journal. https://www.sciencedirect.com/science/article/abs/pii/S0379711218303941.

Simone Scardapane, Steven Van Vaerenbergh, Simone Totaro, Aurelio Uncini, Kafnets: Kernel-based non-parametric activation functions for neural networks, Neural Net- works, Volume 110, 2019, Pages 19-32, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2018.11.002.

Valero, M. M., Jofre, L., amp; Torres, R. (2021, April 13). Multifidelity prediction in wildfire spread simulation: Modeling, uncertainty quantification and sensitivity analysis. Environmental Modelling amp; Software. Retrieved October 26, 2021, from https://www.sciencedirect.com/science/article/pii/S1364815221000931.

Van Wagtendonk, J. W., Root, R. R. (2003). The use of multi-temporal Landsat Normalized Difference Vegetation Index (NDVI) data for mapping fuel models in Yosemite National Park, USA. International Journal of Remote Sensing, 24(8), 1639–1651. https://doi.org/10.1080/01431160210144679.

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., Lettenmaier, D. P. (2019). Observed Impacts of Anthropogenic Climate Change on Wildfire in California. Earth's Future, 7(8), 892–910. https://doi.org/10.1029/2019ef001210

Zhai, C., Zhang, S., Cao, Z., amp; Wang, X. (2020, February 28). Learning-based prediction of wildfire spread with real-time rate of spread measurement. Combustion and Flame. Retrieved October 26, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S001021802030064X.

Entry to the Stockholm Junior Water Prize 2022

Engineering a Surface Flow Constructed Wetland to Evaluate Efficiency for Combating Water Scarcity

Monisha Krothapalli

Nikola Tesla STEM High School

Abstract

A shortage of drinking water has been a problem since far before the 21st century, especially in developing countries that couldn't afford to treat their polluted water bodies, but what if there was a way to fix this problem? This project looks at the efficiency of a cost-effective constructed wetland for treating polluted water bodies, mainly those limited to agricultural runoff, one of the most prevalent forms of freshwater pollution. Using a basic surface flow wetland as a base, a new constructed wetland design was created with features such as a clarifier and tube settlers. The prototype was tested using manually polluted water that was run through the system and measured for a total of seven parameters (pH, temperature, nitrates, nitrites, carbonate hardness, general hardness, and phosphates) before and after treatment to calculate the water filtration efficiency. On average the calculated water quality index (WQI) of the water after undergoing treatment improved by 6.66%. The most prominent pollutants from agricultural runoff (phosphates, nitrates, and nitrites) showed individual improvements of 36.7%, 51.85%, and 58.89% respectively, showing high potential for this model. The information collected from this study can be used to support the instillation of constructed wetlands in areas most effected by agricultural pollution such as Africa, China, and Europe.

Introduction

This project seeks to pose a solution to the problem of water scarcity by investigating the efficiency of low-cost constructed wetlands for filtering polluted water from agricultural runoff. Worldwide, approximately 1.1 billion people don't have reliable access to clean water and 2.7 billion find water to be scarce one month each year (World Wildlife Fund). The largest sources of freshwater (i.e. rivers, lakes, and aquifers) are becoming increasingly polluted and unsuitable for consumption and usage. Climate change has only worsened this ever-growing problem of water scarcity with the increase of extreme precipitation events such as floods and droughts. One of the main causes of this worldwide crisis is water degradation, of which agricultural activities are a primary contributor (UNESCO). Agricultural runoff carries pollutants such as fertilizers, pesticides, and animal feces which are abundant in pathogens and excess nutrients. This mixture of constituents pollutes water, making it unusable and reducing the resource's availability. The modern water treatment technologies in place, such as municipal wastewater treatment plants, are expensive and energy intensive. This makes them very difficult to implement in poorer areas that cannot afford to manage or maintain the financial burden. An alternative to large-scale wastewater treatment would be the implementation of constructed wetlands, a technology that artificially mimics the processes of real wetlands and purifies water at a much lower cost. For example, the Tres Rios pilot project in Arizona cost $3.5 million to construct; conversely, repairing the 91st Avenue treatment plant would have cost $625 million (University of Arizona), showing how much more cost-effective constructive wetlands are in comparison. However, constructed wetlands are not without their drawbacks.

This project addresses issues and limitations previous models could not. One study, which sought to investigate more efficient techniques of wetland construction, exposed the large

land usage that constructed wetlands typically require [1], which is much greater in comparison to conventional treatment plants. Additionally, another concern of constructed wetlands are their low pollutant extraction rates. A study completed in 2016 showed that the median removal rate of total nitrogen and total phosphorus in constructed wetlands was only 37% and 46% respectively [2], which is on the lower side and pales in comparison to municipal treatment. To help address these issues, this project contains a clarifier which integrates the usage of tube settlers and limestone filters. The unique design features of this project improve the efficiency of constructed wetlands for purifying water polluted by agricultural runoff. Efficiency will be determined by calculating the percent increase in the water quality index (WQI) from water prior to treatment compared to post treatment. If proven successful, this research can be used to further support the implementation of this technology in underdeveloped areas to combat water scarcity.

Model Improvements and Description

A typical constructed wetland contains a wetland cell, which is a manmade cell implemented within a water body that contains a bush of emergent plants (ex. Cattails, bulrushes, and sedges) through which water flows. As water flows through this system, the plants slow down the water flow and take up the nutrients through the plant roots, removing them from the water. Between the two main types of wetland cells - subsurface and surface flow cells – a surface flow cell was used. Water travels vertically through subsurface flow cells, causing issues with clogging and management, while it travels horizontally through surface flow, causing issues with large land usage. This prototype contains two types of emergent vegetation - common three-square bulrush and soft stem bulrush were planted.

This prototype builds upon the basic design of a constructed wetland with the novel implementation of a clarifier. The clarifier contains tube settlers which are a series of plates/channels angled at 60 degrees to form a structure that can be observed in Figure 1. As water enters the clarifier and moves upwards through the tube settlers, it forms a counter current flow. As the water moves up, the sediments become heavy and fall downwards toward the sloped edges of the settlers. They then accumulate there to form a sludge which will eventually slide down the device toward the sludge deposit. These devices greatly help increase settling efficiency which can in turn decrease land usage. As proven by Allen Hazen, an engineer, settling efficiency is proportional to the difference in sedimentation velocity and the surface loading rate, which is equal to the water flow divided by tank area. The shape and orientation of the tube settlers greatly increase the total surface area of the tank, thereby increasing tank area and allowing the settling efficiency to increase. By increasing settling efficiency in the clarifier, we are able to decrease the amount of settling area that would be needed in the wetland cell and cut down on the issue of land usage.

The clarifier also contains a limestone filter. Limestone, when mixed with water at an acidic level (a weak acid), dissociates to form calcium and carbonate ions. Limestone also helps greatly increase the pH of the water, bringing it to a slightly basic value. Phosphates, which are one of the most common pollutants in agricultural wastewater, bind to the free calcium ions when the water is at a high enough pH and form an insoluble compound. This compound is then easily settled out, addressing the issue of low nutrient removal.

Materials and Methodology

I. Prototype Construction

The prototype involves the assembly of two elevated plastic bins to form a gravity based flow. One basin acts as a clarifier and the second as a nutrient extraction wetland cell. The clarifier is constructed with a feeder that pushes water directly to the tube settlers. Attached to the mouth of the feeder is a filter with a limestone coating. Twenty-eight small ¼” PVC pipes were then organized in an alternating pattern to create the shape for the tube settler (Fig. 2). The settlers were angled at a 60-degree angle in the center the first basin to optimize space usage (Fig. 1). Plastic walls were then placed on all sides of the feeder settler system to ensure that the water would flow in the desired direction.

For the second basin, the bottom was lined with about two inches of soil to ensure enough

room for plant roots. Ten common three-square bulrush and ten soft-stem bulrush plants were planted in the soil. The basin was then filled with freshwater three inches above its current level to form a surface flow wetland. Against the outlet of the second basin, a gravel filter was made using pebbled pressed against the basin wall.

The first basin was then placed on an elevated surface (e.g., a ladder) and the second basin was placed slightly lower on a different elevated surface. Using PVC pipes, the basins were connected with one leading from the first basin to the second basin, another from the second basin to the collection basin (Fig. 1).

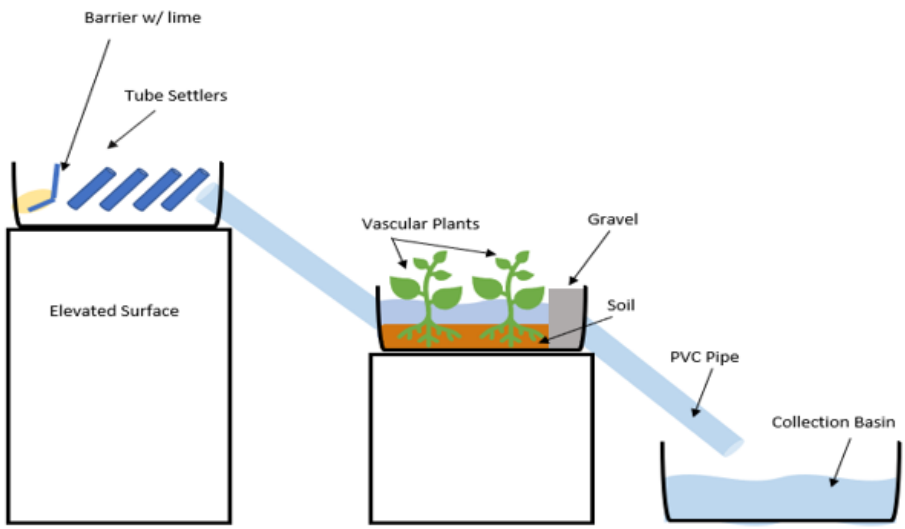


Figure 1. Diagram labels the design of the

$$nse = \left(\frac{\sum excursion}{total\ \#\ of\ tests}\right) \quad excursion = \left(\frac{failed\ test\ value}{guideline\ value}\right) - 1 \quad F_3 = \left(\frac{nse}{0.01nse + 0.01}\right)$$

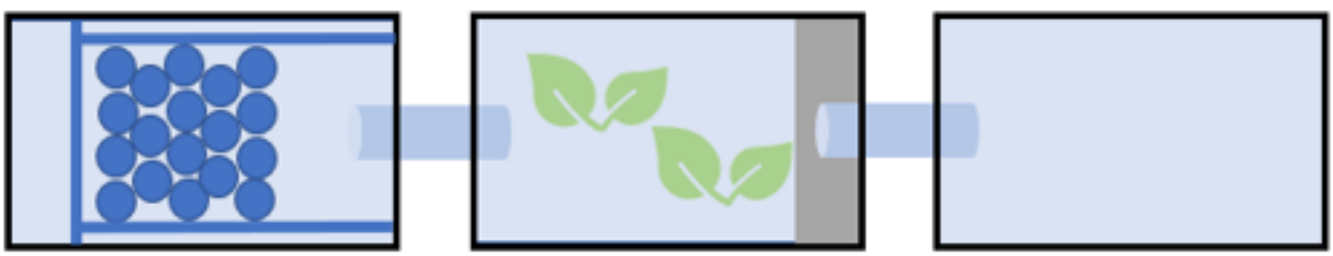


Figure 2. Top view of constructed wetland. Tube settlers angled at 60º to settle out suspended solids and decrease the amount of settling area needed.

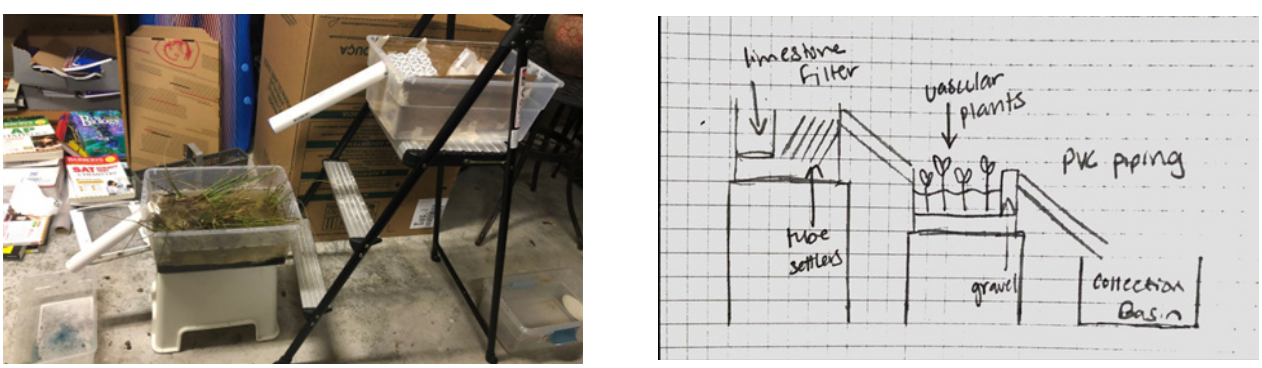


Figure 3. To the left is the constructed prototype and to the right is the sketched diagram.

II. Prototype Testing

To assess the prototype, polluted water was run through the system simulating how the constructed wetland would work in a real body of water. The water was manually polluted with a mixture of fertilizers and pesticides, components that most often are found in agricultural pollution. Prior to treatment, the water sample was tested for seven markers: pH, nitrates, nitrites, phosphates, general hardness, carbonate hardness, and temperature (Table 2). The pH was measured using Litmus strips and temperature with a thermometer. The rest of the markers were measured using commercially available API kits. After the initial water quality testing, the water was poured through the limestone filter at the top of the prototype and the filtered water was collected from the collection basin and the same tests run prior to the treatment were run on the treated water. This procedure was repeated for a total of nine trials.

III. Data Analysis

After the parameters were collected for the pretreatment and post treatment water, a water quality analysis was performed to calculate the water quality index (WQI). The Canadian WQI equation is used to find the WQI of the samples. The overall WQI relies on three calculated factors:

$$WQI = \left(100 - \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}\right)$$

F1 is the scope of the index or the percentage of met parameters.

$$F_1 = \left(\frac{\#failed\ paramters}{Total\ \#\ of\ parameters}\right) * 100$$

F2 is the frequency or the number of individual tests within each parameter that failed to meet the standard

$$F_2 = \left(\frac{\#failed\ tests}{Total\ \#\ of\ tests}\right) * 100$$

F3 is amplitude or the extent to which the failed tests exceed guidelines, which are taken from standard EPA values for freshwater bodies, and this is calculated in several steps as labeled below.

$$nse = \left(\frac{\sum excursion}{total\ \#\ of\ tests}\right) \quad excursion = \left(\frac{failed\ test\ value}{guideline\ value}\right) - 1 \quad F_3 = \left(\frac{nse}{0.01nse + 0.01}\right)$$

The overall rating is determined by which category the final WQI falls in (Table. 1).

Designation	Index value	Description
Excellent	95-100	All measurements are within objectives virtually all of the time
Good	80-94	Conditions rarely depart from natural or desirable levels
Fair	65-79	Conditions sometimes depart from natural or desirable levels
Marginal	45-64	Conditions often depart from natural or desirable levels
Poor	0-44	Conditions usually depart from natural or desirable levels

Table 1. Table outlines the index standards for each water quality level

Results

Raw data for the seven different parameters was collected using a variety of commercially available kits.

	Water Quality Parameters vs. Point of Sampling							
Trial #		pH	Nitrites (mg/L)	Nitrates (mg/L)	Phosphates (mg/L)	General Hardness (ppm)	Temperature (°C)	Carbonate Hardness (ppm)
1	Initial	6.5	40	1	10	120	18.3	80
	Final	7	20	0.5	8	60	14	80
2	Initial	6.5	40	1	10	120	18.3	80
	Final	7	20	0.5	7	30	13.9	60
3	Initial	6.5	40	1	10	120	18.3	80
	Final	6.5	10	0	5	120	14.2	60
4	Initial	6	30	1	8	60	15.9	100
	Final	6.5	10	0.3	7	40	12.8	60
5	Initial	6	30	1	8	60	15.9	100
	Final	6.5	15	0.2	6	50	13	50
6	Initial	6	30	1	8	60	15.9	100
	Final	7	15	0.2	5	60	12.9	80
7	Initial	6	20	0.5	9	140	14.2	40
	Final	6.5	15	0.1	4	120	12.7	60
8	Initial	6	20	0.5	9	140	14.2	40
	Final	6.5	10	0.2	5	80	12.9	30
9	Initial	6	20	0.5	9	140	14.2	40
	Final	6.5	10	0.3	4	80	12.8	30

Table 2. Outlines the raw data parameters that were collected.

The raw data of the seven parameters was then processed through the equation described within the data analysis section of the methodology and the values below were produced for the nine trials. The mean was also calculated from these values.

Trial	WQI
1	48.07
2	50.27
3	48.74
4	48.86
5	51.77
6	51.22
7	53.09
8	48.34
9	49.63
Average	49.99

Table 3. Calculated water quality index post treatment of all water samples.

From the seven parameters, phosphates, nitrates, and nitrites were deemed the most essential for monitoring as they're the most prevalent pollutants within agricultural wastewater. The individual parameter removal efficiency can be clearly viewed in the graph below.

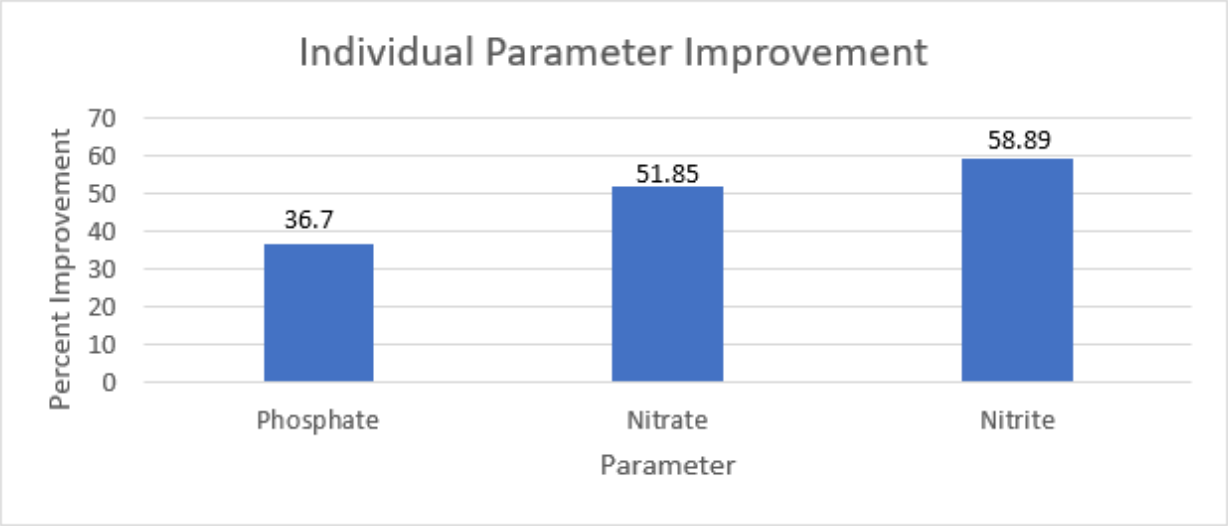


Figure 4. Improvement of most prominent individual parameters.

Additionally, qualitative data from trial four was taken to show the improvement in turbidity and reduction in suspended solids, for which numerical data was not taken.

Prior to Treatment Post Treatment

Qualitative Data

Figure 5. Qualitative data showing difference in turbidity and TSS post and prior to treatment.

Discussion

With the average WQI being 46.87 prior to treatment and 49.99 post, the average improvement in quality was 6.66% overall. While this number is not as high as expected, the individual parameter improvements showed much higher potentials. Both nitrate and nitrite removal levels were high enough that they surpassed the average levels of nitrogen removal for current constructed wetlands outlined in the introduction (Figure 3). The standard deviation of the WQI was 1.722 across all nine trials, a value low enough that we can say with confidence that the nitrogen removal will continue to perform consistently at that removal level. Additionally, the qualitative data demonstrates the high levels of turbidity reduction and total suspended solid removal through the treatment, demonstrating how well the prototype is able to handle high levels of dirt and sediments, another common pollutant in agricultural pollution which can have harmful effects on water bodies.

These results show the potential that this model has to effectively treat agricultural pollution. Being able to effectively treat polluted water allows us to keep water from being unusable for human activities and alleviates the stress of water scarcity. The cost-effectiveness of this solution allows this prototype to be implemented in countries that are commonly plagued by these issues but cannot afford the tools to manage it. Even within our own country, this technology can be implemented within more rural areas where individual farmers operate. Being able to maintain a clean water supply that can be consumed and reused provides the government with an incentive to implement this prototype in such rural areas where larger treatment plants wouldn't be feasible.

Conclusions

1. The prototype was able to achieve nitrate removal levels of 51.85%, nitrite removal levels of 58.89%, and phosphate removal levels of 36.7%. Nitrogen removal levels exceeded the average total nitrogen removal levels of current constructed wetlands.
2. Prototype has high levels of turbidity reduction and suspended solid removal and can handle large amounts of sediments and particles.
3. General trends in sediment removal demonstrate the effectiveness of tube settlers and increase of settling area for filtration.

There are some limitations to the prototype and research conducted. The high flow rate of the device decreased the amount of contact time the water had within each basin and limited nutrient uptake, especially in the wetland cell. The scaled down model also limited settling area and decreased the overall nutrient removal that the limestone filter aided in.

This project can be improved further through the improvement of the phosphorus filter, which was the main combatant for phosphates. By experimenting more with different types of lime, such as hydrated lime, and also adjusting the ratio of lime to water, the effects of the design can be improved. Additionally, by reducing the gradient between the compartments and using gated channels between each section, both the flow rate and retention time will be drastically improved, increasing pollutant removal.

Through some further research and improvement, full scale versions of this model can be produced to be utilized in areas such as Africa, China, and Europe whose water supplies suffer greatly due to agricultural activities. Additionally, because of the pollutants this model targets, it can also be used to prevent algal blooms and dead zones within freshwater bodies, and can also be used to treat water contaminated by sewage.

References

1. Ilyas, H., & Masih, I. (2017). Intensification of constructed wetlands for land area reduction: A Review. Environmental Science and Pollution Research, 24(13), 12081–12091. https://doi.org/10.1007/s11356-017-8740-z
2. Land, M., Granéli, W., Grimvall, A., Hoffmann, C. C., Mitsch, W. J., Tonderski, K. S., & Verhoeven, J. T. (2016). How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? A systematic review. Environmental Evidence, 5(1). https://doi.org/10.1186/s13750-016-0060-0
3. Vasconcelos, J. G., Perez, M., Wang, J., Zech, W. C., & Fang, X. (2017). Evaluation of high-rate settling technology for sediment control in roadway construction sites. Highway Research Center.

Bibliography

Abbasnia, A., Alimohammadi, M., Mahvi, A. H.,

Nabizadeh, R., Yousefi, M., Mohammadi, A. A., Pasalari, H., & Mirzabeigi, M. (2018). Assessment of groundwater quality and evaluation of scaling and corrosiveness potential of drinking water samples in villages of Chabahr City, Sistan and Baluchistan Province in Iran. Data in Brief, 16, 182–192. https://doi.org/10.1016/j.dib.2017.11.003

A Handbook of Constructed Wetlands. (n.d.). Retrieved November 16, 2021, from https://19january2021snapshot.epa.gov/sites/static/files/2015-10/documents/constructed-wetlands-handbook.pdf.

Analyzing a small-scale, constructed wetland for ... (n.d.). Retrieved November 16, 2021, from https://scholar.rose-hulman.edu/cgi/viewcontent.cgi?article=1027&context=undergrad_research_pubs.

Boano, F., Caruso, A., Costamagna, E., Ridolfi, L., Fiore, S., Demichelis, F., Galvão, A., Pisoeiro, J., Rizzo, A., & Masi, F. (2020). A review of nature-based solutions for greywater treatment: Applications, hydraulic design, and environmental benefits. Science of The Total Environment, 711, 134731. https://doi.org/10.1016/j.scitotenv.2019.134731

Environmental Protection Agency. (2000, October). GUIDING PRINCIPLES FOR CONSTRUCTED TREATMENT WETLANDS: Providing for Water Quality and Wildlife Habitat. EPA. Retrieved November 16,2021,from https://nepis.epa.gov/EPA/html/DLwait.htm?url=%2FExe%2FZyPDF.cgi%2F2000ZXNC.PDF%3FDockey

Hydraulic characteristics of ... - wolkersdorfer.info. (n.d.). Retrieved November 16, 2021, from https://www.wolkersdorfer.info/publication/bht/wachniew_wm24.pdf.

Hrozencik, A. (2021, August 27). Irrigation & Water Use. Retrieved August 30, 2021, from https://www.ers.usda.gov/topics/farm-practices-management/irrigation-water-use/

It's Official: Eating Food Watered with (Treated) Grey Water Is Fine. (2018, October 10). Retrieved from https://modernfarmer.com/2015/12/grey-water-study/

Li, X., Wu, S., Yang, C., & Zeng, G. (2020). Microalgal and duckweed based constructed wetlands for swine wastewater treatment: A Review. Bioresource Technology, 318, 123858. https://doi.org/10.1016/j.biortech.2020.123858

Natural materials successfully filter greywater: Society for Science. (2019, December 20). Retrieved from https://www.societyforscience.org/blog/natural-materials-successfully-filter-greywater/

Rosario Klautau Araujo Gomes, R. (1970, January 1). Greywater treatment to drinking water : Challenges and technologies. Brage NMBU. Retrieved November 16, 2021, from https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/2686153.

Sediment forebays. megamanual.geosyntec.com. (n.d.). Retrieved November 16, 2021, from https://megamanual.geosyntec.com/npsmanual/sediment-forebays.aspx.

Sheikh, B., Nelson, K. L., Haddad, B., & Thebo, A. (2018). Grey Water: Agricultural Use of Reclaimed Water in California. Journal of Contemporary Water Research & Education, 165(1), 28-41. doi:10.1111/j.1936-704x.2018.03291.x

Schmitz, K.-U. (2020, August 26). Tube settler design basics - solid settling with lamella clarifiers. Smart Water Magazine. Retrieved March 14, 2022, from https://smartwatermagazine.com/blogs/karl-uwe-schmitz/tube-settler-design-basics-solid-settling-lamella-clarifiers

Soong, H. N., Omar, R., Goh, K. C., & Seow, T. W. (n.d.). The challenges of Implementation Greywater Recycling System in residential buildings. Research in Management of Technology and Business. Retrieved November 16, 2021, from https://publisher.uthm.edu.my/periodicals/index.php/rmtb/article/view/2047.

Tang, S., Liao, Y., Xu, Y., Dang, Z., Zhu, X., & Ji, G. (2020). Microbial coupling mechanisms of nitrogen removal in constructed wetlands: A Review. Bioresource Technology, 314, 123759. https://doi.org/10.1016/j.biortech.2020.123759

Vliet, M. T., Jones, E. R., Flörke, M., Franssen, W. H., Hanasaki, N., Wada, Y., & Yearsley, J. R. (2021). Global water scarcity including surface water quality and expansions of clean water technologies. Environmental Research Letters, 16(2), 024020. doi:10.1088/1748-9326/abbfc3 https://iopscience.iop.org/article/10.1088/1748-9326/abbfc3/pdf

Water Scarcity. (n.d.). Retrieved from https://www.worldwildlife.org/threats/water-scarcity

Water treatment solutions. Lenntech Water treatment & purification. (n.d.). Retrieved March 14, 2022, from https://www.lenntech.com/phosphorous-removal.htm

Watson, J. T., & Hobson, J. A. (2020). Hydraulic design considerations and control structures for constructed wetlands for wastewater treatment. Constructed Wetlands for Wastewater Treatment, 379–391. https://doi.org/10.1201/9781003069850-35

Cellular Biology/ Chemistry

Creating a Structural Model of a saSrtA-Inhibitor Complex to Predict Efficiency of Class A Sortase Enzyme Inhibitor Molecules

Tanvi Sapar and Shivani Godse

Nikola Tesla STEM High School

Abstract

Every year, approximately a thousand new individuals are diagnosed with cases of cystic fibrosis that have the potential to be fatal. A key cause of fatal cystic fibrosis cases are chronic lung infections caused by Staphylococcus aureus biofilm formation. The inhibition of class A sortase enzymes will prevent the formation of molecules that lead to the accumulation of S. aureus biofilm, ultimately preventing biofilm formation. This project aims to create a structural model of the inhibition of Class A sortase enzymes by sortase A inhibitor molecules in Staphylococcus aureus bacteria (saSrtA-inhibitor complex). This model will then be used to observe the binding of S. aureus sortase inhibitors on other sortase A bacteria that have the potential to cause diseases. By exploring methods of disease prevention in different types of sortase A bacteria, diseases such as pneumonia and scarlet fever can be prevented. Six sortase A inhibitors were docked onto four sortase A bacteria through the HADDOCK software. The 24 docking runs were then visualized through the PyMOL software. Inhibitor-bacteria binding affinity was determined based on buried surface area values. The inhibitor, 3,3,3-trifluoro-1-(phenylsulfonyl)-1-propene (TPP), bound to Actinomyces oris SrtA (bacteria) resulted in a buried surface area of 451.9 Å. TPP (inhibitor) bound to Listeria monocytogenes SrtA (bacteria) had a buried surface area of 394.6 Å. Modeling a saSrtA-inhibitor complex and testing binding strength between sortase A bacteria and sortase A inhibitors can provide more insight into preventing various sortase A bacteria-related diseases, one of them being cystic fibrosis.

Introduction

This project aims to engineer a structural model of a saSrtA-inhibitor complex in order to predict the efficiency of class A sortase enzyme inhibitors at binding to the sortase A enzyme that is present in various sortase A bacteria. This can ultimately prevent a wide variety of infections/diseases such as lung infections in cystic fibrosis, meningitis, scarlet fever, pneumonia, and more.

The current widespread prevalence of antibiotic-resistance bacteria prevents antibiotics from being the most effective treatment for certain infections. In the US alone, antibiotic-resistant bacterial infections have caused approximately 99,000 deaths annually (Aslam, 2018). Currently, one of the most researched antibiotic-resistant bacterial infections are chronic lung infections that worsen cases of cystic fibrosis. These lung infections are caused by the Staphylococcus aureus bacteria, which is one of the most researched antibiotic-resistance bacteria. S. aureus was found to be the most common pathogen that was present in the airways of cystic fibrosis patients that dealt with respiratory issues (Esposito, 2019). The lung infections caused by S. aureus bacteria are due to a certain enzyme that is present in the S. aureus bacteria, the sortase A enzyme. Ultimately, the sortase A enzyme is the root cause of the lung infections caused by S. aureus bacteria, eventually leading to worsened cases of cystic fibrosis.

However, this sortase A enzyme is not only present in the S. aureus bacteria, rather it is present in a wide variety of bacteria called "Class A sortase bacteria" or "sortase A bacteria". Some examples of sortase A bacteria include Staphylococcus aureus, Streptococcus agalactiae, Streptococcus pyo-

genes, Listeria monocytogenes, Bacillus anthracis, and Actinomyces oris. The sortase A enzymes that are present in these bacteria can lead to various infections. For example, the Streptococcus pyogenes bacteria can cause scarlet fever, which can later worsen to rheumatic fever. Infections such as these, caused by the sortase A bacteria, have led to economic losses of approximately $20 billion across the country (Aslam, 2018). In addition, owing to the antibiotic-resistant nature of these infections, the US has lost approximately $35 billion annually in productivity, which is time that could've been spent in other health-relate endeavors (Aslam, 2018).

The countless lives and the economic losses caused by the prevalence and growing antibiotic-resistance of sortase A bacteria-related diseases makes it crucial to explore alternate methods to prevent such infections and diseases. One successful method, particularly against the S. aureus bacteria, was the use of sortase enzyme inhibitors (Tan, 2019). Sortase A inhibitors were found to inhibit the function of the sortase A enzymes that were present in S. aureus bacteria, ultimately preventing the chronic lung infections in cystic fibrosis cases. The goal of this study was to extend the use of sortase A inhibitor molecules to other sortase A bacteria, such as Listeria monocytogenes and S. pyogenes, which would be effective since all sortase A bacteria contain a sortase A enzyme that can be inhibited.

This study seeks to create a structural model of a saSrtA-inhibitor complex (Staphylococcus aureus sortase A inhibitor complex), which can be used to explore ways of using sortase A inhibitors to inhibit the sortase A enzyme in other sortase A bacteria. This will ultimately explore more efficient methods to better prevent a wide variety of sortase A bacteria-related infections/diseases. Another goal of this study was to measure the efficacy of the sortase A inhibitors with the other sortase A bacteria, which was done by measuring the binding affinity between the inhibitors and the bacteria. This study aims to create a saSrtA-inhibitor complex model that depicts established-sortase A inhibitors in Staphylococcus aureus and predict the efficiency of those sortase A inhibitors in other sortase A bacteria, ultimately, to better understand and potentially prevent the worldwide infections caused by sortase A bacteria.

Methodology

The engineering procedure consists of gathering a list of various sortase A inhibitors in the form of small, synthetic organic molecules. All Class A sortase structures were classified using the Protein Data Bank, an online database for three-dimensional structural data of large biological molecules. Previously, the structures of the various sortase A inhibitors were modeled manually through software systems such as UniProt. UniProt provided information on proper access to protein structure sequences, information about certain functional groups, and helpful history on how these proteins were derived.

Once all of the specific protein structures had been gathered using a literature search of "S. aureus inhibitor molecules", six specific sortase A inhibitors were chosen: Benzothiazolinone (BIZ), 1,3,4-Thiadiazole (TDZ), Thiazolidinedione (TDI), 2-Phenylthiazole (2PT), Benzo[d]oxazole (BDO), and 3,3,3-trifluoro-1-(phenylsulfonyl)-1-propene (TPP). These six inhibitors were visualized through the PyMOL software (Fig. 1). In addition, four sortase A bacteria were chosen – Staphylococcus aureus, Streptococcus pyogenes, Listeria monocytogenes, and Actinomyces oris. These four sortase bacteria were visualized through PyMOL (Fig. 2).

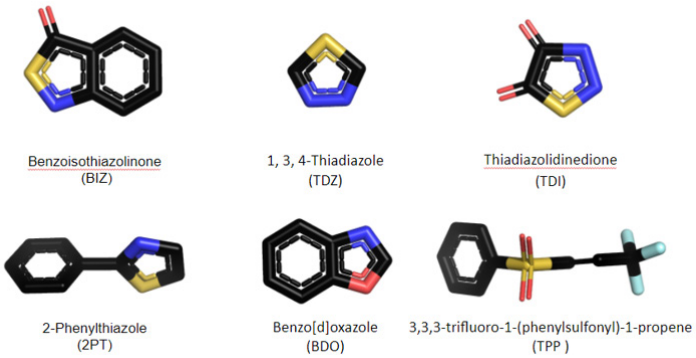


Figure 1. PyMOL diagrams of the sortase A inhibitors that were used in the study.

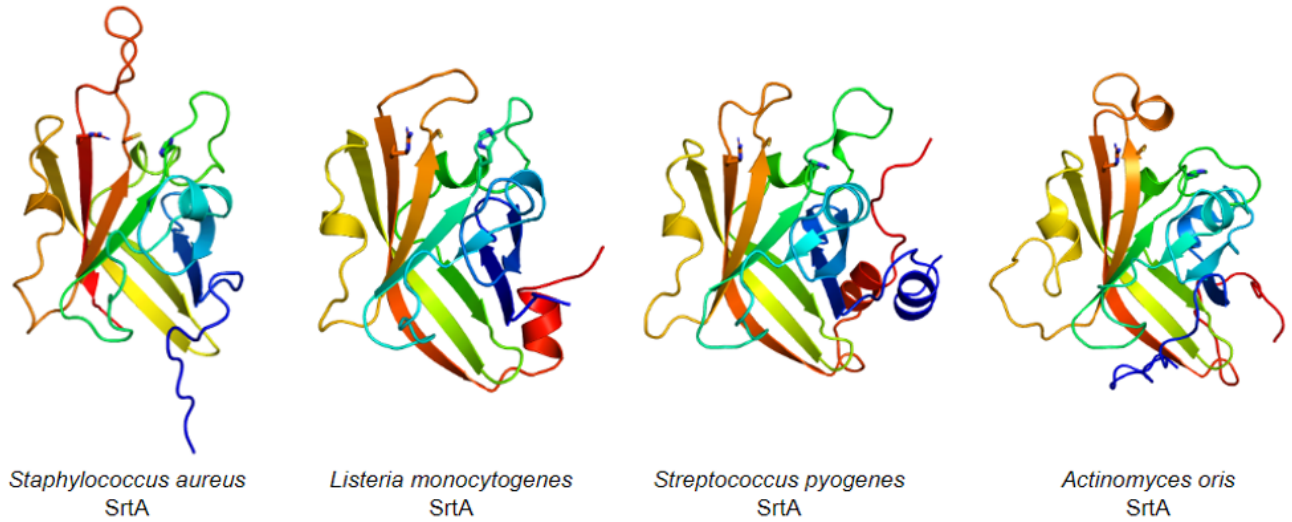


Figure 2. PyMOL diagrams of the four sortase A bacteria that were used in the study – *S. aureus, Listeria monocytogenes, S. pyogenes, and Actinomyces oris.*

Then, ChemDraw was used to generate the SMILES strings of the six sortase A inhibitors and the four sortase A bacteria. Phenix eLBOW, a python-based ligand builder and mapper, was used to convert the ChemDraw SMILES strings into atomic coordinates. Phenix eLBOW provided atomic coordinate information through a .pdb file, which was then inserted into the HADDOCK 2.4 software for bacteria-inhibitor docking purposes. HADDOCK was used to conduct a total of 24 docking runs – four sortase A bacteria with six sortase A inhibitors each. Ten trials were conducted in each of the 24 HADDOCK docking runs were, resulting in a total of 240 trials. After the docking runs, HADDOCK provided quantitate bacteria-inhibitor statistics – one of which was "buried surface area". Buried surface area can be defined as the macroscopic description for hydrophobic interactions (Panagiotis, 2012). Essentially, buried surface area values measure the surface of the enzyme that is buried away from the solvent that the bacteria rests in. After researching about a direct positive correlation between "buried surface area" and binding affinity, "buried surface area" was used to determine the "most efficient" bacteria-inhibitor complex (Table 1).

Table 1. Measuring Binding Affinity of Sortase A Inhibitors and Sortase A bacteria

	2PT	**BDO**	**BIZ**	**TDI**	**TDZ**	**TPP**
S. aureus	299.0	232.6	265.6	248.6	188.7	377.7
S. pyogenes	326.4	282.8	301.9	243.7	205.8	414.9
Listeria monocytogenes	357.7	290.5	303.0	258.9	208.9	394.6
Actinomyces oris	336.2	252.2	260.7	249.2	197.0	451.9
Averages	329.8	264.5	282.8	250.1	200.1	409.8

Additionally, PyMOL was used to visualize all the bacteria-inhibitor docking runs (Fig. 3). Visualization aids in reaching a true conclusion on whether or not a certain molecule is efficient in inhibiting the function of sortase A bacteria. Overall, the PyMOL visualizations and HADDOCK data were helpful in analyzing the sortase A bacteria-inhibitor complexes.

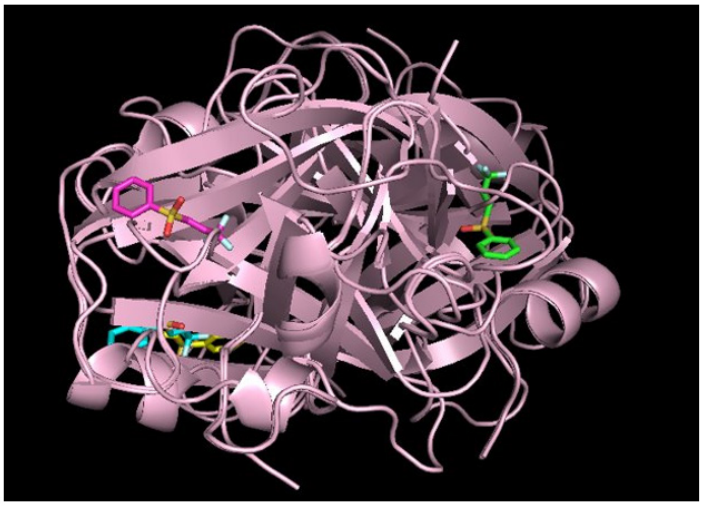


Figure 3. PyMOL visualization of *Listeria monocytogenes* bacteria with TPP Inhibitor

Results

This study met its goals of creating a saSrt-A inhibitor complex model, using the model to test sortase A inhibitor-use in other sortase A bacteria, determining the efficacy of sortase A inhibitors using buried surface area data, and ultimately exploring alternate methods to prevent antibiotic-resistant infections caused by sortase A bacteria. The average efficiency of 2PT with all four bacteria was 329.8 Å, BDO averaged 264.5 Å, BIZ averaged 282.8 Å, TDI averaged 250.1 Å, TDZ averaged 200.1 Å, and TPP averaged 409.8 Å. The median values for efficiency of each inhibitor were 331.3 Å for 2PT, 267.5 Å for BDO, 283.8 Å for BIZ, 248.9 Å for TDI, 202.95 Å for TDZ, and 404.75 Å for TPP. The standard deviation of 2PT is 21.09, 22.34 for BDO, 19.73 for BIZ, 5.51 for TDI, 7.9 for TDZ, and 27.66 for TPP. Our most significant finding was that 3,3,3-trifluoro-1-(phenylsulfonyl)-1-propene (TPP) had the greatest average buried surface area with each bacterium, which correlates to the greatest bacteria-inhibitor binding affinity, implying that it would potentially be the most efficient inhibitor at preventing sortase A bacteria-related diseases.

