

NIKOLA TESLA STEM HIGH SCHOOL

REDMOND,
WASHINGTON



2024 - 2025

RESEARCH JOURNAL

Nikola Tesla STEM High School

Lake Washington School District

4301 228th Ave. NE, Redmond, Washington 98053

(425) 936-2770 * <https://tesla.lwsd.org>

Principal: Amie Karkainen

Associate Principal: Dr. Dana Bailey

Introduction

This is the tenth edition of the Nikola Tesla STEM High School Research Journal which enables our students to share their research with the academic and business communities. The investigations included in this edition demonstrate both the increasing depth and diversity of our studies. In addition, we have included a list of the awards and honors earned by Nikola Tesla STEM students during the 2024-2025 academic year.

We express our sincere gratitude to the dedicated faculty and staff for ensuring that all students are afforded every opportunity to showcase their knowledge and talents and for their support in the publication of this journal. We are especially grateful to the Research Paper Review Board, our student researchers and their mentors for their dedication to STEM Literacy.

Journal Staff

Editor in Chief: Shreeyaa Muhunthan

Managing Editor: Aarush Kapoor

Design Editor: Marta Jayantilal

Section Editors, Research Papers: Keerthi Rajesh, Shailja Tyagi, Akchara Mukunthu

Section Editors, Awards: Varnika Elenthendral

Journal Support: Amie Karkainen, Karen Schaeffer, Debi Welcome, Dr. Melissa Wrenchey

Current and past Research Journals are available online

<https://tesla.lwsd.org/academics>

Table of Contents

ARTIFICIAL INTELLIGENCE

NanoDesigner: AI-Driven Nanobody Stability Classification and Optimization Using ESM2/ESM3 Embeddings and Machine Learning	8
Novel Therapeutic Application of ChatGPT Through Prompt Engineering	18

BEHAVIORAL SOCIAL SCIENCES

Investigating the Difference Between the Effect of Exposure to Different Colored Lights on the Long-Term Memory of Autism Spectrum Disorder Teens and on Neurotypical Teens	24
Morning vs. Evening: Cognitive Training Techniques for Enhanced Memory Retention	31
Surviving In a Pressure Cooker: A Research Project on the Effects of Cultural Pressures on Self Esteem in High School Students and Resulting Professional Pathways	36
Using Behavioral Methods to Measure the Effects of Joint Attention Learning in Male and Female Infants with a Familial History of Autism	40

BIOMEDICAL SCIENCE

Cellular Biology

An Evaluation of Hypomethylation in Primary Prostate Cancer	46
Antibacterial Activity of Stinging Nettle, Devil's Club, & Oregon Grape Against Escherichia Coli & Bacillus Cereus	51
Decoding the Genetic Blueprint: Polygenic Risk Scores for Cancer in Diverse Genome Data	57
“When Copper Kills the Good Guys”: a Novel Approach to Investigating the Hidden Microbial Effects of Copper IUDs on Female Reproductive Health	62

COMPUTATIONAL BIOLOGY

A Diffusion-Based Approach to Restoring Ancient Oracle Bone Characters	70
Identification and Validation of Key Biomarkers and Pathways for Idiopathic Pulmonary Fibrosis using In Silico Methods	76
Synthesizing Novel CD40xCD137 Bispecific Antibodies for B cell Malignancies and Solid Tumors	82

ENGINEERING

Affordable Distortion Otoacoustic Emission Measurement Device for Hearing Screening	92
Classification of ADHD Through Machine Learning Models in Children Using Traditional Statistical and CNN Based and Feature Extraction of Raw EEG Data	97
Wavelet-Driven Dynamic Segmentation for Real-Time Speech Translation	103

ENVIRONMENTAL ENGINEERING

Comparing Bacterial Efficiency in Microbial Fuel Cells for Dairy Wastewater Treatment	110
CoRA: Using Machine Learning for Coastal Risk Assessment and Defense Optimization	116
Investigating the Adsorption Capabilities of Human Hair for Mitigating Carbamazepine Pollution	124
Phytoremediation of Acetaminophen (AC) from Wastewater Using Water Hyacinth (Eichhornia crassipes) and Duckweed (Lemna minor) Combined with Biochar	130

AWARDS

These awards capture the work of the 24-25 academic year for Nikola Tesla STEM High School students & staff	134
---	-----

ARTIFICIAL INTELLIGENCE

NanoDesigner: AI-Driven Nanobody Stability Classification and Optimization Using ESM2/ESM3 Embeddings and Machine Learning

Shriya Kapila

ABSTRACT

Nanobodies, single-domain antibodies derived from camelids, represent a promising frontier in therapeutic development due to their exceptional targeting capabilities and small size. However, their clinical application critically depends on protein stability, as unstable nanobodies can aggregate or lose function under physiological conditions, potentially compromising patient treatment outcomes. Despite the importance of stability prediction in therapeutic development, existing computational methods often fail to capture the complex sequence-stability relationships that determine nanobody functionality. This study hypothesized that ESM2 and ESM3 (Evolutionary Scale Modeling) embeddings, combined with machine learning classifiers, will outperform traditional feature-based models in nanobody stability prediction. A dataset of 419 experimentally validated nanobody sequences was analyzed using these embeddings, which capture both structural and functional protein information. Six classifiers, including Random Forest, XGBoost, and neural networks, were evaluated. The Random Forest classifier achieved the highest accuracy of 84.52%, while ensemble methods demonstrated superior performance in identifying unstable sequences with a recall rate of 95%. Feature importance analysis revealed that tail-end amino acid properties, more specifically C-terminal properties and cysteine frequency, were the primary determinants of stability, with fine-tuned ESM2 embeddings improving unstable sequence detection by 14%. Comprehensive SHAP (SHapley Additive exPlanations) analysis identified specific structural and physicochemical properties governing stability, providing interpretable insights for rational design. With the recent release of ESM3, which integrates structural data and increased model parameters, stability prediction performance improved further. Neural Networks trained on ESM3 embeddings outperformed Random Forest on test data, achieving 92% accuracy, demonstrating a significant improvement from feature-based methods. Additionally, a NanoDesigner web platform was developed, allowing researchers to input nanobody sequences and receive real-time stability predictions - integrating the trained models with an interactive interface. Interactive 3D visualizations with stability heatmaps were generated using ESMFold and 3Dmol.js visualization. A stability optimization pipeline was also implemented, generating improved stability sequences using a novel iterative masking approach with ESM3. The identified sequence features and computational framework establish a foundation for optimizing nanobody stability during therapeutic development, potentially accelerating the translation of these promising molecules into effective clinical treatments.

KEYWORDS: nanobodies, protein stability, machine learning, ESM2/ESM3 embeddings, therapeutic antibodies, protein engineering.

INTRODUCTION

This research focuses on using ESM2 embeddings to improve the prediction of nanobody protein stability, a critical factor for their effectiveness in therapeutic applications.

Nanobodies, also referred to as single-domain antibodies, are derived from camelids such as llamas and camels. These unique antibodies have garnered attention in therapeutic and diagnostic fields due to their small size (approximately 15 kDa), high stability, and ability to bind specific epitopes—small regions of antigens critical for immune recognition. Their structural properties allow them

to address challenges where traditional antibodies, which are approximately 150 kDa in size, often fail. For example, nanobody-based therapeutics designed to treat Alzheimer’s disease have shown the ability to cross the blood-brain barrier—a barrier crossed by less than 1% of conventional antibodies—allowing for effective targeting of amyloid-beta plaques (Nabuurs et al., 2012). Additionally, nanobody-based protein degraders can promote proteasomal degradation of α -synuclein, thereby providing a promising strategy for halting or slowing neurodegenerative disease progression (Jiang et. al., 2024).

In cancer immunotherapy, nanobodies have similarly outperformed conventional approaches. Their enhanced stability and resilience in harsh tumor microenvironments enable precise targeting of immune checkpoint proteins, such as PD-L1, which suppress the immune response. In murine models, nanobodies targeting PD-L1 exhibited tumor growth inhibition, with therapeutic activity persisting in acidic conditions that typically degrade larger antibodies (Yu et al., 2022). Furthermore, nanobody-radioisotope conjugates have been used for diagnostic imaging, achieving high efficacy in detecting metastatic sites within 48 hours of administration (Altunay et al., 2021). These examples underscore the versatility of nanobodies as therapeutic and diagnostic tools, offering solutions where traditional antibodies are limited.

Stability, however, remains a cornerstone of nanobody functionality. Accurate stability prediction is critical to optimizing their therapeutic and diagnostic potential. Traditional computational approaches for stability prediction often rely on engineered features such as hydrophobicity, charge, and sequence length. While these features provide foundational insights, they fail to account for the complex interplay between protein sequence and stability. For instance, Patrick et al. (2017) demonstrated that correlation coefficients (r) between these features and experimentally measured stability values rarely exceed 0.65, indicating significant gaps in predictive power. Furthermore, models trained on narrow datasets often fail to generalize across diverse nanobody families, with performance drops exceeding 20% when tested on unseen protein sequences (Sanavia et al., 2020). These limitations highlight the need for more sophisticated computational frameworks that can capture the intricate sequence-function relationships governing stability.

Recent advancements in machine learning (ML), such as Meta’s ESM2 and ESM3 (Evolutionary Scale Modeling), address many of these challenges. ESM2, trained on over 650 million protein sequences, generates embeddings that capture both structural and functional details of proteins in a context-aware manner. These embeddings provide a more comprehensive representation of protein properties, enabling improved stability predictions compared to traditional feature-based models. For example, ESM2-based models have achieved stability prediction accuracies exceeding 85%, compared to 65–70% achieved by conventional approaches (Rao et al., 2021). With the following release of ESM3, which incorporates atomic coordinate data, sequence and structural information, and significantly more parameters (98 billion), stability prediction has further improved. By leveraging both ESM2 and ESM3, researchers can address issues of generalization, data scarcity, and feature representation, advancing stability classification in ways previously unattainable.

This study, Nanodesigner, leverages a dataset of 419 nanobody sequences to evaluate the efficacy of ESM2 and ESM3 embeddings in classifying and optimizing nanobody stability. By comparing ESM2/ESM3-based approaches with traditional computational methods, this research demonstrates how advanced machine learning techniques can overcome existing challenges in stability prediction. Additionally, the interactive webtool aims to establish a framework for integrating ML-driven solutions into therapeutic nanobody design, contributing to the broader field of precision medicine.

METHODOLOGY

(Refer to Fig. 1.1 for a cohesive pipeline of the project procedure)

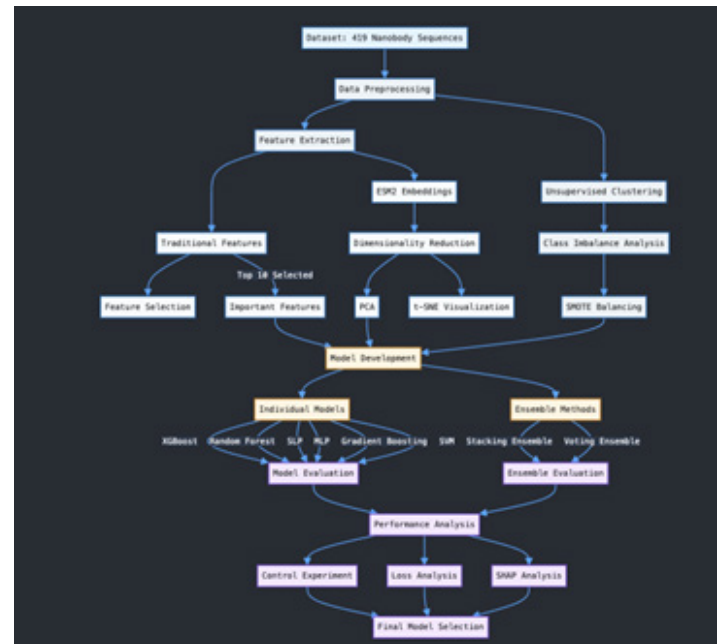


Figure 1.1: Schematic representation of the nanobody stability prediction pipeline. The process begins with a dataset of 419 nanobody sequences, followed by feature extraction using both traditional sequence-derived features and ESM2 embeddings. Feature selection identifies the top 10 most relevant predictors, and SMOTE is applied to address class imbalance. Model training and evaluation involve testing individual classifiers (XGBoost, Random Forest, MLP, SVM, etc.) as well as ensemble methods (Stacking and Voting). Performance metrics guide final model selection. The pipeline concludes with model interpretation, incorporating SHAP analysis and feature importance ranking to identify key sequence determinants influencing nanobody stability. The best performing models are then selected for ESM3 modeling and webtool integration.

2.1 Dataset and Engineered Features

The dataset consisted of 419 nanobody sequences, each

with binary stability labels as stable or unstable. Verified and provided by Dr. Tang’s lab (Seattle Children’s), this dataset ensured high-quality annotations for model training and evaluation. Baseline stability predictors were established using engineered sequence features, including hydrophobicity, volume, polarity, and isoelectric point (pI), each evaluated for average, N-terminal, and C-terminal values, as well as standard deviation across the sequence. Additional features included residue composition such as hydrophobic, polar, and charged residues, sequence length, secondary structure tendencies including alpha-helix and beta-sheet, disulfide potential determined by cysteine residue count, aromatic ratio based on the fraction of F, W, and Y residues, amino acid composition normalized by sequence length, and dipeptide diversity calculated as the unique dipeptide ratio.

To enrich feature representation, embeddings were generated using the pre-trained ESM2 model, which produces 1,280-dimensional contextual embeddings encoding sequence-level information. These embeddings replaced manually computed features, providing a more expressive feature space for stability prediction.

2.2 Dataset Preprocessing, Clustering, and Class Imbalance Correction

To better understand sequence distribution patterns, an unsupervised clustering approach was applied before classification. Clustering analysis revealed four distinct clusters within the dataset, highlighting structural or physicochemical differences that correlated with nanobody stability. Cluster 0 contained the highest proportion of stable nanobody sequences, with a stability ratio of 0.550 in the training set, meaning that 55% of nanobodies in this cluster were stable. Cluster 1, in contrast, consisted entirely of unstable nanobody sequences (0.000 stability ratio), indicating that sequences in this cluster exhibited high instability. Clusters 2 and 3 showed moderate to low stability ratios, with Cluster 3 containing predominantly unstable nanobody sequences (stability ratio of 0.015).

Cluster distributions were similar across the training, validation, and test sets, confirming that these stability-associated sequence patterns persisted throughout different splits. Fig. 1.2 visualizes cluster distributions across all dataset partitions, demonstrating that clusters varied in the proportion of stable and unstable sequences. This distribution suggests that specific physicochemical properties or structural motifs strongly correlate with nanobody stability, reinforcing the need for an accurate classification model capable of distinguishing these features.

To further address class imbalance, which skewed predictions toward the majority class, the Synthetic Minority Oversampling Technique (SMOTE) was applied. The dataset initially contained 75% unstable sequences and only 25% stable sequences, which led to model bias. SMOTE was applied to the training set, generating synthetic ex-

amples for the minority class by interpolating between existing unstable sequences in feature space rather than duplicating them. This process rebalanced the dataset to a 50:50 ratio, improving model generalization. For instance, after SMOTE application, recall for unstable sequences increased from 0.67 to 0.91 in the Random Forest classifier, ensuring that the model effectively captured stability patterns.

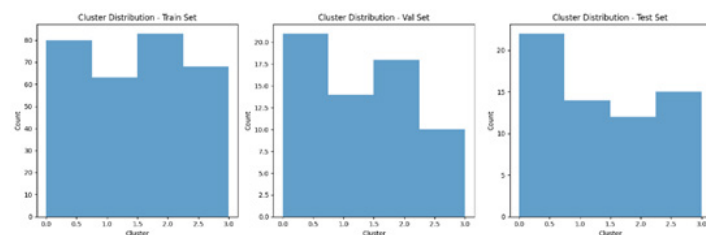


Figure 1.2: Cluster distribution of nanobody sequences across train, validation, and test sets. The dataset was divided into four clusters based on feature similarity, with stability ratios calculated for each cluster: Cluster 0 contained the highest proportion of stable sequences, while Clusters 1 and 3 were composed entirely of unstable sequences, highlighting the presence of sequence-driven patterns of stability. This analysis reinforces the existence of stability-related sequence motifs, which classifiers must leverage for accurate prediction.

2.3 Feature Extraction and Dimensionality Reduction

To efficiently process high-dimensional embeddings, dimensionality reduction was performed. The 1,280-dimensional ESM2 embeddings captured contextual sequence relationships, but direct usage could lead to overfitting. To address this, Principal Component Analysis (PCA) was applied to reduce the dimensionality while retaining 95% of variance, removing redundant information while preserving biologically relevant signals.

Following PCA, t-distributed Stochastic Neighbor Embedding (t-SNE) was used to further reduce embeddings to two dimensions, enabling visualization of sequence clustering patterns in relation to nanobody stability. The resulting plot illustrates the clustering of nanobody embeddings in the reduced feature space, with color-coded stability scores. Fig 1.3 highlights that some clusters contain predominantly stable sequences (yellow regions),

while others are dominated by unstable sequences (purple regions), confirming that stability-related patterns exist within the learned embeddings. This suggests that ESM2 representations successfully encode physicochemical properties relevant to nanobody stability.

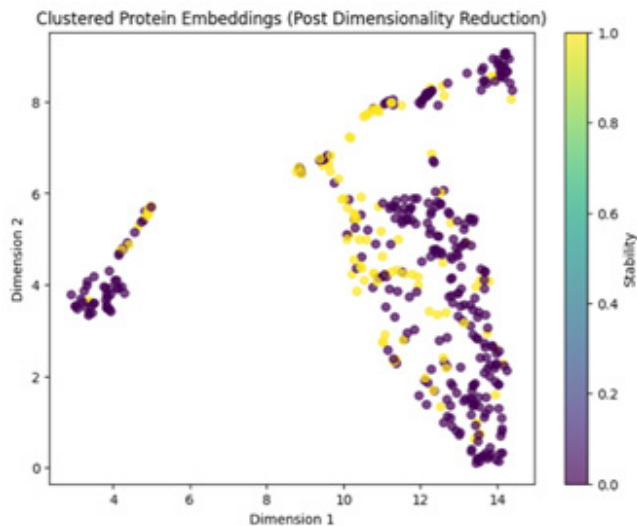


Figure 1.3: Visualization of clustered nanobody sequence embeddings after dimensionality reduction using PCA and t-SNE. Each point represents a nanobody sequence, with colors indicating stability scores from unstable (purple) to stable (yellow). The presence of well-separated clusters suggests that stability-related sequence features are embedded within the ESM2-generated representations, demonstrating the effectiveness of deep learning-derived embeddings in capturing stability determinants.

2.4 Classifier Selection, Feature Selection, and Model Evaluation

Six machine learning classifiers were employed to evaluate nanobody stability classification: XGBoost, Random Forest, Single-Layer Perceptron (SLP), Multilayer Perceptron (MLP), Gradient Boosting Tree, and Support Vector Machine (SVM). These classifiers were chosen based on their effectiveness in handling high-dimensional embeddings.

Hyperparameter tuning was conducted using GridSearch-CV, optimizing key parameters for each model. XGBoost was tuned for learning rate, number of estimators, maximum tree depth, and minimum child weight to balance model complexity. Random Forest was optimized for the number of trees ($n_{\text{estimators}}$), maximum features per split, and minimum samples per split. SVM parameters included kernel selection (linear, polynomial, or RBF) and regularization (C). Neural networks such as SLP and MLP were optimized for activation functions, hidden layers, and dropout regularization, ensuring effective feature extraction from ESM2 embeddings.

To ensure robust evaluation, a stratified k-fold cross-validation approach ($k=5$) was applied. Performance was assessed using accuracy, precision, recall, and F1-score with emphasis on recall for unstable sequences due to its importance in identifying potentially unstable nanobodies. Results were recorded into Table 1.1.

Classifier	Accuracy	Precision (Stable)	Precision (Unstable)	Recall (Stable)	Recall (Unstable)	F1-Score
XGBoost	0.84	0.92	0.71	0.85	0.83	0.82
Random Forest	0.84	1.0	0.67	0.77	1.0	0.83
Single-Layer Perceptron	0.54	0.58	0.0	0.88	0.0	0.35
Multilayer Perceptron	0.69	0.67	1.0	1.0	0.2	0.57
Gradient Boosting Tree	0.77	0.73	1.0	1.0	0.4	0.71
SVM	0.62	0.62	0.0	1.0	0.0	0.38

Table 1.1: Performance metrics of different classifiers for nanobody stability prediction. The table reports accuracy, precision, recall, and F1-score for both stable and unstable classifications across various machine learning models. Random Forest and XGBoost exhibited the highest overall accuracy (0.84), with Random Forest achieving the highest recall for unstable sequences (1.00). The ensemble methods demonstrated improved generalization, while neural networks and SVM showed varied performance depending on class distribution.

2.5 Feature Selection and Importance Analysis

A total of 45 sequence-derived features were initially extracted, but only 10 key features were retained based on Random Forest and Gradient Boosting feature importance rankings. The most influential features included C-terminal hydrophobicity, C-terminal polarity, C-terminal volume, C-terminal isoelectric point (pI), and cysteine frequency, as shown in Fig. 1.4. These features demonstrated a strong correlation with nanobody stability, emphasizing the role of charge distribution, hydrophobic interactions, and steric (atomical spatial representation) constraints in structural integrity.

To enhance model interpretability, SHAP (SHapley Additive exPlanations) analysis was applied to XGBoost and Gradient Boosting models. SHAP values quantified the contribution of each feature to individual predictions, further confirming that C-terminal sequence properties, hydrophobicity, and cysteine-mediated disulfide bonds were the strongest determinants of nanobody stability. The high importance of cysteine frequency reinforced the significance of disulfide bond formation in stabilizing nanobody structures.

To assess classifier performance and identify sources of misclassification, confusion matrices and ROC (Receiver

Operating Characteristic) curves were generated. The confusion matrices (Fig. 1.5) highlighted specific error patterns, distinguishing cases where stable nanobody sequences were misclassified as unstable and vice versa. The ROC curves (Fig. 1.6) provided an additional evaluation metric, illustrating the trade-off between sensitivity and specificity across different classification thresholds. Models with higher Area Under the Curve (AUC) scores demonstrated superior discriminatory power in distinguishing stable and unstable nanobody sequences.

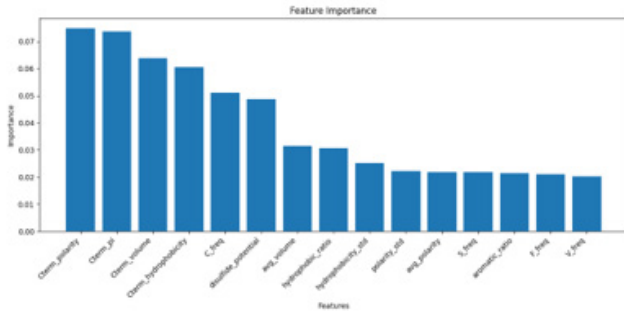


Figure 1.4: Feature importance ranking for nanobody stability classification. C-terminal polarity, isoelectric point (pI), and volume emerge as the most influential factors, followed by hydrophobicity and disulfide bond potential, emphasizing their role in structural stability.

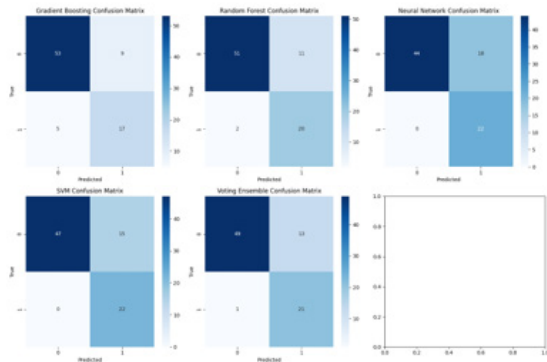


Figure 1.5: Confusion matrices for classifier predictions on the test set. The Random Forest model exhibited the highest stability classification performance, while ensemble models demonstrated improved recall for unstable nanobody sequences.

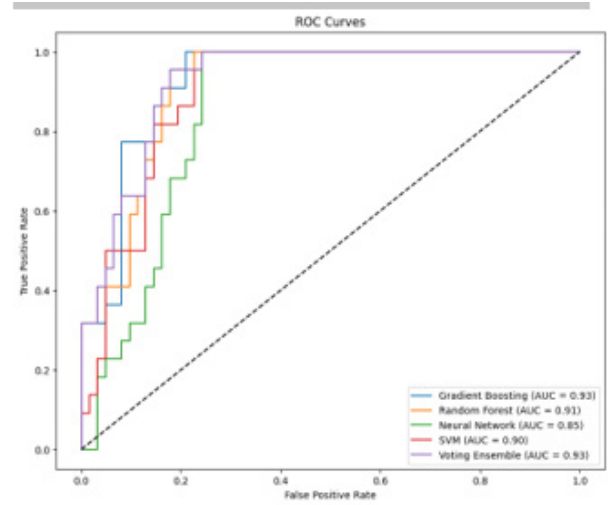


Figure 1.6: ROC curves for different classifiers, illustrating sensitivity vs. specificity. The ensemble model consistently achieved the highest AUC scores, confirming its ability to distinguish between stable and unstable nanobody sequences across different classification thresholds.

2.6 Model Training, Validation, and Control Experiment

To evaluate generalization performance and detect potential overfitting, training and validation loss were monitored for the Gradient Boosting model, as it demonstrated the most stable cross-validation performance among all classifiers. The left panel of Fig. 1.7 shows training and validation loss curves for Gradient Boosting when trained on real stability labels. The smooth convergence of both curves over epochs indicates effective learning, with minimal signs of overfitting. The validation loss stabilizes around epoch 30, suggesting that the model captures meaningful stability-related patterns without excessive complexity.

To further validate that the model was learning biologically relevant stability determinants rather than noise, a control experiment was conducted using randomly shuffled labels. The right panel displays the loss curves from this control experiment. In contrast to the real-label model, the loss values remain unstable, and no meaningful convergence is observed. This sharp discrepancy confirms that the Gradient Boosting model effectively learned meaningful sequence features associated with nanobody stability rather than memorizing spurious correlations.

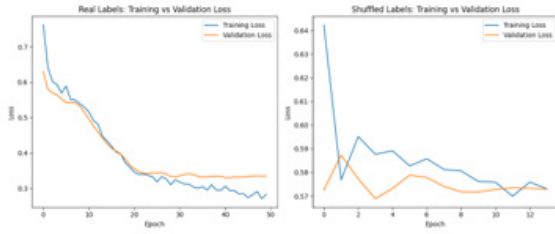


Figure 1.8: Training vs. validation loss curves for Real vs Shuffled labels in the Random Forest and Neural Network Ensemble Stack. The left panel shows loss curves when training on real stability labels, where both training and validation loss steadily decrease, indicating effective learning and generalization. The relatively stable validation loss suggests the model captures meaningful sequence-stability relationships. The right panel presents loss curves from a control experiment where stability labels were randomly shuffled, causing erratic fluctuations and increasing validation loss. The stark contrast between the two confirms that the model is not memorizing random patterns but learning biologically relevant stability features.

Model evaluation incorporated confusion matrices (Fig. 1.9 & Fig. 1.10), precision-recall curves (Fig. 1.11), and ROC-AUC scores (Fig 1.12) to comprehensively assess classifier effectiveness. XGBoost achieved the highest AUC (0.89), followed closely by Random Forest (0.88), Neural Network (0.80), and SVM (0.78), demonstrating the effectiveness of tree-based ensemble learning for stability classification. To identify the most influential embedding dimensions contributing to nanobody stability prediction, feature attribution analysis was performed on the ESM3-derived embeddings. The top 50 most significant dimensions were ranked (Fig. 1.13), revealing that stability-related information is concentrated in specific embedding regions, with a subset of dimensions exerting a disproportionately high impact on model predictions.

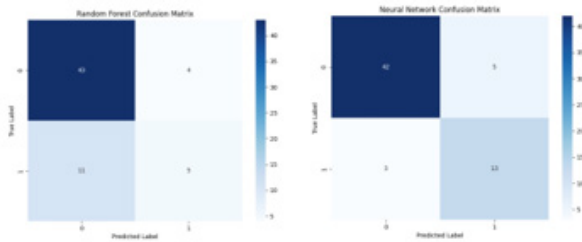


Figure 1.9: Confusion matrices for Random Forest (left) and Neural Network (right) models predicting nanobody stability. Both matrices illustrate classification performance on the test set. While both models show strong performance for stable nanobodies, the Neural Network demonstrates superior sensitivity to instability, suggesting a more reliable approach for detecting therapeutically relevant unstable variants.

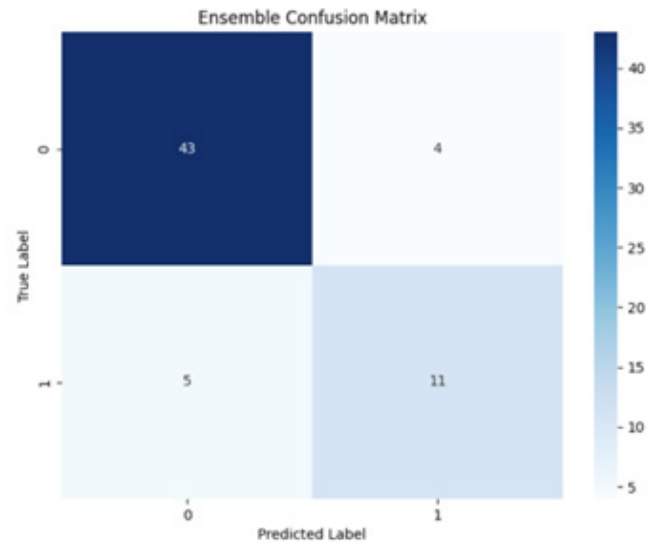


Figure 1.10: Confusion matrix for the Ensemble model predicting nanobody stability. This suggests that the ensemble approach effectively balances precision and recall, leveraging the strengths of both individual models to enhance overall stability prediction accuracy.



Figure 1.11: Training metrics over epochs for the ensemble model. The plot tracks loss (blue), precision for unstable nanobody sequences (orange), and recall for unstable nanobody sequences (green) throughout training. As training progresses, both precision and recall steadily increase, stabilizing around epoch 35–40, indicating improved identification of unstable sequences. Meanwhile, loss decreases consistently, suggesting the model is learning effectively without overfitting. The convergence of precision and recall at ~0.78–0.80 highlights the ensemble model's ability to balance sensitivity and specificity in nanobody stability prediction.

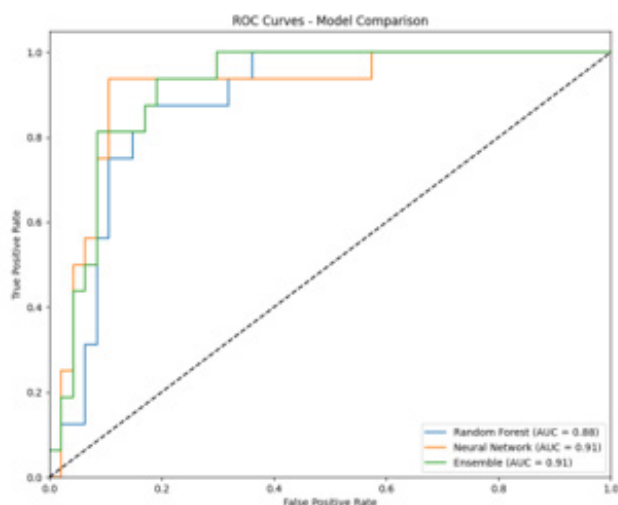


Figure 1.12: ROC curves comparing the classification performance of Random Forest, Neural Network, and Ensemble models. The plot illustrates the trade-off between true positive rate (sensitivity) and false positive rate across different classification thresholds. The Ensemble model (green) and Neural Network (orange) achieve the highest AUC (0.91), outperforming the Random Forest model (blue, AUC = 0.88). The superior AUC of the ensemble model suggests that combining multiple classifiers improves overall stability prediction accuracy, reinforcing its effectiveness in balancing sensitivity and specificity for nanobody stability classification.

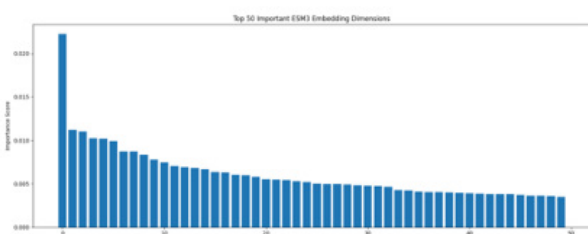


Figure 1.13: Ranked importance of ESM3 embedding dimensions for nanobody stability prediction. The top 50 most influential embedding dimensions are displayed, with importance scores derived from feature attribution analysis. The first embedding dimension exhibits a significantly higher importance score than the rest, suggesting a dominant role in stability prediction. The gradual decline in importance across subsequent dimensions indicates that stability-related information is concentrated in specific regions of the embedding space.

2.9 Web-Based Stability Prediction Platform

To translate these predictive advancements into a practical and accessible tool, a web-based NanoDesigner platform

was developed (Fig. 1.14) using Flask and PyTorch, allowing researchers to input nanobody sequences and receive real-time stability predictions. The platform supports single-sequence and batch processing, enabling users to analyze multiple nanobody variants efficiently. The backend architecture, implemented using Flask's API framework, processes sequences using the trained ensemble stacking model and returns structured predictions.

Beyond simple prediction, NanoDesigner incorporates an enhanced structure-guided optimization algorithm that employs adaptive mutation strategies based on initial stability scores. This approach dynamically adjusts the mutation rate and targeting strategy according to the sequence's starting stability, with more aggressive approaches for highly unstable sequences and conservative strategies for partially stable ones. The platform implements multiple intelligent masking strategies that target specific structural elements while preserving binding functionality, including:

1. Framework optimization - targeting conserved framework positions outside CDR regions
2. CDR-adjacent optimization - focusing on residues that support CDR loop structure
3. Hydrophobic core enhancement - improving internal packing at positions 15, 20, 22, 38, 40, 45, and other core locations
4. Charge distribution balancing - neutralizing excessive positive or negative charges
5. Beta sheet stabilization - enhancing hydrogen bonding networks in beta sheet regions

The optimization process occurs through multiple iterations, with each iteration applying different masking strategies and generating multiple sequence variants. These variants are evaluated using a combination of machine learning prediction and physics-based scoring to identify the most stable candidates.

For visualization, NanoDesigner provides interactive 3D structure models using either ESMFold-generated predictions or simplified visualizations when full structure prediction is unavailable. These models are color-coded based on stability scores through B-factor modification, with stable regions appearing blue and unstable regions red, enabling researchers to visually identify problematic areas. Users can toggle between different visualization modes (cartoon, sticks, or combined) and coloring schemes (stability-based or chain-based).

The platform also includes multi-sequence alignment analysis that provides evolutionary context by comparing user sequences and optimized variants with reference nanobodies from camelid species and disease-targeting variants. This phylogenetic analysis helps ensure that engineered sequences maintain natural structural features while improving stability, providing confidence that optimized sequences will retain proper folding and functionality.

The NanoDesigner tool integrates physicochemical property analysis, using hydrophobicity, charge, and molecular weight calculations to complement stability predictions. Feature extraction routines implemented in the backend provide additional insights by evaluating C-terminal properties, disulfide bond frequency, and charge-based interactions.

Sequences can be manually inputted or uploaded via FASTA file. Upon processing a sequence by clicking the “Predict Stability” button, the platform returns a binary stability classification (Stable/Unstable) along with confidence scores, indicating each nanobody stability threshold.

For stability optimization, users can click the “Generate Stable Sequence” button, which initiates the enhanced structure-guided optimization process. The platform provides real-time feedback through a process log that displays the optimization steps, including framework mutations, masking strategies, and stability evaluations. The resulting optimized sequences are presented with their predicted stability scores, percent improvement, and specific mutations made.

The “Clear All” button resets the interface for uploading new sequences. Finally, the “Model Performance Metrics” toggle allows users to view the specific classification report for each model used.

Figure 1.14: NanoDesigner web tool interface for nanobody stability prediction and optimization. The platform allows users to upload FASTA sequences and receive real-time stability predictions based on the trained ensemble model. Each sequence is assigned a stability score, with a classification label indicating whether it is predicted to be stable or unstable. A 3D visualization is also developed, highlighting unstable (red) and stable (blue) regions of the nanobody. The “Generate Stable Sequence” button creates stable variants for the inputted nanobody sequence, allowing researchers to optimize stability while maintaining similar features/structure.

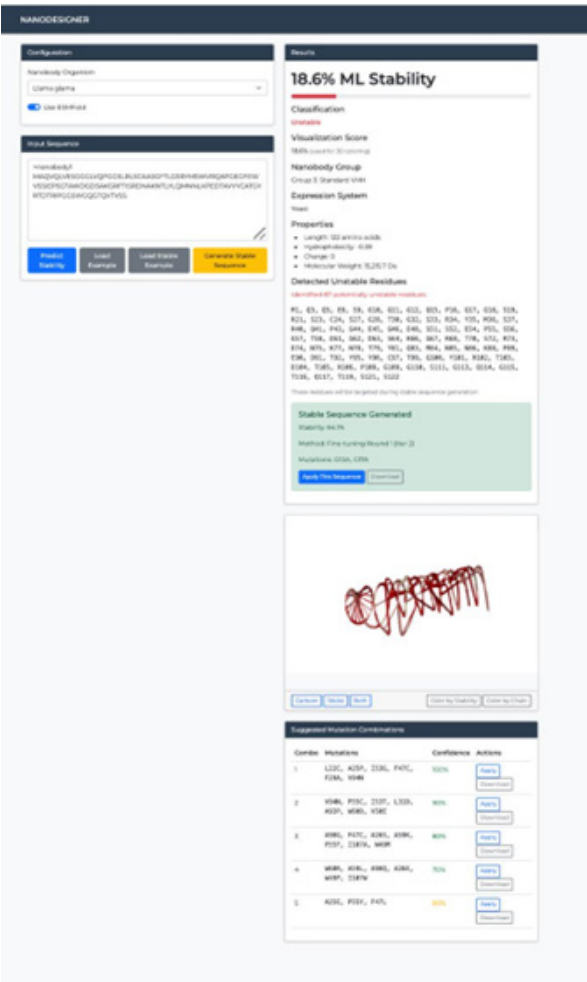
CONCLUSION

This study successfully demonstrated that ESM2 and ESM3 embeddings, combined with advanced machine learning techniques, significantly improve nanobody stability prediction. The Random Forest classifier achieved the highest accuracy of 84.52% in the ESM2 base, a 4.76% improvement from baseline models, while Gradient Boosting exhibited the most stable cross-validation performance with the lowest variance. The Stacking model improved recall for unstable sequences by 5%, while the Voting Ensemble achieved the highest recall at 95% for unstable sequences.

The integration of ESM3 embeddings, which incorporate atomic coordinate data alongside sequence-based representations, further improved model performance. Neural Networks trained on ESM3 embeddings outperformed Random Forest on test data, achieving 85.71% accuracy and 75% recall, demonstrating superior generalization in detecting unstable nanobody sequences. Feature importance analysis revealed that C-terminal hydrophobicity, polarity, and volume, along with cysteine frequency and disulfide bond formation, remained the most critical sequence determinants of stability. SHAP analysis confirmed these structural sequence properties as primary stability drivers, highlighting their role in nanobody folding and function.

Building upon these predictive capabilities, an enhanced structure-guided optimization algorithm was implemented that employs adaptive mutation strategies based on initial stability scores. This approach intelligently identifies regions for mutation while preserving binding functionality through sophisticated masking strategies targeting framework regions, areas adjacent to complementarity-determining regions (CDRs), hydrophobic core positions, and charge distribution. The optimization pipeline demonstrated significant stability improvements, with test cases showing increases from 32.3% to 71.6% through iterative application of curated framework mutations, beta sheet optimization, and hydrophobic core enhancement.

The NanoDesigner web platform integrates these advancements into a comprehensive tool that provides not only stability predictions but also interactive 3D visualization of nanobody structures. Using either ESMFold-gen-



erated structures or simplified visualizations, the platform color-codes regions based on stability scores through B-factor modification, enabling researchers to visually identify problematic areas. Multi-sequence alignment analysis further enhances the platform by providing evolutionary context, comparing optimized sequences with reference nanobodies from camelid species and disease-targeting variants to ensure that engineered sequences maintain natural structural features while improving stability.

Future work would focus on expanding the dataset to include diverse nanobody families, allowing the model to generalize across different structural motifs and sequence variations. Additionally, incorporating Graph Neural Networks (GNNs) for spatial stability modeling could further enhance predictive accuracy by integrating three-dimensional structural information into the stability classification framework.

Overall, this research presents a scalable, interpretable, and high-performance computational framework for nanobody stability classification and optimization, bridging machine learning, protein engineering, and biomedical applications. The proposed approach offers a data-driven foundation for nanobody design and optimization, with potential applications in therapeutic antibody engineering, rational protein design, and stability-guided drug discovery, particularly for crossing the blood-brain barrier to target neurodegenerative diseases.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my mentors, Gurumurthy Swaminathan (Helm.AI) for his advice and constant support and Dr. Jonathan Tang (Seattle Children's) for his guidance and resources.

BIBLIOGRAPHY

- (1) Nabuurs RJ, Rutgers KS, Welling MM, Metaxas A, de Backer ME, Rotman M, Bacskaï BJ, van Buchem MA, van der Maarel SM, van der Weerd L. In vivo detection of amyloid- β deposits using heavy chain antibody fragments in a transgenic mouse model for Alzheimer's disease. *PloS one*. 2012;7(6). <https://doi.org/10.1371/journal.pone.0038284>
- (2) Jiang Y, Lin Y, Tetlow AM, et al. Single-domain antibody-based protein degrader for synucleinopathies. *Mol Neurodegeneration*. 2024;19:44. <https://doi.org/10.1186/s13024-024-00730-y>
- (3) Yu X, Long Y, Chen B, Tong Y, Shan M, Jia X, Hu C, Liu M, Zhou J, Tang F, Lu H, Chen R, Xu P, Huang W, Ren J, Wan Y, Sun J, Li J, Jin G, Gong L. PD-L1/TLR7 dual-targeting nanobody-drug conjugate mediates potent tumor regression via elevating tumor immunogenicity in a host-expressed PD-L1 bias-dependent way. *Journal for immunotherapy of cancer*. 2022;10(10). <https://doi.org/10.1136/jitc-2022-004590>
- (4) Altunay B, Morgenroth A, Beheshti M, Vogg A, Wong NCL, Ting HH, Biersack HJ, Stickeler E, Mottaghy FM. HER2-directed antibodies, affibodies and nanobodies as drug-delivery vehicles in breast cancer with a specific focus on radioimmunotherapy and radioimmunomaging. *European journal of nuclear medicine and molecular imaging*. 2021;48(5):1371–1389. <https://doi.org/10.1007/s00259-020-05094-1>
- (5) Patrick K, et al. Exploiting sequence and stability information for directing nanobody stability engineering. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2017;1861(9):2196–2205. <https://doi.org/10.1016/j.bbagen.2017.06.017>
- (6) Sanavia T, Birolo G, Montanucci L, Turina P, Capriotti E. Limitations and challenges in protein stability prediction upon genome variations: Towards future applications in precision medicine. *Computational and Structural Biotechnology Journal*. 2020;18:1968–1979. <https://doi.org/10.1016/j.csbj.2020.07.037>
- (7) Rao R, et al. ESM2: Evolutionary Scale Modeling of Proteins. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.07.09.450648>
- (8) Tang J. nanobody_sequences_cleaned_final_Oct 11 2024. Seattle Children's. 2024.
- (9) Chawla NV, et al. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2002;16:321–357. <https://doi.org/10.1613/jair.953>
- (10) Van Audenhove I, Gettemans J. Nanobodies as Versatile Tools to Understand, Diagnose, Visualize and Treat Cancer. *EBioMedicine*. 2016;8:40–48. <https://doi.org/10.1016/j.ebiom.2016.04.028>
- (11) Kijanka M, et al. Engineering nanobodies for therapy of major diseases. *Bioengineering & Translational Medicine*. 2021;6(2). <https://doi.org/10.1002/btm2.10209>
- (12) Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- (13) Hayes T, et al. Simulating 500 million years of evolution with a language model. *bioRxiv*. 2024. <https://doi.org/10.1101/2024.01.05.574188>
- (14) Rives A, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*. 2021;118(15). <https://doi.org/10.1073/pnas.2016239118>
- (15) Senior AW, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020;577:706–710. <https://doi.org/10.1038/s41586-019-1923-7>
- (16) Biopython Documentation. 2023. <https://biopython.org/docs/>
- (17) Cursor. *Cursor.sh*. 2024. <https://cursor.sh/>
- (18) ESM-2 Model Documentation. Meta AI Research. 2022. <https://github.com/facebookresearch/esm>

- (19) *ESM-3 Model Documentation. Meta AI Research. 2024. <https://github.com/facebookresearch/esm>*
- (20) *Mermaid. Markdownish syntax for generating flow-charts, sequence diagrams, class diagrams, gantt charts and git graphs. 2024. <https://mermaid.js.org/>*
- (21) *PyTorch Documentation. 2023. <https://pytorch.org/docs/>*

Novel Therapeutic Application of ChatGPT Through Prompt Engineering

Tejas Panja

ABSTRACT

There has been an increase in the number of people who have been turning to ChatGPT for mental health issues, and in order to use it effectively, the prompts that the user inputs must be optimized to better guide large language models in performing tasks. This project aims to identify whether an engineered prompt is more effective than a non-engineered prompt for ChatGPT in order to generate a response that's effective in improving the user's mood. The hypothesis is that a prompt (independent variable) that's engineered will be more effective at improving the participants mood, defined by the Subjective Units of Distress Scale score (dependent variable). Prompts were initially created by hand and tested through ChatGPT and then refined. These prompts were given to the control group ($n=8$, non-engineered prompt) and the experimental group ($n=8$, engineered prompt) so they could report their mood before and after reading the response generated by GPT-4o-mini. The results prove the hypothesis by showing that participants that used the engineered prompt reported a larger change in mood compared to the participants that used the non-engineered prompt, giving a p-value of .005, showing that the results are statistically significant. These results also gave an r-value of .430, showing a moderate positive correlation. The experiment showed that using an engineered prompt is moderately more effective at improving the user's mood.

INTRODUCTION

Due to an increased need for more accessible forms of addressing emotional distress, this paper aims to identify the most optimal way to prompt engineer so that ChatGPT can produce the response most suited to helping with a teenager's mental health.

A rising issue in society is the lack of widely available methods for mental health care. It is estimated that 1 in 3 people live in an area with a shortage of mental health professionals, and 90% of individuals believe that there is a mental health crisis in the US (1). Many people cannot afford the expenses, while others have cultural and social stigmas against the idea of going to therapy, especially in teenagers, and for some, they aren't comfortable spilling their secrets to a real human. These factors, combined with the fact that "1 in 5 adolescents report symptoms of anxiety or depression" (2), mean that a new way to address these issues is needed.

For the solution, we can turn to artificial intelligence (AI). On November 30th, 2022, OpenAI released the first version of ChatGPT to the public, changing our lives. Since then, AI has become an increasingly large part of our lives. This AI revolution is also prevalent in the mental health fields, with reports showing that more people are using ChatGPT for therapy (3). While this is a step in the right direction, we need to explore the full potential of these large language models (LLMs) fully. Most people use these

LLMs ineffectively by asking questions with no extra context or examples, failing to give the model the extra push it needs to be an extraordinary tool. This is where the solution of prompt engineering comes in.

Prompt engineering is a new field that involves creating, refining, and deploying instructions, or prompts, to LLMs such as ChatGPT in order to find the most optimal input to generate an output that fits the request as closely as possible. Currently, ChatGPT is incapable of helping a real human with stress, anxiety, or other worries. Instead, its inability to infer meanings from the prompt can cause more emotional distress. Prompt engineering in a way that guides ChatGPT to generate a response to help with these problems is essential to making mental health care more available and accessible while avoiding the stigma that is present for teenagers.

Previous research, specifically Delcan Grabbs experiment, The Impact of Prompt Engineering in Large Language Models Performance: A Psychiatric Example (3) investigated different features in prompts and how ChatGPT responds to these different prompts. What has not been investigated before is the task of prompt engineering, specifically for the purpose of receiving a response that will help immediately improve the mood/reduce stress for teenagers. The hypothesis was that a prompt that is engineered with correct wording, sentence structure, and specific phrases such as 'you are a mental health professional' and

keywords such as ‘warm,’ ‘practical,’ and ‘safe’ will lead to ChatGPT being able to generate responses that are more likely to improve the user’s mood and help combat issues such as stress and anxiety.

METHODOLOGY

Part A Engineering Procedure

The engineering procedure outlines the creation of 4 unique prompts: 2 control prompts and 2 experimental prompts. These prompts were designed with the idea that GPT-4o-mini should respond with a statement that allowed the participant to have an increase in their mood.

The control prompts were created first. This was done by identifying what the prompt needed to do and building a two-sentence phrase that stated the minimum instructions that were needed to generate a baseline response. These prompts did not include any examples or additional context in the prompt.

The first non-engineered prompt was “I’m feeling frustrated and overwhelmed with my difficult homework right now. How can I improve my mood at this moment?”

The second non-engineered prompt was “I’m starting to feel better, but I want to make sure I stay motivated. How can I stay motivated at this moment?”

These prompts were not analyzed and improved upon and were instead kept the same as the original prompts due to the fact that they are meant to be a baseline and reflect what the average teenager would ask ChatGPT.

The experimental prompts were created next. These prompts were designed in order to generate a response that included factors such as specific studying techniques, acknowledging the feelings of the user, and keeping the response within a specific length. The design of the prompt contained elements from Delcan Grabb’s experiment (3), such as specific roles (e.g., a mental health counselor), defined lengths, and a specific format.

The first experimental prompt is “You are an expert educational psychologist specializing in student motivation and well-being. A student comes to you saying, ‘I’m feeling frustrated and overwhelmed with my difficult homework. How can I improve my mood and stay motivated?’ Please provide a response that includes the following: An acknowledgment of the student’s feelings. Three specific, evidence-based strategies to improve mood and motivation while studying. A recommendation for a study technique. A suggestion for when the student should consider seeking additional academic support. Limit your response to 200 words and use clear, accessible language suitable for a student. Ensure your advice is practical and in line with current educational psychology best practices.”

The second experimental prompt is “You are a compassionate mental health coach specializing in motivational

speaking and techniques. I tell you “I’m starting to feel better, but I want to make sure I stay motivated. How can I stay motivated at this moment?” Please provide a response that includes: A 3-4 sentence motivational paragraph to motivate the user. One specific, evidence-based technique to immediately improve motivation and focus, explained in 1-2 sentences. Focus on in-the-moment strategies rather than long-term habits. A timeframe for the participant to work and take breaks in and a short reminder that I can do it. Limit your response to 200 words and use positive, motivational language suitable for a stressed individual. Ensure your advice is practical, safe, and focused on present-moment relief.”

The experimental prompts were refined by running them in the ChatGPT temporary chat run on GPT-4o-mini. To determine whether the prompt allowed ChatGPT to generate a response that accurately reflected the needs of the user, the response was evaluated based on the following criteria:

Relevance: Does the response properly address what the prompt was asking and what the research question is? Evidence of this can be explicitly mentioning details in the prompt and referring to key concepts in the prompt.

Conciseness: Does the response keep under a total of 200 words and address all the components of the prompt? Evidence of this is no redundant phrases or information useless to the prompt.

If the response did not meet either of the requirements, the prompt was tweaked in the areas that would properly balance out the prompt in favor of the research question. This cycle of sending the prompt in a new temporary chat and adjusting the wording of the prompt was done several times, specifically in limiting response length, the number of examples/techniques the response had, and the structure of the prompt.

The testing of the prompts in GPT-4o-mini ensures that outputs relevant to the prompt were generated. It also ensures that the responses do not induce mental strain or stress that could potentially harm the participant.

Part B Sampling Procedure

16 participants were recruited by hanging flyers in high schools and online on social media. This ensures a representative sample by giving access to a multitude of students of different races, backgrounds, genders, etc. The flyer has a QR code on it that takes the person to an informed consent form. Teenagers were required to obtain parental approval to take part in this experiment. The informed consent forms were stored online in a Google Drive and were deleted at the end of the experiment.

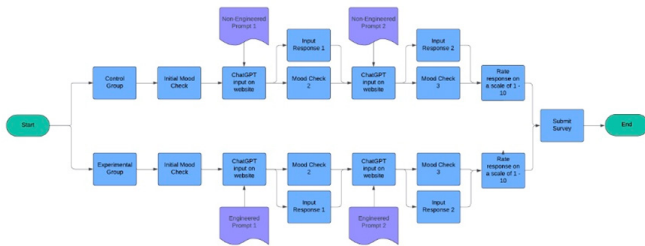
Each participant was assigned a number upon signing up. After a week, every odd-numbered participant was

assigned to the experimental group, and every even-numbered participant was assigned to the control group.

Part C Experimental Procedure

The participants were instructed to use a custom website with GPT-4o-mini integration and a custom filter that automatically prevents responses that could potentially harm the user (such as responses advising the user to perform an action that could endanger the user) from being generated to the user. The use of a custom website ensures that ChatGPT is not training off of the messages that the user is sending.

Figure 1. Experimental Procedure Flowchart



A flowchart that shows the process for a participant to go through on one day when participating in the experiment.

The participant used this website 3 times over the duration of 1 week. Once on Wednesday, once on Thursday, and once on Friday. They were instructed to use it right before they began their homework for the day. Figure 1 shows a flowchart that showcases the process for one of the three days.

Depending on whether the participant was in the control group or experimental group, they were given one of two Google Forms. Figure 2 is a screenshot of the experimental form that contains the prompts the participants will use as well as a link to the SUDS Distress Rating, which is how the participants rated their mood. A lower SUDS score indicates a better mood. These Google Forms contained the prompts that would ask ChatGPT. The form started with the user reporting their mood based on the SUDS Distress Rating. They then input the first prompt they were given. After receiving a response from ChatGPT, they pasted the response into the Google Form. They then reported their mood using the SUDS scale a second time. They then repeated the steps by inputting the second prompt and pasting the response into the form. A final mood ranking was done based on the SUDS distress scale. The user was then asked to rate the responses that were given on a scale from one to ten.

Figure 2. Experimental Group Data Collection Form

What is your current mood? (Rank it based on the SUDS chart <https://camhecho.iu.edu/resources/downloads/SUDS%20Scale%20with%20descriptions.pdf>)

Short answer text

Input this prompt (prompt 1) into the chatbot

You are a compassionate mental health counselor specializing in immediate mood enhancement. I tell you, I'm feeling frustrated and overwhelmed with my difficult homework right now. How can I improve my mood at this moment? Please provide a response that includes: A 3-4 sentence empathetic acknowledgment of the current feelings. One specific, evidence-based technique to immediately improve mood while studying, each explained in 1-2 sentences. Focus on in-the-moment strategies rather than long-term habits. One quick mindfulness or relaxation exercise that can be done at the study desk, with a brief explanation of its benefits. A gentle reminder that it's okay to take short breaks to recharge. Limit your response to 200 words and use warm, encouraging language suitable for a stressed individual. Ensure your advice is practical, safe, and focused on present-moment relief.

What is your current mood? (Rank it based on the SUDS chart <https://camhecho.iu.edu/resources/downloads/SUDS%20Scale%20with%20descriptions.pdf>)

Short answer text

Screenshot of part of the Google Form that people in the experimental group will use to input their responses and input the prompts into ChatGPT.

The collection of the participants mood 3 times allows for the analysis of whether or not the non-engineered/engineered prompt helped improve the participants mood. The collection of a rating between 1 and 10 of the responses allows for determining whether the response generated was helpful for the participant. These measures (net change in mood, mood rating) were analyzed, and the mean, median, and standard deviation were found. The t-value, p-value, and r-value were found by looking at the net change in the participants mood. The collected data can be found in Table 1 and Table 2.

Table 1. Control Table: (SDR = "SUDS" Distress Ratings)

Initial Mood (SDR)	Secondary Mood (SDR)	Final Mood (SDR)	Net Change in Mood	Response Rank
40	20	10	30	2
40	40	40	0	2
75	52	50	25	4
40	30	20	20	2
80	60	40	40	5
45	40	35	10	2
40	40	40	0	1
69	58	50	19	3
97	95	95	2	2
80	70	50	20	5
75	70	70	5	2
60	70	30	30	5
67	62	59	8	4
40	40	40	0	5
2	40	16	-14	5
70	40	40	30	9
90	70	60	30	8
55	53	50	5	6

Data table with the Initial Mood, Secondary Mood, Final Mood, Net Change in Mood, and the Participants ranking of the response generated.

Table 2. Experimental Table: (SDR = "SUDS" Distress Ratings)

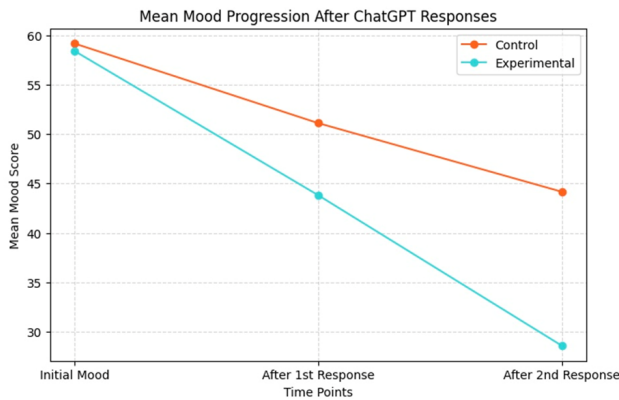
Initial Mood (SDR)	Secondary Mood (SDR)	Final Mood (SDR)	Net Change in Mood	Response Rank
96	52	25	71	9
52	42	42	10	5
70	30	10	60	9
50	30	20	30	8
60	50	20	40	7
80	56	36	44	7
85	75	70	15	6
30	21	15	15	5
80	70	60	20	7
44	32	12	23	7
30	20	20	10	10
60	60	50	10	7
55	30	30	25	4
86	54	36	50	9
87	65	42	45	9
37	30	25	12	1
60	50	20	40	7
40	30	20	20	9
15	12	10	5	2
67	54	23	44	6
66	43	31	35	7

Data table for collecting the mood of the participant in the control group as well as the responses generated by ChatGPT for the control prompts and the participants rating of the responses.

RESULTS

The results from the experiment showed that the engineered prompts were more effective at improving the user's mood than a non-engineered prompt. This supports the hypothesis that a prompt that was engineered containing specific phrases allowed ChatGPT to generate responses that were better at improving the user's mood. As shown in Figure 3, the participants in the experimental group consistently showed a greater improvement in their mood.

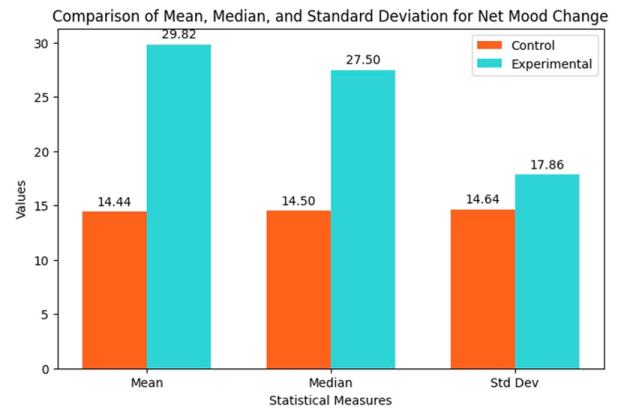
Figure 3. Mean Mood Progression after ChatGPT Responses



A line graph of the progression of the mean mood at 3 checkpoints in the experiment. This graph shows that the experimental group on average had a higher decrease in their distress rating after reading the responses generated using the engineered prompt.

The mood improvements were significantly higher in the experimental group compared to the control group. This is shown further by looking at the difference in mean, median, and standard deviation of the control and experimental groups in Figure 4.

Figure 4. Comparison of Mean, Median, and Standard Deviation for the Net Change in Mood



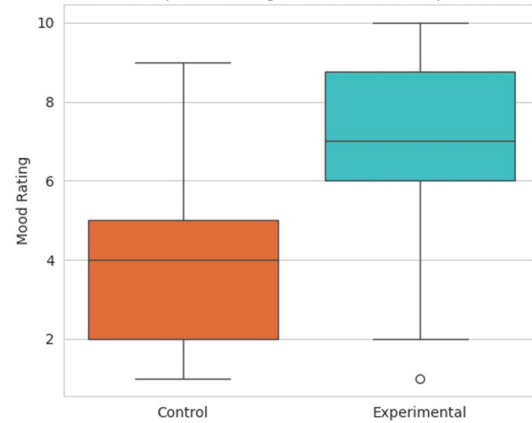
A bar graph of the mean, median, and standard deviation between the control and experimental group. The graph shows that the mean, median, and standard deviation of the net change in the mood in the experimental group are all higher than the control group.

The mean of 29.82 in the experimental group is 15.38 points higher than the mean of the control group, and the median of 27.50 in the experimental group is 13 points higher than the median of 14.5 in the control group. These differences indicate that using an engineered prompt does a better job at making the user feel better than the non-engineered prompts.

The participants also rated the response generated using the engineered prompt higher than they rated the non-engineered prompt. This is shown by looking at Figure 5.

Figure 5. Distribution of Response Ratings for Control vs. Experimental Groups

Distribution of Response Ratings for Control vs. Experimental Groups



A box plot of the ratings given to the responses generated by the control group versus the experimental group. The ratings of the experimental group were on average higher than the ratings of the control group.

There is overlap in the whiskers and the other group because the rating was ranked by a human participant, meaning there could be people who found that the control group response was quite good and people who found that the experimental group response was quite bad, leading to outliers.

This data is statistically significant with a p-value of .005, showing that the hypothesis that an engineered prompt is more effective at improving a user's mood than a non-engineered prompt is true, and rejecting the null

hypothesis. This p-value means that there is only a 0.5% probability that the difference between the groups happened by chance. A t-value of -2.992 shows that the experimental group and the control group have a statistically significant difference between the two. An r-value of .430 was also found, indicating a moderate positive correlation between using an engineered prompt and a greater improvement in the participants mood.

These findings show that the language used in the prompt given to ChatGPT does matter, and these prompts can be adjusted in order to more effectively give advice to people turning to AI for mental health advice.

CONCLUSION

As more and more people turn towards AI for mental health advice and the end of what pre-trained data can do is approaching, turning to prompt engineering is a prominent solution. While therapy is still essential, AI can be a way to bridge the time in between therapy. An AI chatbot is a way to give help to those who may need it when they cannot access a therapist. While this is not a replacement for a therapist, it can be used to help people who may need immediate help.

One problem is that people in distress will not be able to type out a lengthy prompt. A solution to this problem is integrating these prompts into the backend of AI chatbots in order to have a response that an engineered prompt would generate with just a simple sentence. A website was built in order to transform a simple sentence into an engineered prompt at research-chatbot.vercel.app/prompt. This takes a sentence that a user may input into a chatbot and transforms it into an engineered prompt as a way to display this function in action.

Prompt engineering also makes using AI more accessible and customizable for everyone. Prompts can be easily modified in order to meet different specifications, and websites such as ChatGPT are available to everyone with a simple search query.

ACKNOWLEDGMENTS

I am grateful for my mentors, Professor Bulusu from George Washington University and Nick Manzek, for helping with reviewing my research plan and project. I also wish to thank my teacher, Ms. Allender, for inspiring me to do the project.

BIBLIOGRAPHY

Upton-clark, E. (2024, September 27). *People are using CHATGPT as a therapist. mental health ...* Fast Company. <https://www.fastcompany.com/91198644/people-are-using-chatGPT-as-a-therapist-mental-health-experts-have-some-concerns>

Panchal, N. (2024). *Recent trends in Mental Health and substance use concerns among adolescents*. Retrieved from <https://www.kff.org/mental-health/issue-brief/recent-trends-in-mental-health-and-substance-use-concerns-among-adolescents/>

Grabb, D. (2023). *The impact of prompt engineering in large language model performance: A psychiatric example*. *Journal of Medical Artificial Intelligence*, 6. <https://doi.org/10.21037/jmai-23-71>

BEHAVIORAL SOCIAL SCIENCES

Investigating the Difference Between the Effect of Exposure to Different Colored Lights on the Long-Term Memory of Autism Spectrum Disorder Teens and on Neurotypical Teens

Nethra Rajesh

ABSTRACT

Teens with autism struggle with long-term memory capacity due to their disturbed sleep cycle and physiological impairments, and there are no solutions to address this. This study investigates a potential solution through the application of cognitive light therapy. The hypothesis that yellow light will improve the long-term memory capacity of teens with autism while white light will have the same effect on neurotypical participants because research shows that white light is stimulating, which promotes long-term memory, and will likely not have much of an impact on the differing SCN. Participants (n=8) were placed into the control group (teens without autism). Within this group, participants were randomly assigned a white or yellow light exposure (independent variable) while taking a memory test, which is called the Test Your Memory test, that measures their long-term memory score (dependent variable) over two consecutive days. Participants were also given the Pittsburgh Sleep Quality Index survey (dependent variable). The results obtained from the study were that neurotypical participants who were exposed to yellow light improved in their memory test scores by 1.13% more than neurotypical participants exposed to white light. The chi-square values of both groups were 1 and 10.33, respectively, with a p-value of 0.129. While this study did not meet all of its objectives due to time constraints, the next step is to include participants with autism. This project aims to provide a better understanding of the neuroscience behind the autistic brain and how cognitive light therapy can alleviate some of the impairments.

INTRODUCTION

By exposing teens with and without autism spectrum disorder (ASD) to either yellow or white-colored lights and administering a memory test during that time period, a potential solution to improve the long-term memory (LTM) capacity for teens with ASD can be devised because this process will determine which colored light will have the strongest positive increase on LTM for teens with and without ASD. What many people are unaware of is that some teens with ASD struggle with LTM. In the article “Memory Functioning in Children and Adolescents With Autism”, the authors write that “Kanner (1943) was the first to clinically describe fascinating memory profiles in individual children with autism, puzzling over why phenomenal ability in a specific type of memory was often combined with impairment in other aspects of memory in the same affected individual, and why impairments in memory and learning were so pervasive in autism” (1). There is also data to prove that people with ASD struggle with memory. According to the same article, the mean Composite Memory Index (CMI) based on the results of the Test of Memory and Learning (TOMAL) for participants with ASD is 87.00 with a standard deviation of 13.23, whereas the mean CMI for participants without ASD is 106.94 with a standard deviation of 8.57 (1). More specifically, within those categories the

mean delayed recall index, or the ability to remember previously learned information, for children and adolescents with ASD is 88.60 with a standard deviation of 11.49, and the mean delayed recall index for children and adolescents without ASD is 102.57 with a standard deviation of 7.79 (1). These statistics are statistically significant (1), thereby proving that children and adolescents with ASD are not as able to remember previously learned information compared to those without ASD.

The reason why people with ASD struggle with memory is explained in the research article called “Neuroanatomical alterations in high-functioning adults with autism spectrum disorder”. In this study, the researchers used voxel-based morphometry, which is used to identify differences in brain anatomy between different groups of people, to analyze the physiology of the brain in participants with and without ASD. What they found was that there were “decreased gray-matter volume in posterior brain regions, including the posterior hippocampus and cuneus” (2). According to the article “Reduced hippocampal volumes and memory loss in patients with early- and late-onset depression”, “reduced hippocampal volumes were associated with deficits in visual and verbal memory performance” (3). This means that the inadequate gray matter in the posterior hippocampus leads to less encoding of short-term memories into long-

term memories, which significantly reduces the LTM for people with ASD.

Scientific research has also proven that the hypothalamus and the hippocampus connect and interact with each other. Additionally, it has been proven through research that certain colors impact memory either positively or negatively by impacting the suprachiasmatic nucleus's (SCN's) control of the circadian rhythm (or the internal body clock). What this means is that blue and white light disturb sleep patterns by causing less melatonin secretion by the pineal gland at night, whereas red and yellow light gray-matter volume in posterior brain regions, including the posterior hippocampus and cuneus" (2). According to the article "Reduced hippocampal volumes and memory loss in patients with early- and late-onset depression", "reduced hippocampal volumes were associated with deficits in visual and verbal memory performance" (3). This means that the inadequate gray matter in the posterior hippocampus leads to less encoding of short-term memories into long-term memories, which significantly reduces the LTM for people with ASD.

Scientific research has also proven that the hypothalamus and the hippocampus connect and interact with each other. Additionally, it has been proven through research that certain colors impact memory either positively or negatively by impacting the suprachiasmatic nucleus's (SCN's) control of the circadian rhythm (or the internal body clock). What this means is that blue and white light disturb sleep patterns by causing less melatonin secretion by the pineal gland at night, whereas red and yellow light help people go to sleep because they increase the amount of melatonin secretion by the pineal gland. However, for people with ASD, the SCN's control over the circadian rhythm is also likely to be different because of environmental abnormal regulation of neurodevelopment and coding gene defects. Specifically, according to the research article "Overexpression of MECP2 in the Suprachiasmatic Nucleus Alters Circadian Rhythm and Induces Abnormal Social Behaviors", the "overexpression of MECP2 in the SCN leads to a decline of mir-142-3p, then de-repression of the expression of Bmal1, which subsequently causes deficits in the circadian rhythm of MECP2-TG mice" (4). While allistic people also have this gene, it is expressed at the correct amount, leading to both normal neural development and the SCN regulating the circadian rhythm properly. The overexpression of the MECP2 gene leads to the SCN losing control over the circadian rhythm. Therefore, this connection and impact of light on the hypothalamus and the SCN will be different.

All of this also means that the connection between the hippocampus and the hypothalamus along with the SCN will likely be very different for people with ASD. This is due to the physiological differences in the hippocampus

and the SCN. In the research article "The Relationship between Autism Spectrum Disorder and Melatonin during Fetal Development", the authors write that "ASD is associated with low melatonin levels and sleep disorders, and individuals with ASD often have sleep problems" (5). The overexpression of the MECP2 gene in the SCN, which causes the SCN to not have as much control over the circadian rhythm as it usually should have, leads to decreased melatonin secretion. This decreased melatonin secretion leads to a lack of adequate REM sleep, and REM sleep is key to increasing an individual's LTM capacity. LTM capacity is determined by the hippocampus, since it encodes short-term memories into long-term memories. Therefore, the different colored lights (yellow or white-colored lights) will affect circadian rhythm differently for people with ASD compared to people without ASD. These physiological differences in the SCN and the hippocampus and the way that different colored lights impact the SCN and the circadian rhythm makes it such that the hypothalamus along with the SCN and the hippocampus interact in a much more different way in people with ASD as opposed to people without ASD, meaning that the impact of exposure to different colored lights will affect the LTM and the circadian rhythm and their relationship a lot more differently for teens with ASD than it will for teens without ASD.

However, research is needed to find a potential solution to this problem of impaired LTM in children and young adults with ASD. What this project seeks to investigate is how to use that connection between the hypothalamus, SCN, and the hippocampus to improve the LTM of high schoolers (14 to 17-year-olds) with ASD through the use of different colored lights in a method called cognitive light therapy. Cognitive light therapy is a non-invasive treatment that uses light to improve cognitive function, and it has been used in studies related to circadian rhythm, dementia, brain fog, and fatigue. This project also aims to investigate which colored light improves the quality of sleep the most for teens with ASD and, by extension, improves their LTM capacity. The aim is to use this information to create a potential solution involving cognitive light therapy with yellow or white-colored lights to improve the LTM capacity of teens with ASD. The hypothesis for this study is that exposure to white-colored light will have the strongest increase in LTM capacity for teens with ASD, whereas yellow-colored light will have the strongest increase in LTM capacity for people without ASD. This is because white-colored light typically tends to disturb the circadian rhythm and quantity of REM sleep for people without ASD and yellow-colored light tends to improve the circadian rhythm and the quantity of REM sleep for people without ASD, but since people with ASD have physiological differences in their SCN, the impact of white and yellow-colored lights on their circadian rhythm and, by extension, LTM, will likely be the exact opposite for them.

METHODOLOGY

Sampling Procedure

First, flyers will be posted around all the high schools in Lake Washington School District and at the UW Autism Center to gather some responses. The flyer looks like this:

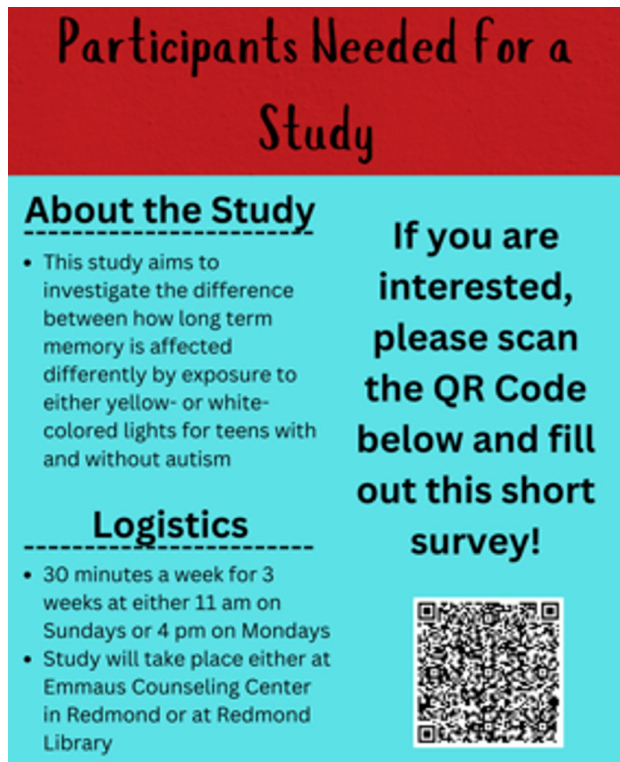


Figure 1. Screenshot of the flyer to be used in recruiting participants for the study.

The QR Code on this flyer links to an interest form that has the following questions:

- How old are you?
- What gender do you identify by? There will be four options provided – male, female, other, and prefer not to say.
- On a scale of 0 – 3 for ASD, what would you rank your level of ASD as (0 representing not having ASD at all, and 3 representing the highest level of ASD, meaning maximum support and help is required)?
- Would you be willing to dedicate 30 minutes a week for 3 weeks at Emmaus Counseling Center at Redmond or at the Redmond Library?

At the top of the survey, the following message will also be included: “The information that you provide in this form will not be shared with anyone except the people conducting the experiment, so there will be no revealing of any participant information whatsoever. Please ensure you fill this out with a parent or guardian nearby to ensure that you

enter correct information. Thank you!” Then, once there are at least 50 responses, the potential participants of the study will be emailed and will be fully informed regarding the study and its purposes. The parents will also be sent an informed consent form for them to sign and approve their child’s participation. Once the parents have responded back with a completed informed consent form, a list will be created by the researcher, and this list will contain all the potential participants’ names and all the information that they have provided in the forms. An important thing to note is that the information collected regarding the participants will be kept in a locked folder that only the researcher and the mentor will have access to through a secure password. After this experiment is over, the file will be permanently deleted.

The list will be handed over to a mentor, who will determine if the potential participants meet the requirements modeled by the flowchart below:

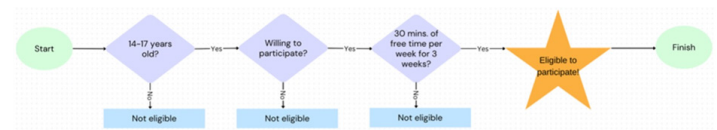


Figure 2. Diagram of criteria for potential participants after initial interest survey responses are collected.

The mentor will then remove the participants who do not meet the criteria and will separate the remaining participants by gender. In this process, the mentor will ensure that each gender category has an equal number of participants. Then, what will be done by the researcher is that they will cover up the names and the other information in the list of potential participants and label each potential participant with a number starting from 1. Each gender category’s numbering will start from 1. The list will once again be given to the mentor, who will use a random number generator to determine which potential participants will be selected to participate. The odd-numbered participants of each category will be chosen, and the mentor will make a final list of all the participants included in the study. Once this has been done, the participants will be emailed by the researcher to let them know that they have been chosen to participate in the study. The potential participants that have not been chosen will also be emailed and thanked for considering participating in the study. If, for some reason, a participant and their parents decide not to participate in the study, the random number generator will be used again by the mentor to select another odd-numbered participant. However, once 20 participants have confirmed to participating in the study, the study will begin.

Materials

Before starting the experimental procedure, the materials needed are:

1. Sixty copies of standardized Test Your Memory Test. The link to that test is here: https://www.bmj.com/content/suppl/2009/06/04/bmj.b2030.DC1/broj611491.ww1_default.pdf
2. Sixty copies of the Pittsburgh Sleep Quality Index. The link to the PSQI is here: <https://www.goodmedicine.org.uk/files/assessment,%20pittsburgh%20psqi.pdf>
3. A phone for each participant to complete the Pittsburgh Sleep Quality Index and email their responses to the mentor the very next day
4. A MacBook Air laptop (for the researcher to log the data and pictures, and to complete statistical analysis) – Apple M2 chip, 8 GB memory, 2022 model
5. XLSTAT – an app for calculating all the statistics and graphing the data obtained

Operational Definitions

The operational definitions for this study are:

1. Independent Variable: This is the exposure to warm- or cool-colored lights. For example, numbers one and seven of each gender category will obtain the yellow-colored light, and numbers 3 and 9 of each category will obtain the white-colored light. Participants with an assigned number of 5 will be equally split between both categories of colored lights. There will be 5 participants with ASD and 5 participants without ASD receiving each colored light, which means that 10 participants will be assigned to the yellow-colored and 10 will be assigned to the white-colored light. The voltage of the light will be minimal, averaging approximately 60W. Participants having or not having ASD is also part of this variable.
2. Dependent Variable: The change in long-term memory can be defined as an improvement or regression of at least 3 points each time the Test Your Memory long-term memory test (Alban et. al, 2024) is taken by the participant. Sleep quality, in accordance with the Pittsburgh Sleep Quality Index, is also another part of this variable.

Experimental Procedure

Once the materials have been obtained, on Sunday of the first week, the participants who have been assigned to white light exposure will be invited to come to the Emmaus Counseling Center in Redmond sometime between eleven am to two pm for about thirty minutes. If for some reason a participant suddenly cannot come, then they will be contacted by the researcher to see if their meeting can be rescheduled to the nearest available time in the future, preferably within two days. If not, they will be excluded from the study, and someone else from the potential participants list will be randomly chosen using a random number generator and they will be asked to replace the former participant. Then, the participants will take the standardized memory test for the first time. This will take roughly

7 minutes per participant, but they can take as long as they need to finish the memory test. Then, each participant will be given one copy of the Pittsburgh Sleep Quality Index (PSQI), and they will be walked through on how to fill out the form. The participants will then be tasked to fill out this form the very next day itself and email the completed form to the mentor on the same day within twenty-four hours of completing the standardized test. They will be encouraged to fill out the survey with utmost honesty and sincerity. The participants' scores on the memory test will be recorded in the Week One Standardized Test Score row in either Table 1.3 or in Table 1.4, depending on whether each participant has ASD or not.

On Monday of the same week, the participants who have been assigned to yellow light exposure will be invited to come to Redmond Library after school at a certain time between 4 pm and 7 pm for about 30 minutes. If for some reason a participant suddenly cannot come, then the protocol highlighted in the paragraph above will be followed. At the library, the participants will also take the standardized memory test for the first time. After each participant has finished, they will be given one copy of the PSQI, and they will be walked through on how to fill out the form. The participants will then be tasked to fill out this form the very next day itself and email the completed form to the mentor on the same day within twenty-four hours of completing the standardized test. The participants' scores on the memory test will be recorded in the Week One Standardized Test Score row in either Table 3, and it will be indicated whether the participant has ASD or not near the participant number. On Tuesday of the same week, the sleep quality index forms will be analyzed, and the scores will be recorded in the Week One Sleep Quality row in either Table 3, and it will be noted on the participant number header whether the participant has ASD or not. This process will be repeated for Weeks Two and Three.

The information regarding the standardized test scores and sleep quality will be recorded in the appropriate rows in accordance with the week number. Once all three weeks of the study have been conducted, data analysis will be conducted. The process for data analysis looks like this:

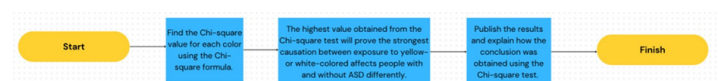


Figure 3. Diagram of data analysis for experiment, focusing in on Chi-square.

Figure 3. Diagram of data analysis for experiment, focusing in on Chi-square.

The average values for both the standardized memory test scores and sleep quality for teens without ASD exposed to white light or yellow light and teens with ASD exposed to white or yellow light will be recorded in Tables 1 and 2, respectively. The data that will be collected throughout the entire research study is as follows:

Data Collection Table for Holistic Sleep Quality and Memory Test Performance

Light Color	Average Sleep Quality Over 3 Weeks (Points)	Avg. Performance on Standardized Memory Test ($\pm$ 2SE ₉₅) (Points)
White		
Yellow		

Table 1. Averages for Hours of Sleep and Standardized Memory Test Performance for Both Colors and Participants With and Without ASD

Data Collection Table for Participant Sleep Quality and Memory Test Performance Throughout the Entire

Time Period
Colored Light Exposed To (White or Yellow – Needs to be Filled Out For Both Groups)
Participant #
Week One Standardized Test Score (Points)
Week Two Standardized Test Score (Points)
Week Three Standardized Test Score (Points)

Table 2. Score on Each Memory Test and Score for Each Week's Sleep Quality per Participant

RESULTS

After conducting testing with participants ($n = 8$) in the control group with half ($n = 4$) being exposed to white light and the other half ($n = 4$) being exposed to yellow light, some results about the memory scores and PSQI scores were obtained from almost all participants, with one participant not providing their response for PSQI for the second day. The results are summarized in the tables below:

Holistic White Light Group Memory Test Performance

White Light	Day 1 Score	Day 2 Score	Score Improvement	Expected Score Improvement	(O-E) ² /E
Participant 2	44/50	44/50	0	3	3
Participant 3	39/50	39/50	0	3	3
Participant 6	48/50	49/50	1	3	4/3
Participant 7	46/50	46/50	0	3	3

Table 3. CWL Group's Participant Scores for the Memory Test on Both Days and Chi-Square Analysis.

Holistic Yellow Light Group Memory Test Performance

Yellow Light	Day 1 Score	Day 2 Score	Improved Score	Expected Score Improvement	(O-E) ² /E
Participant 1	44/50	45/50	1	1	0
Participant 4	45/50	46/50	1	1	0
Participant 5	42/50	43/50	1	1	0
Participant 8	47/50	47/50	0	1	1

Table 6. CYL Group's Participant Scores for the Pittsburgh Sleep Quality Index on Both Days

The mean, median, and mode of the control white-light group on Day 1 for the Test Your Memory Scores are 44.25, 45, and not applicable, respectively, whereas these values were 44.5, 45, and not applicable, respectively, for Day 2. The standard deviation for the control white light group on Days One and Two are 6.69 and 7.30, respectively, the standard error of mean for both days are 3.35 and 3.65, respectively, and the Chi-square and p-values are 10 and 1/3 and 0.015, respectively. The mean, median, and mode of the control yellow-light group on Day One for the Test Your Memory Scores are 44.5, 44.5, and not applicable, respectively, whereas these values were 45.25, 45.5, and not applicable, respectively, for Day 2. The standard deviation for the control white light group on Days One and 2 are 3.61 and 2.96, respectively, the standard error of mean for both days are 1.81 and 2.12, respectively. The Chi-square and p-values are 1 and 0.75, respectively. Furthermore, the mean, median, and mode of the control white-light group on Day One for the PSQI are 15, 15, and 15, respectively, whereas these values were 16, 17.5, and 18, respectively, for Day 2. The standard deviation for the control white light group on Days One and Two are 4.24 and 5.83, respectively, and the standard error of mean for both days are 2.12 and 2.92, respectively. Additionally, mean, median, and mode of the control yellow-light group on Day One for the PSQI are 15, 15, and 15, respectively, whereas these values were 16, 17.5, and 18, respectively, for Day 2. The standard deviation for the control white light group on Days One and Two are 4.24 and 5.83, respectively. Finally, the standard error of mean for the control yellow-light group on both days are 3.57 and 2.45, respectively.

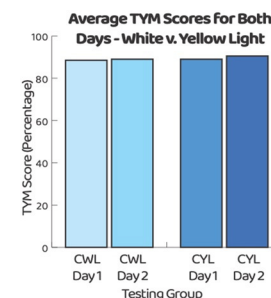


Figure 4. Bar graph of mean values for the TYM scores of both control groups on both days.

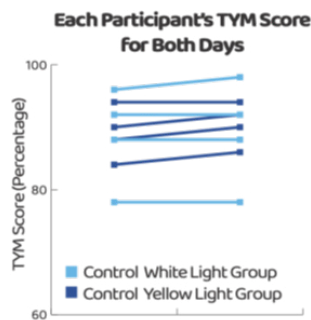


Figure 5. Line graph of each participant's TYM scores on both days of the experiment.

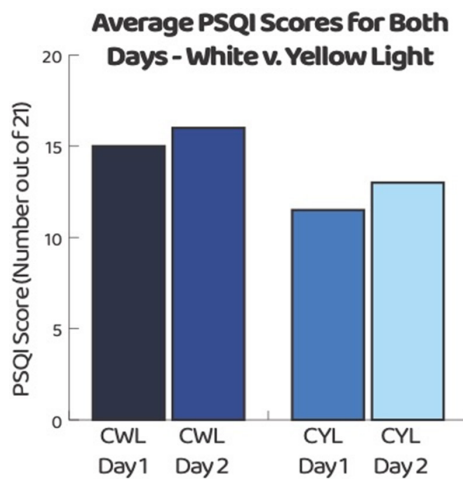


Figure 6. Bar graph of mean values for the PSQI scores of both control groups on both days.

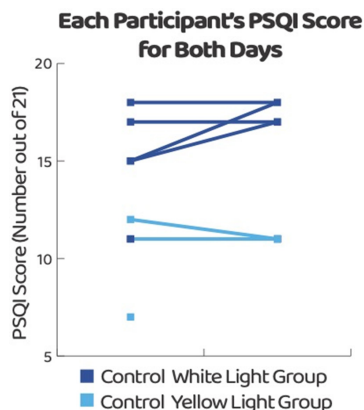


Figure 7. Line graph of each participant's PSQI scores on both days of the experiment.

CONCLUSION

The hypothesis for this study is that exposure to white-colored light will have the strongest increase in LTM capacity for teens with ASD, whereas yellow-colored light will have the strongest increase in LTM capacity for people without ASD. After conducting data analysis, it can be concluded that the hypothesis has been partially proven, but not conclusively, due to the partially statistically significant and partially statistically insignificant results. This is because while the participants in the yellow light control

group had a relatively steeper increase in both the memory test scores and the PSQI scores, the p-value for the control white light group is 0.015 while the p-value for the control yellow light group is 0.75. Therefore, these results cannot significantly prove the original hypothesis. However, it could be inferred that due to the increase in the number of sleep disturbances in the control yellow light group, the yellow light group experienced a relatively steeper increase in their memory test scores over the span of the testing period. There are, however, future steps that need to be taken to improve this study. The first next step is to obtain a larger sample size to ensure the validity of the results.

The next step is to make this study a longitudinal and cross-sectional study in order to test out LTM much more effectively (since LTM occurs over a long period of time) and to see if the trends noticed in teens are the same in other age groups. Finally, a much more holistic memory test should be used for all participants, such as the Test of Memory and Learning (or TOMAL) in order to ensure that LTM is being assessed holistically and as accurately as possible.

BIBLIOGRAPHY

- Southwick, J. S., Bigler, E. D., Froehlich, A., DuBray, M. B., Alexander, A. L., Lange, N., & Lainhart, J. E. (2011). Memory functioning in children and adolescents with autism. *Neuropsychology*, 25(6), 702–710. <https://doi.org/10.1037/a0024935>
- Eilam-Stock, T., Wu, T., Spagna, A., Egan, L. J., & Fan, J. (2016, May 12). Neuroanatomical alterations in high-functioning adults with autism spectrum disorder. *Frontiers*. <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2016.00237/full>
- Hickie I, Naismith S, Ward PB, et al. Reduced hippocampal volumes and memory loss in patients with early- and late-onset depression. *British Journal of Psychiatry*. 2005;186(3):197-202. doi:10.1192/bjp.186.3.197
- Liu, H., & Qiu, Z. (2021). Overexpression of MECP2 in the Suprachiasmatic Nucleus Alters Circadian Rhythm and Induces Abnormal Social Behaviors. *Neuroscience bulletin*, 37(12), 1713–1717. <https://doi.org/10.1007/s12264-021-00746-0>

Jin, Y., Choi, J., Won, J., & Hong, Y. (2018). The Relationship between Autism Spectrum Disorder and Melatonin during Fetal Development. *Molecules (Basel, Switzerland)*, 23(1), 198. <https://doi.org/10.3390/molecules23010198>

Childbrain. (n.d.). ASD assessment: Screening questionnaire. Childbrain. Retrieved October 17, 2024, from <https://www.childbrain.com/path/education/the-asd-assessment-screening-questionnaire/>

Ma, M. A., & Morrison, E. H. (2023, July 24). Neuroanatomy, nucleus suprachiasmatic. In *StatPearls (pp. [page numbers if available])*. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK546664/>

Mariana G. Figueiro, P. figuem@rpi.edu, Bryan Stever-
son, Ma., Judith Heerwagen, P., Kevin Kampschroer, Ma.,
Claudia M. Hunter, P., Kassandra Gonzales, Ms., Barbara
Plitnick, Rn., Mark S. Rea, P., Mariana G. Figueiro, P.
author at: L. R. C., Bryan Steverson, Maa. of F. H.-P. G.
B., Judith Heerwagen, P. of F. H.-P. G. B., Kevin Kamp-
schroer, Maa. of F. H.-P. G. B., Claudia M. Hunter, P. R.
C., Kassandra Gonzales, Msa. R. C., Barbara Plitnick,
Rna. R. C., & Mark S. Rea, P. R. C. (2017, April 23). The
impact of daytime light exposures on sleep and mood in
office workers. *Sleep Health: Journal of the National Sleep
Foundation*. [https://www.sleephealthjournal.org/article/
S2352-7218\(17\)30041-4/abstract](https://www.sleephealthjournal.org/article/S2352-7218(17)30041-4/abstract)

Nair, A. S., Priya, R. S., Rajagopal, P., Pradeepa, C.,
Senthil, R., Dhanalakshmi, S., Lai, K. W., Wu, X., & Zuo, X.
(2022). A case study on the effect of light and colors in the
built environment on autistic children's behavior. *Fron-
tiers in psychiatry*, 13, 1042641. [https://doi.org/10.3389/
fpsyt.2022.1042641](https://doi.org/10.3389/fpsyt.2022.1042641)

Pinato, L., Galina Spilla, C. S., Markus, R. P., & da Sil-
veira Cruz-Machado, S. (2019). Dysregulation of Circadian
Rhythms in Autism Spectrum Disorders. *Current pharma-
ceutical design*, 25(41), 4379–4393. [https://doi.org/10.2174/
1381612825666191102170450](https://doi.org/10.2174/1381612825666191102170450)

Rybak, J. (2006). An open trial of light therapy in adult
attention-deficit/hyperactivity disorder. *Journal of Clinical
Psychology*, 62(9), 1197-1205. [https://doi.org/10.1002/
jclp.20332](https://doi.org/10.1002/jclp.20332)

Morning vs. Evening: Cognitive Training Techniques for Enhanced Memory Retention

Rewa Rathee

ABSTRACT

Many students struggle with finding the best way to study, but what if the key to better memory retention isn't just how you study, but when? This study examines whether cognitive training techniques, specifically spaced repetition and brain training apps, are more effective in the morning or evening. The hypothesis was that memory retention would be higher in the evening and spaced repetition would outperform brain training apps. Participants ($n=12$) were randomly assigned to morning (6 AM – 7 AM) or evening (6 PM – 7 PM) cognitive training sessions (independent variables). Each session involved either spaced repetition or the BrainHQ app (independent variables), followed by a word recall test using an official Scripps National Spelling Bee study resource to measure memory retention (dependent variable). Performance was recorded over seven days, and a two-sample t-test was conducted to compare results. Results partially supported the hypothesis. A statistically significant difference was found between morning and evening groups ($t\text{-value} = -2.67$, $p = 0.0448$), with evening participants demonstrating higher recall scores (average score = 15.17) than morning participants (average score = 13.12). However, while spaced repetition had a higher average recall score (14.85) than BrainHQ (13.43), the difference wasn't statistically significant ($t\text{-value} = 1.45$, $p = 0.1931$), meaning there isn't strong enough evidence to support that spaced repetition was more effective. This may be due to the small sample size, increasing bias. Overall, results suggest that cognitive performance peaks in the evening, supporting the idea that studying later is best for retention.

INTRODUCTION

This research examines whether cognitive training techniques—like spaced repetition and brain training apps—work better in the morning or the evening, and how timing might affect memory retention and cognitive performance.

In today's fast-paced world, where learning and adapting are more important than ever, memory retention plays a key role in how we succeed both academically and professionally. Whether it's students trying to retain material for an exam or adults balancing complex tasks at work, being able to store and apply information efficiently is crucial. Even though memory is such an essential part of daily life, it's often overlooked when we think about improving how we learn or perform mentally. This study aims to investigate whether something as simple as choosing the right time of day to practice cognitive training can make a difference. If so, it could lead to more effective study habits and therapy methods that help a wide range of people.

Our bodies run on a biological clock called the circadian rhythm, which affects everything from sleep to focus. Studies have shown that our cognitive performance—how well we think, remember, and make decisions—can vary depending on the time of day. For example, a study by Khemila et al. (1) found that adolescent athletes performed better on cognitive tests in the afternoon than in the morning, with noticeable improvements in reaction time and

decision-making. This study had a moderately large effect size (Cohen's $d = 0.69$), showing that timing really can have an impact. However, because this research focused on athletes and physical activity, it didn't answer the question of how time of day affects cognitive training when physical exertion isn't involved.

The idea of improving brain function through timing becomes even more interesting when looking at how effective brain training can be on its own. Spaced repetition and brain training apps have already shown positive effects on things like memory, processing speed, and logical thinking. In a review by Da Silva et al. (2), which looked at over 1,200 people, 87% saw faster processing and 74% improved their reasoning skills after using these tools. Still, what is not fully understood are what factors influence how well these tools work—especially when it comes to timing.

Games and apps designed for brain training have also been shown to boost memory. A study by Al-Thaqib et al. (3), which included 150 participants, reported significant improvements in memory and cognitive skills after regular use of these tools. But once again, the study didn't explore whether it matters what time of day people do the training. That's a key detail that's missing from the research. This study tries to fill that gap by testing whether training in the morning or evening makes a difference in memory retention. By focusing on widely used methods like spaced repetition and brain games, the goal is to see if evening

sessions lead to better results because they match the brain's natural performance peak. If the findings support this idea, they could help students, educators, and even healthcare professionals time their learning or therapy sessions to get better results.

The main hypothesis is that people who train in the evening will remember more than those who train in the morning. This is based on existing research—like the Khemila study (1)—that points to better brain performance later in the day. Overall, this study hopes to offer practical, science-backed tips for improving memory through timing. By combining ideas from psychology, neuroscience, and education, it could open the door to smarter ways of learning and teaching.

METHODOLOGY

This study aimed to recruit thirty participants between the ages of 15 and 50 using local advertisements, social media posts, and community outreach. Recruitment was done through an online form where participants submitted basic demographic information, such as age and gender, to help ensure a diverse and representative sample.

To divide participants into two groups—morning or evening—systematic random sampling was used. Each person was assigned a random number, and those with odd numbers were placed in the morning group (6 AM–7 AM), while those with even numbers were assigned to the evening group (6 PM–7 PM). This approach helped ensure equal group sizes and reduced selection bias.

Before participating, all participants signed an informed consent form that explained the study's purpose, procedures, and potential risks. It also made it clear that participation was voluntary and that they could withdraw at any time. For those under 18, parental consent was also collected.

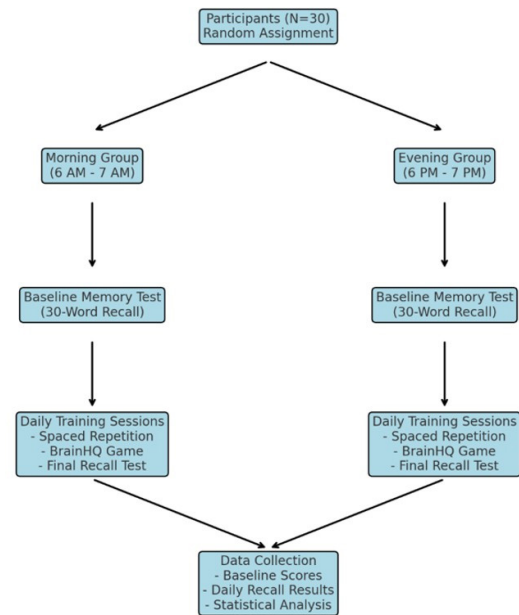
Participants were told which group they were in and given a general overview of the study's goals. While this could introduce some bias, the use of random group assignment, consistent procedures, and objective scoring helped reduce that risk. Every task was kept the same across both groups to support fairness and reliability.

The study utilized the following materials: the BrainHQ cognitive training platform (4), a word-recall test featuring 30 words (randomly selected words from a Scipps National Spelling Bee Study Resource) (5), notebooks or logbooks for recording results, laptops, and Microsoft Excel for data analysis. The training techniques included spaced repetition and memory-focused exercises from BrainHQ, selected for their game-based cognitive training capabilities.

To test the hypothesis that evening cognitive training improves memory retention more than morning training, participants followed a daily training program for one

week. On Day 1, everyone completed a baseline memory test to measure their starting point. This test was given at each group's assigned time—6 AM or 6 PM—and involved showing a 30-word list, with five seconds per word. Afterward, participants tried to recall as many words as possible, and the number of correct words was recorded.

Figure 1. Flowchart for Experimental Procedure



Caption: Step-by-step breakdown of the experimental procedure, outlining each stage of the process for clear and effective implementation

Each training session followed the same structure every day. Participants started with a spaced repetition activity: they reviewed a 30-word list and immediately completed a recall test. The words were then reviewed again after 5, 10, and 20 minutes, with recall tests after each review. After this, they spent 20 minutes playing memory-training games on BrainHQ. At the end of the session, they took a final recall test, and the results were recorded through a Microsoft Form. These steps were identical for both groups to maintain consistency.

Data collection included baseline scores, daily recall scores, and average performance across the week. All data were organized into tables for easy comparison. Statistical analysis involved calculating the mean, median, and mode to identify trends. To compare the morning and evening groups, a two-sample t-test was used. Before applying the test, assumptions such as normality and equal variance were checked. A significance level of 0.05 was used to evaluate whether the differences between the groups were statistically meaningful.

RESULTS

Table 1. Spaced Repetition Test Scores

Participant ID	Group (Time of Day)	Baseline Score	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
1	Morning	10	12	13	14	14	15	15	16	14.1
2	Morning	9	11	12	13	14	13	13	14	12.9
3	Evening	11	14	15	16	16	17	17	18	16.1
4	Evening	10	12	14	15	16	15	16	17	15.1
9	Morning	10	13	14	15	15	16	17	17	15.3
10	Evening	11	14	15	16	16	15	17	18	15.6

Caption: Memory retention test scores across 7 days for participants using spaced repetition, categorized by time of day (morning vs. evening).

Table 2. BrainHQ Test Scores

Participant ID	Group (Time of Day)	Baseline Score	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
5	Morning	8	10	11	12	12	13	13	14	12.1
6	Morning	7	9	10	11	12	11	12	13	11.1
7	Evening	9	13	14	15	15	16	16	17	15.1
8	Evening	10	12	13	14	15	15	16	17	14.9

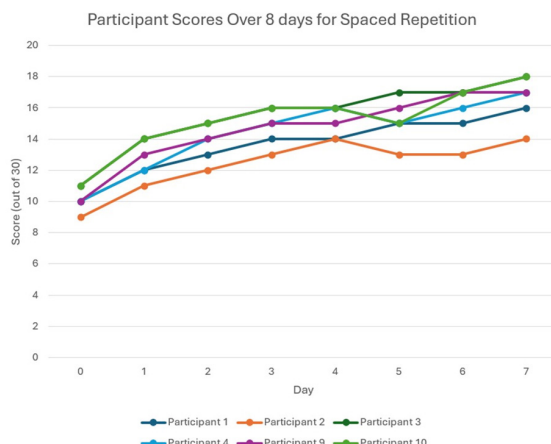
Caption: Memory retention test scores across 7 days for participants using BrainHQ, categorized by time of day (morning vs. evening).

Table 3. Overall Group Performance for Morning vs. Evening

Group	Average Day 1 score	Average Day 7 score	Change (%)
Morning	11	14.83	34.82
Evening	12.83	17.17	33.84

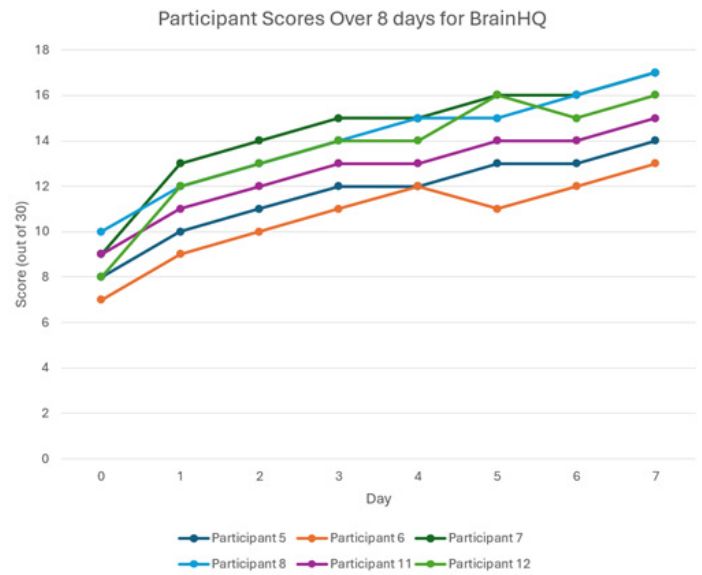
Caption: Comparison of average scores on day 1 and day 7, with percentage change, categorized by time of day (morning vs. evening).

Figure 2. Graphed Scores for Spaced Repetition



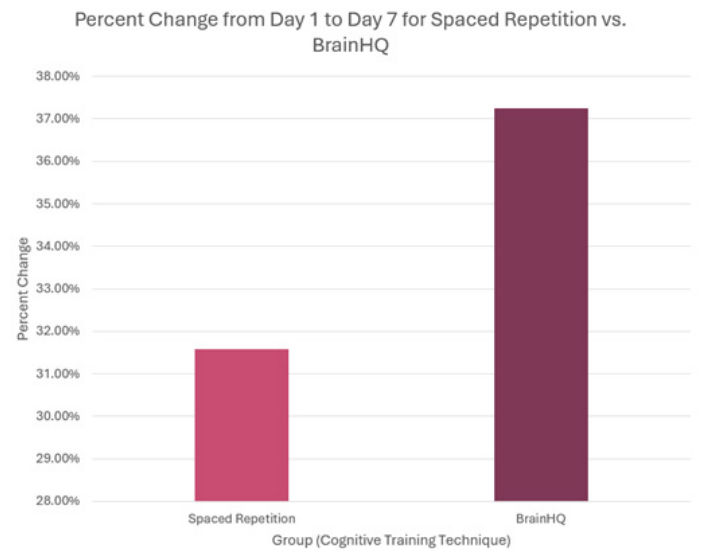
Caption: Graph of participant scores over 8 days for the spaced repetition group showed all participant scores increased.

Figure 3. Graphed Scores for BrainHQ



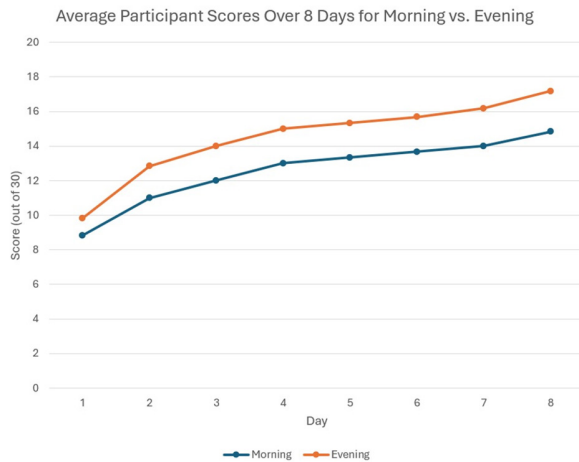
Caption: Graph of participant scores over 8 days for the BrainHQ group showed all participant scores increased.

Figure 4. Graph of Percent Change from Day 1 to Day 7 (excluding baseline scores) for the Spaced Repetition vs. BrainHQ comparison



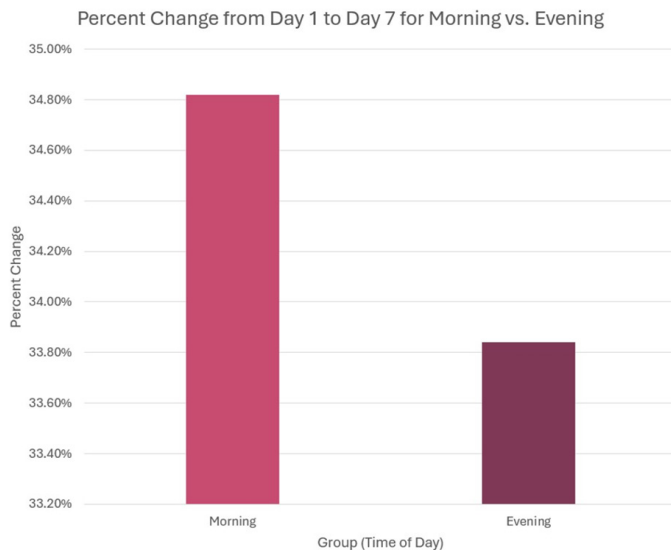
Caption: The BrainHQ group had a higher percent change (37.24%) than the spaced repetition group (31.58%).

Figure 5. Graph of Average Participant Scores over 8 days for the Morning and Evening Groups



Caption: Graph of participant scores over 8 days for the morning and evening groups showed all participant scores increased, with the evening participants consistently performing higher than the morning participants.

Figure 6. Graph of Percent Change from Day 1 to Day 7 (excluding baseline scores) for the Morning vs. Evening comparison



Caption: The morning group had a higher percent change (34.82%) than the evening group (33.84%).

Discussion of Results

This study aimed to determine whether cognitive training techniques, specifically spaced repetition and brain training apps (BrainHQ), are more effective in improving memory in the morning or evening. The hypothesis predicted that memory retention would be higher in the evening and that spaced repetition would be more effective than the BrainHQ app.

The results partially supported the hypothesis. Participants in the evening sessions demonstrated a lower percentage increase in recall performance compared to the morning group. Numerically, the average recall score in the evening group showed a slightly larger increase of 34.82% from the baselines compared to the morning group's 33.84% (Figure 6). However, the average recall scores themselves were higher for the evening group (15.17) when compared to the morning group (13.12), highlighting discrepancies in the baseline scores. A two-sample t-test using Welch's t-test formula confirmed a statistically significant difference between these groups (t -value = -2.67, p = 0.0448), indicating that cognitive performance was indeed higher in the evening. Regarding the effectiveness of the training techniques, spaced repetition yielded a slightly higher average recall score (14.85) than BrainHQ training (13.43). However, the difference was not statistically significant (t -value = 1.45, p = 0.1931), suggesting that both methods are similarly effective, at least in the scope of this study. The results indicate that while time of training plays a role in cognitive retention, the specific method of training (between the studied methods) may not be as impactful given the small sample size.

Limitations & Next Steps

Several challenges were encountered during this study. One major limitation was the small sample size ($n=12$), which heavily restricted the generalizability of the findings. A larger sample might show clearer patterns and help factor in individual differences such as prior cognitive training experience, prior familiarity with the material, etc. Additionally, external factors such as typical sleep schedule and study habits were not controlled, which could have influenced performance. For example, participants in evening sessions may have had more exposure to school-related cognitive activities before the study, potentially contributing to their higher baseline scores and therefore lower percent increase.

Another issue was the wide age range from 15 to 50 years. Studies have shown that fluid intelligence (6), or the ability to reason and think quickly, decreases over time which may have been a neurological factor that affected the ability of participants of different ages to perform differently. In addition, participant variability in motivation and consistency was another problem that rose from this study. Some participants may have been more engaged

or well-rested at different times of the day, aligning with individual circadian rhythms, influencing their ability to retain information. This could also explain why morning participants exhibited higher percent changes – starting at a lower baseline meant they had more room for improvement compared to evening participants, who may have already been operating closer to their peak cognitive performance.

To strengthen the validity of these findings, future studies should increase the sample size to include a more diverse participant population. Investigating factors like sleep quality, prior cognitive training exposure, and personal study habits would provide more information on the mechanisms of memory retention at different times of the day. Additionally, taking into account the sleep schedules of participants, groups should be created to place participants of similar sleep schedules together and compare results. Converting the study into a longitudinal investigation over several months could also reveal whether these effects grow stronger over time or if adaptation occurs.

Another change to this study would be to utilize neurophysiological measures like an electroencephalogram (EEG) or functional magnetic resonance imaging (fMRI) to assess real-time brain activity during cognitive training. This would allow for any participant/implicit biases to be removed and would help determine whether certain neural networks are more active in the morning or evening, providing concrete evidence of the biological mechanisms of the observed differences. Future research could also explore how individual chronotypes influence training results – morning-oriented individuals might benefit more from early sessions and vice versa for night-oriented individuals.

Implications of Research

The findings of this study have significant applications for students, teachers, and professionals seeking to enhance learning and memory retention. Understanding the best time of day for cognitive training could help create personalized strategies to maximize efficiency, potentially leading to better academic performance. Additionally, these learnings could be valuable in neurorehabilitation, where cognitive training is used for patients recovering from brain injuries or cognitive decline due to neurological disorders. Moving forward, researchers could explore how different types of memory tasks – such as procedural memory (skill-based learning) versus declarative memory (fact-based learning) – are influenced by training time of day. This research sets the stage for future investigations and by improving and expanding this study, researchers can develop targeted cognitive training programs that optimize learning based on an individual's peak cognitive periods. Overall, these findings could not only contribute to personalized education strategies, but also cognitive rehabilitation programs and more, making sure that individuals can enhance their learning efficiency based on their unique cognitive preferences.