Potential future studies include the modification of already-existent sortase A inhibitors through the addition of certain functional groups, in order to make them more efficient. After further research, TPP was found to be the most efficient inhibitor due to it having the highest buried surface area and binding affinity. This was because of the high difference in electronegativity that is present in the TPP molecule. TPP contains a fluorine atom, which is one of the most electronegative elements, and this great difference in electronegativity led to a high intrinsic bond energy, and thus a high buried surface area. So, the addition of a fluorine atom to the other inhibitors could potentially increase the difference in electronegativity, leading to an increase in buried surface area and overall efficiency of the inhibitor. However, the addition of a fluorine atom to potentially increase the efficiency of other sortase A inhibitors is something that would have to be tested in a lab setting.

References

Brinkman, C. L., Schmidt-Malan, S. M., Karau, M. J., Greenwood-Quaintance, K., Hassett, D. J., Mandrekar, J. N., & Patel, R. (n.d.). Exposure of bacterial biofilms to electrical current leads to cell death mediated in part by reactive oxygen species. PLOS ONE. Retrieved December 18, 2021, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0168595

Clancy, K. W., Melvin, J. A., & McCafferty, D. G. (2010). Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition. Biopolymers. Retrieved December 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4648256/

Crow, D. (2016, August 1). Sticky bacterial biofilms dissolve on contact with sugar-cutting enzymes. Scientific American. Retrieved January 17, 2022, from https://www.scientificamerican.com/article/sticky-bacterial-biofilms-dissolve-on-contact-with-sugar-cutting-enzymes/

Ghasemian, A., Najar Peerayeh, S., Bakhshi, B., & Mirzaee, M. (2015). The microbial

surface components recognizing adhesive matrix molecules (mscramms) genes among clinical isolates of Staphylococcus aureus from hospitalized children. Iranian journal of pathology. Retrieved December 20, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539745/

Ha, M. W., Yi, S. W., & Paek, S. M. (2020). Design and Synthesis of Small Molecules as Potent Staphylococcus aureus Sortase A Inhibitors. Antibiotics (Basel, Switzerland), 9(10), 706. https://doi.org/10.3390/antibiotics9100706

Harmon, K. (2010, August 10). Drug-resistant staph infections on the decline in U.S. hospitals. Scientific American Blog Network. Retrieved January 18, 2022, from https://blogs.scientificamerican.com/observations/drug-resistant-staph-infections-on-the-decline-in-u-s-hospitals/

Kiedrowski, M. R., Gaston, J. R., Kocak, B. R., Coburn, S. L., Lee, S., Pilewski, J. M., Myerburg, M. M., & Bomberger, J. M. (2018). Staphylococcus aureus Biofilm Growth on Cystic Fibrosis Airway Epithelial Cells Is Enhanced during Respiratory Syncytial Virus Coinfection. mSphere, 3(4), e00341-18. https://doi.org/10.1128/mSphere.00341-18

Ko, Y.-P., & Flick, M. J. (2016, June). Fibrinogen is at the interface of host defense and pathogen virulence in Staphylococcus aureus infection. Seminars in thrombosis and hemostasis. Retrieved December 20, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5514417/

Lévesque, C. M., Voronejskaia, E., Huang, Y.-C. C., Mair, R. W., Ellen, R. P., & Cvitkovitch, D. G. (2005, June). Involvement of sortase anchoring of cell wall proteins in biofilm formation by streptococcus mutans. Infection and immunity. Retrieved December 20, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1111851/

Mandlik, A., Swierczynski, A., Das, A., & Ton-That, H. (2008, January). Pili in gram-positive bacteria: Assembly, involvement in colonization and biofilm development. Trends in microbiology. Retrieved December 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2841691/

Nitulescu, G., Zanfirescu, A., Olaru, O. T., Nicorescu, I. M., Nitulescu, G. M., & Margina, D. (2016, November 22). Structural analysis of sortase a inhibitors. Molecules (Basel, Switzerland). Retrieved December 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6272945/

Pickering, A. C., Vitry, P., Prystopiuk, V., Garcia, B., Höök, M., Schoenebeck, J., Geoghegan, J. A., Dufrêne, Y. F., & Fitzgerald, J. R. (2019, June 19). Host-specialized fibrinogen-binding by a bacterial surface protein promotes biofilm formation and innate immune evasion. PLoS pathogens. Retrieved December 20, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6602291/

Schneewind Olaf and Missiakas Dominique M. 2012Protein secretion and surface display in Gram-positive bacteriaPhil. Trans. R. Soc. B3671123–1139. https://doi.org/10.1098/rstb.2011.0210

Thappeta, K. R. V., Zhao, L. N., Nge, C. E., Crasta, S., Leong, C. Y., Ng, V., Kanagasundaram, Y., Fan, H., & Ng, S. B. (2020, November 14). In-silico identified new natural sortase a inhibitors disrupt S. aureus biofilm formation. International journal of molecular sciences. Retrieved December 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7696255/

Vestby, L. K., Grønseth, T., Simm, R., & Nesse, L. L. (2020, February 3). Bacterial biofilm and its role in the pathogenesis of disease. Antibiotics (Basel, Switzerland). Retrieved December 18, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7167820/

Yu Shichen, Zhu Xiaoshan, Zhou Jin and Cai Zhonghua 2018Biofilm inhibition and pathogenicity attenuation in bacteria by Proteus mirabilisR. Soc. open sci.5170702170702 https://doi.org/10.1098/rsos.170702

Overexposure To Maternal Cortisol During Gestation And Its Influence On The Regulation Of HPA Axis In Female Adults

Sanya Sahni

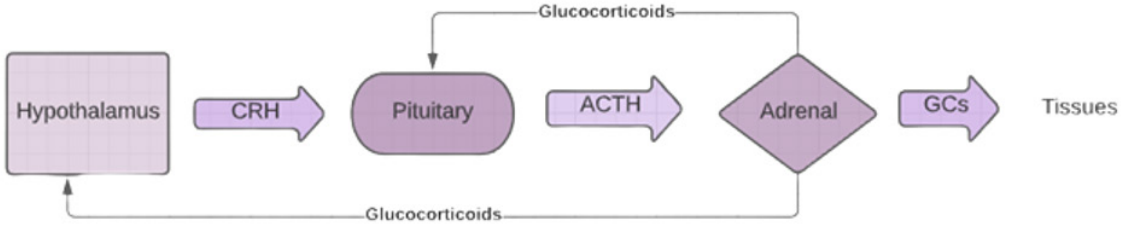
Nikola Tesla STEM High School

Introduction

This study aims to explore the relationship between maternal dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis in the female fetus upon reaching adulthood as measured by perceived stress and cortisol levels. The gestation period is arguably the most pivotal period in a human's lifetime. The quality of pregnancy has a significant impact on an individual's functioning and development, especially the growth of the brain which is essential in determining later life outcomes. The National Comorbidity Survey Replication study estimated that exposure to early-life adversity could account for 32.4% of psychiatric disorders. Genes, nutrition, air, and water quality all compete for significance during the process of "fetal programming". Maternal stress, however, is emerging as a fundamental factor in determining fetal brain growth, the child's response to the world and eventually the neuropsychological profile of the adult. Uncontrolled exposure to maternal cortisol (stress) because of a dysregulated hypothalamic–pituitary–adrenal axis can affect the stress response in the fetus which can have long-lasting effects on brain physiology and behavior (Coussons- Read, 2013).

Cortisol, released by the adrenal cortex, is a vital hormone that is involved in the body's response to stress, modulating immune function and inflammatory processes, and regulating metabolism. During gestation, cortisol plays a role in fetal development, increasing at critical points to aid in neural development (Nath A., 2017). However, the problem occurs when there is too much cortisol passing the placental barrier and subsequently modifying the normal development of the HPA axis of the growing fetus. This is typically the result of stress during pregnancy, and the associated detrimental effects of this can include low birth weight, premature birth, lower cognitive abilities in children, and diabetes in adults. Prenatal stress induces the release of glucocorticoids such as cortisol and cortisone in the mother which enter the fetal circulation and gain access to the developing nervous system, (Coussons- Read 2013). This access to the nervous system disrupts the normal development and results in alterations to the HPA axis of the fetus.

Figure 1.1: the HPA axis and its functions along with the role of glucocorticoids



Note: CRH (corticotrophin releasing hormone) begins a cascade event which leads to the production of ACTH (adrenocorticotropic hormone) which once it gets to the adrenal cortex helps release of glucocorticoids. Glucocorticoids help in the production of tissues, including nervous brain tissue.

According to the Department of Health at the University of Utah, women are twice as more likely to be diagnosed with anxiety disorders than men. It is hypothesized that female fetuses have a much higher susceptibility to external environmental factors as their brains exercise more neuroplasticity when compared to males. Furthermore, researchers have uncovered astonishing evidence supporting the idea that high cortisol levels create different results in male and female fetuses in terms of survival and development. Female fetuses adapt to overexposure to cortisol by increasing their amygdala volume, and for females, higher than usual amygdala volumes have been linked to anxiety in a postnatal environment (Kim et. al, 2017). It is hypothesized that high stress during pregnancy will release increase levels of cortisol in the body of the pregnant woman, which means that more cortisol will cross the amniotic sac and influence the fetal brain, especially the hypothalamic-pituitary-adrenal during pregnancy. This overexposure results in the HPA axis developing a bit differently which could lead to heightened feelings of anxiety in the female adult which will be associated with higher cortisol levels. If there is a high level of perceived stress during a mother's pregnancy, then data will show a positive relationship between maternal stress and the daughter's reported anxiety and cortisol levels.

Methodology

In this study, biological mother daughter dyads were recruited. The daughters were 18 or above and the mothers below the age of 75. Materials needed included informed consent forms, Beck Anxiety Inventory, cortisol salivary home test kits (LetsGet Checked Cortisol Testing), flyers, PSA test (Pregnancy Specific Anxiety). The BAI and PSA test are open source and publicly available for use by everyone. Participants were recruited by emailing daughters at universities and requesting them to contact their mothers. Both the daughters and the mothers were requested to fill out an informed consent form. Half of the daughters were requested to give their current residential address so that the cortisol tests can be delivered to them. And the mothers filled out a Microsoft form which require them to fill out demographic information from the time of their pregnancies. The questions asked about household income during pregnancy, marital and income status and age during pregnancy. An online random number generator was used to match random numbers with each mother daughter duo. This is done to ensure the privacy of the participants; calculations were not made with the participant's name in mind. The daughters were requested to take the BAI and at home cortisol tests, while the mothers were requested to take the PSA test. The cortisol tests need to be taken at 11 AM on the same day without taking any food or drink for thirty minutes before the actual test. For best results, participants were requested to keep the cotton swabs in their mouths for up to two minutes. Daughters were also asked additional questions about health conditions like diabetes etc. This is done to ensure that they have no endocrine irregularities which may impact their levels of cortisol. Additionally, daughters will be questioned about symptoms of irregular cortisol levels. Such as "How often do you experience increased frequency of urination or thirst?" The subjects were informed if the cortisol is outside the normal range, that such results in a home test could represent a false positive, it should not be considered a medical diagnostic test or be considered as medical advice and that consultation with a healthcare provider is recommended. Results for BAI, PSA and cortisol(ng/mol) were added to the data table below (Data Table 1) Simple calculations like mean, mode and median were calculated for the BAI, PSA, and cortisol test results. Since the entire study is carried out remotely, there will be no need to consider Covid 19 restrictions as there is no risk of Covid exposure.

	A	B	C	D	E	F
1	RN	PSA	BAI	Cortisol	Diabetes	Endocrine
2	1	20	31	1.1	Yes	
3	2	19	20	0.89		
4	3	22	26	1.1	Yes	Thyroid
5	4	21	21	1.2		
6	5	17	19	1	Yes	
7	6	12	20	0.76		
8	7	19	22	0.93		
9	8	25	34	1.3		
10	9	28	30	1.7		
11	10	35	30	NA		
12	11	33	28	NA	Yes	
13	12	27	36	NA	Yes	
14	13	30	24	NA	Yes	
15	14	24	30	NA		
16	15	34	23	NA	Yes	
17	16	26	29	NA		
18	17	32	35	NA		Thyroid
19	18	30	31	NA	Yes	

Data Table 1: BAI, PSA and Cortisol test results

Results and Analysis

Average, median and range were calculated for the Beck Anxiety Inventory Test scores, Pregnancy Specific Anxiety Inventory and Cortisol test results. Standard deviation was calculated to get a measure of the variability in the range of values. The standard deviation for the cortisol test results was small because the normal cortisol range is very small. Correlations using Pearson's correlation coefficient were calculated between the cortisol test results of the daughter and the PSA test results of the mother. Additionally, correlations were calculated for the daughter's BAI test results and the cortisol test results of the daughter. Correlations will be calculated using the Pearson's correlation coefficient. A p value of 0.03 was found for the correlation between the Pregnancy Specific Anxiety test scores and the Beck Anxiety Inventory test results. This demonstrates that there is a positive correlation between the daughter's reported anxiety levels and the stress experienced by the mother during gestation. A p value of 0.037 was found between the salivary cortisol test results and the anxiety levels reported by the daughters. Additionally, there was a strong correlation found between the occurrence of type 2 diabetes in daughters and stress experienced by mothers during their pregnancy.

Conclusion and Discussion

Some limitations of the procedure include that the test results for PSA were based on memories of the mothers. The 18 years or more since the mother's pregnancy opens space for error and decreases reliability. A confounding variable could be high cortisol test results and BAI test scores due to major life events. Additionally, only about half of the daughters were able to take the cortisol tests. Results could have been different if all the daughters were administered the cortisol test. At home cortisol tests are less reliable than an in-lab blood sample test. This study only focused on female participants as females and male fetuses have different adaptation techniques. Females are survival focused and better at adapting to stressors received from the mother's system. Male fetuses are growth focused and aren't as good at adapting to such stimuli. This results in female stress response developing a bit differently than males. A better approach to the addressed issue could be to do an experimental study and compare the anxiety in offspring and stress during pregnancy

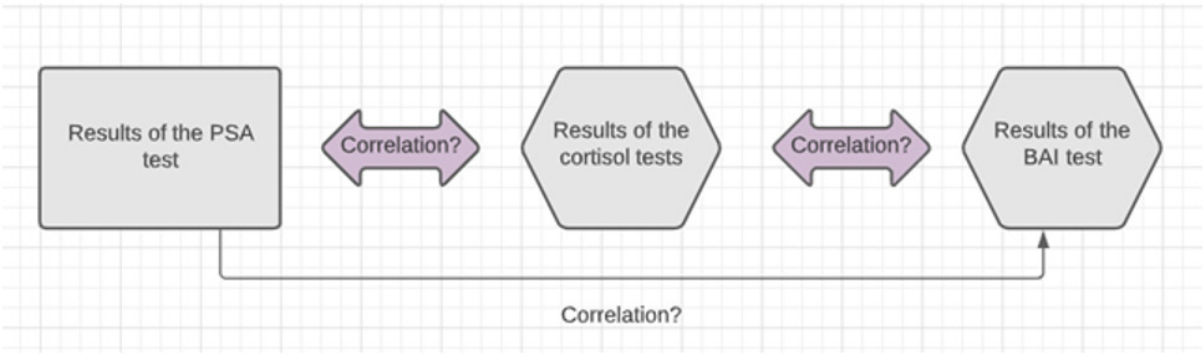


Figure 1.2: Correlation calculations

between males and females. Another possible future research study could include Comparing BAI reports of daughters of mothers with GAD (experimental group) and daughters of mothers without GAD (control group). Further understanding of the impact of the changes to the HPA axis may allow early identification of those most at risk of future disease.

Bibliography

Coussons-Read, M. E. (2013, June). Effects of prenatal stress on pregnancy and human development: Mechanisms and pathways. Obstetric medicine. Retrieved December 3, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052760/.

Nath, A., Murthy, G. V. S., Babu, G. R., & Di Renzo, G. C. (2017, July 17). Effect of prenatal exposure to maternal cortisol and psychological distress on infant development in Bengaluru, southern India: A prospective cohort study - BMC psychiatry. BioMed Central. Retrieved December 2, 2021, from https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-017-1424-x.

Kim, D.-J., Davis, E. P., Sandman, C. A., Sporns, O., O'Donnell, B. F., Buss, C., & Hetrick, W. P. (2017, November 1). Prenatal maternal cortisol has sex-specific associations with Child Brain Network Properties. Cerebral cortex (New York, N.Y. : 1991). Retrieved December 9, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6084613/#:~:text=Elevated%20maternal%20cortisol%20concentrations%20have,manifest%20at%20the%20whole%2Dbrain.

Laplante, D. P., Barr, R. G., Brunet, A., Du Fort, G. G., Meaney, M. L., Saucier, J.-F., Zelazo, P. R., & King, S. (n.d.). Stress during pregnancy affects general intellectual and language functioning in human toddlers. Nature News. Retrieved December 10, 2021, from https://www.nature.com/articles/pr2004225.

Duthie, L., & Reynolds, R. M. (2013, September 19). Changes in the maternal hypothalamic-pituitary-adrenal axis in pregnancy and postpartum: Influences on maternal and fetal outcomes. Neuroendocrinology. Retrieved December 10, 2021, from https://www.karger.com/article/fulltext/354702.

Wood, C. E., & Keller-Wood, M. (2016, January 28). The critical importance of the fetal hypothalamus-pituitary-adrenal axis. F1000Research. Retrieved December 10, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4755437/.

Oxytocin as a Strategy to Enhance the Effectiveness of the Satiation Signal, Cholecystokinin, to Reduce Food Intake in High Fat Diet-Fed Rats

Akhila Bourampeta

Nikola Tesla STEM High School

Abstract

Obesity is a serious chronic disease that is a leading risk factor for premature death contributing to approximately 2.8 million deaths worldwide. One possible treatment for obesity is oxytocin that reduces food intake, in part by enhancing the effectiveness of the satiation signal, cholecystokinin (CCK), to reduce food intake by activating cells in hindbrain areas that control appetite. This project aims to understand the interactive effects of oxytocin and CCK on high fat diet consumption in rats. The goals of this study were to 1) identify areas of the hindbrain that contribute to the effects of CCK-8 and 2) demonstrate that oxytocin enhances the responsiveness of low doses of CCK to reduce high fat diet consumption. Low doses of CCK-8 (0.25, 0.5 nmol/kg) failed to reduce food intake in animals maintained on a high fat diet, but oxytocin enhanced the ability of lower doses of CCK-8 (0.25, 0.5 nmol/kg) to reduce high fat diet consumption. It's determined that CCK reduced low fat diet consumption at both low (0.25 nmol/kg) and high doses (8 nmol/kg) ($P<0.05$). These effects of CCK (8 nmol/kg) were associated with an increased number of c-Fos (+) cells (marker of neuronal activation) within the hindbrain nucleus tractus solitarius (NTS) (control: 6.3±2.7 cells vs. CCK-8: 72.5±31.3 cells) ($P<0.05$). Ultimately, this approach yields new insights into brain sites that could potentially contribute to the ability of oxytocin to enhance the effectiveness of satiation signals to control appetite.

Keywords: Oxytocin, CCK, c-Fos (+) cells, hindbrain

I. Introduction

As millions of individuals suffer from the wide range of consequences from obesity, improved treatments are required to save those lives. This project aims to ensure the survivability by experimenting with oxytocin and cholecystokinin (CCK) that is injected in high fat diet fed rats and examining c-fos (+) cells, resulting in a successful treatment for this condition.

Worldwide, obesity is one of the leading causes of death. More than 332 million adults suffer from obesity in the U.S and close to 2.8 million die each year (WHO, 2021). Treatments such as medication, a change in diet, and surgery have been developed largely to prevent obesity. However, these treatments are generally ineffective for obese patients. The medical costs to treat obesity and its associated complications (such as cardiovascular disease) in the U.S. range from $147 billion to nearly $210 billion per year. If obesity rates continue on their current track, obesity related medical costs are estimated to increase by $48 - $66 billion each year in the U.S., and the loss of economic productivity could be as high as $580 billion yearly by 2030. (Finkelstein et al, 2009)

It is well recognized that oxytocin, a peptide which is expressed within cells of the paraventricular nucleus (PVN) and supraoptic nucleus (SON) within the hypothalamus (an area of the brain linked to the control of body weight), plays an important role in a variety of functions, including reproductive behavior, social bonding, regulation of food intake and body weight. There has been growing excitement over the potential use of oxytocin (OT) as an anti-obesity therapy, as it has recently been found to reduce body weight following administration in overweight and obese humans without any

adverse side effects (Zhang et al, 2013).

In the streptozotocin-induced diabetic mouse model from the National Library of Medicine, scientists were intrigued about the use of oxytocin as a treatment for obesity and these effects have been translated from rat and mouse models of genetic obesity and diet-induced obesity to diet-induced obese nonhuman primates such as Rhesus monkeys (J.E. Blevins et al, 2015) and overweight humans (Zhang et al, 2013). This illustrates the need for enhanced treatments that results in a significant reduction of obese. Increased oxytocin action may reduce meal size, by enhancing the effectiveness to short-term meal-related eating behavior signals (satiation signals) such as cholecystokinin (CCK).

According to the hypothesis, oxytocin treatment increases the effectiveness of low doses of CCK-8 to reduce food intake in high fat diet-fed rats by activating areas in the hindbrain that are involved in the control of food intake. This experiment determines the extent to which oxytocin may identify areas within the hindbrain (Nucleus Tractus Solitarius, Area postrema) that express c-Fos (marker of neuronal activation) in response to low and high doses of CCK-8. These findings will help 1) determine the mechanism by which oxytocin reduces food intake in diet-induced obesity in animals maintained on high fat diets, 2) identify the areas in the hindbrain that may be linked to these effects. Furthermore, the experiment should be feasible, and the resulting combination should also be beneficial in reducing consumption of high fat diet in rats. This study will help identify appropriate dose combinations of oxytocin and CCK that will eventually lead to the development of future treatments to help prevent the progression of diet-induced obesity in high fat diet-fed rats.

Any significant change in the number of c-Fos (+) cells above that of control animals will represent an increase in neuronal activation. More cells represent the stronger degree as it's expected that the high dose of CCK will elicit a more robust or heightened c-Fos response (# of cells) compared to the low dose.

II. Methodology

This project investigates the interactive effects of oxytocin and CCK on short-term food intake (high fat diet consumption) in male CD-IGS rats from Charles River. At the conclusion of the study, hindbrain tissue was studied for c-Fos (+) in response to oxytocin alone or in combination with CCK. The number of c-Fos (+) cells were counted at 4 separate levels of the caudal hindbrain and the treatment groups were counted using a bright field Nikon 80i microscope.

The c-fos study was done to measure the changes in c-Fos within the hindbrain in response to CCK treatment. Oxytocin was given into the brain into an area called the third ventricle.

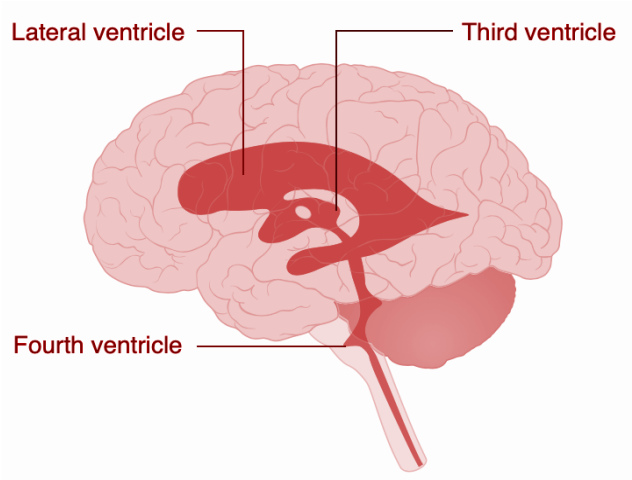


Figure 1. Illustration of the lateral, third and fourth ventricles of the adult rat brain (S. Vermilyea et al, 2014)

This is an open cavity within the brain that allows oxytocin to reach areas in the brain where it can act to reduce food intake or body weight. It acts as oxytocin receptors which are found in brain sites next to the third ventricle and throughout the brain. It is thought that oxytocin diffuses to these receptors and act to reduce food intake.
CCK was given into the peritoneal cavity or abdomen where it can then enter the bloodstream or at CCK receptors at the vagus nerve, a nerve that relays messages from the gut to the brain to activate hindbrain areas to reduce food intake and increase c-Fos (+) cells. The slides were prepared by having rats that received injections of CCK at (0.25 or 8 nmol/kg) or control (saline). Then they were humanly euthanized according to the approved animal care protocols and perfused with saline and 4% paraformaldehyde at 1.5-hour post injection and hindbrains were collected and processed for c-fos immunocytochemistry, which is a technique that is used to visualize the location of a protein or

antigen in cells by using a primary antibody that binds to it. Finally, the frozen coronal sections at 14 micrometers were obtained using a cryostat, a machine that maintains the temperature of samples placed inside.

Figure 2. Bright Field Nikon 80i microscope used to count the number of c-Fos (+) cells

III. Experimental Procedure

A. Analyzing c-Fos (+) cells from pictures

Mapping the distribution of neurons that expressed c-Fos protein in response to CCK-8, the caudal Nucleus Tractus Solitarius, the sensory nucleus in the medulla oblongata and lower pons in the brain of each rat was sampled between four coronal levels. Smaller structures such as the Area postrema on the medulla oblongata provided data for fewer levels per animal across two coronal levels posterior to the interaural line. Within each of the anatomic regions on each section sampled, all of the neurons that had a c-Fos (+) nucleus were counted at each of the levels. A neuron was considered to be c-Fos (+) if the c-Fos (+) nucleus was located within the boundary of the cytoplasm, as determined by adjusting the focus using the bright field microscope. c-Fos (+) cells were identified by the presence of dark brown color. The region of interest was placed visually so as to include c-Fos (+) cells within the visual field of the Area postrema. The Nucleus Tractus Solitarius of the injected CCK and the control-treated (Saline) was analyzed separately. This required counting more than one field for the Area postrema and Nucleus Tractus Solitarius. Analysis of c-Fos immunostaining in the Nucleus Tractus Solitarius following CCK-8 or control was together and at constant at 10x and 20x magnification using the microscope. Data was grouped at each of the anatomically matched levels for all of brains of rats injected with CCK-8 and separately for those of the control-injected (Saline). To compare the effect of CCK-8 on c-Fos expression at each level analyzed, the data was expressed as Linear Model equation, Estimates of Effects B = (X'X)-1X'Y, Analysis of Variance, and Fisher's Least-Significant-Difference Test equation to assess the number of c-Fos (+) cells of the CCK-injected group compared to control-injected.

Figure 3A- B. Examples of c-Fos (+) cells (marker of neuronal activation) under microscopes:

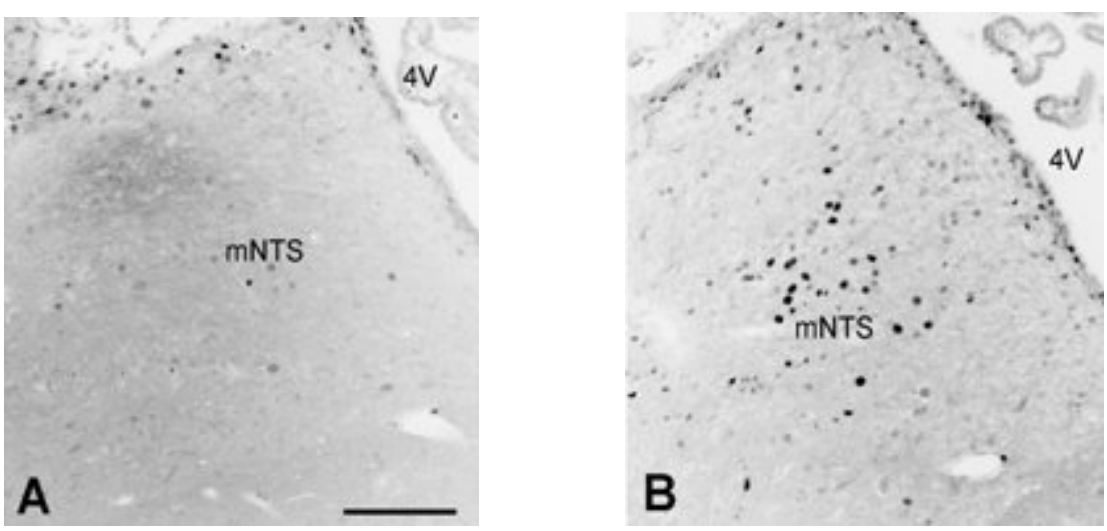


Figure 3A-B. This is a representative microscope picture of a coronal section (14 mm) from the medial nucleus tractus solitarius (mNTS) of a rat hindbrain. Images were taken from 1 side of the mNTS, with the fourth ventricle (4V) shown in top right.

Figure 3A: Control-treated control showed few Fos (+) cells.

Figure 3B: Peripheral injection of CCK-8 produced numerous Fos (+) neurons in mNTS.

CCK-8 (1.2 nmol/rat) or control (0.1% BSA in saline) was given in the abdominal organs into a group of adult male Wistar rats (310-409 grams) at 1.5 hours prior to tissue collection. CCK-8 (Blevins et al, 2004).

Images were captured in gray scale and then reversed for clarity in this figure with Photoshop. Images were taken at 40x magnification using a Zeiss Akioskop fluorescence microscope.

B. Data Analysis

To measure the cells, a Linear Model equation, Estimates of Effects B = (X'X)-1X'Y, Analysis of Variance, and Fisher's Least-Significant-Difference Test was used for each treatment. The table below shows the levels of the dose and the number of cells in the Nucleus Tractus Solitarius and in the Area Postrema.

NTS = nucleus tractus solitarius
NTSFOS = # of c-Fos (+) cells in NTS
AP = area postrema
APFOS = # c-Fos (+) cells in AP

Table I. ***General Linear Model***

Variables	Levels		
DOSE (3 levels)	0.000	0.250	8.000

Dependent Variable	NTSFOS
N	12
Multiple R	0.623
Squared Multiple R	0.388

Table II. ***Estimates of Effects B = $(X'X)^{-1}X'Y$***

Factor	Level	NTSFOS
CONSTANT		34.417
DOSE	0	-28.167
DOSE	0.25	-9.917

Table III. ***Analysis of Variance***

Source	Type III SS	df	Mean Squares	F-Ratio	p-Value
DOSE	9,368.167	2	4,684.083	2.850	0.110
Error	14,792.750	9	1,643.639		

Post Hoc Test of NTSFOS
Using least squares means.
Using model MSE of 1,643.639 with 9 df.

Table IV. ***Fisher's Least-Significant-Difference Test***
This is a two-step testing procedure for pairwise comparisons of several treatment groups. In the first step of the procedure, a global test performed for the null hypothesis that the expected means of all treatment groups under study are equal.

DOSE(i)	DOSE(j)	Difference	p-Value	95% Confidence Interval	
				Lower	Upper
0	0.25	-18.250	0.540	-83.100	46.600
0	8	-66.250	0.046	-131.100	-1.400
0.25	8	-48.000	0.128	-112.850	16.850

The control vs 0.25 nmol/kg CCK-8 is not significantly different as this is likely due to small number of rats/groups.

Table V. ***General Linear Model***

Variables	Levels		
DOSE (3 levels)	0.000	0.250	8.000

Dependent Variable	APFOS
N	12
Multiple R	0.581
Squared Multiple R	0.338

Control vs 8 nmol/kg CCK-8 produced a significant increase in the number of c-Fos (+) cells within the nucleus tractus solitarius.

IV. Results

A. Data Collection

Results were measured by counting the number of c-Fos (+) cells by using the bright field Nikon 80i microscope. Dose of CCK-8 (0.25 or 8 nmol/kg) or control were used in rats.

Key:

AP (Area Postrema): Located on the medulla oblongata. It detects toxins in the blood and acts as a vomit-inducing center.

NTS (Nucleus Tractus Solitarius): Sensory nucleus located in the medulla oblongata and lower pons in the brain. It generates the peristaltic activity (muscle contractions that move food through digestive tract) during swallowing.

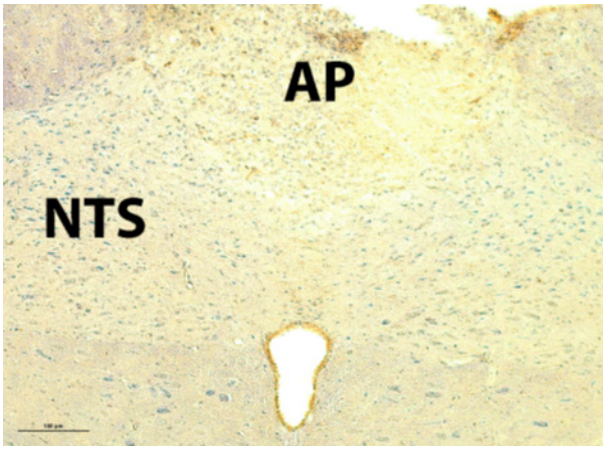


Figure 4. **Control (10x)**

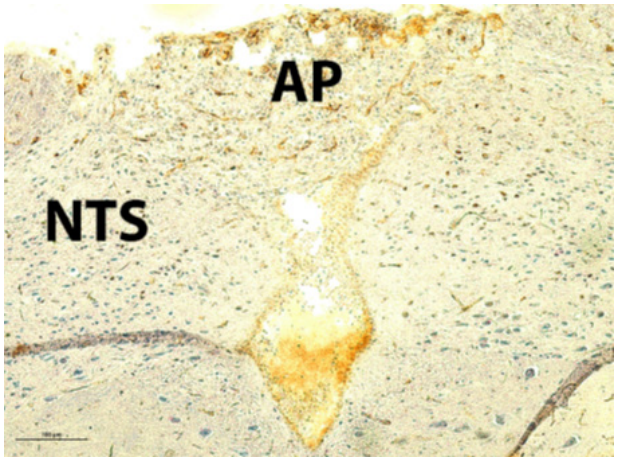


Figure 5. **CCK (10x)**

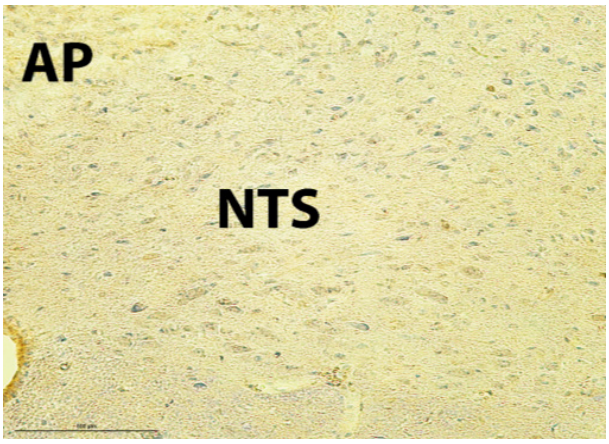


Figure 6. **Control (20x)**

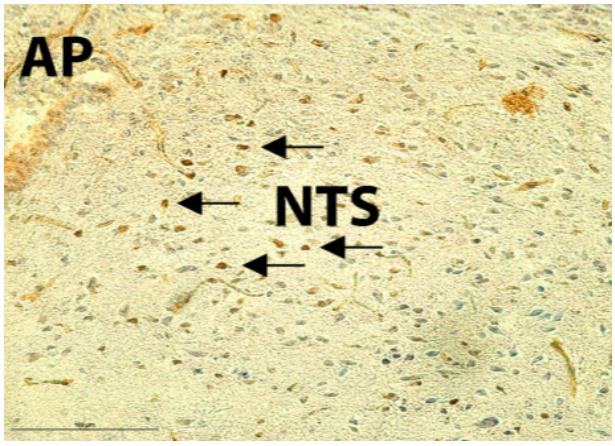


Figure 7. **CCK (20x)**

The figures are images of hindbrain slide on a bright field Nikon 80i microscope taken at both 10x and 20x magnification. The results were measured by counting the number of c-Fos (+) cells. These cells are defined as an indicator of neuronal activity and is associated with a number of neural and behavioral responses.

The CCK treatment slide showed multiple c-fos(+) cells which indicate a stronger degree of activation. This was expected given that these animals received a high dose of CCK and the increased doses of

CCK will produce a great c-Fos (+) response. *Goal 1* was accomplished, which was to identify areas of the hindbrain that contribute to the effects of CCK-8 through microscopic analysis of c-Fos (+) cells.

Table VI. Data collection table to assess number of c-Fos (+) cells in response to systemic [intraperitoneal (IP)] injections of CCK (0.25 or 8 nmol/kg) or control in rats.Curing Cancer by Cancelling Cell Growth Curing Cancer by Cancelling Cell Growth

Rat ID	Dose (CCK-8)	# of c-Fos (+) Cells Rat Nucleus Tractus Solitarius	# of c-Fos (+) Cells Rat Area Postrema
2,320.00	0	7	11
2,318.00	0.25	67	14
2,324.00	8	142	95
2,319.00	8	103	25
2,325.00	8	45	28
2,315.00	0	5	9
2,316.00	0.25	0	0
2,317.00	0	13	10
2,321.00	8	0	3
2,322.00	0.25	29	7
2,323.00	0	0	1
2,326.00	0.25	2	5

[1] Number of c-fos (+) cells are counted for each treatment.

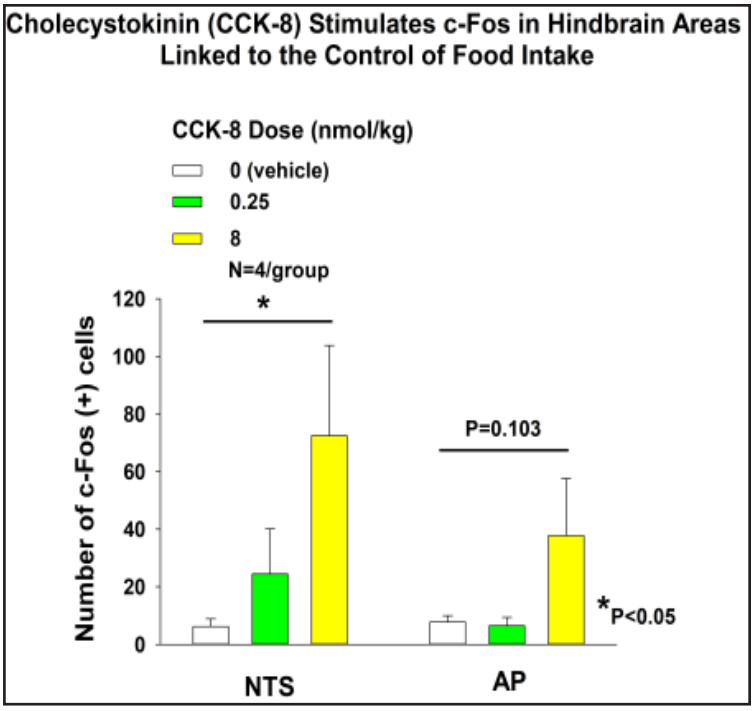


Figure 8. CCK-8 Activates Neurons in Key Hindbrain Areas Linked to the Control of Food Intake and Meal Size

There is no statistically significant difference in the number of c-Fos (+) cells within the nucleus of the solitary tract between control vs CCK at the lower dose of (0.25 nmol/kg). This may have been due to the small number of rats per group. In summary, CCK at low doses of 0.25 nmol/kg failed to cause a significant elevation of c-Fos relative to the control group whereas CCK at higher doses (8 nmol/kg CCK-8) produced a significant 11.5-fold increase in the number of c-Fos (+) cells within the Nucleus Tractus Solitarius. Goal 2 was accomplished, which was to demonstrate that oxytocin enhances the responsiveness of low doses of CCK to reduce high fat diet consumptions.

V. Conclusion / Discussion

A. Conclusion

Consistent with previous studies, (M. Covasa et al, 2001) CCK appeared to be more sensitive at reducing food intake in chow-fed rats relative to high fat diet-fed rats. CCK reduced chow consumption at a dose (0.25 nmol/kg) in the current findings that was found to be ineffective when administered into high fat diet-fed rats (J.E. Blevins et al, 2016). Overall, Oxytocin enhanced CCK-8 to reduce food intake. This produced a dose-dependent increase in the number of c-Fos (+) cells within hindbrain areas (NTS) linked to the control of calorie intake. The experiment inhibited food intake in chow-fed rats but was without effect in age-matched rats fed a high-fat diet. Collectively these findings support the hypothesis that consumption of high-fat fat diet impairs the response to meal-related satiation signals such as CCK.

B. Discussion

Future studies will determine if oxytocin can restore the effectiveness of low doses of CCK to activate neurons within the hindbrain of High fat diet fed animals. Ultimately, this approach yields new insights into the brain that could contribute to the ability of oxytocin to enhance the effectiveness of satiation signals to control appetite.

Bibliography

[1] Animals, Committee for the Update of the Guide for the Care and Use of Laboratory, and National Research. Council. Guide for the Care and Use of Laboratory Animals. National Academies Press, 1900.

[2] Arletti, R., et al. "Oxytocin Inhibits Food and Fluid Intake in Rats." National Library of Medicine, 1990, pubmed.ncbi.nlm.nih.gov/2087513.

[3] B.R. Olson, M.D. Drutarosky, E.M. Stricker, and J.G. Verbalis, Brain oxytocin receptor antagonism blunts the effects of anorexigenic treatments in rats: evidence for central oxytocin inhibition of food intake. Endocrinology 129 (1991) 785-91.

[4] C. Blouet, Y.H. Jo, X. Li, and G.J. Schwartz, Mediobasal hypothalamic leucine sensing regulates food intake through activation of a hypothalamus-brainstem circuit. J Neurosci 29 (2009) 8302-11.

[5] C.L. Ogden, M.D. Carroll, B.K. Kit, and K.M. Flegal, Prevalence of obesity in the United States, 2009-2010, NCHS data brief, 2012.

[6] C.M. Hales, M.D. Carroll, C.D. Fryar, and C.L. Ogden, Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017-2018. NCHS data brief (2020) 1-8.

[7] C.M. Lokrantz, K. Uvnas-Moberg, and J.M. Kaplan, Effects of central oxytocin administration on intraoral intake of glucose in deprived and non-deprived rats. Physiol Behav 62 (1997) 347-52.

[8] D.G. Baskin, F. Kim, R.W. Gelling, B.J. Russell, M.W. Schwartz, G.J. Morton, H.N. Simhan, D.H. Moralejo, and J.E. Blevins, A new oxytocin-saporin cytotoxin for lesioning oxytocin-receptive neurons in the rat hindbrain. Endocrinology 151 (2010) 4207-13.

[9] E.A. Finkelstein, J.G. Trogdon, J.W. Cohen, and W. Dietz, Annual medical spending attributable to obesity: payer-and service-specific estimates. Health affairs 28 (2009) w822-31.

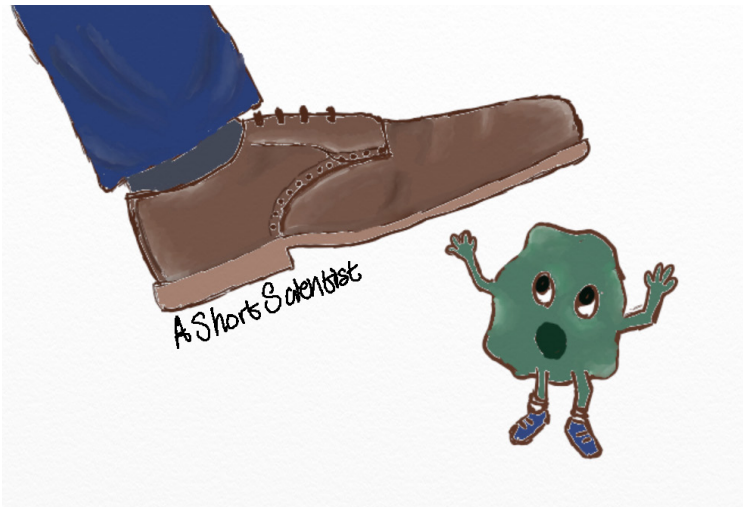
[10] E.A. Lawson, D.A. Marengi, R.L. DeSanti, T.M. Holmes, D.A. Schoenfeld, and C.J. Tolley, Oxytocin reduces caloric intake in men. Obesity (2015).

Curing Cancer by Cancelling Cell Growth

Adhya Konaw

A Short Scientist
Bringing Science to everyone

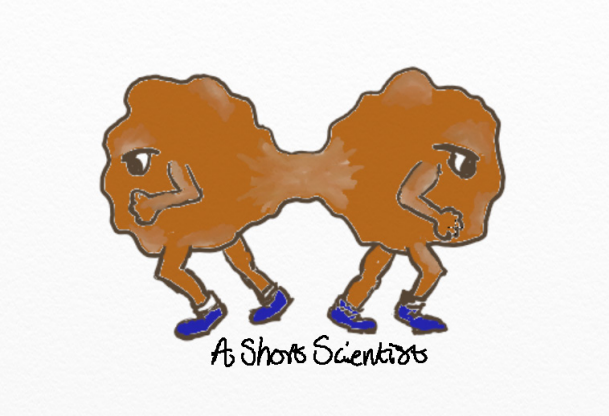
Adhya is the second winner of our science writing competition, and entered the 13-15 year old category. Adhya is 15 years old and from the USA, our first winner outside of the UK! Adhya's piece is all about how some cancer treatments prevent cell growth. It was an insightful read, and incredibly impressive for someone so young. Congratulations Adhya.



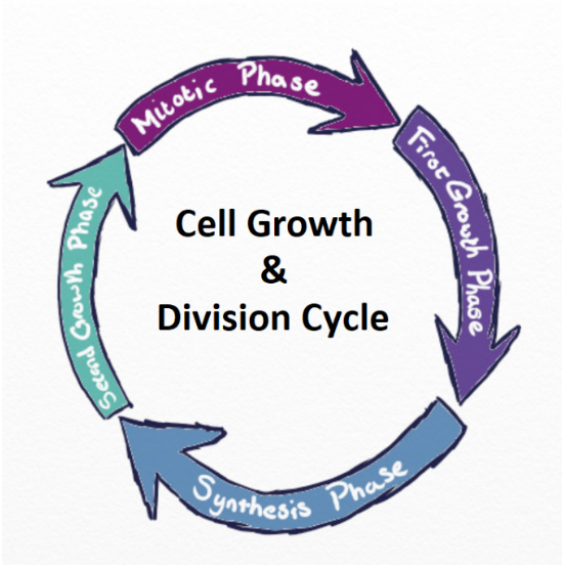
Hey everybody! Today we're going to talk about something called a thymidylate synthase inhibitor, a substance that works against the effects of cancer and growth of tumours in the body.

Before we get started— whatever you're thinking right now, there's no need to be intimidated by the large words. I'm going to give you all the context you need to understand our topic, even if you've never studied biology before, and I'll take you step-by-step through the details of how and why this treatment works. So, without further ado; Let's dive in.

Our DNA, for those of you who don't yet know, is what carries the instruction for all of our body's functions— telling our cells to make proteins that allow for the completion of tasks, and carrying out the jobs of our internal organs. One of the cellular tasks that gets executed most frequently is cell division and replication, which as we all know, is a good thing— because it's what lets our body grow and maintain a supply of healthy cells as older ones die off. But there are times, however, when cell division isn't such a good thing and cancer, as it occurs in humans and other species all over the world, is a prime example of this.

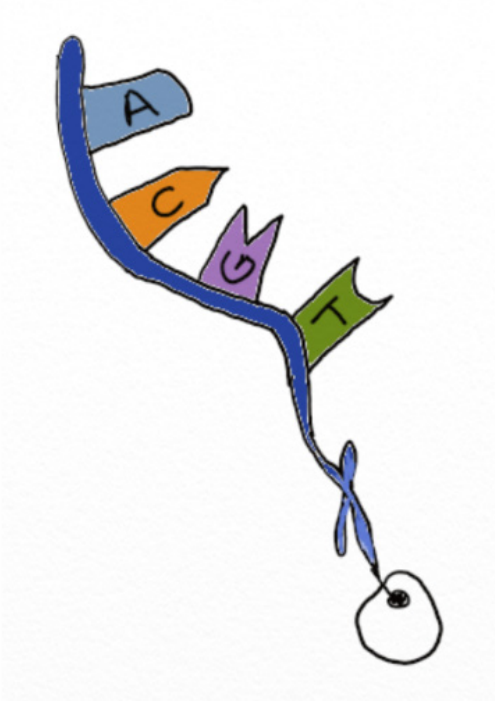


Tumours, the defining characteristic of a cancerous condition, occur in people when a portion of their cells don't go through the process of mitosis, or cell division, properly; dividing too much/too fast in a way that they shouldn't, and building up harmfully in a group. Stopping tumour growth is one of the primary challenges of treating and/or potentially curing cancer, and so many treatments that scientists work on and oncologists (cancer doctors) administer to their patients are substances called antimetabolites, which interfere with the synthesis phase of the cycle of cell growth and division. This is because if tumour cells can't divide, then the tumour can't grow, and the cancer can't get worse. And without any new cells, the tumour — and thereby the cancer — will eventually die out.



So how exactly do thymidylate synthase inhibitors, which are a type of antimetabolite, stop cells from dividing? The key lies in mitosis, and specifically: the very same DNA that enables mitosis to occur.

Every sequence of DNA is made up of a pattern of four distinct nitrogenous bases, (called **adenine, guanine, cytosine and thymine**), that can form discrete pairs and create sections that encode certain genetic information. When cells divide and allow a mass of the body to grow, including tumor mass, they replicate the DNA inside them to give it to the cell copy that's being produced, which means that all of these bases have to be replicated as well. This, my friends, is where the strategy of cancer researchers comes into picture.



In the synthesis of thymine, one of DNA's four bases, there is an enzyme called thymidylate synthase, which makes changes to a handful of other substances (the names of which are long, but the properties of which are fascinating if you decide to do more research) and plays a key role in allowing this base to be replicated. However, a thymine synthase inhibitor binds to this enzyme, and prevents it from being able to interact with the other components of the synthesis process, and thus prevents new thymine from being successfully assembled. What this means is that the DNA itself cannot form, since it won't have all of the parts it needs to construct itself, and without DNA, a cell can't successfully divide or live past division; meaning that any new cells the treated area tries to spawn will die, and become unable to add to or maintain its mass.

Simply put, this treatment is one of many that seek to dampen tumour growth by targeting cell growth, and one of many that might one day lead to a potential cure for certain types of cancer as a whole. The method isn't perfect, but today, it's already in use. So will you and I ever see the day that cancer gets cured for good? Maybe we will, and maybe factors stopping cell growth will play a vital role in that.

Behavioral and Social Sciences

Evaluating the Relationship Between Stress and Impulsivity in a Mathematical Context

Aastha Shah

Nikola Tesla STEM High School

Abstract

Impulsivity largely influences the way individuals diversely process and react to challenges, but could this difference in cognition be caused by an involuntary stress response rather than simply an aspect of personality? This project aims to find a difference in cardiovascular response between highly impulsive and slightly impulsive individuals before, during, and after they perform challenging mathematical calculations. First, impulsivity scores (independent variable) were calculated for all the participants; those with scores above the 80th percentile or below the 20th percentile were separated and 5 were randomly selected from each group. Those 10 participants had their blood pressure and heart rate (dependent variable) taken in minute increments before, during, and after they worked on 20 math problems, to establish a pattern of cardiovascular activity. It was discovered that the variation in systolic blood pressure for the non-impulsive group was very significant with a P value of 0.0001, whereas for the impulsive group, variation was insignificant, with a P value of 0.72. Additionally, there was significantly more variation in heart rate for the less impulsive individuals (with a 24% increase during the task) than the more impulsive individuals (only an 11% increase). These findings imply that a lack of impulsivity may cause a more active stress response, thus establishing a difference in the unconscious ways impulsive and non-impulsive adolescents innately respond to difficulty. These results can open up pathways to reveal larger relationships between stress, biological reactions, and other aspects of cognition.

Introduction

This project seeks to investigate how innate impulsivity may relate to cardiovascular activity, and thus activation of the sympathetic nervous system, which occurs while one is processing and computing mathematical problems.

As times progress, we are becoming more aware of the diversity that surrounds us and more accommodating to those who are different from us. A significant change to note is the shift towards being more inclusive of students with different learning abilities and personality types. ADHD, a learning disability, is characterized by unusually high levels of impulsivity, to the point where it can impair one's day-to-day processing. However, the idea that impulsivity can cause problems with learning and comprehending information does not only apply to those with ADHD: we are now learning that these inherent struggles with academics can manifest in all kinds of students. Not only is a student's ability to do well inhibited, this leads to them having less of an interest in academics, particularly in STEM fields (Marriott et. al, 2019). Countless studies have shown an inverse relationship between impulsivity and academic performance. As a consequence, intelligent individuals are driving themselves away from fields where they could make a difference and change the world for the better.

In addition, studies have determined that impulsivity is correlated with heart rate and blood

pressure; the rate of heart rate increase is lower in highly impulsive individuals (Allen et al., 2019). However, these findings fell short of having full certainty in their conclusion; instead, they mentioned that the results likely would have become more significant with a more potent mental task. This follows a similar line of reasoning that many other works have described, such as Lorenz and Kray's (2019) conclusion that impulsivity, especially in adolescents, is highly dependent upon the context of the situation.

This concept further unfolds when stress is taken into account. Cortisol, the stress hormone, is greatly beneficial in limited quantities, as it increases energy, focus, and oxygen sent to the brain. With this, it may be possible that, because impulsive individuals experience a lessened spike in heart rate (as mentioned earlier), they also experience lessened stress levels in testing scenarios, and that may be explaining their poor academic performance.

With these several findings acting in junction, one cannot help but wonder if the variables involved – a test-taking scenario, cardiovascular activity, and impulsivity – may be related. This research will determine the relationship between these three variables, which may aid in confirming the inherent, unconscious differences in students' ability to perform well on an exam, and introduce the notion that impulsivity is more complex than just a simple aspect of personality.

These studies support the hypothesis that that cardiovascular activity must differ among varying levels of impulsive individuals while doing an academic, mathematical, task. They also support that the resting cardiovascular activity of highly impulsive individuals would be significantly higher/more active than that of their less impulsive peers. While performing the calculations, however, both groups may exhibit similar cardiovascular responses. In doing so, the less impulsive group would experience a more dramatic increase in activity compared to the more impulsive group.

Methodology

The materials used in this study include: (1) a document informing participants and their parent/guardian(s) of the project and what it entails, (2) an online survey collecting emails and impulsivity levels (using BIS-11 scores), (3) a set of 30 algebra problems, printed onto 10 one-sided copies, (4) a timer, (5) an adult supervisor, (6) spare pencils, and (7) one heart rate/blood pressure monitor (BP/HRM).

Only systolic blood pressure levels will be collected, as they correlate strongly with one's mental state and stress levels. For the sake of conciseness, systolic blood pressure will still be described as "blood pressure" or "BP".

Sampling procedure. Participants were 60 students taking Algebra 2 at Tesla STEM High School, across two class periods. They were provided an informed consent form that detailed the motives and methodology of this research. Participants were then instructed to take an impulsivity quiz (the BIS-11) and submit their numerical score via a survey.

Correlational procedure, part 1. After collecting the preliminary impulsivity data, participants with BIS-11 percentiles at/above the 80th percentile and participants with BIS-11 percentiles at/below the 20th percentile were separated into two distinct groups. Five participants were randomly selected from each group (using a random number generator), resulting in a total of ten participants moving on to the secondary study (Fig. 1).

Correlational procedure, part 2. These ten participants were connected to an arm band blood pressure / heart rate monitor (abbreviated BP/HRM) and seated at a desk. Each participant had to spend two minutes with their eyes closed, focusing on slowing their breathing, and a stopwatch was immediately started. The participant's BP/HR was taken twice, once at the start of the deep breathing (time 0:00), and once after one minute of deep breathing (time 1:00). Then, the participant was given the worksheet and was instructed to begin working on the problems. Their BP/HR was taken every two minutes for ten minutes (five intervals, at times 2:00, 4:00, 6:00, 8:00, and 10:00). Once the time ran out, the test was collected, and the participant was told to repeat the two minutes of controlled breathing and their BP/HR was taken twice again, once when the test was taken away (time 12:00) and once after one minute of deep breathing (time 13:00). Once the two minutes of deep breathing were complete (time 14:00), the stopwatch was stopped, and the participant was disconnected from the BP/HRM.

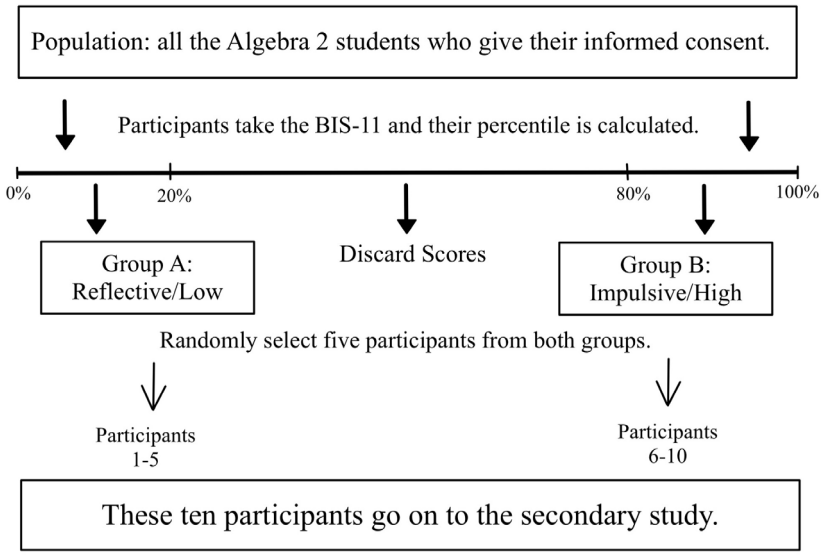


Figure 1. Flowchart of how the ten participants were selected.

Table 1. Average heart rate (BPM) for both the low and high impulsivity groups, separated by their cognitive states.

Impulsivity Group	Heart Rate While Resting (Prior to Task)[1]	Heart Rate While Active (Performing Task)[2] (BPM)	Heart Rate While Resting (After Task)[3] (BPM)
Low	65.13	80.55	73.38
High	77.50	86.73	81.67

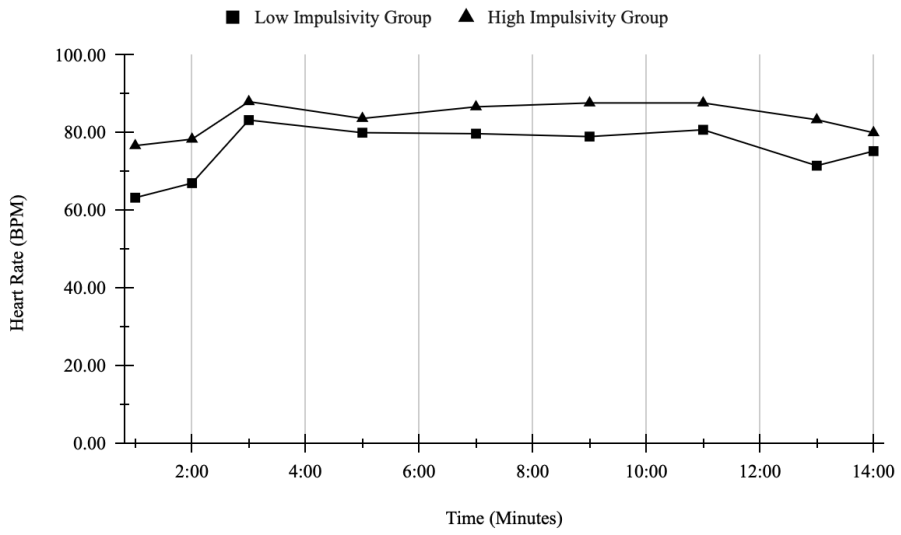
[1] Minute 0:00 to 1:59; a state of focused, deep breathing, prior to being given the problems.
[2] Minute 2:00 to 11:59; a state of computation and mathematical processing. Stress is expected to be seen here.
[3] Minute 12:00 to 14:00; a state of focused, deep breathing, after completing the problems and/or the timer running out.

Table 2. Average systolic blood pressure (mmHg) for both the low and high impulsivity groups, separated by their cognitive states.

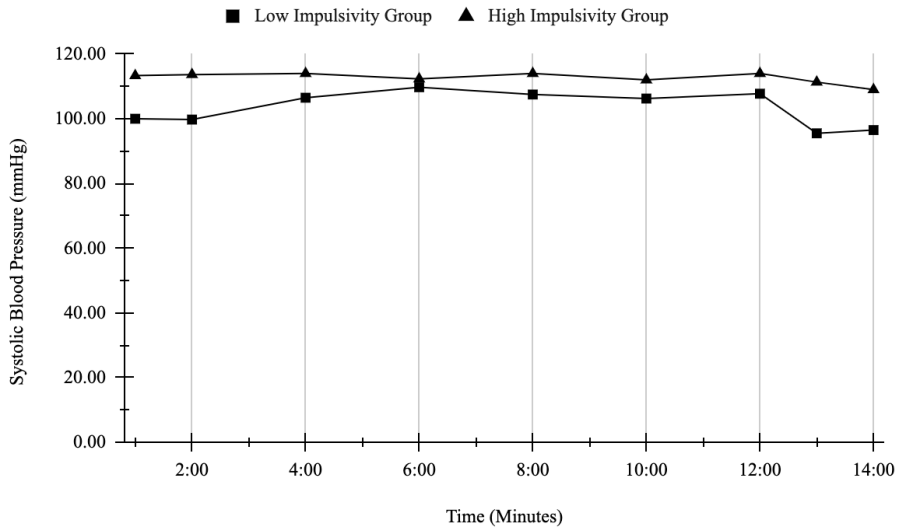
Impulsivity Group	Systolic Blood Pressure While Resting (Prior to Task)	Systolic Blood Pressure While Active (Performing Task)	Systolic Blood Pressure While Resting (After Task)
Low	99.88	107.55	96.00
High	113.50	113.27	110.17

Data Analysis

Results were visualized in two ways: heart rate/blood pressure over numerical time (Fig. 2) and heart rate/blood pressure over categorical time intervals (Fig. 3). Once again, diastolic blood pressure levels were not collected (or analyzed) due to their relative ineffectiveness at predicting stress levels or state of mind. The graphs of heart rate and blood pressure over (numerical) time (Fig. 2) showed a considerable spike at minute 2:00 and a general dip around minute 12:00. The expected results, which were (1) that highly impulsive individuals would have higher resting heart rate/blood pressure levels and (2) that the less impulsive group would exhibit a greater relative increase in heart rate at the beginning of the task period, were successfully exhibited in the data collected. There is also an evident difference between the heart rates/blood pressures of the low and high impulsivity groups' heart rates during the pre-task period, although this difference appears to shrink in the during task period and increases only slightly during the post-task period (Fig. 3); this trend was also hypothesized. Other quantifiable results were determined after further data analysis.

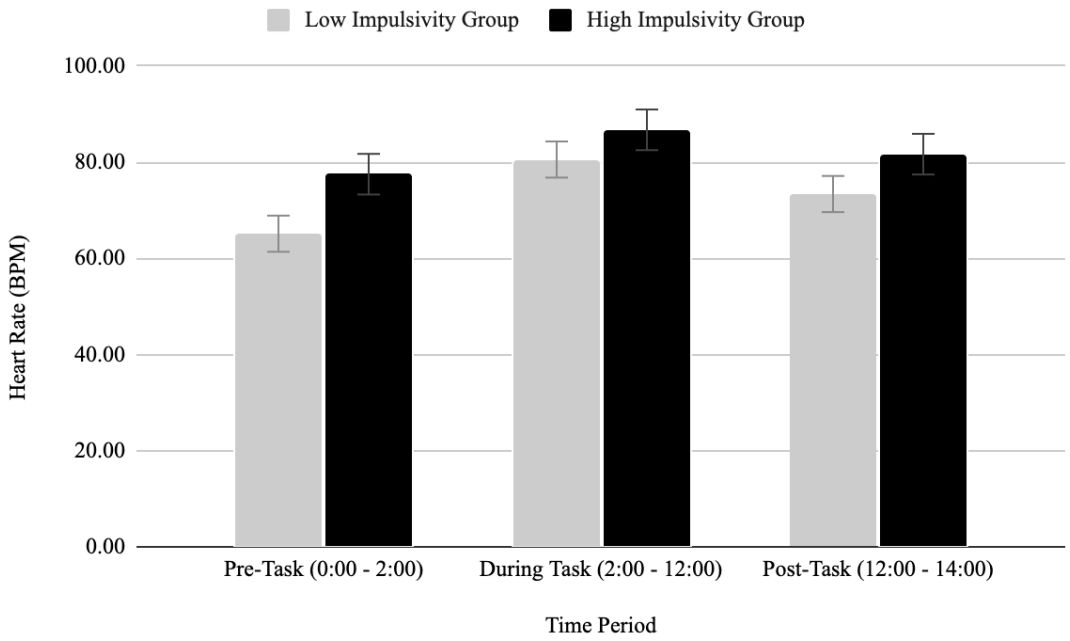


A.

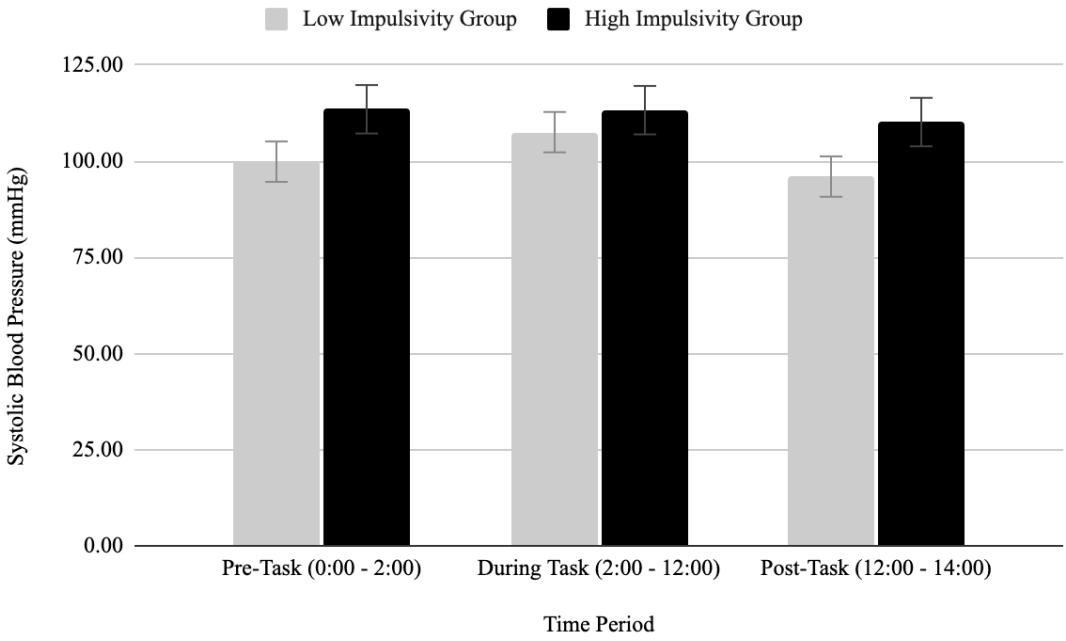


B.

Figure 2. Trends in (A) heart rate and (B) systolic blood pressure over time for both impulsivity groups. Minute 0:00 to 2:00 was a period of deep breathing exercises, minute 3:00 to 12:00 was the duration of the mathematical computations, and minute 13:00 to 14:00 was another period of deep breathing exercises.



A.



B.

Figure 3. Average (A) heart rate and (B) systolic blood pressure in each situation for both impulsivity groups.

Determining statistical significance. Data was analyzed for within-group differences (comparing pre-task, during-task, and post-task heart rate/blood pressure for both the low and high impulsivity groups) and between-group differences (comparing the low and high impulsivity groups' heart rate/pressure for the pre-task, during-task, and post-task time intervals), and P values were calculated to determine statistical significance. All significance was determined at the 95% confidence level.

A one-way ANOVA was conducted to determine if there was a statistically significant within-group difference, with three categories for the three different time periods. For heart rate, the low impulsivity group had a P value of less than 0.00001, and the high impulsivity group had a P value of 0.04207. Both these values are statistically significant; this implies that there is variation between the pre-task, during-task, and post-task heart rates in both groups. For systolic blood pressure, the low impulsivity group exhibited statistical significance ($P = 0.0001$) while the high impulsivity group did not ($P = 0.72$). From this, it is observed that there is significant variation in blood pressure for less impulsive individuals, while there may not be any variation in more impulsive individuals.

Next, an unpaired T test was used to determine if there was a significant difference between the low and high impulsivity groups (each of which was its own group in the t-test). For heart rate, the difference between the low and high impulsivity groups' heart rates throughout the study was statistically significant ($P < 0.00001$, $P = 0.006$, and $P = 0.001$ for the pre-task, during-task, and post-task differences). The difference in systolic blood pressure across the entire study was also statistically significant ($P = 0.006$, 0.008, and 0.014 for each time period, respectively).

Percent change. To assess the specific variation in heart rate and blood pressure for both the low and high impulsivity group, three percent changes were calculated for each impulsivity group, one to assess the increase as the task began, another to assess the decrease as the task ended, and the final one to assess the overall change in heart rate/blood pressure.

The low impulsivity group exhibited a heart rate increase of 23.6% and a decrease of 8.9%. The high impulsivity group, on the other hand, exhibited a heart rate increase of 11.9% and a decrease of 5.8%. Reflecting the other results and the prediction, the low impulsivity group experienced a more drastic change in heart rate, but it appears that this group also took relatively longer for their heart rate to decline back to their baseline.

For systolic blood pressure, the low impulsivity group exhibited the expected pattern of a rise and then a fall (increase of 7.7%, decrease of 12%), while, unexpectedly, the high impulsivity group only decreased (with a net decrease of 2.9%). It is interesting to note how the more impulsive group's blood pressure was entirely unaffected by the stressful task; this may relate to the lower variation in heart rate exhibited by this group as well.

Results

Heart rate. As seen across the various data analysis procedures, the highly impulsive group experienced less variation in heart rate compared to the less impulsive group. This is best explained through the extreme significance ($P < 0.0001$) of the variation in the non-impulsive group, but the moderate significance ($P = 0.04$) of the variation in the impulsive group, as well as the significantly higher ($P = 0.00003$) percent increase in the less impulsive group. Because the low impulsivity group exhibited a lower baseline heart rate, this increased spike resulted in the two groups' heart rates becoming approximately even (Fig. 2A, Fig. 3A). These two deductions are in line with the prediction, which stated that the low impulsivity group would have a lower resting heart rate, but a bigger spike than that of the high impulsivity group; therefore, heart rate behaved exactly as predicted.

Blood pressure. To begin, there was a significant difference (P = between the two impulsivity groups' systolic blood pressure throughout the entire study, which was hypothesized (and relates to the pattern in heart rate). Moreover, the systolic blood pressure exhibited a pattern of increasing (8%) and then decreasing (12%) for the less impulsive group, but it only decreased (3% in total) for the more impulsive group. While this technically agrees with the prediction (in the sense that the less impulsive group displayed a greater degree of variability), it was unexpected for the impulsive

group to exhibit no spike or significant variance in systolic blood pressure. This may provide further evidence towards the hypothesis that impulsive individuals have a decreased stress response in a testing environment, as it appears to be an extreme result (considering it is a lack of a stress response). However, it still provides necessary support for the desired conclusion.

Discussion

The purpose of this research was to establish a relationship between impulsivity and cardiovascular activity levels in an academic test-taking setting. It was hypothesized that (1) impulsive individuals would have a higher resting heart rate/systolic blood pressure, (2) the heart rates and systolic blood pressures of both groups would be similar while they were performing the calculations, and (3) less impulsive individuals would exhibit a higher degree of variability (increase/decrease) in their heart rate and systolic blood pressure. The trends of heart rate and systolic blood pressure (Fig.2, Fig.3), as well as the quantitative statistical analysis conducted, provide support for those statements.

The high impulsivity group exhibited a significantly greater heart rate ($P = 0.00001$, 0.007, and 0.001) and systolic blood pressure ($P = 0.006$, 0.008, and 0.015) throughout the duration of the study. Fig.2 and Fig.3 show the relative heart rate/blood pressure similarity between the two groups, and the statistical analysis points towards the conclusion that less impulsive individuals do exhibit a greater degree of heart rate variability compared to their more impulsive peers.

While this study attempted to remove confounding variables with random sampling, the small sample size may have resulted in a slight skew. Perhaps doubling or tripling the sample size to 20 or 30 would create a more generalizable result. Additionally, the monitor used to measure heart rate and blood pressure was not medical-grade; this may have caused a few minor errors in the measurements. Finally, the time of day the study was conducted varied from participant to participant (either before school or after school), and while this was random for the most part, time of day does have an influence on blood pressure readings. Keeping this variable consistent may provide a more accurate reading and reduce confounding.

This study introduced the possibility that impulsivity, previously considered an aspect of personality, extends deeper within our unconscious nervous system, questioning the previous deduction that impulsive students simply aren't as academically competent. Instead, impulsivity may cause an inherent difference in the way individuals interpret, process, and react to information. The decreased variability in heart rate and systolic blood pressure in impulsive individuals implies that impulsivity lowers one's stress response and prevents the heightened functioning that cortisol usually provides to an individual. These findings can extend beyond the field of psychology, adding further incentive to decrease the use of exams in academics and introducing the concept that personality and identity can influence the way our body reacts to certain stimuli. Overall, the results of this study open up a plethora of doors towards further research involving the relationship cognition, reactivity, and other aspects of the self.

Acknowledgements

I'd like to thank Dr. Suzanne Mitchell, my mentor, and Ms. Allender, who guided me through the entire process.

References

Allen, M. T., Hogan, A. M., & Laird, L. K. (2009). The relationships of impulsivity and cardiovascular responses: The role of gender and task type. International Journal of Psychophysiology, 73(3), 369–376. https://doi.org/10.1016/j.ijpsycho.2009.05.014

ČAvojová, V., & Mikušková, E. B. (2015). Does Intelligence Predict Academic Achievement? Two Case Studies. Procedia - Social and Behavioral Sciences, 174, 3462–3469. https://doi.org/10.1016/j.sbspro.2015.01.1019

Ismaeel, D., & al Mulhim, E. (2021). The influence of interactive and static infographics on the academic achievement of reflective and impulsive students. Australasian Journal of Educational Technology, 147–162. https://doi.org/10.14742/ajet.6138

Li, M., Lauharatanahirun, N., Steinberg, L., King-Casas, B., Kim-Spoon, J., & Deater-Deck-

ard, K. (2019). Longitudinal link between trait motivation and risk-taking behaviors via neural risk processing. Developmental Cognitive Neuroscience, 40, 100725. https://doi.org/10.1016/j.dcn.2019.100725

Lorenz, C., & Kray, J. (2019). Are Mid-Adolescents Prone to Risky Decisions? The Influence of Task Setting and Individual Differences in Temperament. Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.01497

Marriott, L. K., Coppola, L. A., Mitchell, S. H., Bouwma-Gearhart, J. L., Chen, Z., Shifrer, D., Feryn, A. B., & Shannon, J. (2019). Opposing effects of impulsivity and mindset on sources of science self-efficacy and STEM interest in adolescents. PLOS ONE, 14(8), e0201939. https://doi.org/10.1371/journal.pone.0201939

Mathias, C. W., & Stanford, M. S. (2003). Impulsiveness and arousal: heart rate under conditions of rest and challenge in healthy males. Personality and Individual Differences, 35(2), 355–371. https://doi.org/10.1016/s0191-8869(02)00195-2

Mathias, C. W., Stanford, M. S., Liang, Y., Goros, M., Charles, N. E., Sheftall, A. H., Mullen, J., Hill-Kapturczak, N., Acheson, A., Olvera, R. L., & Dougherty, D. M. (2018). A test of the psychometric characteristics of the BIS-Brief among three groups of youth. Psychological Assessment, 30(7), 847–856. https://doi.org/10.1037/pas0000531

Schwabe, L., Wolf, O. T., & Oitzl, M. S. (2010). Memory formation under stress: Quantity and quality. Neuroscience & Biobehavioral Reviews, 34(4), 584–591. https://doi.org/10.1016/j.neubiorev.2009.11.015

Tan, T. X., Liu, Y., Damjanovic, V., Ledford, E., Li, G., & Li, Y. (2021). Inattention, hyperactivity/impulsivity, and academic competence: Findings from three cohorts. British Journal of Educational Psychology, 92(1), 82–104. https://doi.org/10.1111/bjep.12439

Vigil-Colet, A., & Morales-Vives, F. (2005). How Impulsivity is Related to Intelligence and Academic Achievement. The Spanish Journal of Psychology, 8(2), 199–204. https://doi.org/10.1017/s1138741600005072

Combatting Depression using the Gut-Brain Axis via Probiotic Treatments

Divya Singh

Nikola Tesla STEM High School

Abstract

In 2017, around 264 million people globally experienced or were experiencing depression. What if there was a way to regulate depression levels with a simple probiotic? The research problem in question is focusing on how adding probiotics to the gut microbiome can be utilized to minimize depression levels. This study also investigated the placebo effect to see if depression levels could be lowered without probiotic treatment. Over the course of 2 weeks, participants ingested either a probiotic (independent) or placebo pill (control) daily; on day 14, participants completed the Beck's Depression Inventory to obtain self-reported levels of depression. At day 19, 5 days after the conclusion of administering the placebo or probiotic pill, participants took the Beck's Depression Inventory a third time. The statistical results of the experimental group showed a p-value of 0.006 indicating that the experimental results are statistically significant and can be applied in further studies. Overall, the experimental group displayed a 36% decrease in self-reported depression levels from day 0 to day 19, compared to the control group. The experiment showed that continuous use of probiotics can assist in reducing depressive symptoms and can produce lasting effects that aid in maintaining mood and depressive tendencies. Further studies may include conducting an extended study, over the course of 6-8 weeks, that studies how long the effects of the probiotics last.

Introduction

This research focuses on how the gut microbiome can be utilized to minimize depression levels within the human brain using probiotic treatments.

The COVID-19 pandemic has shown record-breaking levels of depression within adults and teens, affecting more than 30% of the USA's population (Pikul, 2021). The pandemic continues to cause significant changes in people's lifestyles impacting their physical and mental health. As the world is slowly returning to the pre-pandemic conditions, it is important to reduce and restore the depression levels within adults worldwide as well. Theoretically, antidepressants and a healthier lifestyle is the solution to this problem, but what if there was a cure with less side-effects and more accessibility? Often, phrases such as "gut feeling" and "butterflies in my stomach" are used without people knowing the scientific reason why these feelings occur. A gut feeling can be described as nausea, stomach pain, or twinges. The purpose of this study is to understand the gut-brain connection more in-depth, and learn how reliable the body's "second brain" is. For more context, the gut is responsible for controlling digestion and excreting waste, as well as communicating with the brain via the enteric nervous system. This system consists

of millions of nerve cells that line the digestive tract. When emotions are triggered by psychosocial factors like loneliness and work environment, this can cause feelings of anxiety, depression, and stress to manifest and activate the enteric system's response. One study by Ghannoum et al., 2021, explained that around 40% of depressed patients do not respond to initial antidepressants treatments and 20% never show any reduction of depressive symptoms even after use of multiple drugs. After a look through various studies discussing the link between COVID-19 and depression, the microbiome (specifically, Gut-Brain Axis) was the answer to a healthier life since probiotics rebalance the gut through the introduction of new bacterial strains. Another metanalysis study by Wallace et al., 2017 pointed out that among the many uses of probiotic supplementation for better mood effects that Lactobacillus casei was the most frequently used strain. The study mentioned that participants using probiotics start showing effects after three weeks of continuous supplementation. Although there is sufficient research on whether probiotics help reduce depression levels in adults/teens, this investigation will specifically focus on whether treatment effects will last after treatment or solely through continued treatment (Noonan et al., 2020), which has not been done before.

Methodology

In this study, participants were collected as they showed interest based on flyers (Figure 1) posted in various clinics around the Seattle area. They were required to sign an informed consent form after being debriefed on the purpose and procedure of the experiment. All participants were required to show proof of COVID-19 vaccination in the initial "interest" form to qualify for the study. After around 22 participants showed interest, using a random number generator, participants were assigned a number from 1-22. Every other number was assigned to treatment group 1 (probiotic pill) (1,3,5,…) and all remaining participants were assigned to treatment group 2 (placebo pill). To ensure single blindness, both treatments were disguised to look the same and only one person knew which treatment was administered to each group.

Table 1. Measurements of criteria from Beck's Depression Inventory collected for each participant. Population size, $n = 45$.

Beck's Depression Inventory Criteria	Pre-Treatment	Post-Treatment	No Continued Treatment
Severity of Depression (minimal/mild/moderate/severe)	xxx	xxx	xxx
Cognitive-Affective Depressive Symptoms (1-44)	xxx	xxx	xxx
Somatic Depressive Symptoms (1-24)	xxx	xxx	xxx

All participants were asked to complete a survey containing a series of questions about their health and lifestyle. They were asked to fill out the Beck Depression Inventory (Figure 2) from which results will be collected and recorded (Table 1). Each participant was asked to take one tablet of their respective treatment, twice a day, over the span of two weeks, and document any significant changes in their mental/physical wellbeing (i.e., diarrhea, nausea, lethargy). After two weeks of ongoing treatment, participants were asked to take the survey and Beck Depression Inventory again, and results were recorded (Table 1).

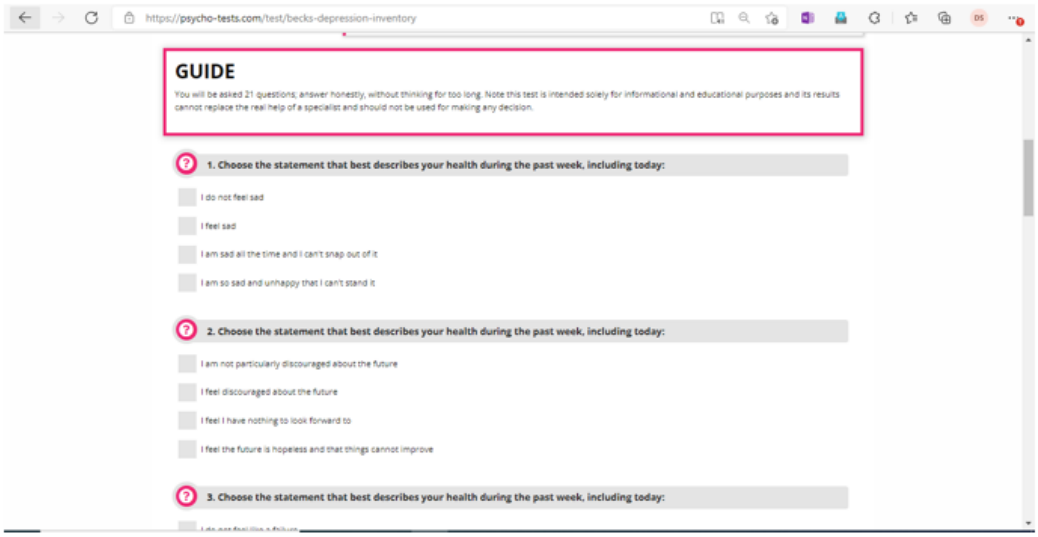


Figure 2. Screenshot of first few questions to online Beck's Depression Inventory. Questions noted will address the three data subtypes in Table 1. Photo courtesy of psychotests.com.

Figure 1. The recruiting flyer that will put out to initiate participant recruiting. Will be sent out in combination with emails advertisements and word of mouth. QR code will be linked to interest form to know if they qualify for participation.

Quick checkups conducted on participants to monitor their physical/mental health. The first checkup (day 0) was before treatment began, then after treatment ended (day 14), then after one week of discontinued treatment (day 19). The study was extended one week for participants to record (via survey) any symptoms or changes in themselves without continuous treatment to see if there were lasting effects of the probiotics. Once all the data was collected, each treatment group's data was averaged and recorded in the data tables to compare results. All observational data was inputted into a separate document and categorized based on treatment group. From this, we could prove that probiotics can be used as a lasting treatment to reduce depression levels using experimental statistical procedures.