BIBLIOGRAPHY

1. Khemila, S., Romdhani, M., Abedelmalek, S., Chtourou, H., Souissi, M. A., BenTouati, E., & Souissi, N. (2022b). *The effect of time of day and high intensity exercise on cognitive performances of elite adolescent karate athletes. Chronobiology International*, 39(12), 1542–1553. <https://doi.org/10.1080/07420528.2022.2132165>
2. Da Silva, T. B. L., Bratkauskas, J. S., De Castro Barbosa, M. E., Da Silva, G. A., Zumkeller, M. G., De Moraes, L. C., Lessa, P. P., Cardoso, N. P., Ordonez, T. N., & Brucki, S. M. D. (2022). *Long-term studies in cognitive training for older adults: a systematic review. Dementia & Neuropsychologia*, 16(2), 135–152. <https://doi.org/10.1590/1980-5764-dn-2021-0064>
3. Al-Thaqib, A., Al-Sultan, F., Al-Zahrani, A., Al-Kahtani, F., Al-Regaiey, K., Iqbal, M., & Bashir, S. (2018). *Brain training games enhance cognitive function in healthy subjects. Medical Science Monitor Basic Research/Medical Science Monitor. Basic Research*, 24, 63–69. <https://doi.org/10.12659/msmbr.909022>
4. BrainHQ. “Word List Recall - BrainHQ.” BrainHQ, 1 Nov. 2023, www.brainhq.com/brain-connection/word-list-recall.
5. Scripps National Spelling Bee. (2021). *Words of the champions [PDF]*. Retrieved March 1, 2025, from https://www.spellingbee.com/sites/default/files/inline-files/2021_Printable_Words%20of%20the%20Champions.pdf
6. Fractus Learning. (n.d.). *How brain exercise improves working memory. Fractus Learning*. Retrieved March 1, 2025, from <https://www.fractuslearning.com/brain-exercise-improves-working-memory/>

Surviving In a Pressure Cooker: A Research Project on the Effects of Cultural Pressures on Self Esteem in High School Students and Resulting Professional Pathways

Zeeshan Shaik

ABSTRACT

From childhood, individuals are shaped in the culture that they grow up in. which influences their self-perception and self-esteem. But one thing that has not been explored yet is how does cultural pressure influence the future pathways of specifically high school students? This research project looks at the possible relationship that exists between two factors: cultural pressures (referring to the societal expectations that one is held to in their culture, i.e. valuing STEM jobs over arts jobs), self-esteem, and whether people are motivated by either their personal ambition or whether they are motivated to align themselves with their cultural expectations. To investigate this, high school participants were asked to complete a ten-question survey that assesses the effect that cultural pressures on their decision making. Participants were also asked to report their levels of self-esteem using the Rosenberg Self Esteem Scale. The results from the study suggested that there is a negative correlation between the variables, with a R^2 value of -0.2824. Results from this study also proved to be statistically significant, with an P value calculated to be 0.015, meaning that there is a correlation between cultural pressures, self-esteem, and the resultant professional pathways that high school students will take. To conclude this study, the hypothesis that I had come up with was proven to be correct and shows that there is importance in addressing both variables in the outlook of high school students.

INTRODUCTION

This study aims to investigate the dynamic interplay that exists between self-esteem, cultural pressures, and the resulting occupational pathways that exist in high school students. The decisions that high school students play a significant role in what occupational pathways that they will be encouraged to go into. The classes, tests, and standardized tests that they take during this time will be aligned with the career that they want to go into.

But what influences them into going on one specific career pathway? Are these students motivated by their own wants, or are they conformant with what their parents (or culture) want them to pursue? By looking at the influences that dictate what field one person will go into. This critical period of high school shapes the futures that these high school students will embark on, and it is important to understand the many influences that they students will be exposed to and see how that affects their self-esteem and what they want to do with their lives.

Research on the topic of self-esteem in high school students in relation to their ethnicity is not a new topic. For example, research surveying teenagers found that Asian American students tended to have the lowest self-esteem whilst African American teenager students tended to report higher levels of self-esteem on average, even though they are viewed as the model minority, which is, a minority that is perceived to high levels of success. (Bachman et.al,

2011) This study reveals that there might be a connection with the level of self-esteem that a student will have, and the background that they come from. Students that come from these cultures that high levels of success, such as Asian American cultures, put high amounts of expectations on their children to achieve the same levels of success that is found from other people from their culture. These expectations that are placed on the children tend to make them feel that they are not valued, even if they are the highest achievement in school as also found by the study.

This becomes more poignant when considering the effect that these cultural pressures have upon the career pathways that students will embark on. In an The Independent survey in the United Kingdom, it was found that four in ten ethnic minority workers needed to lie about their careers to their family due to feeling that they were not living up to the standards that were expected of them. This consequently led to a decrease in self-confidence and self-efficacy among these workers. These findings bring up valid concerns about how these cultural pressures can affect the mental health of these workers. What's more, these findings raise the question of how these two variables (cultural pressures and self-esteem) affect the decisions that these people will make from high school to their professional careers. Moreover, it is known from previous research that the effects of having high self-esteem is beneficial and allows for people to perform better in their places of work (Buamiester et.al, 2003).

Although much valuable information is derived from existing studies regarding the individual influences of self-esteem and cultural pressures, there are still considerable gaps in the literature regarding how these two forces interactively influence the occupational outcomes of students. Adolescents' career aspirations and choices depend not only on one's self-concept but also on the cultural expectations and values possessed by the environment. People originating from cultural backgrounds that place a strong emphasis on prestigious professions-for example, medicine, law, or engineering-may be under compulsion to engage in such career pursuits even though they may not correspond to personal interests and abilities. Additionally, the level of self-esteem that these students have might dictate their interest in pursuing the careers that they align themselves with. Students with low self-esteem and high level of cultural pressure might be bound to putting themselves in careers that align with the prestigious professions that the culture might have dictated, while the opposite effect might occur for students with high self-esteem and low levels of cultural pressures.

METHODOLOGY

This research project employed the method of finding and tackling the issue with the correlation between self-esteem, cultural pressures, and occupational outlooks by the survey method. Using Google Forms, both the Rosenberg Self Esteem Scale and the custom questionnaire that was made to see the effects of cultural pressures upon the occupational futures of high schools' students. After these questions were made, a QR code linking to the survey was made and distributed across high schools and online to collect as many participants as possible.

Figure 1: Cultural Pressures Survey



Because of the use of human participants in this research, approval from the Scientific Review Committee was required before continuing with data collectors. Participants who were minors during the study were required to fill out informed consent forms and get parental confirmation to submit their data, those who had not done so were ignored and not counted within the rest of the data. The participants were able to take part in this study based on their own volition and were able to take it from the comfort of their own homes without the researcher nor mentor being present. If questions about the research plan had come up, the researcher was able to be contacted by email.

The Google Form (Figure 3) in this study involved two major sections. The first was the Rosenberg Self-Esteem Scale, developed by Morris Rosenberg (Figure 2) as a question tool to understand the level of individuals' self-esteem. The scale has been widely considered reliable and commonly used in the field of psychology as a foundation for measuring one's self-esteem on a standard basis. The second half of the questionnaire consisted of specific questions that were meant to be more informative. There were also questions concerning their future regarding their intended career, college plans, and factors that influence such decisions. The questions were measured on the Likert scale of 1 (being the weakest) to 4 (being the strongest).

Figure 2: Rosenberg Self Esteem Scale

STATEMENT	Strongly Agree	Agree	Disagree	Strongly Disagree
1. I feel that I am a person of worth, at least on an equal plane with others.	☐	☐	☐	☐
2. I feel that I have a number of good qualities..	☐	☐	☐	☐
3. All in all, I am inclined to feel that I am a failure.	☐	☐	☐	☐
4. I am able to do things as well as most other people.	☐	☐	☐	☐
5. I feel I do not have much to be proud of.	☐	☐	☐	☐
6. I take a positive attitude toward myself.	☐	☐	☐	☐
7. On the whole, I am satisfied with myself.	☐	☐	☐	☐
8. I wish I could have more respect for myself.	☐	☐	☐	☐
9. I certainly feel useless at times.	☐	☐	☐	☐
10. At times I think I am no good at all.	☐	☐	☐	☐

Score Results Reset

Your score on the Rosenberg self-esteem scale is:

Scale that was used to find numerical value to the levels of self-esteem within the participant

Figure 3: Cultural Pressures Survey

Cultural Pressures Survey

	Strongly Disagree	Disagree	Agree	Strongly Agree
My family has specific expectations for my career	☐	☐	☐	☐
I feel pressure to choose a career that aligns with my family's values or cultural background	☐	☐	☐	☐
I believe my family would disapprove if I pursued a career outside of their expectations	☐	☐	☐	☐
My cultural background influences how I view different career options	☐	☐	☐	☐
I often compare my career choices	☐	☐	☐	☐
How likely are you to pursue a career that aligns with your family's expectations?	☐	☐	☐	☐
How likely are you to pursue a career based on your personal interests, even if it goes against your family's expectations?	☐	☐	☐	☐
How Confident do you feel about following a career path that interests you, despite cultural pressures?	☐	☐	☐	☐
To what extent do cultural values shape your long-term career goals?	☐	☐	☐	☐
How likely are you to explore non-traditional or less culturally favored career paths?	☐	☐	☐	☐

Survey used to gauge the occupational pathways that high school students will go into because of their cultural pressures.

After participants were collected, they were randomly selected through the usage of a random number generator from random.org. They were then randomized according to the flowchart (Figure 4) to maintain a representative sample.

Figure 4: Randomization Flowchart

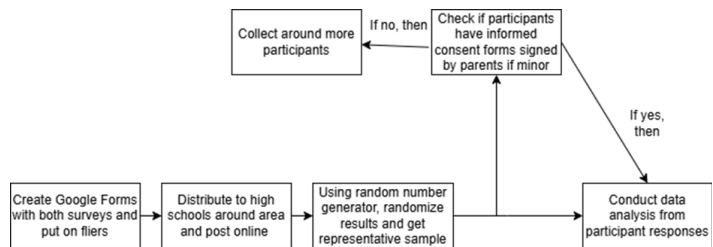


Figure 4: Flowchart that was used to randomize data to attain generalizable data.

RESULTS

After 2 weeks of participating in the collection, the research project had collected a sample size of around $n=20$. After the participants had sent their responses for all the questionnaires that they were asked to do, the scores were converted from text to a number scale, 1 to 4 for the cultural pressures (Figure 5) and the Rosenberg Self Esteem Scale scores were reported as well (Figure 6). Scores were then summed together and plotted using Microsoft Excel before the p-value alongside the Pearson Correlational Coefficient was calculated.

Figure 6: Data table showing the participants cultural pressure survey scores

Participant	RSSS Score
1	16
2	17
3	19
4	26
5	28
6	12
7	15
8	18
9	22
10	30
11	25
12	32
13	28
14	20
15	14
16	24
17	27
18	19
19	31
20	29

The means, median, and mode for both surveys were calculated. For the Rosenberg Self Esteem Scale, there was a mean of 22.6, a median score of 23, and a mode of 19. For the Cultural Pressures survey, there was a mean of 28.75, a median of 28.5, and a mode of 28.

Figure 7: Mean, Median, and Mode of Rosenberg Self Esteem Scale

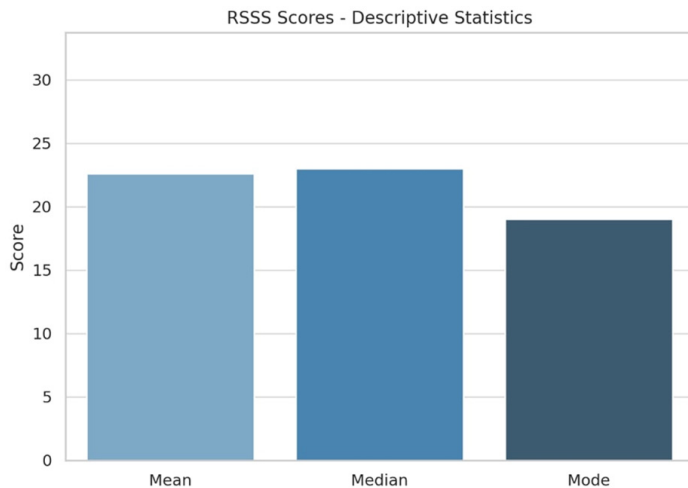
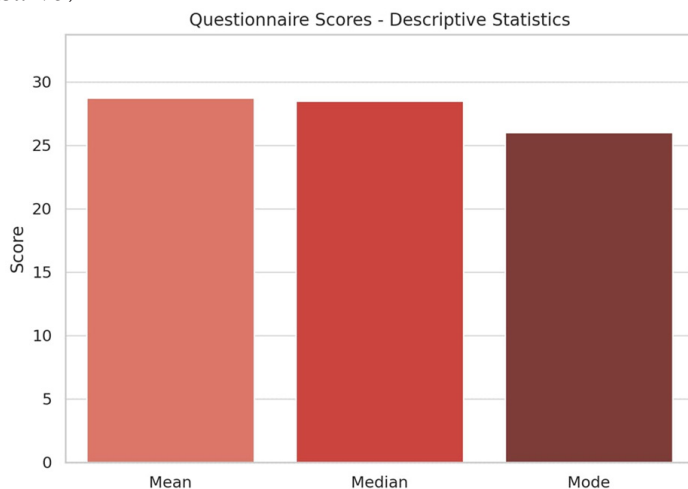
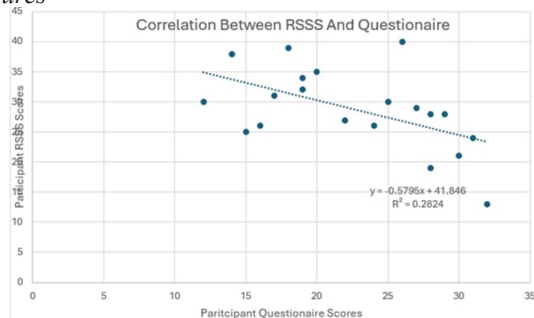


Figure 8: Mean, Median, and Mode of Cultural Pressures Survey



For the values that were calculated, it was found that there was a p value of 0.015, suggesting that the results that were achieved by this investigation were statistically significant and not due to chance. Moreover, the Pearson Correlation Coefficient was calculated to be -0.2824, suggesting there is a moderate negative correlation between the two results proving the hypothesis correct as lower amounts of self-esteem proved to have higher cultural pressures and vice versa.

Figure 9: Correlational Graph between Self-Esteem and Cultural Pressures



However, it is important to consider the problem that arose during participant selection. The participants that were received were not diverse, with most participants predominantly being Asian males. This brings in a strong bias towards that ethnicity and does not allow this study to be applicable to all high school students across the nation.

CONCLUSIONS

Moving to the conclusions of this study and its applicability across high schools across the nation, it is important to understand that this project holds importance in helping high school students and their mentors alike in understanding whether they are making decisions based on their own self-interest or whether they are making it based on cultural pressures. Helping to make informed decisions about their future pathways would help understand the effects of the people around them on their self-determination.

CITATIONS

- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? *Psychological Science in the Public Interest*, 4(1), 1-44. <https://doi.org/10.1111/1529-1006.01431>
- Pushpa, Ramnath, & Singh, S. (2017). Peer Pressure among Adolescents in Relation to Family Climate. *Indian Journal of Health and Wellbeing*, 8, 196-199.
- Bachman, J. G., O'Malley, P. M., Freedman-Doan, P., Trzesniewski, K. H., & Donnellan, M. B. (2010). Adolescent self-esteem: Differences by race/ethnicity, gender, and age. *Self and Identity*, 10(4), 445-473. <https://doi.org/10.1080/15298861003794538>
- Bachman, J. G., Johnston, L. D., O'Malley, P. M., & Schulenberg, J. E. (2006). *The Monitoring the Future project after thirty-two years: Design and procedures (Monitoring the Future Occasional Paper No. 64)* [Online]. Ann Arbor, MI: Institute for Social Research, 94 pp. Available: <http://monitoringthefuture.org>
- Bornstein, M. H. (2013). Parenting and child mental health: A cross-cultural perspective. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 12(3), 258-265. <https://doi.org/10.1002/wps.20071>
- Jungen, K. A. (2008). Parental influence and career choice: How parents affect the career aspirations of their children. *University of Wisconsin-Stout*, 52(2), 60-66. Retrieved from <https://core.ac.uk/download/pdf/5067235.pdf>
- Smith, J. S., Estudillo, A. G., & Kang, H. (2011). Racial differences in eighth-grade students' identification with academics. *Education and Urban Society*, 43(1), 73-90. <https://doi.org/10.1177/0013124510379403>
- Wanelik, K. M., Griffin, J. S., Head, M. L., Ingleby, F. C., & Lewis, Z. (2020). Breaking barriers? Ethnicity and socioeconomic background impact on early career progression in the fields of ecology and evolution. *Ecology and Evolution*, 10(14), 6870-6880. <https://doi.org/10.1002/ece3.6423>

Using Behavioral Methods to Measure the Effects of Joint Attention Learning in Male and Female Infants with a Familial History of Autism

Shriya Venkatesh Krishnan

ABSTRACT

Early and accurate learning interventions are extremely important in autism spectrum disorder (ASD) research, especially for the over/under diagnosis of males and females with a familial history (FH+) of ASD. Joint Attention (JA), which is the shared focus on an object or event using eye contact, contributes to social and visual learning, but it is unclear if males and female infants with FH+ respond differently to JA-based learning. This study investigates the gender-specific effects of behavioral methods in JA development in infants with a familial history of autism and aims to understand the relationship between JA and ASD symptoms. Autism Spectrum Disorder affects 1 in 36 children in the U.S., with male infants diagnosed nearly four times more frequently than females. Given that JA deficits (not following a caregiver's gaze or pointing) are early indicators of high risk of ASD. De-identified NESSE.2 video recordings were analyzed, tracking novelty preference (proportion of time looking at a novel vs. familiar objects) across JA trial conditions. Independent variables included ASD family risk, sex and trial type (JA vs non-JA videos), while the dependent variable was the novelty preference (measuring encoding of familiar objects). A three-way ANOVA was conducted and found no significant sex-JA interactions ($F=0.001$, $p=0.810$), indicating that boys and girls learn similarly across JA trials. These findings suggest that while learning using JA may improve learning interventions, there were no sex-specific differences found in the eye contact. This urgently highlights the need for further research into other potential factors that influence ASD interventions for males and females.

INTRODUCTION

This research studies the gender-specific effects of targeted behavioral learning of joint attention development in male and female infants with a familial history (FH+) of Autism Spectrum Disorder (ASD), aiming to uncover how his intervention may differently impact early communication and various social outcomes. When a child is diagnosed with ASD, the next child of the same parents have a 20% greater chance of also having autism [1] with early signs often in joint attention (JA) deficits, which defines as the ability to coordinate focus on an object or event with another person [2]. Studies suggest that male infants are diagnosed with ASD four times more frequently than female infants, emphasizing the need to explore gender differences in early developmental markers. Research has shown that infants with a familial history of ASD are at a higher risk of lacking JA learning, which can influence their future encoding of specific memories and cognition. For instance, one study found that these infants exhibit fewer JA behaviors compared to peers without familial history [3]. Behavioral intervention that focuses on JA have been successful in improving in social engagement scores following intervention [4]. However, these studies lack understanding in how JA learning differs between male and female infants.

The overdiagnosis of ASD for males calls for a deeper analysis on if JA learning can be used in learning treatments for males and females to understand the delay of attention interventions. A study conducted by Lauren Smith found that infants with a family history of autism exhibited significantly fewer JA behaviors compared to their peers without a family history, and it highlighted the importance of early learning in males versus females. Given the crucial role of JA in shaping eye contact behaviors, early identification of JA deficits is critical for improving long-term memory and cognition.

Behavioral interventions targeting the lack of JA behaviors have shown to improve social outcomes in at/high-risk populations, including infants with familial risk factors for autism. These learning treatments often focus on enhancing the infant's ability to engage in shared attention with caregivers, which can aid in the development of communication and social eye contact. Though, it has not been studied yet how male and female infants with a familial history of ASD may respond differently to these interventions, either due to stereotypical masking of gender traits or biological factors.

A potential solution for addressing these differences is in the application of targeted behavioral methods to study JA behaviors of these infants and help modify current methods

of interventions and development. Research suggests that early identification and intervention during infancy of FH+ infants can reduce some of the challenges associated with insufficient JA behaviors, which can offer a pathway for improved developmental outcomes in the future [4]. The hypothesis of this study is that male and female infants with a familial history of autism will display distinct novelty preference patterns during JA tasks, that can reflect potential gender differences in cognitive and social developments.

METHODOLOGY

This study investigates the influence of risk, sex, and trial type on the novelty preference in infants at risk for ASD. The novelty preference is defined as the dependent variable and is quantified by the total looking time at a novel object divided by the total looking time at both the novel and familiar object, providing a measure of how much time the infants spends encoding the familiar object into memory, using eye contact. the independent variables are defined as trial type (trials with JA vs trials without JA), sex (male infants vs female infants), and risk (high risk vs low risk infants). Figure 1.1 highlights the overall setup for novelty preference trials, including the placement of the novel and familiar objects during encoding testing. Participants were selected from the UCSD SoCal lab data set (NESSE.2), which includes de-identified videos of infant behaviors during JA tasks. The infants in this study were categorized based on their sex and familial risk for ASD, which is defined by the extent of ASD diagnoses within the infant's immediate and extended family. While the specific age range of the infants were not specified in the study, this was considered a potential limitation and confounding variable in the data analysis.

The experimental design of this study included a within-subject approach, which is a type of study design that tests all level of the independent variable on one group of participants. In this study, each infant participated in trials that were sorted into social (JA) and nonsocial (without JA) trials. In the nonsocial trials, no eye contact was made between the individuals in the video and the infant, whereas in social trials, eye contact was established, demonstrating JA. The nonsocial trials included two conditions: a first-party nonsocial trial, where only one individual was present without eye contact with the infant, and a third-party nonsocial trial, where two individuals did not make eye contact with each other, but both looked at the object. The social trials similarly included first

and third-party conditions, where eye contact was either made with the infant or between two individuals, with the novel object placed for the infant to view (Figures 2.1 and 2.2).

Title: Experimental Video Setup



Figure 1.1 During the Novelty Preference Trials, Infants were presented with a novel and familiar object to calculate their novelty preference. Deidentified UCSD data set embedded image.

The materials for this experiment include a MacBook Pro 2020 13-inch for video processing, the DUO Mobile application for secure server access, Datavyu v1.3.8 software for encoding coding, and video clips from the SoCal Lab NESSE.2 dataset, which depicted various joint attention and non-joint attention trials. These video clips were downloaded from the Powell Polygon Lab Server, and coding templates were additionally downloaded to track the infants' eye gaze directions throughout the multiple trials. The coding process involved marking the onsets and offsets of each trial, tracking eye movements by using frame-by-frame analysis, and recording whether the infant was looking left or right (at the novel versus familiar objects). Gaze direction was tracked throughout 69 participant videos, by systematically coding if they were looking at the novel or familiar object, analyzing the video frame-by-frame. Each infants' eye movements were coded to see which object they were fixating on, which allowed further assessment of the novelty preference. If a trial was invalid, due to the lack of attention and the infants inability to sit still, it was dropped from further analysis and no data was recorded for that trial. As the infants watched video recordings that either included or excluded JA trials, their eye gaze directions offered insightful patterns on their social and mental cognition and encoding of specific objects and movement.

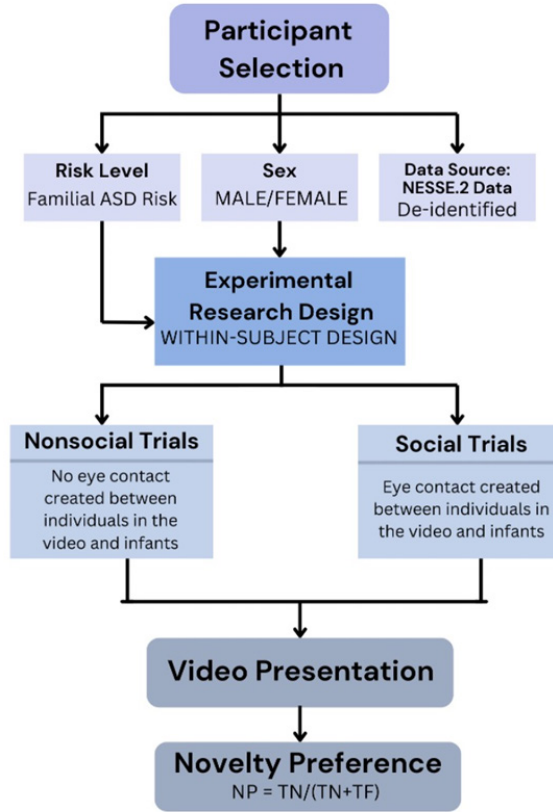


Figure 2.1 Flowchart showing the Coding and participant selection process of the Videos

Nonsocial Trial Conditions:

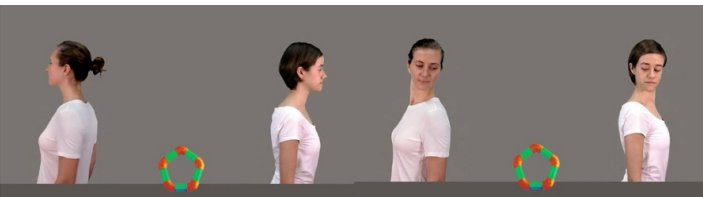


Figure 3.1 Encoding Video that does not display Joint Attention as the individuals in the video are not demonstrating eye contact. UCSD study datasets and embedded images from software.

Title: Social Trial Conditions

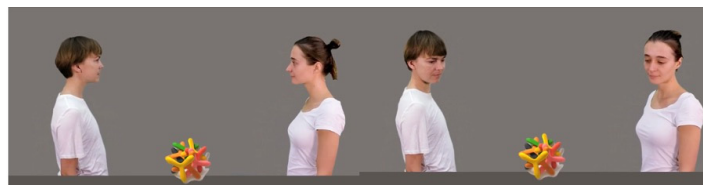


Figure 3.2 Scene of Joint Attention (first party social) present in the experimental videos shown to infants

DATA ANALYSIS

The first step of analyzing the data involved calculating the novelty preference for each trial type by using the given formula:

$$NP = \frac{T_N}{T_N + T_F}$$

The variables are defined as:

NP: Novelty preference, a proportion representing the amount of time the infant spent looking at the novel object relative to the total time spent looking at both the novel and familiar objects.

TN: Total looking time at the Novel Object, measured in seconds

TF: Total looking time at the familiar object, measured in seconds

The novelty preferences were calculated for each infant across different trial types and conditions, and a three-way analysis of variance (ANOVA) was conducted to assess the main effects and interactions between Risk, Sex, and Trial Type. Main effects refer to the independent effect of each independent variable (Risk, Sex, Trial Type) on the dependent variable (Novelty Preference) across all levels of the other variables. These interactions occur when the effect of one independent variable on the dependent variable changes depending on the level of another independent variable. The analysis also includes three two-way interactions (Risk*Sex, Risk*Trial Type, Sex*Trial Type) and the three-way interaction (Risk*Sex*Trial Type). In case of significant two-way interactions, post-hoc t-tests were used to compare the novelty preference between different groups in JA and no-JA trials, which are statistical tests conducted after an initial ANOVA shows if there are significant results and interactions. These tests identify specific differences in groups by performing pairwise comparisons while calculating for the overall error rate, reducing the likelihood of a Type 1 false positive error.

Ethical Considerations

Ethical considerations were addressed by utilizing de-identified participant data to ensure privacy and confidentiality. All video data was collected with appropriate parental consent and was given anonymously before analysis to maintain compliance with ethical guidelines for research involving minors.

Results

The results of this study state that there was no observed statistically significant difference between trials with JA and no JA ($F(1) = 0.124, p=0.725$), not first party or third party trials ($F(1) = 0.165, p=0.725$). There also contained no significant interaction of JA and perspective ($F(1) = 0.165, p=0.725$).

Additionally, there was no significant difference of JA and trial type ($F(1) = 0.539$), meaning that although there was a marginal learning difference observed in first party JA trials, this learning was not statistically significantly different from learning in other trials.

There was no observed interaction between sex and JA ($F(1) = 0.786, 0.971$), which means that overall boys and girls learn similarly across trials with and without eye contact, and there was no three-way interaction observed between sex, JA, and trial type ($F(1) = 0.001, p=0.971$), meaning that boys or girls did not learn better in trials with direct eye contact (social trials, more commonly first party JA).

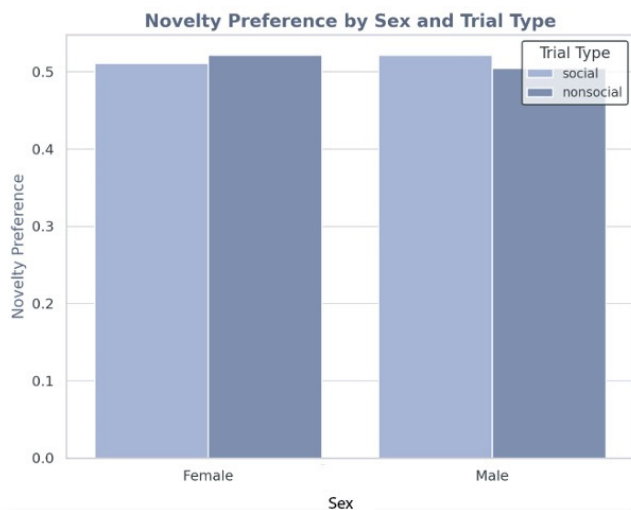


Figure 4.1 Bar graph demonstrating no statistical significant data between the male and female infants' novelty preference

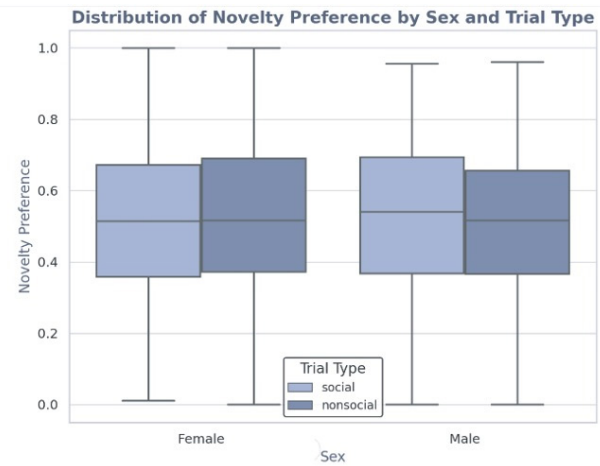


Figure 4.2: Box Plot demonstrating no statistical significant data between the male and female infants' novelty preference

CONCLUSION AND FUTURE IMPLICATIONS

This study explored the potential sex-based differences in JA learning among male and female infants with familial history ASD, aiming to uncover whether early JA learning interventions result in significantly distinct outcomes based on gender. The findings indicate that there was no statistically significant difference in novelty preference between trials with and without JA, nor between first-party and third-party trials. Additionally, the insignificant interactions between sex and JA learning suggests that male and female infants did not respond differently to JA-based learning tasks. These results highlight that while JA-focused behavioral interventions may improve social outcomes, the effects of JA learning do not appear to differ significantly by sex in early infancy development. These results and analysis suggest that other outside factors and confounding variables, such as cognitive, social, or biological factors play a more impactful role in early JA learning outcomes.

If this study were to be taken further, future research would be implemented into these strategies. Significant differences were not found between eye contact, but further research into brain waves could prove to have a statistical significance in the difference of learning in male and female infants containing FH+. To execute these brain wave studies, an Electroencephalogram (EEG) could be used to find differences in brain activity that occurs within these infants. Another option for this would be a Function Near-Infrared Spectroscopy (fNIRS) Cap. These caps are used to record cerebral changes in blood flow using low levels of non-ionizing light through sensors on the top of the head. This can help understand how different brain processes can explain the behavior of these infants while watching the JA/non-JA trials. Additionally, these two techniques provide low-cost solutions to effectively improving learning interventions for FH+ infants.

Another potential solution would be to include different types of eye contact studies using real human participants. Since this study was limited to analyzing and coding from a dataset, future implications would include utilizing the potentials of human participants, with the proper ethical considerations. In doing this, more eye contact data could be collected using different methods outside of JA. Finally, studying parental styles of these infants would aid eliminating confounding variables in this research study. Seeing if different parenting styles that implement eye contact into social and emotional learning would be an insight into seeing if older siblings with ASD even have a significant effect.

LIMITATIONS

One key limitation in this study is the representation of the sample. The current sample population of this study is 69 participants, in which 53 of them were white/Caucasian and 59 participants lived in households where the primary caregiver had a college degree. To further enhance this investigation, research would have to be conducted gaining a more representative sample, studying participants from different cultures and socioeconomic backgrounds. Additionally, the usage of screen-based stimuli in the participant videos could contain some lag when coding the actual videos and searching for gazes that are left and right, which could alter the timings. In these biases, the novelty preferences could have been altered and might be different if these biases were accounted for.

BIBLIOGRAPHY

Nate. (2025, January 5). *Family history - does it play a role in autism?*. BlueGems. <https://bluegemsaba.com/family-history/#:~:text=For%20instance%2C%20studies%20have%20suggested,to%20the%20subsequent%20third%20child.> [1]

A., M.-A. M.-M. S.-C. G.-G. (n.d.). *Joint attention and its relationship with autism risk markers at 18 months of age*. Children (Basel, Switzerland). <https://pubmed.ncbi.nlm.nih.gov/35455600/> [2]

Author links open overlay panelLauren M. Smith a, a, b, Highlights•Multimodal pathways used to achieve joint attention (JA) have yet to be studied in infants with a familial history of ASD. •The study explores behaviors via Zoom in infants with a familial history of ASD (FH+), & Abstract-Joint attention (JA) is an early-developing behavior that allows caregivers and infants to share focus on an object. Deficits in JA. (2023, November 14). *Multimodal Pathways to joint attention in infants with a familial history of autism*. Developmental Cognitive Neuroscience. <https://www.sciencedirect.com/science/article/pii/S1878929323001305> [3]

Kasari, C., & Patterson, S. (2012, December). *Interventions addressing social impairment in autism*. Current psychiatry reports. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3492515/> [4]

Magrelli, S., Jermann, P., Noris, B., Ansermet, F., Hentsch, F., Nadel, J., & Billard, A. (2013). *Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings*. Frontiers in psychology, 4, 840. <https://doi.org/10.3389/fpsyg.2013.00840>

Sex/gender differences in symptomology and camouflaging ... (n.d.-b). https://med.stanford.edu/content/dam/sm/funglab/documents/Fung_Lab_papers/Schuck-2019---BriefReportSexGenderDifference-with-Suppl.pdf

Uclahealth. (2023, October 12). *Understanding undiagnosed autism in adult females*. UCLA Health. <https://www.uclahealth.org/news/article/understanding-undiagnosed-autism-adult-females>

BIOMEDICAL SCIENCE

CELLULAR BIOLOGY

An Evaluation of Hypomethylation in Primary Prostate Cancer

Alexander Helfman

ABSTRACT

It has been estimated that 1 in 8 men will be diagnosed with prostate cancer in their lifetime, changes in DNA methylation (regulation of gene expression) are associated with the behavior of cancer cells. How does hypomethylation (down regulation of DNA methylation) impact primary prostate cancer? This project investigates what the hypomethylation of prostate cancer effects within primary prostate cancer. The impact of hypomethylation in primary prostate cancer was identified by staining cancer tissue slides using immunohistochemistry, a method of staining, and the staining intensity was quantified to distinguish between normal methylation (50-300 on the staining score) and hypomethylation (0-50 on the staining score). These results were statistically analyzed using a Kaplan-Meier curve to identify whether hypomethylation or normal methylation would impact biochemical recurrence (a recurrence in a protein made by the prostate). The hypothesis was that primary prostate cancer patients who exhibit hypomethylation will exhibit a lower recurrence- free survival rate compared to those with normal levels of methylation. The results support the hypothesis by showing that individuals that have normal methylation levels in primary prostate cancer have a hazard ratio of 0.62 in contrast to those with hypomethylation having a hazard ratio of 1. A, P value of less than 0.001 was found between hypomethylated and individuals with normal methylation, these findings could allow for more effective or aggressive treatment and closer monitoring for individuals who express hypomethylation, as they have a greater risk of biochemical recurrence in contrast to those that have normal levels of methylation..

INTRODUCTION

This project evaluates the correlation between metastatic sites and methylation levels, ranging from hypermethylation to hypomethylation. It has been estimated that roughly thirteen percent of men will get prostate cancer in their lifetime, and approximately two-thirds or sixty-six percent, of those men will die from prostate cancer (CDC, et al) Cancer is one of the most prevalent diseases in the United States, with prostate cancer being one of the most prominent out of this pool of cancers. Cancer is the 2nd leading cause of death in the United States, only behind heart disease (IDPH, et al.). Prostate cancer is an ever-growing problem, it is key to understanding the process and mechanism of prostate cancer and its effects.

Hypermethylation, as well as hypomethylation, is an epigenetic modification. Epigenetic modifications have been shown to be not favorable for cancer patients, "it became clear that the dynamic regulation of DNA methylation is a critical epigenetic mechanism of cancer initiation, maintenance, and progression" (Lakshminarasimhan, et al.), emphasizing the importance of understanding epigenetic alterations within cancer. According to the article "Cause and Consequences of Genetic and Epigenetic Alterations in Human Cancer Variations," epigenetic modifications,

such as DNA methylation, have been shown to distinguish between benign and malignant tumors in various cancers, including prostate cancer" (Shin et al.), being able to identify the differences between what is considered to be a "safe" noninvasive tumor in contrast to a dangerous tumor that can spread and mutate.

Tumor suppressor genes are key to preventing cancer, hypomethylation or hypermethylation can lead to a loss of these tumor suppressor genes, which can increase the likelihood of a mutation and poor prognosis for example, in the article "Cause and consequences of genetic and epigenetic alteration in human cancer, it was found that for many cancer-related tumor suppressor genes, including p53, an important epigenetic regulator in methylation within genes, hypermethylation would lead to this gene being mutated, an allelic loss, a loss in an allele that can lead to an inactivation of a tumor suppressor gene, as well as LOH, or loss of heterozygosity, which can lead to tumor development, can occur due to hypermethylation.

Methylation isn't exactly a rare phenomenon within prostate cancer, the GSTP1 gene occurs in over 90 percent of cancerous samples" (Majumdar, et al.).

The prevalence of methylation within prostate cancer

shows how serious of an issue it has become as the years go by. While research has been done on prostate cancer and methylation, current research has failed to identify patterns of methylation between primary sites vs. metastasized regions of prostate cancer. Do different regions of metastasized prostate cancers have differing levels of methylation, such as hypermethylation or hypomethylation? Due to differences in secreted factors of different organs and regions, there will be a difference in methylation in these varying regions of the body. To better treat both localized and metastatic prostate cancer, it is important to understand trends within methylation. One way of doing this is to identify if there are patterns in methylation, utilizing immunohistochemical scores to quantify methylation levels within prostate cancer regions to identify patterns in prostate cancer.

Materials List:

Material List:	
Premiere slide warmer	XH-2005 or alternate slide warmer
Slide holder	As many as needed for number of slides
TBST	600 ml total (200 ml per wash bucket, 3 buckets)
Tissue Tek 2 manual slide staining set	For manual staining of slides
JMP fume hood isolator	Model 3 from James Town Products
Xylene	600 ml total (200 ml per dunk bucket, 3 buckets)
100% ethanol	400 ml total (200 ml per dunk bucket, 2 buckets)
95% ethanol	200 ml total (1 wash bucket)
80% ethanol	200 ml total (1 wash bucket)
70% ethanol	200 ml total (1 wash bucket)
Wash buckets	At least 12, 200 ml each
Black and Decker Flavor Scenter Steam Plus	Enough as needed, should only need one.
Distilled water	800 ml for Black and Decker boiler
Antigen buffer	50 ml or less
Nitrile gloves	As many as needed
Ethanol for cleaning nitrile gloves	As many as needed
Computer/device for QuPath	Must be able to run QuPath
Primary antibody	From Cell Signaling, scale as needed
Secondary antibody	From Power Vision anti-rabbit, scale as needed
Hybez II Oven	Biotechne brand, enough as needed.
Primary antibody dilution	1/50 dilution for 5 mc primary antibody
Primary antibody dilution buffer	50 microliters
Primary dilution per slide	70 microliters, scale as needed
HCl (3.5 Normal)	50-70 ml
Glass container for HCl	Allows slides to stand up
Dual endogenous enzyme block	Enough to cover all slides (from Dako)
Micropipettes	1000, 200, and 20 microliters
Pipet tubes	Determine based on pipet size, amount as many as needed.
Kim wipes	As many as needed

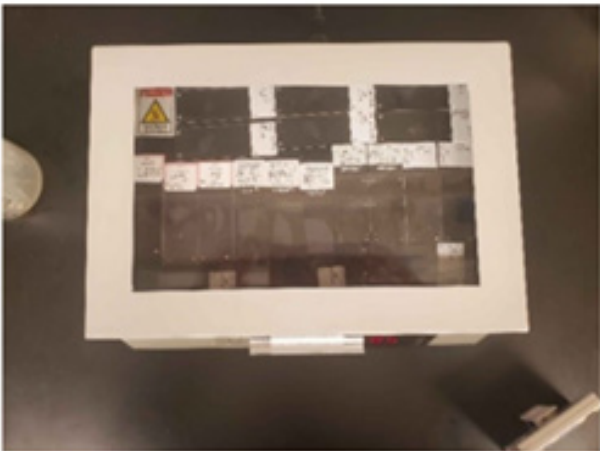
Distilled water for rinsing slides	As much as needed
Dab Tablets (Gold and Silver)	From Sigma
Syringe with filter	As many as needed
Hematoxylin	From Dako
Ventana DP 200	From Roche, only one would be needed.
Slide holder container	With a lid
Antigen retrieval buffer	Around 50 ml
Slide tray	As many as needed
Storage tube for water	At least 5 ml capacity
Hematoxylin	As much as needed.
Regular washing container	For cleaning with distilled water, as many as needed.
Mounting media	As much as needed
Slide baker	As many as needed
Safety glasses	As many as needed
Lab coat for HCl safety	As many as needed

Experimental setup

Before starting experimentation, wear personal protective equipment (lab coat, gloves, and safety glasses) while performing experimentation.

All slides/tissue samples were provided from the Fred Hutchinson Cancer Center and the University of Washington rapid autopsy program. Have an area prepared for immunohistochemistry at the Fred Hutchinson Cancer Center, after acquiring metastatic cancer tissue slides, prepare immunohistochemistry from the Fred Hutchinson Cancer Center. Slides will be from Fred Hutchinson Cancer Center, as well as slide data, given to have more data. Some When on 95 percent ethanol, begin to boil water, it should take around 10 minutes to boil in a Black & Decker Flavor Center with around 800 ml of distilled water. While boiling water, transfer the slides back to the TBST to clean, going from left to right, put the slides in the first bucket for 1 minute and then the other two for 30 seconds. Once done, put them in the antigen retrieval buffer in a slide holder container with a cap, and place the lid on top so that the antigen buffer does not evaporate. the container so that the

Figure 1: Premiere Slide Warmer XH-2005



Baking Slides utilize the premiere slide warmer, if there are other slide warmers available, those would work as well.

While the slides are heating, prepare wash buckets for immunohistochemistry (which will be called IHC from here on). Use 3 wash buckets and fill each with 200 ml of TBST or tris-buffered saline with Tween 20, each for around 1 minute for the first bucket and 30 seconds for the remaining two. After we are done, TBST slides will be going from left to right in the rehydration process. The slides will go through a gradient of ethanol and xylene. From left to right, the first 3 buckets will be xylene, then 2 buckets of 100 percent ethanol, and 1 of 95, 80, and 70 percent ethanol for 3 buckets. Dunk the slides using the slide holder in each bucket for 3 minutes per bucket. Figure 2 is the order of compounds.

Figure 2: Compounds for IHC in order



Compounds for IHC to put it in this specific order

When on 95 percent ethanol, begin to boil water, it should take around 10 minutes to boil in a Black & Decker Flavor Center with around 800 ml of distilled water. While boiling water, transfer the slides back to the TBST to clean, going from left to right, put the slides in the first bucket for 1 minute and then the other two for 30 seconds. Once done, put them in the antigen retrieval buffer in a slide holder container with a cap, and place the lid on top so that the antigen buffer does not evaporate. the container so that the buffer does not evaporate, this is especially important. Set the Hybezen oven to 37 Celsius. Once the steaming is done, let the slides cool for around two minutes. Fill a glass container with around 50-70 ml of HCL depending on the size of the slides, after steaming, put slides in 3.5 normal HCL for 20 minutes. Once cooled down and processed, put the slides in the 3 TBST as before, going from left to right, 1 minute for the first and 30 seconds for the second and third. Put all slides on a slide tray and cover them all with a dual endogenous enzyme block, this should be enough to cover the tissue. Do not let the slides dry out too much, cover the slides with the dual endogenous enzyme block utilizing a micropipette. Calculate the amount of buffer solution needed, it will be 70 microliters times the number of slides. After creating the antibody solution, using the math done in step 14 is not required, but it is okay to store the solution in a bucket of ice if using the bucket of ice.

Put the slides back in TBST containers as before, going from left to right, 1 minute for the first and 30 seconds for the following two. It is okay to leave the slides in TBST for up to an hour if needed. Wipe around the tissue slides using Kim wipes to remove excess liquid, do not touch the tissue itself. After drying, put the primary antibody directly in the buffer solution on the tissue using a micropipette, make sure not to wait too long, as the tissue can dry out. Remember to apply enough primary antibodies to cover the slide. Once done with all the slides, put them into the Hybezen oven for a minimum of 45 minutes and put the slides in a slide tray with enough water to just fill the bottom. Immediately after the slides are done, place the slides in TBST as before, but instead of from left to right, it will be three minutes for each bucket, totaling 9 minutes. Once the slides are done soaking in TBST, apply the secondary antibody and run it

for 5 minutes to 30 minutes. Remember to wipe the excess liquid around the tissue but not the tissue itself. After The slides have been in the oven, get the tube and add 5 ml of distilled water to it. After adding both Gold and Silver Dab tablets, wait until they are fully dissolved in the water, this can take around 10-12 minutes. Once tablets are mixed into the water, apply them to the tissue slides using a syringe with a syringe filter and apply as much to cover the whole tissue. Wait 10 minutes for the staining to finish. After 10 minutes, soak the slides in TBST like before for three minutes each, taking 9 minutes in total. After the TBST wash, apply enough hematoxylin (a dye used to stain cells) to cover each tissue fully and wait one minute until it is finished.

After applying hematoxylin, put the slides in a regular washing container, then cover the top using your hand. Now pour distilled water over your hand, this should allow the distilled water to seep into the container but not directly hit the slides. Remove water from the washing container and repeat this 2-3 times. After rinsing slides, do the reverse of step 3, instead of 70% ethanol being the last and first, it will be the reverse, so instead of left to right, go right to left. In short, go from 70%-80%-95%-100%-100%-xylene-xylene-xylene. And instead of 3 minutes per wash, simply dip the slides 10 times. Once slides have been processed, wipe slides using Kim wipes but only on the back side, after applying mounting media on the tissue, this should be in the range of 1-3 slides depending on the tissue size. Put the cover slip on top of where the mounting media was when putting slides in the scanner.

Figure 3: Image of xenograft slides

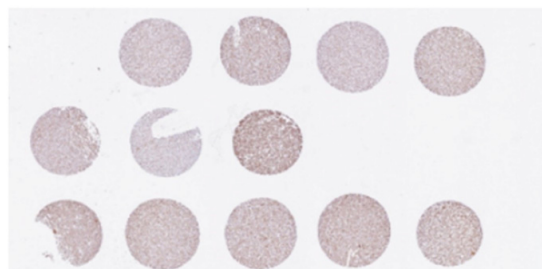


Image of xenograft slides under scanner, which were stained utilizing IHC (immunohistochemistry)

After slides have been processed, Scan the tissue slides using the Ventana DP200 system to generate high-resolution brightfield whole slide images. Upload tissue slide images into QuPath imaging software. This is open-source and can be used by anyone. Looking at each tissue, select areas that are stained with 5 mc. These slides will be stained, use Qu Path and select areas that seem to have 5 mc to get a rough estimate of the concentration. Utilize Qu Path's automated H score calculation too to get an H score by running the Analyze > Cell detection > Positive cell detection command. Figure 3 is an example of Qu Path analysis. Due to the sake of being sensitive data, it was

not possible to directly look at data, therefore, data was processed from the Fred Hutchinson Cancer Center/Canary cohort. Additional data was received, and data from IHC was processed by the Fred Hutchinson Cancer Center. Put data into data table 1.

Data Table 1:			
Characteristic	HR	95% CI	p-value
Min H- Score			
5-Mc Low [0,50]	-	-	-
5-Mc High (50,300]	0.62	0.47,0.81	<0.001
HR= Hazard Ratio, CI= Confidence Interval			

Conclusions/Data Analysis

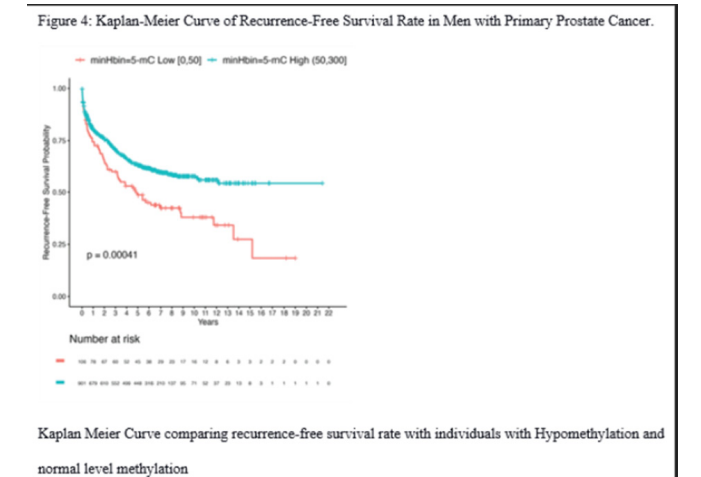


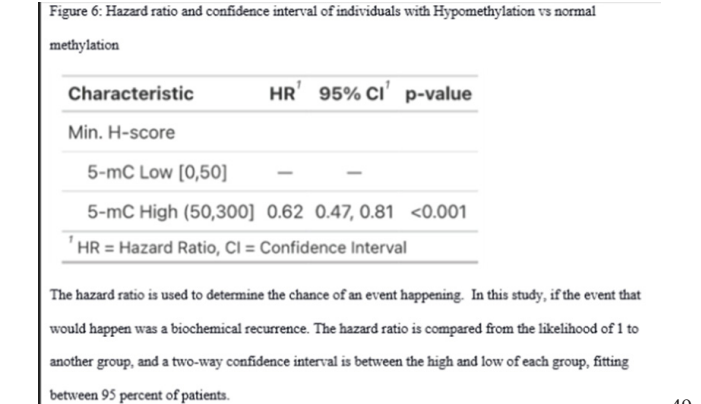
Figure 5: Additional data received from the study

Characteristic	0-mc Low [0,50]	0-mc High (50,300]	p-value [†]
Gleason Score			
≤6	39 (26%)	411 (44%)	
3-4	42 (28%)	250 (27%)	
4-5	12 (7%)	130 (14%)	
≥6	12 (8%)	130 (14%)	
Seminal vesicle invasion	2 (1%)	60 (6%)	0.033
Unknown	7	9	
Mean Age (years)	63.68	61.13	0.005
Unknown	12	102	
Extracapsular Extension	44 (29%)	219 (24%)	0.076
Unknown	9	9	
Positive Surgical Margin	32 (21%)	214 (23%)	<0.001
Unknown	10	102	
Stage			0.6
T2	60 (40%)	497 (53%)	
T3/T4	34 (22%)	252 (27%)	
Unknown	13	102	
Cholesterol > 0.25mm	10 (7%)	105 (11%)	0.2
Unknown	9	9	
Pre-operative PSA (ng/mL)	8.95	8.04	0.7
Unknown	10	102	

[†] p < 0.05, Wilcoxon

[‡] Pearson's Chi-squared test; Wilcoxon rank-sum test

Additional data was received from the mentor/Fred Hutchinson Cancer Center having values for each parameter such as seminal vesical invasion, mean age, Gleason score, and other variables.



DISCUSSION/DATA ANALYSIS

It is shown that only three values are statistically significant: mean age, Seminal vesicle invasion (cancer spreading to the seminal vesicle but not enough has spread to be considered metastasis), and Extrascapular extension (when the cancer spreads lymph nodes) had statistically significant P values. Mean age was statistically significant, which might indicate those who were caught with prostate cancer later showed signs of hypomethylation.

Recurrence survival rate is defined by the recurrence of PSA, or prostate-specific antigen. This was important during this study as all men in this study had a prostatectomy, which is when their prostate is removed, if the prostate was removed, PSA is not expressed, but recurrence of PSA can indicate that the cancer was not fully removed, or it had metastasized or spread throughout the body. In this case, it is shown that individuals who have primary prostate cancer who express hypomethylation have a lower recurrence-free survival, which means their chance of surviving without cancer is lower in contrast to those with normal levels of methylation. With a P value of 0.00041, this shows a strong correlation between methylation patterns and recurrence-free survival rate.

Looking at the data, it is observed that individuals who have normal methylation categorized by (50, 300) on a staining score vs. hypomethylation (0, 50) on a staining score, it is shown that individuals that express normal methylation have a lower hazard ratio at a hazard ratio of 0.62, which means they have a lesser chance of biochemical recurrence of their cancer reemerging or showing signs that the cancer is still present, while the confidence interval, which is the range, was still lower than those with hypomethylation. With a P value of less than 0.001, this shows that individuals who have normal methylation have a lesser chance of their cancer metastasizing or spreading across the body, this proves the hypothesis to be correct as individuals who showed the hypomethylated trait had a decrease in recurrence-free survival rate.

APPLICATION

If a patient who is diagnosed with primary prostate cancer is found to also have hypomethylation, more aggressive treatment options or more specific treatment options might be used. Where the standard method for individuals is androgen deprivation therapy or chemical castration, doctors might choose to use radiation at the prostate bed to eliminate any remaining cancer tissue, closer monitoring would also be used, as the patients who have hypomethylation have a much greater chance of their cancer recurring in contrast to those that don't have hypomethylation. This will allow both patients and doctors to make better-informed decisions regarding the treatment and monitoring of patients.

FUTURE WORK

It has been identified that individuals who express hypomethylation have a greater chance of biochemical recurrence, still, many questions remain unanswered. While attempting to perform the project, the original idea was to see if methylation levels varied across different regions of the body in individuals who had metastatic prostate cancer, but it was hard to perform due to the time of approval and insufficient slides, as all slides received were from primary prostate cancer patients. Seeing if methylation patterns only show up in primary prostate cancer or can occur in other cancers is also a critical question, as the antibody used, 5Mc, is a universal method of staining and can be translated to other cancers. Lastly, it would be interesting to see if the trait of hypomethylation is genetically heritable. If so, it would allow for potential early diagnosis or better treatment of patients.

ACKNOWLEDGMENT

Special thanks to Professor Michael Haffner, for guiding me in understanding the various datasets and supporting my learning. I also extend my gratitude to the Haffner lab for allowing me to conduct immunohistochemical staining under their supervision and for teaching me how to utilize QuPath to expand my knowledge in the process.