For data analysis, we plotted the Beck's Depression Inventory scores for each participant into an Excel spreadsheet. Then, determined the sample mean, sample standard deviation based on the sample sizes. Based on this, the t-score was calculated for the data using the t-score formula ($t=(\bar{x}-\mu)/\sqrt{(s^2/n)}$) and the critical t-score was identified based on t-score table (Figure 3) for each treatment group. The calculated t-score was compared to the critical t-score, and the calculated t-score was greater than the critical t-score then the sample mean was statistically different than the population mean. An additional single-factor ANOVA test was conducted for each treatment group, and it revealed a p-value of 0.006 for the experimental (probiotic) group and a p-value of 0.09 for the control (placebo) group. Percent decrease in each treatment group's data was calculated by taking initial sample mean and subtracting it from the final sample mean of Beck's scores and dividing it by the initial mean.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figure 3. Full t-table to determine critical t-values based on calculations of sample mean and standard deviation. Photo courtesy of sjsu, faculty.

Conclusion

The results of the study revealed that the initial hypothesis – continuous & discontinuous use of probiotics can effectively lower depression levels within adults and warrants further investigation – is supported. The data showed that the experimental group displayed a downward trendline – proving that probiotics are effective. The control group exhibited raised depression levels after the discontinuation of treatment due to the placebo effect. Statistical results from the single factor ANOVA test conducted on both treatment groups produced a p-value of 0.006 (experimental) and 0.09 (control). This indicates that results are not significant for control group, and we fail to reject the null hypothesis that the placebo will not provide lasting effects on mood. As for the experimental group, results indicated significance and the rejection of the null hypothesis that probiotics don't provide any lasting effects on mood. The percent decrease from day 0 to day 19 showed a 36% decrease in depression levels for the experimental group and a 3% increase in depression levels for the control group. A few limitations of the study include a limited sample size (single factor ANOVA could not be conducted for each factor from Beck's Inventory), not clinically depressed participants (results could not be generalized to that population), and positive reporting bias (with self-reported scores, participants may have felt pressured to make them seem "less depressed"). Future studies might include a longitudinal study over the period of 4-6 weeks of continuous treatment and 2-3 weeks of discontinued treatment to truly measure exactly how long discontinued effects last.

Bibliography

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). *The gut-brain axis: Interactions between enteric microbiota, Central and Enteric Nervous Systems.* Annals of gastroenterology. Retrieved November 5, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4367209/.

Ghannoum, M. A., Ford, M. K., Bonomo, R. A., Gamal, A., & McCormick, T. S. (1AD, January 1). *A microbiome-driven approach to combating depression during the COVID-19 pandemic.* Frontiers. Retrieved September 3, 2021, from https://www.frontiersin.org/articles/10.3389/fnut.2021.672390/full.

Harmer, C. J., Duman, R. S., & Cowen, P. J. (2017, May). *How do antidepressants work? New Perspectives for Refining Future Treatment Approaches.* The lancet. Psychiatry. Retrieved September 30, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5410405/.

Karakula-Juchnowicz, H., Rog, J., Juchnowicz, D., Łoniewski, I., Skonieczna-Żydecka, K., Krukow, P., Futyma-Jedrzejewska, M., & Kaczmarczyk, M. (2019, August 31). *The study evaluating the effect of probiotic supplementation on the mental status, inflammation, and intestinal barrier in major depressive disorder patients using gluten-free or gluten-containing diet (SANGUT study): A 12-week, randomized, double-blind, and placebo-controlled clinical study protocol.* Nutrition journal. Retrieved October 22, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6717641/.

Lee, C.-H., & Giuliani, F. (2019, July 19). *The role of inflammation in depression and fatigue.* Frontiers in immunology. Retrieved October 13, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6658985/.

Noonan, S., Zaveri, M., Macaninch, E., & Martyn, K. (2020, December 1). *Food & Mood: A review of supplementary prebiotic and probiotic interventions in the treatment of anxiety and depression in adults.* BMJ Nutrition, Prevention & Health. Retrieved September 10, 2021, from https://nutrition.bmj.com/content/3/2/351.

Pikul, C. (2021, October 5). *Depression rates tripled and symptoms intensified during first year of COVID-19.* Brown University. Retrieved January 25, 2022, from https://www.brown.edu/news/2021-10-05/pandemic-depression

Wallace, C. J. K., & Milev, R. V. (2021, February 15). *The efficacy, safety, and tolerability of probiotics on depression: Clinical results from an open-label pilot study.* Frontiers in psychiatry. Retrieved November 12, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7917127/.

Development of the Educator Implicit Association Test

Praharsha Manda

Nikola Tesla STEM High School, WA

Abstract 1.0

Although racism against people of color has reduced significantly, in this modern world, it still exists, affecting healthcare, workplaces and even classrooms. On most occasions, these actions are not overt, but they are the implicit biases that act in the classroom. Implicit biases are unconscious favoritism or prejudice towards a certain group of people. This project seeks to determine whether educators can get accurate results about their implicit biases from a test measuring their association level. The purpose of creating the Educator IAT is to simulate the stimulus that teachers have to react in, this hopefully can lead to more accurate results. The Educator IAT uses the reaction time of the teacher to solve for a D-score, the final statistical factor in determining the association level strength. Three teachers completed the Educator IAT, Race IAT and the self-reported biases to report their results. The results of the Educator IAT were compared to determine the validity of the test which resulted in a p-value of ($p = 0.001$) statistical significance. But because the number of participants was so low, it was determined that the results are not statistically significant. By bringing awareness to these biases, teachers hopefully create safe, engaging classroom environments to help enable all students to reach their full potential. This study seeks to identify the implicit bias that educators may have while in a teaching environment by creating a test that can be used by educators to reduce the effects of implicit bias.

1.1 Background on Implicit Bias

Racism is defined as the prejudice and discrimination directed against a person or people on the basis of their characteristics, typically one that is a minority. A recent global situation, the COVID-19 pandemic has unearthed the systematic racism that people of color face in healthcare, in work, and in school.

It was found in May 2020, that African American people were 3.57 times more likely to die from COVID-19 than White Americans (Gross, 2020), possibly due to the lack of accessible healthcare caused by systemic racism. In healthcare, there have been reports of dehumanizing services which affect the ability for African Americans to contact mental health services. In the workplace, microaggressions, or indirect acts of prejudice, are common as many people associate affirmative action with promotions rather than intelligence (Williams, 2014). In schools, it's been found that predominantly non-white communities get $23 billion less than their white counterparts (Director, 2021) leading to less educational opportunities.

1.2 Teachers and Bias

Teachers are facilitators for the growth for a student. They play an important role in student's lives which is why teachers are given the responsibility of treating children fairly despite differences. The racism problem stems from the existence of implicit biases, they tamper with the

classroom environment and affect how the teacher teaches, this majorly impacts the student and possibly slows their education. It has been found that African American students are less likely to be put into gifted-programs and are more likely to be suspended or given detention with white teachers (Will, 2020), a possible impact due to implicit biases acting on the educators.

These unconscious associations can truly determine the student's path in further education and its exactly why further steps have to be taken to reduce these biases. This why the Educator IAT was selected as a possible solution to this problem.

1.3 Background of Problem

This inaction on implicit biases doesn't just add to the racial institutions that exist in America, but it also negatively impacts their students of color. With a higher chance of being accused and suspended, African American students are put at a disadvantage as they miss more school days which leads to less academic achievement. Along with consistent degradation through microaggression, many students of color may believe that they are not capable of higher academic achievement. This plus the lack of opportunities due to lack of funding for schools can lead to lower numbers of students being able to get into higher education and adding to the stigma around students of color and their achievements.

Research shows that taking tests like the Project Implicit's Race Bias Implicit Association Test (IAT) can uncover implicit biases (Greenwald et al., 1998). The IAT is a test that uses reaction time to determine the results, though someone may not be explicitly racist, they may subconsciously categorize certain groups together, making them categorize the two groups together more quickly, which is the methodology behind the IAT.

1.4 Background of Solution

The original IAT was created on the basis that humans naturally create associations between objects, terms and people. Evolutionarily, creating associations between objects or food would've been very helpful in identifying harmful objects or items versus unharmful ones. But in classrooms, if a teacher tends to correlate an African American student with cheating apps or cheating in general, they may be more prone to accusing the student of plagiarism. To create a test that is applicable to a large population, such as teachers, the IAT must test on stimuli that the user can quickly associate with a category. If not, the results may be inaccurate.

When taking the IAT, pictures of the chosen stimuli show up on screen one after another, the user is prompted to categorize each image into its respective group before continuing the test. Ex. An image of a White American would be sorted into the category White Americans, which aligned on the left side of the screen, would be matched with it (Model 2). If the image is incorrectly categorized, a red X mark shows up on screen and the image remains on screen till the correct key is pressed. The number of incorrect answers is not used in the final D-score, but rather the time it took to categorize them. These aspects are important as they prevent practice-based accurate results.

1.6 Project Goals (Condensed)

If the Educator IAT is created and used, then educators should be able to get a better understanding of their level of associations between white students and students of color with cheating and studying apps, as the environment that is simulated is similar to that in the classroom. This, if accommodated into teacher training, can hopefully help reduce the impact of implicit biases and aid students of color.

Methodology

2.1 Stimuli Used in the E. IAT

The new version of the test displays several pictures as stimuli for the teacher, featuring stimuli that the educator should be familiar with. The first and second set of images shown to the user are colored images of people ranging from a young age to young adult, there are twelve such pictures, half of which are pictures of White Americans and half are Black Americans. The third and fourth set

of images are pictures of computer websites and mobile applications that are traditionally used to study and the pictures of computer websites and mobile applications that are traditionally used to cheat.

2.2 Algorithm behind the E. IAT

The test is created using a 7 – block structure and the 'E' and 'I' keys on the keyboard, these two keys were used in other IAT tests, so to reduce differences they are used in this test as well. These blocks are also categorized into 20-image blocks or 40-image blocks, the 20 image blocks represent practice blocks, these allow the user to adjust to the process of categorizing. These blocks are not used when determining the participant's final score. The 40-image blocks are the test blocks and are used in the determination of the final score (see Model 1).

The first block requests the user to categorize cheating app images and studying app images, and the second asks the user to categorize images of White Americans and images of Black Americans. The third block (20 – images) requires the user to categorize White Americans with Studying apps and Black Americans with Cheating apps. The fourth block is the same as the third, but it has a set of 40 images. The fifth block (20 – images) asks the user to categorize the images of Black Americans with the opposite key, if they previously used the 'E' key for the left side, they now use the 'I' key. The sixth and seventh ask the user to categorize White Americans with Cheating Apps and Black Americans with Studying Apps (See Model 1).

2.3 Data Analysis (Condensed)

The data collected after participation was analyzed to show a p-value that is not statistically significant and relation to the scores of the Race IAT (2.4). To find the Educator IAT (E. IAT) score, the *D*-score algorithm is used, this uses the numerical reaction time to return a value that will fall into a certain category – finally determining the result. The data is first cleaned, removing the responses higher than 300 milliseconds and those below 10 milliseconds.

The *D*-score is calculated as follows:

a. Use the data from Block 3, 4, 6, and 7 (Model 1)
   i. Eliminate the trials completed by the user if the latencies are greater than 10,000 milliseconds
   ii. Delete trials with latency less than 400 milliseconds
   iii. Compute the mean and standard deviation of each block
   iv. Compute full standard deviation for all blocks
   v. Find the average of the resulting values
   vi. Compute the difference [Block 6 - Block 3 and Block 7 - Block 4]
   vii. Divide the difference of the two blocks by the Standard deviation
   viii. The final score is then categorized into a range representing level of correlation: going from little to none, slight, moderate, strong – the response is also dependent on which two categories the participant more strongly correlates.

2.4 Collection of Participants

There was a total of 3 teachers recruited through emails to participate, all of whom teach in the Lake Washington School District. Before participating, the teachers were asked to fill out an informed consent form and were debriefed of the experiment. Those who volunteered were chosen using a random number generator and a single-blind procedure, this allows for anonymity after contact with the participants. The participant took three tests, in the order of the self-reported bias, the Race IAT and then the Educator IAT.

2.5 Experiment

Before the participants took any test, they were told the following message: "This is a research study searching to find information about implicit biases in educators, the first test that you will be asked to take is a standard test used in social cognition research; the second test that you are asked to take is experimental and created to aide in determining implicit associations in teachers. As a participant you have the option throughout the entire procedure to opt out of participating. Your information will not be kept after you fill the Informed Consent form. These forms are present to confirm your participation and will not be kept after filled. The study requires you to have a computer that has access to the internet and can access webpages. You will be asked to download an application called Inquisit Lab to run the second experimental test. As well as opening one website that will take an estimated 10-15 minutes to complete."

Inquisit Lab was used to collect data and return the results to the participant. The results of the Educator IAT and additional survey information was sent to an online database where they are stored and could later be accessed. The test is hosted on the Inquisit Player which the participant must download to access, this is mentioned in the debriefing.

The teacher was asked to complete the standard Race IAT, supported by the Project Implicit website, that determines how much the user associates race with positive and negative terms. They were also asked to complete a self-reported bias survey; this asks the participant of their warmth/cold-ness towards African Americans and White Americans. These two tests were used during the analysis of the data to determine the accuracy of the E. IAT.

Results

The participants were contacted over the course of February and March. The results showed less correlation between the scores of the Educator IAT and the Race IAT though on its own the Educator IAT scored a p-value of 0.00. But due to the small sample size, this value is not statistically significant. This can lead to the assumption that if there is enough data, the test can be used after being improved to be more applicable.

Sample Data for a participant would include their inputted personal data and their results. Below is actual data from the data collection (doesn't include personal data on ethnicity and subjects teaching):

implicit_cheating_iat	implicit_race_iat	explicit_black_bias	explicit_white_bias
-0.194	0.947	0.67	0.00
-0.137	-0.087	0.00	0.00
0.287	0.389	0.67	0.67

Under the Educator IAT (implicit cheating), the further the *D*-score is from 0 in the positive direction, the more association of Black Americans with cheating and White with studying, vis versa in the more negative direction. Under the Race IAT, the further away from 0 in the positive direction means stronger association with White Americans and words with good connotations and Black Americans with words of bad connotations, vis versa in the negative direction. For the both explicit measures, the more positive the more warm the participant feels towards the respective race mentioned.

3.1 Confounding Variables with Participation

There were confounding variables to the data collection. As the targeted group for the study was teachers in different schools, the number of responses would automatically be less in frequency. Additionally, with Covid-19 also having a big impact on the curriculum and the changes around graduation requirements, many of the teachers and principals didn't reply to participate.

The popularity of each application may have also impacted the association times which would in turn also impact the scores for each educator. To examine this, there can be a post-study done to survey teachers about their biases that will check their familiarity with the apps.

3.2 Conclusions

With the stimuli, possibly keeping the applications on the sides of the screens to make the test less reliant on the participants knowledge of the applications either cheating or studying. Changing the application stimuli to be more relatable to an older group who may not recognize these apps, adding pictures of objects more associated to cheating/studying during a test, like hidden notes or pictures of people studying.

3.3 Discussion

This test can be used in parallel to empathy training done by students who are aspiring to be teachers or those who are educators/tutors to learn about their own implicit biases. This can allow for people to look more into their own biases if any and can be more aware of them if they do exist. Hopefully if this is more commonly used, the rates for children of a certain skin color being discriminated against due to implicit biases can reduce significantly in the future leading to an increase in educational opportunities.

This project can be applied to not only teachers but to several parts of the education administration and the images can be replaced by different stimuli that can then also be compared to the Race Attitude IAT, allowing for an analysis of the importance of the stimuli.

Block (Step)	Number of images	Type	Categories and Sets (left hand key)	Categories and Sets (right hand key)
1	20	Practice	Students of African descent	Students of European descent
2	20	Practice	Cheating images	Studying images
3	20	Practice	Cheating Images and students of African descent	Studying images and students of European descent
4	40	Test	Cheating Images and students of African descent	Studying images and students of European descent
5	20	Practice	Students of Europeans descent	Students of African descent
6	20	Practice	Cheating images and students of European descent	Studying images and students of African descent
7	40	Test	Cheating images and students of European descent	Studying images and students of African descent

Model 1: The diagram displays the program of the Educator IAT; the 7-step design was originally created by the Project Implicit IAT, it has been adopted to fit into the procedure of this study. The result uses only Step 4 and Step 7 as Steps 1 to 3, 5 and 6 acts as practice sets for the participant.

Survey 1	2 Multiple Choice Questions	check box below
#1	How do you feel towards students of European descent:	
	-very warm	O
	-moderately warm	O
	-warm	O
	-neither warm nor cold	O
	-cold	O
	-moderately cold	O
	-very cold	O
#2	How do you feel towards students of African descent:	
	-very warm	O
	-warm	O
	-neither warm nor cold	O
	-cold	O
	-very cold	O

Survey 1: This survey will be used at the very beginning of the test. It takes the answers by participants in the form of multiple choice, to leave the data quantifiable and will be compared to the test values that the participant will receive

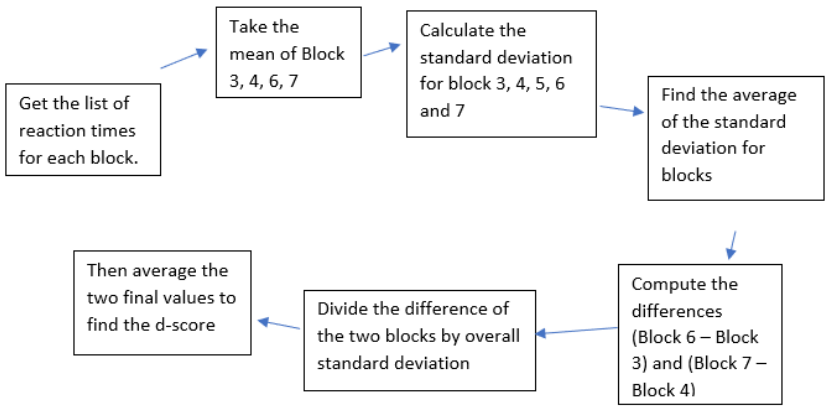


Image 1: a visual representation of calculating the IAT score

After participant data is collected, the t – value will be calculated: $t = [\bar{x} - \mu] / [s / \sqrt{n}]$ this will later be used to calculate the p-value. Below are the values of the variables:

- $\bar{x}$ sample mean
- μ population mean
- s sample standard deviation
- n sample size

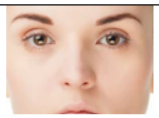
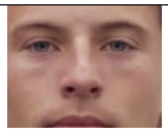
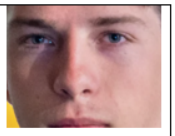
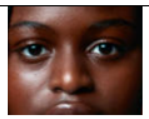
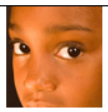
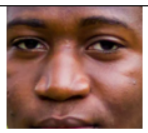
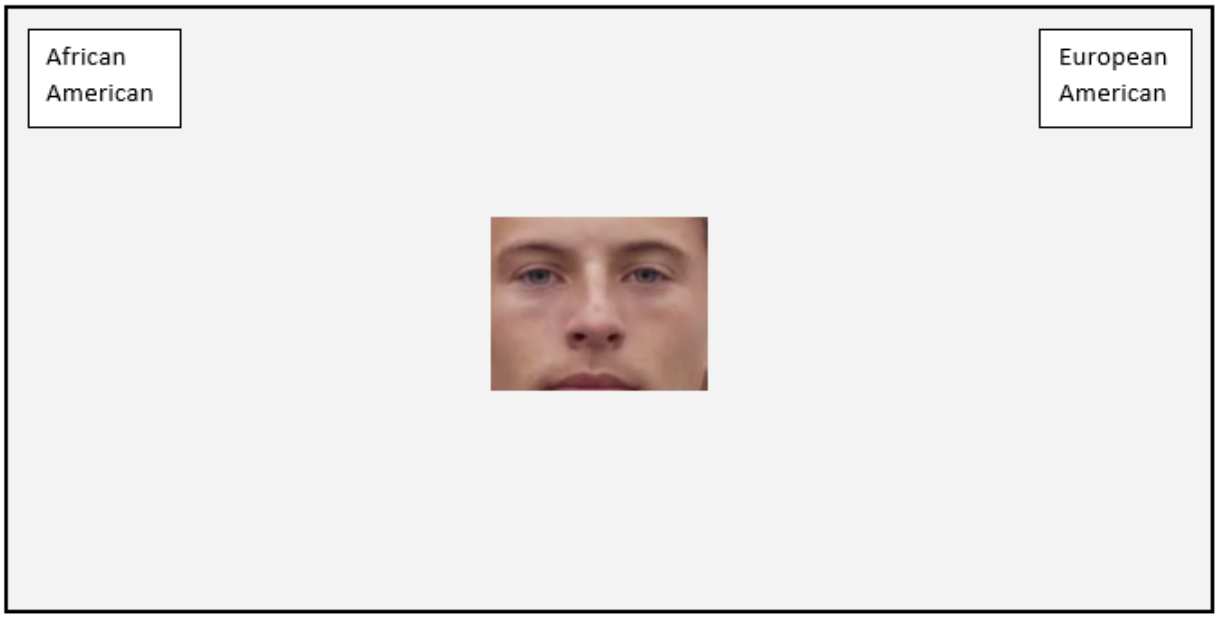
Pictures of European Descent						
Pictures of African Descent						
Pictures of Study Images						
Pictures of Cheat Images						

Table 1: The data table shows the images that will be shown to the educators. These images will show up on the screen of the teacher during all blocks



Model 2: This is the screen layout that is seen by the user while taking the test, it is the same layout for the Race IAT and the E.IAT.

References

Barrett, Nathan, et al. "Discipline Disparities and Discrimination in Schools." *Brookings,* Brown Center Chalkboard, 9 Jan. 2018, www.brookings.edu/blog/brown-center-chalkboard/2017/11/20/discipline-disparities-and-discrimination-in-schools/.

Boys, H. (2001). "Racism and the Administration of Justice. Racism and human rights: Racism and the administration of Justice." Retrieved September 19, 2021, from https://www.hrw.org/legacy/campaigns/race/criminal_justice.htm.

Cvencek, D., Meltzoff, A., Maddox, C., et al. (2019). *Meta-analytic use of balanced ... - UW faculty web server.* Personality and Social Psychology Bulletin . Retrieved February 24, 2022, from https://faculty.washington.edu/agg/pdf/Cvencek&al.PSPB+supplement(in%20press).pdf

Director, J. C. S., Cusick, J., et al. (2021, November 3). *Fighting systemic racism in K-12 education: Helping allies move from the keyboard to the school board.* Center for American Progress. Retrieved January 25, 2022, from https://www.americanprogress.org/article/fighting-systemic-racism-k-12-education-helping-allies-move-keyboard-school-board/

Gilliam, W., Maupin, A., Reyes, C., Accavitti, M., & Shic, F. (2016, September 28). *Do early educators' implicit biases regarding sex and race* ... Yale University Child Study Center. Retrieved February 24, 2022, from https://medicine.yale.edu/childstudy/zigler/publications/Preschool%20Implicit%20Bias%20Policy%20Brief_final_9_26_276766_5379_v1.pdf

Gross, C. P., Essien, U. R., Pasha, S., Gross, J. R., Wang, S.-yi, & Nunez-Smith, M. (2020). Racial and ethnic disparities in population level covid-19 mortality. https://doi.org/10.1101/2020.05.07.20094250

Helm, Rebecca K., et al. "Can Implicit Associations Distinguish True and False ..." *Can Implicit Associations Distinguish True and False Eyewitness Memory? Development and Preliminary Testing of the IATe,* Wiley Online Library, 2017, www.uwyo.edu/psychology/research/can-implicit-associations-distinguish-true-and-false-eyewitness-memory.pdf.

Henkel, E. (2020, December 16). How unconscious bias affects our perceptions. Nonprofit Risk Management Center. Retrieved September 19, 2021, from https://nonprofitrisk.org/resources/e-news/say-what-how-unconscious-bias-affects-our-perceptions/.

KA;, H. R. K. C. S. J. B. (2017). *Can implicit associations distinguish true and false eyewitness memory? development and preliminary testing of the IATE.* Behavioral sciences & the law. Retrieved February 24, 2022, from https://pubmed.ncbi.nlm.nih.gov/28116810/

Lor, K. B., Truong, J. T., Ip, E. J., & Barnett, M. J. (2015, March 25). *A randomized prospective study on outcomes of an empathy intervention among second-year student pharmacists.* American Journal of Pharmaceutical Education. Retrieved February 24, 2022, from https://www.ajpe.org/content/79/2/18

McEntarffer, R. (2020, August 17). Summary – Confronting Bias Through Teaching: Insights from Social Psychology. Nature news. Retrieved September 19, 2021, from https://npjscilearncommunity.nature.com/posts/summary-confronting-bias-through-teaching-insights-from-social-psychology.

Oliver, M. N., Wells, K. M., Joy-Gaba, J. A., Hawkins, C. B., & Nosek, B. A. (2014). Do physicians' implicit views of African Americans affect clinical decision making? *The Journal of the American Board of Family Medicine,* 27(2), 177–188. https://doi.org/10.3122/jabfm.2014.02.120314

Quinn, P. C., Zhang, L., Qian, M., Heyman, G. D., & Lee, K. (2019). *Racial categorization predicts implicit racial bias in preschool children.* Child development. Retrieved February 24, 2022, from https://pubmed.ncbi.nlm.nih.gov/28605007/

Research, H. (2019, March). *The impact of implicit bias training.* www.hanoverresearch.com. Retrieved February 24, 2022, from https://f.hubspotusercontent00.net/hubfs/3409306/The-Impact-of-Implicit-Bias-Training.pdf

Resmovits, Joy. "1 In 4 Black Students with Disabilities Suspended from School." *HuffPost,* Black Voices, 13 Mar. 2014, www.huffpost.com/entry/school-discipline-race_n_4952322.

Will, M. (2020, November 29). *Teachers are as racially biased as everybody else; study shows.* Education Week. Retrieved January 25, 2022, from https://www.edweek.org/teaching-learning/teachers-are-as-racially-biased-as-everybody-else-study-shows/2020/06

Williams, A. (2014). Modern-day racism in the workplace: Symbolic diversity or real change? *FROM SCIENCE TO PRACTICE: ORGANIZATIONAL PSYCHOLOGY BULLETIN*. https://doi.org/10.19099/fstp.081401

Assessment of Yoga Therapy or Mindfulness-Based Therapy Interventions on Symptom Severity and Quality of Life in a Cohort of Adults with Inflammatory Bowel Disease

Aubrey C. Grove

Nikola Tesla Stem High School

Abstract

Inflammatory Bowel Disease is linked to a mind-gut brain connection and non-invasive alternative medicine could be promising to alleviate IBD symptoms and side effects associated with invasive medicines. This project seeks to investigate if Yoga Therapy or Mindfulness Based Therapy significantly improves IBD symptoms and quality of life measures in adults with usual care in a short-term module. Four groups of participants consisted of A - Control/Waitlist, B - Guided Intervention, C - Mindfulness, D – Yoga. Each group participated for four weeks and received a different procedure including readings, yoga or mindfulness videos, or no specific instructions. They submitted IBD symptom severity (fatigue, bowel Symptoms, abdominal and bodily discomfort) forms after each week and two weeks post study. They also submitted Anxiety severity (nervous, worrisome, etc.) forms week 1, 3, and 6. The results did not support my hypothesis because the p-value of 0.806 for group B (guided intervention) and p-value of 0.745 for group A (control) indicate the data is not statistically significant enough to reject the null hypothesis. This experiment did not prove a short-term module is effective in reducing IBD symptoms through alternative medicine, proving it is worth investigating this of area of medicine further in terms of optimal module lengths. A contributing factor to the data not being generalizable is the limited number of participants (two of four dropped out) and the control group being in remission.

Introduction

This experiment will seek to expand our understanding of the potential role of MBT and YT in IBD treatment in form of a short-term module. Inflammatory Bowel Diseases (IBD), consisting of Crohn's disease (CD) and Ulcerative Colitis (UC) are chronic inflammatory conditions. The condition is digestive system based and requires lifelong therapies/care to reduce inflammation. IBD occurs at all ages and debilitating symptoms such as diarrhea, abdominal pain, bloody stools, weight loss, and fatigue are sometimes present. In 2017 it was recorded that 6.8 million people had IBD in form of Ulcerative Colitis and Crohn's Disease, globally (Sudabeh et al., 2020). This demonstrates there is a significant portion of society affected by UC or CD. One major concern is the increase of IBD diagnoses in developing countries without the infrastructure to provide quality/affordable treatment plans. Additionally, other studies suggest IBD particularly in children is rapidly increasing (Benchimol et al., 2014). IBD's affect is increasing which indicates a need for more diverse treatment plans adaptable to socioeconomic backgrounds.

While IBD symptoms are affected by physical factors, including location, extent, and severity, research has shown psychosocial factors, including stress and anxiety, may impact symptom severity, reduce quality of life, and even precipitate disease exacerbations (Purnima et al., 2015; Sun

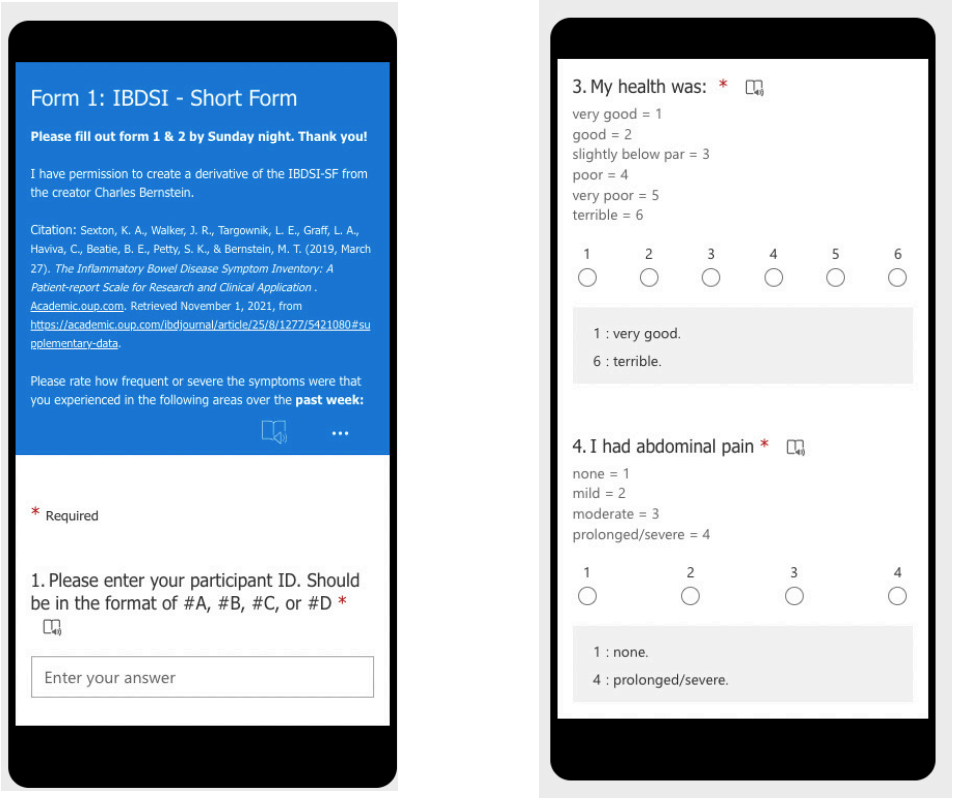
et al., 2019). Gastrointestinal conditions involve brain-gut connections meaning treatment directed towards cognition and mental state will affect gastrointestinal activity (Gaylord et al., 2021). Human brains adjust neurotransmitters released as a result of emotional state. Yoga and mindfulness release neurotransmitters such as GABA, dopamine, endorphins, etc. These helps regulate stress, anxiety, mood, nervous system activity, and pain in the human body (Lymn 2019). Therefore, mindfulness and yoga therapy may be an effective form of adjunctive treatment for reducing/managing IBD symptoms and improving quality of life. For example, there would be potential for reduction in symptoms such as abdominal pain, irregular bowel movement, nausea, level of appetite, fatigue, and joint pain in form of arthritis.

Researchers Steven and Adam (2018) found the, "estimated prevalence of complementary and alternative medicine (CAM) use in patients with IBD is high, ranging between 21% and 60%," showing there is an interest in for CAM in the IBD community. This project seeks to investigate if MBT or YT reduces symptoms severity and improves quality of life in adult patients living with IBD. Further, this project will assess these patients' interest in these therapeutic modalities and experience participating in them. Given the known interest of patients with IBD in CAM and the potential for non-pharmacologic therapies to improve symptom and psychosocial outcomes, this project's purpose is to expand our understanding of the potential role of MBT and YT in IBD treatment. The question explored is: Does Yoga Therapy or Mindfulness Based Therapy significantly improve IBD symptoms and quality of life measures in adults with usual care? Data supports, both adjunctive MBT and YT will significantly reduce IBD symptoms and improve quality of life measures in adult patients with IBD while receiving usual care. This hypothesis is supported not only by research suggesting stress and psychological factors affect IBD-related symptoms but also by MBT and YT have been proven to reduce stress and promote acceptance.

Methodology

This study includes participants over the age of 18 who were able to commit roughly two hours a week for four weeks. Participants were recruited via electronic flyers *(Figure 2.1)* on websites such as Crohn's and Colitis Foundation (CCF). Private GI clinics in the Puget Sound area were also contacted and given the electronic flyers. They also signed an informed consent form prior to starting week 1. This project will be single blind, and only the study supervisor will have access to two passwords protected excel files. One containing the participants emails and another with participant ID. Participants were given a random participant ID via a random number/letter generator *(Figure 3.1)*. The letter corresponded with one of the four groups — A - Control/Waitlist, B - Guided Intervention, C - Mindfulness, D - Yoga). After their groups were assigned each participant completed pre-study surveys (Anxiety & IBDSI-SF) to gather baseline data. Then, emails were sent out weekly with forms (IBDSI-SF and Anxiety form every 2 weeks) to be completed by Sunday afternoon and their weekly plan *(Figure 3.2).* For safety concerns, if they had a physical or mental disability preventing them from participating in the modified YT or MBT program, they were excluded from the study. Participants were required to have access to a device capable of connecting to internet and streaming videos. They completed a 4-week program and completed post-study surveys. After their programs ended each participant was sent a debrief email to explain the purpose of the study *(Figure 3.3).*

Figure 1.1 The pictures depict the IBDSI-SF layout from the perspective of a mobile device. The picture on the left shows a citation for the original form and instructions as to when to complete the form. The picture on the right shows how the questions look (scales), additionally there are three sections with different question categories (Symptoms, Complications, General Health).



The control group did not receive any additional resources and were advised to continue with their current treatment plan (unique for each participant) in addition to filling out the weekly forms. The Mindfulness group watched 2 20–40-minute videos per week from the MBSR program in conjunction with current treatment plans. The Yoga Therapy group watched 2 40-minute videos from University of Chicago's GI Research Foundation in conjunction with current treatment plans. If a participant wasn't able to physically complete certain poses, they were allowed to omit them from their session or solely use the restoration video with less demanding poses. This portion of the procedure was modified for any participant with a physical disability. The Guided Intervention Group spent 30 minutes each week reading 2-3 excerpts from the MBSR program in addition to their current treatment plan. Each group completed the Anxiety form pre- and post-study and the IBDSI-SF form weekly. They also completed an End of Research Study survey post-study. Lastly, the participants were required to consistently space out their sessions with a minimum of 2 days in between (for MBT and YT). After, each form was completed the researcher inserted the data into the corresponding data tables in preparation for data analysis.

In terms of language barriers, emails were sent in English, but participants were able to use online translators. Additionally, the interest form asked which language they would like the forms/surveys to be in. The Microsoft forms had an option to add translations for a wide variety of languages, and this feature was adjusted based on participants' responses. Each group had a specific protocol for translating their resources, which included receiving translated video transcripts, translating readings with online translators, or watching videos with captions in their language. Lastly, for accessible materials participants were encouraged to use towels, books, and pillows in replace of the yoga therapy supplies.

Figure 1.2 Picture of University of Chicago GI Research Foundation's yoga therapy video accessible through YouTube. The yoga therapy group participants will access these videos

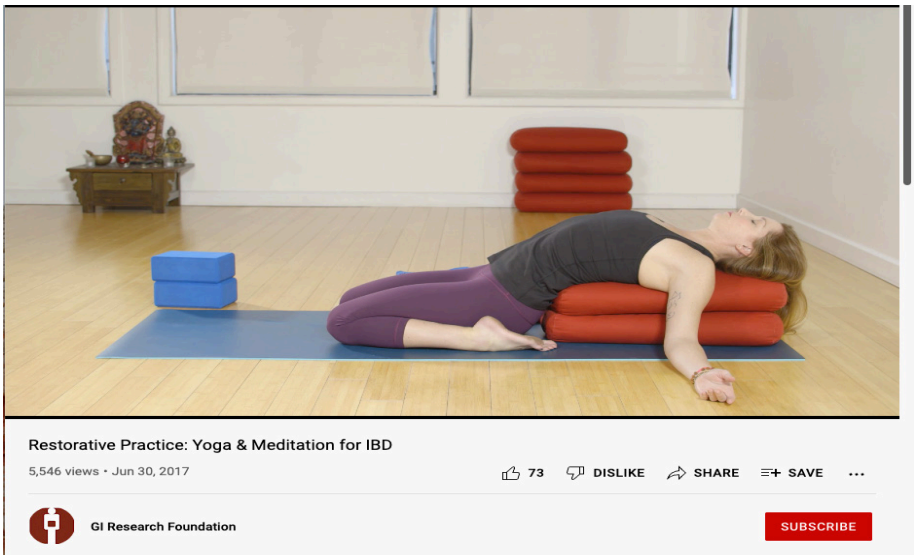
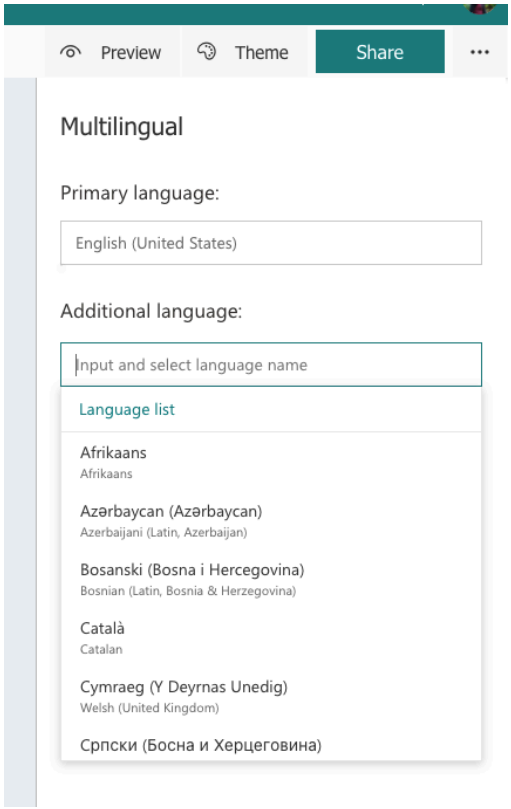
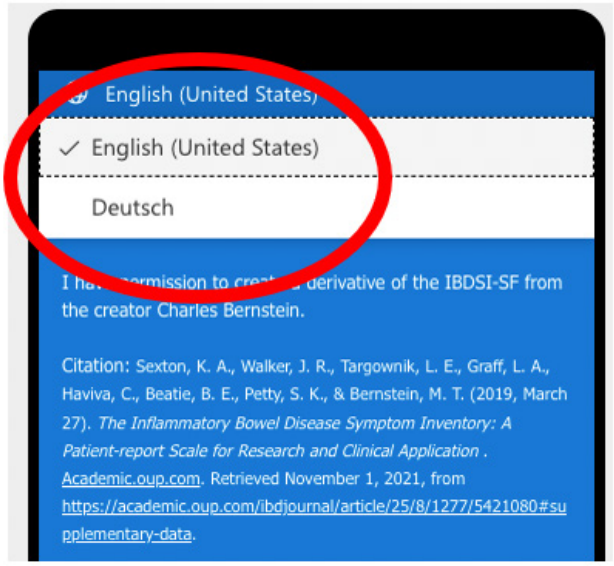


Figure 1.3 This picture shows the search tool integrated in Microsoft forms, used to provide options for a translated version of the form.





Results/Discussion

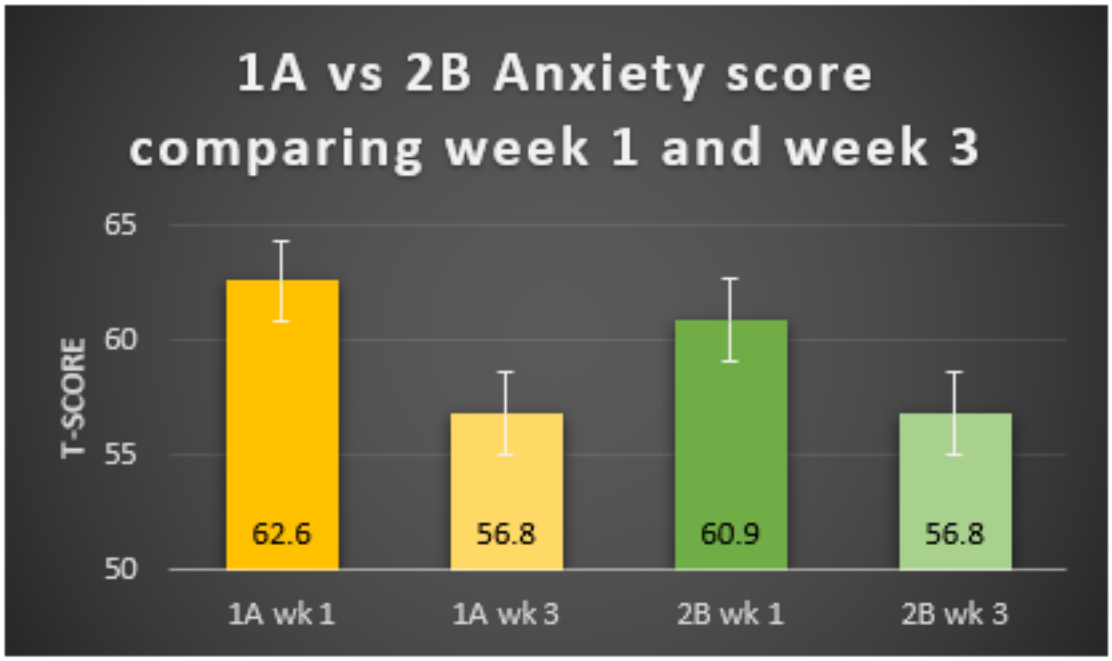
The results from the IBDSI data did not support my hypothesis because the p-value of 0.806 for group B (guided intervention) and p-value of 0.745 for group A (control) indicate that data is not statistically significant enough to reject the null hypothesis. Additionally, the results from the Anxiety scores from week 1 to week 3 had a p-value of 0.480 for the control group and 0.255 for the guided intervention group. These values are also greater than 0.05 indicating the null hypothesis cannot be rejected. This experiment did not produce generalizable results or have a representative population meaning it cannot prove if a short-term module is effective in reducing IBD symptoms through alternative medicine. This shows it is still worth investigating this area of medicine in terms of optimal module lengths with a more generalizable population. One contributing factor to the data not being generalizable is the limited number of participants (two of four dropped out). Another factor that affected results was that the control group participants was currently in remission, meaning their inflammation was at a reduced to unnoticeable level before starting the study. This was apparent in the IBDSI scores when their initial scores were 7 to 17 points lower than the guided intervention group's score.

For further application this project will seek to reach a generalizable number of participants, use an app to house all features related to the study, and lastly collect two additional forms of data through fitness devices: sleep and BPM.

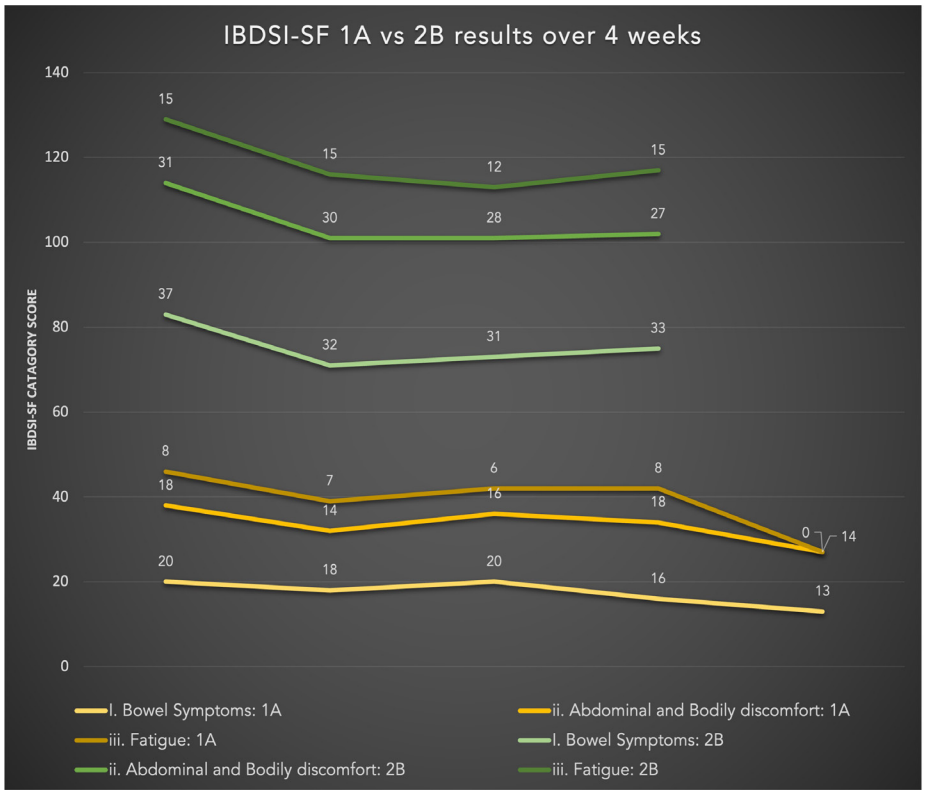
Acknowledgments

This research project was completed with guidance and edits from Dr. Person Hannibal, a Pediatric Gastroenterologist at Seattle Children's Hospital. The participants were sourced through help of Nicole Kittleson from Chron's and Colitis Foundation. Permission to use additional resources were given by GI Research Foundation (University of Chicago – Medicine), Charles Bernstein, and Kiranjit Sran.

Graph 1.1 Anxiety Scores QOL for Control Group (1A) and Guided Intervention (2B) from week 1 to week 3. SE is from 1.7 to 1.8 and t-scores are calculated using raw scores.



Graph 1.2 IBDSI-SF Raw Scores by segment (bowel symptoms, fatigue, abdominal and bodily discomfort) for Control Group (1A) and Guided Intervention (2B) from week 0 to week 4.



References

Alatab, S., Sepanlou, S. G., Ikuta, K., Vahedi, H., Bisignano, C., & Safiri, S. (2019, October 21). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Gastroenterology & Hepatology. Retrieved November 1, 2021, from https://www.thelancet.com/journals/langas/article/PIIS2468-1253(19)30333-4/fulltext.

Gaylord, S. A., Palsson, O. S., Garland, E. L., Faurot, K. R., Coble, R. S., Mann, J. D., Frey, W., Leniek, K., & Whitehead, W. E. (2011, June 21). Mindfulness training reduces the severity of irritable bowel syndrome in women: Results of a randomized controlled trial. The American journal of gastroenterology. Retrieved October 1, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6502251/.

Kavuri, V., Selvan, P., Malamud, A., Raghuram, N., & Selvan, S. R. (2015, November 11). Remedial yoga module remarkably improves symptoms in irritable bowel syndrome patients: A 12-week randomized controlled trial. European Journal of Integrative Medicine. Retrieved from https://www.sciencedirect.com/science/article/pii/S1876382015300561.

Labus, J. S., Bolus, R., Chang, L., Wiklund, I., Naesdal, J., Mayer, E. A., & Naliboff, B. D. (2004). The Visceral Sensitivity Index: development and validation of a gastrointestinal symptom-specific anxiety scale. Alimentary pharmacology & therapeutics, 20(1), 89–97. https://doi.org/10.1111/j.1365-2036.2004.02007.x

Lin, S. C., & Cheifetz, A. S. (2018). The Use of Complementary and Alternative Medicine in Patients With Inflammatory Bowel Disease. Gastroenterology & hepatology, 14(7), 415–425.

Lymn, E. (2019, December 4). How yoga changes your brain. Yoga Medicine. Retrieved January 1, 2022, from https://yogamedicine.com/how-yoga-changes-your-brain/

Purnima Sharma, Gopal Poojary, Sada Nand Dwivedi, Kishore Kumar Deepak; Effect of Yoga-Based Intervention in Patients with Inflammatory Bowel Disease. Int J Yoga Therap 3 January 2015; 25 (1): 101–112. doi: https://doi.org/10.17761/1531-2054-25.1.101

Potter, D. (2021, December 12). Online

MBSR/mindfulness (free). Online MBSR/Mindfulness (Free). Retrieved December 13, 2021, from https://palousemindfulness.com/

Sexton, K. A., Walker, J. R., Targownik, L. E., Graff, L. A., Haviva, C., Beatie, B. E., Petty, S. K., & Bernstein, M. T. (2019, March 27). The Inflammatory Bowel Disease Symptom Inventory: A Patient-report Scale for Research and Clinical Application . Academic.oup.com. Retrieved November 1, 2021, from https://academic.oup.com/ibdjournal/article/25/8/1277/5421080#supplementary-data.

Statistics How To. (2021, November 20). T test (student's t-test): Definition and examples. Statistics How To. Retrieved November 22, 2021, from https://www.statisticshowto.com/probability-and-statistics/t-test/.

Streeter, C. C., Whitfield, T. H., Owen, L., Rein, T., Karri, S. K., Yakhkind, A., Perlmutter, R., Prescot, A., Renshaw, P. F., Ciraulo, D. A., & Jensen, J. E. (2010). Effects of yoga versus walking on mood, anxiety, and brain GABA levels: a randomized controlled MRS study. Journal of alternative and complementary medicine (New York, N.Y.), 16(11), 1145–1152. https://doi.org/10.1089/acm.2010.0007

Yeh, A. M., Wren, A., & Golianu, B. (2017, April 3). Mind–body interventions for pediatric inflammatory bowel disease. MDPI. Retrieved November 13, 2021, from https://www.mdpi.com/2227-9067/4/4/22/htm.

Tables

Table 1. IBD Symptom Inventory – Short Form (IBDSI-SF) before and after.

Participant ID	Participant Group	IBDSI-SF Score before sessions begin	IBDSI-SF Score week 4
1A	Control	46	42
2B	Guided Intervention	83	75
29D	Yoga	56	N/A

Table 2. Participant information.

Participant ID	sex	Current treatment plan	Family income	Relation ship status	Race	age	Education level
1A							
2B	Female	Stelara every 8 weeks & stem cell trial for anal fistulas (pain meds, Imodium, Zofran, JPouch)	70,000	married	Caucasian	37	Bachelor of Science Degree
3C	Female	Stelara and Methotrexate	125,000	single	Caucasian	21	undergraduate
29D	male	Stelara every 8 weeks	225,000	dating	Caucasian	42	Bachelor's Degree

Table 3. Anxiety Score before and after study.

Participant #	Participant Group	Anxiety Raw Score before sessions begin (out of 105)	Anxiety Raw Score week 3 (out of 105)
1A	Control	29	22
2B	Guided Intervention	27	22
3C	Mindfulness	21	n/a
29D	Yoga	33	n/a

Table 4. Completion Rate for entire study.

Group	Total Number of sessions attended	Total Number of sessions missed	Total # of sessions / total # of session attended (round to 3sf)
Wait-list (Control) **sessions = forms*	8	0	100%
Mindfulness (experimental)	0	8	0.00%
Yoga therapy (experimental)	1	8	11.1%
Guided Intervention (control)	4	1	80.0%

Table 5. IBD Symptom Inventory – Short Form (IBDSI-SF) weekly raw scores

Participant number	Baseline score	Week 1 score	Week 2 score	Week 3 score	Week 4 score	2 weeks After sessions ended: score
1A	46	39	42	42	34	n/a
2B	83	77	71	75	n/a →	
29D	56	n/a →				
MEAN SCORE:	n/a →					

Table 6. Anxiety Scores – QOL - weekly t-score & SE

Participant number	Baseline Mean of scores	Week 3 mean of scores	2 weeks After sessions ended: mean of scores
1A	62.6±1.7	56.8±1.8	n/a
2B	60.9±1.8	56.8±1.8	n/a
3C	n/a →		
29D	55.9±1.8	n/a →	

Table 7. End of Study Survey Responses (week 6)

Participant ID	How would you rate your experience 1-10	Was the study worth your time?	Was it an enjoyable experience?	Have you noticed any changes in your physical well being? If so what?	Would you continue the symptom intervention you participated in after this study?	Would you recommend the symptom intervention you participated in to someone else in the IBD community?
1A	7	Yes	Yes, it was nice I was being held accountable for moving throughout the week.	I definitely felt less stressed during this study.	Yes	Yes

Figure 2.1 Electronic Flyer for Participant Recruitment. Was used to advertise study on Crohn's and Colitis Foundation as well as private GI clinics in the Puget Bay Area. The flyer has general information about what is expected of participants, qualifications, supplies needed, and benefits. It also has contact information, two ways to access the interest form, and acknowledgement of SRC/IRB approval.

Figure 2.2 Visual representation of procedure conducted at participants homes. Shows the Mindfulness group watching videos while participating, the Yoga group moving along with video cues, and the Guided Intervention group reading via four possible methods.

Figure 3.1 First email sent to participants after interest form is submitted. Contains crucial information such as participant ID, first week plan (including links to videos/readings), and the pre-study surveys. Important information is bolded, and the email is broken into easy-to-read sections.

To: Particpant email

Cc:

Bcc:

Subject: IBD Research Project Info and Week 1 plan

Greetings!

Your participant number is **_A2_**

Thank you for participating in this study! If you have any questions or concerns, please reach out to me at s-agrove@lwsd.org or contact my adult supervisor at Hannibal.Person@seattlechildrens.org.

You are part of the Yoga Therapy group which will require roughly 2 1-hours sessions each week for 4 weeks. You can decide which days you practice yoga, just be consistent with spacing out sessions with a minimum of 2 days apart.

Yoga therapy instructions/resources:

(1)Restorative Practice -Yoga & Meditation for IBD: https://www.youtube.com/watch?v=axSxpLfF464 (41 minutes)
(2)Active Practice: Yoga & Meditation in Inflammatory Bowel Disease (IBD): https://www.youtube.com/watch?v=3TdjCZpX3Dk (43 minutes)

For your first session complete the **Restorative practice (1)**, for you second complete the **Active practice (2).**

To complete the lessons, you can use rolled up towels, blankets, and pillows in replacement for the equipment seen in the video.

IMPORTANT
Before starting the study, I would like to collect some baseline data. Please fill out each of these forms by ***Friday, January 21st.***

Participant Information: https://forms.office.com/r/XJGLzQ21yV
Anxiety Form (sourced from Health Measures) : https://forms.office.com/r/rvw2YY3EL6
IBDSI-SF: https://forms.office.com/r/ZMZExN8piE

Sincerely,
Aubrey Grove

11th grade
Tesla STEM High School
2019-2023

Senior Editor of Yearbook

Contact me at: s-agrove@lwsd.org

Figure 3.2 Email containing the IBDSI-LF end of week survey. This will be sent to every participant and on the second week will contain the Anxiety form. This will be sent out Saturday (day before) at 9:00am.

Bcc:

Subject: IBD Research Study Survey - Due January 30th

Greetings!

I hope your first week went smoothly. Please reach out if you have any questions about the procedure or accommodations.

Please complete the following survey by **11:59 pm on Sunday, January 30th.**

Microsoft Form version of the IBDSI-SF: https://forms.office.com/r/ZMZExN8piE

Remember to use your participant number received at the beginning of this study.

Sincerely,
Aubrey Grove

Figure 3.3 Email sent for participant debrief on the Monday after the study is complete. Contains background information about the study's purpose, research supporting the hypothesis, and a final thanks for participating.

IBD Research Study Wrap up • s-agrove@lwsd.org

Message Options

Send Paste Calibri (Bo... 11 Attach File Insert Link Table

To:

Cc:

Bcc:

Subject: IBD Research Study Wrap up

Greetings!

I am happy to announce the IBD Research Study you have been participating in has finished! Thank you so much for participating for these past four weeks and taking your time to complete surveys. You may be wondering what I was trying to accomplish with this project. Here is some background information/the purpose.

Inflammatory Bowel Diseases (IBD), consisting of Chron's disease (CD) and Ulcerative Colitis (UC) are chronic inflammatory conditions. The condition is digestive system based and requires life-long therapies/care to reduce inflammation. IBD occurs at all ages and debilitating symptoms such as diarrhea, abdominal pain, bloody stools, weight loss, and fatigue are sometimes present. In 2017 it was recorded that 6.8 million people had IBD in form of Ulcerative Colitis and Crohn's Disease, globally (Sudabeh et al., 2020). Additionally, other studies suggest IBD particularly in children is rapidly increasing (Benchimol *et al.,* 2014). While IBD symptoms are affected by physical factors, including location, extent, and severity, research has shown psychosocial factors, including stress and anxiety, may impact symptom severity, reduce quality of life, and even precipitate disease exacerbations (Purnima *et al.,* 2015; Sun *et al.,* 2019). Gastrointestinal conditions involve of brain-gut connections meaning if you regulate the brain's response you can alter the messages relayed between the systems (Gaylord *et al.,* 2021). Human brains adjust signaling based on their emotiona state. Yoga and mindfulness release neurotransmitters such as GABA, dopamine, endorphins, etc. These helps regulate stress and anxiety levels in the body (Lymn 2019). Therefore, mindfulness and yoga therapy may be an effective form of adjunctive treatment for reducing/managing IBD symptoms and improving quality of life. Researchers Steven and Adam (2018) found the, "estimated prevalence of complementary and alternative medicine (CAM) use in patients with IBD is high, ranging between 21% and 60%," showing there is an interest in for CAM in the IBD community. This project seeks to investigate if meditation or yoga is the more effective CAM for IBD symptoms. This project seeks to investigate if MBT or YT reduces symptoms severity and improves quality of life in adult patients living with IBD. Further, this project will assess these patients' interest in these therapeutic modalities and experience participating in them. Given the known interest of patients with IBD in CAM anc the potential for non-pharmacologic therapies to improve symptom and psychosocial outcomes, this project's purpose is to expand our understanding of the potential role of MBT and YT in IBD treatment.

Please let me know if you have any questions about the information above. Thank you for being a part of this study, it has helped add the research for the IBD community.

Sincerely,
Aubrey Grove

11th grade
Tesla STEM High School
2019-2023

Senior Editor of Yearbook

Contact me at: s-agrove@lwsd.org

The Effect of a Cognitive Behavioral Therapy App (Happify) on Adolescent Anxiety

Ananya S. Soni

Nikola Tesla STEM High School

Abstract

Anxiety is a serious mental health condition in the United States. Many people in low-income communities often cannot afford professional mental health care, so how can mental health care be made more accessible using an efficacious treatment intervention? The purpose of this study is to investigate whether a free cognitive behavioral therapy app (Happify) lowers adolescent anxiety levels as measured through the Beck Anxiety Inventory. The hypothesis being tested is whether at least 70% of participants will experience a reduction in anxiety symptoms after using a cognitive behavioral therapy app. After human participants were found and given the pre-Beck Anxiety Inventory, participants in the experimental group were given the cognitive behavioral therapy app Happify (independent variable), while participants in the control group were not exposed to any behavioral therapy. After 4-weeks, all participants were given the post-Beck Anxiety Inventory (dependent variable). The post-Beck Anxiety Inventory values were compared to the pre-Beck Anxiety Inventory values in both the control and experimental groups and statistically analyzed. The Cohen's d-value found was 0.707 which shows that the cognitive behavioral therapy app Happify had a medium effect size and reduced levels of adolescent anxiety. All participants in the experimental group experienced a reduction in self-reported levels of anxiety as opposed to 50% in the control group. These results indicate that a cognitive behavioral therapy app could lower self-reported levels of adolescent anxiety which further provides support for an evidence-based treatment intervention for anxiety that is free and widely assessable to many communities.

Introduction

The purpose of this study is to investigate how a free cognitive behavioral therapy app (Happify), used without other treatment interventions, lowers self-reported levels of adolescent anxiety as measured through the Beck Anxiety Inventory (an inventory used to assess cognitive and physical symptoms of anxiety). Anxiety is an emotion characterized by consistent fear, worry, and tension. Some symptoms of anxiety include sweating, increased blood pressure, and rapid breathing. However, anxiety is not considered a disorder until it interferes with an individual's ability to cope with daily situations. Anxiety disorders are a common condition in the United States and are in fact the most common mental health disorder in adolescents living in the United States (ACOG). An astounding 6.3 million teenagers in the United States have an anxiety disorder according to a study done by National Institute for Mental Health in 2015. In addition, according to Our World in Data, an estimated 284 million people in the world experienced an anxiety disorder in 2017 (Dattani 2). The objective behind this project is to figure out how to make mental health care for anxiety more accessible. Many people in low-income communities cannot afford professional mental health care. Mental Health America reports that 11.1% of Americans with a mental illnesses are uninsured (MHA). Cognitive Behavioral therapy apps are a

potential solution. Cognitive behavioral therapy is a is a form of psychotherapy used to reduce symptoms of mental health illnesses by challenging negative thought patterns. Cognitive behavioral therapy consists of three components: thoughts, behaviors, and emotions, which are each linked to one another. Cognitive behavioral therapy is considered an evidence-based treatment intervention for anxiety. However, cognitive behavioral therapy is typically delivered through a professional therapist. This study seeks to analyze the effects of cognitive behavioral therapy delivered through an app. Prior research has been done on the effects of a cognitive behavioral therapy apps. In one study researchers attempted to build and test an app (MindClimb) for adolescents with anxiety for use during cognitive behavioral therapy intervention. The intent of the researchers for the app was for it to be used by adolescents outside of cognitive behavioral therapy sessions for exposure activities, relaxation strategies, and self-reward. Participants reported that the app helped them become comfortable around new people, use relaxation sounds in the car, and remember to do their therapy homework. The average satisfaction score of the app among adolescents was 11.3, which was considered to be high (Newton 6). Another study tested a mobile app integrating Cognitive Behavioral Therapy, called Pacifica. The researchers found that participants who were assigned to use the app Pacifica in the Pacifica group, had greater improvements in self-reported levels of anxiety, stress, and depression compared to individuals assigned to the control group and that it is effective in reducing symptoms of mild-to-moderate anxiety. They also found that the app is effective in increasing feelings of self-efficacy when compared to the control group. Lastly, they found that the app had a Cohen's d effect size of .38, which can be interpreted as a medium effect size (Moberg 11). In another study researchers attempted to build and test an app called SmartCAT 2.0 to see whether it is effective in delivering Cognitive behavioral therapy skills to participants aged (9-14) with generalized/ social/ separation anxiety in between brief CBT sessions. They found that 97% of youth and parents were "mostly satisfied" or "very satisfied" with the treatment. Additionally, researchers found that 86.2% of participants were free of all qualifying anxiety diagnoses at the 2-month follow-up after using the app SmartCAT 2.0 (a cognitive behavioral therapy app) in conjunction with brief CBT sessions (Silk 13). Similarly, this study aims to find the change in anxiety levels in human participants. However, unlike other studies, this study aims to find the effect on anxiety levels through a cognitive behavioral therapy app alone, to see whether cognitive behavioral therapy can be effective when delivered through an app without any other treatment interventions. This leads to the hypothesis that if the cognitive behavioral therapy app Happify is used without any other treatment, at least 70% of participants will experience a reduction in self-reported anxiety symptoms.

Methodology

The cognitive behavioral therapy app Happify, Microsoft forms, Word, Excel, G-power software, and computers were used to conduct the study. The G-power calculator (Faul 3) was used by inputting a Cohen's d-value from a similar study looking at the effect of a cognitive behavioral therapy app on anxiety to find the target number of participants needed for this study. Since the Pacifica Study found a medium effect size for their app, a medium effect size of 0.5 was inputted into the G-power calculator and a target number of 50 participants was found. Participants were then recruited. During the recruiting period, local psychology professors were requested to tell their students about the study through email. Seniors in local high schools in Redmond were also emailed about the study. An initial 2-week period was set for recruitment of participants. After the end of the initial recruitment period, an extension of 1-week was set to the recruitment period due to there being less than 100 people willing to participate. After the end of the recruiting period for participants, the names of all willing participants and background information were collected, and the names were put into a random number generator. Participants were randomly assigned to the control or experimental group as this study was a double blind experiment. This minimized any experimenter bias that could skew the results found. After collecting background information on all the participants, the participants were given the Informed Consent form through email and Microsoft forms. The participants were then given the Beck Anxiety

Inventory and were asked to provide their age and gender. All the Beck-Anxiety inventory values were collected and classified as pre-BAI values. The experimental group was then given instructions to use the cognitive behavioral therapy app, Happify, while the control group was asked to do what they normally do. After a period of 4-weeks, all participants were given the Beck Anxiety inventory again. All the Beck-Anxiety inventory values during this time were classified as post-BAI values. Afterwards, all data collected throughout the study was inserted into Data Table 1.1.

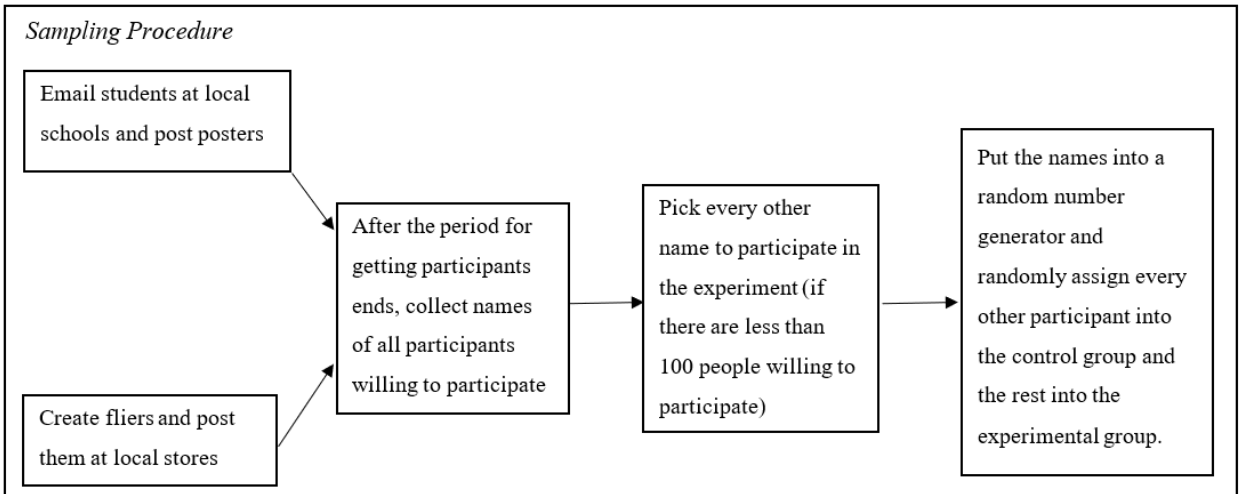


Figure 2. This model depicts the process of the sampling procedure. It shows the steps from recruiting participants, to randomly assigning them to the control or experimental group.

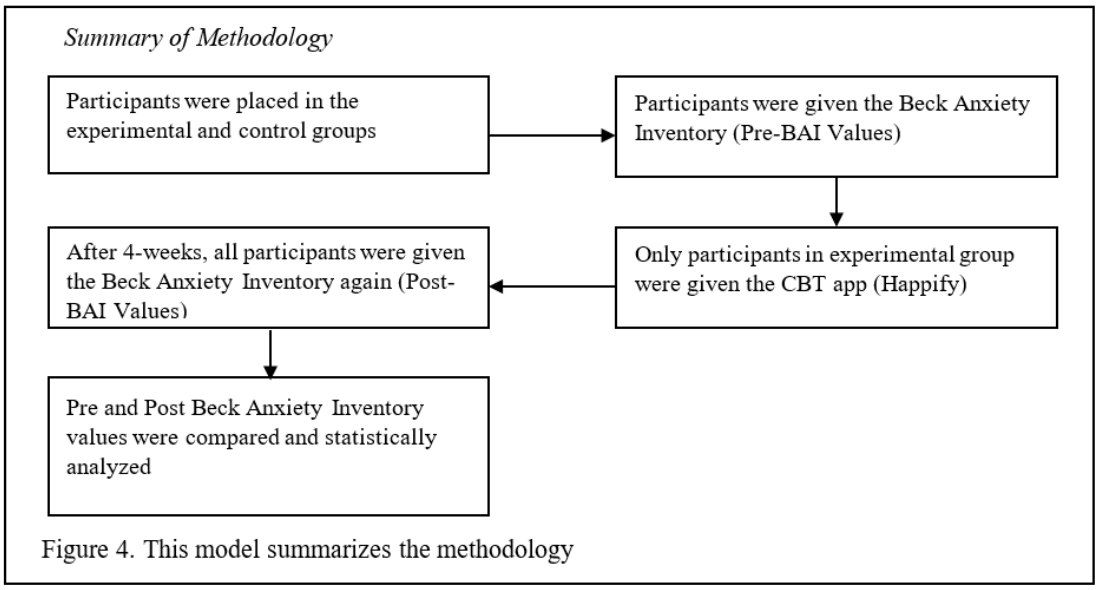


Figure 4. This model summarizes the methodology

G-Power

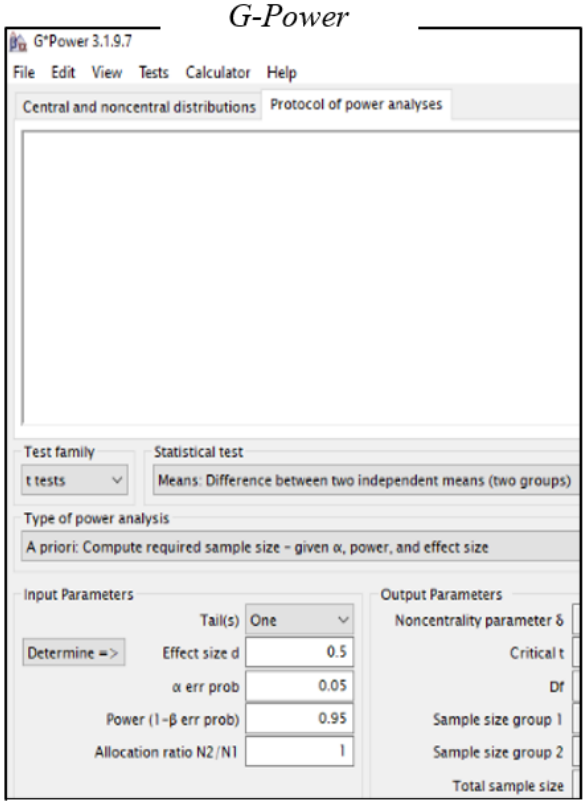
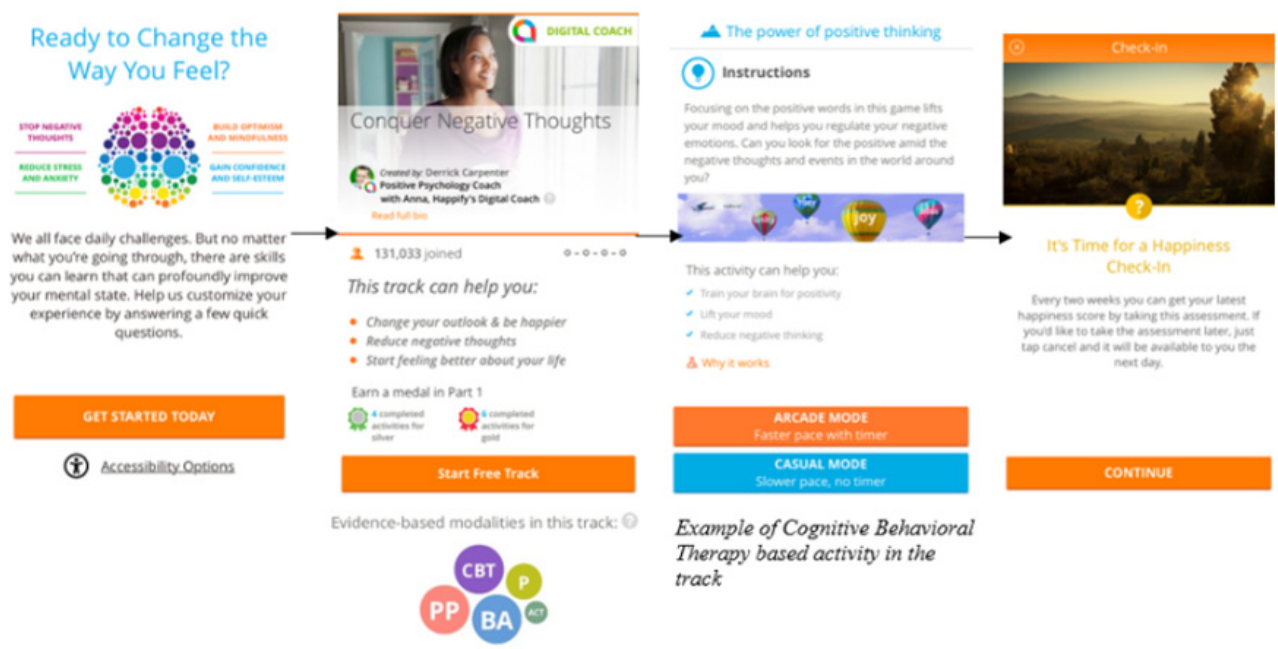


Figure 1. This calculator takes effect size input (d-value) and can give output of the target sample size.

App Procedure



Example of Cognitive Behavioral Therapy based activity in the track

Figure 3. *Shows some of the important screens of the app that follow the App Procedure*

Data Table 1.1

(Participant #)	*Age*	*Gender*	*Control/Experimental Group*	*Pre-BAI Value*	*Post- BAI Value*	*Change in BAI Value*
1	*18*	*Female*	*Control*	*15*	*11*	*-4*
2	*18*	*Female*	*Experimental*	*12*	*9*	*-3*
3	*18*	*Gender-fluid*	*Control*	*11*	*11*	*0*
4	*18*	*Female*	*Experimental*	*10*	*7*	*-3*

Results and Conclusion

After the pre and post Beck Anxiety Inventory values were collected, an online calculator was used to conduct statistical tests on the data. The average pre-BAI value of the control and experimental group were found along with the average post-BAI value. Additionally, the average change in the BAI value was found across the experimental and control groups. The Cohen's d-value was collected to find the effect size for the cognitive behavioral therapy app (Happify) on the experimental group. Lastly the p-value was collected to see how likely the results occurred by random chance. All values collected from the data analysis were put into Data Table 1.2. The p-value found was 0.352 which indicates that the results found are not statistically significant. This was likely due the small sample size in the study. This p-value means that the hypothesis being tested cannot be accepted. However, a Cohen's d-value of 0.707 was found which indicates that the CBT app Happify had a medium to high effect size and shows that there was a significant difference between the mean change in BAI-value in the experimental and control groups. In addition, a higher percent decrease in the Beck Anxiety Inventory value was found in the experimental group, which suggests that the app does lower symptoms of anxiety. All participants in the experimental group experienced a reduction in self-reported levels of anxiety as opposed to only 50% in the control group. This provides some evidence for the hypothesis being tested because more than 70% of the participants experienced a reduction in self-reported anxiety symptoms after using the CBT app (Happify). However, the p-value found indicates that these results occurred by random chance so the hypothesis cannot be accepted. Further testing, with additional human participants, needs to be done to gain more reliable results.

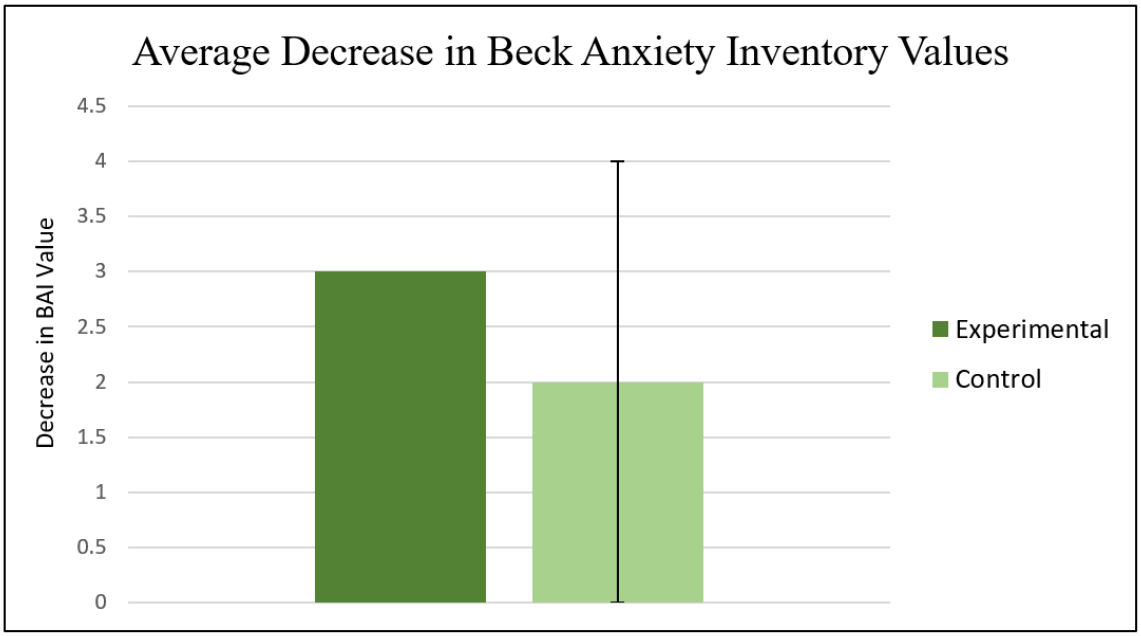
Data Table 1.2

Group	*Average Pre-BAI Value*	*Average Post-BAI Value*	*Average Decrease in BAI Value*	*Percent Decrease in BAI Value*	*Cohen d-value*	*P-value*
Control	*13*	*11*	*2*	*15.38%*	*0.707*	*0.352*
Experimental	*11*	*8*	*3*	*27.27%*		

Discussion

There were some limitations to this project. One limitation was the limited number of human participants in this study. With very few human participants, the results aren't as reliable as they could have been with more participants. The low number of human participant also likely led to the insignificant p-value found in this study. Another limitation in this study was possible human error when transferring the raw data to make calcula-

tions. However, despite these limitations, cognitive behavioral therapy apps are still worth looking into as an alternative to traditional psychotherapy. Further testing on CBT apps could be done on a larger scale with additional participants for more reliable results. In addition, in a future study specific CBT features in an app could be investigated to see which features are the most effective in reducing symptoms of anxiety. Lastly AI implementation could be used with CBT apps. For example, an app could use a smart chat-bot feature that is able to deliver cognitive behavioral therapy skills. AI could also be used to target CBT therapy towards each user. All of these suggestions are potential ideas for future research to continue to explore the implementation of CBT through an app.



Bibliography

Addepally, S. A., & Purkayastha, S. (2017). Mobile-application based cognitive behavior therapy (CBT) for identifying and managing depression and anxiety. *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety,* 3–12. https://doi.org/10.1007/978-3-319-58466-9_1

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of consulting and clinical psychology,* 56(6), 893–897. https://doi.org/10.1037//0022-006x.56.6.893

BYJU'S. (2021, June 15). *P-value (definition, formula, Table & Example): Significant P-value.* BYJUS. Retrieved December 12, 2021, from https://byjus.com/maths/p-value/.

Dattani, S., Ritchie, H. and Roser, M., 2021. *Mental Health.* [online] Our World in Data. Available at: <https://ourworldindata.org/mental-health> [Accessed 11 December 2021].

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical Power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders,* 218, 15–22. https://doi.org/10.1016/j.jad.2017.04.046

Grist, R., Porter, J., & Stallard, P. (2017). Mental Health Mobile Apps for preadolescents and adolescents: A systematic review. J*ournal of Medical Internet Research,* 19(5). https://doi.org/10.2196/jmir.7332

Happify. Happify | Anxiety and Depression Asso ciation of America, ADAA. (n.d.). Retrieved December 12, 2021, from https://adaa.org/node/2809.

Julian, L. J. (2011). Measures of anxiety: State-trait anxiety inventory (STAI), Beck anxiety inventory (BAI), and hospital anxiety and depression scale-anxiety (HADS-A). *Arthritis Care & Research,* 63(S11). https://doi.org/10.1002/acr.20561

A Scientific Look into Mental Visualization and Improved Performance in Taekwondo

Kai T. Vanderlip

Nikola Tesla STEM High School

Abstract:

Mental visualization is a practice common by athletes. This project investigated image training's effectiveness in increasing taekwondo students' accuracy and speed in kicks. A group of taekwondo students from varying ages and belt levels were randomly selected into groups and asked to perform image training- except for the control group which had no image training-, for 30 sec or 60 sec, in between sets. The accuracy and speed of the kicks were measured to test for the highest improvement of speed and accuracy. After compiling the results, a 20% increase in accuracy and a 41% increase in speed were recorded in one participant from the group with sixty seconds of image training. The results a p-value of 0.346 for accuracy when comparing the 60, 30, and 0 second image training groups, and a p-value of 0.369 for speed. This experiment was successful in finding a link between increased taekwondo performance and image training, which increases knowledge in how athletes can improve sports performance.

Introduction:

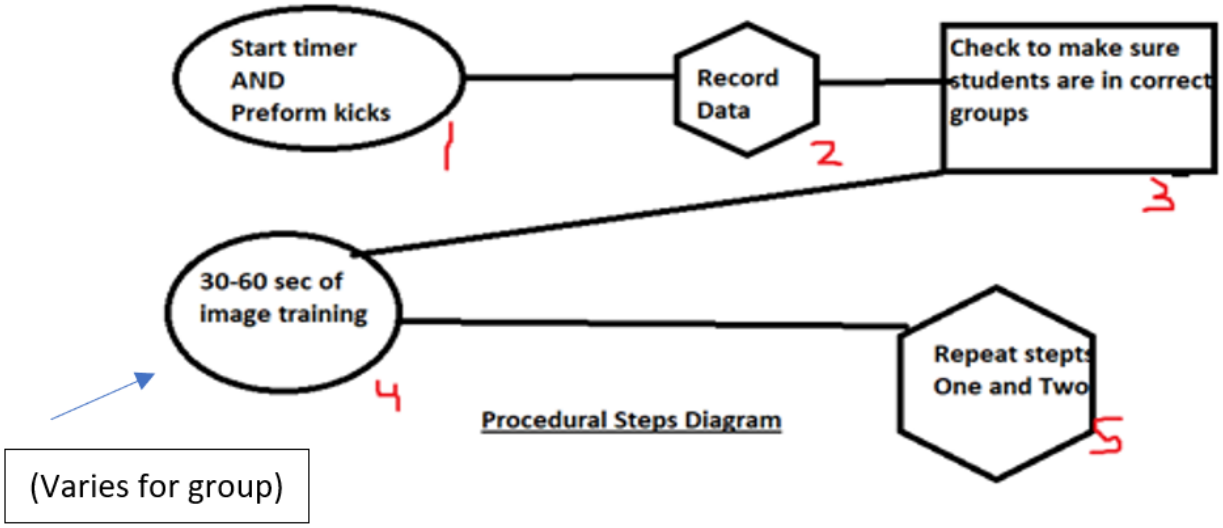
This research paper seeks to investigate a possible causation effect between image training and an improved performance in Taekwondo. Relax, breathe, don't be tense: these are some of the most common things Taekwondo Masters tell students as feedback. In addition, "freezing-up" is the biggest enemy of any martial artist and/or athlete when it comes time to perform. Stress and fear, which can cause "freezing-up", are both universal experiences and roadblocks for athletes, making a potential remedy for these ailments universally relevant. One study by Newmark (2012) studied anxiety in a gymnast who had performed "excellent in practice, but never [did] well during competition … her heart would race fast, she would sweat and tremble, and completely lose focus." The gymnast later engaged in therapy with guided imagery techniques: she relaxed her muscles, then visualized herself in a competitive event and had to verbally describe, in detail, her surroundings and watching herself execute all moves perfectly. "Her relaxation and focus was to the point to where she was having more fun during competition. With this happening, results were achieved and she improved her scores/performance during competition." (Ekeocha, Osborne, Seay, Huber, 2015). Additionally, a study done on basketball players revealed that image training was almost as effective, over the course of ten weeks, as physical training (Uludağ, Dorak, Vurgun, Yüzbaşioğlu, Ateş, (2021). This project is similar to the basketball experiment, as it involves measuring the amount of improvement in a sport after the practice of image training. This project differs as it seeks to see if an extraordinarily brief amount of mental visualization (30 or 60 seconds) will improve physical performance. Similarly, a sports psychologist team in Denmark established mental skills as an important piece in an individuals perforce. (Henriksen, Diment, Hansen, 2011). A group of music students showed that after an eight-week image training program, their performance anxiety was less than those without the training (Munroe-Chandler, & Guerrero, (2017)). If a group of Taekwondo students practice positive image training1 for 30 or 60 seconds before performing a set of kicks, their accuracy and speed

will be greater than those who do not.
1: Image training/mental visualization is the practice of imaging yourself performing an activity in your mind, including as many sensory details as possible. You should be executing your task perfectly (succeeding) and sticking to positive thinking only.