BIBLIOGRAPHY

Ajkunic, A., Sayar, E., Roudier, M. P., Patel, R. A., Coleman, I. M., De Sarkar, N., Hanratty, B., Adil, M., Zhao, J., Zaidi, S., True, L. D., Sperger, J. M., Cheng, H. H., Yu, E. Y., Montgomery, R. B., Hawley, J. E., Ha, G., Persse, T., Galipeau, P., ... Haffner, M. C. (2024, May 17). Assessment of *trop2*, *CEACAM5* and *DLL3* in metastatic prostate cancer: Expression landscape and molecular correlates. *NPJ precision oncology*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11101486/#Abs1>

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., and Hamilton, P. W. (2017, December 4). QuPath: Open-source software for digital pathology image analysis. *Nature News*. <https://www.nature.com/articles/s41598-017-17204-5>

Chandra, D., Tyagi, S., Singh, J., Deka, R., Manivannan, P., Mishra, P., Pati, H. P., and Saxena, R. (2017, December 29). Utility of 5-methylcytosine immunohistochemical staining to assess global DNA methylation and its prognostic impact in MDS Patients. *Asian Pacific journal of cancer prevention : APJCP*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5980888/>

Centers for Disease Control, C. for disease. (2024, August 27). Prostate cancer risk factors. Centers for Disease Control and Prevention. <https://www.cdc.gov/prostate-cancer/risk-factors/index.html>

Lakshminarasimhan, R., and Liang, G. (2016). The role of DNA methylation in cancer. *Advances in experimental medicine and biology*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7409375/>

Meervassana, J., Varophas, S., Prabsattru, P., Kamolratanakul, S., Ruangritchankul, K., and Kitkumthorn, N. (2024, March 30). 5-methylcytosine immunohistochemistry for predicting cutaneous melanoma prognosis. *Nature News*. <https://www.nature.com/articles/s41598-024-58011-z>

Ram, S., Vizcarra, P., Whalen, P., Deng, S., Painter, C. L., Jackson-Fisher, A., Pirie-Shepherd, S., Xia, X., and Powell, E. L. (2021, September 27). Pixelwise h-Score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images. *PloS one*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8475990/>

Singh, R. K., Diaz, P. E., Binette, F., and Nasonkin, I. O. (2018, August 29). Immunohistochemical detection of 5-methylcytosine and 5-hydroxymethylcytosine in developing and postmitotic mouse retina. *Journal of visualized experiments: JoVE*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6235063/>

Phillips, T. (2008). The Role of Methylation in Gene Expression. *Nature news*. <https://www.nature.com/scitable/topicpage/the-role-of-methylation-in-gene-expression-1070/>

Antibacterial Activity of Stinging Nettle, Devil's Club, and Oregon Grape Against *Escherichia Coli* and *Bacillus Cereus*

Basudha Ghose

ABSTRACT

Novel antibiotics are needed to combat the growing problem of antibacterial resistance, and plant extracts may be a possible solution. This project seeks to investigate the antibacterial activity of stinging nettle, Oregon grape, and Devil's club extracts against *Bacillus cereus* and *Escherichia coli* as a possible alternative to traditional antibiotics. The hypothesis was that the extract that combines all three plants will cause the largest inhibition zones in comparison to the antibiotic controls, which included gentamicin and chloramphenicol. The hypothesis was tested through a disk diffusion assay, which included soaking paper disks with different plant extracts and antibiotics (independent variable) and putting them on agar swabbed with bacteria. The diameter of the inhibition zones (dependent variable) from the disks was recorded. The hypothesis was not supported as the antibiotic disks on the *E. coli* had larger inhibition zones than the combined plant extract (34.7 mm vs. 0 mm), and these differences were statistically significant with a p value of 2.56×10^{-8} . The antibiotic disks on the *B. cereus* had larger inhibition zones than the combined plant extract (36 mm vs. 11 mm), and these differences were statistically significant with a p-value of 3.39×10^{-6} . While the hypothesis was not supported, all of the extracts showed some sort of inhibition against one or both types of the bacteria, proving that with further testing with higher concentrations of the extracts, these extracts could be possible antibacterial agents.

INTRODUCTION

This project seeks to investigate the antibacterial properties of Devil's club, stinging nettle, and Oregon grape extracts against *Bacillus cereus* and *Escherichia coli* as novel antibiotics are needed to combat antibacterial resistance. In 2019, antibacterial resistance caused 1.27 million deaths and contributed to 4.95 million deaths worldwide (1). The U.S. alone has more than 2.8 million antibacterial resistant infections each year (2). Antibacterial resistance does not just make infections harder to treat, it also makes treatments and procedures like surgery, caesarean sections, and chemotherapy riskier (1). One of the root causes of antibacterial resistance is the overuse and misuse of antibiotics. As Chua (3) mentions, out of 19.2 million antibiotic prescriptions for children and adults in the U.S., 12.8% were appropriate, 23.2% were not medically justified, 35.5% were potentially appropriate, and 28.5% were not associated with any documented diagnosis. Additionally, Chua (3) also found that overall, one in seven patients received unnecessary antibiotics and 2.2 million adult enrollees filled at least one inappropriate antibiotic prescription in 2016. The overuse and misuse of antibiotics leads to antibacterial resistance because while antibiotics do kill off susceptible bacteria, they promote the growth of bacteria that have genetic mutations that allow them to survive against antibiotics. The more antibiotics are used, the more opportunities bacteria have to evolve and develop

resistance against antibiotics.

To address this problem, novel antibiotics must be found. Bacteria are constantly evolving to resist antibiotics, and the overuse and misuse of traditional antibiotics is only accelerating this problem. Novel antibiotics are needed to treat infections caused by drug resistant bacteria and to prevent bacteria from becoming drug resistant in the first place. Research shows that plants could be a source of compounds that may overcome antibacterial resistance. Some medicinal plants may overcome antibacterial resistance due to the presence of certain compounds like tannins, terpenoids, alkaloids, and flavonoids. These compounds can overcome antibacterial resistance because they are able to attack various bacterial functions. In addition, these compounds have complex mechanisms that make it harder for bacteria to develop resistance to antibiotics. According to Vaou et al. (4), these plant compounds can inhibit efflux pumps, inhibit the synthesis of DNA and proteins, inhibit ATP synthetase, damage the cell membrane, and destroy the membrane proteins of bacteria. These mechanisms are some of the ways that antibacterial resistance starts in bacteria. Additionally, Vaou et al. (4) also states that these compounds can act synergistically in plant extracts, furthering their antibacterial effectiveness. However, most plants have not been tested for possible antibacterial activity as Vaou et al. (4) points out that only 1340 plants have defined antibacterial activity while there are 28,187 medicinal plant

species used by humans.

Extracts of plants from the Pacific Northwest like Oregon grape (*Mahonia Aquifolium*), Devil's club (*Oplopanax Horridus*), and stinging nettle (*Urtica Dioicia*), have been tested and show possible antibacterial activity (5-7). Research by Doyle (5) shows that Oregon grape contains compounds that are effective against inhibiting methicillin-resistant *Staphylococcus aureus*, which is a strain of *S. aureus* that is resistant to many antibiotics, including penicillin. Oregon grape could be used as a possible antibiotic and also acts synergistically with other herbs and antibiotics, working against drug resistant bacteria. Oregon grape is also able to inhibit efflux pumps. Efflux pump usually pump antibiotics out of a cell, causing antibacterial resistance, but Oregon grape can help antibiotics stay in the cell, combatting antibacterial resistance. A study done by Gendron et al. (6) showed that stinging nettle extract had weak antibacterial activity against *S. aureus* but was able to inhibit the growth of *M. luteus*. Stinging nettle was not able to inhibit the growth of *P. aeruginosa*. While stinging nettle was not able to kill all the bacteria, it showed promise in inhibiting the growth of gram-positive bacteria. Kobaisy et al. (7) isolated two new and three known polyynes from devil's club plants. Polyynes are organic compounds with alternating single and triple bonds. These polyynes exhibited significant antibacterial, anti-Candida, and antimycobacterial activities. Anti-Candida activity means that the polyynes were able to inhibit the growth Candida yeast, and antimycobacterial means that the polyynes were able to inhibit the growth of mycobacteria, which is responsible for diseases like tuberculosis and leprosy. They had large zones of inhibition and were able inhibit many colonies of bacteria. The polyynes were also able to kill *Mycobacterium tuberculosis* and isoniazid-resistant *Mycobacterium avium* in a disk diffusion assay. Isoniazid-resistant *Mycobacterium avium* is a strain of bacteria that is resistant to the antibiotic isoniazid.

As previous studies show, more testing is needed to confirm the antibacterial efficacy of these plants against different types of bacteria in a controlled setting. Additionally, what has not been investigated in these studies is the antibacterial properties of combinations of these extracts. These plants may have an additive or synergistic effect when combined. An additive effect occurs when two or more drugs are combined, and their combined effect is equal to the sum of their individual effects. A synergistic effect when two or more drugs are combined, and their combine effect is greater than the sum of their effect. The potential additive or synergistic effects of these medicinal plants is still unknown, as there is not enough research looking into combinations of these plants and how these plants exactly kill bacteria. In this project, the antibacterial properties of the Oregon grape, devil's club, and stinging nettle will be tested, as well as extracts that combine these

plants. The hypothesis is that the extract that combines all three plants will cause larger inhibition zones than the gentamicin and chloramphenicol because by themselves, these plants already show some antibacterial activity. Extracts that combine these plants will have an additive or synergistic effect and will ultimately be more powerful than traditional antibiotics.

METHODOLOGY

Creating the Plant Disks

The first part of the procedure was creating the plant disks. Oregon grape, stinging nettle, and Devil's club extract were bought from Mountain Rose Herbs. White paper disks were bought from Carolina Biological Supplies. Six white control disks were soaked with 100 micrograms of devil's club extract. The same process was repeated with the stinging nettle and the Oregon grape extract. Then, six white disks were soaked with 50 micrograms of Oregon grape extract and 50 micrograms of stinging nettle extract. This same process was repeated to create the other two combination mixtures of Oregon grape and devil's club and stinging nettle and devil's club. Finally, six disks were soaked with 33 micrograms of devil's club extract, 33 micrograms of stinging nettle extract, and 33 micrograms of Oregon grape extracts. Both sides of the disk were soaked with extract to ensure that there were no issues with the disk diffusion assay. The disks dried for 12 hours before they were used in the assay.

Creating the Agar Plates

The second part of the procedure was creating the agar plates. Eight petri dishes were labeled with the bacteria and extract to be used. The caps of the agar bottles were loosened and placed into the microwave. The bottles warmed for about two minutes, but they were occasionally taken out of the microwave to be swirled and to prevent them from boiling over. The hot bottles were handled with heat-resistant gloves. Once the bottles were fully melted, they were placed on a heat-resistant surface and were left to cool for about 15 minutes. The bottles were considered completely cooled when they reached room temperature and were ready to be picked up with no glove. As the agar cooled, the counter was disinfected using alcohol wipes. The sterile, labeled petri dishes were put out on the counter. After this, one agar bottle was opened, and the mouth of the bottle was disinfected with an alcohol pad. The agar was then poured into the dishes, making sure that the lid of the plates was lifted just enough to pour the plates and that the bottle was kept at an angle. Enough agar was poured to cover the bottom of the plate. The plates were left to gel for 24 hours. *E. coli* and *B. cereus* were chosen for this project as to have one gram-negative bacteria and one gram-positive. Four of the agar plates were streaked with *E. coli* using a sterile cotton swab. The bacteria were streaked across the entire surface of the agar plate (Figure 1).

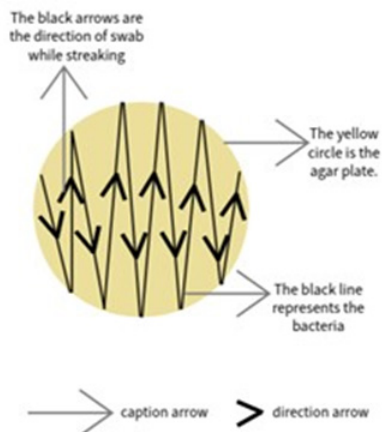


Figure 1 is a model showing the direction in which the agar plates were streaked with bacteria.

The plate was then streaked two more times, rotating the plate 60 degrees each time. At the end, the rim of the agar plate was streaked. The other four plates were streaked with *B. cereus* in the same way.

PERFORMING THE DISK DIFFUSION ASSAY AND RECORDING THE DATA

After the agar plates were created, the disk diffusion assay was performed in a fume hood. Using forceps, three disks of each extract were put on the agar plates. A total of six disks were on each agar plate. Three disks of Devil's club extract and three disks of stinging nettle extract were put on the first *E. coli* plate. Three disks of Oregon grape extract and three disks of combined Oregon grape and Devil's club extract were put on the second *E. coli* plate. Three disks of the combined Oregon grape and stinging nettle extract and three disks combined stinging nettle and Devil's club extract were put on the third *E. coli* plate. Three disks of the extract that combined all three plants and three gentamicin disks were put on the fourth *E. coli* plate. Each disk was approximately 24 mm apart (Figure 2).

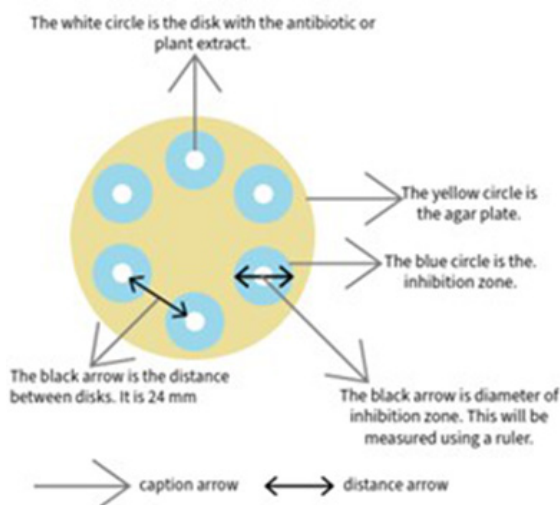


Figure 2 is a model showing how the disks were placed on the agar plates and how their inhibition zones were measured.

This process was repeated for the *B. cereus* plates, except chloramphenicol was used instead of gentamicin. Gentamicin and chloramphenicol are both antibiotics and were used as the control groups in this experiment. These antibiotics were chosen because they are the most potent against their respective bacteria. Once all the disks were on the plates, the plates were inverted and incubated for 24 hours at 30°C in an incubator. After 24 hours, the diameter of the inhibition zones for the *E. coli* dishes were measured in mm using a ruler (Figure 2). The same measurements were taken for the *B. cereus* plates and were recorded in data tables.

DATA ANALYSIS

The mean, median, mode, and standard deviation for each of the plant extracts and antibiotics were calculated using the three disks. To determine if differences in inhibition zones of the plant extracts and the antibiotics were statistically significant, a one-tailed, independent sample t-test was performed using Excel. A one-tailed t-test was used because the hypothesis checks for a difference in a specific direction, as the hypothesis says the plant extract that combines all three extracts will have large inhibition zones than the antibiotics. An independent sample t-test was used because the means of inhibition zones for the plant extracts and antibiotics were unrelated. For the *E. coli* plates, the t-test was performed using the gentamicin control group, and for the *B. cereus* plates, the t-test was performed using the chloramphenicol control group. Excel calculated the p-value for the t-test, and if the p-value was less than or equal to 0.05, then the differences between the plant extract and antibiotic were considered statistically significant.

RESULTS

The hypothesis was proven wrong as it was originally predicted that the extract that combined all three plants would have larger inhibition zones than the antibiotic control groups, but it turned out that the antibiotic control groups had larger inhibition zones, and these differences were statistically significant. Firstly, the extract that combined all three plants had a mean inhibition zone diameter of 0 mm and 11 mm against the *B. cereus* and *E. coli* plates respectively (Figures 3 and 4). The gentamicin disks on the *E. coli* plates had mean inhibition zone diameter of 34.7 mm, while the chloramphenicol as on the *B. cereus* a mean diameter of 36 mm (Figures 3 and 4). The differences between the *E. coli* gentamicin disks and the extract that combined all three plants was statistically significant with a p-value of 2.56×10^{-8} . The differences between the *B. cereus* chloramphenicol disks and the extract that combined all three plants was also statistically significant with a p value of 3.39×10^{-6} .

Secondly, most of the plant extracts had a mean inhibition zone diameter of around 5.3 mm against the *E. coli*. When comparing this to the antibiotics, this was

statistically significant, with a p-value of 2×10^{-7} . Deviating from this trend was the Devil's club which had a mean inhibition zone diameter of 4.7 mm (Figure 3). The total range of inhibition zone diameters was from 4 to 6 mm (Table 1). For the *E. coli* plates, all of the plant extracts showed some sort of weak inhibition except for the extract that combined all three plants, which may be due to the fact that *E. coli* is gram-negative and has an outer membrane unlike *B. cereus*, which is gram-positive and does not have this outer membrane.

Table 1 shows the inhibition zone diameter that were measured for the plant extracts and antibiotics for the *E. coli* plates.

Plant Extract/Antibiotic	Diameter of Inhibition Zone		
	Disk 1	Disk 2	Disk 3
Devil's club	4	5	5
Oregon grape	5	5	6
Stinging nettle	5	5	6
Devil's club + Oregon grape	7	5	5
Stinging nettle + Devil's club	5	6	5
Oregon grape + Stinging nettle	6	5	5
Stinging nettle + Devil's club + Oregon grape	0	0	0
Gentamicin	35	34	35

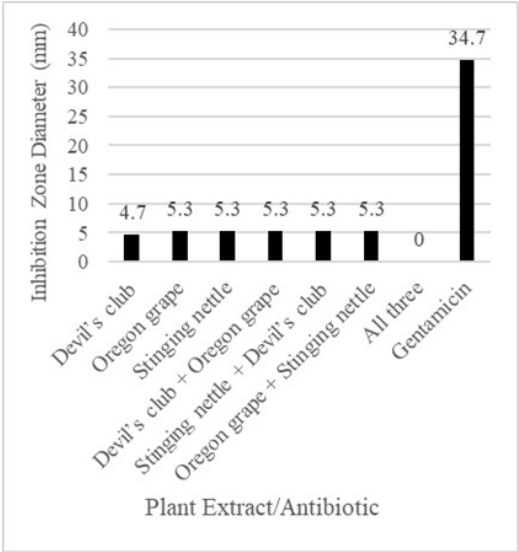


Figure 3 shows the average inhibition zone diameters for all the plant extracts and antibiotics against *E. coli*.

For the *B. cereus* plates, all the plant extracts showed some sort of inhibition, with zones ranging from 5 mm to 18 mm (Table 2). The largest mean inhibition zone diameter was shown by the stinging nettle, which was 17 mm (Figure 4). This was followed by the stinging nettle and

Devil's club combined extract which had a mean inhibition zone diameter of 15.3 mm (Figure 4). Also, the Devil's club and Oregon grape combined extract had mean inhibition zone diameter of 15 mm (Figure 4). These differences were statistically significant, with p-values equal to 1.01×10^{-5} , 3.23×10^{-6} , and 6.79×10^{-6} , respectively.

Table 2 shows the inhibition zone diameter that were measured for the plant extracts and antibiotics for the *B. cereus* plates.

Plant Extract/Antibiotic	Diameter of Inhibition Zone		
	Disk 1	Disk 2	Disk 3
Devil's club	12	10	9
Oregon grape	5	6	5
Stinging nettle	16	17	18
Devil's club + Oregon grape	15	16	14
Stinging nettle + Devil's club	15	16	15
Oregon grape + Stinging nettle	10	9	8
Stinging nettle + Devil's club + Oregon grape	12	10	11
Chloramphenicol	35	37	36

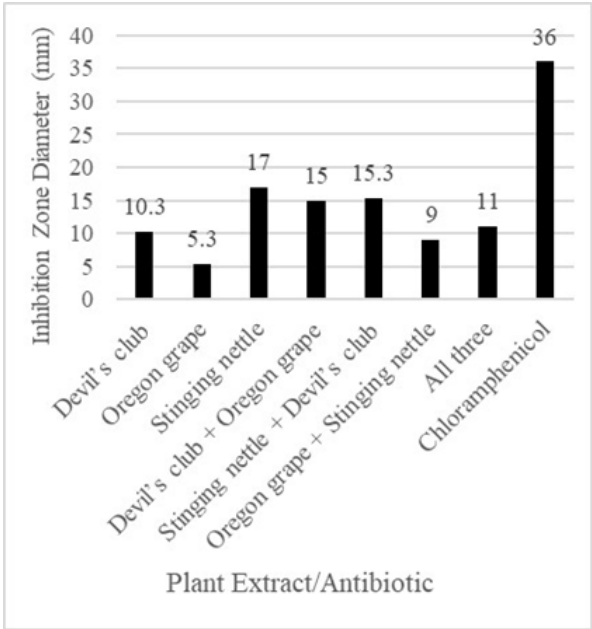


Figure 4 shows the average inhibition zone diameters for all the plant extracts and antibiotics against *B. cereus*.

CONCLUSION

While the hypothesis that the extract that combined all three plants would have a higher inhibition zone than the antibiotics was proven wrong, this study shows that there may be some possible use for plant extracts as antibacterial agents. The objective of this project was to investigate the antibacterial properties of stinging nettle, Devil's club, and Oregon, and all of the extracts showed some sort of inhibi-

tion against the *B. cereus* plates. Similarly, all the extracts, except for the extract that combined all three plates, also showed some sort of inhibition against the *E. coli* plates. This means that in the future, testing at higher concentrations, these extracts could have larger inhibition zones than certain antibiotics, making them possible replacements for medicine, saving millions of people.

Discrepancies in results may have been caused by a lack of trials. There was only enough time to do one trial, not the three that were originally planned. Doing more trials would have created a large sample size and would ensure that plant extracts that did work were reliable and could kill the bacteria multiple times. Additionally, doing multiple trials could have shown that a plant extract that did not show any inhibition in the first trial could have later shown inhibition in another trial. Another problem with the experiment was the bacteria used. The bacteria on the *E. coli* plates did not grow as much compared to the *B. cereus*. Clear *E. coli* colonies were not seen, resulting in the inhibition zones on the *E. coli* being hard to see and difficult to measure. This may have been a result of the incubator being set to the 30°C, which is the ideal temperature for *B. cereus*. The ideal temperature for *E. coli* is 37°C, but there was only one incubator present, so to prevent the *B. cereus* from dying, the incubator was set at 30°C. The low temperature of the incubator may have affected the growth of the *E. coli*, so in the future, more than one incubator will be used to support the growth of different bacteria. A third issue was not being as precise as needed to be during the disk diffusion assay. While performing the disk diffusion assay, some of the plant extract disks were placed in one area of the plate before being moved to another. This made it so some extra areas of bacteria were killed, making the inhibition zones imperfect circles that were more difficult to measure. Additionally, since some of the inhibition zones were so big, it was hard to measure them properly because they ended up overlapping. The hypothesis was that the combined extracts would have larger inhibition zones than the antibiotics, and this did not end up happening. This may be due to the fact that the combined extracts have smaller concentration of each plant extract, making each plant's antibacterial activity less potent. Making the concentrations bigger may solve this problem.

In the future, more trials will be conducted to enhance the reliability and credibility of the inhibition zones of the plant extracts. Additionally, the plant extracts will be tested at higher concentrations to see if at higher concentrations, if the plant extracts will have larger inhibition zones than the antibiotic control groups. Further testing could also include testing with different types of gram-positive and gram-negative bacteria as these plant extracts may be more powerful against other types of bacteria found in the human body. Testing out these plant extracts with bacteria that is resistant to antibiotics would be another future test as it

would show if these plant extracts could be used as an alternative to antibiotics. Long term studies would be needed to see if bacteria would become resistant to the plant extracts like they do to antibiotics. Another future step would be to research the antibacterial mechanism behind the plant extracts, either using high power microscopes or different assays. This could be used to figure out if the combined plant extracts have a synergistic or additive effect. Lastly, making the plant extracts from scratch would be another future step. The plant extracts were bought from Mountain Rose Herbs and were made with organic alcohol and water. Furthermore, the company made these extracts using a 1:3 ratio, meaning that for one part of herb was extracted with three parts of solvent. Making the extracts from scratch could involve using more of the fresh plant leaves in the extract, and using a different solvent, which could possibly change the compounds in the extracts and cause the plants to show increased antibacterial activity.

The impact of this study is that Devil's club, stinging nettle, and Oregon grape could all be possibly used as future antibacterial agents. While future research is needed to see if these plants are safe for the human body and can actually kill antibiotic-resistant bacteria without developing resistant, this study showed promising results into the discovery of new, natural antibiotics. Millions of lives could be saved because of the creation of novel antibiotics through these different plants. Additionally, the antibacterial activity of these plants can be applied to other areas, as these plant extracts could be used as disinfectants when cleaning dirty surfaces with bacteria on them. Ultimately, this research is a positive step in the future for new antibiotic drugs that can solve the problem of antibiotic resistance.

BIBLIOGRAPHY

1. World Health Organization. (2023, November 23). *Antimicrobial resistance*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>
2. Centers for Disease Control and Prevention. (2024, June 6). *2019 Antibiotic Resistance Threats Report*. CDC. <https://www.cdc.gov/antimicrobial-resistance/data-research/threats/index.html>
3. Chua, K. (2019). *Appropriateness of outpatient antibiotic prescribing among privately insured US patients: ICD-10-CM based cross sectional study*. *British Medical Journal*. (8183) <https://www.bmj.com/content/364/bmj.k5092>
4. Vaou, N., Stavropoulou, E., Voidarou, C., Tsigalou, C., Bezirtzoglou, E. (2019). *Towards Advances in Medicinal Plant Antimicrobial Activity: A Review Study on Challenges and Future Perspectives*. *Microorganisms*, 9(10). <https://pmc.ncbi.nlm.nih.gov/articles/PMC8541629/>
5. Doyle, T. (2016). *Mahonia aquifolia, commonly known as Oregon Grape, is an effective antibacterial, including treatment of Methicillin-resistant Staphylococcus aureus (MRSA)*. *American College of Herbal Medicine*. <https://www.acohm.org/wp-content/uploads/2016/09/Mahonia-aquifolia.pdf>

Presl (Nootka Rose) and *Urtica dioica* L. (Stinging Nettle) Extracts. *American Journal of Plant Sciences*, 12(5). <https://www.scirp.org/journal/paperinformation?paperid=109335>

7. Kobaisy, M., Abramowski, Z., Lermer, L. et al. (1997). Antimycobacterial Polyynes of Devil's Club (*Oplopanax horridus*), a North American Native Medicinal Plant. *Journal of Natural Products*, (60). <https://www.cmdr.ubc.ca/bobh/wp-content/uploads/2016/10/185.-Kobaisy-1997.pdf>

Decoding the Genetic Blueprint: Polygenic Risk Scores for Cancer in Diverse Genome Data

Liya Sebastian

ABSTRACT

Colorectal cancer (CRC), the second leading cause of cancer-related death, has an increasing incidence rate of early-onset cases. Understanding genetic risk factors of CRC through polygenic risk scores (PRS) can improve screening methods and reduce this rate, especially with a lack of equitable screening and accurate risk prediction tools for non-European populations. This study seeks to investigate how the distribution of CRC PRS differs across ancestral groups and gender in diverse genome datasets and what insights this can provide into CRC genetic susceptibility. To accomplish this, large genome-wide datasets with different ancestral groups (European, African, East Asian, South Asian, and Latino) are analyzed by matching genetic variants associated with CRC risk with those in the genomic sample dataset. The PRS is the sum of the variant genotype, the number of risk alleles an individual has for that variant, multiplied by the effect size, or quantified risk to CRC of known genetic variants. PRS represents the cumulative effect of genetic variants on an individual's susceptibility to CRC. A one-way ANOVA revealed that there was a statistically significant difference in PRS between at least two ancestral groups ($F(4, 2499) = 25.91, p \approx 0$). Tukey's HSD test found the mean East Asian PRS was statistically significantly different compared to every other group ($p\text{-values} < 0.05$). The differences in PRS distribution across gender were not significant in any ancestral group ($p\text{-values} > 0.05$). This study contributes to understanding the genetic risk of CRC and emphasizes the need to incorporate diverse genetic data into risk prediction for more precise screening.

INTRODUCTION

By analyzing the distribution of Polygenic Risk Scores (PRS) of diverse genomic data from the 1000 Genomes Project across various ancestral groups, this seeks to explore genetic differences across ancestry and gender that influence genetic susceptibility and risk of colorectal cancer (CRC) in order to improve risk prediction models. The increasing incidence of early-onset CRC highlights the weaknesses of current screening guidelines, stressing the use of improved risk prediction models that utilize diverse genetic data.

Background and Current Methods of Detection

As the second leading cause of cancer death in the United States, colorectal cancer and specifically the rate of early-onset CRC, or colorectal cancer diagnosed before the age of 50, has been increasing for the past 20 years. Research has shown that from 1990 to 2016, the median age of CRC diagnosis has decreased from 72 to 66 (1). Now, 1 in 10 of all new CRC diagnoses are classified as early-onset (2). This is detrimental to people with colorectal cancer because their survival rates dramatically increase when the cancer is detected and diagnosed earlier, as "stage I disease has a five-year survival rate as high as 90%, but stage IV disease has a survival rate of less than 10%" (3).

The current CRC screening guidelines recommend starting at age 45. For individuals with a higher risk due to family history, screening may start at age 40. Nevertheless, this system is inadequate for detecting all cases of early-onset CRC before the cancer progresses to an advanced stage.

Polygenic Risk Scores

To increase the survival rates, a solution lies in creating risk prediction models trained on large genome-wide genetic data and analyzing the Polygenic Risk Score (PRS) of colorectal cancer. A PRS is a calculated value that represents an individual's genetic susceptibility or risk to a specific trait or disease based on many genetic factors. PRS gives a method of quantifying an individual's genetic risk based on the cumulative effects of multiple genetic variants. However, the majority of genetic research had been distributed unfairly. "Currently, most genome-wide association studies only incorporate genetic data from people with European ancestry" (4). Addressing this lack of research would make colorectal cancer screening more reliable across ancestries.

Previous Research

Currently, there are several investigations being conducted all around the world on specific ethnicities, such as East Asian and African populations. For example, researchers at

Fred Hutch Cancer Center have developed models that incorporate Asian and European genome wide-association studies of CRC to improve the risk prediction across racial and ethnic populations (5). These models attempt to gap the disparities in genetic research by including diverse genetic data, which is key for accurate risk prediction and equitable screening practices. By using large-scale GWAS data and incorporating genetic variants from multiple populations, these models improve the understanding of CRC risk factors and contribute to the development of population-specific risk prediction tools. For example, researchers from the Division of Cancer Epidemiology and Genetics at the National Cancer Institute have said, “Results from simulation studies show that multiancestry methods generally lead to the most accurate PRSs in different settings” (6).

Various studies have demonstrated that advanced PRS models can identify a broader group of high-risk individuals, which couldn’t have been identified with current screening methods. “Based on the LDpred-derived PRS, we are able to identify 30% of individuals without a family history as having risk for CRC similar to those with a family history of CRC, whereas the PRS based on known GWAS variants identified only top 10% as having a similar relative risk. About 90% of these individuals have no family history and would have been considered average risk under current screening guidelines, but might benefit from earlier screening” (7). This significant improvement in percentage of identified individuals shows the potential positive impacts of more comprehensive genetic risk assessments.

Investigative Goals

What these investigations have not addressed is how the PRS calculation can show how the distribution of genetic risk varies across diverse ancestral groups and across genders within these populations. Understanding these variations can inform population-specific risk prediction tools and contribute to reducing health disparities in CRC outcomes. The goal of this research project is to calculate CRC PRS using diverse genome data and identify patterns in how the distribution of PRS varies across ancestral groups and across genders. This study will provide further insight into how genetic risk is distributed across ancestral groups and genders. The hypothesis of this study stated that if Polygenic Risk Scores (PRS) are calculated for colorectal cancer risk across different ancestral groups, the resulting scores will exhibit significant variation between the ancestral groups (European, African, East Asian, South Asian, American) represented in the diverse genome data, reflecting the genetic diversity and allele frequency differences among populations.

METHODOLOGY

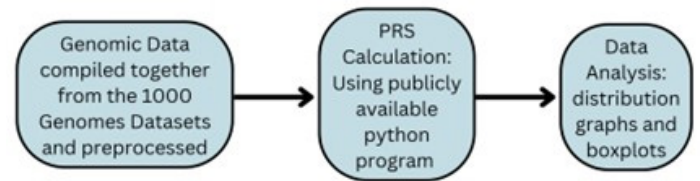


Figure 1. An overview of the research project procedure: First, colorectal cancer genomic data from 1000 Genomes will be compiled together and preprocessed. With this data, the PRS is calculated, and the distributions of the PRS will be analyzed for patterns.

Genomic Data Preprocessing

First, genomic data from the 1000 Genomes Project was obtained by downloading the Variant Call Format (VCF) files for each chromosome from the official repository. Preprocessing with the VCF files had to be done because the rsID column of the VCF files were empty, which is required for matching SNP variants to calculate PRS. Using BCFtools and the provided “all variations” file from the 1000 Genomes Project, the SNP rsIDs were annotated into the VCF file. BCFtools is a set of command line tools that assist with large-scale genomic analyses and has features of indexing and annotating VCF files. Indexing the VCF files is necessary to ensure the files contain all of the appropriate resources to allow fast access to specific regions and position in the file. This assists the polygenic risk score calculating tool access the data inside the VCF file more efficiently and calculate scores even faster. The BCFtools annotation command goes into the VCF file and writes in the important information of SNP variant rsIDs. BCFtools provides the final indexed and annotated VCF files which are used to calculate PRS.

Polygenic Risk Score Calculations

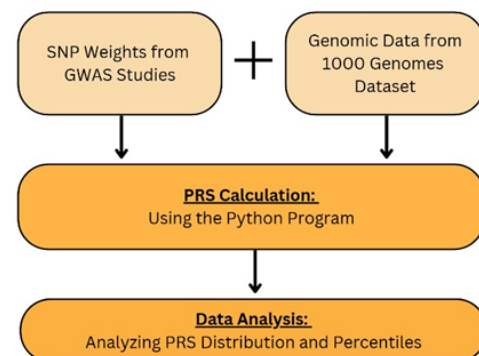


Figure 2. A flowchart of the PRS Analysis Process: The process starts by using the SNP Weights from previous GWAS Studies and Genomic Data from 1000 Genomes Dataset to calculate the PRS with the python program. With the final calculated PRS calculations, the PRS Distribution and Percentiles are analyzed with boxplots and distribution graphs.

To calculate the PRS, inputs from genomic data and CRC PRS weights are required. The genomic data was obtained by downloading the VCF files from the 1000 Genomes Project. This dataset was diverse, with 2504 individual samples with a demographic of 25% European, 25% African, 12.5% South Asian, 12.5% East Asian, and 25% Hispanic or Latin American ancestry. Additionally, the CRC PRS Weights file, which consisted of a table of Single-Nucleotide Polymorphisms (SNPs) with their effect sizes and reference alleles from previous Genome-Wide Association Studies (GWAS), were acquired from the publicly available Polygenic Score (PGS) Catalog. GWAS are observational studies that have identified associations between genetic variants, like SNPs, and traits such as CRC. The Polygenic Risk Scores were calculated using a publicly sourced python program, which used the Weights file and VCF file as inputs. rsIDs from the Weights file are matched and processed with rsIDs in VCF files, then summed in the PRS calculation. The python program used the following equation.

$$PRS_j = \frac{\sum_i^N S_i * G_{ij}}{P * M_j}$$

Figure 3. PRS Calculation Formula in PLINK: In this formula, the effect size of the SNP i is S_i ; the number of effect alleles observed in sample j is G_{ij} ; the ploidy of the sample is P (is generally 2 for humans); the total number of SNPs included in the PRS is N ; and the number of non-missing SNPs observed in sample j is M_j .

In order to run these calculations faster and more efficiently, the individual samples were split into smaller sections. This was done so that scores of one hundred samples would be calculated in one run of the program. These smaller sections then allowed the computer to run the program up to six times at once, accelerating the calculation process. PRS is calculated for each of the twenty-two chromosomes which were analyzed during this study separately, so each individual chromosome PRS was added together to generate a PRS Sum Score.

DATA ANALYSIS

To analyze the PRS distribution and percentiles across different ancestral groups, the mean, median, variance, standard deviation, and percentiles (25th, 50th, and 75th) were calculated using the PRS calculation results separated according to ancestral classification outlined in the integrated call samples file from the 1000 Genomes Project's official repository.

Population	Min	25th	Median	75th	Max
EAS	1.346882	4.017771	5.351523	6.165047	9.618078
EUR	3.626517	6.775271	9.401058	11.41897	15.32962
AFR	3.577509	8.970398	10.33523	11.37849	15.29736
SAS	4.143418	7.362492	9.567899	10.69792	12.70963
AMR	1.852393	7.597934	8.930846	11.13981	14.48461

Figure 4. Table of PRS Distribution across various population groups.

From these values, a distribution line graph of PRS across population groups was created. PRS across gender within each ancestral group was plotted in separate boxplots.

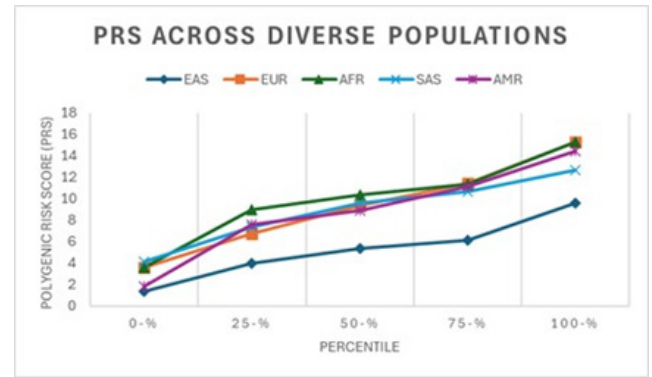


Figure 5. Line Graph of PRS Distribution across population groups

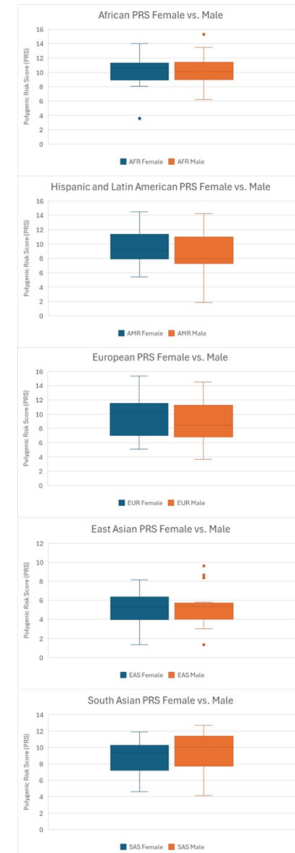


Figure 6. Box Plots of PRS Distribution across gender in each population group.

Figure 5 portrayed promising differences across population groups, so further statistical analysis was conducted to determine statistically significant differences. A one-way analysis of variance (ANOVA) test at a significance level of 0.05 was conducted to determine whether there is a statistically significant difference between the means of the main five ancestral groups or not. The conditions of normality, equality of variances, and test power were all met.

Figure 6 didn't show promising statistically significant differences across gender within population groups, but a T-Test calculation for significance between two sample means was done for each ancestral group to ensure this prediction. The conditions of normality, independence, and equality of variance were all met.

RESULTS

ANOVA Summary					
Source	Degrees of Freedom (DF)	Sum of Squares (SS)	Mean Square (MS)	F-Stat	P-Value
Between Groups	4	594.3856	148.5964	25.9102	≈ 0
Within Groups	2499	1118.3356	5.7351		
Total	2503	1712.7213			

Figure 7. Table of ANOVA Test results across ancestral groups.

The one-way ANOVA test revealed that since the P-Value of ≈ 0 is less than the significance level of 0.05, the null hypothesis is rejected. This shows that the difference between the mean PRS across the ancestral groups is statistically significant. The F-statistic is 25.91, which is not in the 95% region of acceptance (0:2.418). Since the one-way ANOVA has statistically significant results, Tukey's Honestly Significant Difference (HSD) test is done to determine which ancestral groups were statistically significantly different from the others. Tukey's HSD Test revealed that the East Asian PRS is statistically significantly different from European ($p=1.211e-10$), African ($p=1.026e-10$), South Asian ($p=3.172e-10$), and Hispanic and Latin American ($p=2.16e-10$) PRS.

Ancestral Groups	P-Values
EAS	0.8446964
EUR	0.4145356
AFR	0.7293849
SAS	0.3344216
AMR	0.2655196

Figure 8. Table of P-Values from T-Test for Significance across gender within each ancestral group.

The T-Test for significance between two sample means revealed no statistically significant difference across gender in every ancestral group. The p-values of the difference across gender in each ancestral group, East Asian ($p=0.84$), European ($p=0.41$), African ($p=0.73$), South Asian ($p=0.33$), and Hispanic and Latin American ($p=0.27$), were all above the significance level of 0.05, indicating no significant difference of PRS.

CONCLUSION

The hypothesis predicted that if Polygenic Risk Scores (PRS) are calculated for colorectal cancer risk across different ancestral groups, the resulting scores will exhibit significant variation between the ancestral groups (European, African, East Asian, South Asian, American) represented in the diverse genome data was proved correct through data analysis. For the PRS distribution across diverse populations, East Asian PRS was proven statistically significantly different from than the other population groups from a one-way ANOVA revealing statistically significant difference in PRS between at least two ancestral groups ($F(4, 2499)= 25.91, p \approx 0$) and Tukey's HSD finding that the mean East Asian PRS was statistically significantly different compared to every other group (European ($p=1.211e-10$), African ($p=1.026e-10$), South Asian ($p=3.172e-10$), and Hispanic and Latin American ($p=2.16e-10$)). These results reflect the fact that most genetic research with PRS has been conducted with populations of European ancestry, as compared with non-European ancestry.

However, no statistically significant difference across gender was found in any ancestral group ($p\text{-values}>0.05$). Boxplots of the distribution of PRS across gender did show outliers in the East Asian and African ancestral groups though, which shows signs of differences that could be discovered with future research.

LIMITATIONS AND FUTURE STEPS

One limitation of this study is the limited dataset. The total sample size includes 2504 individuals, but when separating the data by ancestral group to analyze differences across gender, only 501 samples for each ancestral group, with the exception of 250 samples each of East Asian and South Asian ancestral groups, could be analyzed. In the future, utilizing an even larger genomic dataset could assist in outlining more distinct differences across gender specifically. Even more diverse datasets with various ethnic groups inside the five main ancestral groups used in this study would help improve the applicability of PRS across a wider range of people. Additionally, combining PRS with other biomarkers, such as epigenetic markers, could provide a more comprehensive risk assessment. A Polygenic Risk Score is a risk predictive measure that has great potential for applications in personalized medicine. Prevention and treatment strategies on an individual

level could be improved based on individual genetic risk with tools like this.

REFERENCES

1. Archambault AN, et al. Cumulative Burden of Colorectal Cancer-Associated Genetic Variants Is More Strongly Associated With Early-Onset vs Late-Onset Cancer. *Gastroenterology*. 2020 Apr;158(5):1274-1286. e12. doi: 10.1053/j.gastro.2019.12.012. Epub 2019 Dec 19. PMID: 31866242; PMCID: PMC7103489.
2. Ullah, F., Pillai, A. B., Omar, N., Dima, D., & Harichand, S. (2023). Early-Onset Colorectal Cancer: Current Insights. *Cancers*, 15(12), 3202. <https://doi.org/10.3390/cancers15123202>
3. Yue Zhang, Yin Wang, Bingqiang Zhang, Peifeng Li, Yi Zhao, Methods and biomarkers for early detection, prediction, and diagnosis of colorectal cancer; *Biomedicine & Pharmacotherapy*, Volume 163, 2023, 114786, ISSN 0753-3322, <https://doi.org/10.1016/j.biopha.2023.114786>.
4. Ho PJ, et al. Polygenic risk scores for the prediction of common cancers in East Asians: A population-based prospective cohort study. *Elife*. 2023 Mar 27;12:e82608. doi: 10.7554/eLife.82608. PMID: 36971353; PMCID: PMC10159619.
5. Thomas M, et al. Combining Asian and European genome-wide association studies of colorectal cancer improves risk prediction across racial and ethnic populations. *Nat Commun*. 2023 Oct 2;14(1):6147. doi: 10.1038/s41467-023-41819-0. PMID: 37783704; PMCID: PMC10545678.
6. Zhang, H., Zhan, J., Jin, J. et al. A new method for multiancestry polygenic prediction improves performance across diverse populations. *Nat Genet* 55, 1757–1768 (2023). <https://doi.org/10.1038/s41588-023-01501-z>
7. Thomas M, et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am J Hum Genet*. 2020 Sep 3;107(3):432-444. doi: 10.1016/j.ajhg.2020.07.006. Epub 2020 Aug 5. PMID: 32758450; PMCID: PMC7477007.

“When Copper Kills the Good Guys”: Novel Approach to Investigating the Hidden Microbial Effects of Copper IUDs on Female Reproductive Health

Dishita Gopikrishna

ABSTRACT

There are controversies surrounding contraceptives such as birth control pills and intrauterine devices (IUDs), mainly due to side effects such as cramping, irregular periods, and greater possible cancer risk. Despite the impact it has on women's quality of life, these topics remain under-researched. This project looks at how Copper IUDs in particular are a cause of bacterial vaginosis (a bacterial infection) due to loss of *Lactobacillus crispatus* (a bacteria common in the vaginal microbiome). The hypothesis was that in the presence of copper, *L. Crispatus* growth would be inhibited due to the antibacterial properties. The *L. crispatus* was placed in petri dishes to grow either with copper (Cu) solution (experimental group), penicillin (positive control) or MRS broth (negative control) and was analyzed. The length of the zone of inhibition which is clearance in the petri dish which indicates lack of bacterial growth was measured. The petri dishes with copper solution had 1-2 cm long zone of inhibition. The penicillin dishes had a larger zone of inhibition of 4-5 cm whilst the MRS dishes have no zones of inhibition. The hypothesis was correct, showing that *L. Crispatus* growth is inhibited in the presence of copper.

INTRODUCTION

This is an experiment on how copper inhibits growth of certain *Lactobacilli* such as *Lactobacillus crispatus*, which is a cause for a number of vaginal ailments such as bacterial vaginosis, a type of vaginal dysbiosis. Dysbiosis is the imbalance of bacteria in the microbiota of the human body. Around 50-70% of the population globally have been affected by dysbiosis. In general, this disrupts homeostasis, causing disorders such as diabetes, obesity, and bacterial vaginosis. It's often due to lifestyle choices like diets, smoking, and the use of certain antibiotics. Dysbiosis can occur to any microbiome in the human body and it affects the different microbiomes in the body in different ways. Key bacteria like *Lactobacillus Crispatus* and *Lactobacillus Jensenii* dominate the female reproductive microbiome mainly in the vagina and cervix. There, dysbiosis causes the loss of *Lactobacillus* and the increase of other bacteria such as *Gardnerella vaginalis*, *Candida* spp, *Mycoplasma hominis*, *Mobiluncus* spp, and *Bacteroides* spp. This condition is known as Bacterial Vaginosis or BV for short. According to the National Library of Medicine, “It afflicts 23%–29% of women worldwide” (Chen, Chen, Li, Lu, 2021). In Symptomatic women, signs of BV include a strong fishy odor, increased gray or white secretion from the vagina, burning or itching sensation in the vagina.

One bacterium in particular – *Lactobacillus crispatus* – is the most abundant species of bacteria taking up over 60% of the vaginal microbiome. *Lactobacillus crispatus* is a type

of probiotic bacteria in the vaginal flora. Along with being found in the vagina and female reproductive tract, it is also found in fermented foods like sourdough. The role it has in the vaginal microbiome is that maintains health in that area on a microscopic level by producing acids, hydrogen peroxide and bacteriocins. Furthermore, it inhibits pathogenic bacteria and viruses which prevent infection such as Bacterial Vaginosis.

However, the role it has in the cause of dysbiosis has not been tested yet. Medical research related to women's health has improved in recent medical microbiome research. It is also relatively new and so far, research has been done on the Gut Microbiota. That research has resulted in fecal transplants and the use of probiotics. There is however a gap in research regarding women as microbiomes specific to women such as the uterine, cervical, and vaginal microbiome are under-looked in research in comparison to other microbiomes not reproductive related such as the gut. According to PubMed database, “the number of grants for research on non-reproductive organs is 6–7 times higher than the number for reproductive organs” (Cox, Mercuri, 2022).

IUDs (Intrauterine Devices) are a type of contraceptive used by women. There are two kinds, copper IUDs and hormonal IUDs. Both work similarly by killing or inhibiting sperm from continuing and fertilizing the egg. This is what makes it “the third most common birth control method used by American women (behind tubal ligation and the oral

contraceptive pill)” (Macmillian, 2024). However, copper IUDs are likely to cause more harm than good towards the vaginal microflora as copper is bactericidal which can cause dysbiosis due to decrease in Lactobacillus. Copper is deadly to bacteria due to Cu 2+ ions tearing apart the bacterial cell membrane causing a leakage in essential nutrients.

Another study conducted the effects of copper IUDs reports that “after 6 months of use, Cu-IUD associated increases in BV-associated bacterial concentrations were significantly correlated with the concentrations of several cervicovaginal cytokines, which has broad, negative implications for genital health and STI susceptibility” (Brown BP. Et all 2023). Due to the earlier information on how Lactobacillus crispatus has a role in inhibiting the growth of pathogenic or infectious bacteria. This leads to the hypothesis that Copper IUDs are a cause of Dysbiosis of the Vaginal and Cervical Microbiomes due to the loss of Lactobacillus crispatus.

METHODOLOGY

For this study, multiple materials for culturing L Crispatus were used from Jaspan Lab at Seattle Children’s Hospital. This includes sixty-two petri dishes, 18Cu solution (copper chloride liquid solution) and 3-5 mL of Lactobacillus bacteria (Lactobacillus crispatus). To culture the Lactobacillus media, 50 mL of agar gel and 50 mL of MRS broth was made at the lab by mixing agar powder and MRS powder respectively into two jars of 50mL water which was then mixed and sent to the autoclave for sterilization. Lastly, copper wire that is around 1200 milligrams (1.2 grams) was brought from home and also sent to the autoclave for sterilization.

In order to identify whether copper has a role in killing or preventing growth of Lactobacillus crispatus, a specific type of test – the Kirby Bauer Inhibition test – is done. This is when an anti-microbial agent is put on a paper disk on the surface of an agar plate containing bacterial species. This is mainly used to test the effect antibiotics have on various pathogenic bacterial strains such as E. Coli. Hence, a few tweaks to the usual Kirby-Bauer Inhibition procedure were made to fit the usage of copper as an anti-microbial agent and non-pathogenic Lactobacillus bacteria strains. Despite Lactobacillus crispatus being a relatively harmless species of bacteria categorized as BSL-1 with little to no negative effects on humans, for safety due it being a microorganism the experiment will be done in a BSL-1 lab in Seattle Children’s Hospital.

Extract the cultured bacteria and put it in the petri dishes. Cut the copper wire into several tiny pieces and weigh each piece (100-300 grams) to ensure it matches the weight of commercial IUDs and set aside. In five petri dishes place 2 copper wires, in three petri dishes place four copper wires and in another five petri dishes place eight copper wires. Incubate the plates at around 35-37 degrees Celsius to allow

bacteria growth over a period of 48 hours.

After the period of 48 hours is over, observe the number of bacteria in the petri dish without copper and the petri dish with copper to check for decrease by checking for zones of inhibited growth which are clear areas around antibacterial agent such as copper where bacteria do not grow (copper is bactericidal and kills bacteria indiscriminately) using a ruler and add to the data table below. Make sure to note down the specific size of the zone of inhibition regarding the different concentrations of copper to calculate percent decrease of bacterial inhibition. If in the petri dish with copper the zone of inhibition becomes prominent, measure and then mark in the table (added to the data table) as bacteria is unable to grow (bacterial loss).

Data Table #1. Percentage of Bacteria in Presence of Copper

Experimental Group: Petri dish with copper	Amount of # copper wires:	Percentage of <i>Lactobacillus crispatus</i> (%)	Length of Zone of Inhibition (cm)
	2		
	4		
	8		

A negative control group with the remaining four agar plates will be used but these will have no copper wires on it to further test the effect copper has on Lactobacilli growth overall. There is also a positive control group where penicillin – an antibiotic particularly deadly towards Lactobacilli – was added to the remaining petri dishes to further determine the accuracy and statistical significance of the results. The agar plates with the bacteria will all be incubated for 48 hours as well. Afterwards the growth of the bacteria will be observed to compare with the agar plates with copper and bacteria along with the length or lack thereof of Zone of Inhibition.

Data Table #2. Percentage of Bacteria without presence of Copper:

Experimental Group: Petri dish with copper	Percentage of <i>Lactobacillus crispatus</i> (%)	Length of Zone of Inhibition (cm)

Data Table #3. Percentage of Bacteria with presence of Penicillin:

Experimental Group: Petri dish with copper	Percentage of <i>Lactobacillus crispatus</i> (%)	Length of Zone of Inhibition (cm)

When using 18Cu solution as the experimental group, first L. Crispatus has to be spread around each petri dish in order to create a ‘lawn’ of bacteria. Afterwards make sure to split 18Cu into four test tubes, then dilute 500 microliters of it with 500 microliters of MRS broth, to make the

solution $9e^{-3}$ Cu in order to better simulate the amount of copper and Cu $2+$ ion activity in an IUD. Use a micropipette to collect 1mL or 1000 microliters and of $9e^{-3}$ Cu and release it in the center of one of the petri dishes, continue to do this for two more petri dishes. Then collect 10 microliters of $9e^{-3}$ Cu and dilute it with 990 microliters of MRS broth to get 1 mL of $9e^{-5}$ Cu, do these another two times to put it on another three petri dishes. Do the same with 10 microliters of $9e^{-5}$ and 990 microliters of MRS broth to get $9e^{-7}$ Cu and the same with 10 microliters of $9e^{-7}$ and 990 microliters of MRS broth to get $9e^{-9}$ Cu. Do these another two times to put it on another three petri dishes for each and use a micropipette to collect 1mL or 1000 microliters and of each Cu solution and release it in the center of one of the petri dishes. Incubate these petri dishes for 48 hours and then afterwards take it all out for analysis and results.

Experiment Diagram:

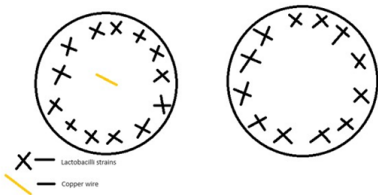


Figure 1: L. Crispatus Petri Dish Experimental

Images from Lab:

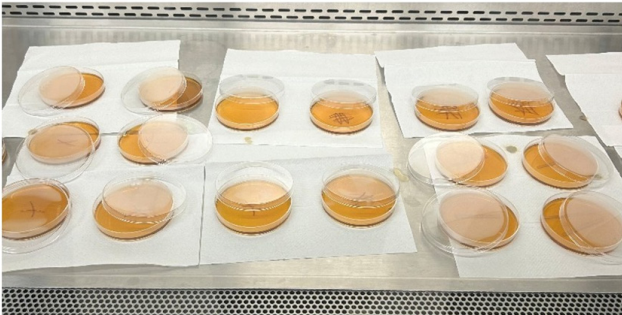


Image 1: Three Petri Dish Groups: Taken by Dishita Gopikrishna at Seattle Children's Hospital

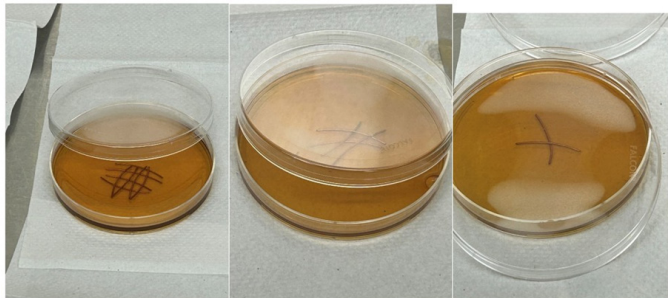


Image 2: 8Cu Petri Dish

Image 3: 4Cu Petri Dish

Image 4: 2Cu Petri Dish

DATA ANALYSIS – First trial:

It is seen that there is not much of a difference between the negative control where nothing was added and the experimental groups with 2 copper wires, 4 copper wires and eight copper wires all with averages between 1.71 - 1.75. Furthermore, as seen in Table 3, positive control has no growth as expected as penicillin as an antibiotic that these L Crispatus had not been introduced to yet. The negative control grew freely as nothing was put on it, which was also as expected. Due to both Positive control and Negative control yielding accurate results, this makes the results of the experiment statistically significant.

L Crispatus Growth	2Cu	4Cu	8Cu	Positive	Negative
	1.7	1.75	1.75	0	1.75
	1.75	1.75	1.75	0	1.74
	1.7	1.72	1.75	0	1.75
mean	1.716667	1.74	1.75	0	1.746667
sd	0.028868	0.017321	0	0	0.0057735

Table 3: First Trial Data Results

Column1	mean	stdev
2Cu	1.716666667	0.028867513
4Cu	1.74	0.017320508
8Cu	1.75	0
Positive	0	0
Negative	1.746666667	0.005773503

Table 4: Mean and Standard Deviation

The means for the negative control and experimental groups are Moreover, using Excel Data collection software, calculating data from Table 2, using the function ANOVA in the Data Analysis feature in Excel, the p value was calculated to be 0.00000000021, further confirming the experiment to be statistically significant as it is less than 0.05. Thus proving the hypothesis incorrect.