Body-Methodology:

This hypothesis is tested by recruiting Taekwondo students to perform a set of ten snap kicks, followed by ten roundhouse kicks. Participants were assigned to one of three groups at random. To accomplish this, participants were given random numbers via a random number generator. Group one performed another set of snap and roundhouse kicks following their initial set. Group two did 30 seconds of image training exercises, then did another set of the kicks. Group three followed the same steps as group two, but performed image training for 60 seconds instead (Fig. 1). To test if groups improved, speed and accuracy were measured.

Figure 1 – Procedural Steps Diagram



Participants were recruited through e-mails to Taekwondo school masters. From there, the observation monitored students throughout the experiment over a virtual meeting to maintain safety with an increasing risk of COVID-19. This project recruited 10 participants, 3 of which were placed into the control group (0 seconds), 4 were in experimental group one (30 seconds), and the last three in experimental group two (60 seconds). An informed consent form is provided to each participant, minors need an adult to sign.

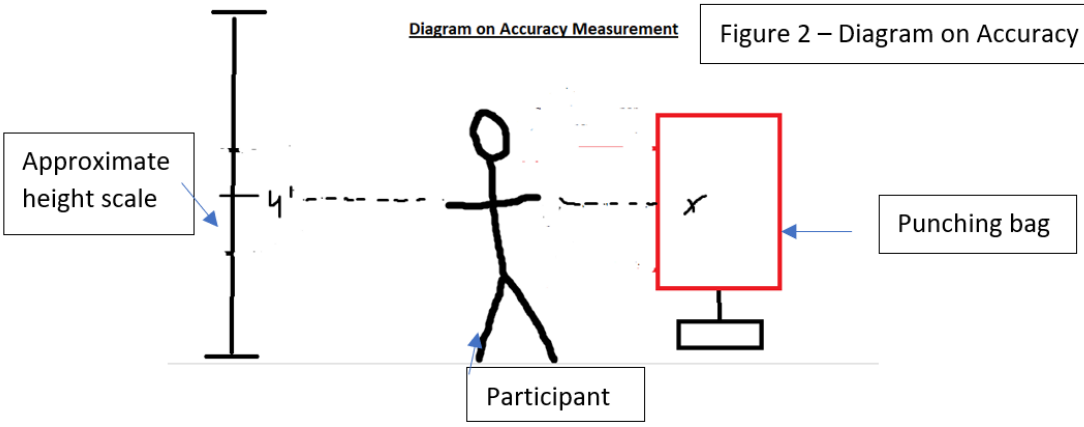


Figure 2 – Diagram on Accuracy Measurement

Accuracy is measured by taping a "x" onto a punching bag, four feet off of the floor, and taking note of how many times the student missed or hit (Fig. 2).

Table 1 – Preliminary Data Table

Mental Visualization Time (Sec)	Belt Level (Color/degrees)	Accuracy (Hit or miss w/ X)	Speed (Sec)	Male/Female (Legal sex)

(Data Table of Mental Visualization Time, Accuracy, Power, And Belt Level)

Mental visualization/image training will be measured using a timer and will be recorded in seconds (Table 1). The belt level of students will be recorded for better comprehension of data, as it allows to check for correlation between belt level and improvement. The higher the belt, the more vivid the image training should be, and therefore the improvement should be greater. This can be seen by a similar study in karate students, which found the more competitive athletes performed better at mental visualization (Guarnera, Stummiello, Cascio, Corrado, 2016) Accuracy is measured as binary, 1 if they hit, 0 if they miss. Speed will also be recorded in seconds, this speed being the amount of time it takes participants to execute a set of 10 kicks.

Table 2 – Finalized data table used during experimentation

Belt Level	Image Training Time (Sec)	Roundhouse (Rh) time (Sec), Snap kick (Sk) time (Sec)	Rh Accuracy (0 to 10, hit /miss), Sk Accuracy (0 to 10, hit /miss)	Rh time after training, Sk time after training	Rh Accuracy after training, Sk accuracy after training

Results:

After compiling the accuracy and speed measurements into bar graphs, there was a clear trend of improvement for before and after image training. However, the null hypothesis that image training would increase Taekwondo students' accuracy and speed is rejected as this project was unable to prove statistical significance. For the group that performed 60 seconds of image training, their mean for accuracy (kicks hit out of 10) was 8.7, with the median of 8.5, and the mode of 8. After going through image training, the mean was 9.7, and the median and mode were 10.

Using a one-way ANOVA, a P-value of 0.346 was recorded for accuracy and 0.369 for speed. Therefore, results were statistically insignificant. The next step of this study would be to continue research and experimenting with a larger sample size. With only 10 participants, 3 were assigned to control, 4 to the 30 second experimental group, and 3 to the 60 second experimental group. More participants are needed for accurate and statistically significant results. In addition, this project was done remotely due to concerns with COVID-19, this is something that could be changed in further studies. Therefore, future studies, will be in person with a larger sample size, if possible.

Figure 3 – Bar graph chart for the 60 second accuracy group

(Note: Snap kicks depicted left of grey line, roundhouse to the right. Blue is before image training orange is after)

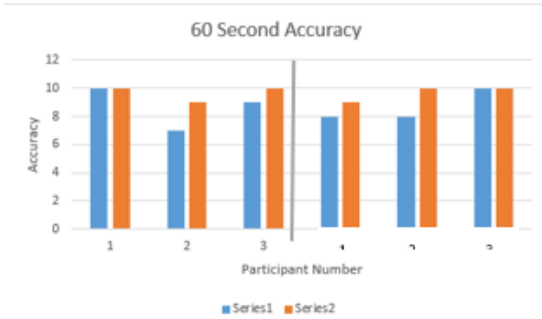


Table 3 – Finalized data table with data from experimentation

Belt Level	Image Training Time (Sec)	Roundhouse (Rh) time (Sec), Snap kick (Sk) time (Sec)	Rh Accuracy (0 to 10, hit /miss), Sk Accuracy (0 to 10, hit /miss)	Rh time after training, Sk time after training	Rh Accuracy after training, Sk accuracy after training
3rd Black	60	6.78, 8.27	10, 8	5.68, 7.85	10, 9
2nd Black	30	8.81, 10.47	10, 9	7.58, 9.38	10, 10
Brown	0	8.96, 9.31	9, 9	7.58, 9.38	9, 10
Black	0	9.55, 9.91	9, 10	9.23, 8.45	10, 10
3rd Black	30	23.7, 10.48	10, 9	9.77, 9.91	10, 10
Red	0	14.83, 16.73	10, 10	19.32, 16.46	10, 10
Black	60	15.75, 11.68	7, 8	11.18, 11.18	9, 10
Brown	30	10.67, 10.76	3, 9	9.35, 9.60	7, 10
Black	30	6.87, 6.69	10, 7	4.87, 6.68	10, 10
Black	60	7.01, 8.03	9, 10	5.89, 6.83	10, 10

Works Cited:

1. Ekeocha, T. C., Osborne, R., Seay, O., & Huber, J. (2015). T*HE EFFECTS OF VISUALIZATION & GUIDED IMAGERY IN SPORTS PERFORMANCE.* https://digital.library.txstate.edu/bitstream/handle/10877/5548/EKEOCHA-THESIS-2015.pdf?sequence=1

2. Guarnera, M., Stummiello, M., Cascio, M. I., & Corrado, D. D. (2016). Vividness and transformation of mental images in karate. International Journal of Kinesiology & Sports Science, 4(3) Retrieved from https://www.proquest.com/scholarly-journals/vividness-transformation-mental-images-karate/docview/1878750066/se-2

3. Henriksen, K., Diment, G., & Hansen, J. (2011). Professional philosophy: Inside the delivery of sport psychology service at team denmark. *Sport Science Review*, Xx(1-2), 5. doi:http://dx.doi.org/10.2478/v10237-011-0043-6

4. Munroe-Chandler, K. J., & Guerrero, M. D. (2017). Psychological Imagery in Sport and Performance. Oxford Research Encyclopedia of Psychology.

5. Newmark, T. (2012). Cases in Visualization for Improved Athletic Performance. *Psychiatric Annals,* 42(10), 385–387. https://doi.org/10.3928/00485713-20121003-07

6. Uludağ, S., Dorak, F., Vurgun, N., Yüzbaşioğlu, Y., & Ateş, E. (2021). Effects of 10 weeks of imagery and concentration training on visual focus and free-throw performance in basketball players. *Journal of Physical Education and Sport,* 21(4), 1761-1768. doi:http://dx.doi.org/10.7752/jpes.2021.04223

Stand up to Stigma: ADHD in Women

Nuha Rahman, Madhumitha Rasiah

Nikola Tesla STEM High School

Introduction:

This research paper aims to demonstrate the effects of stigma and lack of knowledge in the referrals and diagnosis of Attention/Deficit Hyperactivity Disorder (ADHD) in women. ADHD is a neurodevelopmental disorder that is commonly genetic, and effects of this disorder are seen in the frontal lobe. Women are three times more likely to go undiagnosed for ADHD compared to men and only about 5.6% of all women with ADHD are eventually diagnosed (CDC 2020). One of the main contributors to this low rate of referrals/diagnoses is the stigma surrounding ADHD and mental health as a subject. In 2001 the World Health Organization reported an estimated 25% of the entire population have a mental illness or a behavioral disorder sometime in their life, showing the high prevalence of this issue, and yet a considerable number of those people do not seek help. The reason for this is the stigma surrounding all mental health issues, including the focus of this study, ADHD, especially in women.

A recent study found that awareness programs implemented into schools saw higher rates of knowledge of mental health disorder symptoms and referrals for those symptoms (Salerno, 2017). In addition, perceptions of peers, teachers, and parents were found to influence the diagnosis of ADHD (Hamed et al., 2015). These two studies demonstrate that the involvement of background knowledge of disorders, and in the case of ADHD, can heavily change the referral rates for ADHD if our society is well educated and has open conversations about mental health. It is important to note when discussing women with ADHD, that actively and consciously our society should take notice of the internalized nature of the ADHD symptoms in women (Young, 2020). This internalized nature appears in the form of more easily masked symptoms such as inattention and lower levels of hyperactivity and impulsivity as well as in the form of comorbidities (more than one disease simultaneously acting with another) with emotional lability and anxiety. The lack of knowledge combined with the stigma surrounding ADHD in women, specifically with functional impairments, or the applications of symptoms, has possibly led to the decreased referrals and diagnosis of ADHD and especially ADHD among women (Nguyen & Hinshaw, 2020). It is hypothesized that the variable of stigma (an often negative preconception that can affect attitudes and reactions and misconception) and the variable of lacking knowledge (accuracy of answers to questions about factual things such as specific functional impairments and symptoms) can negatively impact our society as women with unconventional symptoms of ADHD such as inattention and low-severity impulsiveness, go unreferred and undiagnosed forming hardships in life that could have been avoided.

Methodology:

The study was partially modified from a survey that can be used to quantify a person's knowledge of ADHD and ADHD in women showing misconceptions or correct assumptions of ADHD with permission from lead researchers (Meza et al., 2019). Novel questions that assessed stigma against ADHD and knowledge of ADHD, both in regards to women, were added as the main questions that were assessed. The modified survey also aimed to observe relations to different demographics as well as to bring awareness to this subject. Participants were gathered via voluntary action through recruitment flyers, emails, and word of

mouth, and were only recruited with signed confirmation of informed consent forms. To reduce bias, all participants were assigned a randomized number to mask identities and to prevent the experimenters from knowing the participants. (Figure 1)

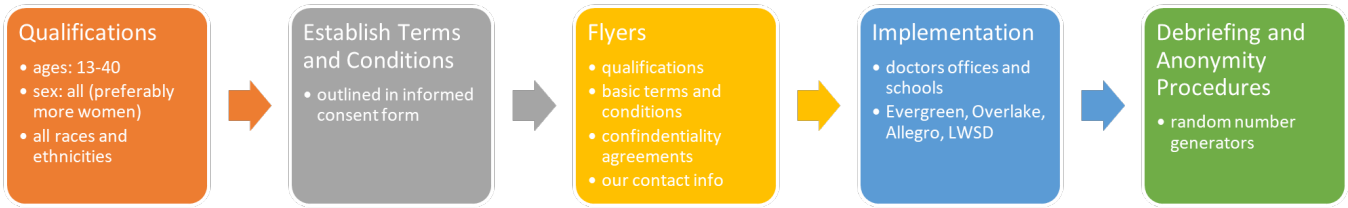


Figure 1. sampling procedure of human participants

52 participants' data was collected with differing ethnicities, ages 15 and older, and a female to male ratio of 74:26. The participants took the modified survey through a Google survey link (https://forms.gle/5oJCXk3MPyRiS4Zw7) which took around 20 minutes on average. The data was then transferred to an Excel Spreadsheet and cleaned. Attention questions that assessed whether the participant was thoughtfully answering the questions decided whether to void a participant's data. Then the responses were assessed. Qualitative responses – Agree, Somewhat Agree, Neither Agree nor Disagree, Somewhat Disagree, Disagree, and yes or no questions – were quantified via a 5-point scale and a 0-or-1-point scale. Values from 1-5 were assigned as points per question and an answer suggesting high tendency for that variable would receive a 5 and depending on the accuracy of the participant's answer the points would decrease from there. Yes or no responses were assigned either a 1 or 0, respectively. (Table A).

Table A. overall organization of questions and responses from each participant

Table A	**Question 1**	**Question 2**	**Question 3**	**(Question #)**
Participant 1	response	response	response	Etc.
Participant 2	response			
Participant 3	response			
(Participant #)	Etc.			

Quantified values for questions of each variable were added up to get a level of that variable for each participant. (Table B)

Table B. the format of which the quantified data for responses of each variable's questions are recorded and organized, sample values from knowledge questions given.

Table B	Question 1	Question 2	(Question #)	Level of (Variable)
Participant 1	5	5	Etc.	5+5+...
Participant 2	4	4	Etc.	4+4+...
(Participant #)	(Quantified response)	(Quantified response)	(Quantified response)	(sum of row values)
Average	**(Mean of column values)**	**(Mean of column values)**	**(Mean of column values)**	**(Mean of column values)**

Overall levels of each variable per participant were gathered into a combined table (Table C) for basic assessment and further data analysis.
Then the data was mathematically analyzed using R-values, p-values, and t-values to show correlation and significance of misconceptions or stigma in participants towards ADHD in women or in ADHD in general.

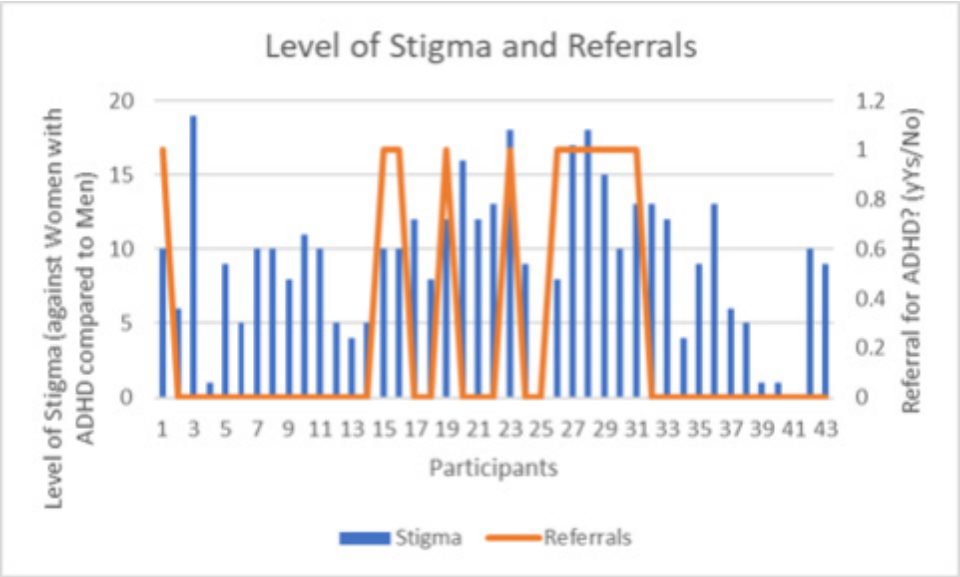
Results/Discussion:

Using R-studio, the data was analyzed for correlations and statistical significance. The t-value comparing Level of Stigma against men vs women came out to 12.71 indicating a significant difference in Level of Stigma between sexes as well as the p-value coming out to be less than .001, confidently rejecting the null hypothesis and showing statistical significance in Level of Stigma between sexes. The variables of Level of Knowledge and Referrals proved to have a statistically significant correlation with a p-value less than .05. The variables of Level of Stigma against women with ADHD and Level of Knowledge indicated no statistically significant correlation possibly due to limitations within the study. The biggest limitations for this study included a lack of sufficient participants. Of 52 participants, only 6 had ADHD, 5 being women. This is disproportional to real-world statistics where men are seen to have more instances of ADHD than women and so the representative spread of ADHD among sexes is not accurate. The data is not representative of all communities as many participants are educated in medical knowledge (due to many participants coming from medical careers in education), skewing the data. To prevent this in the future, the scope of where are participants are coming from needs to be increased to include a 1:1 ratio of females to males as well as get more participants of differing ages and backgrounds. This could be implemented through sharing the survey and the poster for the survey among middle schools, high schools, office buildings, and medical facilities.

Table C. overall levels of each variable after cleaning of participants' data

Participant#	Level of Stigma against Anthony	Level of Stigma against Amanda	Compared Level of Stigma against Women	Level of Knowledge	Referred to a professional for ADHD?
2	10	20	10	106	1
3	12	18	6	107	0
4	6	25	19	101	0
5	12	13	1	97	0
6	10	19	9	97	0
7	12	17	5	94	0
9	10	20	10	101	0
10	11	21	10	0	0
11	12	20	8	88	0
12	8	19	11	111	0
13	13	23	10	101	0
17	15	20	5	105	0
18	11	15	4	116	0
19	13	18	5	105	0
20	10	20	10	130	1
21	10	20	10	105	1
22	13	25	12	104	0
23	9	17	8	112	0
24	12	24	12	102	1
25	7	23	16	121	0
26	12	24	12	110	0
27	7	20	13	106	0
28	5	23	18	117	1
29	16	25	9	121	0
30	11	11	0	123	0
31	11	19	8	94	1
32	5	22	17	103	1
33	7	25	18	115	1
34	9	24	15	112	1
35	11	21	10	112	1
36	6	19	13	110	1
37	11	24	13	119	0
38	9	21	12	115	0
39	18	22	4	126	0
40	10	19	9	119	0
41	8	21	13	115	0
42	13	19	6	106	0
45	11	16	5	93	0
46	14	15	1	102	0
47	15	16	1	123	0
48	15	15	0	117	0
49	7	17	10	97	0
51	11	20	9	105	0
Averages	**10.6511**	**19.8837**	**9.2325**	**106.1162**	**0.2558**

Graph D. visual representation of calculated levels of stigma vs. Referrals for each participant



Bibliography

Crawford, N. (2003). ADHD: *A Women's issue.* Monitor on Psychology. Retrieved November 5, 2021, from https://www.apa.org/monitor/feb03/adhd.

Hamed, A. M., Kauer, A. J., & Stevens, H. E. (2015). *Why the diagnosis of attention deficit hyperactivity disorder matters.* Frontiers in psychiatry. Retrieved January 18, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4659921/

Hinshaw, S. P., Nguyen, P. T., O'Grady, S. M., & Rosenthal, E. A. (2021). Annual Research Review: Attention-deficit/ hyperactivity disorder in girls and women: underrepresentation, longitudinal processes, and key directions. Association for Child and Adolescent Mental Health.

Meza, J.I., Monroy, M., Ma, R. et al. Stigma and attention-deficit/hyperactivity disorder: negative perceptions and anger emotional reactions mediate the link between active symptoms and social distance. ADHD *Atten Def Hyp Disord* 11, 373–382 (2019). https://doi.org/10.1007/s12402-019-00302-x

Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., & Todd, R. D. (2010). *Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11*. Journal of the American Academy of Child and Adolescent Psychiatry. Retrieved November 5, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3101894/.

Wilens, T. E., & Spencer, T. J. (2010). *Understanding attention-deficit/hyperactivity disorder from childhood to adulthood*. Postgraduate medicine. Retrieved November 5, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724232/.

Young, S., Adamo, N., Ásgeirsdóttir, B. B., Branney, P., Beckett, M., Colley, W., Cubbin, S., Deeley, Q., Farrag, E., Gudjonsson, G., Hill, P., Hollingdale, J., Kilic, O., Lloyd, T., Mason, P., Paliokosta, E., Perecherla, S., Sedgwick, J., Skirrow, C., … Woodhouse, E. (2020). *Females with ADHD: An expert consensus statement taking a lifespan approach providing guidance for the identification and treatment of attention-deficit/hyperactivity disorder in girls and women.* BMC psychiatry. Retrieved November 5, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7422602/.

The Impact of Credibility When Hitting "Purchase Now"

Karina Kejriwal

Nikola Tesla STEM High School

Introduction

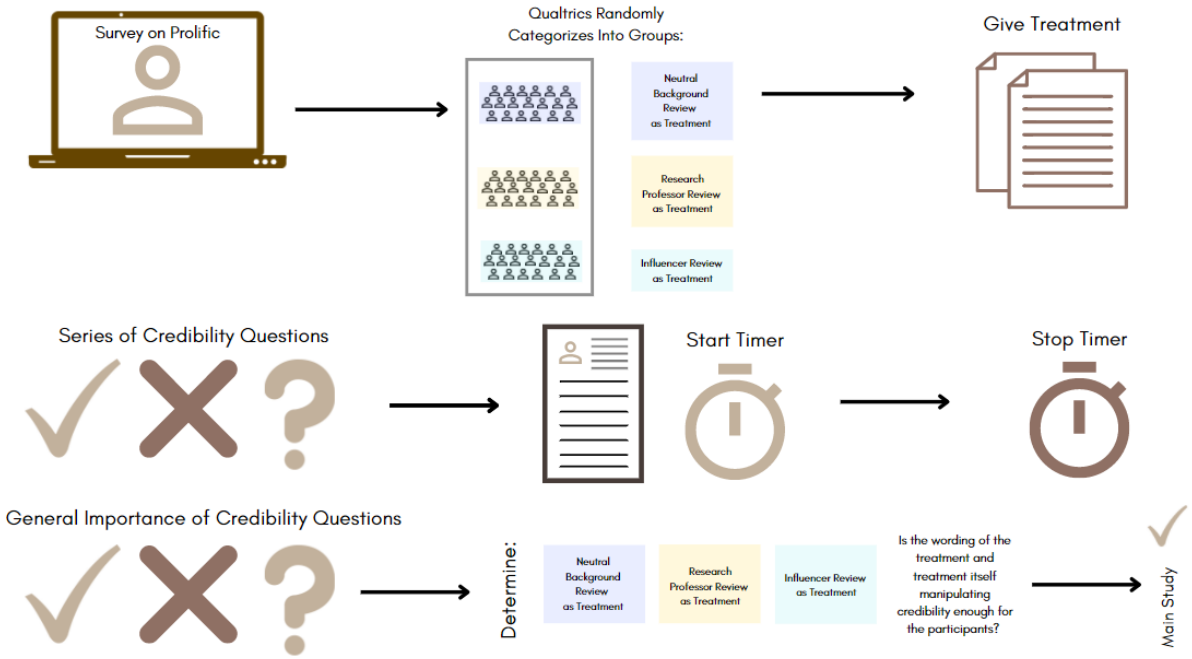
This research seeks to understand the factors individuals take into consideration when reading a product review and an analysis of the thought process which determines the perceived reviewer credibility. In today's age, 84% of individuals have reported that they trust online reviews as much as they trust a personal connection's review of a retail product (Forbes). With one click, consumers can view online reviews about a retail product from several individuals. These reviews can be positive, negative, neutral, and can be written by anyone. When consumers use online platforms for their purchases, the increased ease of access within e-commerce to view reviews has caused customers to become reliant on others' opinions and forming their own purchase decision based on reviews. In fact, Trustpilot says that 9 out of 10 make an effort to read the product reviews before purchasing and 6 out of 10 still are not convinced after a five-star rating (Dykstra, John, 2017). For the seller, this makes a substantial difference in sales because reviews written by any individual, without any context, can make a difference on a future consumer's purchase decision (Kreutzberg, G. W, 2005). As the e-commerce industry expands, consumers are relying more on reviews to have a broader and more detailed understanding of the product and see if other purchasers found the product to meet quality expectations, rather than simply trusting only the seller itself. In fact, this dependency is becoming so extreme, consumers are finding themselves influenced by any review and tend to make the decision to not purchase the product if the reviews are low or negative (Wei, L., 2016). Due to uncertainty, if a negative review is displayed, customers immediately begin to reconsider. A study by Bazaarvoice shows that 54% would not purchase products if reviews were negative or seemed fake (Liu, Gang., 2020). Currently, it is less known how credibility and professional background of the reviewer can impact the consumer's opinion in terms of how long it takes them to make a purchase decision. In this space, all research is directed towards how social media, credibility, and word of mouth impacts others' opinions in general; there is nothing specific regarding time for decision making in the retail industry. When there is a review on a product, different individuals with varied backgrounds can give their thoughts. It is perceived that the reviewer's qualifications factors into their review due to their own background knowledge. But, it is unknown that if the credibility is perceived as higher by the customer, will they instantly trust the reviewer within a short period of time? Decision timing is important to look at for two reasons. First, it helps to see how much the participant trusts the author. If they read the review and purchase quickly, it can be inferred that they read the review by skimming because the bio of the author was so credible to them that their review must be accurate. Second, the purpose of a review is to help participants make a purchase decision, and hopefully in a faster manner than without the review because there is more information about the product presented. Through reviews written by different professional backgrounds - high profile research professors, influencers, and typical individuals with a neutral background – we can determine whether different professional backgrounds influence perceived credibility of the review and the time it takes

to make that decision. This project investigates a wide array of retail shoppers coming from different demographics (gender and age group) and analyzes their processes and duration of time taken to determine a purchase decision on a retail product. It is expected that if an individual is given a review about a product that is written by reviewers from different professional backgrounds, they will be able to make a purchase decision the quickest after reading the review written from the highest perceived expertise/ credibility (ex: research professor) and side with them when purchasing a retail product.

Methodology

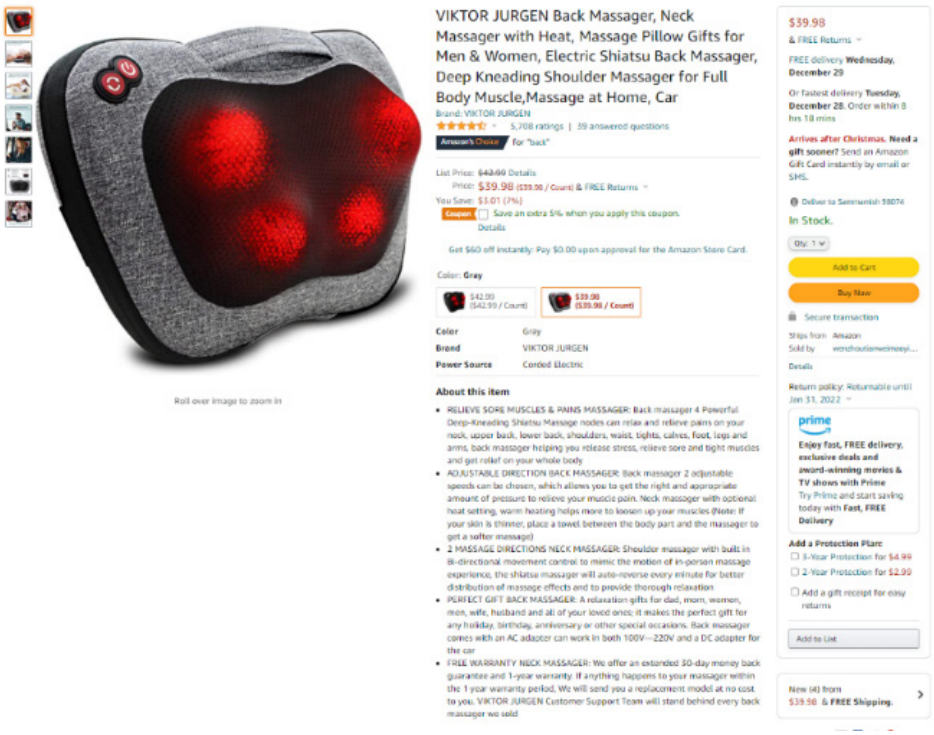
For the procedure of this experiment, participants (18 and above) were first recruited as a part of the sampling procedure. In this two-part study, the Pre-Test included 150 participants and they were recruited through Prolific Academic. This platform allows users to launch their study onto the site and reach specific demographics to fill out the survey. In the Pre-Test, participants were asked to answer a series of questions about a review (treatment for the main study). Survey questions included standard credibility questions that analyze how the customer feels about the reviewer in terms of trustworthiness, accuracy, knowledgably, and expertise. The review presented was determined by a random number generator within the Qualtrics form and each review was shown to an equal number of individuals. The Pre-Test results helped determine if the stimuli being used in the main study was manipulating credibility and expertise enough. It is important to ensure that the reviews acting as treatments in the experiment were worded properly so that participants in the study would recognize a significant difference between each of the reviews and their author's background while completing the main study. After positive results were received for the Pre-Test, the stimuli were ready for the Main Study. To recruit participants (18 and above) the sampling procedure consisted of putting up flyers at local grocery stores and Microsoft campus and sharing with colleagues, friends, family. Each participant was asked to fill out an informed consent that confirmed their approval to be a part of the study and that they understand all possible side effects. After recruiting 300 participants, each one was assigned a random number through the Qualtrics form. This study is a single blind study though the experimenter is not directly assigning the treatment, it can be checked which treatment each participant received.

Figure 1. Procedural Design: Model of Pre-Test survey purpose and procedure laid out step-by-step with questions and break down.



For the study procedure, all participants are firstly being timed while given a description of a retail product – back massager – and asked to determine how likely they are to purchase the product (1 being not at all and 7 being very likely). This is an initial standard question that all participants have to answer and is recorded as their initial response to the product in the study. An initial response is necessary to determine because then final response and initial response can be compared. If only final response is asked in the study, several variables can factor in such as reading time and then cause differences in data and it would be unknown as to why a participant's purchase decision duration was the length reported. When comparing initial and final, all these variables are taken into account and only the difference in time is analyzed.

Figure 2. Initial Image: Shown to all participants with product and product description



After, each participant is randomly given a treatment. These treatments were equally shown amongst all participants. Each treatment is composed of a standard worded review and then a manipulated reviewer biography and description. The experiment has 3 treatments: influencer background, research background, and neutral background. The review itself is worded the same on each treatment when the author biography is altered based on which professional background the treatment is required to display. Participants are then timed again while they read the review, have the option of re-reading the initial product description, and making a final purchase decision of 1-7 likelihood of purchasing (1 being not at all and 7 being very likely). Participants are also asked to give an in-depth explanation behind their purchase decision and the factors that contributed to their final decision.

Figure 3. Treatment Image: Image with review from a neutral professional background shown to randomly selected participants

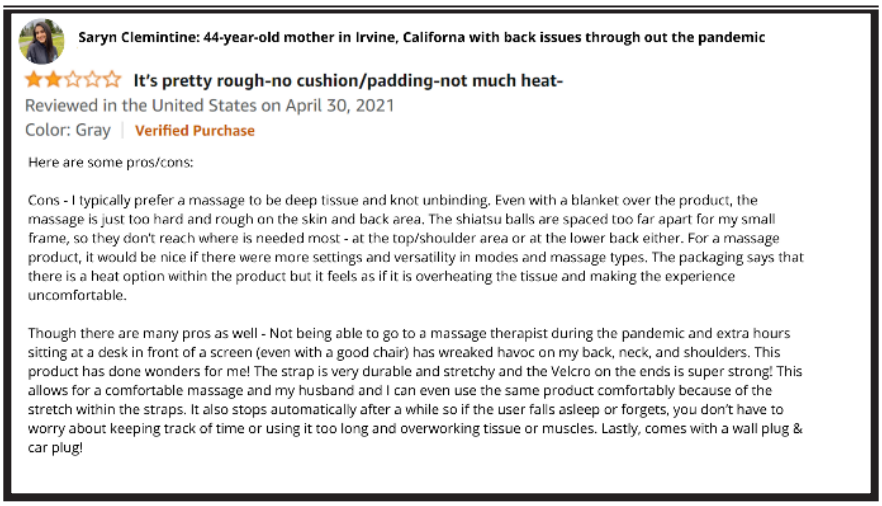


Figure 4. Treatment Image: Image with review from professor professional background shown to randomly selected participants

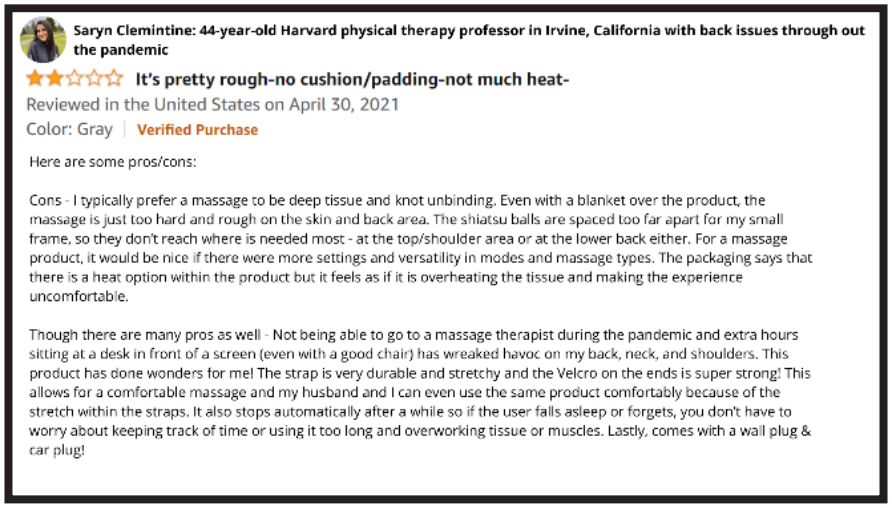


Figure 5. Treatment Image: Image with review from influencer professional background shown to randomly selected participants

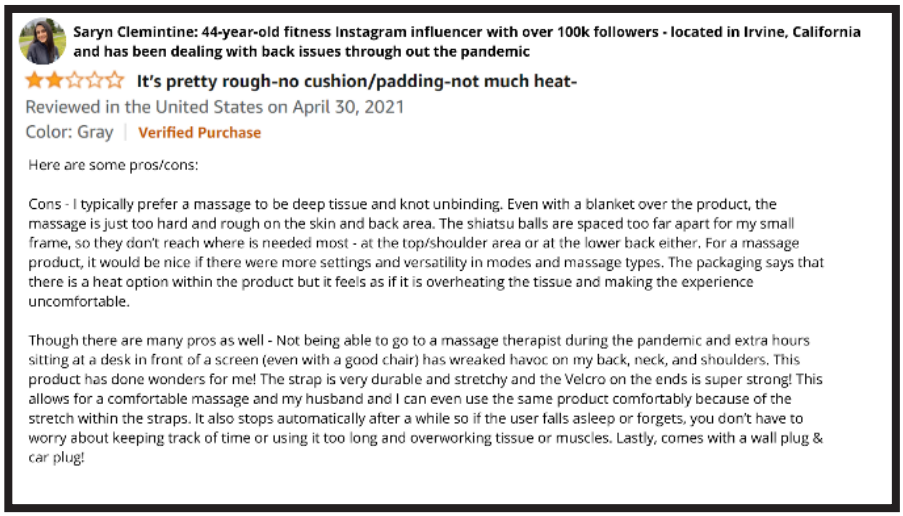
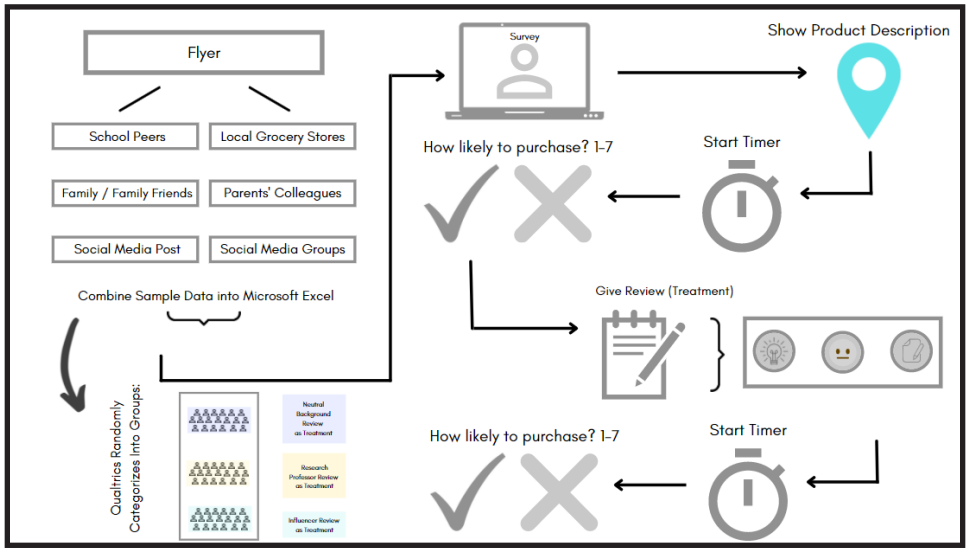


Figure 6. Treatment Design: Model of Assigning Treatment to participants and Experimental questions for main study



The data in this experiment is analyzed to understand the variations in time taken to make a purchase decision between each of the three professional backgrounds (experimental treatments). By finding the mean, median, and mode of initial and final time, the difference in time taken, and a T-Test value, it can be determined whether there was a statistical difference between each of the treatments – if the manipulated reviews contributed to time taken to make a purchase decision and if this was different amongst professional backgrounds.

Tables

Data Table: 1.1 – recorded for each participant

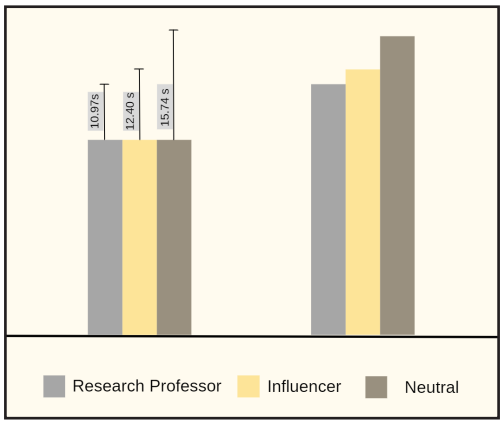
Participant Number:	*Initial Response (0-7 Scale)*	*Initial Time Elapsed (Seconds)*	*Initial Response Reasoning*	*Final Response (0-7 Scale)*	*Final Time Elapsed (Seconds)*	*Final Response Reasoning*
Treatment Category:						

Data Analysis Table: 1.2

	Mean Initial Time (Seconds)	*Mean Final Time (Seconds)*	*Difference in Time (Final – Initial Mean) (Seconds)*	*T Test Value*	*Statistical Difference? (Yes/No)*
Treatment 1: Neutral	55.47	71.21	15.74	0.0321	**Yes**
Treatment 2: Influencer	55.47	67.87	12.40	0.0265	**Yes**
Treatment 3: Research Professor	55.47	66.44	10.97	0.2314	**Yes**

Results

In conclusion, 300 participants show that there is a correlation between the professional background of the product reviewer with the time taken to make a purchase decision. This proves my hypothesis as correct. I predicted that if given a review written by different professional backgrounds for a standard product, the highest perceived credible author would influence the time taken to make a purchase decision by lowering it. With results from the Pre-Study, it can be concluded that the Professor had the highest perceived credibility. Difference in time between initial mean and then with review was 11 seconds (professor), 16 seconds (neutral), and 12 seconds (influencer).



With three treatments and their mean time, a one-way ANOVA test could be conducted to ensure significance between results. The p-value calculated showed a significance of 0.003, which is statistically significant. With these results, retail companies should ensure that the product reviews listed on their site include comments from individuals with professional backgrounds to ensure larger product traction amongst potential customers.

References

Dykstra, John. "The Critical Role of Reviews in Internet Trust - Trustpilot Business Blog." Trustpilot, 2017, https://business.trustpilot.com/guides-reports/build-trusted-brand/the-critical-role-of-reviews-in-internet-trust#downloadreport.

Kim, Sojung. "Credibility Cues in Online Shopping: An Examination of Corporate Credibility, Retailer Reputation, and Product Review Credibility." Research Gate, UT Austin, June 2012, https://www.researchgate.net/profile/Sojung-Kim-8/publication/264820999_Credibility_cues_in_online_shopping_An_examination_of_corporate_credibility_retailer_reputation_and_product_review_credibility/links/5510d5030cf20352196cab39/Credibility-cues-in-online-shopping-An-examination-of-corporate-credibility-retailer-reputation-and-product-review-credibility.pdf.

Kreutzberg, G. W. (2005, May). Scientists and the marketplace of opinions. scientific credibility takes on a different meaning when reaching out to the public. EMBO reports. Retrieved October 15, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299311/.

Limei, C., & Wei, L. (2020, December 10). The impact of reviewers' creditworthiness on consumers' purchase intention in edge path: Implications for the coronavirus disease 2019 pandemic. Frontiers in public health. Retrieved October 21, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7758319/.

Liu, Gang, et al. "An Empirical Study on Response to Online Customer Reviews and e-Commerce Sales: From the Mobile Information System Perspective." Mobile Information Systems, Hindawi, 28 Sept. 2020, https://www.hindawi.com/journals/misy/2020/8864764/.

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013, November 15). Social influence and the collective dynamics of opinion formation. PLOS ONE. Retrieved November 2, 2021, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0078433.

Wei, L. (2016, May 25). Decision-making behaviours toward online shopping. International Journal of Marketing Studies. Retrieved November 11, 2021, from https://www.ccsenet.org/journal/index.php/ijms/article/view/59367.

Using the Art of Kuchipudi to Treat Symptoms of Attention Deficit Hyperactivity Disorder

Bhavya S. Nandikanti

Nikola Tesla STEM High School

Abstract

Treatment options for attention deficit hyperactivity disorder (ADHD) currently are medication, therapy, or a combination of both. Medications often come with several negative side effects and only help users in the short term. Therapy can be expensive, and the methods learned in counseling have been reported to only help in the moment, which are not beneficial in the long term. Wouldn't it be amazing if there was a long-term, sustainable solution to help the 8.7% of the population struggling to complete daily tasks? This study looks at the effects of Kuchipudi (an Indian Classical Dance Form) on attention levels and impulsivity. This quasi-experiment sought adolescent female Kuchipudi dancers, who were required to complete four hours of dance practice weekly, for four weeks (Independent Variable). At the end of each week of dancing, they filled out two surveys testing their attention level and impulsivity (Dependent Variable). The control group was composed of adolescent females who were not practicing Indian Classical Dance. The hypothesis was that increased practice of Kuchipudi dance practice over time will increase attention levels and decrease impulsivity. The results supported my hypothesis by rejecting the null as there was a p-value of 0.001, proving the statistical significance of Kuchipudi dance practice improving attention levels. Additionally, the p-value of 0.003 proved the relationship between Kuchipudi practice and lowered impulsivity levels as statistically significant. These results support the holistic treatment approach of using Indian Classical Dance as a new form of Dance Movement Therapy to specifically treat symptoms of ADHD.

Introduction

The focus of this experiment is to see how the Indian classical dance form, Kuchipudi, affects attention levels and impulsivity, which are two major symptoms of attention deficit hyperactivity disorder. Many studies done in India show that Indian classical dance can have a positive impact on children suffering from ADHD. However, there are very few studies and experiments that correlate Kuchipudi (a form of Indian Classical Dance) specifically to attention and impulsivity. According to a study that was conducted in 2016, there are 3.3 million children in the United States between the ages of 12 and 17 that are diagnosed with attention deficit hyperactivity disorder (CDC). There were 25 million children in the United States as of 2017, meaning 13% of children were diagnosed with ADHD, and with ADHD being a difficult disorder to identify and diagnose it is likely that there may have been more children with it (Child Stats).

Currently, there are two treatment options for ADHD: medication and counseling. When a child is diagnosed with ADHD, their doctor will most likely recommend that they start their treatment with counseling, where therapists will work with the family along with the patient to develop techniques to cope with their symptoms. One form of therapy used is behavior therapy where the patient is taught behavior management and the therapist uses operant conditioning as a form of learning by rewarding the patient for good behavior. Coun-

seling is also used for social skills training where counselors work with the patient and are taught how to behave in certain situations that could aggravate their symptoms. The main idea with counseling is to help the patient learn to cope with their symptoms but there are several problems associated with it. Many ADHD patients have reported that coping methods are helpful in the moment but are not a long-term solution because there are not reliable. Additionally, researchers have found it most beneficial if used in combination with medication, which does not work for everyone. Additionally, therapy or counseling is very expensive and sometimes insurance cannot cover most of it, meaning low-income families will not be able to afford it. Therefore, there is a need for an inexpensive and long-term route for treatment.

The other treatment option is medication. There are two different broad categories under which the medication that could be prescribed lie: stimulants and non-stimulants. Stimulants are the best-known ADHD medication and according to the Mayo Clinic, 70-80% of children using them have fewer symptoms than if they took non-stimulants (Mayo Clinic). The two most common stimulants are Methylphenidate and Lisdexamfetamine. Methylphenidate is a central nervous system stimulant that is known to increase the activity in the user's brain specifically in the frontal lobe which is known to be the attention and behavior areas. Lisdexamfetamine is also a central nervous system stimulant that is reported to improve concentration, reduce impulsive behaviors, and is usually taken once a day. Additionally, it is important to note all ADHD medication prescribed is not a permanent cure (Mayo Clinic); they only improve the symptoms of the individual at the moment and are not good long-term substitutes. Additionally, there is not a perfect algorithm for prescribing ADHD medication because every patient responds differently to it. So, doctors start with the lowest dose, see how the patient responds and slowly monitor the side effects to note down if the dose needs to be increased or decreased. It takes a long time for the dosage to get perfected and as soon as it is, it might need to change again because a child's body is always growing and changing. Overall, in addition to the list of side effects that patients could experience, there are many other drawbacks to the use of medication.

The article written by Dr. Vijaya Pathloth called Therapeutic Healing Through Dance, is a theoretical study done on dance movement therapy (DMT) but specifically focused on Indian classical dance as the form of DMT. He says, "Dance therapy or dance movement therapy is the psychotherapeutic use of movement and dance for emotional, cognitive, social, behavioral, and physical conditions" (Pathloth). But, what is the difference between DMT currently in the States and the DMT mentioned in India? According to the American Dance Therapy Association, DMT is a form of expressive therapy where movement is used to describe emotion. It is used to help process feelings and is about making a connection between movement and emotion. Right now, according to the "Effects of Dance Movement Therapy and Dance on Health-Related Psychological Outcomes" (a meta-analysis update), DMT is now specifically designed to help with depression, dementia, post-traumatic stress disorder, chronic pain, arthritis, and anxiety. Dance movement using an Indian classical dance form has not yet been introduced to the United States. Additionally, DMT, in general, has not been used to treat symptoms of ADHD here, mostly because Indian Classical Dance (ICD) hasn't been introduced as DMT. Like Dr. Pathloth says "[Indian Classical] Dance is a great device to enhance and increase concentration levels" (Pathloth) meaning ICD is the key to treating attention levels.

To elaborate on the idea of ICD as DMT, "Effects of Indian Classical Dance Form-Kathak in Diagnosed Cases of ADHD Coming to a Tertiary Hospital Setup In Rural Area," conducted by Dr. Maneesha Deshpande and others in Nagpur, focused explicitly on the influence of Kathak on motor skills, and interestingly they used patients that were diagnosed with ADHD as their subjects. They saw that there was a statistically significant effect of increased motor skills with Kathak practice. Similarly, this experiment seeks to investigate the effect of Kuchipudi (an Indian Classical Dance Form) specifically, on attention and impulsivity levels. But what is different about Kuchipudi? Kuchipudi is one of the eight major Indian Classical Dance forms, and it is rooted in the ancient Hindu Sanskrit text of Natya Shastra. It is originated from a village called Kuchipudi located in Andhra Pradesh (a state in India). Like any other dance form, Kuchipudi is an expressive way to

share a story through specific movements. "Physical & Mental Health for Indian Classical Dance" discusses the several beneficial effects on the brain and body (Deekshitulu). Kuchipudi is known for its quick footwork, dramatic characteristics such as eye movement and hand gestures. It is a beautiful composition of precise, elegant, and flowy movement. Kuchipudi forces the dancer to have a good sense of vestibular and kinesthetic sense. Also, to perform the movements of Kuchipudi it's incredibly important to have a presence of mind because that's the first step to improving balance and being aware of your body. Eye movement is a simple movement in Kuchipudi but can demonstrate the essence of Kuchipudi in general: It starts by slowly and very subtly jerking the neck to the right, the neck must not be completely turning. This is done with the rest of the body in perfect uniform, meaning both feet are together, the knees are touching, with both hands in Anjali Mudra (Figure 1). As you jerk your neck to the right, you must move both eyeballs to the corner of the eye socket in the direction of your neck. The same thing must be done to the left side to complete the movement. Each jerk must be done to the beat (or thalam as they say in Sanskrit). This one simple movement forces the individual to pay attention to every part of their body.

If Kuchipudi is practiced for a long period, it could have long-term effects on an individual's attention levels, significantly improving their quality of life. This study believes that Kuchipudi will have a statistically significant increase in attention and impulsivity levels, and in theory, become a different form of dance movement therapy for patients struggling with attention deficit hyperactivity disorder.

Figure 1. Example of Anjali Mudra. Photo courtesy of Ditya Zemli from Alamy Stock Vector.

Methodology

Gathering participants. In this experiment there are children from the ages of 10 through 17, who are all female and are all Indian. This is a quasi-experiment because the selection of participants isn't completely random, and it only uses one gender. Additionally, this experiment will not be single-blind or double-blind but due to the use of questionnaires and the collection of quantifiable data the possibility of experimental bias will be minimized.

The sampling procedure started by finding a Kuchipudi Dance Academy with a Kuchipudi dance teacher that was willing to participate. One important material that's needed to replicate this experiment is the Experimenter Booklet, which was written by me. It has all the template messages that should be sent to the Academy, the teacher and potential participants. Additionally, it has copies of the impulsivity questionnaire used which is the UPPS-P Impulsivity Behavior Scale (Lynam, Smith, Whiteside, & Cyders, 2006) which has 59 questions where the participant is required to scale their agreement with a statement from one to four. It also has the attention survey, which is the ATTC, Attention Control Scale, containing 20 questions requiring participant to do perform a similar task. Once the teacher agreed to assist with the experiment, an email was sent to her with all the information for advertising the study to her students that met the requirement for age.

As participants from the academy were filling out the interest form, a consent form and final confirmation message was sent to them. The consent form was to ensure that the participants knew that their dance classes will continue to stay online, and the entire study will be done in the comforts of their home due to the COVID-19 protocols. Additionally, it makes sure that the participant understands what the study exactly is before confirming. While the participants for the experimental group were being recruited, the process for recruiting the control group also began with sending the flyer to everyone at the elementary and middle schools. Additionally, it was sent to other tutoring organizations and called the attention of all females, between ages 10-17, who were non-dancers. The first 25 kids, for the experimental and control group, that sent their consent form back, confirming their participation in the study were the final participants. All their information was kept in an excel which will be destroyed after the study to keep confidentiality. They were all assigned a unique number as reference to them throughout the study using a random number generator. Once all 50

participants were ready a final message was sent to them with their Participant Booklet which included their weekly schedules, tests for each week, rules of the study, and a note mentioning that they can leave the study whenever they want for whatever reason.

Data Collection. All the participants started the study on the same Sunday making it easier to keep track of data collection. For the experimental group the task for every Sunday was to create a schedule for the week (Figure 2), this way they knew ahead of time which day they had dance class for an hour (weekly twice) and which day they had to practice for 30 minutes. The goal was to complete 3 hours of practice every week to see the improvement of attention levels and impulsivity. After the schedule for the next week was created, they were required to send me a picture of it so I could keep track of their progress and send reminder messages for practice and classes.

After they completed their practice throughout the week, they were required to take the two tests for attention and impulsivity on Saturday and send pictures of it completed. Figure 3 outlines the process of data collection for the experimental group. These tests together took them 30 minutes to complete. As soon as the pictures were sent, I started to calculate their scores using the steps lined out in Figure 4 to score the impulsivity test and the steps in Figure 5 to calculate their score for the attention level test. To calculate the impulsivity test, there were six sub-section scores that described how good specific parts of their impulsivity were. The scores for these subsections and the overall average score were recorded in Table 1. Similarly, the attention level test only had two sub-sections and that was recorded in Table 2.

For the control group, they had to do the exact same thing but didn't have to do any dance practice. They were just required to take the two tests every Saturday and send me a picture of it completed. Once the study was over everyone received a certificate for participating, volunteering hours for the total hours spent on the study, and a thank you message. For the experimental group I told them their percent increase for attention levels, hopefully motivating them to continue the practice of dance. Figure 6 summarizes the process of data collection in a condensed flow chart.

Week 2 Schedule

- Schedule in one day of 30 minutes practice
- Schedule in your two-hourly dance classes
- Check the "Done" box once you are done with the activity on each day

Day	Activity	Done?
Monday		☐
Tuesday		☐
Wednesday		☐
Thursday		☐
Friday		☐
Saturday	- Complete test 1 ***(Pg 13 - 14)*** - Complete test 2 ***(Pg 15 - 18)*** - *Send picture of completed tests* - *Send picture of that week's schedule*	☐
Sunday	- Fill out **Week 3 Schedule** ***(Pg 19)*** - *Send picture of the filled out schedule* - *Send picture of completed* ***Week 2 Schedule***	☐

Figure 2. Example of a template schedule for the participant for week 2. Instructions for how to use this schedule are noted in the image.

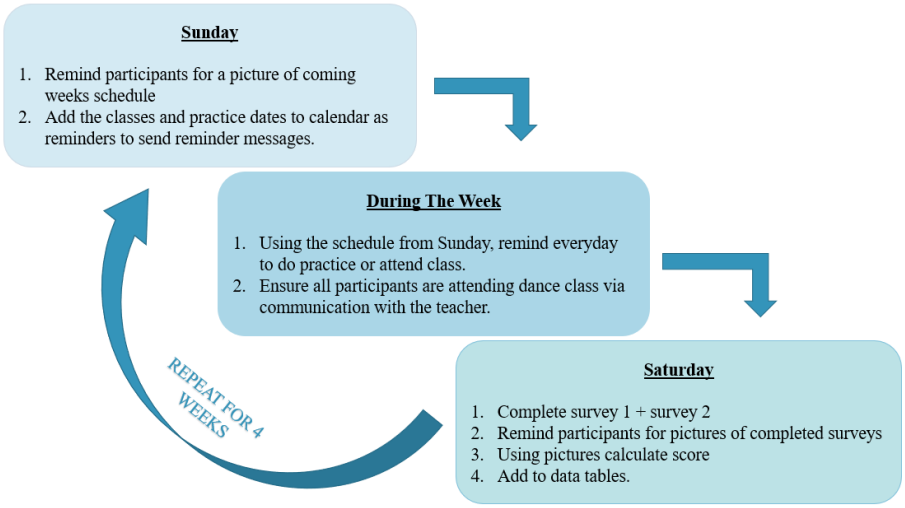


Figure 3. Flowchart of how data will be collected from each participant from the experimental group.

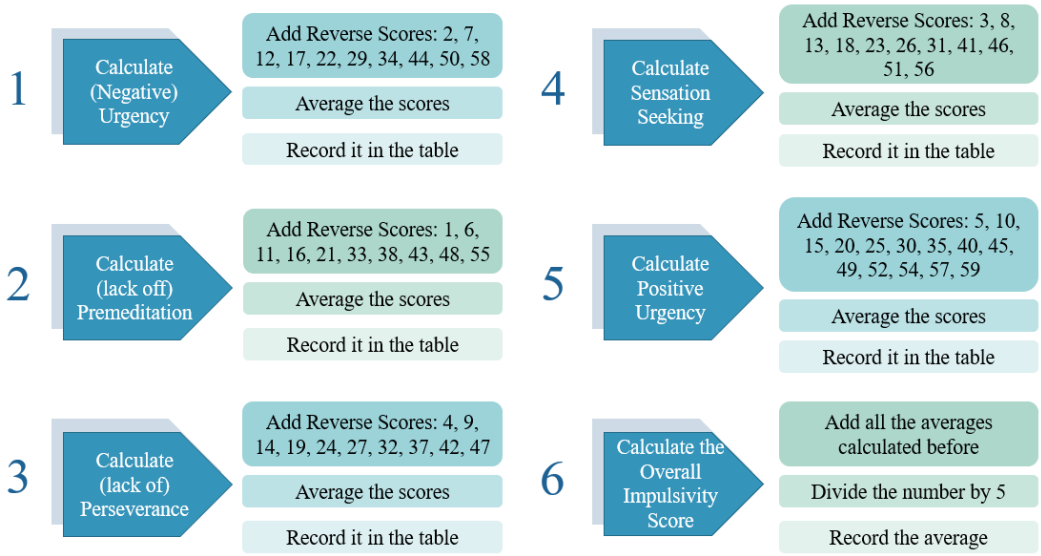


Figure 4. Flowchart of exactly how the scores need to be calculated for the Attention Levels questionnaire that all participants are required to complete.

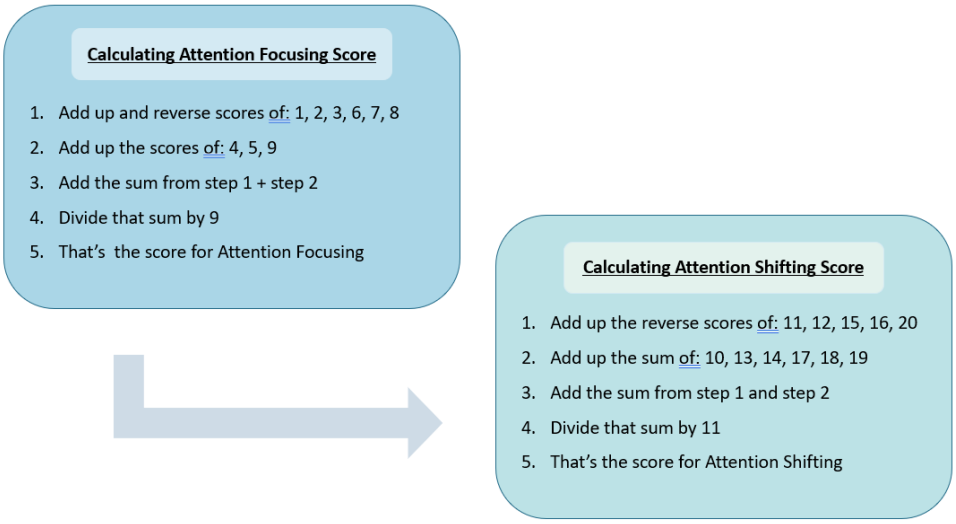


Figure 5. Flowchart of how to collect the scores of the Attention Levels questionnaire. There are two sperate scores that needs to be calculated individually before finding the overall score for attention level.

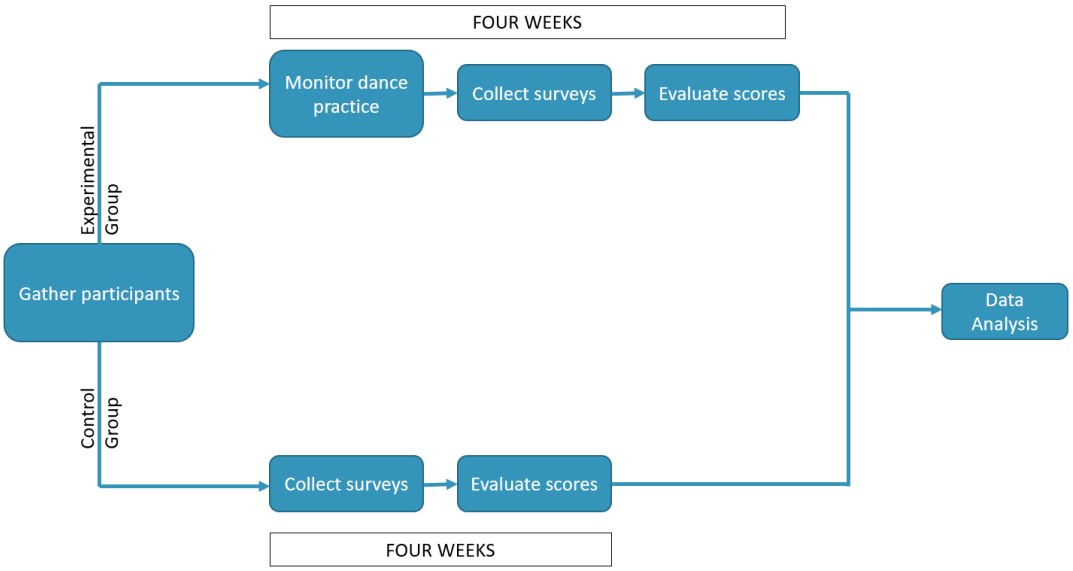


Figure 6. Condensed flow chart to visualize the steps taken to complete the data collection.

Table 1. Calculated impulsivity scores are stored in this data table. Each participant receives their own number which is how they are referred to throughout the experiment. Average score is all the section specific scores averaged to get the overall impulsivity score for that particular individual.

Calculating Impulsivity Scores						
Participant Number	Negative Urgency	Lack of Premeditation	Lack of Perseverance	Sensation Seeking	Positive Urgency	Average Score

Table 2. Calculated attention level scores are stored in this data table. Each participant receives their own number which is how they are referenced throughout the experiment. Average score is the Attention focusing score plus the attention shifting score averaged for an overall for individual's attention level.

Calculating Attention Level Scores			
Participant Number	Attention Focusing Score	Attention Shifting Score	Average Score

Results

Data analysis to calculate the results for the study started with data collection. The scores for all four weeks were calculated for both the UPPS survey (which was testing for impulsivity) and the ATTC survey (which was testing for attention levels). The higher the score for ATTC the higher their attention levels and the higher their score for UPPS the better their impulsivity control is. These scores were calculated for all the participants in both the control group and experimental group. To see the difference from their initial survey and the very last survey they took, I decided to take the percent increase between their scores from week 0 and week 3. In the control group there was barely any difference, so the percent increase was almost always less than 1%. However, in the experimental group there was a percent difference for both the variables which was between 3 and 15 percent. All the percent increases that were collected were plotted on a scatter plot to visualize the difference between control group and experimental group (Figure 7 & Figure 8). To find the statistical significance between attention levels/impulsivity and prolonged Kuchipudi practice, I decided to carry out a Chi Square test. For the variable of attention my null hypothesis was that a prolonged period of Kuchipudi dance practice will have no effect on the participant's attention levels. With my chi square value being 0.001 I was successfully able to reject my null hypothesis, proving that my data is statistically significant. For my impulsivity variable I had a similar null hypothesis and got a chi square value of 0.001 and was able to reject my null once again.

My data is statistically significant proving that a prolonged period of Kuchipudi dance practice has an affect on the symptoms of ADHD, specifically a person's attention level and impulsivity. Although there were a few data limitations that I experienced. One main limitation would be my sample size, I had 30 participants in each my control and experimental group. But all these participants were Indian females meaning it's hard for my data to be generalized to the general population. Which is a problem that should be tackled in future studies. In addition to this the reason my chi square value could be so low, is probably because a lot of my participants in the control group showed 0 percent increase driving my chi square down. Finally, as mentioned before, the participants that were used were not diagnosed with ADHD clinically and were tested in the symptoms of ADHD. Although this is a substantial impact on the study, I believe that it is the best first step for future studies. Now that we know that there is a major difference in the symptoms with the use of Kuchipudi, we can be confident in applying this as a reason to further study the impact of Kuchipudi on clinically diagnosed participants.

Discussion

My long-term aspiration for this project is for a new branch of dance movement therapy to be created. As mentioned in the introduction, right now dance movement therapy is specifically focused on depressive disorders plus the mind and body connection. There's a fixed mindset about how dance movement therapy can be applied to mental health disorders which is all about flowy movements and feeling your emotions through your movements. Although I believe that dance is more than that, it can be used to treat a multitude of other disorders. Dance isn't just about connecting your mind to your soul, through dance you can learn discipline, structure and how to balance that with elegance and free movement. Through learning these skills, the dancer is benefiting everywhere in life. Their attention levels will allow them to be better listeners, a better friend, and a better student. Controlling their impulses will help them make careful decisions.

In addition to this I wanted to bring in a new perspective into medicine. The art of Kuchipudi and many other Indian classical dances go unnoticed in our Western world. Not only are they beautiful to watch but they teach the viewers and the dancers so much. They are the perfect mixture of structure and elegance. A lot of Indian parents push their children to join Indian Classical Dance to stay connected to their culture, but little do they realize that they're helping their child build lifelong skills that will make them a better person.

Let this study be the first step into revolutionizing dance therapy and adding a whole new branch specifically to treat attention disorders through using other dance forms, like Kuchipudi.

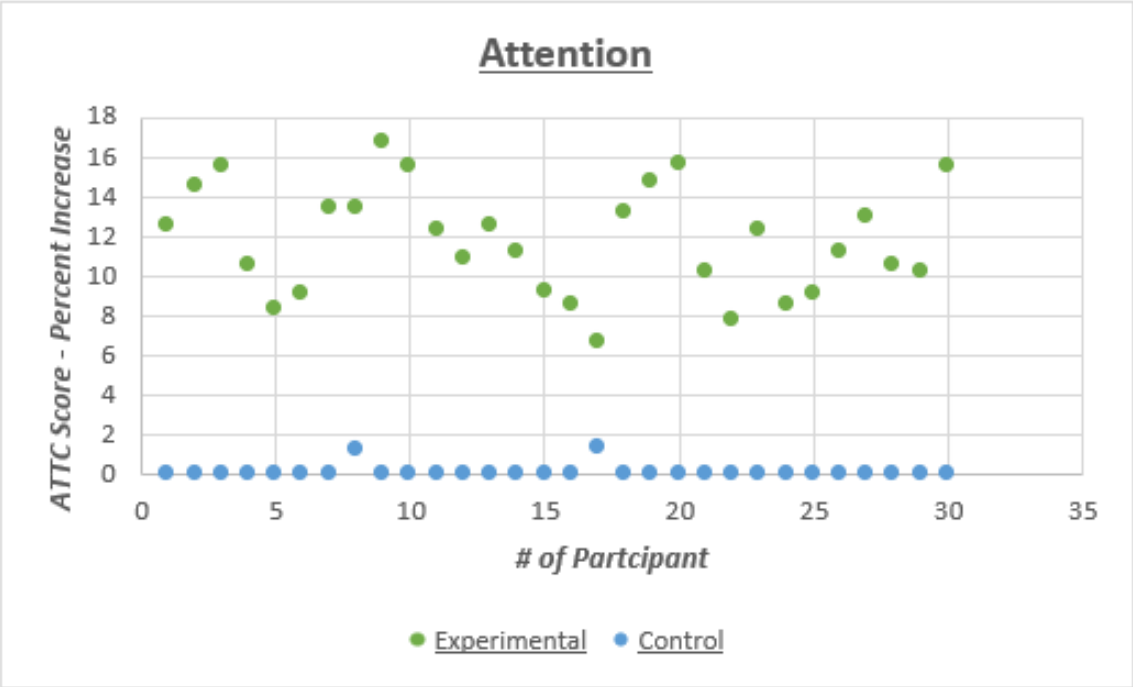


Figure 7. Percent increase of all the participants from control group and experimental group plotted on a scatter plot for Attention Level from the ATTC survey.

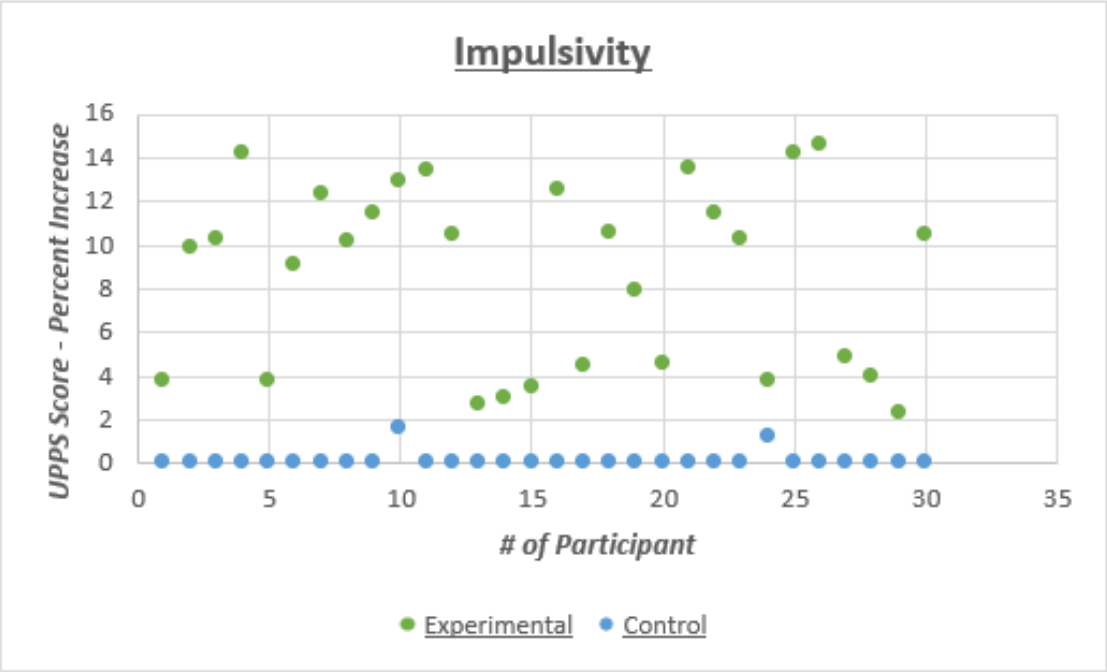


Figure 8. Percent increase of all the participants from control group and experimental group plotted on a scatter plot for Impulsivity Level from the UPPS survey.

Bibliography

Centers for Disease Control and Prevention. (2021, September 23). *What is ADHD?* Centers for Disease Control and Prevention. Retrieved October 25, 2021, from https://www.cdc.gov/ncbddd/adhd/facts.html

Deekshitulu , B. (2019, January 9). Physical & Mental Health for Indian Classical Dance. Tirupati; American Journal of History and Culture.

Deshpande, M., Joshi, M., & Thakar, C. (2019, February). Effect Of Indian Classical Dance Form-Kathak In Diagnosed Cases Of ADHD Coming To A Tertiary Hospital Setup In Rural Area. Nagpur; Indian Journal of Applied Research.

Inman, H. (2021, May 16). Dance Mo Dance Movement Treatment Therapy for Clients with A y for Clients with Attention Deficit attention Deficit Hyperactivity Disor activity Disorder: A Liter der: A Literature Review. Expressive Therapies Capstone Theses.

Koch, S. C., Riege, R. F. F., Tisborn, K., Biondo, J., Martin, L., & Beelmann, A. (2019, August 20). Effects of Dance Movement Therapy and Dance on Health-Related Psychological Outcomes. A Meta-Analysis Update. Frontiers in Psychology.

Mayo Clinic Staff. (2019, June 25). *Attention-deficit/hyperactivity disorder (ADHD) in children*. Mayo Clinic. Retrieved October 27, 2021, from https://www.mayoclinic.org/diseases-conditions/adhd/diagnosis-treatment/drc-20350895

Shetty, S. J., & Shetty, A. P. (2015, January). Effectiveness of dance movement therapy on attention deficit hyperactivity disorder children aged between 6-12 years. Mangalore; Manipal Journal of Nursing and Health Sciences.

Achievements/Awards

65th Annual Washington State Science and Engineering Fair (WSSEF)

March 2022

Finalist judging on April 2, 2022.
(Tesla students won sixty awards, twenty-one of them were 1st place awards)

We had 3 student projects receive ISEF (International Science & Engineering Fair) bids and place in the top seven best projects:

Anjali Sreenivas *received a bid to ISEF and CSRSEF,*

Druhin Bhowal *and* **Arihant Singh** *also received an ISEF bid at CSRSEF, and*

Evan Kim *and* **Harish Krishna Kumar** *received ISEF bids at WSSEF.*

Cindy Chen, *Facial and Emotional Recognition Among Caucasian and East Asian Adolescents*
1st place in Behavioral and Social Sciences Category, Senior Division

Akul Garg and Luka Jovcic, *Correlation Between Social Media and Anxiety in High School Students Using the SCARED*
1st place in Behavioral and Social Sciences Category, Senior Division

Aubrey Grove*, Assessment of Yoga Therapy of Mindfulness-Based Therapy Interventions of Symptom Severity and Quality of life in a Cohort of Adults with Inflammatory Bowel Disease*
2nd place in Translational Medical Sciences Category, Senior Division

Druhin Bhowal and Arihant Singh, *It's Flaming Out: Using Artificial Intelligence to Emulate Critical Aspects of Wildfire Growth*
1st place in Earth and Environmental Sciences Category, Senior Division
Sigma Xi Scientific Research special category award

Special Note - advanced to 2nd round of Finalist judging on April 2nd. Only 1 of 20 projects out of over 300 that competed in fair to advance to this round for consideration in awarding an ISEF bid.

Sneha Murali, *Using an Epidemiological Model to Determine Impact of Manual and Digital Contact Tracing on Preventing COVID-19 Disease Spread*
2nd place in Computational Biology and Bioinformatics category, Senior Division

Danyuan Wang, *Using Machine Learning Methods to Estimate the Distribution of Harmful Invasive Phytoplankton*
1st place in Plant Sciences Category, Senior Division
Statistical Thinking special category award

Ananya Soni, The *Effect of Cognitive Behavioral Therapy App (Happify) on Adolescent Anxiety*
2nd place in Behavioral and Social Sciences Category, Senior Division

Anjali Sreenivas, *A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease*
1st place in Computational Biology and Bioinformatics category, Senior Division
Sigma Xi Scientific Research special category award
Yale Scientific and Engineering Association Award
Wolfram Research, Inc. Mathematica Certificate of Achievement
Silver Medalist Winner - placed 2nd in Fair - Regeneron ISEF Finalist

Anya Vaish, *Bioinformatic Analysis to Find a Universal Sequence from SARS-CoV-2 Genomes for the Potential Development of a Pancoronovirus Vaccine*
1st place in Computational Biology and Bioinformatics category, Senior Division

Hahwon Chue, *Rubber Tire Leachate Effect in Lake Washington and Lake Sammamish*
1st place in the Earth and Environmental Sciences Category, Senior Division
Wolfram Research, Inc. Mathematica Certificate of Achievement

Thanushree Karunagaran *Mobile Application to Improve the Cognition of Dementia and Alzheimer's Disease Patients*
1st place in Biomedical Engineering Category, Senior Division

Karina Kerjiwal, *The Impact of Credibility When Hitting "Purchase Now"*
1st place in Behavioral and Social Sciences Category, Senior Division

Evan Kim, *Utilizing a Generative Adversarial Network to Predict Hypothetical Superconductors*
1st place in the Materials Science category, Senior Division
Olympic College STEM Award
Wolfram Research, Inc. Mathematica Certificate of Achievement
One of top 7 projects in the fair - Regeneron ISEF Finalist

Praharsha Manda, *Development of the Education IAT*
2nd place in Behavioral and Social Sciences Category, Senior Division

David Gershony, *Combatting Anxiety in Dental Offices Using Virtual Reality*
1st place in Translational Medicine Category, Senior Division
WA State Dental Association of Oral Health award

Emily Liu, *Type 1 Diabetes Alters Perineural Net Assembly Surrounding Hypothalamic Glucoregulatory Circuitry*
1st place in Biomedical and Health sciences category, Senior Division

Shivani Godse and Tanvi Sapar, *Creating a Structural Model of SaSrtA-Inhibitor Complex to Predict Efficiency of Class A Sortase Enzyme Inhibitor Molecules*
1st place in the Chemistry category, Senior Division

Special Note - advanced to 2nd round of Finalist judging on April 2nd. Only 1 of 20 projects out of over 300 that competed in fair to advance to this round for consideration in awarding an ISEF bid.

Adhya Kona, *An Evolutionary Origin Study Based on Genetic Homology*
2nd place in Computational Biology and Bioinformatics category, Senior Division

Harish Krishna Kumar, *Analysis of Ring Galaxies Detected Using Deep Learning with Real and Simulated Data*
1st place in Physics and Astronomy category, Senior Division
Wolfram Research, Inc. Mathematica Certificate of Achievement
One of top 7 projects in the Fair - **Regeneron ISEF Finalist**

Monisha Krothapalli, *Engineering a Surface Flow Constructed Wetland to Evaluate Efficiency for Combatting Water Scarcity*
1st place in the Earth and Environmental Sciences Category, Senior Division
US Stockholm Junior Water Prize
Water Purveyor Association for Water Quality special category award

Bhavya Nandikanti, *Using the Art of Kuchipudi to Treat Symptoms of Attention Deficit Hyperactivity Disorder*
2nd place in Behavioral and Social sciences category, Senior Division
Dr. Devagupta - Henrietta Lacks special category award

Siddhant Porwal, *Predicting Waterfall Forms Through Feature Engineering and Model-Selection Machine Learning*
1st place in the Environmental Sciences category, Senior Division
MU Alpha Theta special category award
NASA Langley Research award
US Stockholm Junior Water prize

Nuha Rahman and Madhumitha Rasiah, *Stand Up To Stigma: ADHD in Women*
3rd place in Behavioral and Social sciences category, Senior Division

Rhea Kuppa, *3D Printing Bio-Inspired Heart Valves to Increase Efficiency of Perioperative Care*
1st place in Biomedical Engineering category, Senior Division
ASM Materials Education Foundation special category award
Statistical Thinking special category award

Divya Singh, *Combatting Depression Using the Gut-Brain Axis Via Probiotic Treatments*
1st place in Biomedical and Health sciences category, Senior Division
Wolfram Research, Inc. Mathematica Certificate of Achievement

Miles Martinez, *Memory of Notes by Hand Versus Keyboard*
2nd place in Behavioral and Social sciences category, Senior Division

Aastha Shah, *Evaluating the Relationship Between Stress and Impulsivity in a Mathematical Context*
1st place in Behavioral and Social sciences category, Senior Division

Pierce Zhang, *Computational Prediction of Oncogenic Mutations*
1st place in Computational Biology and Bioinformatics Category, Senior Division

Edward Zhou, *Dampening Phantom Shockware Jams in Mixed-Autonomy Traffic Via Deep-RL Control of Autonomous Vehicles*
1st place in the Robotics and Intelligent Machines category, Senior Division
City of Bremerton Mayors award
Special Note - advanced to 2nd round of Finalist judging on April 2nd. Only 1 of 20 projects out of over 300 that competed in fair to advance to this round for consideration in awarding an ISEF bid.