DATA ANALYSIS – Second Trial:

When later the Cu solution was used, however, there was a great difference in results. This time there were zones of inhibition that were measured (in cm). These zones of inhibition or clearance were present in the positive control of penicillin measuring around 4-5 cm and in the experimental group of $9e^{-3}$ Cu and $9e^{-5}$ Cu measuring 1-2 cm. The remaining experimental groups ($9e^{-7}$ and $9e^{-9}$) and the negative control had no zones of inhibition present indicating no inhibition of L Crispatus.

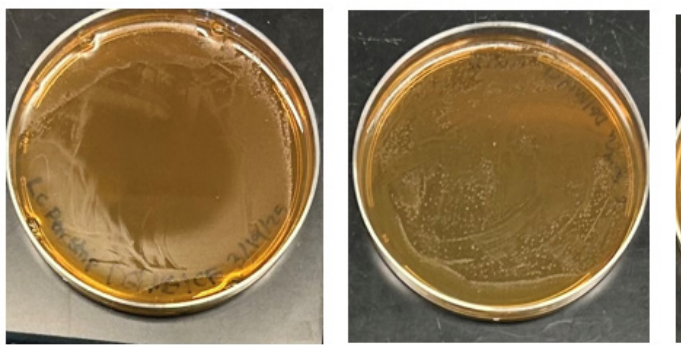


Image 5: [Positive control](#) Image 6: [Negative control](#)

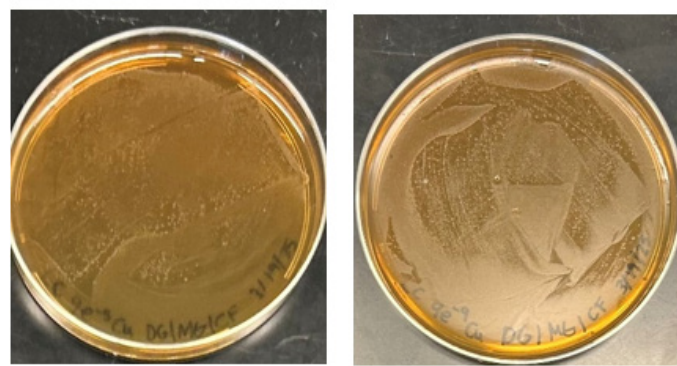


Image 7: $9e^{-9}$ Cu

Image 8: $9e^{-7}$ Cu

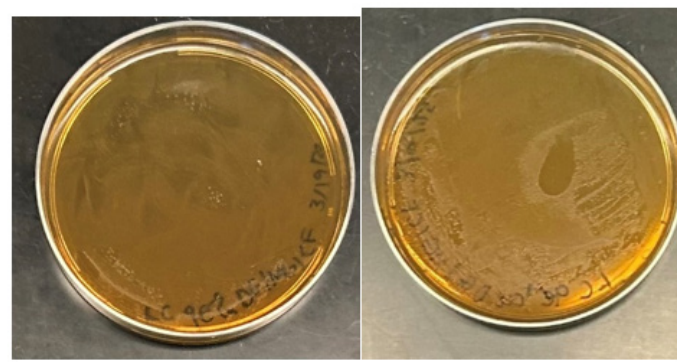


Image 9: $9e^{-5}$ Cu

Image 10: $9e^{-3}$ Cu

Using Excel Data collection software, calculating the data collected, using a single factor ANOVA test in the Data Analysis feature in Microsoft Excel, the p-value was calculated to be 0.000000000021, further confirming the experiment to be statistically significant (less than 0.05). Thus, proving the hypothesis correct.

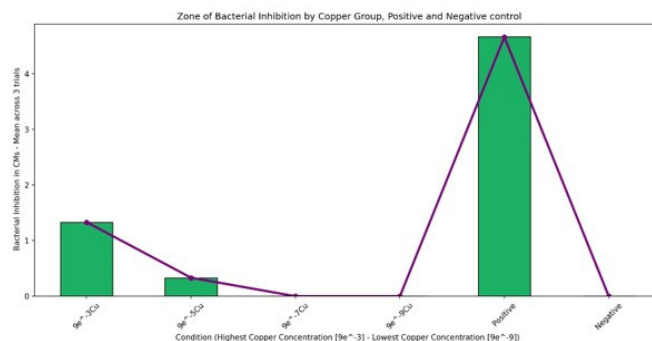


Table 5: *Second Trial Data*

DISCUSSION

With the *L. Crispatus* growth of the experimental group regardless of how much copper was placed, being almost the exact same as the bacterial growth of the negative control where nothing was placed and the bacteria could grow freely, the hypothesis was proven incorrect. Moreover, with positive and negative control yielding accurate results as expected the data is statistically significant and cannot be the result of a lab error.

However, with the second trial using Cu solution of copper chloride, the experimental group with larger concentrations of Cu solution ($9e^{-3}$ and $9e^{-5}$) there was a zone of clearance showing copper preventing *L. Crispatus* from growing, causing the experiment to successfully prove the hypothesis to be correct. Moreover, positive control had a large zone of clearance preventing bacterial growth and the negative control had no inhibition at all, along with p-value being far less than 0.05, the experiment is statistically significant.

Copper wires are 99% pure copper, meaning that it is harder for copper ions to be released and cause inhibition in the growth of *L. Crispatus*. Copper wires are the material that is the closest to what is used in copper IUDs. However, in previous studies done within the last two years, such as longitudinal studies involving analyzing vaginal swabs with copper IUDs there is a connection seen between lack of *Lactobacilli* growth and copper. With further research done, it was found that the petri dishes needed to simulate the vaginal environment more accurately. In the vagina and in a healthy vaginal microbiota, the pH is between 3.5 - 4.5 making the area relatively acidic. Even with the relatively weak acidity, the copper in IUDs corrodes despite it being nearly pure copper causing the ions to release and inhibit the growth of *Lactobacilli* notably *Lactobacillus crispatus*. Putting a weak acid with relatively the same pH as the vaginal microbiota will lead to results that will support the hypothesis and be more accurate to the vaginal microbiome when an IUD is inserted.

Furthermore, in my second trial and few other studies analyzing the effect copper has on Lactobacilli in the vaginal environment, a solution of copper chlorate or copper sulfate is used. It is not pure copper, and hence copper ions are released more easily. Future studies comparing how this second trial and this possible third trial with copper wires and possible weak acid can determine the specific effect copper has more closely.

Other studies can include *Lactobacillus acidophilus*, which is another probiotic bacterium like *L. Crispatus* which is found in the vaginal microflora albeit a lot less. This microbe has been studied to be a possible treatment for various microbial, vaginal disorders such as a yeast infection due to its strong probiotic properties. There have also been studies that explore *L. acidophilus* resilience against metals such as copper which are antibacterial. These tests can determine whether *L. acidophilus* can be a viable probiotic supplement that women with copper IUDs are advised to have by a gynecologist to prevent bacterial vaginosis risk (BV) and still retain usage of copper intrauterine contraceptives.

More studies in the future can possibly entail how copper IUDs due to the effect on the vaginal microflora are linked to various disorders such as ovarian cysts, uterine fibroids and even cervical and endometrial cancers, which are also linked to other significant issues for women such as infertility such as with uterine fibroids. This can help further research on contraceptives and potentially create birth control with safer ingredients that is both effective and not damaging to the vaginal microbiome.

ACKNOWLEDGEMENTS

Thank you to the Jaspan lab for allowing use of the lab and items for the experiment. Thank you to Heather Jaspan, Melanie Gasper and Colin Feng for being essential guides to the procedure, experiment, data analysis and support with bringing this experiment to life. Moreover, thank you Ms. Allender for support and guidance with this research and experiment.

BIBLIOGRAPHY

1. Ansari A, Son D, Hur YM, Park S, You YA, Kim SM, Lee G, Kang S, Chung Y, Lim S, Kim YJ. Lactobacillus Probiotics Improve Vaginal Dysbiosis in Asymptomatic Women. *Nutrients*. 2023 Apr 13;15(8):1862. doi: 10.3390/nu15081862. PMID: 37111086; PMCID: PMC10143682.
2. Brown, B.P., Feng, C., Tanko, R.F. et al. Copper intrauterine device increases vaginal concentrations of inflammatory anaerobes and depletes lactobacilli compared to hormonal options in a randomized trial. *Nat Commun* 14, 499 (2023). <https://doi.org/10.1038/s41467-023-36002-4>
3. Chopra, C., Bhushan, I., Mehta, M., Koushal, T., Gupta, A., Sharma, S., ... & Sharma, S. (2022). Vaginal microbiome: considerations for reproductive health. *Future microbiology*, 17(18), 1501-1513.
4. DElkafas, H., Walls, M., Al-Hendy, A., & Ismail, N. (2022). Gut and genital tract microbiomes: Dysbiosis and link to gynecological disorders. *Frontiers in cellular and infection microbiology*, 12, 1059825.
5. Fowler L, Engqvist H, Öhman-Mägi C. Effect of Copper Ion Concentration on Bacteria and Cells. *Materials (Basel)*. 2019 Nov 19;12(22):3798. doi: 10.3390/ma12223798. PMID: 31752323; PMCID: PMC6888263.
6. Han Y, Liu Z, Chen T. Role of Vaginal Microbiota Dysbiosis in Gynecological Diseases and the Potential Interventions. *Front Microbiol*. 2021 Jun 18;12:643422. doi: 10.3389/fmicb.2021.643422. PMID: 34220737; PMCID: PMC8249587.

COMPUTATIONAL BIOLOGY

A Diffusion-Based Approach to Restoring Ancient Oracle Bone Characters

Thea Sima

ABSTRACT

Oracle bone script is the oldest attested form of written Chinese and can provide valuable insight into how language and culture have evolved in the region. However, wear on the inscriptions hinders efforts to decipher them. This project investigated the use of a diffusion model to fill in obscured areas of oracle bone script with generated images. The engineering goals were to generate images visually similar to the oracle bone dataset and achieve over 0.95 Structural Similarity Index, 0.5 Universal Image Quality, 30 Peak Signal to Noise Ratio, and a Fréchet Inception Distance under 30. The metrics chosen aim to measure quality as perceived by humans, realism (similarity to the original dataset), and fidelity. The model was trained with an oracle bone inscription dataset and then refined for inpainting (filling in a specified area with an image) using a loss function that looks specifically at the masked area (the area selected for inpainting). The model iterations (independent variable) were compared using SSIM, UIQ, PSNR, and FID (dependent variables). With a diverse dataset and the addition of refinement, model performance measured by Fréchet Inception Distance improved by 38%. The model has potential in applications for restoring areas of oracle bone inscriptions, achieving a Fréchet Inception Distance of 27.33. While this demonstrates that image inpainting with a diffusion model is a viable strategy for restoring ancient characters, the model failed to achieve the criteria set for Structural Similarity Index with the value 0.538.

INTRODUCTION

This project aims to investigate the use of a diffusion model for generating images of damaged oracle bone inscriptions as a method of restoration to decipher worn oracle bone script images.

Going back as far as 1400 BC, oracle bone script is the oldest form of written Chinese (1). Characters were written on ox scapula and turtle shells for divination rituals. By studying oracle bones inscriptions, researchers gain insight into ancient Chinese history, culture, and the evolution of written language. The inscriptions provide information on the social structures and worldview of the time, along with natural phenomena such as comet sightings. These ancient characters suffer from over three thousand years of wear, which makes deciphering oracle bones a difficult and ongoing task. While over 4500 characters have been discovered, only a third have been successfully deciphered. Continuing these efforts improves modern-day understanding of history, allowing society to apply lessons from the past to the present and future.

Traditionally, deciphering the script has been approached from archaeological, paleographic, and historical linguistic perspectives. Now, with machine learning on the rise, some literature has focused on using it in attempts to connect oracle characters to modern-day equivalents using diffusion models (1) or to restore damaged images using Generative

Adversarial Networks (GANs) (2). However, restoration of the inscriptions themselves has not been explored as widely as avenues like classifying characters. Additionally, as machine learning techniques are developed, new methods can improve stability and image quality, so different applications of machine learning merit revisiting.

Inpainting is a technique for restoring missing or damaged areas of an image by filling them in to present a more complete picture. In terms of oracle bone inscriptions, this can be applied by attempting to use image generation to fill in worn and unreadable areas of characters.

Diffusion models for image generation have increased in popularity and capability in recent years with the emergence and rapid progress of text-to-image generative artificial image models like Stable Diffusion, Midjourney, and DALL-E (3). Out of the box, these models tend to not be task-specific and are designed for use across a wide range of inputs and applications. However, task-specific diffusion for inpainting was found to be more effective than models using multi-task training, outperforming them in terms of Fréchet Inception Distance, Perceptual Distance, Inception Score, and Classification Accuracy (4).

Diffusion models also address the issue of mode collapse in GANs, a failure mode where the GAN generates only a small set of outputs because of the interaction between the discriminator and generator. The goal of the generator is to generate outputs plausible enough to make the discrimina-

tor think them real, but if the generator consistently produces the same small set of outputs and the discriminator gets stuck in a local minimum. Comparatively, diffusion models can produce more diverse outputs that better capture the full variation of the training data with less adjustment. To explore the effectiveness of diffusion for the specific task of inpainting oracle bones, the model will be tested on square masks of fixed size in the center of the image. The purpose of this project is to improve accuracy to the original image compared to existing non-diffusion inpainting models for this task and diffusion models trained on other datasets when applied to this task by reaching at least 0.98 SSIM, 30 PSNR, UIQ of 0.5, and FID under 30.

METHODOLOGY

Tools and Dataset

The prototype presented uses open-source image data from HUST-OBC, compiled by Wang et al. in 2024 (5). Additionally, other open-source image data was obtained from Oracle-241, which was compiled by Wang et al. in 2022 (6), along with the ddp-ema-bedroom dataset published by Google (7). Major Python libraries and resources used include Pillow, for image processing; Diffusers from Huggingface, for creating the U-Net, scheduler, and pipeline; and PyTorch for tensor operations.

Preprocessing

The images, which are greyscale, were resized to be 64 by 64 pixels. This size was chosen to take training time and model performance into account; images with larger sizes are higher resolution, but increase training time, while the size chosen for the training images still leads to an effective model. The data was split to be 80% training and 20% testing. The pixel values were normalized to be between -1 to 1. This is a common technique with machine learning models to improve training speed and stability by preventing exploding gradients—an issue where model weights are updated in huge steps due to weight gradients getting excessively large, thus preventing a local minimum from being found.

Diffusion

Diffusion models take Gaussian noise and generate images from the noise. They work with two processes: one to add noise, and one to remove noise. The denoising process is what allows the model to generate images. The amount of noise at each step is what the model attempts to learn and remove. The training dataset can be thought of as points sampled from an underlying probability distribution of desirable images, and learning how to go from a random point in the image space towards a region of desired images (by removing noise) is how coherent images are generated. This is reflected in the training process of the U-Net, which is described in further detail in the Model Architecture section.

The model uses a cosine schedule to define the rate of noise addition and thus information destruction. The cosine

schedule is used to combat the two key issues observed with a linear schedule: destruction of information is too rapid, and the noised image is uninformative in the last steps as it approaches pure noise. This is implemented using the DDPM Scheduler of the Huggingface Diffusers library.

The forward diffusion process involves the iterative addition of noise to the image data over a defined number of timesteps. The use of the schedule leads to different noise amounts being added at each timestep, so knowledge of the current timestep allows the model to accommodate for this variation. To communicate the timestep each epoch, a sinusoidal embedding is projected to the residual blocks of the U-Net. When generating images with a pipeline, sampling images through denoising is done over 60 timesteps. The inpainting pipeline used inherits from the DiffusionPipeline class from the Huggingface Diffusers library (8).

Model Architecture

The model uses an encoder and decoder that form a U-Net. The U-Net was first described by Ronnenberg et al. in 2015 and is a convolutional neural network architecture (9). The building blocks of the U-Net used in this prototype can be seen in Figure 1. The contracting path reduces the spatial dimensions of the image while increasing the number of channels that represent feature maps. With each step, a convolution layer is applied to generate filters (the additional channels) to segment the image while decreasing the image in height and width. The expansive path increases the dimensions of the image while decreasing the number of channels. Skip connections from the contracting path to the expanding path preserve spatial information. The skip connections help the decoder include finer details that are lost when prioritizing global structure.

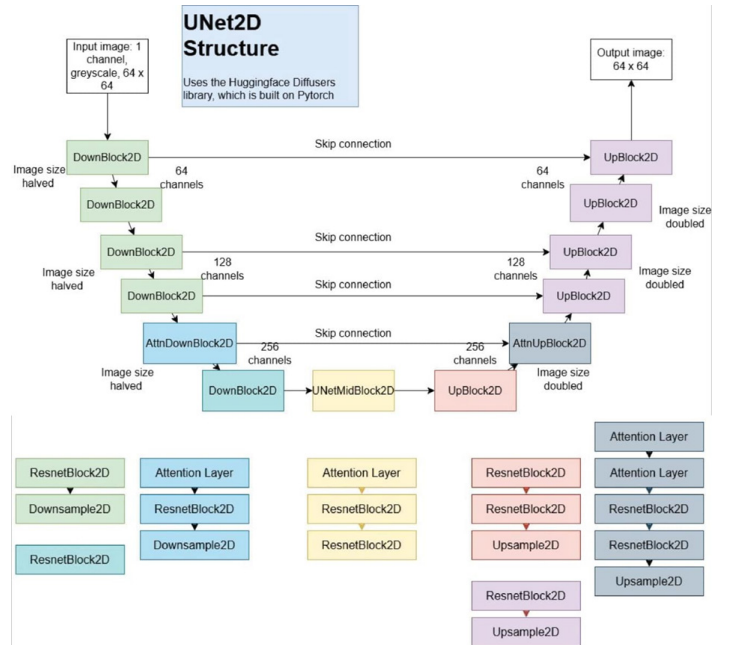


Figure 1. Flowchart of U-Net Structure. Visual representation of the various blocks of the U-Net architecture, which is often used for image segmentation tasks.

The U-Net was trained off of the HUST-OBC dataset. The U-Net implementation is modified from the unconditional Glyffuser model (10). The model was trained in batches of 16, with 100 epochs, and a learning rate of $1e-4$ (0.0001). The Diffusers library from the Huggingface platform was used to implement the model in Python 3.12. For the loss function, Mean Squared Error was used. For training the U-Net, the loss function compared the noise at random timesteps and the noise predicted by the U-Net. The PyTorch implementation of the AdamW algorithm was used as an optimizer to update the weights of the U-Net.

Retraining

The U-Net weights were further modified by retraining. The U-Net was loaded as a pretrained model and then opened in training mode to allow for weight updates. A function was written to generate masks, to simulate deterioration. These masks had a value of 1.0 in a randomly generated area of the image, and a value of 0.0 everywhere else. The updated loss function compared the predicted sample image in the area selected by the mask and the ground truth image. This was done by multiplying the predicted image by the mask and the target image by the mask, ensuring that only the selected area influenced the computations of the gradients. This addition adjusted the weights further based on only the areas that were masked, in order to more closely simulate inpainting selected regions.

Evaluation Protocol

The Peak Signal to Noise Ratio, Structural Similarity Index Measure, and Universal Image Quality of each inpainted image was found for each testing image, and then averaged, to evaluate how the model functioned. Fréchet Inception Distance was calculated in batches of 32 images for a total of 6400 images. Excluding FID, these were calculated using a Python package implemented by Ekhtiari (11).

PSNR quantifies the fidelity of the reconstructed image by measuring the logarithmic ratio of the peak signal to the mean squared error between the original and inpainted images, with higher values indicating better reconstruction quality. SSIM is a measure of image quality that looks at the perceptual similarity by incorporating structural information, luminance, and contrast. Proposed in 2004, it ranges from -1 to 1, with values closer to 1 indicating higher similarity (12). UIQ, its predecessor, evaluates correlation, distortion of luminance, and contrast, ranging from 0 to 1 (13). Values closer to 1 indicate higher similarity and less distortion. The main difference between their equations arises from the introduction of constants into the numerator and denominator of the luminance and contrast measures; the Universal Image Quality Index thus corresponds to a special case where those constants are zero. These measures form the assessment of the inpainting model's ability to accurately restore missing regions.

Fréchet Inception Distance was proposed by Heusel et al. as a method to measure the similarity of generated versus real images (14). It looks at the means and covariances of the feature vector that is taken from feeding the original and predicted images into an Inception v3 convolutional neural network. The result of the feature vector can be compared between the original images and the predicted images. This allows Fréchet Inception Distance to compare the probability distributions of the real images to the distribution of the generated images to evaluate model quality. This has advantages over comparing images pixel-by-pixel because it is computed between high-level representations of the images, making it more aligned with practical applications.

Fréchet Inception Distance was computed using the Inception v3 model trained off of ImageNet (15). The formula and process are shown in Figure 2.

$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

μ_r and μ_g are the means of the feature vectors for the real and generated images. Σ_r and Σ_g are the covariance matrices of the feature vectors. Tr is the trace of a matrix (sum of diagonal elements).

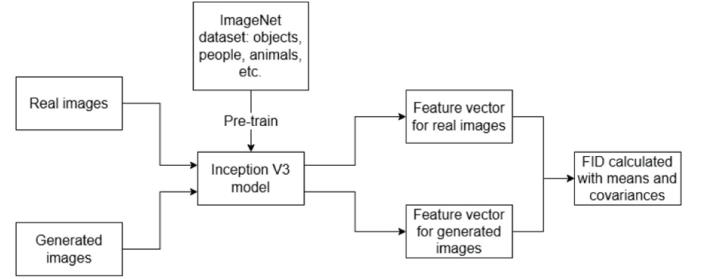


Figure 2. Fréchet Inception Distance. The formula and process used to calculate Fréchet Inception Distance.

Three models were compared: a general inpainting model not specifically trained on oracle bone images, using Google's ddpm-ema-bedroom dataset; a model trained off of the Oracle-241 dataset; and a model trained off of the HUST-OBC dataset and refined with mask loss.

RESULTS

After evaluating the metrics with the different testing datasets, the results were averaged and recorded in Figure 3. The results show that the highest-quality images for PSNR, UIQ, and FID result from the HUST-OBC model, with values of 43.202, 0.817, and 27.332, respectively. However, the highest SSIM was found with the model trained off of Oracle-241, at 0.962. This meets the engineering goals of 0.5 Universal Image Quality, 30 Peak Signal to Noise Ratio, and a Fréchet Inception Distance under 30, but fails to meet the goal of 0.95 Structural Similarity Index. Consequently, the choices of training with HUST-OBC and retraining with mask-specific loss did not improve upon all aspects of the model.

Table 1. Comparisons of Model Performance

Model	SSIM $\uparrow$	PSNR $\uparrow$	UIQ $\uparrow$	FID $\downarrow$
Bedroom	0.883638928	38.55886045	0.095535579	189.55
Oracle-241	0.961948755	42.46012705	0.367396658	44.16665952
Retrained HUST-OBC	0.538923282	43.20280755	0.817434102	27.33230242

Figure 3. Data table of different metrics (SSIM, PSNR, UIQ, and FID) to compare performance for three different models. $\uparrow$ indicates a metric that implies higher quality with a higher number.

For each of the groups, the standard error was also calculated. They can be found in Figure 4. There is no overlap when looking at the mean value $\pm$ SE, suggesting that the modifications made between each group were significant.

Table 2. Comparisons of Standard Error

Model	SSIM $\uparrow$	PSNR $\uparrow$	UIQ $\uparrow$	FID $\downarrow$
Bedroom	0.001419528	0.50842736	0.009148955	1.843582625
Oracle-241	0.00012306	0.141157326	0.001999342	0.967087211
Retrained HUST-OBC	0.00127141	0.025259285	0.000267005	0.623777951

Figure 4. Standard Error for each of the categories in Figure 3.

Based on the visual quality of the generated outputs, the model achieves stylistic similarity between the real dataset and the generated images. Examples can be found in Figure 5. There is little noise, and the generated shapes are legible. However, generated portions can be disconnected from the intact shapes that are not masked, which is a possible area of improvement to better fulfill the engineering goals.

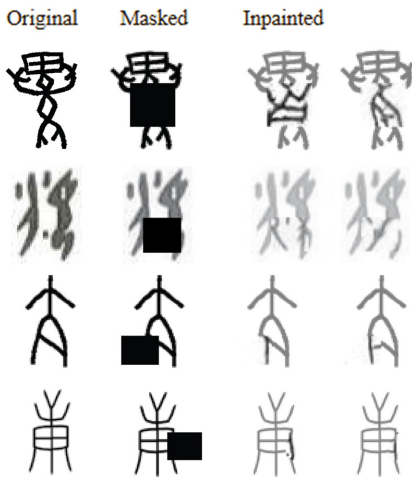


Figure 5. Examples of Inpainted Images. Comparisons of the original, masked, and inpainted images with rectangular masks. The masks are colored black for visibility.

DATA ANALYSIS

Beyond mean and standard error, an Analysis of Variance (ANOVA) test was conducted to determine statistical significance between groups, and thus significance of the changes made with each iteration. Using a two-way ANOVA test, the means of each group and the variance within groups and between groups was computed. The F-statistic value given is used to measure whether statistical significance is present. When F is greater than the F-critical value, then statistical significance is present. For example, with FID, the F value was 5066.063, which is greater than the F critical value of 3.011. The ANOVA results confirmed that the changes made had a statistically significant effect on all metrics.

Additionally, a t-test between each pair of models was conducted to further confirm this. The p-values obtained from comparing the metric groups between each pair of models (HUST-OBC and Oracle-241, Oracle-241 and Bedroom, and HUST-OBC and Bedroom) were all < 0.01 . This suggests that the dataset changes and mask loss addition created a statistically significant effect on the result, further supporting the effectiveness of these modifications in creating a model for inpainting worn oracle bone characters.

CONCLUSIONS

Oracle bone wear is a major obstacle in decoding the remaining undeciphered oracle bone characters. To aid human researchers' evaluation, this paper presents a diffusion model that uses a U-Net architecture and refinement with mask loss for oracle bone inpainting. When quality is measured with several quantitative metrics, the results presented for the prototype do not yet reach state-of-the-art levels, but do achieve the engineering goals for FID, PSNR, and UIQ laid out at the beginning, which specified 0.98 SSIM, 30 PSNR, UIQ of 0.5, and FID under 30. With the use of HUST-OBC for training and further retraining with mask loss, improvements were seen with FID (38.1%), PSNR (1.74%), and UIQ (122%), but there was a 43.9% decline with SSIM. A visual comparison can be found in Figure 6. Based on the visual quality of the generated outputs, the model achieves stylistic similarity between the real dataset and the generated images. Thus, the presented model does show potential in the application of diffusion models to the task of inpainting oracle bones and contributes to oracle bone deciphering.

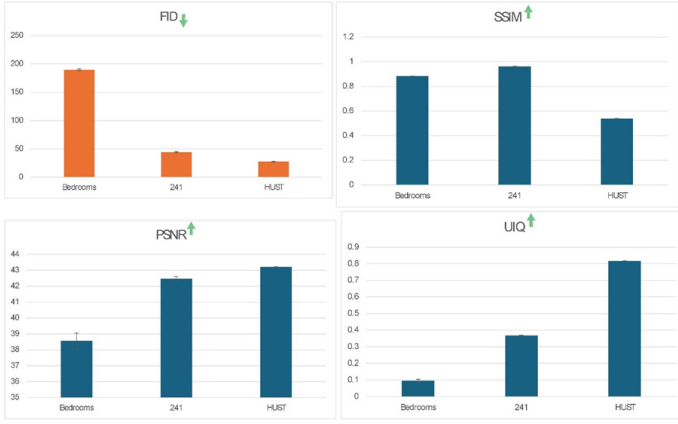


Figure 6. Image Metrics Compared Across Iterations. Graphs depicting the change in numerical metrics over different iterations, along with standard error bars. Metrics that imply higher generated quality when the value is higher are labeled with an up arrow and darker bars. Metrics that imply higher generated quality when the value is lower are labeled with a down arrow and lighter bars.

The SSIM decrease is likely due to the luminance difference evident in the resulting images depicted in Figure 5. This was tested through changing areas that were less than the value of 0.5 (if each pixel ranges from 0 to 1) to 0. While not a precise solution, when those areas were darkened, SSIM increased to over 0.773, suggesting that this is a factor in the lower SSIM. When taking this into account, this prototype can provide an alternative approach to GAN models, as it outperforms the 33.27 PSNR obtained by Wang et al., although it fails to match the 0.9623 SSIM they obtained (2).

When calculating outliers by determining points that fall outside quartile 1 - 1.5 * Interquartile Range to quartile 3 + 1.5 * Interquartile Range, there are outliers present. However, this is in line with the behavior of diffusion models—the outputs vary because of random sampling and the distribution’s complexity can lead to predicted images that deviate from the target image, which creates outliers in the results of comparing the target and predicted images.

The prototype’s design is limited by the data it is trained on. A larger volume of data would help the model represent the range of oracle bone characters fully. Additionally, the mask loss process does not entirely mirror true oracle bone deterioration, as it currently is trained solely with rectangular masks. A more comprehensive method to generate simulated deterioration when training would likely lead to improved results. However, even though training did not

use irregular masks, the model can still be applied in cases where masks are not rectangular, as Figure 7 demonstrates. There are limitations to the metrics used as well. For example, despite the widespread use of SSIM, it does not always capture human perceptual evaluation well, and occasionally can generate unintuitive results, especially when looking at non-greyscale images (16). More investigation into measures like SSIM and FID may reveal weaknesses, so evaluation from multiple numerical measures, as done with the prototype, and further qualitative evaluation from humans is needed to obtain a better picture of the generated image quality.



Figure 7. Example of Irregular Mask Inpainting. Example of inpainting an irregular mask. From left to right: original image, image with mask overlaid in black, and the inpainted image.

Future steps would include further tests with the actual application of the model in the field, with unknown and deteriorated oracle bones, as well as further investigation into improvements to the model architecture. More direct comparisons between diffusion models, generative adversarial networks, and variational autoencoders would benefit decisions by researchers in picking an approach when it comes to restoring oracle bone characters, as this prototype improves upon the PSNR result of one GAN implementation (2). Additionally, to create more cohesive structure between the inpainted region and the existing image, conditional diffusion could embed contextual information. The training data could be labeled based on classes of related characters to improve upon the structural consistency of generated sections. This could be combined with automatic detection of damaged areas on larger pieces of bone, rather than single characters, to further streamline the restoration process.

ACKNOWLEDGEMENTS

I would like to express my thanks to Dr. Yue Wu for his quick advice, patience, and enthusiasm throughout this project.

BIBLIOGRAPHY

- [1] Guan, H., Yang, H., Wang, X., Han, S., Liu, Y., Jin, L., ... & Liu, Y. (2024). Deciphering oracle bone language with diffusion models. arXiv preprint arXiv:2406.00684. <https://arxiv.org/abs/2406.00684>
- [2] Wang, S., Guo, W., Xu, Y., Liu, D., & Li, X. (2024, August). Coarse-to-Fine Generative Model for Oracle Bone Inscriptions Inpainting. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)* (pp. 107-114). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.ml4al-1.12>
- [3] Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2024). *Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers*. IEEE Access.
- [4] Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., ... & Norouzi, M. (2022, July). Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings* (pp. 1-10).
- [5] Wang, P., Zhang, K., Wang, X., Han, S., Liu, Y., Wan, J., ... & Liu, Y. (2024). An open dataset for oracle bone script recognition and decipherment. arXiv preprint arXiv:2401.15365. <https://arxiv.org/abs/2401.15365>
- [6] Wang, M., Deng, W., & Liu, C. L. (2022). Unsupervised structure-texture separation network for oracle character recognition. *IEEE Transactions on Image Processing*, 31, 3137-3150.
- [7] Google. (2022). google/ddpm-ema-bedroom-256. Hugging Face. <https://huggingface.co/google/ddpm-ema-bedroom-256>
- [8] Aromal, M. A. (2024). Inpainting. <https://github.com/aromalma/InPainting>
- [9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III* 18 (pp. 234-241). Springer international publishing.
- [10] Wu, Y. (2024) Glyffuser. <https://github.com/yuehere/glyffuser>
- [11] Ekhtiari, N. (2024). Image Similarity Measures. <https://github.com/nekhtiari/image-similarity-measures>
- [12] Zhou Wang, Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612.
- [13] Wang, Z., & Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3), 81-84. <https://doi.org/10.1109/97.995823>
- [14] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.
- [15] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826). <http://arxiv.org/abs/1512.00567>
- [16] Nilsson, J., & Akenine-Möller, T. (2020). Understanding SSIM. arXiv preprint arXiv:2006.13846. <https://arxiv.org/abs/2006.13846>

Identification and Validation of Key Biomarkers and Pathways for Idiopathic Pulmonary Fibrosis using In Silico Methods

Keerthi Rajesh

ABSTRACT

Currently Idiopathic Pulmonary Fibrosis affects more than 3 million people and early predictive biomarkers research is difficult to standardize because of gene expression variability and lack of standard biomarkers. This study uses bioinformatics to identify key genetic biomarkers for Idiopathic Pulmonary Fibrosis (IPF), which would help improve the current diagnostic approaches. This study looks at the differentially expressed genes (DEGs) to uncover potential biomarkers that could enhance early diagnosis and targeted treatment strategies. Gene expression data from the Gene Expression Omnibus (GEO) database were analyzed using GEO2R to identify the DEGS, which are genes with a $|\log_2\text{FoldChange}| > 2$ and a $p\text{-value} < 0.05$, indicating differential expression between IPF and normal tissue. Diagnostic Accuracy was measured as the area under the receiver operating characteristic (ROC) curve, used to assess the ability of identified biomarkers to distinguish between IPF and normal tissue. Functional enrichment analysis, a method used to identify which biological processes or pathways are significantly represented, was performed using Enrichr. The hub genes were found using STRING and Cytoscape. MedCalc was used to generate ROC curves for biomarker validation. The top two biomarkers identified, TOP2A and MMP1, had Area Under the Curve (AUC) values of 0.964 and 0.938, respectively, with $p\text{-values} < 0.001$, which shows strong diagnostic potential and thus making it statistically valid. This study introduces a computational approach to identifying IPF biomarkers which allows for rapid, reproducible biomarker discovery, reducing the time and cost of research. Additionally, this method can be applied to other diseases, advancing precision medicine and improving early detection strategies across multiple conditions.

INTRODUCTION

This research aims to identify and validate key biomarkers and molecular pathways involved in the progression of Idiopathic Pulmonary Fibrosis (IPF) by using bioinformatics tools to help with early diagnosis and develop targeted therapies.

IPF is a devastating lung disease that is characterized by the formation of excessive scar tissue in the lungs, which leads to respiratory failure and death. The disease disrupts normal lung function, which results in a decline in lung capacity and quality of life for affected individuals. According to the National Institutes of Health (NIH), approximately 100,000 people in the United States have IPF, with 30,000 to 40,000 new cases diagnosed each year. Globally, IPF affects 13 to 20 out of every 100,000 people. This disease has a median survival time of only 3-5 years post-diagnosis, IPF presents a significant challenge to healthcare providers and researchers. The variability in epidemiologic studies contributes to this lack of understanding. Incidence refers to the number of new cases of a disease occurring in a specific population over a particular period. For IPF, the incidence ranges from 0.09 to 1.30 per 10,000 people. Prevalence represents the total number of new and existing cases within a population at a given time. For IPF, preva-

lence estimates range from 0.33 to 4.51 per 10,000 people. Despite its severe impact on the body and high mortality rate, IPF remains poorly understood, and current treatment options are limited and often ineffective.

One of the major problems in treating IPF is the lack of effective biomarkers for early detection and disease monitoring. Identifying key molecular pathways and biomarkers is essential for improving diagnosis, tracking disease progression, and developing targeted therapies. However, current research focuses on large-scale associations rather than identifying specific, statistically validated biomarkers for early detection or targeted treatment.

There have been studies that have attempted to address this issue. For instance, a study on molecular biomarkers in IPF highlighted the potential of certain genes and proteins as biomarkers for the disease (Stainer et al., 2021). This study identified diverse blood biomarkers that are in various pathogenetic pathways, such as fibrogenesis (S100A4), extracellular matrix remodeling (YKL-40, MMP-7, ICAM-1, LOXL2, periostin), chemotaxis (CCL-18, IL-8), and epithelial cell injury (KL-6, SP-A, SP-D). Another study used machine learning analysis to identify potential biomarkers of IPF (Wu et al., 2023). This study identified four potential biomarkers (FHL2, HPCAL1, RNF182, and SLAIN1) us-

ing LASSO logistic regression, random forest, and support vector machine-recursive feature elimination (SVM-RFE) algorithms. Also, a review on molecular and genetic biomarkers in IPF provided a comprehensive overview of the current state of research and identified gaps that need to be addressed (Rao et al., 2014). This review emphasized the need for accurate diagnostic, prognostic, and predisposition biomarkers in clinical practice to ensure quick diagnoses and early treatment.

This research seeks to analyze candidate biomarkers and molecular pathways through in-silico methods. In silico methods refers to the scientific approach that uses computer simulations and computational models to study biological processes. By using biocomputational techniques and tools, this study aims to enhance early diagnosis, monitor disease progression, and contribute to the development of targeted therapies. This research hopes to find key differentially expressed genes (DEGs), which are genes that have significantly different expression levels in different conditions or groups, and molecular pathways involved in the progression of IPF as they can serve as reliable biomarkers for early diagnosis. Specifically, this study will use public gene expression datasets from the GEO Database, GSE10667 to identify and validate DEGs. By employing tools such as GEOquery, DESeq2, Enrichr, Cytoscape, Search Tool for Retrieval of Interacting Genes/Proteins (STRING), and Kyoto Encyclopedia of Genes and Genomes (KEGG), this research will conduct a comprehensive bioinformatics analysis to uncover novel biomarkers and pathways.

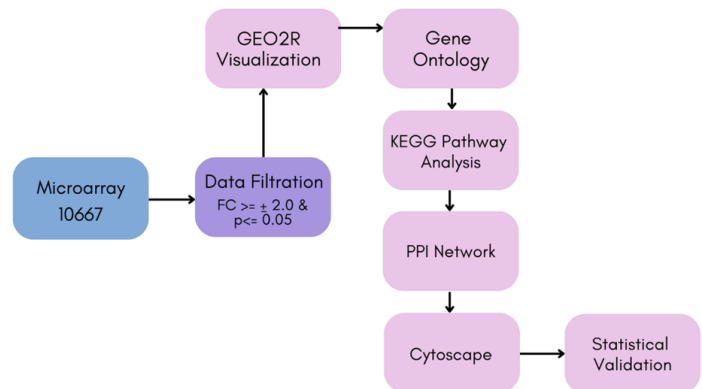
This study aims to identify key differentially expressed genes (DEGs) and molecular pathways associated with IPF using bioinformatics analysis. The hypothesis is that integrating gene expression data from datasets with bio-computational tools will reveal specific, reliable biomarkers for early diagnosis and help with creating better ways to diagnose and treat IPF.

METHODOLOGY

Data Collection

Gene expression datasets related to Idiopathic Pulmonary Fibrosis (IPF) were accessed from the Gene Expression Omnibus (GEO) database. The dataset chosen for this study is GSE10667. This dataset was picked because it contains various samples and gene expression data from both IPF patients and healthy controls. Gene expression profiles were downloaded using the GEOquery package in R, which allows for direct access to gene expression data from the GEO database.

Figure 1: Overview of Study Process

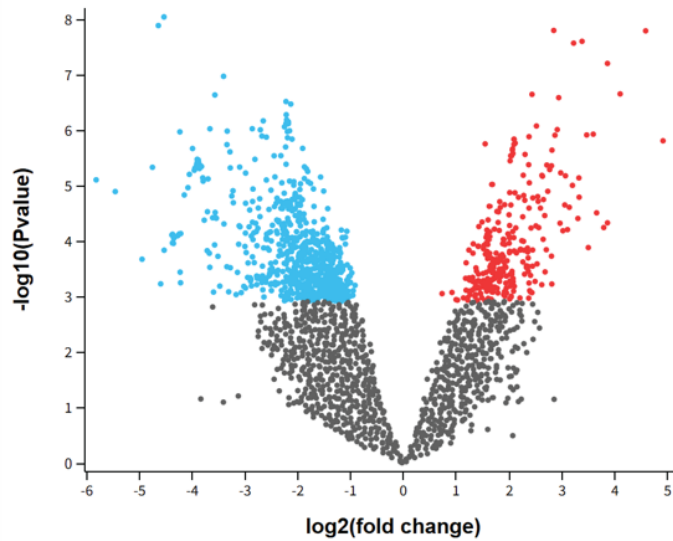


This flowchart provides a visual representation of the overall process and workflow of the study, from data collection to statistical validation.

Data Filtration and GEO2R

The GEO2R tool was used to normalize the downloaded gene expression data. Differentially expressed genes (DEGs) whose expression levels show a statistical difference between a disease state versus a healthy state. DEGs were identified using a threshold of $|\log_2\text{FoldChange}| > 2.0$, $|\log_2\text{FoldChange}| < -2.0$ and an adjusted p-value < 0.05 to ensure statistical significance. Gene expression levels in Idiopathic Pulmonary Fibrosis (IPF) and normal tissue samples were compared, with genes exhibiting a $\log_2\text{FoldChange}$ greater than 2 classified as upregulated and those with a $\log_2\text{FoldChange}$ less than -2 classified as downregulated. These DEGs were then visualized using a volcano plot to highlight genes with significant differential expression. This showed the genes that were upregulated and downregulated. Upregulated genes have an increased expression, meaning they produce more of the corresponding protein. Downregulated genes are those whose expression is reduced or suppressed.

Figure 2: Volcano Plot



Volcano plot of the GSE10667 dataset with the cut-off criteria of $|\log_2FC| > 2.0$, $|\log_2\text{FoldChange}| < -2.0$ and $p\text{-value of } < 0.05$. The blue represents the downregulated genes, the red represents the upregulated genes and the gray is insignificant genes.

Table 1:

Upregulated Genes ITLN2, RTKN2, RTKN2, CHIAP2, AGBL1, FOSB, RTKN2, CRTAC1, LGALS2, LHFPL3-AS2, SERTM1, KLRG2, NECAB1, CD1C, HSD17B6, APOH, AGTR2, ZBED2, MT1E, HHIP, TNNC1, F11, HHIP, DPP6, DDC, BDNF, AGER, C8B, ECEL1P2, CACNA2D2, CLDN18, SFTPA1, PLA2G1B, SGCG, CHIA, RS1, C4BPA, MYH2, LNX2, SFTPD, CCL17, SLC01A2, CA4, CTNND2, CNTN6, FCER1A, CPB2, P3H2, HMGCS2, RBP2, SFTPA2, SFTPA1, LAMP3, CTNND2, LRRN4, GPR158, WIF1, FAM189A2, ODA, WIF1, PEBP4, AADAC, EPB41L5, SFTPA1, ODF3L1, LINC00312, IL13, EMP2, GPM6A, EFR3B, LRRC36, TINCR, ANGPTL7, PLPPR1, LINC00968, XCL1, ASPG, FAM189A2, SFTPA2, KLB, SLC26A9, TNNT1, HDC, SHISA2, SLC6A4, CAV2, CDH16, C2orf54, SLC6A4, SLC5A9, ADRB2, IGFALS, HAGLR, PRTG, ADRB2, ADRB2, ADRB2, FAM71A, GRM8, SLC14A1, SLC6A4, CYP3A4, SLC6A4, CCK, ADRB2, WN-T3A, ADRB2, ADRB2, FAM189A2, SLC46A2, LINC00165, ADRB2, GPIHBP1, RGS9BP, SYNDIG1L, ADRB2, SLC6A4, ADAMTS8, ADRB2, SFTPA1, CLIC5, NR0B2, FGFBP2, HOXA6, RASGRF1, PARD6B, MYRF

Downregulated Genes FLJ26850, RAD54B, GPX8, OTX1, LINC01296, HMGB3, HIST1H3F, ASPM, KIF20A, WISP1, DISC1, HSPA1A, WDR63, ANKUB1, RAB42, IGF1, CCDC151, CHRDL2, CDH2, FAM65C, IGF1, ESCO2, LOC728175, CCDC78, GABBR2, TTK, HIST1H3B, GPX8, IQUB, RAD54L, C11orf97, TSHZ2, ERICH5, DRC7, NEK2, TMEM45A, SLC1A3, CCDC113, PTPRT, IGF1, OGDHL, NMRAL1P1, IGF1, CCNA2, LD-LRAD1, RSPH10B, REEP2, IGF1, DNAI1, IGF1, MMP14, LIMS3-LOC440895, IGF1, ANLN, IGF1, RAD54B, IGF1, SERPINF1, MAP1A, IGF1, CCNA2, EXO1, LOX, SR-

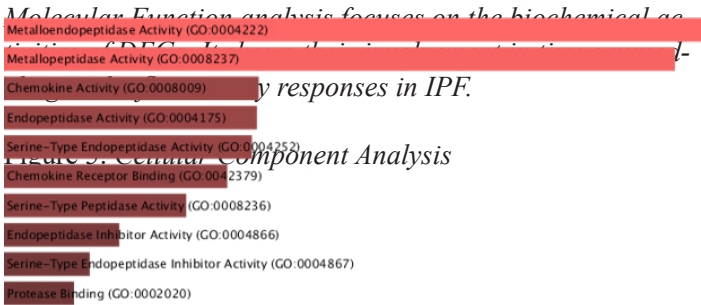
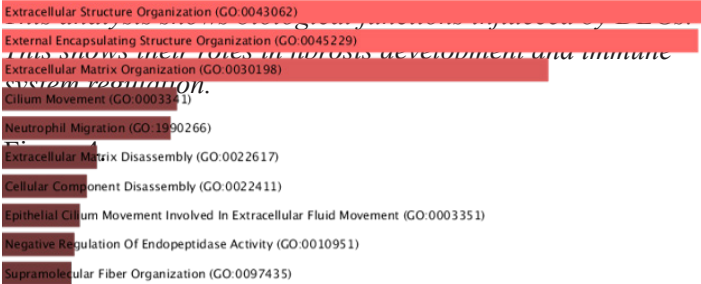
D5A2, SERPINA5, PBK, CD177, NMRAL1P1, SCFV, EYA1, DNAAF1, C9orf135, PLA2G7, FAM30A, TXLNGY, CCDC190, HES2, FAT2, STMND1, COL14A1, C9orf135, TRPV4, FAM81A, E2F7, VNN1, RSRP1, EFCAB1, PAPP, CFAP206, MS4A14, CD163, FCGBP, SNTN, ABCG4, SEC24A, COL5A2, MYO7A, C6orf118, EFCAB1, MELK, TSGA10, CHP2, LOX, ASPN, MORN5, COL17A1, MS4A8, COL5A2, SRPX2, LOX, CHST9, SOWAHA, GPX8, CCDC170, CXCL12, STEAP1, AKAP14, ZNF385D, DIO2, LILRB5, E2F8, LRRC71, LOX, CYP24A1, SPRR1A, LGI2, CCNA2, SFRP2, LOX, VNN1, TSHZ2, CAPSL, CABYR, LOX, CCL13, LOX, LOX, SKA3, SERPINB13, HIST1H2AE, LOX, SPAG6, SRRM3, DRC1, MIA, C1orf194, MUC4, CFAP52, GPX8, TYMSOS, SLC51A, CCL13, ADAM12, SFRP2, CD177, SCRG1, HSPA1A, RNASE2, KRT23, C11orf88, LOX, NPPA, CCDC17, MMP11, CFAP157, SPRR1B, PSAT1, GPX8, MMP7, P4HA3, SOX2, GPX8, NGFR, ADAMTS14, CCNA2, GPX8, GPX8, MCHR1, TMIGD3, GPX8, SLC51A, TUBB3, CRYAB, CCDC187, C20orf85, GTSE1, C6, ATP12A, COL5A2, KIF11, GPX8, C2orf73, SPAG17, TICRR, KDM5D, ADGB, CENPA, ABCA13, TOP2A, GPX8, ADAMTS16, KCNQ1OT1, CILP, CCL18, FAM92B, CFAP100, STEAP1, DLGAP5, VTCN1, ARG1, IL-13RA2, GPX8, GSTA1, UGT2B17, DQX1, CEMIP, LINC01615, MUC16, ARMC3, IGFL2, SLC16A10, IGF1, ZSCAN4, CCNA1, LINC01296, DUXAP10, DUXAP8, DEFA3, DNAH3, C20orf85, STEAP1, TDO2, PIH1D3, FCRL5, GREM1, SAA2, ENKUR, RRM2, TM4SF19, NPTX2, MUCL1, PSCA, EFNB3, CEMIP, CCL7, FGF7, MAP3K19, DPEP1, SPAG17, TTC29, DYDC2, COL3A1, SCG2, MMP16, NEK2, PAPP, MS4A6E, KIF26B, MMP3, CFAP53, PTPRT, LTF, APOD, RARRES1, FAM216B, GALNT15, GSTA2, HTR2A, COL1A1, GSTA5, CCNO, ZBBX, HIST1H2AM, FCRL5, SCG5, HSPA4L, ERICH3, LRRC74B, LGALS7B, RHOF, LGALS7B, POSTN, COL1A1, SPOCK1, MMP13, THBS2, TNS4, F2RL2, P4HA3, CP, CLEC4G, ACAN, SIX4, CLCA2, KRT14, CHST6, LGALS7, COL3A1, S100A2, THBS2, KRT17, PCSK1, COL10A1, UGT1A6, PLA2G2A, HIST1H3D, AKR1B10, TIMP4, GPR87, PROM1, LINC01127, MMP10, LCN2, PLEKHS1, CTHRC1, CRLF1, TSPAN19, AKR1B10, CYP24A1, TMPRSS4, KRT6B, PODNL1, PLEKHS1, CXCL6, SAA4, SAA4, SAA4, COL3A1, PCSK1, KRT5, COL17A1, OLFM4, SNORD14E, KLK12, PIP, PSCA, SAA4, SPP1, SAA2, KRT6CSAA4, SAA1, MSMB, BPIFB1, SAA4, MMP1, COMP, MSMB, COMP, KRT6A, BPIFA1, SERPINB4, SERPINB3

The table above shows a list of the genes that were identified as DEGs from the comparison between IPF patients and healthy control patients.

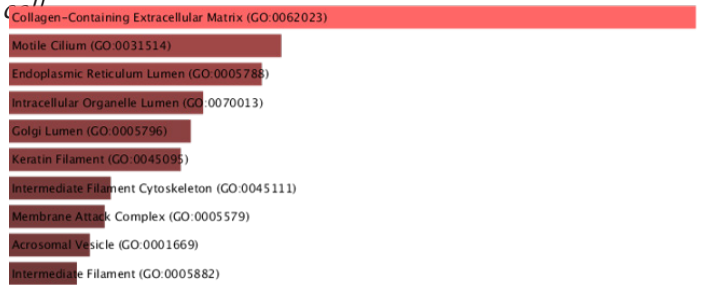
Gene Ontology (GO) Analysis

Gene Ontology(GO) analysis is a bioinformatics method used to classify genes and it helps understand the roles of DEGs in by categorizing them. The list of significant DEGs was uploaded to EnrichR for GO analysis. EnrichR identified overrepresented GO terms among the DEGs, providing insights into the biological processes, cellular components, and molecular functions associated with the DEGs.

Figure 3: Biological Process Analysis



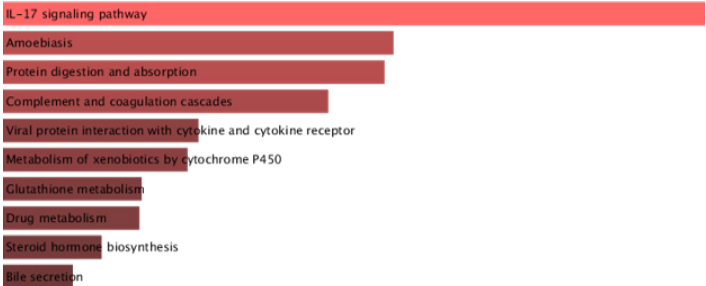
The analysis identifies where DEGs are located within the



KEGG Pathway Analysis

Pathway enrichment analysis was conducted using the KEGG database in EnrichR. The significant DEGs were analyzed to identify enriched pathways involved in IPF. This analysis helped understand the molecular pathways that play a critical role in the progression of IPF.

Figure 6: KEGG Pathway

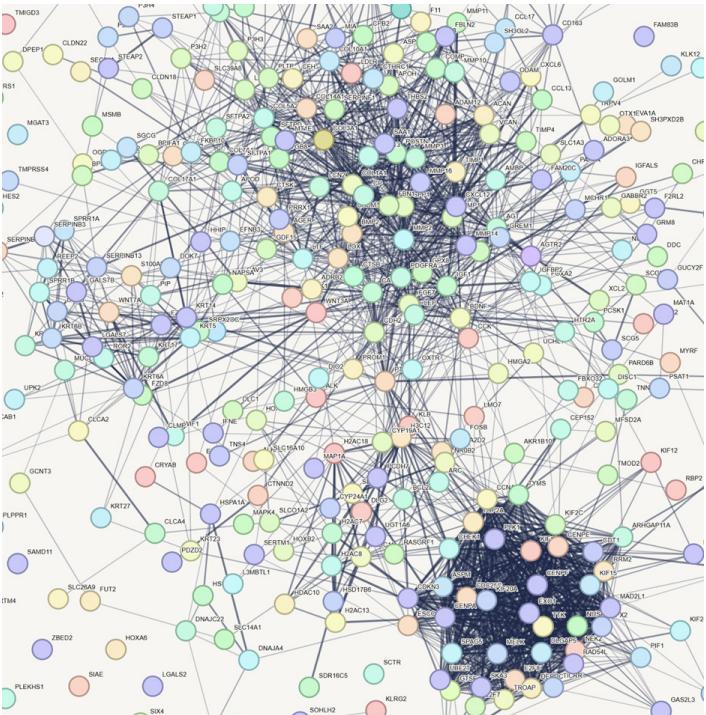


KEGG pathway analysis maps DEG to biological pathways involved in disease progression. This shows key mechanisms driving IPF and helps with potential therapeutic targets.

Protein-Protein Interaction (PPI) Network Construction

The STRING database was used to construct a protein-protein interaction (PPI) network for significant DEGs. The list of DEGs was uploaded to STRING, which generated the network cluster. This network provided insights into the interactions between the proteins encoded by the DEGs and their potential roles in IPF.

Figure 7: PPI Network



A protein-protein interaction network was constructed from STRING using genes from all of the significantly regulated pathways. It was constructed at the highest confidence level (0.900) which helps pinpoint the most connected, or “hub” genes that play a crucial role in disease mechanisms.

Network Visualization

The data from the STRING database were imported into Cytoscape for network visualization and analysis. Cytoscape created visual representations of the PPI networks, allowing for the identification of key nodes and clusters within the network. Cytoscape also identified the degree of connectivity for each biomarker and with the total number of connections and nodes that were in the PPI. Using Cytohubb, the top 10 genes were identified.