Rithani Saravanakumar, *Machine Learning Guided Selection of Bioinks for 3D Printing of Tissue Scaffolds*
Absci-Scientific Unlimited Special category award

Ghousia Shahzad, *Impact of Physical Activity on Personality and Emotions*
2nd place in Behavioral and Social sciences Category, Senior Division

Akhila Bourampeta, *Oxytocin as a Strategy to Enhance the Effectiveness of the Satiation Signal, Cholecystokinin, to Reduce Food Intake in High Fat Diet-Fed Rats*
3rd place in Biomedical and Health sciences Category, Senior Division

Kai Vanderlip, *A Scientific Look into Mental Visualization and Improved Performance in Taekwondo*
2nd place in Behavioral and Social Science category, Senior Division

Laya Nair, *Parkinson's Disease: A Risk Prediction Model Using Inflammatory Bowel Disease*
2nd place in Biomedical and Health Sciences category, Senior Division

Sanya Sahni, *Overexposure to Maternal Cortisol during Gestation and its Influence on the Regulation of HPA axis (relationship with stress) in Female Adults*
2nd place in Behavioral and Social Sciences category, Senior Division

Gauri Raman, *Using Bright Light Therapy to Effectively Treat Night Eating Syndrome*

3rd place in Behavioral and Social Science category, Senior Division

Vijaya Sripada, *Using Various Inputs of Smell to Reduce Depression Score in Alzheimer Patients*
2nd place in Behavioral and Social Science category, Senior Division

Central Sound Regional Science & Engineering Fair (CSRSEF)

March 2022

24 awards, 2 grand prize winners moving on to the Regeneron International Science and Engineering Fair (Regeneron ISEF).The world's largest international pre-college science competition

Grand Prize winners - Advanced to ISEF

Anjali Sreenivas, *A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease*
Arihant Singh and **Druhin Bhowal**, *It's Flaming Out: Using Artificial Intelligence to Emulate Critical Aspects of Wildfire Growth*

Behavioral and Social Sciences
1st place - **Cindy Chen**, Facial and Emotional Recognition Among Caucasian and East Asian Adolescents
2nd place - **Aastha Shah**, Evaluating the Relationship Between Stress and Impulsivity in a Mathematical Context
3rd place - **Vijaya Sripada**, Using Various Inputs of Smell to Reduce Depression Score in Alzheimer Patients
4th place - **Praharsha Manda**, Development of the Education Implicit Association Test (IAT)

Biomedical Engineering and Translational Health
1st place - **Rhea Kuppa,** 3D Printing Bio-Inspired Heart Valves to Increase Efficiency of Perioperative Care
2nd place - **Aubrey Grove,** Assessment of Yoga Therapy of Mindfulness-Based Therapy Interventions of Symptom Severity and Quality of life in a Cohort of Adults with Inflammatory Bowel Disease
3rd place - **David Gershony**, Combatting Anxiety in Dental Offices Using Virtual Reality

Software, Embedded Systems, Robotics
2nd place - **Edward Zhou**, Dampening Phantom Shockware Jams in Mixed-Autonomy Traffic Via Deep-RL Control of Autonomous Vehicles

Computational Biology & Bioinformatics (COMP)
1st place - **Anjali Sreenivas**, *A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease*

3rd place - **Anya Vaish**, *Bioinformatic Analysis to Find a Universal Sequence from SARS-CoV-2 Genomes for the Potential Development of a Pancoronovirus Vaccine*

Organismal Biology

3rd place – **Danyuan Wang**, *Using Machine Learning Methods to Estimate the Distribution of Harmful Invasive Phytoplankton*

Chemistry, Earth, Environmental Sciences and Technology
1st place - **Arihant Singh and Druhin Bhowal**, *It's Flaming Out: Using Artificial Intelligence to Emulate Critical Aspects of Wildfire Growth*

2nd place - **Evan Kim,** *Utilizing a Generative Adversarial Network to Predict Hypothetical Superconductors*

Biomedical, Molecular and Cellular Biology
3rd place - **Tanvi Sapar and Shivani Godse**, *Creating a Structural Model of SaSrtA-Inhibitor Complex to Predict Efficiency of Class A Sortase Enzyme Inhibitor Molecules*

Physics, Astronomy, and Mathematics
1st place - **Harish Krishna Kumar,** *Analysis of Ring Galaxies Detected Using Deep Learning with Real and Simulated Data*

Mechanical Engineering
3rd place - **Anisha Vaish**, *Automated Pill-Dispenser to Improve Medication Compliance Efficiency and Accuracy*

Special Category Awards:

Regeneron Biomedical Sciences Award ($500)
Anjali Sreenivas, *A Machine Learning Approach to Identifying Blood-Based Biomarkers for Differential Diagnosis of Alzheimer's Disease*

American Psychological Association (APA)Award
Karina Kejriwal, *The Impact of Credibility When Hitting "Purchase Now"*

American Meteorological Society Award
Suyash Mothukuri, *Diagnosing Precipitation from IR Satellite Images*

Materials Science Special Category Award
Evan Kim, *Utilizing a Generative Adversarial Network to Predict Hypothetical Superconductors*

Stockholm Junior Water Prize
Siddhant Porwal, *Separating Microplastics from Sand using Terminal Velocity and Density Differences*
Monisha Krothapalli, *Engineering a Surface Flow Constructed Wetland to Evaluate Efficiency for Combatting Water Scarcity*

Future Business Leaders of America (FBLA) Regional Conference

February 2022

Presentation Events:

Banking and Financial Systems: Pia Dave and Kaavya Trivedi 5th place

Broadcast Journalism: Omid Farahdel and Anirudh Seshadri 6th place

Business Management: Diya Karthic, Mehek Sathe, Nihan Tatli 3rd place; Eesha Kunisetty 2nd place

Data Analysis: Kavya Srikumar 2nd place

Digital Video Production: Kshitij Rao, Anirudh Seshadri, Rohan Singh 5th place

E-Business: Annie Adhikary 6th place; Omid Farahdel, Pranav Sayal, Brian Yin 4th place

Electronic Career Portfolio: Cindy Chen 2nd place

Entrepreneurship: Emma Shi, Osheen Tikku 3rd place

Graphic Design: Diya Karthic, Mehek Sathe, Nihan Tatli 6th place; Kshitij Rao, Manovay

Sharma, Ayush Singh 5th place

International Business: Divij Aswinkumar, Farzad hasan 5th place

Introduction to Business Presentation: Vasudah Narayanan, Anjana Punniamoorthy 5th place; Prakshi Shukla, Nikki Taleghani 3rd place

Introduction to Public Speaking: Nitya Kakulamarri 6th place; Annie Adhikary 4th place,

Anshu Aggarwal 3rd place

Introduction to Social Media Strategy: Abhinav Nadupalli, Alekhya Nidadavolu 6th place; Ahanya John, Megha Ramachandran, Alexa Rorrer 5th place

Job Interview: Eesha Kunisetty 2nd place

Marketing: Rishi Dilip, Aahil Irfan, Sourish Singh 6th place; Shivani Godse, Sneha Murali 5th place

Public Service Announcement: Gayathri Pillai, Shivani Sama, Shriya Tiku 2nd place; Sarah Wen 1st place

Public Speaking: Nikki Taleghani 2nd place

Publication Design: Aaryan Bondre, Tanush Contractor, Gyan Dave 4th place

Sales Presentation: Rhea Kuppa, Shreya Pandey 5th place; Karina Kejriwal, Bhavya Nandikanti 4th place; Divij Aswinkumar, Saathvik Somujayabalan, Harini Vijeyanandh 3rd place

Social Media Strategies: Cindy Chen, Karina Kejriwal, Bhavya Nandikanti 3rd place; Gauri Raman, Divya Singh 1st place

Sports and Entertainment Management: Gyan Dave, Aarav Khanna, Isayiah Lim 4th place

Website Design: Kohki Kita 3rd place

Objective tests:

Accounting I: Kohki Kita

Advertising: Sneha Murali 4th place; Hrudai Thungathurthi 5th place

Agribusiness: Katelyn Ye 1st place *(from LWHS but part of our chapter)*

Computer Problem Solving: Rohit Karthik 1st place; Michael Yu 4th place; Keshav Acharya 5th place

Cybersecurity: Michael Yu 3rd place

Economics: Isayiah Lim 5th place

Human Resource Management: Eesha Kunisetty 3rd place; Felix Hallmann 4th place

Introduction to Business Communication: Vasudha Narayanan 1st place

Introduction to Financial Math: Johan Karukayil 5th place; Divya Ramu 6th place

Introduction to Information Technology: Johan Karukayil

Journalism: Shriya Tiku 6th place

Network Infrastructures: Michael Yu 3rd place; John Karaca 5th place

Organizational Leadership: Isha Rajpul 5th place

UX Design: Kohki Kita 4th place

FBLA State conference

April, 2022

Students who placed 1 – 4 automatically advance to Nationals in Chicago this summer. Fifth – tenth place have a possibility of going.

Out of 1500 competitors throughout the state in 60 different events, here are our results:

Coding & Programming: Rajat Sengupta; Arihant Singh	4
Coding & Programming: Rishi Dilip; Aahil Irfan; Sourish Singh	5
Computer Game & Simulation Programming: Keshav Acharya	6
Data Analysis: Kavya Srikumar; Katelyn Ye	4
E-business: Omid Farahdel; Pranav Sayal; Brian Yin	1
Mobile Application Development: Vaishnavi Rayaprolu	3
Website Design: Prisha Agarwal; Shivani Sama	4
Future Business Leader: Harini Vijeyanandh	2
Business Management: Diya Karthic; Mehek Sathe; Nihan Tatli	2
International Business: Divij Aswinkumar; Farzad Hasan	2

Marketing: Rishi Dilip; Aahil Irfan; Sourish Singh	6
Sports & Entertainment Management: Gyan Dave; Aarav Khanna; Isayiah Lim	5
Advertising: Hrudai Thungathurthi	6
Computer Problem Solving: Rohit Karthik	1
Computer Problem Solving: Keshav Acharya	6
Economics: Isayiah Lim	8
Introduction to Financial Math: Divya Ramu	5
Introduction to Information Technology: Johan Karukayil	7
Organizational Leadership: Isha Rajpure	7
Broadcast Journalism: Omid Farahdel; Anirudh Seshadri	1
Electronic Career Portfolio: Cindy Chen	1
Graphic Design: Diya Karthic: Mehek Sathe	1
Graphic Design: Kshitij Rao; Manovay Sharma; Ayush Singh	5
Introduction to Business Presentation: Prakshi Shukla; Nikki Taleghani	5
Public Service Announcement: Sarah Wen	5
Public Service Announcement: Gayathri Pillai; Shivani Sama; Shriya Tiku	8
Publication Design: Aaryan Bondre; Tanush Contractor; Gyan Dave	7
Sales Presentation: Divij Aswinkumar; Saathvik Somujayabalan; Harini Vijeyanandh	1
Sales Presentation: Karina Kejriwal; Aarav Khanna; Bhavya Nandikanti	2
Sales Presentation: Rhea Kuppa; Shreya Pandey	3
Social Media Strategies: Gauri Raman; Divya Singh	3
Social Media Strategies: Aaryan Bondre; Tanush Contractor; Bhavya Nandikanti	9
3-D Animation: Diya Karthic; Mehek Sathe; Nihan Tatli	3
Digital Video Production: Kshitij Rao; Anirudh Seshadri; Rohan Singh	6
Business Financial Plan: Alekhya Nidadavolu; Harshita Raja	6
Public Speaking: Nikki Taleghani	2

FBLA National Conference

June/July, 2022

Tesla STEM High School took 44 students to Chicago, IL to compete in the FBLA National Conference. There were approximately 11,000 students competing. Our results are below:

Rohit Karthik 9th place Computer Problem Solving
Divya Ramu 4th Place Introduction to Financial Math
Bhavya Nandikanti, **Karina Kejriwal, Aarav Khanna** 3rd place in Sales Presentation.

Holocaust Center for Humanity's Writing, Art and Film Contest

May 13, 2022

This year, the center received almost 600 entries representing 46 schools! Winning entries are posted on the Holocaust Center for Humanity website

ART 9th – 12th Grade:

Cameron Yetzer	1st Place

FILM 9th – 12th Grade:

Maxwell Soh	1st Place
Phillip Araujo	2nd Place
Rohan Iyer	3rd Place

Intel International Science and Engineering Fair (ISEF)

May 2021

American Meteorological Society

Second Award of $1,000

EAEV080 — *Predicting the Extent of Late Seasonal Wildfires in Washington Using the Correlation Between the Standardized Precipitation Index and Burned Area Data*

- **Saketh Dhulipalla**

National Oceanic and Atmospheric Administration – NOAA

First Award of $1,500

EAEV057 — *Path Dependence of Atlantic Meridional Overturning Circulation Weakening: A Geostrophic Shear Approach*

- **Yuchen Li**

Behavioral and Social Sciences, sponsored by Society for Science

Fourth Award of $500

BEHA030 — *Identification of Co-Expressed Genes to BDNF and trk-B as Major Depressive Disorder Related Biomarkers Using Microarray Data*

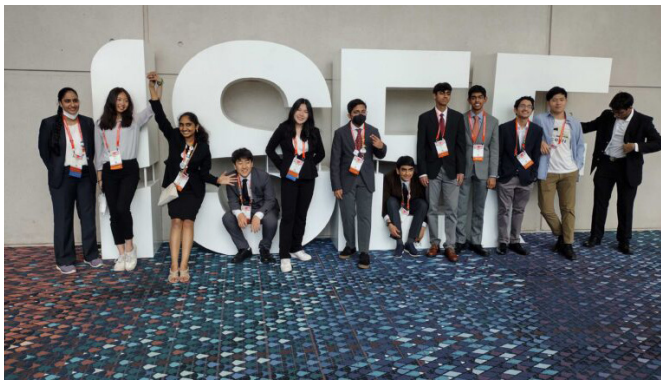
- **Jennifer Hu**

Computational Biology and Bioinformatics, sponsored by Regeneron

Fourth Award of $500

CBIO060 — *Predicting Cancer Stem Cell Biomarkers with Machine Learning*

- **Julia Liu**



Featured on **GeekWire**

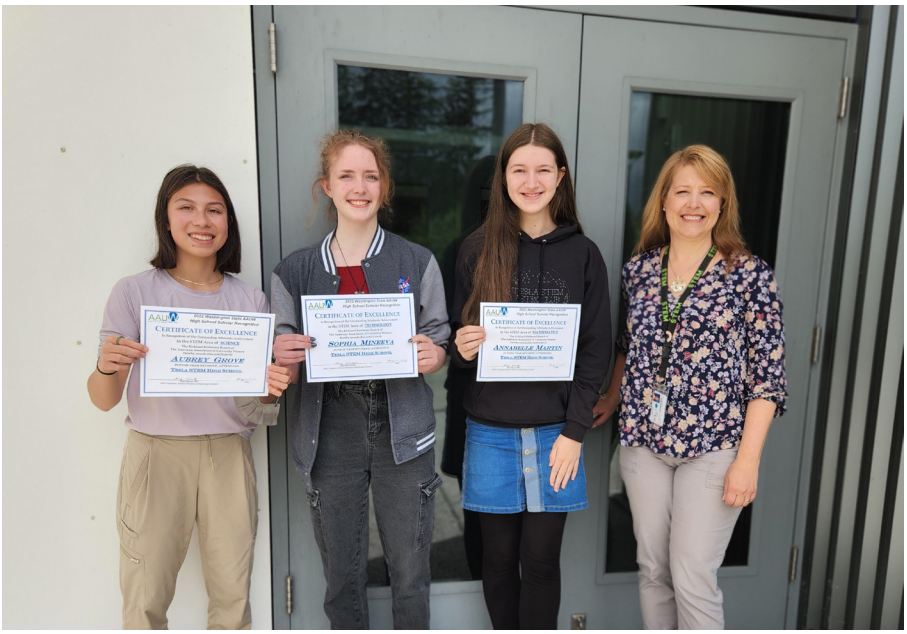
https://www.geekwire.com/2022/meet-the-seattle-area-teen-geeks-that-just-won-awards-at-an-international-science-fair/

American Association of University Women (AAUW) Awards for Outstanding Female Students in the World of STEM

Math - Alexia Friedman
Science - Lauren Kim
Technology - Christine Lee

American Association of University Women (AAUW) High School Scholar Award

Aubrey Grove
Alex Sophia Mineeva
Annabelle Martin



The Seattle Youth Poet Laureate (YPL) program

The Seattle Youth Poet Laureate (YPL) program aims to identify youth writers and leaders committed to poetry, performance, civic and community engagement, education, and equity across the Puget Sound region.

Adhya Kona - selected to join the Seattle Youth Poet Laureates (YPL) cohort for the 2022-23 year.

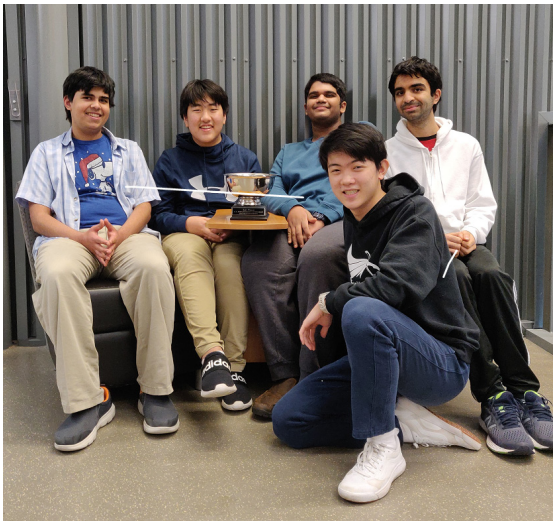
A video of Adhya's Folklife reading is here: https://drive.google.com/file/d/1bxsn-P1LYP4BsR1X2j33VyyP_BhM53zx/view

2020 Pacific Northwest Regional Science Bowl

Team 1 won the competition! The team of **Sebi Kumar, Evan Kim, Sami Badal Khan, Yuchen Li, and Rohan Iyer** had the highest score in each of the seven rounds.

Team 2 finished in 4th place! This team, **Prerana Kota, Cheguvijay Aashray, Reese Long, Tejas Pakalapati, and Arihant Singh,** was the highest-finishing second team.

Team 3 finished in 7th place! This team, **Eva Jain, Danyuan Wang, Shuhui Yang, Aakarsh Balla, and Johan Karukayil,** was the only third team to advance to the round of 7 teams (the quarterfinals).



Team 1:

2022 National Science Bowl

Final results

Sixty-one teams advanced to the National Science Bowl. This year, instead of flying all of them to Washington, DC, the Federal Government hosted an online tournament to determine the final 8 teams that will fly to the Capitol and compete.
*On May 27th, the Tesla STEM Science Bowl team--**Yuchen Li, Rohan Iyer, Evan Kim, Sebi Kumar, and Sami Badal Khan**--survived four elimination rounds to qualify as one of the final 8 teams in the country.*

Four members of the team, **Sami Badal Khan, Sebi Kumar, Yuchen Li, and Evan Kim,** traveled to Washington, DC and competed in July. They finished in **3rd place.**

2022 Washington Science Olympiad: Northwest Regional

36 teams participated in the Northwest Regional of the Washington Science Olympiad

Tesla STEM Green finished in 2nd place and qualified for the State Championship.
Tesla STEM Magenta finished in 6th place and qualified for the State Championship.
Tesla STEM Blue finished in 18th place and would have qualified for the State Championship (by finishing in the top half of all teams), but only two teams from any one school are allowed to move on.

Science Olympiad is played by teams of 15 students who compete in pairs and trios across 23 total events. They are stack-ranked in each event and the team with the lowest sum of rankings in the 23 events wins. The State Championship will be held on Saturday 9 April.

Of the 23 events at Regionals, Tesla STEM teams won medals in 20 events for finishing in the Top 8, as follows:

Anatomy & Physiology: Rohan Iyer & Anjali Sreenivas, 7th (Green)

Astronomy: Harish Krishnakumar & Evan Kim, 1st (Green)

Cell Biology: Rohan Iyer & Maxwell Soh, 1st (Green); Shuhui Yang & Sahaana Sridhar, 7th (Blue)

Chemistry Lab: Sebi Kumar & Sami Badal Khan, 1st (Green)

Code Analysis: Evan Kim & Aayush Sheth, 3rd (Green); Matthew Lam & Farzad Hasan, 6th (Magenta)

Codebusters: Annanya Unnikrishnan, Ananya Soni & Matthew Lam, 8th (Magenta)

Detector Building: Divij Aswinkumar & Farzad Hasan, 2nd (Magenta)

Disease Detectives: Anjali Sreenivas & Maxwell Soh, 2nd (Green), Eva Jain & Krish Katre, 8th (Magenta)

Dynamic Planet: Yuchen Li & Aashray, 1st (Green); Ashwin Kaliyaperumal, 6th (Blue); Matthew Lam & Ethan Liao, 8th (Magenta)

Environmental Chemistry: Yuchen Li & Sebi Kumar, 1st (Green); Prerana Kota & Siddhant Porwal, 7th (Magenta)

Experimental Design: Anjali Sreenivas, Aayush Sheth & Harish Krishnakumar, 3rd (Green)

Forensics: Rithani Saravanakumar & Anjali Sreenivas, 1st (Green); Prerana Kota & Ananya Soni, 6th (Magenta)

Green Generation: Divij Aswinkumar & Prerana Kota, 1st (Magenta); Rohan Iyer & Maxwell Soh, 2nd (Green)

It's About Time: Sami Badal Khan & Evan Kim, 5th (Green)

Ornithology: Katya Rukhlinskaya & Shuhui Yang, 5th (Blue), Vidit Khandelia & Matthew Lam, 8th (Magenta)

Remote Sensing: Yuchen Li & Aashray, 2nd (Green)

Rocks & Minerals: Yuchen Li & Aashray, 1st (Green)

Trajectory: Siddhant Porwal & Abhiram Yakkali, 2nd (Magenta); Katie Chan & Shivank Dutt, 6th (Green)

WiFi Lab: Sami Badal Khan & Evan Kim, 2nd (Green); Eva Jain & Vidit Khandelia, 6th (Magenta)

Wright Stuff: Marcin Anforowicz & Abhiram Yakkali, 3rd (Magenta)

The team gratefully acknowledges the support of Ms. Duenas, Mr. Crowe, Ms. Townsend, Mr. Leslie, and Mr. Chan in providing building access, lab and equipment access, and coaching.

2022 Washington State Science Olympiad

April 2022

Coach: Andy Christensen

Tesla STEM Green finished in 1st place and qualified as the Washington State representative to the National Science Olympiad.
Tesla STEM Magenta finished in 7th place and was the highest-finishing second team in the State Tournament.

Science Olympiad is played by teams of 15 students who compete in pairs and trios across 23 total events. They are stack-ranked in each event and the team with the lowest sum of rankings in the 23 events wins. Each high school may send a maximum of two qualifying teams to the State Tournament; each state may send only one team to the National Science Olympiad.

Of the 23 events at State, Tesla STEM Green finished 1st in 10 events, Our teams won medals for finishing in the Top 5, as follows:

Astronomy: Evan Kim & Yuchen Li, 1st (Green), Harish Krishnakumar & Annanya Unnikrishnan, 4th (Magenta)

Bridge: Marcin Anforowicz & Abhiram Yakkali, 4th (Magenta)

Cell Biology: Rohan Iyer & Maxwell Soh, 1st (Green)

Chemistry Lab: Sebi Kumar & Sami Badal Khan, 1st (Green)

Code Analysis: Evan Kim & Aayush Sheth, 3rd (Green)

Codebusters: Phillip Araujo, Aayush Sheth & Olivia Wang, 3rd (Green)

Detector Building: Divij Aswinkumar & Tejas Pakalapati, 5th (Magenta),

Disease Detectives: Anjali Sreenivas & Maxwell Soh, 1st (Green)

Dynamic Planet: Yuchen Li & Aashray, 1st (Green)

Environmental Chemistry: Yuchen Li & Sebi Kumar, 1st (Green)

Experimental Design: Siddhant Porwal, Anjali Sreenivas & Aayush Sheth, 2nd (Green)

Forensics: Rithani Saravanakumar & Anjali Sreenivas, 1st (Green)

Gravity Vehicle: Farzad Hasan & Siddhant Porwal, 5th (Green)

Green Generation: Rohan Iyer & Maxwell Soh, 2nd (Green)

Ornithology: Rohan Iyer & Olivia Wang, 5th (Green)

Remote Sensing: Yuchen Li & Aashray, 1st (Green)

Rocks & Minerals: Yuchen Li & Aashray, 1st (Green)

Trajectory: Farzad Hasan & Siddhant Porwal, 3rd (Green)

WiFi Lab: Sami Badal Khan & Evan Kim, 1st (Green)

Write It CAD It: Phillip Araujo & Katie Chan, 2nd (Green)

Tesla STEM SciOly has only existed for 3 years. Our students went from not placing in Year 1, to 3rd place in Washington State in Year 2, to 1st place this year. It is a testament to the vision, dedication, time and effort of the students who lead this organization.

The team gratefully acknowledges the support of Ms. Duenas, Mr. Crowe, Ms. Townsend, Mr. Leslie, and Mr. Chan in providing building access, lab and equipment access, and coaching.

2022 National Science Olympiad

This year, Tesla STEM qualified for the National Science Olympiad for the first time by finishing 1st at the Washington State Tournament.

In the National Tournament, they were coached by Ms. Rebecca Townsend, with help from Ms. Duenas and Mr. Crowe.

A total of 60 teams competed in the National Tournament and Tesla STEM placed 24th. Ten groups that placed in the top 15 in the nation in their events:

Astronomy: Evan Kim & Yuchen Li, 12th

Chemistry Lab: Sebi Kumar & Yuchen Li, 15th

Dynamic Planet: Aashray Chegu Vijay & Yuchen Li, 1st

Forensics: Rithani Saravanakumar & Anjali Sreenivas, 10th

Remote Sensing: Aashray Chegu Vijay & Yuchen Li, 4th

Rocks and Minerals: Aashray Chegu Vijay & Yuchen Li, 1st

Trajectory: Farzad Hasan & Siddhant Porwal, 10th

Cybersecurity: Marcin Anforowicz & Aayush Sheth, 8th

Digital Structures: Marcin Anforowicz & Divij Aswinkumar 12th

Write It CAD It: Prerana Kota & Harish Krishnakumar, 10th

Tesla STEM Science Olympiad: Boyceville, Bird & Southern California Invitationals

In a few national-level competitions, our "Tesla STEM Green" team has finished:

7th (Boyceville, WI, out of 85 teams),
3rd (BirdSO, out of 159 teams), and
12th (Southern California, out of more than 250 teams)

2022 USA Earth Sciences Olympiad

In April, three Tesla STEM students took the USA Earth Sciences Olympiad (USESO) Open Exam. One Tesla STEM student, **Chegu Vijay Aashray** was among the top 40 students who took the exam. Aashray earned an invitation to the USESO training camp, which is an event where top scholars learn from distinguished alumni (including Tesla STEM's own Yuchen Li, '22) and leading researchers in earth science.

Particular thanks to Mr. Arny Leslie who helped the students study for the exam and gave them new ways of looking at earth systems interactions.

2022 Orca Bowl

Four Tesla STEM scholars, **Aashray, Atharv Dixit, Yuchen Li, and Maxwell Soh**, competed in the University of Washington's Orca Bowl.

Twelve teams competed. Tesla STEM won all 9 matches they played to capture their second straight Orca Bowl championship and qualify for the National Ocean Sciences Bowl in May.

2022 National Ocean Sciences Bowl

Over two weekends in May, Tesla STEM competed in the National Ocean Sciences Bowl. (They qualified by winning the ORCA Bowl at the University of Washington.) After surviving several elimination rounds, the team finished in 4th place nationally (one place higher than last year).

Yuchen Li
Aashray Chegu Vijay
Maxwell Soh
Atharv Dixit

Tesla STEM High School 2022 Division B Science Olympiad Invitational

Tesla STEM High School (TSHS) hosted its first B-division Science Olympiad Invitational on January 30. More than 250 students from 22 middle school teams across the country participated in this event, including LWSD teams from Redmond, and International Community School, and Evergreen middle schools, as well as teams from California, Michigan, and Alabama.

The goal of an invitational is to help students prepare for regional, state, and national competitions. Normally, an invitational is held in-person at the hosting school, and while the TSHS team made every effort to host a joint event for both in-person and online competitors, the Omicron variant of COVID-19 forced the invitational to be online-only. All middle school teams competed using the same online platform that will be used in the Science Olympiad regional and state events, which allowed TSHS to grade, rank, and announce results on the day of competition!

You can view the awards ceremony at http://www.youtube.com/watch?v+=OKSiqJx2dvs. Thanks to everyone who supported this effort, from ringleaders Yuchen Li, Anjali Sreenivas, and Katie Chan, to event supervisors and graders, including Evan Kim, Aayush Sheth, Olivia Wang, Sebi Kumar, Rithani Saravanakumar, Prerana Kota, Cheguvijay Aashray, Grace Zhao, Sahaana Sridhar, Saharsh Bhargava, Siddhant Porwal, Divij Aswinkumar, Harish Krishnakumar, Sami Badal Khan, Phillip Araujo, Shuhui Yang, and Maxwell Soh.

2022 East Puget Sound National History Day Regional Competition

March 2022

Theme: Debate and Diplomacy in History
Each of the Tesla STEM entries finished 1st or 2nd and will progress on to the State competition.

Anson Chen, Senior Historical Paper: Lines in the Sand: Failures and Consequences of the Partition of the Ottoman Empire (2nd Place)

Alec Rothkowitz, Benjamin Moskalensky, Connor O'Rourke, David Soto, Senior Group Exhibit: Forming New Frontiers in Seattle: The Debate for Fair Employment in Boeing from 1935-1949. (1st Place)

Druhin Bhowal, Senior Individual Documentary: Shattered Land: How Diplomacy, Debate, and the Round Table Conferences Shaped the Fate of India 1930-1932.(2nd Place)

William Gong, Ziang Chen, Teddy Dong, Ethan Yu, Arihant Singh Senior Group Documentary: A Welcome Beacon in the Dark Night: How Debate and Diplomacy in Czechoslovakia Sparked the Passage of the Marshall Plan. (2nd Place)

2022 USA Biology Olympiad

A total of 8 bio club students participated in the USABO Open Exam, the first of two rounds of challenging exams. Following the Semifinal, twenty Finalists are invited to a residential training program where they learn advanced biological concepts and exacting lab skills at the USABO National Finals. Ultimately, four students earn the right to represent the USA at the International Biology Olympiad (IBO), a worldwide competition involving student teams from over seventy countries.

We have three students who qualified as semifinalists for the USABO Open Exam, meaning that they scored in the top 10% of students nationally.

Maxwell Soh
Rohan Iyer
Aashray Chegu Vijay

Toshiba/NSTA ExploraVision

The following students received Honorable Mention awards for the Toshiba/NSTA ExploraVision contest. The contest has students predict the evolution of a technology (for my class, a biomedical tech) into ten years in the future. Congratulations to the following students, who placed in the top 10% nationwide!

3D Bioprinted Blood Vessels for Surgical Transplant
Suhas Kannam
Claire Boshaw
Jasmine Ku
Meagan Pearson

A Novel Diagnostic Tool for Measuring Anterior Cruciate Ligament Tension
Irene Lau
Lauren Kaplita
Neelum Jawanda

Integrating Nanoparticles into a Transcranial Electromagnetic Cap to Treat Alzheimer's Disease
Tayla Belikoff
Esha Bangur
Garan Tantasirikorn
Jessica Wang

The Fantastic Future of Fishy First Aid
Catherine Park
Jolie Chan

Microbubble Technology's Future in Alzheimer's Disease Treatment and Gene Therapy
Noelle Hall
Maile Hori
Nidhi Munikote

TSA: Washington State Competition

The following students placed at the TSA State competition.

Sci Vis: Adhya Kona, 2nd place

Digital Video Production: Kshitij Rao & Steven He, 4th place; Siddhant Porwal, Manas Kanakala & Mohit Soni, 5th place

Engineering Design: Gaya Pillai, Rishika Gautam, Shivani Sama & Annie Adhikary, 2nd place

Board Games: Anjana Punniamoorthy & Vasudha N, 2nd place

Children's Stories: Anusha Sharma, 1st place

Data Science & Analysis: Rishika Gautam, 1st place

Geospacial Tech: Druhin Bhowal, Siddhant Porwal & Mohit Soni, 1st place

Places 1-3 will compete at the National competition this summer.

Washington Music Educators Association All-State Honor Ensembles

February 2022

All State Chamber Orchestra

- Guhan Kumar, Violin

All State Concert Band

- Manami Chiba, Flute
- Varsha Mantravadi, Percussion
- Alec Rothkowitz, Clarinet
- Nakul Rajpal, Oboe

All State Jazz Band

- Alex Takagi-Stewart, Piano

All State Symphony Orchestra

- Chris Ma, Violin

All State Treble Choir

- Diya Modi, Soprano 1
- Natalie Qiu, Alto 1

All State Wind Ensemble

- Davis Yang, Clarinet
- Anson Chen, Clarinet

Technology Student Association (TSA) - National Top 10 High Schools

July 2021

10th place – Board Game Design

2022 WA State High School Chess Championship

Anthony He

Anthony is an International Master of Chess and is also the reining **WA State Chess Chanmpion**

LWSD Teen Startup Challenge

22 initial entries across the district, Tesla STEM contributed 6 entries. Of our 6 entries, 3 made it into the next round with investors.

Tesla students advancing to the semi-finals: J**osh Shergill, Leonardo Ribeiro de Brito, Alexander Potra, Mohammed Mendahawi, Alex Mineeva, Anisha Vaish**

US Innovator Challenge

Senior **Ateeq Ramlan** worked with two colleagues from Redmond High School to develop Whale Watchers, an app to help Puget Sound orca populations. The team's prototype placed in the top ten in the U.S. They will participate in the Global Innovator Challenge in Fall 2021.

Wharton School of Business Global High School Investment Competition

Regional Competition Round:

The Tesla STEM Economics Club, advised by Ms. Ryfka Schafer, competed in the University of Pennsylvania Wharton School of Business investment competition. Out of 1300 teams from across the world, one of our teams earned a place in the top 50 teams that will participate in the global semi-finals.

The team included 11th graders **William Gong, Samiullah Khan Badal Khan, Sidharth Ganesh, Harini Vijeyanandh, Pranav Teegavarapu, and Gautham Iyer.**

In the semifinals, this team will present their investment strategies and answer questions from a panel of investors and bankers. They will compete against teams from Indonesia, New Zealand, Canada, India, and China.

2022 NCWIT Awards for Aspirations in Computing

Please join us in congratulating our outstanding young women who received 2022 NCWIT Awards for Aspirations in Computing:

Anjali Sreenivas, National Honorable Mention

The NCWIT Award for Aspirations in Computing builds a talent pool for the growing technical workforce and helps academic and corporate organizations celebrate diversity in computing by honoring young women at the high-school level for their computing-related achievements and interests. Award recipients are selected based on their aptitude and aspirations in technology and computing; leadership ability; academic history; and plans for post-secondary education.

Seattle Times Features Tesla STEM Student work for Day of Remembrance

Tesla STEM 11th-grade student **Kai Vanderlip** was featured in the Sunday Seattle Times for his work, *Day of Remembrance- Japanese Internment Camps,* helping educate elementary school students in our district.

https://www.seattletimes.com/entertainment/books/this-local-teen-wants-elementary-schools-to-have-books-about-the-incarceration-of-japanese-americans/

2022 Washington State National History Day (NHD)

April 2022

The Washington State National History Day competition was held at Lake Washington High School with over 500 students in attendance. 11 TESLA Stem students competed as individuals and groups in several categories in this year's competition, after having succeeded at a regional competition held in February. Two of the four entries took first place and qualify for the National Competition in June.

Senior Group Exhibit category: **Connor O'Rourke, Ben Moskalensky, Alec Rothkowitz, & David Soto**
1st place and qualify for Nationals, also won $100 award for best use of local archives.
The Exhibit, on the subject of the first African Americans hired at Boeing is titled, *Forming New Frontiers in Seattle: The Debate for Fair Employment at Boeing from 1935 – 1949*

Senior Group Documentary: **William Gong, Teddy Dong, Ethan Yu, Arihant Singh and Ziang Chen**
1st place and qualify for Nationals.
Their Documentary is titled, *A Welcome Beacon in the Dark Night: How Debate and Diplomacy in Czechoslovakia Sparked the Passage of the Marshall Plan*

Senior Historical Paper: **Anson Chen,**
3rd place
Title of his Paper: *Lines in the Sand; Failures and Consequences of the Partition of the Ottoman Empire*

Senior Individual Documentary: **Druhin Bhowal,**
7th place
Title of his Documentary: *Shattered Land: How Diplomacy, Debate and the Round Table Conferences Shaped the Fate of India, 1930-1932*

Thank you to Karen Schaeffer and Melissa Wrenchy for design advice and help with printing the materials for the Senior Exhibit entry.

2022 ArcGIS Washington State Map Story Contest

Congratulations to **Siddhant Porwal and Druhin Bhowal**, who were the Washington State High School Winners for the ArcGIS Online Competition for 2022.

Their work will be submitted to ESRI to compete against the winners from other states. ESRI will make a decision later in May or early June.

See their work at Giants of the Puget Sound: The Southern Resident Orca Whales (arcgis.com)

AYLUS President's Volunteer Service Award

Tesla STEM student **Keshav Acharya** won a Silver Medal as part of the 2022 President's Volunteer Service Awards. As part of the Alliance of Youth Leaders in the United States (AYLUS), advised by Ms. Brooke Ranieri, Keshav participated in volunteering events around the community, including food drives, clearing invasive plants from trails, and collaborating with the City of Sammamish on their projects, eventually earning nearly a hundred total hours throughout the year. The PVSA was created to award outstanding members of the community who achieve set goals for community service.

3rd Annual STEM Expo (2021-22)

Tesla STEM hosted the 3rd Annual STEM Expo and the first one that was a joint, in-person event for the Startup Track and the Research Track. There were five judges for the startups and three judges for the research projects, and we successfully livestreamed the entire event.
After each startup and research project presented, the judges deliberated extensively and awarded prizes to the following winners:

Startup Track

Kobra (**Pranav Teegavarapu**) earned Best Startup for their visual programming language that makes machine learning easy to learn.

DBall Defensive Basketball Training (**Ryan Chan, Ben Moskalensky, and Cameron Yetzer**) earned the Entrepreneurship Award for their mobile app that helps basketball players learn defensive positions.

FriendFlix (**Triana Fernando, Tanvi Kamat, and Emma Shi**) earned Audience Favorite for their online platform and community where friends can share and recommend new TV shows to each other.

Research Track

1st Place: **Anya Vaish** - Bioinformatic Analysis to Find a Universal Sequence from SARS-CoV-2 Genomes for the Potential Development of a Pancoronavirus Vaccine

2nd Place: **Cindy Chen** - Facial and Emotional Recognition Among Caucasian and East Asian Adolescents

3rd Place: **Laya Nair** - Parkinson's Disease: A Risk Prediction Model Using Inflammatory Bowel Disease

Atlas Fellowship Award

The Atlas Fellowship is a free program for talented high school students to develop big-picture thinking about science, philosophy, and the future

Pranav Teegavarapu
has been awarded a scholarship of $50,000/- towards his college tuition expenses.

MIT-Most Influential Teacher Award 2022

Each academic year MIT ask students admitted to MIT to share the name of a teacher who has been especially influential in their development.

It is a remarkable achievement and a great honor to change a life, so we congratulate you on being named this year by:

Julia Liu--**Bethany Kankelborg**
Emma Shi--**Arny Leslie**

College Board Advanced Placement (AP) Computer Science Female Diversity Award

Tesla STEM High School

HOSA (Health Occupations Students of America) State Leadership Conference

Individual Events

Medical Spelling
Nidhi Munikote-7th place

Medical Terminology
Rithani Saravanakumar-4th place

Medical Assisting
Thanushree Karunagaran-1st place

Pharmacy Science
Aparna Iyer-7th place

Epidemiology
Nimai Belur-5th place

Leadership Events

Extemporaneous Writing
Hannah Chen-7th place

Researched Persuasive Writing and Speaking
Secondary
Priyanka Kudallur-2nd place

Research Poster
Sophia Rodriguez-1st place
Inaya Chowdhury-10th place

Teamwork Events

Biomedical Debate
Secondary
Priyanka Kudallur-2nd place
Ajeet Venkatesh-2nd place
Ryan Chung-2nd place
Megha Ramachandran-2nd place
Shravya Daggumelli-5th place
Divya Ramu-5th place
Yue Ding-5th place
Prakshi Shukla-5th place
Nimai Belur-6th place
Tejas Pakalapati-6th place
Venkata Nookal-6th place

Community Awareness
Sanjana Katti-6th place
Saahiti Mamidipaka-6th place
Priyanka Kudallur-6th place
Rajiv Arvind-9th place
Emily Huang-9th place
Alekhya Nidadavolu-9th place
Abhinav Nadupalli-9th place

Forensic Science
Secondary
Ritu Doshi-5th place
Tejas Pakalapati-5th place

Health Career Display
Tejas Pakalapati-3rd place
Talah Albadawi-3rd place
Harshita Raja-10th place
Samyuktha Krishnan-10th place

Health Education
Secondary
Deva Junu-6th place
Anjana Punniamoorthy-6th place
Farzad Hasan-6th place
Laya Nair-10th place
Gauri Raman-10th place
Bhavya Nandikanti-10th place

Medical Innovation
Farzad Hasan-1st place
Milana Veera-1st place
Sruthika Saravanakumar-1st place
Shivani Sama-2nd place
Sanjana Desai-2nd place
Prisha Agarwal-2nd place
Angelina Durbin-4th place
Shivani Godse-4th place
Nitya Kakulamarri-8th place
Moomal Siddiqui-8th place
Isha Rajpure-8th place
Sophia Rodriguez-8th place

Public Service Announcement
Ateeq Ahmedramlan-2nd place
Vijaya Sripada-2nd place
Gauri Raman-2nd place
Barath Saravanan-2nd place

Recognition Events

Healthcare Issues Exam
Keshav Iyer-9th place

Top 3 In Events --Welcome to Attend HOSA's International Leadership Conference

Medical Assisting
Thanushree Karunagaran-1st place

Researched Persuasive Writing and Speaking Secondary
Priyanka Kudallur-2nd place

Research Poster
Sophia Rodriguez-1st place

Secondary
Priyanka Kudallur-2nd place
Ajeet Venkatesh-2nd place
Ryan Chung-2nd place
Megha Ramachandran-2nd place

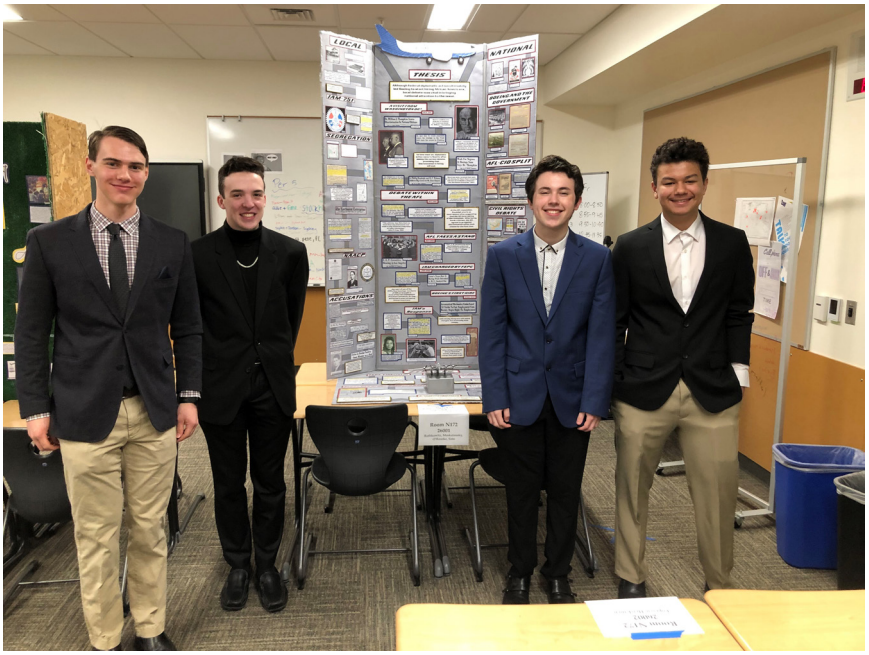
Health Career Display
Tejas Pakalapati-3rd place
Talah Albadawi-3rd place

Medical Innovation
Farzad Hasan-1st place
Milana Veera-1st place
Sruthika Saravanakumar-1st place
Shivani Sama-2nd place
Sanjana Desai-2nd place
Prisha Agarwal-2nd place

Public Service Announcement
Ateeq Ahmedramlan-2nd place
Vijaya Sripada-2nd place
Gauri Raman-2nd place
Barath Saravanan-2nd place

2022 National History Day (NHD) National Competition

First Place, **Senior Group Exhibit:** *Forming New Frontiers in Seattle: The Debate for Fair Employment in Boeing from 1935-1949*

The group was comprised of **Ben Moskalensky, and Connor O'Rourke, Alec Rothkowitz, and David Soto**

Their exhibit was also selected to be part of The National Museum of American History Exhibit Showcase, and was available to view on the Museum's website.