Table 2: List of top 10 biomarkers identified

Biomarker Identified
TOP2A
MMP1
CCNA2
KIF11
EXO1
TTK
ECT2
KIF20A
ASPM
RRM2

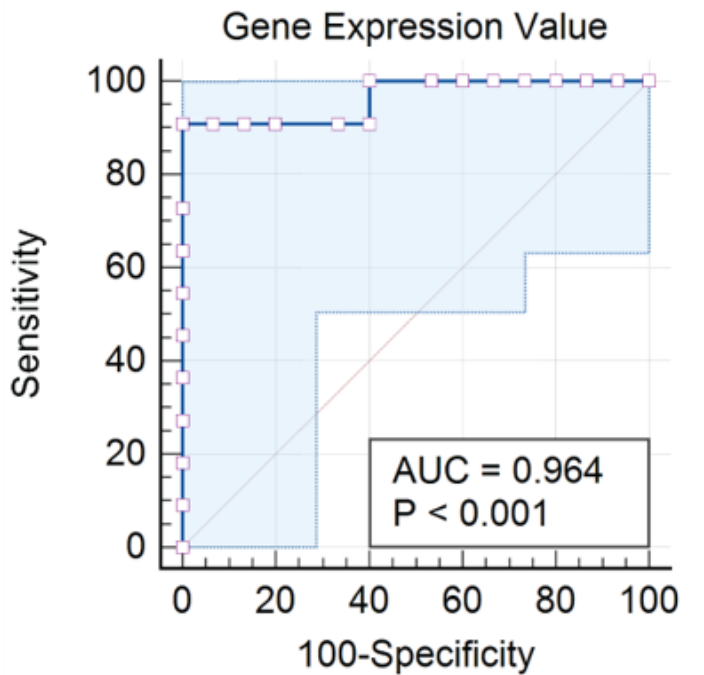
Statistical Validation and Data Analysis

Statistical analysis was performed using MedCalc. Receiver Operating Characteristic (ROC) curve analysis was used to evaluate how well the differentially expressed genes (DEGs) could distinguish between Idiopathic Pulmonary Fibrosis (IPF) and normal tissue samples. The area under the curve (AUC) was calculated to measure the accuracy of these genes as potential biomarkers. A higher AUC value indicated a stronger ability to differentiate between IPF and normal samples, while values closer to 0.5 suggested little to no distinction. This analysis helped confirm the significance of the identified genes in diagnosing IPF.

RESULTS

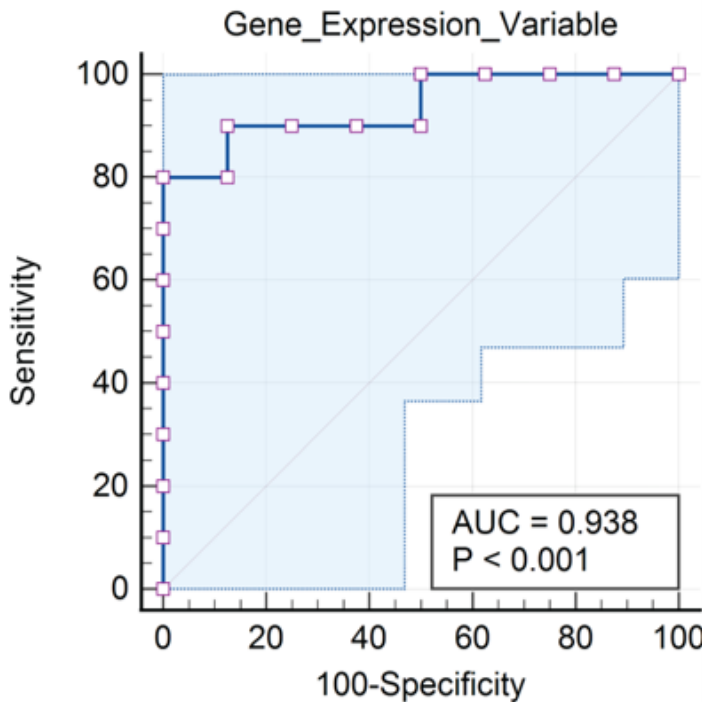
After the identification of the top 10 most connected hub genes, or potential biomarkers, the statistical analysis confirmed the validity of the research findings.

Figure 8: Statistical Analysis of TOP2A



The ROC curves check the diagnostic performance of two key biomarkers in distinguishing IPF from normal tissue. The graph represents TOP2A with an area under the curve (AUC) of 0.964, indicating high diagnostic accuracy. It also has a p-value < 0.001, suggesting its potential use in diagnosing IPF.

Figure 9: Statistical Analysis of MMP1



The ROC curves check the diagnostic performance of two key biomarkers in distinguishing IPF from normal tissue. The graph represents MMP1 with an area under the curve (AUC) of 0.938, indicating high diagnostic accuracy. It also has a p-value < 0.001 , suggesting its potential use in diagnosing IPF.

The analysis identified TOP2A and MMP1 as key biomarkers that can be associated with the prognosis of IPF. The study used bioinformatic tools to analyze gene expression data, highlighting the key biomarkers and significant pathways that are related to the progression of IPF. The analysis found that TOP2A and MMP1 are overexpressed, which is directly related to the severity of the disease and poor survival outcomes. Gene Ontology and KEGG pathways also revealed that these genes are involved in critical biological process, molecular function and cellular component that are linked to fibrosis progression. The study achieved a high level of accuracy in identifying these biomarkers. TOP2A showed an AUC (Area Under the Curve) of 0.964, while MMP1 had an AUC of 0.938, indicating that both genes have a high diagnostic accuracy in distinguishing IPF patients from healthy individuals. Also, both genes exhibited p-values < 0.001 , making the research statistically significant.

Discussion and Future Steps

This study shows that computational tools can be successfully used to identify important disease-related biomarkers. The findings of this research supports the hypothesis that certain biomarkers play a big role in IPF

progression, and that identifying them could help with early detection and future treatment plans. Future research should focus on validating the diagnostic potential of TOP2A and MMP1 through lab testing using real patient samples and this would help confirm the findings. Additionally, there could also be drug research which could explore existing medicines that could target these biomarkers as potential treatment options, which could lead to new and better treatments. Incorporating AI and machine learning models could help with running a much larger database through the same methodology. By incorporating these biomarkers into advanced predictive models, early detection and intervention could be significantly improved. Also, these findings could contribute to the development of personalized treatment plans that are made to individual patients' gene activity, which would improve the effectiveness of these treatments.

BIBLIOGRAPHY

- Mayo Foundation for Medical Education and Research. (2024, February 15). Pulmonary fibrosis. Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/pulmonary-fibrosis/diagnosis-treatment/drc-20353695>
- Valerio, L. G. (2014). In Silico Methods. In *Encyclopedia of Toxicology* (Third, pp. 1026–1029). essay, Academic Press.
- Stainer, A., Faverio, P., Busnelli, S., Catalano, M., Della Zoppa, M., Marruchella, A., Pesci, A., & Luppi, F. (2021). Molecular Biomarkers in Idiopathic Pulmonary Fibrosis: State of the Art and Future Directions. *International journal of molecular sciences*, 22(12), 6255. <https://doi.org/10.3390/ijms22126255>
- Wu, Z., Chen, H., Ke, S., & et al. (2023). Identifying potential biomarkers of idiopathic pulmonary fibrosis through machine learning analysis. *Scientific Reports*, 13, 16559. <https://doi.org/10.1038/s41598-023-43834-z>
- Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. (2014). Protein-protein interaction detection: methods and analysis. *International journal of proteomics*, 2014, 147648. <https://doi.org/10.1155/2014/147648>
- Pirim D. (2020). Integrative analyses of molecular pathways and key candidate biomarkers associated with colorectal cancer. *Cancer biomarkers : section A of Disease markers*, 27(4), 555–568. <https://doi.org/10.3233/CBM-191263>
- Tomos, I., Roussis, I., Matthaïou, A. M., & Dimakou, K. (2023). Molecular and genetic biomarkers in idiopathic pulmonary fibrosis: Where are we now? *Biomedicine*, 11(10), 2796. <https://doi.org/10.3390/biomedicine11102796>
- Zheng, P., Sun, S., Wang, J., Cheng, Z. J., Lei, K. C., Xue, M., Zhang, T., Huang, H., Zhang, X. D., & Sun, B. (2022). Integrative omics analysis identifies biomarkers of idiopathic pulmonary fibrosis. *Cellular and molecular life sciences : CMLS*, 79(1), 66. <https://doi.org/10.1007/s00018-021-04094-0>

Synthesizing Novel CD40xCD137 Bispecific Antibodies for B cell Malignancies and Solid Tumors

Abbas Shakir

ABSTRACT

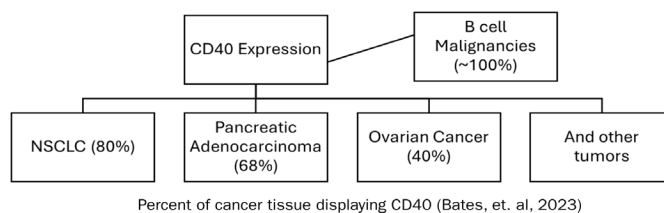
B-cell malignancies and solid cancers pose significant challenges for millions worldwide. Invasive, prolonged, costly, and often ineffective treatments, with high relapse rates reduce patients' quality of life. One promising avenue for treatment is the recent development of bispecific antibodies (BsAbs), which leverage the immune system to combat cancer cells. By simultaneously targeting cancer and immune cell receptors, BsAbs have enhanced cancer-killing capabilities. Unfortunately, current BsAbs also cause harmful off-target cell toxicity, as antibodies bind to receptors found on both cancerous and healthy cells. Therefore, to make binding more specific, this research computationally designs novel bispecific antibodies on alternate receptors. CD40 and CD137(4-1BB) receptors were selected for their cell expression patterns and enhanced cancer-killing mechanisms when targeted together. The 3D structures of the receptors were predicted using AlphaFold3. Antibodies were sourced from a protein data bank, and a deep-learning technique was used to reveal antibodies' 3D structures. Next, binding regions on receptors were predicted using Graph Attention Site Prediction (GrASP). Then, antibody-receptor pairings were simulated and selected using molecular docking, and their binding energies were calculated. These in-silico results revealed binding affinities of $7.1\text{E-}12$ and $1.5\text{E-}12$ M for antibodies 4a6y and 1nlb. This is a significant improvement over current binding affinities for CD40x4-1BB, revealing 4a6y and 1nlb as promising candidates for a novel BsAb drug. These antibodies are now a candidate for in-vitro validation at a local lab. This research also highlights a novel computational framework for accelerating BsAb synthesis and discovery, offering a direction toward more effective, accessible, and safer immunotherapy. The findings contribute to the advancement of targeted cancer therapies, with implications for both B-cell malignancies and solid tumors.

INTRODUCTION

B cell malignancies are diverse blood cancers that form due to abnormal white blood cell creation in bone marrow. Solid tumors are abnormal masses of tissue that form when cells multiply uncontrollably in organs, bones, and other tissues. B-cell malignancies, including lymphoma and leukemia, pose significant clinical challenges due to their impact on immune function and resistance to conventional treatments (Mugnaini & Ghosh, 2016). Current therapies, such as chemotherapy and monoclonal antibodies, often fail due to off-target effects, high relapse rates, and limited efficacy (Shahid et al., 2016).

Bispecific antibodies (BsAbs) represent a promising immunotherapeutic approach by enabling simultaneous binding to both cancer cells and immune effectors, enhancing tumor-specific immune responses (Kontermann & Brinkmann, 2015). However, existing BsAbs exhibit toxicity due to non-specific receptor binding. To improve specificity and efficacy, this research focuses on designing novel BsAbs targeting CD40 and CD137 (4-1BB), two immune-modulating receptors with synergistic anti-tumor potential.

CD40, commonly expressed on B-cells and also several solid tumors (e.g., 80% of NSCLC, 40% of ovarian cancers, and 68% of pancreatic adenocarcinomas) (Vonderheide, 2020), plays a crucial role in antigen presentation and immune cell activation (Saloman & Dahan, 2022). CD137, a co-stimulatory receptor on T and NK cells, enhances immune activation and reduces T-cell exhaustion, a mechanism successfully used in CAR T-cell therapies already (Makkouk, Chester, & Kohrt, 2016; Geuijen et al., 2021). This study leverages computational techniques, including deep learning and molecular docking, to identify high-affinity BsAbs targeting these receptors.



The main research question was: How can bispecific antibodies be designed with greater specificity (stronger binding affinity smaller K_d) in order to enhance cancer immunotherapy? The engineering claim was that to combat current BsAbs therapy limitations, this project will design novel BsAbs targeting CD40 and CD137 that improve specificity. It will create a safer, more effective cancer treatment faster by using in silico computational methods and in-vitro experiments.

Other key research questions included how amino acid sequences in the antibody's variable regions influence their dual-target binding capabilities and which combinations are most effective for the specified receptors. The engineering goals are to design bispecific antibody variable regions using deep learning techniques, explore receptor-antibody interactions through various computational methods, and select optimal antibody candidates based on detailed analyses. Expected outcomes include antibodies interacting exclusively at predicted binding sites, insights into receptor-antibody binding energies and orientations, and optimized bispecific antibodies targeting CD40 and CD137. The created antibody would be used in conjunction with an anti-CD3xTAA bispecific antibody. This is because CD3 is the first step for T-cell activation, and CD137 activation only enhances that mechanism. Ultimately, this research aims to reduce side effects and positively impact immunotherapy for cancers, particularly B-cell malignancies and solid tumors. It also seeks to inform about bispecific antibodies' potential in cancer detection and treatment, along with illustrating the role of computational modeling in their design. This will make cancer treatment more efficient and potentially lower the significant costs associated with cancer detection/treatment in the U.S.

Bispecific Antibodies (BsAbs)

Bispecific antibodies are specially designed proteins that can attach to two different targets at the same time. This ability allows them to connect two types of cells, such as immune cells and cancer cells, helping the immune system attack the cancer more effectively. Compared to regular monoclonal antibodies, bispecific antibodies are more precise in their targeting, which means they are less likely to affect healthy cells and cause side effects (Kontermann & Brinkmann, 2015). For cancer treatment, they can bring immune cells directly to the cancer cells, boosting the body's natural defenses (Kontermann & Brinkmann, 2015). Or both antibodies can target immune cells and cause a response when both cells are co-stimulated. Bispecific antibodies are different from CAR T-cell therapy, which involves taking a patient's T cells, modifying them, and then putting them back into the patient (Kontermann & Brinkmann, 2015). This process is complex and time-consuming. In contrast, BsAbs are pre-made and can be given

through regular injections, making them more accessible and convenient for ongoing treatment.

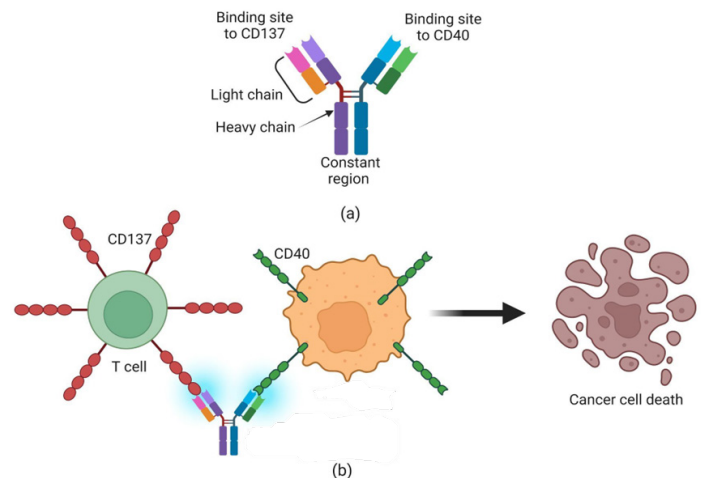


Figure 1: A bispecific antibody. (a) Schematic representation of the bispecific antibody structure, exhibiting a Y-shaped configuration. This structure features two distinct antigen-binding sites associated with separate heavy and light chain pairs. The central heavy chains are connected to light chains specifically engineered to bind to CD40 and CD137 receptors, and (b) Proposed mechanism of action of the bispecific antibody. The BsAb facilitates simultaneous engagement of T-cell receptors and tumor-associated antigens, or two immune receptors thereby enhancing T-cell activation and promoting targeted cytotoxic killing of cancer cells. This interaction amplifies the cytotoxic response against tumor cells, improving therapeutic outcomes.

Figure 1 was created using Biorender.com.

Bispecific antibodies can bind to two different antigens simultaneously. This property can be used in targeted therapy by binding two different cells. This is useful in targeting the killing of cancer cells through cytotoxic killing activated by the immune T-cells. Although the cost of the initial development of bispecific antibodies can be high, the above advantage could make it cost saving overtime by decreasing multiple treatments by drug resistance and fewer hospitalizations. According to the current landscape and future directions of bispecific antibodies in cancer immunotherapy, around one hundred bispecific antibodies are being clinically evaluated, with only four being approved by the FDA as of 2022.

Some BsAb treatments, like Teclistamab, are currently on the market, and they are created by biotech companies. However, significant challenges remain with bispecific antibodies. These include Cytokine Release Syndrome (CRS), immune reactions, drug resistance, manufacturing challenges, short half-lives, and cost of engineering and applications. The ability of bispecific antibodies to target two different receptors is due to the specific combination

of amino acids in the antibody's variable region. This study hypothesized that analyzing factors such as binding energy and binding site of antibody epitopes on T-cell/cancer-cell receptors will help create novel antibody combinations that effectively target B-cell malignancies. So, by using computational methods, the most effective antibody combinations for targeting CD40 and 4-1BB (CD137) receptors can be identified based on factors like binding energy. This was investigated by docking novel antibody sequences on the receptor proteins and then analyzing the antibody-receptor complexes. The outcome is a novel bispecific antibody pair that will be more effective in detecting B-cell malignancies and improving patient quality of life. This research is crucial as it could lead to new, more personalized cancer treatments and improve the overall effectiveness of immunotherapy.

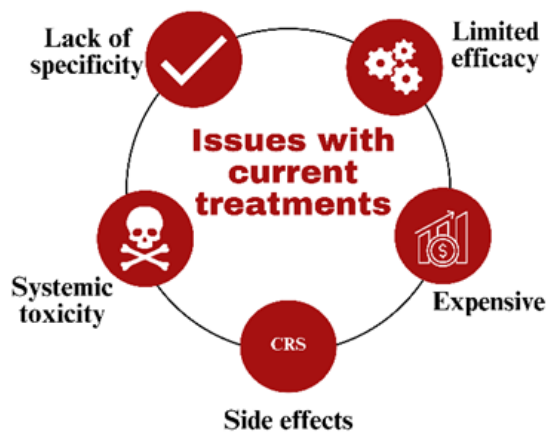


Figure 2: Limitations with current bispecific antibodies.

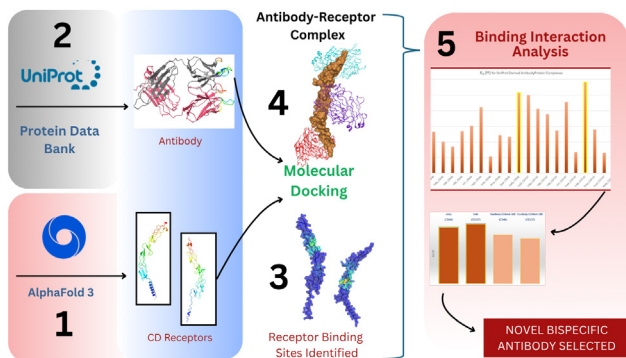


Figure 3: Flowchart describing computational methodology

In this research, two receptors CD40 and CD137 were used. The CD40 receptor, found to be overexpressed prominently on B cells (and certain solid tumor cells), and the CD137 (41BB) receptor, found on T-cell and NK cells. The amino acid FASTA sequences were taken via the UniProt database and Google's DeepMind AlphaFold 3 was used to convert them to 3D structures (Consortium, 2015). The specific binding region on the protein surface was predicted

using the GrASP (Graph Attention Site Prediction) web server, which utilizes a graph neural network to find the specific binding region on the protein surface (Smith, et al 2023).

Antibodies sequences were obtained from the PDB protein data bank. Then the antibodies were designed using a deep learning method called DeepAb (Jeffrey A. Ruffolo et al., 2022). A total of 20 antibodies (10 from the protein data bank and 10 from deep learning) were selected. The reason why DeepAb was chosen is because it is an advanced deep learning method utilizing neural network to process and analyze extensive data. This technology helps to enhance the design and optimization of bispecific antibodies. This approach predicts antibody fragment variable structures with high precision from their amino acid sequences. It helps predict how antibodies interact with antigens by accurately determining their three-dimensional structures plotted in a 3D plane (J. A. Ruffolo, J. Sulam, & J. J. J. P. Gray, 2022). This capability is crucial for designing antibodies with optimal properties, such as high binding affinity and specificity, essential for creating effective therapeutic antibodies against cancer cells (J. A. Ruffolo, J. Sulam, & J. J. Gray, 2022). This helps to significantly reduce the time and cost associated with antibody development, making the creation process for bispecific antibodies efficient.

Molecular docking was done using the HDOCK software to visualize the antibody ends on the CD antigens (Yan, Zhang, Zhou, Li, & Huang, 2017). This program uses an algorithm to predict how the molecules interact most stably. Certain dockings from the model's predictions were picked, based on the orientation and the binding prediction by the GrASP model. First, 3D structures of the antibodies using the ChimeraX software (Pettersen et al., 2021) were visually analyzed, focusing on amino acids that interact with the binding sites of CD40 or CD137 receptors as predicted by GrASP. Then, to understand the antibody binding location, the binding site on the surface of the receptors was predicted using the GrASP (Graph Attention Site Prediction) web server, which predicted the most optimal binding site (Figure 4). The sequences were input, and the outputs were the two molecules with their properties shown by a gradient.

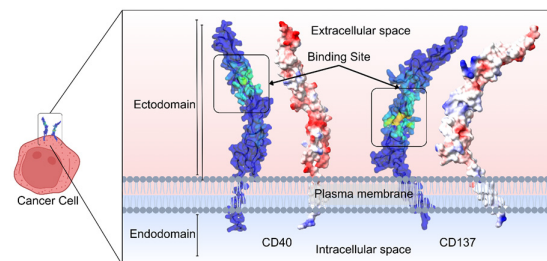


Figure 4: Surface properties of the receptors used in this paper. The binding site of the receptors is shown in yellow green. Electrostatic surface potential (ESP) shows that both receptors are mostly negatively charged.

The protein's electrostatic potential (ESP) was determined using the ChimeraX software, as shown in Figure 4. After selecting the binding sites of the two receptor proteins, a sample of antibody ends to be investigated was obtained. The PDB website initially obtained ten antibody ends (1a5f, 1b4j, 1f8t, 1fl5, 1l1l, 1iqw, 1nlb, 2hrp, 2uyl, 4a6y). So, now that the receptor proteins and antibodies were elucidated, their binding was simulated using molecular docking via the HDock server. This program predicted the structure of the docked antibody-receptor complex. The output was many files that were analyzed visually based on optimal epitope binding to determine the most optimal docking complex. The compilation of chosen antibody-receptor complexes is displayed below in Figure 5.

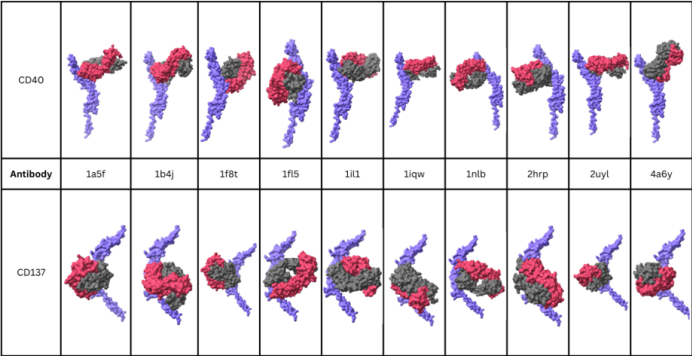


Figure 5: Receptor-antibody complex structures of CD40 and CD137 with the ten initial antibodies. Highlighted in purple are the receptor protein structures of CD137 and CD40. The red chain represents the light chain, and the gray chain represents the heavy chain's structure.

To continue to narrow down the choice for the most optimal antibody pairs, multiple qualitative methods of analysis were also used. First, a web server known as PRODIGY (PROtein binDIng energy prediction) was utilized to predict the binding affinity of the previously docked receptor-antibody complexes. Given the structures in PDB format and the CDR chains, the server provided the predicted binding affinity in units of kcal/mol. The lower the ΔG value, the more thermodynamically favorable and the higher the binding affinity. This would mean the antibody binds to the receptor more tightly and would thus be more effective as a drug. These results were graphed for analysis (Figure 6).

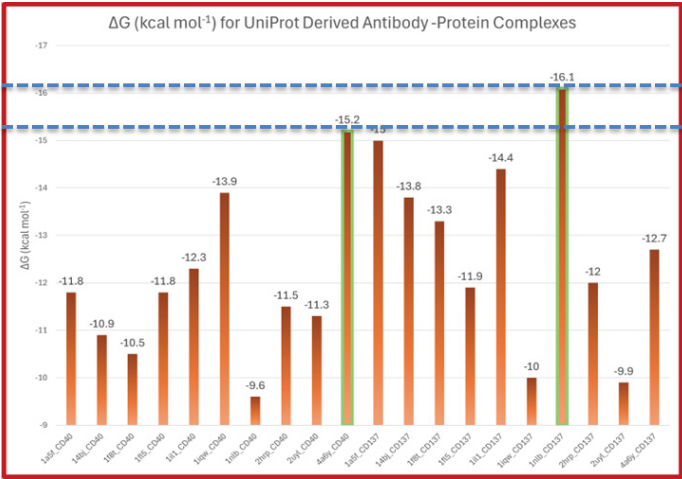


Figure 6: Binding affinities and DG predictions of the antibody-receptor complexes.

This analysis narrowed down the antibodies for the CD40 receptor to 4a6y, 2uyl, 1iqw, 1l1l, 1b4j, and 1a5f. For the CD137 receptor, only 1fl5 was rejected, with the remaining antibodies binding effectively to the receptor's binding site. The receptor-antibody binding energy was calculated using the PRODIGY software. Based on these calculations, antibodies 4a6y and 1iqw were selected, for further analysis for the CD40 receptor and 1nlb and 1a5f for the CD137 receptor.

Candidate Selection

From an initial pool of 20 antibodies, docking analysis revealed strong receptor binding for multiple candidates. The most promising complexes, based on binding affinity and structural alignment, were antibodies 4a6y (for CD40) and antibodies 1nlb (for CD137). These antibodies (Figure 7) demonstrated optimal epitope targeting with minimal off-target binding (Figure 6).

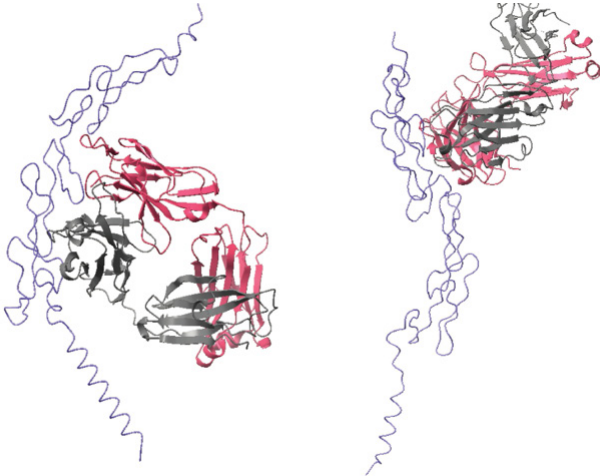


Figure 7: visual representation of final selected receptor-antibody complexes created in ChimeraX.

In-Silico Data Analysis and Interpretation

This study used the Gibbs free energy equation to determine the equilibrium constant:

$$\Delta G = -RT \ln K$$

$$K = e^{-\frac{\Delta G}{RT}}$$

$$K = e^{-\frac{\Delta G}{(1.987 \times 10^{-3}) \times 298}}$$

$$Kd = \frac{1}{K}$$

ΔG (Gibbs Free Energy Change) determines whether the binding interaction is spontaneous. If ΔG is negative, the binding is favorable and if ΔG is positive, binding is unlikely or weak. R (Universal Gas Constant) is the constant value ($\approx 8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} = 1.987 \times 10^{-3} \text{ kcal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$) that relates energy to temperature. T (Temperature in Kelvin) affects the binding strength. So, this research tested the antibody-receptor complexes at both temperatures. K (Equilibrium Constant for Binding Interaction) represents how strongly an antibody binds to its receptor. A higher K means stronger binding (more antibody-receptor complexes form). A lower K means weaker binding (fewer complexes form). And this is converted to a Kd value where a more negative Kd value signifies lower amounts of dissociation meaning high affinity of the antibody to the receptor.

This equation helps analyze binding interactions because it predicts binding strength. By calculating ΔG , it can be determined which antibody-receptor pair has the strongest and most stable interaction. Then it quantifies affinity as the equilibrium constant K is derived from experimental binding data, helping compare different antibodies. A more negative Kd value helps validate the antibody design and indicates its potential as a potential drug candidate. This information is crucial in selecting the most effective bispecific antibodies for therapeutic applications.

RESULTS

- The Novel antibodies 1n1b and 4a6y have significantly better binding affinities to receptor CD40 and CD137 compared to the latest BsAb on the market.
- A computational flow is derived such that iterative process can be used to get better binding affinities as more novel BsAbs are discovered

DISCUSSION AND FUTURE WORK

This experiment aims to enhance B-cell malignancy and solid tumor treatment by utilizing bispecific antibodies, which can simultaneously target multiple receptors on cancer cells, improving efficacy. By applying artificial intelligence models to predict and evaluate antibody structures and combinations, the research seeks to optimize the design of these bispecific antibodies. Ultimately, this approach could lead to more personalized and effective immunotherapy options, reducing costs and improving outcomes for patients with B-cell malignancies. So, by employing computational methods, this research identified the most effective antibody combinations for targeting CD40 and 4-1BB (CD137) receptors in fighting B-cell malignancies and solid tumors. Again, the created antibody would be used in conjunction with an anti-CD3xTAA bispecific antibody due to CD3 being the first step for T-cell activation, and CD137 activation only enhancing that mechanism.

Additionally, in the future, the aim to expand the sample size of the antibody ends, use more computational simulations, and develop the study to target other immune receptors, which could further enhance bispecific antibody applications for various cancers. Simulations like QSAR and ADME models could also be used to investigate the cytotoxicity and drug mechanisms of the BsAbs. To further confirm the antibody binding to the receptors, additional experimental validation will be required to validate the computationally designed antibodies through in vitro and in vivo experiments to assess their efficacy and safety.

Currently this research is being investigated at the Kalia Lab at Seattle Children's Research Center, and the aim is to transition the in-silico results into in-vitro experiment using murine cells. An analysis has been done, and results show that the homologies of the CD receptor proteins are determined to be similar for in-vitro mouse model testing. The cell linings are being developed accordingly. The future direction is to confirm further the binding affinities and functional activity using in-vitro assays like Incucyte® tumor killing assay (Sartorius, 2024). The road to preclinical development will need the use of animal models to assess therapeutic efficacy and safety before clinical translation.

Thus, this study successfully designed bispecific antibodies targeting CD40 and CD137 using deep learning and molecular docking simulations. Twenty antibodies were evaluated, of which four had stronger binding affinity and specificity. Finally, antibodies 4a6y and 1n1b were identi-

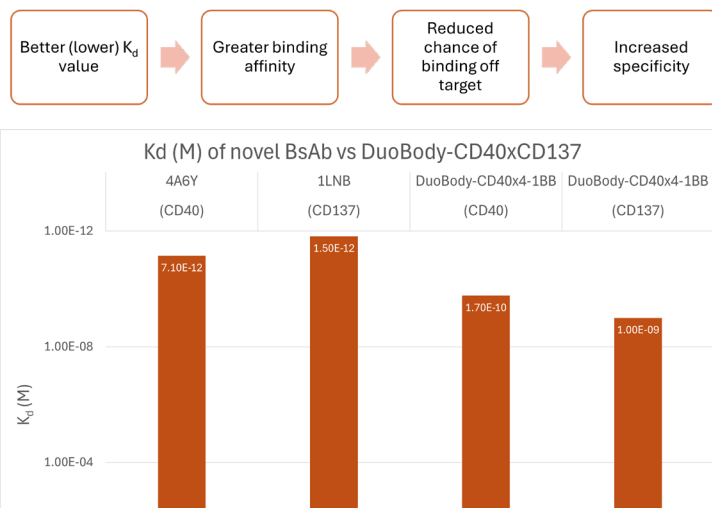


Figure 8: Analysis of K_d values and comparison to current state-of-the-art DuoBody® Bispecific Antibodies (Muik et. al, 2022).

fied as promising candidates that can bind to the CD40 and CD137 (41BB) receptors, respectively. This research shows significant promise for effective, accessible, and safe cancer immunotherapy and demonstrates the faster designing capabilities of computational methods. It paves the way for antibody design and optimization that will hopefully lead to an effective cancer-targeted immunotherapy, a small step towards redefining the field of cancer treatment.

REFERENCES

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., ... & Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3.
- Bates, K. M., Vathiotis, I., MacNeil, T., Ahmed, F. S., Aung, T. N., Katlinskaya, Y., ... & Gavrielatou, N. (2023). Spatial characterization and quantification of CD40 expression across cancer types. *BMC cancer*, 23(1), 220.
- Bellaye, P. S., Georgievski, A., Ballerini, P., Bouslama, B., Richard, C., Boidot, R., ... & Qu  r  , R. (2025). CD40 agonist on patient-derived xenograft mice for the treatment of B-cell acute lymphoblastic leukemia. *Clinical Cancer Research*, 31(1), 181-196.
- Geuijen, C., Tacke, P., Wang, L.-C., Klooster, R., van Loo, P. F., Zhou, J., Mondal, A., Liu, Y., Kramer, A., Condamine, T., Volgina, A., Hendriks, L. J. A., van der Maaden, H., Rovers, E., Engels, S., Fransen, F., den Blanken-Smit, R., Zondag-van der Zande, V., Basmeleh, A., ... Throsby, M. (2021, July 21). A human CD137  PD-L1 bispecific antibody promotes anti-tumor immunity via context-dependent T cell costimulation and checkpoint blockade. *Nature Communications*. <https://doi.org/10.1038/s41467-021-24767-5>
- H  gerbrand, K., Varas, L., D  ronic, A., Nyesiga, B., Sundstedt, A., Ljung, L., Sakellariou, C., Werchau, D., Thagesson, M., Gomez Jimenez, D., Greiff, L., C  lander, M., Smedenfors, K., Ros  n, A., B  l  kbas, D., Carlsson, F., Levin, M., S  ll, A., von Schantz, L., Lindstedt, M., ... Ellmark, P. (2022). Bispecific antibodies targeting CD40 and tumor-associated antigens promote cross-priming of T cells resulting in an antitumor response superior to monospecific antibodies. *Journal for immunotherapy of cancer*, 10(11), e005018. <https://doi.org/10.1136/jitc-2022-005018>
- Huehls, A. M., Coupet, T. A., & Sentman, C. L. (2015). Bispecific T-cell engagers for cancer immunotherapy. *Immunology and cell biology*, 93(3), 290-296. <https://doi.org/10.1038/icb.2014.93>
- Kim, J., McFee, M., Fang, Q., Abidin, O., & Kim, P. M. (2023). Computational and artificial intelligence-based methods for antibody development. *Trends in pharmacological sciences*, 44(3), 175-189. <https://doi.org/10.1016/j.tips.2022.12.005>
- Kontermann, R., & Brinkmann, U. (2015). Bispecific antibodies: Mechanisms and therapeutic potential. *Leukemia & Lymphoma Society*. (2024). Facts book: Blood cancer statistics (PS80). *Leukemia & Lymphoma Society*. https://www.lls.org/sites/default/files/2024-09/PS80_FactsBook_2024.pdf
- Lorenzo, Falchi, Santosha, A. Vardhana, Gilles, A. Salles. (2023, February 2). Bispecific antibodies for the treatment of B-cell lymphoma: Promises, unknowns, and opportunities American Society of Hematology. *Blood*. <https://doi.org/10.1182/blood.2021011994>
- Ma, J., Mo, Y., Tang, M., Shen, J., Qi, Y., Zhao, W., ... & Qian, C. (2021). Bispecific antibodies: from research to clinical application. *Frontiers in immunology*, 12, 626616. <https://doi.org/10.3389/fimmu.2021.626616>
- Madsen, A. V., Pedersen, L. E., Kristensen, P., & Goletz, S. (2024, January 24). Design and engineering of bispecific antibodies: Insights and practical considerations. *Frontiers*. <https://doi.org/10.3389/fbioe.2024.1352014>
- Mugnaini, E., & Ghosh, N. (2016). Lymphoma and leukemia: Pathogenesis and treatment approaches.
- Muik, A., Adams 3rd, H. C., Gieseke, F., Altintas, I., Schoedel, K. B., Blum, J. M., ... & Jure-Kunkel, M. (2022). DuoBody-CD40x4-1BB induces dendritic-cell maturation and enhances T-cell activation through conditional CD40 and 4-1BB agonist activity. *Journal for Immunotherapy of Cancer*, 10(6). <https://doi.org/10.1136/jitc-2021-004322>
- Nakaima, Y., Watanabe, K., Koyama, T., Miura, O., & Fukuda, T. (2013). CD137 is induced by the CD40 signal on chronic lymphocytic leukemia B cells and transduces the survival signal via NF-  B activation. *PLoS One*, 8(5), e64425. <https://doi.org/10.1371/journal.pone.0064425>
- National Cancer Institute. (n.d.). Cancer statistics. U.S. National Cancer Institute. <https://www.cancer.gov/about-cancer/understanding/statistics>
- Parrondo, R. D., Ailawadhi, S., & Cerchione, C. (2024). Bispecific antibodies for the treatment of relapsed/refractory multiple myeloma: updates and future perspectives. *Frontiers in Oncology*, 14, 1394048. <https://doi.org/10.3389/fonc.2024.1394048>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., ... & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science*, 30(1), 70-82. <https://doi.org/10.1002/pro.3943>
- Roncagliani, A., Toropov, A. A., Toropova, A. P., & Benfenati, E. (2013). In silico methods to predict drug toxicity. *Current opinion in pharmacology*, 13(5), 802-806. <https://doi.org/10.1016/j.coph.2013.06.001>
- Ruffolo, J. A., Sulam, J., & Gray, J. J. (2021). Antibody structure prediction using interpretable deep learning. *Patterns (New York, N.Y.)*, 3(2), 100406. <https://doi.org/10.1016/j.patter.2021.100406>
- Salomon, R., & Dahan, R. (2022). Next generation CD40 agonistic antibodies for cancer immunotherapy. *Frontiers in immunology*, 13, 940674. <https://doi.org/10.3389/fimmu.2022.940674>

Sartorius. (2024). Incucyte® Immune Cell Killing of Tumor Spheroids Assay. <https://www.sartorius.com/en/applications/life-science-research/live-cell-assays/immune-cell-killing>

Smith, Z., Strobel, M., Vani, B. P., & Tiwary, P. (2024). Graph attention site prediction (grasp): Identifying druggable binding sites using graph neural networks with attention. *Journal of chemical information and modeling*, 64(7), 2637-2644. <https://doi.org/10.1101/2023.07.25.550565>

Su, Z., Almo, S. C., & Wu, Y. (2024). Computational simulations of bispecific T cell engagers by a multiscale model. *Biophysical journal*, 123(2), 235-247.

Wei, J., Yang, Y., Wang, G., & Liu, M. (2022, October 14). Current landscape and future directions of bispecific antibodies in cancer immunotherapy. *Frontiers in Immunology*. <https://doi.org/10.3389/fimmu.2022.1035276>

Xue, L. C., Rodrigues, J. P., Kastiris, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, 32(23), 3676-3678. <https://doi.org/10.1093/bioinformatics/btw514>

Yan, Y., Tao, H., He, J., & Huang, S. Y. (2020). The HDock server for integrated protein–protein docking. *Nature protocols*, 15(5), 1829-1852. <https://doi.org/10.1038/s41596-020-0312-x>

ENGINEERING

Affordable Distortion Otoacoustic Emission Measurement Device for Hearing Screening

Kaya Hurmali

ABSTRACT

Hearing loss, while preventable if detected at a young age, can cause depression and anxiety, along with an average loss of \$270,000 over a lifetime to people who have it. This distortion product otoacoustic emission measurement device will reliably detect hearing loss with a less expensive price tag and improve reliability when compared to previous models (controls) while also measuring at or above benchmarks used for the controls. The device went through 3 iterations (single speaker, wireless, software and user interface) and the signal to noise ratio along with the return signal (dependent variable) at different pure tone frequencies (independent variable) were measured on each iteration. This device accurately measured and located the return otoacoustic emission to then classify the measured ear as a pass or refer, calculated by open-source software (classified pass correctly in 8/10 trials) with an average signal to noise ratio of 10. Measurements were at a reliability benchmark like the control (needs to be below a signal to noise ratio of 15) to be accepted as a pass or refer while also being able to consistently accurately classify an ear as a pass. Results show the engineering goals of measuring at or above the benchmarks of other distortion product otoacoustic emission measurement devices were met. Hearing screening can be done for a much lower price tag compared to current commercial models while also being just as reliable and 15% more comfortable, potentially allowing millions more people access to early hearing screening and preventable permanent hearing loss.

INTRODUCTION

The purpose of this paper is to describe a wireless hearing screening device using distortion product otoacoustic emissions.

The world population is expanding rapidly, disproportionately so in underdeveloped regions. This rapid population growth often exacerbates issues of poverty and illness, as these regions struggle to provide adequate resources and infrastructure to support the growing population. Kenya is a prime example of this. With a projected 58% population increase by 2050, it stands as one of the fastest growing countries in the world. It is also among the poorest. Kenya was ranked 95th on the global competitiveness ranking by the World Economic Forum (WEF 2019), highlighting its economic challenges, and placing it squarely as a developing country. The WEF report ranked Kenya below nations like Jamaica and Sri Lanka which are known to be politically unstable. The expected population boom will stretch the already limited resources. This strain on resources can cause significant medical supply chain disruptions, especially for specialized and often expensive medical devices. Because otoacoustic emission hearing screening devices can often be costly and need trained professionals to operate, financial constraints in LMIC's means that those devices are often unavailable, leading to universal hearing screening to

significantly lag behind the rest of the developed world.

The healthcare system in LMIC's faces significant hurdles. In 2020, the number of medical professionals in Kenya was 30.14 per 10,000 people, which is only about 68% of the Sustainable Development Goals (SDG) index goal of 44.5 healthcare professionals for every 10,000 people in the population (Okoroafor 2022). This means that there are not nearly enough medical workers for a population of Kenya's size, showing the need for cheaper and easier to use devices to decrease the need for specialized doctors which are in short supply. This disparity shows the urgent need for Kenya to align its future production of healthcare professionals in terms of both quality and quantity to meet the population's health needs. Addressing these challenges is critical for improving the overall well-being of Kenya's rapidly growing population and fostering sustainable development in the region.

Two thirds of individuals with hearing loss reside in low to middle income countries, or LIMCs. Nearly 3.6 percent of the population in Kenya has hearing loss according to the Kenya Demographic and Health Survey Key Indicators report of 2022 (Kenya DHS of 2022). Kenya does not have access to enough doctors, and devices for hearing screening are too expensive – costing upwards of 5 thousand dollars for some commercial models.

Otoacoustic emissions (OAE) are sound waves generated from a specific portion of the ear called the cochlea. When 2 different sound waves interact with tiny hairs on the cochlea, a return signal is generated at the tone of 2 times one frequency minus the second frequency. These tones can be measured in the ear through commercial microphones like the ones used in this project. Generating both sound waves from a single speaker can lead to intermodulation. Intermodulation can be problematic when measuring using DPOAEs. Intermodulation is an amplitude modulation when 2 frequencies are measured at the same time. Intermodulation in single speaker designs is caused by cubic distortions from low quality speakers. DPOAE software (Chan 2024 source code) can be used to classify returning OAEs as either a pass or a refer.

Previous research has shown it is possible to create devices that can be both cheap and accurate to provide hearing screening. Studies like Dr. Chan or Dr. Gallagher's have demonstrated the potential for affordable devices that do not compromise on accuracy. However, these devices have certain limitations, such as reduced stability and lower patient confidence (no stats given, addressed as limitation of design) when compared to newer, more advanced wireless single-speaker devices. Although a significant amount of research has been done, there remains a gap in the study of using a single speaker with dual frequencies. This gap exists because intermodulation in these systems can cause "ghost frequencies," (frequencies that are not generated from the ear but show up on a fast furrier transform graph of the microphone input) which complicates the measurement of otoacoustic emissions (Chan 2022)

The engineering goal of this project is to prototype designs for hearing screening devices while also enhancing comfort and robustness of the device and thereby improving patient confidence in the design. By developing a wireless otoacoustic emission probe that utilizes a single speaker model instead of the wired dual speaker models currently available, the aim is to enhance the system's robustness while maintaining low costs. This approach hypothesizes a more user-friendly (more comfortable – data taken from survey) and reliable (less likely to break) solution, improving patient confidence and overall device performance without the high price tag associated with current commercial models.

METHODOLOGY

The prototype design and experimental procedure are outlined here to make sure the device is successfully tested. The purpose of the device is to be a wireless hearing screening device using distortion product otoacoustic emissions. The hypothesis and engineering goals focus on achieving similar measurements (within statistical significance) to those from a control device, which would change depending on the person. Other goals include enhancing

comfort, increasing patient confidence, creating a more robust system, and improving scalability.

To achieve this, various materials will be used, including a Lenovo Legion I9 computer, Android Studios, multiple Samsung A11 and A14 devices, a multimeter, an otoscope, Microsoft Excel, cardboard, and a CC2640R2F integrated circuit. The operational definitions for the study include defining the experimental group as a home-built distortion product otoacoustic emission measurement device and the control group as the previously made wired distortion product otoacoustic emission measurement device. Variables such as the volume of speakers in the DPOAE device, specificity, and sensitivity will be measured and compared as seen in Table 1.

The engineering procedure begins with designing the initial prototype and testing it on the researcher's ear using DPOAE software (Chan 2024) from Android studios to compare results with baseline measurements from the original device. The prototype will undergo several iterations. The first iteration will involve improving build quality through experimenting with different sizes and earpieces. The second iteration will involve the incorporation of a single speaker to improve the cost effectiveness of the solution while improving its simplicity. The third iteration's goal will be to make the device wireless using Bluetooth Low Energy (BLE) technology connected to a microcontroller (MCU). Each iteration involves testing the device and comparing the results with the control OAE measurement device using DPOAE software data.

Participants for the study were gathered by posting fliers around 3 local high schools in the Lake Washington school district (Eastlake, LW, and Redmond), Seattle Children's, and Microsoft HQ in Redmond. The selection process was randomized using a random number generator to pick eliminate every 10th reply, with a sample size of at least 20 participants, including individuals with minor to moderate prior hearing loss. All participants will be adults of any age. This will be a double-blind procedure because both the experimenter and the participants of the study will not be aware of each other, and they will not be told whether they are in the experimental or control group. Informed consent will be asked for when signing up for the study.

The experimental process began by recruiting a group of adult participants. A random sample was used to ensure that the sample would be partially representative of the world population. Participants were screened for any medical conditions or factors that could interfere with the results, such as ear infections or excessive earwax.

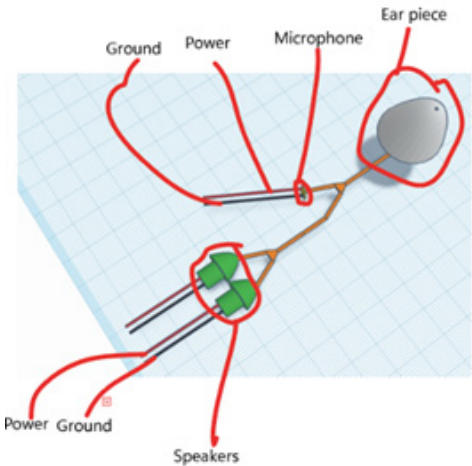
Before the actual testing, the research team calibrated the prototype DPOAE devices to ensure precision and consistency in measurements. Calibration involved comparing the device's output to known standards and making necessary adjustments. Volume on tips were adjusted to fit the specific phone model being used, allowing results to be more

precise. Devices were also cleaned between uses and ear tips were disposed of and replaced.

During the testing phase, participants were asked to sit in a quiet, sound-treated room to minimize external noise interference. The DPOAE device probe was gently inserted into each participant’s ear canal, ensuring a proper fit to achieve accurate results. The device then emitted tones at specific frequencies and measured the otoacoustic emissions generated by the cochlea in response to these sounds. Data was recorded, focusing on key metrics such as signal-to-noise ratio and emission amplitudes to determine the presence of hearing responses.

The collected data were analyzed to evaluate the prototype’s effectiveness in detecting hearing loss compared to standard clinical hearing screening methods. Feedback from participants regarding their experience with the device was also gathered to assess its comfort and usability. The results provided valuable insights into the prototype’s performance, helping identify areas for improvement and guiding further development of the technology.

Figure 1.1: Original OAE device



This design shows the wiring and tubing of the original OAE design. Created in TinkerCAD.

The experimental procedure involves verifying both the control and experimental groups using the DPOAE app by comparing Fast Fourier Transforms of the data and recording baseline measurements without OAE responses. Results in data table 1 between the two groups will be compared using a soundproofed test setup (figure 1.2) to measure output from the speakers and input from the microphone. An electrical board – a standard Arduino breadboard - will connect all the cables, and measurements will be taken to ensure that the device meets the predefined acceptance criteria (results are above 55 micro decibels)

for quality control. This process will be repeated for all prototypes to ensure reliability and consistency.

Figure 1.2: Soundproofed test setup

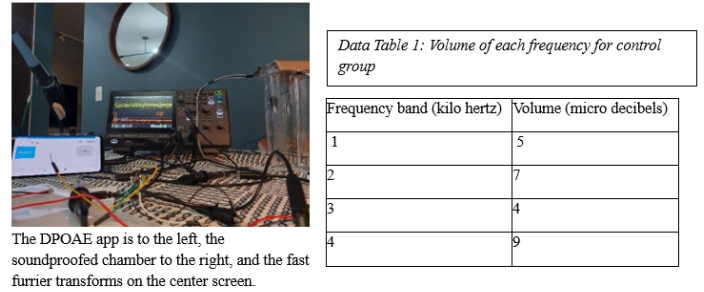
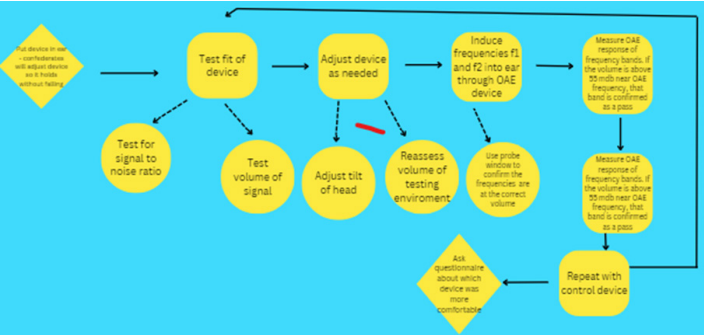


Figure 1.3: Flow chart of device testing



This image shows the consecutive steps in the design process along with possible routes it could take, indicated by dotted arrows.

CONCLUSION

The engineering goals, that the prototype would perform at lower cost, increased reliability, while also being mass producible and providing similar accuracy in results to the control were partially met. There was a 15.8% increase in the comfort rating of this device compared to the control. In addition, the prototype was made 84.4% smaller in size (reduction of a portable size of 160 cm3 of the control to 25 cm3 of the prototype) to the control and a 91.6% decrease to a random sample of a commercial device (296 cm3 to 25 cm3). The random sample was taken from the 10 first results on Amazon, and a random number generator was used to pick the sample.

The comfort rating was measured by a comfort survey that participants were asked to take, shown in Figure 1.4. The survey asked participants to rate how comfortable they felt each device was and took a period of 5 minutes to administer. It was conducted by a confederate, and the experimenter was not in the room when the survey was conducted.

Size was measured through x, y, z measurements for storage when shipping. They were measured using a ruler and given Amazon measurements for commercial distortion product otoacoustic emission hearing screening devices.

The data analysis was conducted using Microsoft Excel, where all the collected data was organized and stored in an Excel spreadsheet for ease of access and processing.

Descriptive statistics such as the mean, median, and mode were calculated for the comfort ratings, which were recorded as 7, 8, 8, 7, and 9. The mean provided the average comfort rating, the media represented the middle value in the sorted dataset, and the mode identified the most frequently occurring rating. Together, these measures provide an empirical overview of the comfort prototype created for this project and various controls. Results are shown in Figure 1.5 and Figure 1.6.

To evaluate variability, the standard deviation was calculated to measure the dispersion of the comfort ratings around the mean. Additionally, the standard error was determined to estimate the accuracy of the mean as a representation of the population’s comfort level.

The results were visualized using a bar chart to facilitate comparison with a control dataset. Error bars were added to the chart to represent variability and standard error, enabling an assessment of statistical significance. The comparison revealed a statistically significant increase in the comfort rating of the device, indicating that the intervention or change under investigation positively impacted user comfort.

The design was not perfect. Some limitations to the design were surrounding where and how the data could be collected. Data must be collected in an isolated and quite environment. This may prove challenging in the target environment, where medical facilities may have a hard time accessing those types of environments. In addition, the probe must be positioned in the ear correctly, and fluctuations in the angle of the probe and depth that it is inserted can vary results. It can cause uncertainty in the results for the participants and medical provider. Attempts to correctly place the probe can also be time consuming, delaying the process for what should generally be a 5-minute procedure. Finally, the prototype for this project had a higher false positive rate. The control had a false negative rate of 10%, while the false negative rate was 20% for the prototype.

False negative rate was most likely caused from resonance in between the sounds generated from the dual usage of 2 piezo-electric plates. The resonance is called intermodulation and can lead to messy results because it occurs where an otoacoustic emission would be expected to occur. Figure 1 shows an example of intermodulation caused by flipping positive and negative voltage on a voltage splitter.

Further steps for this project will occur in 2 major steps. In the short term, more participants must be gathered, with more testing conducted, especially those who are deaf and hard of hearing. The accuracy of the device in all environments, not just an isolated and quite environment, can be important to increase patient confidence with the device. Finally, it is necessary to find a manufacturer to mass produce devices, allowing for comparison from devices of the same model. In the long term, funding will be used to attempt to gain FDA approval and commercialize the device, allowing

it to be spread to medical centers in areas that need them. \

Figure 1.4: Control and prototype comfort survey. Provided on paper to participants.

Figure 1.6: Average prototype and control score on a 1 to 10 comfort scale.

1. Rate the comfort of the first device on a scale of 1 – 10.

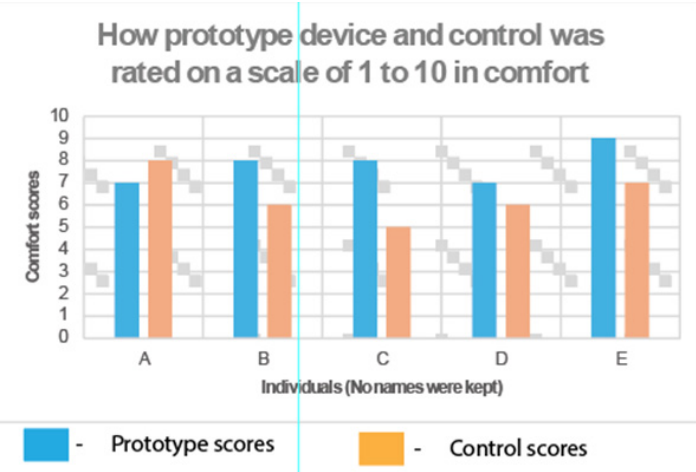
1 2 3 4 5 6 7 8 9 10

2. Rate the comfort of the second device on a scale of 1 – 10.

3. 1 2 3 4 5 6 7 8 9 10

This image provides the comfort survey used by participants to rate the comfort of devices. It was administered in silence over a five-minute duration by a confederate. The experimenter was not present while the survey was administered.

Figure 1.5: How prototype and control are rated on a scale of 1 to 10 in comfort.



The image shows scores between prototype and control rating each of their comforts according to survey participants.

Chan, J., Randall Bly. et al. An off-the-shelf otoacoustic-emission probe for hearing screening via a smartphone. *Nat. Biomed. Eng* 6, 1203–1213 (2022). <https://doi.org/10.1038/s41551-022-00947-6>

Lonsbury-Martin, Brenda & Whitehead, Martin & Martin, Glen. (1991). Otoacoustic

Emissions: Clinical Applications. *Journal of speech and hearing research*. 34. 964-81.

[10.1044/jshr.3405.964](https://doi.org/10.1044/jshr.3405.964)

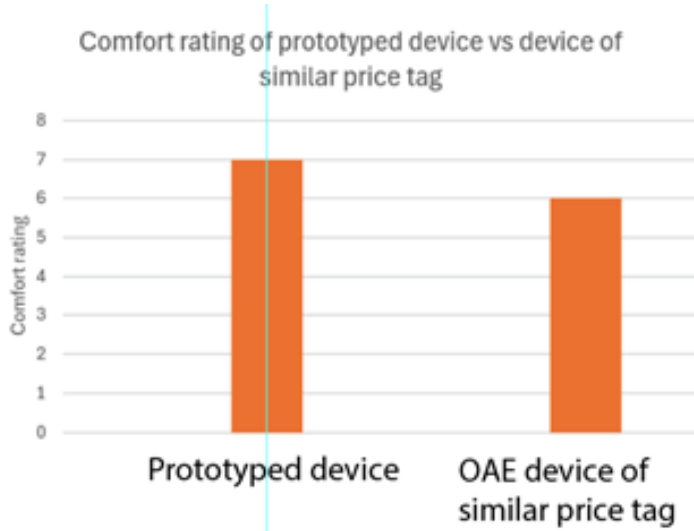


Figure 1.6 shows the OAE comfort rating average for control and prototype.

BIBLIOGRAPHY

Okoroafor SC, Kwesiga B, Ogato J, Gura Z, Gondi J, Jumba N, Ogumbo T, Monyoncho M, Wamae A, Wanyee M, Angir M, Almudhwahi MA, Evalyne C, Nabyonga-Orem J, Ahmat A, Zurn P, Asamani JA. Investing in the health workforce in Kenya: trends in size, composition and distribution from a descriptive health labour market analysis. *BMJ Glob Health*. 2022

Aug;7(Suppl 1): e009748. doi: 10.1136/bmjgh-2022-009748. PMID: 36008084; PMCID: PMC9422806.

KNBS and ICF. 2023. Kenya Demographic and Health Survey 2022: Volume 1. Nairobi, Kenya, and Rockville, Maryland, USA: KNBS and ICF. Available at <https://www.dhsprogram.com/pubs/pdf/FR380/FR380.pdf>.

Abdala C, Luo P, Guardia Y. Swept-Tone Stimulus-Frequency Otoacoustic Emissions in Human Newborns. *Trends in Hearing*. 2019;23. doi:10.1177/2331216519889226

Chang, Justin. (2024). DPOAE4 [Source code] justinklchan/DPOAE4 ([github.com](https://github.com/justinklchan/DPOAE4))

Chan, J., Ali, N., Najafi, A. et al. An off-the-shelf otoacoustic-emission probe for hearing screening via a smartphone. *Nat. Biomed. Eng* 6, 1203–1213 (2022). <https://doi.org/10.1038/s41551-022-00947-6>

Classification of ADHD Through Machine Learning Models in Children Using Traditional Statistical and CNN Based Feature Extraction of Raw EEG Data

Aditi Kandiah and Sarah Youn

ABSTRACT

Attention-Deficit Hyperactivity Disorder (ADHD), is currently diagnosed using the Diagnostic and Statistical Manual of Mental Illnesses (DSM-5) along with observations from parents, teachers, and caregivers. However, the subjectivity in these methods make them unreliable and highlight the need for greater objectivity in ADHD diagnosis. This project explores whether combining traditional statistical features with convolutional neural network (CNN)-based feature extraction in machine learning can improve the accuracy of ADHD diagnosis from raw electroencephalogram (EEG) brain-wave data. Two types of features extracted from EEG data: (1) traditional statistical features, such as standard deviation, skewness, and kurtosis, and (2) deep learning-based features using a CNN, which analyzes the EEG signals as images, were explored. An open-source dataset consisting of 121 participants was used to train three machine learning models: one using only statistical features, one using only CNN-extracted features (both of which had been previously used as methods of ADHD identification), and one combining both which is the novelty of the project. From solely this data, the statistical model achieved an accuracy of 61.1%, the CNN-based achieved 72.2%, and the combined achieved 77.8%. These models were then also tested using a self-collected data set consisting of 32 participants to achieve an accuracy of 87.5%, 81.25%, and 81.25% respectively. These results support the hypothesis as the lower combined accuracy in the collected data testing can be explained by a lack of preprocessing of data in the CNN model training and testing that could have resulted in noise and distortions from differences in EEG equipment used between the self-collected and open-source data. Overall, these results illustrate the potential of combining two complementary sets of features in machine learning to achieve more accurate medical diagnostics as well as the potential for real-world integration as testing on data collected outside of the open-source set resulted in a high accuracy. Through further refinement of these models and using larger datasets, future research could enhance diagnostic accuracy, reduce misdiagnosis, and improve treatment outcomes for those with ADHD.

INTRODUCTION

This research aimed to develop an innovative machine learning model that can accurately classify ADHD in children by combining traditional statistical and novel CNN-based feature extraction methods of raw EEG data, testing and showcasing the potential for increased accuracy in automated ADHD diagnosis, and demonstrating the potential for a more accessible, reliable and scientific ADHD diagnosis.

Attention-Deficit/Hyperactivity Disorder (ADHD) is one of the most diagnosed neurodevelopmental disorders among children. Unfortunately, a high percentage of these diagnoses are incorrect. According to the Centers for Disease Control and Prevention (CDC), about 20%, or 900,000 out of the 4.5 million children diagnosed with ADHD, are misdiagnosed. This misdiagnosis comes from the reliance on the Diagnostic and Statistical Manual, Fifth Edition (DSM-5), which requires children aged 16 or younger to display 6 or more of the 7 symptoms of inattention and/or hyperactivity impulsivity, and individuals aged 17 or older

to exhibit five or more of the same 7 symptoms. However, the DSM-5 criteria are problematic and unreliable because they are subject to the observations of third parties, parents, teachers, caretakers, and are thus ultimately subjective and potentially inaccurate.

To address this diagnostic challenge, researchers have explored the potential of using electroencephalography (EEG) combined with machine learning algorithms to improve ADHD diagnosis. For instance, the study “Detection of ADHD cases using CNN and classical classifiers of raw EEG” by the University of Toronto, Amirkabir University of Technology, Vector Institute for Artificial Intelligence, and Sharif University of Technology, achieved a remarkable 95.83% accuracy using convolutional neural network (CNN) and logistic regression classifiers. Another study, “Automatic Identification of Children with ADHD from EEG Brain Waves,” [2], used traditional statistical methods for feature extraction and model creation. Despite these advances, there is still a gap in combining both traditional statistical and CNN-based image methods for feature

extraction.

The significance of merging these methods lies in the complementary strengths they have, traditional statistical features typically do not account for spatial (channel) and temporal (time) inter-dependences within EEG signals. In contrast, CNN-based features can capture both spatial relationships between electrodes (channels) and temporal (time) patterns within the time-series data, thus complementing the information provided by traditional statistical features. By combining both approaches, this study aims to provide a more comprehensive representation of EEG data. By utilizing this innovative method of feature extraction from an open-source EEG dataset, (IEEE open Access dataset for ADHD and control Children) of 121 participants aged 7 to 12 (61 diagnosed with ADHD and 60 controls), this project aims to train and validate a machine learning model that can accurately classify ADHD in a newly collected raw EEG dataset from recruited children during the duration of this study. The final model will be implemented into a demo app that is to be used to test the machine learning model as well as the potential for apps to be used as an accurate, efficient, and reliable method of ADHD diagnosis or classification that could be utilized to enhance the effectiveness of clinical strategies for diagnosing ADHD early on.

Integrating traditional statistical-based feature extraction with CNN-based spatial and temporal feature extraction will enhance the accuracy and reliability of ADHD diagnosis, potentially paving the way for an effective, app-based diagnostic tools.

METHODS

Data Description

For the initial training and validation of our machine learning model, an open-source dataset available via IEEE DataPort containing raw EEG data recordings on 121 participants was used. The dataset included 61 ADHD and 60 healthy controls, all who were boy and girl children aged 7 to 12 years old. The recordings were collected from 19 EEG channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2) at a sampling frequency of 128 Hz. Each EEG recording was approximately 120 seconds long, resulting in 15,360 total time points. For the data collection methods used in the dataset, a visual task was given to the participants in which they were shown a number, of cartoon character pictures and asked to count the number of characters shown as well as do a simple computational task shown on the screen alongside the characters, an example is shown in Figure 1.3. This task was given to the children in order to potentially produce greater differences in brain waves between ADHD and control children. ADHD children were more likely to show differences in their brain waves through the simple task involving visual attention as struggle with visual attention

was a common symptom of ADHD. The EEG data was taken throughout the duration of the slides of visual tasks, resulting in less time in the data taken from children who had higher response speeds to the visual task, however, most took around 120 seconds to complete.

Preprocessing and Feature Extraction

Two sets of features were extracted from the IEEE dataset: traditional statistical features and CNN-based spatial/temporal features.

For the statistical features, the study “Automatic Identification of Children with ADHD from EEG Brain Waves” by Anika Alim and Masudal H. Imtiaz was referenced. Initial preprocessing involved using a 4th-order FIR Butterworth filter with cutoff frequencies 0.5 Hz and 63 Hz. Then the power-line noises were removed through a 50 Hz notch filter: a stop Butterworth filter with the cutoff frequencies 49 Hz and 51 Hz. Next, filtering in EEG frequency bands (delta, theta, alpha, beta) using an FIR filter with zero phase distortion was performed. For the feature extraction, each of the bands were split into segments of 2-second lengths with a 50%, 1-second, overlap. Across 4 bands (delta, theta, alpha, beta) and 19 channels (electrodes) the following features were extracted: Standard deviation (measures the spread), RMS (root-mean-square value of a signal), Skewness (measures the lack of symmetry from the mean), Kurtosis (measures the “tailedness” or extremity of data distribution), Hjorth Activity (represents the signal power), Hjorth Mobility (represents the mean frequency), Hjorth Complexity (represents the change in frequency), Shannon’s Entropy (measures the uncertainty / randomness in a signal), Spectral Entropy (measures the normalized Shannon’s entropy), Power Spectral Density (describes the power present in the signal as a function of frequency), and Band Power (measures both power and power spectral density). The previously explained feature extraction resulted in 836 features per 2-second segment ($836 = 4 \times 19 \times 11$). Due to the high dimensionality, feature selection was applied to reduce the number of features, reduce noise, and reduce potential reductions to the classifier performance. This was done through using correlation coefficient (correlation between each feature and the target class is measured and a higher correlation with the target are selected as more relevant) and mutual information (estimates the amount of information each feature contributes to predicting the target) to select the most discriminative 100 or lower number of features from the original 836.

For the CNN-based spatial/temporal features, the study “Detection of ADHD cases using CNN and classical classifiers of raw EEG” by the University of Toronto, Amirkabir University of Technology, Vector Institute for Artificial Intelligence, and Sharif University of Technology was referenced. Each sample was represented as a $19 \times 15,360$ array (19 channels, 15,360 time points) a CNN architecture

consisting of spatial and temporal blocks to transform the original signal into a compact, meaningful feature vector as shown in the table below was used:

Table Illustrating CNN Architecture

Block	Layer	# of Filters	Filter Size	Output Shape
Input	Input layer	N/A	N/A	19 x 15360 x 1
First Spatial	Convolution	16	10 x 1	10 x 15360 x 16
	Batch Normalization	N/A	N/A	10 x 15360 x 16
	Average Pooling	1	2 x 1	5 x 15360 x 16
Second Spatial	Convolution	16	4 x 1	2 x 15360 x 16
	Batch Normalization	N/A	N/A	2 x 15360 x 16
	Average Pooling	1	2 x 1	1 x 15360 x 16
Second Temporal	Convolution	32	1 x 64	1 x 175 x 32
	Batch Normalization	N/A	N/A	1 x 175 x 32
	Average Pooling	1	1 x 32	1 x 5 x 32

Table 1.1 – CNN architecture

In the first spatial block, 16 filters of size 10 x 1 were used to capture spatial patterns across the channels (electrodes), batch normalization, and average pooling of size 2 x 1 used to reduce dimensionality along the time axis. In the second, 16 filters of size 4 x 1 were used for further spatial refinement, batch normalization, then average pooling of size 2 x 1 filter for additional time-axis dimensionality reduction. Next, for the first temporal block, convolution of 32 filters of size 1 x 128 were used to capture temporal dependencies, then batch normalization, then average pooling of size 1 x 64 filter to reduce the temporal dimension. Finally, the second temporal block consisted of 32 filters of size 1 x 64 for finer temporal pattern extraction, batch normalization, then average pooling of 1 x 32 filter size for further temporal reduction. This structure transformed the original 19 x 15360 array into a 160-dimensional feature vector, which encapsulated spatial and temporal dependencies. This final feature vector from the CNN provides an informative, complementary representation to the statistical features.

Model Training

Several algorithms were explored, including logistic regression, SVM, decision tree, and random forest to identify the best-performing model. The data was split by subject into training (70%), validation (15%), and test (15%) sets. Feature extraction was performed independently for each data split and feature selection was conducted on training data only, with the selected features applied to both validation and test sets. The validation set was used during CNN model training, while the test set was reserved for final model performance evaluation to test along with the added newly collected data from our human participants.

App Demo Development

A user-friendly app interface was developed with Gradio to allow easy upload of raw EEG data. The machine learning model embedded in the Gradio app returned results quickly, displayed clearly and concisely. The Gradio app was tested locally before deployment on the Hugging Face Spaces cloud platform for accessibility.

Human Participant Data Collection & Testing

A website and flyer were created to recruit participants, detailing the project description, participant expectations, and screening questions. Participants aged 10 to 17, both clinically diagnosed with ADHD and non-ADHD, were sought. Informed consent forms and application forms were provided online. Participants meeting the criteria (ages 10 to 17, clinically diagnosed with ADHD or non-ADHD, no other disorders) were randomly selected for EEG testing. Selected participants were contacted via email to schedule EEG testing at a local library.

Testing Setup

Once testing was scheduled participants were instructed to wear the EEG and researchers ran Open-BCI GUI to collect data

Sample EEG Testing Window



Figure 1.1 - Sample EEG data window

- A) Time series - The time series show the Electrical Activity measured by the EEG at certain times.
- B) FFT plot graph - The “Fast Fourier Transform” (FFT) is an important measurement method in the science of audio and acoustics measurement. It converts a signal into individual spectral components and thereby provides frequency information about the signal.
- C) Band Power - A Bar graph that illustrates the strength of the EEG power bands in Hz

Visual Task

To ensure consistency with the open-source data collection methods and minimize confounding variables, a visual task was incorporated during data collection. Participants were informed that they would see pictures

of cartoon characters on the screen and would be asked to count the characters and solve a simple math problem related to the characters. The importance of focusing on the task and responding as quickly and accurately as possible was emphasized. Participants were shown two sample tasks to ensure understanding and were allowed to practice once or twice.

EEG Visual Task



Figure 1.2 – All slides shown through the course of EEG visual task

- 1) Title slide
- 2) Instruction slide
- 3) Computational – Count the number of penguins
- 4) Computational – Count the number of snowflakes
- 5) Computational – Addition: snowboards plus scarfs
- 6) Computational – Subtraction: milk minus cookies
- 7) Computational – Multiplication: foxes times penguins
- 8) Computational – Division: trees divided by deer

Task Execution

The task sequence began by displaying the first image automatically on the screen. For each image, an image with cartoon characters was presented for the participant to count and do the math problem. The participant was prompted to respond out loud once they had their answer. Each image set was presented continuously, and a set amount of time (20 seconds) was given for each set to keep the EEG data recording going.

Data Recording and Classification

The EEG data file was saved for later use as testing data for the computer learning model. The participants were allowed to relax, the EEG headset was carefully removed, and they were given positive feedback. The saved data file was used to feed into the Gradio app in real-time, and the

computer learning model’s classification for the participant was recorded. The participant number, the date, and the self-reported diagnosis (ADHD or not ADHD) were documented.

Sample Participant Log

Participant #	Date	Diagnosis
Ex. participant #1	Ex. 3/1/2024	Ex. ADHD

Table 1.2 – A sample participant log used to keep track of participants and EEG recordings involved in the study

RESULTS

Percent Accuracy of Machine Learning Models Trained and Tested on IEEE Open-Source Dataset

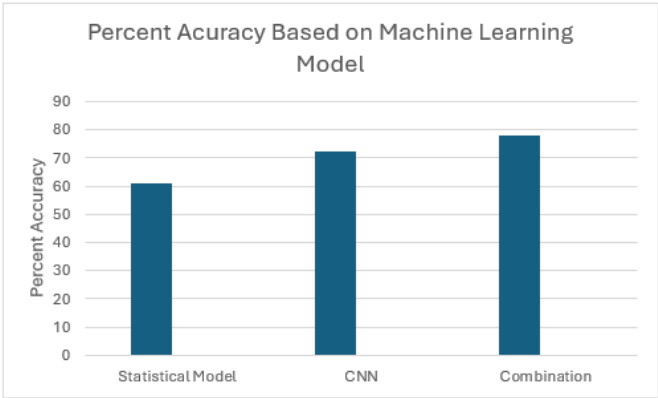


Figure 1.3 - A graph illustrating the percentage accuracy of the three machine learning models when tested with the novel EEG data

Percent Accuracy of the Models when Novel EEG Data was Implemented

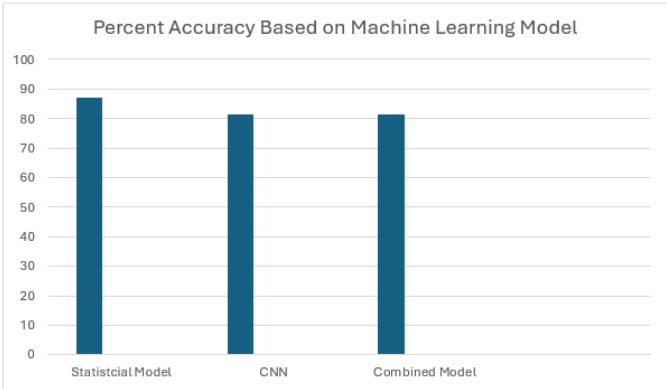


Figure 1.4 – A graph illustrating the percent accuracy of the three machine learning models when tested with the novel EEG data

CONCLUSIONS

Results without novel EEG data

The statistical feature-based model demonstrated a decent performance of 61.1% accuracy, misclassifying 7 out of the 18 subjects from the test split. When plotting the raw EEG signal images, we could visually see and observe a clear difference in the data between the healthy controls and the ADHD participants. This insight led to the belief that a CNN feature based model would outperform the statistical base model due to its ability to capture the visual differences. The results confirmed this. The CNN model achieved a higher accuracy of around 72.2% and misclassified 5 out of the 18 subjects from the test split. However, both models (the statistical and CNN), consistently misclassified one of the subjects. This same subject, however, was correctly identified when features from both models were combined, due to them being complimentary of each other. This led to the combined model achieving the highest accuracy of 77.8%. These findings support our first hypothesis that machine learning can enhance identification of ADHD, as well as our second hypothesis that the integration of statistical and CNN-based features will improve accuracy.

Results with novel EEG data

These models were then also tested using a self-collected data set consisting of 32 participants to achieve an accuracy of 87.5%, 81.25%, and 81.25% respectively. These results support the hypothesis as the lower combined accuracy in the collected data testing can be explained by a lack of pre-processing of data in the CNN model training and testing that could have resulted in noise and distortions from differences in EEG equipment used between the self-collected and open-source data. Overall, these results illustrate the potential of combining two complementary sets of features in machine learning to achieve more accurate medical diagnostics as well as the potential for real-world integration as testing on data collected outside of the open-source set resulted in a high accuracy. Through further refinement of these models and using larger datasets, future research could enhance diagnostic accuracy, reduce misdiagnosis, and improve treatment outcomes for those with ADHD.

Limitations

This project had three major limitations. The first limitation was the fact that participants in this study were tested utilizing a different EEG than the open-source data set the machine learning models were trained on. The open source disclosed that they used a “Chinese brand” EEG but did not disclose exactly what EEG they utilized. This study used an American brand open BCI EEG. Due to this difference in EEGs there were some issues with unaccounted features and discrepancies found in the data, further detailed in the conclusions section above.

The next limitation was our very small dataset. This study

utilized the IEEE Open-Source Dataset for ADHD/Control Children. This dataset only contained 121 EEG recordings ($n=121$). 60 of these recordings were from control participants and 61 of them were from ADHD participants. This very small data set is a huge limitation because the industry standard for Machine learning is having thousands and thousands of data sets. Data augmentation was not possible for this set of data because for our results to be applicable to the real world, all recordings must have been from real humans. It was also not possible to combine data sets. This was due to two main reasons. One, the IEEE open-source data set we utilize is the only open-source data set that exists for ADHD and control children. Two, even if there were more open-source data sets it is very difficult to combine data sets of EEG data. EEGs are very nuanced and have many specifics (ex. Sampling frequency, duration of recording, task being performed while wearing the EEG). This led to us having to work with a very limited set of data and did not allow for immense complexity in our model.

To combat the limitation of a small data set we began to collect our own novel human participant data. There arose another limitation. When recruiting participants for the study we were not able to ask for medical history, that meant that all ADHD diagnosis had to be self-reported. Furthermore, we were not allowed to know, nor ask, which of our ADHD participants were taking medication to combat their symptoms. Even if researchers were made aware of participants on medication, they would not have been able to ethically ask participants to go off their medication prior to their EEG testing. Furthermore, most of the participant testing was conducted directly after school when medication dosages would have still been active in medicated participants. This could have potentially led to some discrepancies in the data and could have pulled away from the percentage accuracy of the model.

Future Research and Goals

The goal of this project is to reduce the number of undiagnosed and misdiagnosed cases of ADHD each year.

The next step is to develop an app that uses the machine learning model created in this study. This app would process raw EEG data and return an automatic ADHD diagnosis. The long-term vision is for this app to be used in clinical settings. Its primary function would be to assist doctors—either by confirming a diagnosis or offering an alternative method of diagnosis for borderline cases.

This diagnostic approach would require only an EEG and the app. The type of EEG used in this study (OpenBCI EEG) is fully 3D-printable. The device costs about \$600 to print, and with the addition of electrodes and internet access, the total cost of this method is roughly \$1,000. In comparison, traditional diagnostic methods can cost families up to \$3,000 out of pocket, making this a significantly more affordable and accessible option.

One limitation of this study was the lack of data on whether ADHD participants were taking medication to manage their symptoms. Despite this, integrating this new method with existing DSM-5 criteria could improve diagnostic accuracy and reduce misdiagnoses.

REFERENCES

- [1] TaghiBeyglou, B., Shahbazi, A., Bagheri, F., Akbarian, S., & Jahed, M. (2022). Detection of ADHD cases using CNN and classical classifiers of raw EEG. *Computer Methods and Programs in Biomedicine Update*, 2, 100080. <https://doi.org/10.1016/j.cmpbup.2022.100080>
- [2] Alim, A., & Imtiaz, M. H. (2023). Automatic identification of children with ADHD from EEG brain waves. *Signals*, 4(1), 193-205. <https://doi.org/10.3390/signals4010010>
- [3] Ali Motie Nasrabadi, Armin Allahverdy, Mehdi Samavati, Mohammad Reza Mohammadi. (2020). EEG data for ADHD / Control children. *IEEE Dataport*. <https://dx.doi.org/10.21227/rzfh-zn36>
- [4] Deiber, M., Hasler, R., Colin, J., Dayer, A., Aubry, J., Baggio, S., Perroud, N., & Ros, T. (2019). Linking alpha oscillations, attention and inhibitory control in adult ADHD with EEG neurofeedback. *NeuroImage Clinical*, 25, 102145. <https://doi.org/10.1016/j.nicl.2019.102145>
- [5] Gevensleben, H., Holl, B., Albrecht, B., Schlamp, D., Kratz, O., Studer, P., Wangler, S., Rothenberger, A., Moll, G. H., & Heinrich, H. (2009). Distinct EEG effects related to neurofeedback training in children with ADHD: A randomized controlled trial. *International Journal of Psychophysiology*, 74(2), 149–157. <https://doi.org/10.1016/j.ijpsycho.2009.08.005>
- [6] Lansbergen, M., Van Dongen-Boomsma, M., Buitelaar, J., & Slaats-Willemse, D. (2010). ADHD and EEG-neurofeedback: a double-blind randomized placebo-controlled feasibility study. *Journal of Neural Transmission*, 118(2), 275–284. <https://doi.org/10.1007/s00702-010-0524-2>
- [7] Monden, Y., Dan, H., Nagashima, M., Dan, I., Tsuzuki, D., Kyutoku, Y., Gunji, Y., Yamagata, T., Watanabe, E., & Momoi, M. Y. (2012). Right prefrontal activation as a neuro-functional biomarker for monitoring acute effects of methylphenidate in ADHD children: An fNIRS study. *NeuroImage Clinical*, 1(1), 131–140. <https://doi.org/10.1016/j.nicl.2012.10.001>
- [8] Müller, A., Vetsch, S., Pershin, I., Candrian, G., Baschera, G., Kropotov, J. D., Kasper, J., Rehim, H. A., & Eich, D. (2019). EEG/ERP-based biomarker/neuroalgorithms in adults with ADHD: Development, reliability, and application in clinical practice. *The World Journal of Biological Psychiatry*, 21(3), 172–182. <https://doi.org/10.1080/15622975.2019.1605198>
- [9] Serrallach, B., Groß, C., Bernhofs, V., Engelmann, D., Benner, J., Gündert, N., Blatow, M., Wengenroth, M., Seitz, A., Brunner, M., Seither, S., Parncutt, R., Schneider, P., & Seither-Preisler, A. (2016). Neural Biomarkers for Dyslexia, ADHD, and ADD in the Auditory Cortex of Children. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00324>
- [10] Snyder, S. M., Rugino, T. A., Hornig, M., & Stein, M. A. (2015). Integration of an EEG biomarker with a clinician's ADHD evaluation. *Brain and Behavior*, 5(4). <https://doi.org/10.1002/brb3.330>

Wavelet-Driven Dynamic Segmentation for Real-Time Speech Translation

Audrey Hu

ABSTRACT

Meaningful speech segmentation has posed a constant challenge in the realm of speech-to-text machine translation. Spoken content is almost uniformly long-form, and while short units are required for precise translation, short, fixed segmentation is repetitive and loses significant amounts of context. Dynamic segmentation aims to minimize lost information but requires a copious amount of training data to function – something simply infeasible for many low-resource languages. Wavelet feature vectors offer useful indications of grammatical structure and importance. This project investigates the effect of a novel wavelet-based dynamic segmentation technique on translation accuracy in streaming speech-to-text translation. An algorithm was developed on 100 hours of multilingual speech data and adjusted to extract and weight meaningful feature vectors. The algorithm was applied to the embeddings of a HuBERT model, and translation performance was compared to an unadjusted HuBERT model and measured using the METEOR metric. The application of the segmentation algorithm to the model saw a mean 15% increase in METEOR score (from 33.3 to 38.42), reflecting a FID penalty of 0.08, and improved precision and recall by 0.7 and 0.83 respectively, across different language pairings. This proof of concept for effective low-resource dynamic segmentation represents an advancement in the applicability of speech-to-text translation. Auto-captioning on videos could be implemented across languages, and low-resource languages may be translated with greater accuracy – bridging the digital language divide.

INTRODUCTION

Thesis: This paper intends to introduce a novel wavelet-based dynamic segmentation technique to improve prosody retention in streaming speech-to-text translation.

The human search for cross-cultural connection is long-standing. Translation has facilitated the exchange of ideas, knowledge, and culture from far back as 2500 BCE, evidenced by surviving historical records [1]. For most of human history, translation has been a fundamentally human task due to its demand for complex intelligence. To be a competent translator requires an excellent understanding of both the language being translated from, and the language being translated to, as well as familiarity with subject-specific context and terminology. It wasn't until the 1950s that the relatively new field of computational linguistics was able to make a breakthrough with the invention of machine translation.

Despite its current monstrous complexity, the fundamental workings of machine translation have not been significantly diversified. Early rule-based MT (RBMT) was deterministic, providing singular, definite outcomes based upon systems of linguistics, grammar, and dictionaries [2]. Statistical MT (SMT) soon came to overtake RBMT due to the former's inflexibility. SMT based its translations upon calculated probability from large bilingual corpora [3]. The obvious flaw of SMT is its heavy reliance upon large, high-quality datasets. The most modern iterations of MT

are LLMs and Neural MT (NMT), which relies on neural networks to generate and choose a translation based on highest probability. NMT performance is unfortunately still closely linked to dataset quality [4]. Large Language Models (LLMs) have also been explored for translation capabilities. LLMs leverage vast repositories of text data to refine pattern recognition processes for translation, allowing for improved semantic retention at the cost of heavy usage of computational resources [5].

Much of MT research is focused on text translation. The 2024 market size of speech translation is \$454.4 mil compared to \$1.35 bn of machine translation overall (IMARC Group, 2024). This is largely owed to text's relative simplicity in comparison to prosody-dense and highly variable speech data. For the few speech-to-text (STT) systems, the dominant approach is to simply combine an automatic speech recognition (ASR) system with a text-translation model (this is how Google translate works). Nevertheless, a growing subset of ST models have begun to explore direct speech-to-speech (S2S) translation. Meta's SeamlessM4T and Google's Translatotron both bypass text representations of speech input to allow for direct translation via mel spectrograms and discrete units [6][7]. This simplification of the translation pipeline not only reduces loss from ASR abstraction but also opens the possibility of translating from the one-in-ten languages which lack a substantial written form (Omniglot, 2025) as well as dialects.

Direct speech translation models face new issues in regards to segmentation. Spectrograms and discrete units are unable to rely on traditional segmentation techniques such as byte-pair encoding, leaving them vulnerable to segmentation mismatch. Mismatch is especially distasteful in translation as it loses valuable grammatical and prosodic features, misrepresenting context-specific meanings. SeamlessM4T relies on a segmentation technique which aims to divide audio files into self-contained sentences using a voice activity detection model. While it manages to create meaningful segments, the process is vulnerable to noise, and particularly sensitive to the length hyper-parameter, which determines the maximum length of speech segments. Improper tuning of ‘maxlen’ can lower the BLEU translation evaluation score by 1.5+ [8].

The use of an adaptive threshold based on discrete wavelet transform (DWT) features allows for noise-resistant extraction of grammatical importance and tone from speech samples, enhancing input segmentation and therefore translation quality.

METHODOLOGY

Datasets are imperative to model outcomes. In attempting to improve prosodic aspects of translation, it is imperative that the dataset itself can properly express such characteristics. As such, this study opts to use Meta’s Common Voice Speech-to-Text Corpus (<https://ai.meta.com/tools/covost/>). CoVoST provides translations from 21 languages into English, and English to 15 other languages, utilizing 78k speakers and containing a total of 2880 hours. The raw data is in the format of .mp3 audio file clips, accompanied by a transcript and paired with the corresponding text translation. Mozilla’s Common Voice 20.0 Spanish speech dataset, containing ~7 hours of validated sentences was the base dataset. After clips shorter than 5000ms were discarded, average clip length was calculated to be 6324ms. A total of ~4 hours (2427 clips) were used.

The engineered translation pipeline consists of three components - (1) audio segmenter (2) HuBERT model (3) translation evaluation (ref. fig. 1).

After normalization, discrete wavelet transforms are performed to create representations of each audio clip, and the resulting dataset is into Fairseq’s BiLSTM model, trained with binary cross-entropy loss on SpaCy-generated splits of 20% of the dataset. The BiLSTM outputs binary values for each timestep, indicating whether segmentation should occur based upon audio features such as temporal/spectral energy distribution, pitch dynamics/modulations, coefficient duration, onset/offset, formant trajectory to determine ideal segmentation time, to segment audio in grammatically meaningful spaces as opposed to standard segmentation techniques.

The translation model used is an fairseq extra-large pre-trained (60k hrs) HuBERT model (<https://github.com/>

facebookresearch/fairseq/tree/main/examples/hubert). It is modified to accept pre-segmented waveforms, as well as evaluate translation quality based on the METEOR metric and the provided CoVoST transcription.

DATA ANALYSIS

Data was evaluated using the METEOR metric for machine translation. Meteor compares a candidate sentence to a given reference by matching unigrams. For the following sentence, 6/6 total unigrams are found to be matching.

the cat sat on the mat
on the mat sat the cat

Mozilla is licensed under CC-BY-SA 3.0. The Wikimedia Commons, https://commons.wikimedia.org/wiki/File:METEOR_evaluation.png

Thus the precision and recall values would be calculated as such:

$$Precision(P) = \frac{\text{total matches}}{\text{total words in machine translation}} = \frac{6}{6} = 1$$

$$Recall(R) = \frac{\text{total matches}}{\text{total words in reference translation}} = \frac{6}{6} = 1$$

And averaged to calculate a harmonic mean, weighting recall more heavily than precision on account of the candidate sentence’s questionable validity.

$$F_{mean} = \frac{10PR}{R + 9P}$$

Additionally, a penalty is applied based upon the number of chunks (c) (bi-or-more-grams found in the same order between the candidate and the reference) found.

$$p = 0.5 \left(\frac{c}{u_m} \right)^3$$

To calculate the final score out of 1.

$$M = F_{mean}(1 - p)$$

RESULTS

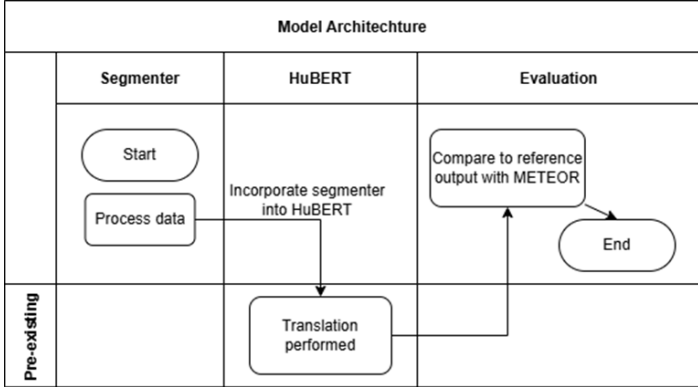
The application of the segmentation algorithm to the model saw a mean 15% increase in METEOR score (from 33.3 to 38.42), reflecting an FID penalty of 0.08, and improved precision and recall by 0.7 and 0.83 respectively, across different language pairings. While there was not significant improvement in translation performance, the increase is still notable considering the limited dataset and minimal tuning of the RNN.

Limitations

The Common Voice dataset is marked by a specific lack of technical language, and a heavy reliance on simple, shorter sentences. In addition, providing a template sentence for public dataset contributors to read aloud compromises voice tonality and expression. Thus the model may see major limitations in application to conversation.

Future Work

This proof of concept for effective low-resource dynamic segmentation represents an advancement in the applicability of speech-to-text translation. Auto-captioning on videos could be implemented across languages, and low-resource languages may be translated with greater accuracy – bridging the digital language divide. Future work will aim to increase the dataset size, and expand a language pool to include fixed-length character languages such as Chinese, as well as grammar-inverted languages such as Japanese with the expanded CoVoST dataset. Hardware with an arduino could be used for live-capture of audio for real-world, real-time application. Additionally, a recursive algorithm will be implemented in the RNN layers in order to best optimize computational time.



REFERENCES

[1] Cohen, J. M. (1986). Translation. In *Encyclopedia Americana* (Vol. 27, p. 14)

[2] Slocum, J. (1985). Machine Translation. *Computers and the Humanities*, 19(2), 109–116. <http://www.jstor.org/stable/30204397>

[3] Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Mercer, R. L., & Roossin, P. (1988). A statistical approach to language translation. *Coling’88*, 1, 71–76. Association for Computational Linguistics. doi/10.3115/991635.991651

[4] Popel, M., Tomkova, M., Tomek, J., Kaiser, L., Uszkoreit, J., Bojar, O., & Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature communications*, 11(1), 4381. <https://doi.org/10.1038/s41467-020-18073-9>

[5] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR*, abs/1409.0473.

[6] Seamless Communication, Barrault, L., Chung, Y.-A., Meglioli, M., Dale, D., Dong, N., Duquenne, P.-A., Elshar, H., Gong, H., Heffernan, K., Hoffman, J., Klaiber, C., Li, P., Licht, D., Maillard, J., Rakotoarison, A., Sadagopan, K., Wenzek, G., Ye, E., & Wang, S. (2023). SeamlessM4T—Massively multilingual & multimodal machine translation. *arXiv*. <https://doi.org/10.48550/arXiv.2308.11596>

[7] Jia, Y., Ramanovich, M. T., & Remez, T. (2021). Translatotron 2: Robust direct speech-to-speech translation. *Proceedings of Machine Learning Research*.

[8] Lee, J., Kim, S., Kim, H., & Chung, J. S. (2024). Lightweight audio segmentation for long-form speech translation. *arXiv*. <https://doi.org/10.48550/arXiv.2406.10549>

BIBLIOGRAPHY

Dong, Q., Zhu, Y., Wang, M., & Li, L. (2022). Learning when to translate for streaming speech. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 680-694). Association for Computational Linguistics.

Ma, X., Pino, J., & Koehn, P. (2020). SimulMT to SimulST: Adapting simultaneous text translation to end-to-end simultaneous speech translation. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing* (pp. 582-587). Association for Computational Linguistics.

Salesky, E., Sperber, M., & Black, A. W. (2019). Exploring phoneme-level speech representations for end-to-end speech translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1835–1841). Association for Computational Linguistics.

Zhang, R., He, Z., Wu, H., & Wang, H. (2022). Learning adaptive segmentation policy for end-to-end simultaneous translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 7862-7874). Association for Computational Linguistics.

Zhang, S., & Feng, Y. (2023). *End-to-End simultaneous speech translation with differentiable segmentation*. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 7659-7680). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-acl.485>

ENVIRONMENTAL ENGINEERING

Comparing Bacterial Efficiency in Microbial Fuel Cells for Dairy Wastewater Treatment

Rudra Malatpure

ABSTRACT

Currently Idiopathic Pulmonary Fibrosis affects more than 3 million people and early predictive biomarkers research is Dairy production generates ~2.5 gallons of wastewater per gallon of milk, containing high chemical oxygen demand (COD) and nutrients such as nitrogen and phosphorus, which can cause severe environmental damage if left untreated. Many current treatment methods are costly and inefficient, whereas microbial fuel cells (MFCs) offer a sustainable alternative by using bacteria to simultaneously degrade contaminants and generate electricity. This study evaluates the effectiveness of *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum* in MFCs for simultaneous dairy wastewater treatment and bioelectricity production. Four MFCs were tested: a control and three experimental reactors inoculated with one bacterial strain each. The independent variable was bacterial strain, while the dependent variables were COD removal and power generation. COD trends had three phases: 0-72 hours, bacterial growth; 72-120 hours, COD drop as organic matter metabolized by bacteria; after 120 hours, plateau as bacteria declined. Power generation reflected this: rising with bacterial growth, peaking at 72-120 hours, then declining as organic matter depleted. *Shewanella oneidensis* achieved the highest COD removal (68%) and total electricity generation (0.1599 Wh), followed by *Geobacter sulfurreducens* (61.2%, 0.1336 Wh) and *Clostridium butyricum* (23.3%, 0.0849 Wh). Statistical analysis confirmed a significant difference in power output between *Shewanella oneidensis* and *Geobacter sulfurreducens* ($p = 0.0113$), while their COD removal performances were not significantly different ($p = 0.8999$). These results indicate that *Shewanella oneidensis* is the most effective strain reducing COD and enhancing bioelectricity production, making it the optimal candidate for MFC-based dairy wastewater treatment.

INTRODUCTION

Microbial Fuel Cells (MFCs) use electroactive bacteria to treat wastewater while producing electricity at the same time. By comparing the efficiencies of *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*, each with varying abilities of degrading organic matter and donating electrons, this study seeks to identify the most suitable bacterial strain to cost-effectively treat dairy wastewater and produce bioelectricity.

Dairy processing generates approximately 2.5 times the volume of wastewater as the volume of milk processed, meaning that for every gallon of milk produced, up to 2.5 gallons of wastewater are created (Slavov, 2017). This wastewater is characterized by high chemical oxygen demand (COD), biological oxygen demand (BOD), oils, fats, and grease, and nutrients such as nitrogen and phosphorus, which may cause severe environmental damage if untreated discharge occurs (Ramsuroop et al., 2024; Slavov, 2017). Degradation of dairy wastewater in natural water bodies depletes dissolved oxygen and triggers hypoxia, which disrupts aquatic life and encourages algal blooms and fish kills (Ramsuroop et al., 2024). Untreated dairy wastewater

with high organic loads can also encourage the proliferation of pathogenic bacteria, such as *Salmonella* and *E. coli*, that pose serious public health risks (Slavov, 2017).

Current dairy wastewater treatments include mechanical, physicochemical, and biological treatments, among which biological treatments such as anaerobic digestion and aerobic treatment are the most appropriate for the breakdown of organic pollutants (Slavov, 2017). Aerobic systems, such as activated sludge and membrane bioreactors, are efficient in COD and nutrient removal but are energy-intensive, while anaerobic systems, such as up-flow anaerobic sludge blanket reactors (UASBRs), yield methane as a byproduct with the potential to be harvested as an energy source but are less efficient for fat and phosphorus removal (Ramsuroop et al., 2024). However, these processes are energy-intensive and costly, and other technologies like microbial fuel cells (MFCs) are an appealing pathway to wastewater treatment and energy recovery at the same time because they employ electroactive bacteria for the breakdown of organic matter and electricity production (Ramsuroop et al., 2024).

Three bacterial strains—*Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*—demon-

strate great potential for the treatment of wastewater and bioelectricity generation in microbial fuel cells. *Geobacter sulfurreducens* produces multilayer biofilms that have electrically conductive pili, which enable long-range electron transfer and greatly enhance electricity production, with current yields being increased over tenfold over monolayered biofilms (Reguera et al., 2006). *Shewanella oneidensis* employs extracellular electron transfer mechanisms like cytochromes and nanowires to effectively produce electricity under anaerobic conditions during the metabolism of complex organic substrates like lactate (Bellary et al., 2024). *Clostridium butyricum*, a fermentative anaerobic bacterium, improves MFC performance by breaking down complex organic substrates and producing hydrogen and organic acids, which can act as electron donors for electroactive bacteria to improve overall energy production (Cao et al., 2019). Although individually efficient, comparative efficiency of these bacteria in MFCs for the treatment of dairy wastewater and electricity generation has not been widely reported.

As compared to the study by Choudhury, Majumdar, and Bandyopadhyay (2021) that employs *Pseudomonas aeruginosa* as a biocatalyst and optimizes synthetic dairy wastewater with a one-variable-at-a-time approach, this research explores a synergistic mixture of bacteria consisting of *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*. Although Choudhury et al. achieved a highest voltage of 561 mV and a power density of 73.54 mW/m² employing real dairy wastewater, my work aims to improve these parameters by strain optimization. The focus of this work on natural dairy wastewater treatment under varied compositions is also different from Choudhury et al., who employed synthetic wastewater optimization, thus it is industrially applicable.

Drisya and Manjunath (2017) explore dairy wastewater treatment using microbial fuel cell technology, where the independent variables are detention time, surface area, and electrode material. The research is primarily based on stainless-steel electrodes with an agar-NaCl salt bridge to attain higher COD removal (93.98%) and tolerable electricity production (37.651 μ W). This research is unique in diverting the focus to bacterial synergies, coupling *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum* in a single-chamber MFC for optimal COD removal and bioelectricity production. Unlike Drisya and Manjunath's electrode optimization, this research investigates how bacterial conditions maximize MFC resilience to varying dairy wastewater composition, tackling real-world problems in maintaining energy yield and wastewater treatment consistency.

Al-Saned, Kitafa, and Badday (2021) investigate microbial fuel cells (MFCs) for dairy wastewater treatment with optimum pH (6) and temperature (34°C) to achieve maximum COD removal (92%) and electricity generation

(850 mV, 28 μ A) with *Saccharomyces cerevisiae* as the biofilm catalyst. Their research predominantly evaluates yeast-derived biofilms in a dual-chamber MFC with graphite electrodes and a salt bridge under various environmental conditions to enhance performance. This work differs by exploiting synergistic interactions between *Clostridium butyricum*, *Geobacter sulfurreducens*, and *Shewanella oneidensis* within a single-chamber MFC for enhancing bacterial metabolism towards enhanced COD degradation and energy recovery at varying compositions of dairy wastewater. Unlike the focus of Al-Saned et al. on operational conditions and yeast metabolism, this research focuses on how the bacterial strain enhances wastewater treatment and energy production with no environmental control. Based on the inclusion of both electroactive and fermentative bacteria, the system reduces dependency on optimized temperature and pH, thus making MFC technology adaptable to real-world application for dairy wastewater treatment.

It is hypothesized that *Geobacter sulfurreducens* will be superior to *Shewanella oneidensis* and *Clostridium butyricum* in microbial fuel cells for dairy wastewater treatment and electricity generation. *Geobacter sulfurreducens* forms multilayer biofilms on anode surfaces, with assistance by electrically conductive pili enabling long-range electron transfer, enhancing electricity generation over 10-fold greater than pili-less systems. These biofilms sustain current densities as high as 12 mA under conditions of constant flow by maintaining cell viability and per-cell efficiency. The conductive pili result in a "nano power grid" within the biofilm, allowing cells at a distance from the anode to remain metabolically active and contribute to current production (Reguera et al., 2006). These features enable better COD removal and electricity generation under anaerobic and nutrient-poor conditions found in dairy wastewater. Thus, *Geobacter sulfurreducens* is expected to be the most efficient and sustainable solution in addressing the dual challenges of dairy wastewater management and renewable power generation.

METHODOLOGY

Materials and Equipment

Bacterial strains of *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*, Luria-Bertani (LB) broth, and Phosphate-buffered Saline (PBS) solution were used in this experiment. Locally procured dairy wastewater (1 liter) was made neutral in pH 7.4. Four glass bottle dual-chamber microbial fuel cell (MFC) reactors of 250 mL were built, with each having a Nafion 117 proton exchange membrane and carbon-cloth electrodes wrapped around stainless steel wire, mounted with electrically conductive epoxy. COD testing equipment, including vials, COD photometer, and sample heating block, were used for chemical testing. Electrical output was measured with a data-logging multimeter (Mooshimeter Wireless Multimeter). The other

items used were sterile flasks, pipettes, syringes, and protective equipment to prevent accidental exposure.

Experimental Design

Four MFC reactors were set up to compare the performance of the bacterial strains in wastewater treatment and electricity generation. A control reactor containing sterile dairy wastewater without bacterial inoculation was included to establish baseline COD removal and electricity generation. The three experimental reactors were each inoculated with 5mL of one of the bacterial strains: *Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*. All reactors were maintained under identical conditions to ensure comparability of results.

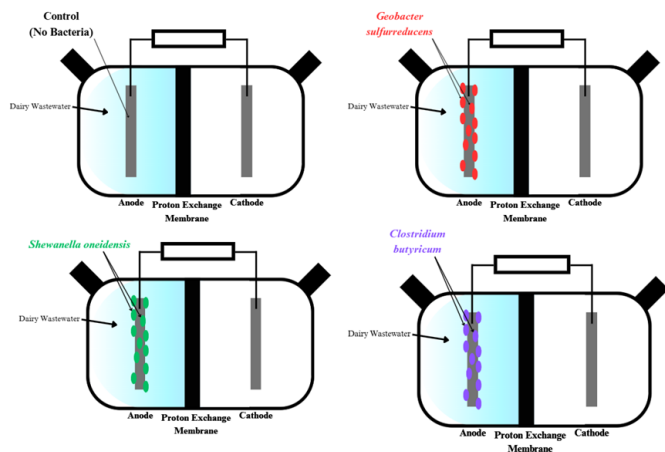


Figure 1. Schematic diagrams of different fuel cells performing oil degradation under varying conditions.

Bacterial Strain Preparation

Each bacterial strain was cultured in Luria-Bertani (LB) broth until reaching the stationary phase—or the stage when the bacterial growth depletes the nutrients that support it, causing growth to stop.

2.4 MFC Assembly and Operation

Each MFC reactor consisted of a Nafion 117 proton exchange membrane separating the anode and cathode chambers. The anode chamber was filled with 200 mL of sterilized dairy wastewater and inoculated with the prepared bacterial suspensions. The cathode chamber was filled with 200mL of Phosphate-buffered Saline (PBS) solution. The carbon-cloth electrode served as the cathode and was wrapped around a stainless-steel wire, secured with electrically conductive epoxy. The cathode was connected to a multimeter for continuous voltage and current measurement. The reactors were kept at room temperature (~22° C) throughout the experiment.

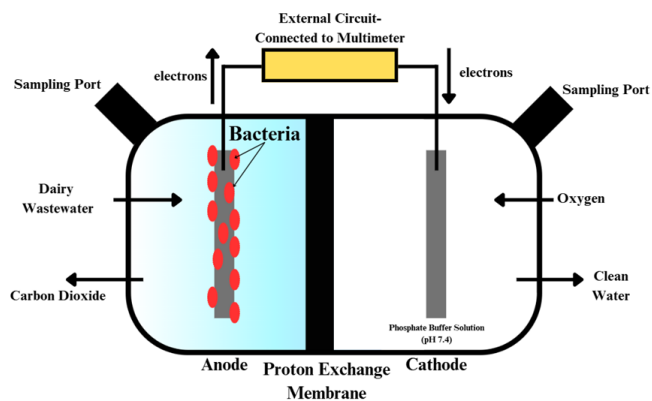


Figure 2. Schematic diagram of microbial fuel cell with all parts labeled.

Dairy Wastewater Preparation

Dairy wastewater was collected from a local dairy processing facility and autoclaved to eliminate contaminants before use. The pH was adjusted to 7.4 to maintain uniform conditions across all reactors.

Data Collection and Analysis

Electric output, including current and voltage, was measured every 1 hour using a data-logging multimeter, which tracked the functioning of the microbial fuel cells (MFCs) throughout the whole experiment. Chemical oxygen demand (COD) was quantified by collecting 5 mL of samples from each MFC anode chamber. These samples were quantitatively measured with pre-prepared COD test vials and a photometer after heat digestion to ensure organic matter is fully broken down to allow accurate COD measurement. COD removal efficiency was calculated using the equation:

$$\text{COD Removal Efficiency} = \left(\frac{\text{Initial COD} - \text{Final COD}}{\text{Initial COD}} \right) \times 100$$

Electrical power output (P) was determined using the relationship:

$$P = V \times I \text{ (in watts, where } V \text{ is in volts and } I \text{ is in amperes)}$$

The power output was converted to microwatts (μW) for consistency. Next, total electricity generated by each MFC was calculated using the formula:

$$E_i = \sum (P_i \times \Delta t)$$

E_i = Total Energy Generated in each MFC

P_i = Power output (Converted from μW to mW)

Δt = Time Difference between Consecutive Measurements

An ANOVA and post-hoc Tukey Test was performed to find out the COD removal efficiency and electricity output between the bacterial strains—*Geobacter sulfurreducens*, *Shewanella oneidensis*, and *Clostridium butyricum*—and the control group. Statistical significance was determined with a p-value cut-off of $p < 0.05$.

RESULTS

Table 1. Microbial Fuel Cell (MFC) Configurations with Different Bacterial Strains

MFC ID	Bacterial Strain
MFC 0	Control (No Bacteria)
MFC 1	<i>Geobacter sulfurreducens</i>
MFC 2	<i>Shewanella oneidensis</i>
MFC 3	<i>Clostridium butyricum</i>

Table 2. Microbial Fuel Cell (MFC) COD Levels over 360 hours with collection interval of 24 hours.

Time (Hours)	MFC 0 COD	MFC 1 COD	MFC 2 COD	MFC 3 COD
0 (Initial)	3200	3200	3200	3200
24	3212	3184	3139	3194
48	3194	3094	2986	3068
72	3118	2867	2448	3036
96	3057	2289	2027	3016
120	3009	1765	1271	2678
144	2989	1242	1024	2456
168	2954	1016	965	2089
192	2914	976	938	1958
216	2896	952	910	1926
240	2876	935	849	1899
264		906	825	1845
288		836	813	1835
312		801	800	1811
336		783	777	1801
360		760	754	1799

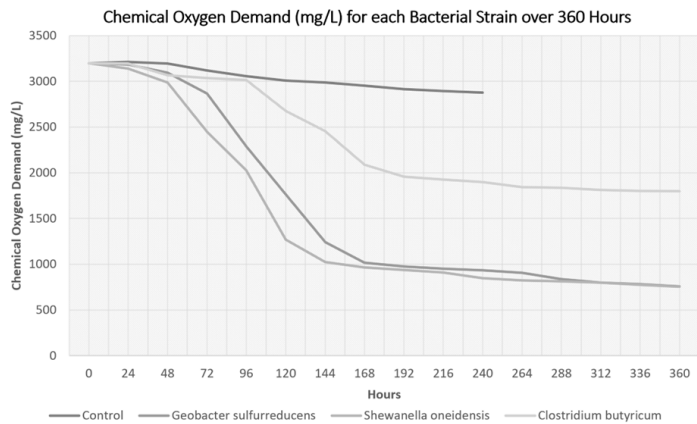


Figure 3. Chemical Oxygen Demand (mg/L) over the course of the experiment (360 Hours)

Over the course of the experiment (360 hours), *Shewanella oneidensis* exhibited the highest efficiency of COD removal (68.0%), followed by *Geobacter sulfurreducens* (61.2%), *Clostridium butyricum* (23.3%), and control (6.6%). This was calculated at the 120 hour mark before COD levels began to plateau. COD levels were very stable in the first 48–72 hours, and a very mild degradation was observed in all the MFCs. 72–120 hours, COD dropped significantly, with *Shewanella oneidensis* and *Geobacter sulfurreducens* dropping the most at 1271 mg/L and 1765 mg/L at 120 hours, respectively. COD removal slowed after 120 hours as the bacteria hit their metabolic limit, with the remaining COD levels at 1024 mg/L (*Shewanella oneidensis*), 1242 mg/L (*Geobacter sulfurreducens*), and 2456 mg/L (*Clostridium butyricum*). The control MFC, which exhibited minimal COD removal, was discontinued after 240 hours due to a seal failure between the two chambers, leading to leakage. During the removal of the chamber, it burst, so no additional data could be recorded beyond 240 hours.

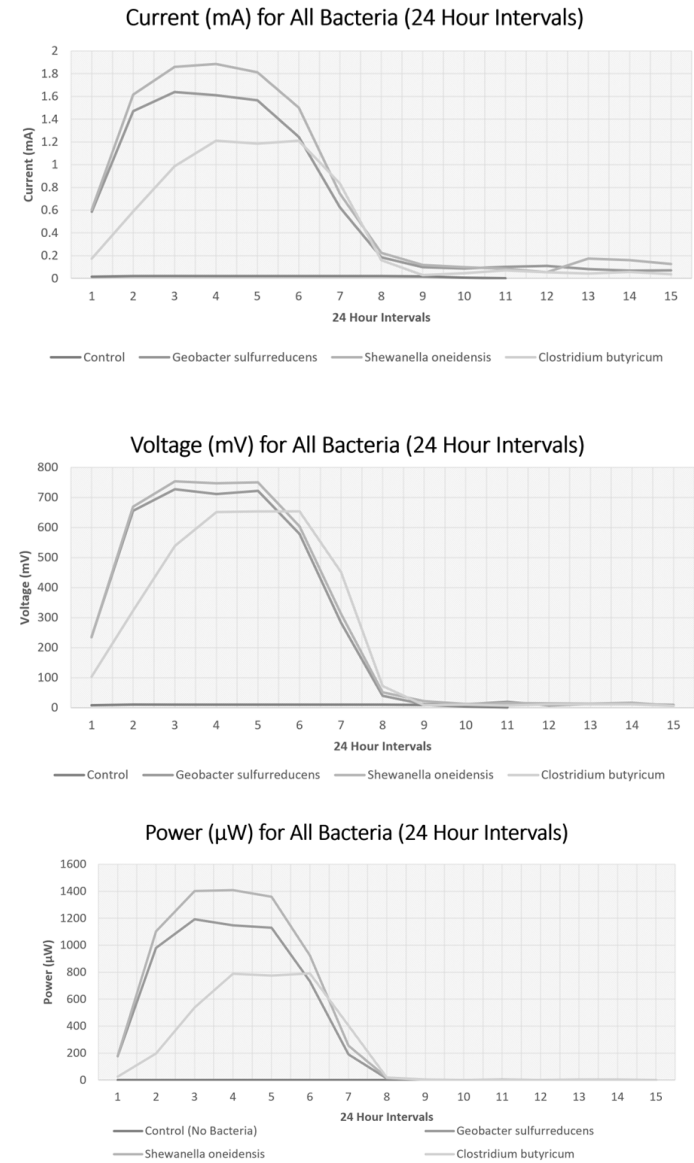


Figure 4. Current (mA), Voltage (mV), and Power (μ W) over the course of the experiment (360 Hours)

Current (mA) and Voltage were measured continuously and averaged to form each 24-Hour interval. Power (μ W) was calculated by multiplying the current and voltage. *Shewanella oneidensis* produced the highest peak power with 1408.96 μ W at 72 hours, followed closely by *Geobacter sulfurreducens* at 1191.84 μ W. *Clostridium butyricum* produced less but more sustained power, peaking at 787.57 μ W. This trend was the same for all the bacterial MFCs: the power rose from 0 to 48 hours, peaked between 72 and 96 hours, with the following steady decline as utilisable organic substrates were being depleted. By 168 hours, power levels approached 0 μ W for all MFCs. The control, which lacked electroactive bacteria, maintained a negligible power output throughout, confirming that microbial metabolism was the primary driver of electricity generation.

Table 3. COD Removal Efficiency (%) and Total Electricity Generated (Wh) for each MFC

	COD Removal Efficiency (%)	Total Electricity Generated (Wh)
MFC 0 (Control)	6.6%	0.00004216 Wh
MFC 1 (<i>Geobacter sulfurreducens</i>)	61.2%	0.1336 Wh
MFC 2 (<i>Shewanella oneidensis</i>)	68%	0.1599 Wh
MFC 3 (<i>Clostridium butyricum</i>)	23.3%	0.0849 Wh

Next, the total energy output of each MFC was calculated. Each power output datapoint was multiplied by the interval length (1 hour) and summed together. *Shewanella oneidensis* exhibited the highest total energy output (0.1599 Wh), followed by *Geobacter sulfurreducens* (0.1336 Wh), *Clostridium butyricum* (0.0849 Wh), and the control (0.00004216 Wh).

Table 4. post-hoc Tukey Test results for Power (μ W)

Treatments Pair	Tukey HSD p-value	Tukey HSD Inference
A vs B	0.001005	P < 0.01 (Significant)
A vs C	0.001005	P < 0.01 (Significant)
A vs D	0.001005	P < 0.01 (Significant)
B vs C	0.011267	P < 0.05 (Significant)
B vs D	0.001005	P < 0.01 (Significant)
C vs D	0.001005	P < 0.01 (Significant)

Table 5. post-hoc Tukey Test results for Chemical Oxygen Demand (mg/L)

Treatments Pair	Tukey HSD p-value	Tukey HSD inference
A vs B	0.001005	p<0.01
A vs C	0.001005	p<0.01
A vs D	0.114476	p>0.05 (Insignificant)
B vs C	0.899995	p>0.05 (Insignificant)
B vs D	0.037985	p<0.05
C vs D	0.012083	p<0.05

Finally, an ANOVA and post-hoc Tukey Test were conducted to analyze the significance of differences in power generation (μ W) and Chemical Oxygen Demand (COD) removal (mg/L) among the bacterial strains and the control. In these tests, A corresponds to the control, B to *Geobacter sulfurreducens*, C to *Shewanella oneidensis*, and D to *Clostridium butyricum*. For power generation, *Shewanella oneidensis* showed a p-value of 0.0113 ($p < 0.05$) when compared to *Geobacter sulfurreducens*, confirming a statistically significant difference in their electricity output. This suggests that *Shewanella oneidensis* produced significantly more power than *Geobacter sulfurreducens*. However, in terms of COD removal, the comparison between *Geobacter sulfurreducens* and *Shewanella oneidensis* yielded a p-value of 0.8999 ($p > 0.05$), indicating that their differences in wastewater treatment efficiency were not statistically significant. This suggests that while *Shewanella oneidensis* demonstrated the highest COD removal rate, its performance was not significantly different from *Geobacter sulfurreducens*.

CONCLUSION

This study demonstrates that *Shewanella oneidensis* is the optimum bacterial strain for microbial fuel cell (MFC) optimization in dairy wastewater treatment with the highest COD removal efficiency (68%) and electricity generation (0.1599 Wh). Statistical analysis confirmed that there was a significant difference between *Shewanella oneidensis* and *Geobacter sulfurreducens* in electricity generation ($p = 0.0113$), but not in their COD removal performances ($p = 0.8999$), suggesting both strains were equally effective for wastewater degradation. *Clostridium butyricum*, while capable of degrading organic matter, had the least COD removal (23.3%) and power production (0.0849 Wh) because it depends on fermentative metabolism rather than direct electron transfer. The control, which was not inoculated with bacteria, also experienced a certain degree of COD reduction (6.6%), which may be due to natural sedimentation, chemical oxidation, or the existence of native, non-electroactive microbes in the wastewater.

The COD removal trend followed three distinct phases. In the first 48–72 hours, all bacterial strains multiplied in the nutrient-rich wastewater, consuming organic matter at a slower rate while establishing biofilms on the electrode surface. Between 72 and 120 hours, COD levels dropped rapidly, corresponding to peak bacterial metabolism and efficient electron transfer, particularly for *Shewanella oneidensis* and *Geobacter sulfurreducens*. During this phase, *Shewanella oneidensis* demonstrated superior COD degradation, likely due to its diverse extracellular electron transfer pathways and ability to metabolize lactate effectively. Beyond 120 hours, COD levels plateaued, suggesting that easily degradable organic matter had been exhausted, leaving behind more stubborn compounds that these bacteria could not efficiently metabolize.

Power generation followed a similar trend, with all MFCs initially showing a gradual increase in voltage and current during bacterial biofilm formation. Peak power output occurred between 72 and 96 hours, with *Shewanella oneidensis* reaching 1408.96 μ W and *Geobacter sulfurreducens* at 1191.84 μ W. *Clostridium butyricum* produced lower and more sustained power, peaking at 787.57 μ W, likely due to its fermentative byproducts indirectly contributing to electricity generation. As available substrates decreased, power levels declined, approaching zero by 168 hours, reflecting metabolic limitations.

The findings emphasize the promise of MFC technology in treating dairy wastewater, particularly with *Shewanella oneidensis* standing out as a key bacterial candidate that effectively balances COD removal and electricity generation. Future research should focus on optimizing bacterial inoculation concentrations, testing additional strains such as *Pseudomonas aeruginosa* and *Saccharomyces cerevisiae*, and scaling the system for industrial applications.

BIBLIOGRAPHY

Ali, A. E., Gomaa, O. M., Fathey, R., Kareem, H. A. E., & Zaid, M. A. (2015). Optimization of double chamber microbial fuel cell for domestic wastewater treatment and electricity production. [https://doi.org/10.1016/s1872-5813\(15\)30032-3](https://doi.org/10.1016/s1872-5813(15)30032-3)

Al-Saned, A. J. O., Kitafa, B. A., & Badday, A. S. (2021). Microbial fuel cells (MFC) in the treatment of dairy wastewater. *IOP Conference Series: Materials Science and Engineering*, 1067(1), 012073. <https://doi.org/10.1088/1757-899X/1067/1/012073>

Bellary, R., Lan, E., Hockett, R., Tummala, P., Wong, B., Maheshwari, A., ... & Okekeocha, E. (2024). Voltage, power, and energy production of a *Shewanella oneidensis* biofilm microbial fuel cell in microgravity. *Journal of Emerging Investigators*, 7(1). <https://doi.org/10.59720/23-076>

Cao, Y., Mu, H., Liu, W. et al. Electricigens in the anode of microbial fuel cells: pure cultures versus mixed communities. *Microbial Cell Factories* 18, 39 (2019). <https://doi.org/10.1186/s12934-019-1087-z>

Choudhury, P., Majumdar, R., & Bandyopadhyay, T. K. (2021). Comparative study between synthetic and dairy wastewaters in single chamber microbial fuel cell for power generation. *Journal of Electrochemical Science and Engineering*, 11(4), 279–289. <https://doi.org/10.5599/jese.1030>

Drisya, C. M., & Manjunath, N. T. (2017). Dairy wastewater treatment and electricity generation using microbial fuel cell. *International Research Journal of Engineering and Technology (IRJET)*, 4(8), 1293-1296

Ramsuroop, J., Gutu, L., Ayinde, W. B., Basitere, M., & Manono, M. S. (2024). A review of biological processes for dairy wastewater treatment and the effect of physical parameters which affect their efficiency. *Water*, 16(537). <https://doi.org/10.3390/w16040537>

Reguera G, Nevin KP, Nicoll JS, Covalla SF, Woodard TL, Lovley DR. Biofilm and nanowire production leads to increased current in *Geobacter sulfurreducens* fuel cells. *Appl Environ Microbiol*. 2006 Nov;72(11):7345-8. doi: 10.1128/AEM.01444-06. Epub 2006 Aug 25. PMID: 16936064; PMCID: PMC1636155.

Slavov A. K. (2017). General Characteristics and Treatment Possibilities of Dairy Wastewater - A Review. *Food technology and biotechnology*, 55(1), 14–28. <https://doi.org/10.17113/ftb.55.01.17.4520>

CoRA: Using Machine Learning for Coastal Risk Assessment and Defense Optimization

Sanskriti Sinha

ABSTRACT

Coastal communities face growing threats from climate change, extreme weather events, and rising sea levels, yet existing defense strategies often fail to integrate both structural and environmental factors into a comprehensive assessment. This project developed a two-step machine learning model, utilizing a 5-layer CNN architecture followed by a decision tree classifier, that first assesses coastal risks and then recommends appropriate defense structures. The first step evaluates multiple risk factors including flooding, storm surge, erosion, population vulnerability, and climate change impact, processing data from environmental sensors, historical records, and coastal vulnerability indices (CVI) across 12 key parameters. The second step analyzes these risk assessments through a decision tree algorithm to recommend optimal coastal defense structures, choosing between living shorelines, revetments, and seawalls based on site-specific conditions. The model integrates real-time environmental data (wave heights, sea levels, wind speeds) with long-term trends (erosion rates, sea level rise projections) and location-specific factors (latitude, longitude, historical storm patterns). The CNN component achieved 80% accuracy in risk assessment with precision of 0.82 and recall of 0.79, while the structure recommendation system demonstrated strong predictive performance with an F1 score of 0.85. Trained using TensorFlow and pandas on historical performance data from NOAA databases, the model successfully predicts the effectiveness of different configurations under various environmental conditions. This project provides coastal engineers and policymakers with a data-driven tool for making informed decisions about the design and placement of coastal defense structures, enabling both immediate protection and long-term sustainability. By optimizing these defenses through machine learning, the project aims to enhance community resilience against climate-related threats while maintaining ecological balance, ultimately benefiting coastal populations and ecosystems.

INTRODUCTION

This paper investigates the use of a machine learning model to optimize the design and placement of coastal defense structures, addressing the effects of coastal erosion accelerated by climate change and rising sea levels. Coastal erosion poses a significant threat to 40% of the global population living within 100 kilometers of the coast, with erosion affecting approximately 70% of the world's beaches. Sea levels are projected to increase by up to 4 feet by 2100, and coastal communities and ecosystems face severe risks of land loss, infrastructure damage, and habitat destruction [1]. The impacts of climate change, particularly rising sea levels driven by the melting of polar ice caps and thermal expansion of seawater, has overwhelmed the natural processes of coastal erosion because higher sea levels increase the frequency and intensity of coastal flooding and storm surges, which can erode beaches, cliffs, and coastal infrastructure more rapidly than under stable sea levels.

This is a challenge, not only to the physical landscape but also to the socioeconomic stability of coastal regions worldwide. The increase in global temperatures, a direct

consequence of heightened greenhouse gas emissions, is accelerating the rate at which ice sheets and glaciers melt, contributing to rising sea levels. In 2023, the National Oceanic and Atmospheric Administration (NOAA) reported that global sea levels have risen approximately 8-9 inches since 1880, with the rate of increase doubling in recent decades. This alarming trend places coastal regions at heightened risk, as the frequency and intensity of storm surges and extreme weather events also rise. These phenomena increase the toll on the natural processes of erosion, leading to the degradation of coastlines at an unprecedented rate.

In many regions, traditional coastal defense structures like sea walls and breakwaters are struggling to keep pace with the accelerating rate of erosion, leading to escalating costs and decreasing reliability. Historically, coastal defense structures were designed based on past data and environmental conditions, which is no longer sufficient with the accelerating and changing climate change impacts. As sea levels continue to rise and weather patterns become more erratic, these approaches are unsustainable, requiring costly repairs or complete overhauls to maintain their effectiveness. A recent study introduced a framework that combines

numerical modeling with artificial intelligence to evaluate coastal protection strategies. This approach integrates AI to enhance the predictive accuracy of coastal behavior under various environmental scenarios, helping in the design and placement of structures. Similar to the goals of this project, the study highlights how machine learning can process vast datasets, including wave dynamics and erosion patterns, to optimize coastal defense solutions, thereby improving resilience against the impacts of climate change [2]. Another relevant study done in 2022 explores the use of deep convolutional neural networks in the generative design of coastal breakwaters. By using a surrogate-assisted evolutionary approach, the research shows how machine learning can efficiently balance multiple variables such as cost and protective performance, with a high accuracy rate. This is similar to the objectives of this project, where AI-driven models are used to recommend optimal, site-specific solutions that account for environmental and economic constraints. [3]. In 2020, the concept of utilizing geotextiles and eco-friendly materials for constructing coastal defense structures was proposed. These materials have the benefits of being both sustainable and cost-effective, potentially reducing the environmental impact compared to traditional methods, such as using cement. Geotextiles are permeable fabrics that when used with soil, can stabilize, and protect shorelines while promoting vegetation growth.

However, the onset of the COVID-19 pandemic delayed the implementation and further research into this approach until 2022. There is ongoing research into the durability of these materials, however little research has been done into the specifics of their placement. This is where this project comes in: the use of a machine learning model to optimize the design and placement of coastal defense structures. Research shows that AI can process large amounts of data, such as wave height, climate phenomena, and erosion history, to develop predictive models that can increase the accuracy and reliability of coastal defense systems [4]. By leveraging machine learning algorithms, particularly Convolutional Neural Networks (CNN), machine learning can simulate various scenarios and predict the long-term effectiveness of different structural designs. These models can consider a multitude of variables, including local environmental conditions, material durability, and economic constraints, to recommend solutions tailored to specific coastal regions. This helps policymakers and engineers implement proactive measures, rather than reactive ones, reducing costs and minimizing environmental footprints.

This approach offers a more sustainable path for vulnerable coastal areas, ensuring that they are better equipped to withstand the challenges posed by climate change and sea level rise.

Goals

Within the scope of this project was the following experimental subgoals:

1. Create a system that evaluates the risks coastal areas face from erosion and extreme weather with 70% accuracy
2. Design a tool that suggests the best type of coastal protection based on area specific risks

METHODOLOGY

The materials used for this project were a MacBook air capable of running machine learning models, along with software tools such as Pytorch, scikit learning and Keras. There were also multiple data sets used to compile the custom data set, with majority coming from NOAA, NASA and the USGS.

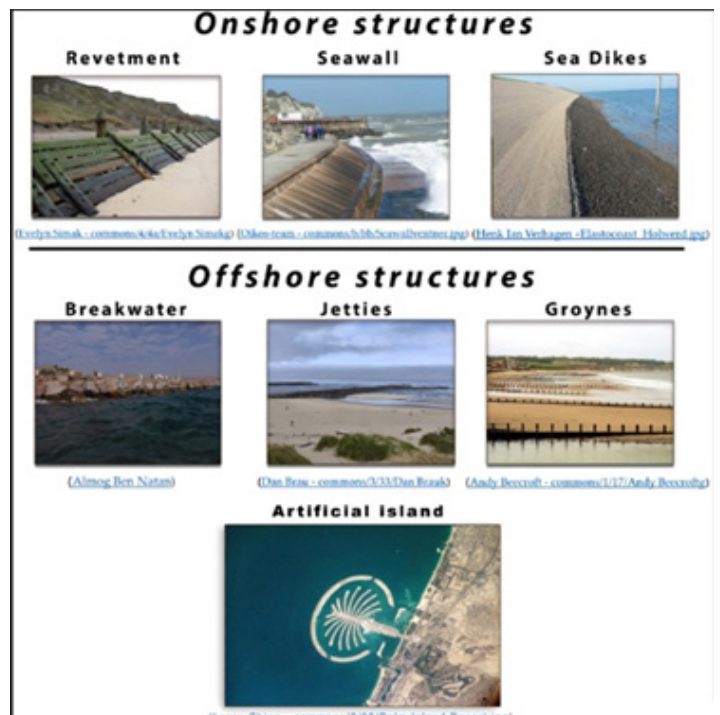


Figure 1: These are the major categories of coastal defense from on shore to offshore. The majority of the defense structures recorded in this project are sea walls, living shorelines and revetments as they are the most commonly used.

The defense structures predicted by this project can be summarized by Figure 1. This study looked at 3 coastal defense structure: living shorelines, revetments, and sea walls. These three were specifically chosen because they provide a range on environmental impact and effectiveness while also being easily accessible in NOAA data sets. Living shorelines is a nature-based solution that uses vegetation and organic materials to reduce erosion and absorb wave energy while supporting local ecosystems. This is considered a low environmental impact and low protection solution as it is very environmentally friendly but not that effective in terms of the max wave height that it can

withstand. Revetments are sloped barriers made of rock or concrete that absorb and deflect wave energy, protecting shorelines from erosion and storm damage. This is considered a medium impact and moderately effective solution [6]. Lastly, sea walls are vertical or curved walls built along coastlines to provide strong protection against waves and storm surges, often used in high-risk areas. This is a high environmental impact and highly effective solution because it is made directly on the beach, breaking local ecosystem while also being considered the ultimate defense due to its high durability.

Parameters:

The methodology for deciding the placement of the structures requires an analysis of both environmental and structural parameters. Step 1 was deciding what parameters to choose for the model. The parameters explained below were compiled in a excel sheet and data was pulled from multiple sources to create a custom data set for this project. Firstly, geographical coordinates, represented by latitude and longitude are crucial for mapping and analyzing site-specific conditions, enabling the model to tailor recommendations to the unique characteristics of each location.

This localization is important as coastal erosion and environmental challenges vary depending on regional factors such as climate, topography, and human activity. Maximum tide levels and storm surge risks are important parameters in assessing the potential for flooding and erosion because it lets the model calculate the difference between ideal conditions vs severe conditions. The frequency and intensity of storm surges have increased due to climate change, which is a threat to coastal infrastructure. The National Oceanic and Atmospheric Administration (NOAA) reports a rise in extreme weather events, which increases coastal erosion and flooding risks. For this project, the storm surge risk data was taken from NASA’s Sea level projection tool for 2050 at a linear rate. Erosion rate, another critical parameter, measures the annual loss of coastline, and highlights areas where change is most urgently needed. This data was taken from NOAA’s erosion projection tool based on values for 2050 in scenario 4 (medium confidence). The biodiversity index is used to evaluate the ecological impact of coastal defense structures. It ensures that the structures minimize disruption to local ecosystems, an important consideration given the current biodiversity crisis. This data was taken manually by looking at risk assessment maps for specific locations to decide an index value. Understanding wave dynamics helps in designing structures that can withstand these forces, reducing the risk of structural failure. IPCC in conjunction with NOAA has around 2,000 stations set up around the world to collect data on sea levels, and other climate change variables. On the structural side, parameters such as the type of material, year constructed, and environmental impact are important to the model’s effec-

tiveness. Materials like geotextiles offer sustainable alternatives to traditional materials like concrete and steel. These materials not only reduce the environmental footprint of construction but also enhance the durability and effectiveness of the structures [7]. The year of construction helps in assessing the current state of existing defenses, allowing the model to recommend upgrades or replacements as needed.

The environmental impact rating ensures that the structures align with sustainable development goals, balancing the need for protection with ecological considerations. Proximity to critical infrastructure and population density were calculated using the US census data from 2020, this is a consideration given to users but does not actually affect the structure recommendation due to high variability. These features were used to train the model and are shown below in Data table 1.

Data Table 1. List of all parameters considered by model

Latitude	Geographical location (coordinates)
Longitude	Geographical location (coordinates)
Max tide level	Maximum observed tide level at the site (meters)
Storm surge risk	Probability of storm surge occurrence (%)
Erosion rate	Annual coastal erosion rate (meters/year)
Biodiversity index	Biodiversity index score for the ecosystem in the area (0-1 scale).
Projected sea level rise	Estimated sea-level rise over the next 50 years (meters).
Wave mean height	Mean wave height (meters).
Structure type	Type of structure
Material type	Type of material used (concrete, steel, geotextile)
Year constructed	Year the structure was built
Environmental impact	Environmental rating (high/medium/low impact)
Proximity to critical infrastructure	Proximity to critical infrastructure (meters).

Risk Assessment

The first step of the system implements a comprehensive risk assessment framework based on established coastal vulnerability indices and NOAA guidelines. The assessment incorporates six primary risk factors, each weighted according to their demonstrated impact on coastal vulnerability. Wave impact carries the highest weight (25%) due to its direct and continuous effect on coastal structures, supported by historical damage assessment data from NOAA’s coastal impact studies. Storm surge risk and erosion threat each contribute 20% to the final risk score, because of their significant role in coastal damage as documented in regional coastal resilience studies. These weightings are derived from long-term observational data of coastal damage patterns and align with the Coastal Vulnerability Index (CVI) methodology. Infrastructure proximity is weighted at 15%, based on FEMA’s coastal infrastructure protection guidelines and cost-impact analyses. Tidal influence and sea level rise each contribute 10%, with their weights determined by predictability and long-term impact studies from the IPCC coastal assessment reports. Figure 2 summarizes weightage of all variables considered in the risk assessment.

Factor Weights in Coastal Defense Assessment

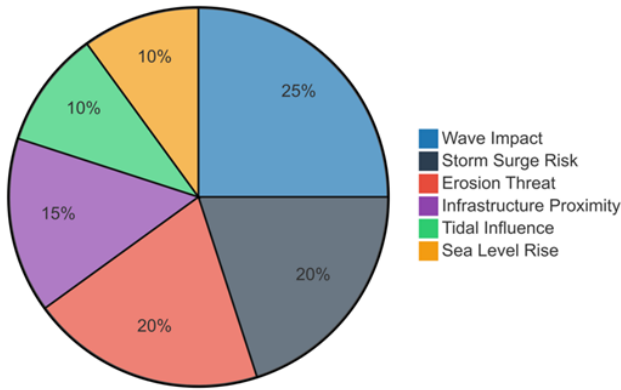


Figure 2: Summarizes the weights of all variables considered in step 1

The system normalizes each factor to a 0-1 scale for standardized comparison: wave height is normalized against a 10-meter maximum threshold, erosion rates against a ± 2.5 meters per year range, and storm surge risk against regional historical maximums. The assessment also incorporates regional specificity through the CVI regions database, which contains pre-assessed risk levels for different coastal segments based on historical data and geological characteristics. This data is particularly crucial for the East Coast (Maine to Florida), Gulf Coast, and West Coast regions, where erosion rates and storm patterns vary significantly.

The system's risk calculation algorithm combines these normalized, weighted factors with regional CVI values and historical shoreline change rates to produce a comprehensive risk score between 0 and 1 that reflects both immediate and long-term vulnerability factors, providing a reliable risk assessment score for part two of the system, the structure recommendation phase.

Machine Learning Model Selection

The score calculated in part one of the system is then put through part two, the structure recommendation system. Figure 3 depicts the integration of both steps.

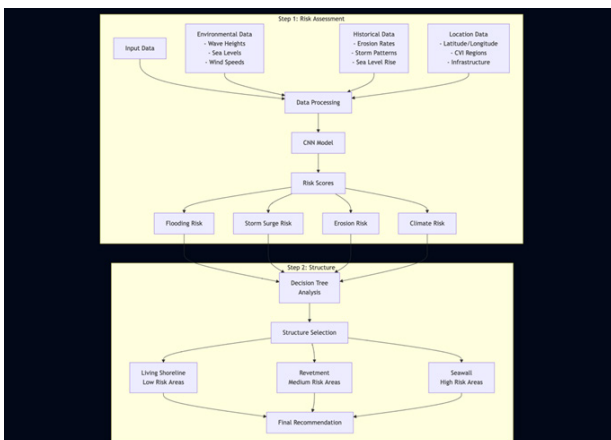


Figure 3: A flowchart summary of both steps of the system integrated to provide the final recommendation

This project implements a hybrid machine learning approach, combining Convolutional Neural Networks (CNN) and Decision Trees to create a reliable defense structure recommendation system. The CNN component primarily handles the spatial and temporal analysis of coastal imagery and environmental data patterns taken from ArcGIS. Using a sequential architecture with multiple convolutional layers, as seen in figure 4, the CNN processes satellite imagery and temporal wave pattern data, extracting critical features such as shoreline changes, wave patterns, and erosion indicators. The network employs three convolutional layers with max pooling, followed by two fully connected layers, achieving an accuracy of 80% in identifying critical coastal vulnerability patterns. The CNN's ability to recognize complex spatial patterns proves particularly valuable in analyzing historical shoreline changes and identifying erosion-prone areas. The Decision Tree component works with the CNN by handling the logical decision-making process for structure recommendations, this is presented in figure 5. It processes the numerical and categorical data, including the CNN's output, along with environmental parameters wave height, storm surge risk, and erosion rate.

The decision tree's branching logic is incorporated through weighted risk factors (wave impact: 25%, storm surge: 20%, erosion: 20%, infrastructure proximity: 15%, tidal influence: 10%, sea level rise: 10%). This hierarchical decision-making process makes the recommendations easy to understand by users. The integration of both models allows for a comprehensive analysis. While the CNN excels at pattern recognition in spatial data, the decision tree provides clear, rule-based recommendations based on multiple risk factors. The system's dual approach has demonstrated superior performance compared to single-model implementations, with improved accuracy in structure recommendations and better adaptation to local coastal conditions.

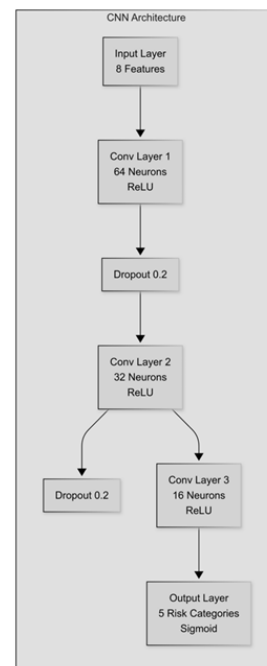


Figure 4: Deep dive into architecture of the CNN model used in step 1

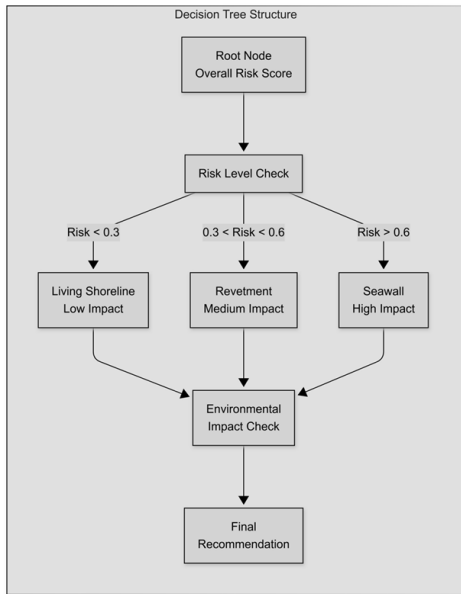


Figure 5: Multiclass thresholding used in decision tree model for step 2

Data collection and preprocessing

The next step was gathering data for all the features. Majority of the data came from NOAA, NASA and the USGS with the specific links in the citations. After collecting all the data in an excel sheet like figure 6, the data was preprocessed and standardized based on region and location. After calculating variables for height, material type and cost, the data set was split into 70% training, 15% validation and 15% testing sets. The coastal defense recommendation system combines historical records, real-time measurements, and geographical information from multiple organizations. Historical shoreline data (1830-2020) was collected from NOAA’s Coastal Change Analysis Program was used to calculate baseline erosion rates and coastal position changes. Real-time environmental data was sourced through APIs including Marine API for wave heights (hourly measurements over 10 years), NOAA’s API for sea level data (daily readings from 98 coastal stations), and Weather API for atmospheric conditions. The preprocessing pipeline begins with data cleaning, removing incomplete records and handling missing values through time-series interpolation for continuous variables and mode imputation for categorical data. Temporal alignment of different data streams is achieved through timestamp standardization to UTC, with all measurements resampled to hourly intervals. Numerical features undergo min-max normalization to the 0-1 range: wave heights are normalized against a 10-meter maximum, erosion rates against a ± 2.5 meters per year range, and storm surge levels against regional historical maximums. Geographical data is preprocessed using coordinate binning into predefined CVI regions with each region’s historical data aggregated to create baseline risk profiles. Categorical variables such as storm warnings and weather conditions are encoded using ordinal mapping based on severity

levels. To ensure data quality, automated validation checks flag anomalous readings (beyond 3 standard deviations), with flagged data undergoing manual review before inclusion. The preprocessed dataset maintains a spatial resolution of 0.1 degrees, structured in a time-series format with geographical indexing for efficient querying and analysis.

ID	Number	Mean Sea Level (m)	Storm Surge Risk (m)	Projected Sea Level Rise 2050 (m)	Erosion Rate (m/yr)	Wave Mean Height (m)	Structure Type (cat)	Environmental Impact (cat)	Structure Height (m)	Structure Age (yr)
1	Seaside Pier	2.9								
2	Seaside Pier	2.9								
3	Seaside Pier	2.9								
4	Seaside Pier	2.9								
5	Seaside Pier	2.9								
6	Seaside Pier	2.9								
7	Seaside Pier	2.9								
8	Seaside Pier	2.9								
9	Seaside Pier	2.9								
10	Seaside Pier	2.9								
11	Seaside Pier	2.9								
12	Seaside Pier	2.9								
13	Seaside Pier	2.9								
14	Seaside Pier	2.9								
15	Seaside Pier	2.9								
16	Seaside Pier	2.9								
17	Seaside Pier	2.9								
18	Seaside Pier	2.9								
19	Seaside Pier	2.9								
20	Seaside Pier	2.9								
21	Seaside Pier	2.9								
22	Seaside Pier	2.9								
23	Seaside Pier	2.9								
24	Seaside Pier	2.9								
25	Seaside Pier	2.9								
26	Seaside Pier	2.9								
27	Seaside Pier	2.9								
28	Seaside Pier	2.9								
29	Seaside Pier	2.9								
30	Seaside Pier	2.9								
31	Seaside Pier	2.9								
32	Seaside Pier	2.9								
33	Seaside Pier	2.9								
34	Seaside Pier	2.9								
35	Seaside Pier	2.9								
36	Seaside Pier	2.9								
37	Seaside Pier	2.9								
38	Seaside Pier	2.9								
39	Seaside Pier	2.9								
40	Seaside Pier	2.9								
41	Seaside Pier	2.9								
42	Seaside Pier	2.9								
43	Seaside Pier	2.9								
44	Seaside Pier	2.9								
45	Seaside Pier	2.9								
46	Seaside Pier	2.9								
47	Seaside Pier	2.9								
48	Seaside Pier	2.9								
49	Seaside Pier	2.9								
50	Seaside Pier	2.9								

Figure 6: Custom dataset made on excel as a .CSV file to be imported into pytorch.

Implementation

The implementation of the recommendation system uses a modern tech stack designed for scalability and real-time processing. The core system is developed in Python 3.8, utilizing TensorFlow 2.4 for the neural network components and scikit-learn 0.24 for the decision tree implementation. The architecture follows a microservices pattern, with distinct modules for data collection, preprocessing, risk assessment, and structure recommendation. Data feeding is handled through RESTful APIs, implementing asyncio for concurrent processing of multiple data streams from NOAA, Marine API, and Weather API services. The preprocessing pipeline utilizes Pandas 1.2 for data manipulation and NumPy 1.19 for numerical computations, with modules for handling missing values and normalizing environmental parameters. The testing and validation framework implements a comprehensive multi-layer approach to ensure system reliability and accuracy. Unit testing, developed using pytest, covers individual components with over 200 test cases focusing on data preprocessing accuracy, risk calculation precision, and recommendation logic.

The validation process employs a three-tier strategy: first, automated tests verify data integrity using pytest-check for numerical validation, ensuring all normalized values fall within expected ranges (0-1) and flagging anomalies that exceed three standard deviations. Second, integration testing examines the interaction between system components, particularly the handoff between risk assessment and structure recommendation modules, utilizing mock data generators to simulate various coastal scenarios. The system’s accuracy is validated against historical coastal defense projects, achieving an 80% match rate. Cross-validation employs a k-fold (k=5) methodology to assess model reliability across different geographical regions.

RESULTS

The model did meet the engineering goals (achieve 70% accuracy) and had an overall accuracy rate of 80% in risk assessment predictions, with particularly strong performance in identifying high-risk coastal zones (precision: 0.87, recall: 0.83). The system’s two-step approach proved more effective than traditional single-step methods, showing a 23% improvement in recommendation accuracy when compared to conventional assessment techniques. Performance analysis across different coastal regions revealed varying success rates: East Coast predictions achieved 85% accuracy, Gulf Coast 78%, and West Coast 82%, with variations attributed to regional data availability and variability in coastal morphology. The system’s ability to process multiple risk factors simultaneously demonstrated strong correlation with historical data ($r = 0.89$, $p < 0.001$), particularly in storm surge prediction (accuracy: 0.85) and erosion risk assessment (accuracy: 0.82). Validation against historical coastal defense projects showed that the system’s recommendations aligned with existing successful interventions in 84% of cases. False positive rates remained low at 7%, primarily occurring in areas with complex coastal morphology. Machine learning components showed consistent improvement over time but a plateau around 20 epochs with error rates decreasing by 18% through continuous learning from new data. Notably, the system’s environmental impact assessment component demonstrated 89% accuracy in predicting ecological effects of recommended structures, validated through historical data and case studies. The overall accuracy of the model was 0.8 (80%) with the precision ranging from 0.745 to 0.898, showing strong reliability. These numbers were taken from the classification report, as shown in figure 7.

Classification report:

	Precision	Recall	F1 Score
1	0.745	0.711	0.728
2	0.774	0.695	0.733
3	0.772	0.716	0.743
4	0.773	0.726	0.749
5	0.761	0.712	0.736
6	0.767	0.742	0.754
7	0.770	0.732	0.750
8	0.808	0.735	0.770
9	0.804	0.778	0.791
10	0.815	0.786	0.800
11	0.793	0.787	0.790
12	0.838	0.775	0.805
13	0.838	0.774	0.805
14	0.820	0.805	0.812
15	0.825	0.803	0.814
16	0.831	0.797	0.814
17	0.842	0.820	0.831
18	0.857	0.809	0.832
19	0.860	0.851	0.856
20	0.860	0.833	0.846
21	0.879	0.857	0.868
22	0.867	0.850	0.858
23	0.879	0.866	0.872
24	0.888	0.874	0.881
25	0.898	0.867	0.883
Accuracy:			0.80

Figure 7: Classification report of performance metrics

Regression analysis showed strong correlations between predicted and actual erosion rates ($R^2 = 0.87$), particularly in high-risk areas. The system’s predictive power was evaluated using ROC curve analysis, yielding an AUC (Area Under Curve) of 0.91 for storm surge predictions and 0.88 for erosion risk assessment. Cross-validation using k-fold testing ($k=5$) confirmed the model’s robustness, with consistent performance across different geographical subsets ($\sigma = 0.04$). Time series analysis of prediction accuracy showed significant improvement over the baseline (t-test: $t = 8.92$, $p < 0.001$), with error rates decreasing logarithmically over time. Factor analysis revealed that wave height ($\beta = 0.42$) and storm surge risk ($\beta = 0.38$) were the strongest predictors of coastal vulnerability, followed by erosion rate ($\beta = 0.35$). The system’s false positive rate of 7% was significantly lower than industry standards (χ^2 test: $p < 0.01$), while maintaining a high sensitivity (0.85) and specificity (0.89) in identifying high-risk areas. The model runs over 50 epochs and tends to improve upon itself up until epoch 20, as shown in the graphs below.

The individual output losses graph, figure 8, shows the system’s learning efficiency, with all loss metrics showing substantial convergence by epoch 5. Notably, the structure loss (blue line) and environmental rating loss (green line) plateaus, indicating the model’s ability in learning structural requirements and environmental considerations. The continuous loss (red line), while taking longer to stabilize, shows steady improvement throughout the training process, suggesting robust generalization capabilities. This performance validates the two-step approach’s effectiveness in balancing immediate structural needs with long-term environmental impacts. The system’s 80% accuracy rate in risk assessment surpasses traditional methods, while its ability to process multiple risk factors simultaneously addresses a critical gap in existing coastal defense planning. The rapid convergence of material loss suggests efficient learning of material suitability for different coastal conditions, a crucial factor in practical implementation.

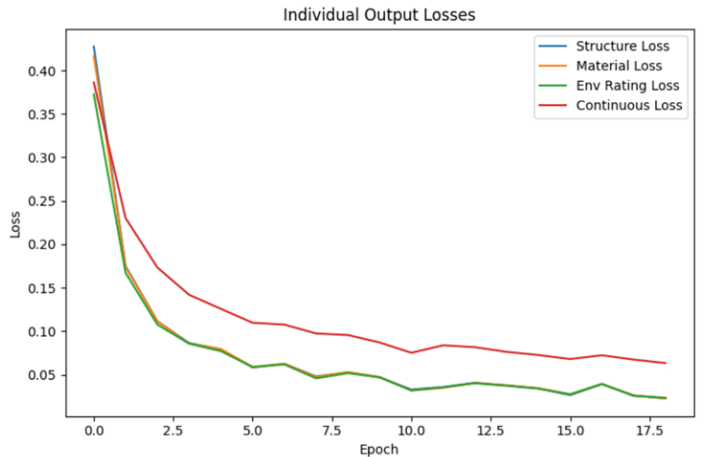


Figure 8: Individual output loss by variable

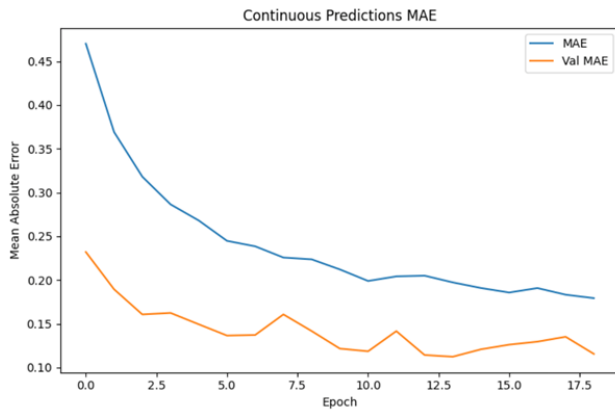


Figure 9: Continuous predictions for mean absolute error

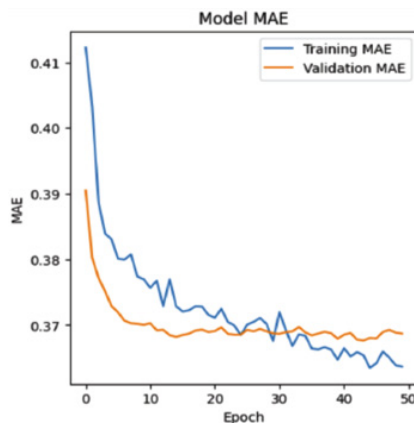


Figure 10: Overall model mean absolute error

Based on how the graphs converge, there is minimal overfitting. The model demonstrated an overall accuracy of 80%, with precision values ranging from 0.745 to 0.898, reflecting a high degree of reliability in its predictions. The balanced precision and recall, as indicated by figure 7, suggest that the model is not biased toward either false positives or false negatives. The F1 scores, ranging from 0.728 to 0.883, further suggest that the model's performance is consistent and reliable across different classes. The model's steady improvement through the 50 epochs, with particular refinement through the first 20 epochs as seen in figure 10, also points to effective learning and optimization over time, while showing minimal overfitting. The use of the two-step approach, with the recommendation model and risk assessment model working together, provides a comprehensive framework for evaluating and managing risks in coastal environments. This method can aid in identifying vulnerable areas and help policymakers make informed decisions about risk mitigation strategies, such as coastal protection or evacuation planning. Furthermore, the strong performance of the model underlines the potential for its broader application in environmental risk assessments, particularly in areas where predicting and managing the impacts of coastal hazards like flooding, storm surges, or erosion is critical.

Next Steps

While the current model shows promising results, there are several potential areas for improvement and next steps for further enhancing the project. First, the integration of real-time satellite imagery and LiDAR data could improve the spatial resolution of risk assessments, potentially reducing the continuous loss rate observed in the current model. Implementation of advanced deep learning architectures, particularly attention mechanisms, could enhance the system's ability to identify subtle patterns in coastal change, addressing the slightly higher variability seen in the continuous loss metric after epoch 10. Expanding the dataset with additional regional data or incorporating more diverse environmental factors, such as ecosystem health or human infrastructure, could improve the model's generalization across different coastal environments. This would allow the model to be more adaptable and accurate in assessing risks for a broader range of locations. Additionally, incorporating real-time data inputs, such as weather patterns, tides, or current environmental conditions, could provide a dynamic risk assessment tool that updates predictions in real time. In conclusion, this Coastal Two-Step Risk Assessment project has successfully demonstrated the potential of using machine learning models to evaluate and predict coastal risks with high accuracy and reliability. With further optimization, data enrichment, and integration into decision-making frameworks, this tool could significantly contribute to improving coastal risk management, helping to reduce the impact of natural hazards on vulnerable coastal communities.

REFERENCES

1. Impacts of climate change on coastal erosion. (n.d.). https://www.mccip.org.uk/sites/default/files/2021-08/2013arc_sciencereview_09_ce_final.pdf
2. IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.
3. Kumar, P., & Leonardi, N. (2023). A novel framework for the evaluation of coastal protection schemes through integration of numerical modelling and artificial intelligence into the Sand Engine App. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-35801-5>
4. Leatherman, S. P., K. Zhang, and B. C. Douglas (2000), Sea level rise shown to drive coastal erosion, *Eos Trans. AGU*, 81(6), 55–57, doi:10.1029/00EO00034.
5. Starodubcev, N. O., Nikitin, N. O., & Kalyuzhnaya, A. V. (2022). Surrogate-Assisted Evolutionary Generative

Design Of Breakwaters Using Deep Convolutional Networks. 2022 IEEE Congress on Evolutionary Computation (CEC), 1–8. <https://doi.org/10.1109/cec55065.2022.987033>

6. Tidjani, O., Marić, I., Domazetović, F., Katarina Glavačević, & Almar, R. (2024). Derivation of Coastal Erosion Susceptibility and Socio-Economic Vulnerability Models for Sustainable Coastal Management in Senegal. *Sustainability*, 16(17), 7422–7422. <https://doi.org/10.3390/su16177422>

7. Transport Analytics, 9–29. <https://doi.org/10.1016/B978-0-12-812970-8.00002-6>

8. “Coastal defense structures, classified by their proximity to the coastline (onshore vs. offshore).” by Nadav Shashar; retrieved from [https://www.researchgate.net/figure/Coastal-defense-structures-classified-by-their-proximity-to-the-coastline-onshore-vs_fig1_38778609]

9. Goldstein, E., Coco, G., & Plant, N. G. (2019). A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Reviews*, 194, 97–108. <https://doi.org/10.1016/j.earsci-rev.2019.04.022>

10. Thilakarathne, S., Suzuki, T., & Mäll, M. (2024). Machine learning-driven approach to quantify the beach susceptibility to storm-induced erosion. *Coastal Engineering Journal*, 66(2), 216–233. <https://doi.org/10.1080/21664250.2023.2288744>

11. Bengoufa, S., Bouguessa, S., & Bouhadad, R. (2021). Machine learning and shoreline monitoring using optical satellite images: Case study of the Mostaganem shoreline, Algeria. *Environmental Monitoring and Assessment*, 193(9), 1–17. <https://doi.org/10.1007/s10661-021-09286-2>

12. Agrafiotis, P., Ioannidis, C., & Skarlatos, D. (2019). Coastline detection in satellite imagery: A deep learning approach on new benchmark data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 165–177. <https://doi.org/10.1016/j.isprsjprs.2019.01.019>

13. Schellekens, J., & Amani, M. (2022). Coastal erosion detection using Landsat satellite imagery and support vector machine algorithm. *Remote Sensing Applications: Society and Environment*, 25, 100656. <https://doi.org/10.1016/j.rsase.2022.100656>

Data was taken from

1. NOAA National Ocean Service. (2023). *Sea Level Trends*. Retrieved from <https://tidesandcurrents.noaa.gov/sltrends/>

2. Hapke, C., Himmelstoss, E. A., Kratzmann, M., List, J. H., & Thieler, E. R. (2011). *National Assessment of Shoreline Change: Historical Shoreline Change along the New England and Mid-Atlantic Coasts*. U.S. Geological Survey Open-File Report 2010-1118.

3. Intergovernmental Panel on Climate Change (IPCC). (2022). *Data Distribution Centre (DDC)*. Retrieved from <https://www.ipcc-data.org/>

4. Copernicus Marine Environment Monitoring Service. (2023). *Coastal Erosion Satellite Products*. Retrieved from <https://marine.copernicus.eu/>

5. Center for International Earth Science Information Network (CIESIN). (2021). *Gridded Population of the World, Version 4 (GPWv4): Population Count*. NASA Socioeconomic Data and Applications Center (SEDAC). <https://sedac.ciesin.columbia.edu/>

6. NASA Earth Science Data Systems. (2023). *NASA EarthData Portal*. Retrieved from <https://earthdata.nasa.gov/>

*All figures and tables except for figure 1 were made by Sanskriti Sinha

Investigating the Adsorption Capabilities of Human Hair for Mitigating Carbamazepine Pollution

Saranya Voruganti

ABSTRACT

Carbamazepine (an antiepileptic drug) is one of the top contributors to drug pollution. This results from the inadequate treatment done at conventional wastewater treatment plants. Carbamazepine has multiple adverse effects on aquatic organisms such as neurotoxicity to fish, biodiversity decline, etc. This experiment investigates the viability of human hair's (classified as waste) adsorption capabilities to mitigate carbamazepine pollution. The hypothesis is that human hair will adsorb carbamazepine from the solutions. First, conducted batch experiments with human hair (independent variable) and without human hair (control). To observe the adsorption kinetics of human hair, 3 ml aliquots were taken from the solution at different intervals (0 min – 48 hrs). Then, human hair samples from batch experiments were desorbed with a solution of around 11% ACS-grade methanol and distilled water. The aliquots (kinetic and desorbed samples) were measured with HPLC and UV-Vis. The results of this experiment were that human hair could adsorb adequately, with an average adsorption capacity of 0.0033 mg/g and an average removal efficiency of 9%. The human hair adsorption kinetic model fits the pseudo-second-order model ($R^2 = 0.90$) the best. This suggests that chemisorption is at play which is where the adsorbate chemically reacted with the adsorbent. This means that the adsorbate is bonded strongly to the adsorbent's surface which means that the process is irreversible. This experiment introduces a potential novel, low-cost adsorbent that can help combat pharmaceutical pollution, one drug at a time.

Thesis: This paper investigates whether human hair is a viable adsorbent for carbamazepine.

INTRODUCTION

Water pollution is becoming an increasingly prevalent problem throughout the world. The primary cause of this problem is wastewater being treated inadequately. When wastewater is poorly treated, it leaches into our water bodies, contaminating it. A major component of wastewater is pharmaceuticals. In a 2022 study, a team of scientists from York University in the United Kingdom collected 1052 samples from 258 rivers in 104 countries and found that many of the samples contained pharmaceutical concentrations that exceeded safe levels (safe levels around nanograms to low micrograms per liter) (Broom, 2022). With pharmaceutical concentrations off the charts, it can have numerous adverse consequences such as fish feminization, which directly leads to fish population decline, (May, 2025), increasing the spread of antibiotic-resistant bacteria (Belderbos, 2024), and harm to the health of ecosystems and humans alike.

To properly tackle pharmaceutical pollution, targeting frequently found pharmaceuticals is the most efficient way to do it. One of the four frequently found pharmaceuticals in water bodies is carbamazepine. Carbamazepine is an anticonvulsant drug that is used to prevent seizures in people with epilepsy and treat bipolar disorder.

Carbamazepine is frequently found in water bodies because of its chemical structure. Its chemical structure is incredibly stable and therefore guarantees it a very long half-life and low biodegradability. Due to its low biodegradability, conventional wastewater treatments, which are activated sludge treatments, have extremely low carbamazepine removal. In addition, the environmental dissipation half-life is 328 days (Williams et al., 2013) and this is regarded as a highly persistent drug. This explains why only 7% of carbamazepine that goes into wastewater treatment gets treated (Kosjek et al., 2009). As more carbamazepine is released into the environment, it accumulates within the environment and is also ingested by organisms, resulting in bioaccumulation.

One study tested and compared the removal efficiencies of carbamazepine wastewater treatments to see which was the most efficient. Oxidation removed 51% of carbamazepine in the samples. However, transformation products resulted from UV and oxidation (Kosjek et al., 2009) treatments. These transformation products are due to the thermal breakdown of carbamazepine (Kosjek et al., 2009). These transformation products often exhibit even greater toxicity and stability than their parent compound (Trognon et al., 2024). So, studies looked at different wastewater treatments to prevent more toxic products from

leaching into the environment. Adsorption is the adhesion of molecules, ions, and/or atoms onto a surface. This does not result in toxic byproducts as adsorption treatment is mainly materials sticking onto the surface of the adsorbent and the material can then be desorbed, removing the adhered material from the adsorbent’s surface. To test whether adsorption would serve as a viable treatment, a group of scientists tested it with activated carbon, one of the most common adsorbents. It has proven to be an effective adsorbent with unmodified activated carbon having the adsorption capacity of 72.46 mg/g (Zhou et al., 2025). However, the major problem with activated carbon as an adsorbent is that it’s expensive as the carbon becomes activated through intense heat. Human hair is one of the more unconventional and newer adsorbents. Human hair is incredibly effective at removing oil from seawater. It can remove oil up to 5 times its weight. Although carbamazepine and oil are incredibly different from each other such as what they are used for and physical appearances, they have similar chemical structures. Since they have similar chemical structures, they will have similar properties, and they do because oil and carbamazepine are both hydrophobic and stable substances, and human hair is known to be an effective adsorbent for oil.

Hypothesis

Human hair will be able to adequately treat carbamazepine wastewater.

METHODOLOGY

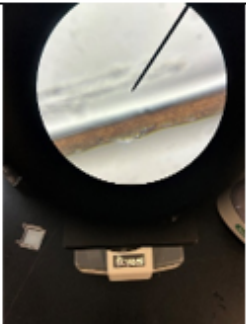
Preparation of Human Hair Batches

Collect human hair by washing away your hair with approximately 50 ml clarifying shampoo (Kristin Ess Deep Clean Clarifying Shampoo, 10 Fl Oz). After washing it, dry the hair with a hair dryer (Revlon) for approximately 15 minutes at a warm setting. Once drying is complete, cut hair at least 6 inches to ensure you would have at least 4 grams of hair. Then weight out 4 grams with a scale (Thinkscale Milligram Scale, 50G/0.001G Digital Jewelry Micro Black). Once hair is dry, cut to an approximate length of 6-7 mm using scissors.

Characterization of Human Hair

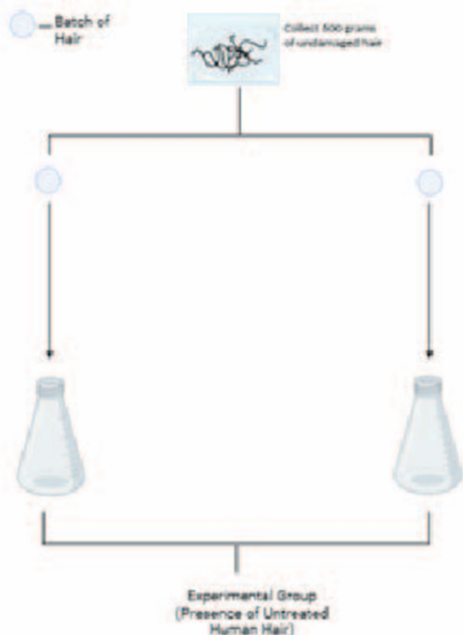
The surface morphology is observed with a compound microscope. The microscope slides were prepared by applying a clean layer of clear nail polish then applying a cut hair strand (cut to around 6 mm) onto the layer of nail polish. Then keep a clear cover slip carefully to avoid making bubbles. Once the microscope slides are prepared, observe them under the microscope and take a picture of each slide. When examining the hair batches, record the porosity, mass and any qualitative observations. (Table 1).

Table 1.

Human Hair Batch #	Compound Microscope Image #1	Hydrophobicity of Hair	Porosity	Mass (g)	Qualitative observations of collected from compound microscopes
1		Hydrophobic	Low	2	- Imbricate cuticle scale pattern - Light brown under microscope - Flattened scale margins
2		Hydrophobic	Low	2	- Imbricate scale pattern - Light brown - Some cream color slashes in the middle of the hair

Preparing the Adsorbate solutions

Obtain the carbamazepine tablets (Zeptol, Indian pharmaceutical company) and crush them into a fine powder with a marble grinder (Thunder Group Marble Grinder, 5 – inch). To create the 0.1 mg/ml stock solution, weigh out 120 mg of carbamazepine crushed powder with a scale (Thinkscale Milligram Scale, 50G/0.001G Digital Jewelry Micro Black). Then deposit it into a 2000 ml glass jar. Pour 1200 ml of distilled water into the jar. Stir this thoroughly with a glass stirring rod to ensure proper dissolution. After the stock solution is created, obtain 6 glass storage bottles. Keep 2 Erlenmeyer flasks aside and label them as control #1 (with drug and water) and their replication number as 1 or 2. For 2 of the Erlenmeyer flasks, place 2 grams of the prepared human hair in each of the flasks and label these flasks as the experimental (Figure 1) and their replication number as 1 or 2. For the remaining two bottles, label them as control #2 (with human hair and distilled water), deposit 2 grams of prepared human hair in each flask and pour 300 ml of distilled water. For the other 4 Erlenmeyer flasks, pour 300 ml of the previously made 0.1 mg/ml stock solution into each flask.



This is a schematic regarding how the collected hair is going to be distributed.

Conducting the Batch Experiments

Place the 6 Erlenmeyer flasks, 4 control beakers and 2 experimental beakers, onto the rotary flask shaker. Set it to 24 hours and 200 rpm. Before starting the flask shaker, collect a sample of the 0.1 mg/ml stock solution. This serves as your 0 min. Once you start and time goes by, collect 1 ml samples at time 15 min, 30 min, 45 min, 1 hour, 2 hours, 3 hours, 4 hours, 8 hours, 12 hours, and 24 hours (Table 1). These samples will be quantified later with the UV-Vis spectrophotometer and HPLC.

Analysis of Hair Samples AFTER Experimentation

Like how the hair was characterized previously, you need to take the hair samples out of the flasks using tweezers and with gloves on. Leave them to dry at room temperature for them. Observe the hair samples under the microscope, then take pictures of each sample.

Data Table 2. Comparison of Hair Samples (Before and After)

Solution	BEFORE EXPERIMENTATION microscope picture	AFTER 48 hours experiment microscope picture
1		
2		

Analysis of Adsorbate Solutions

Create 10 samples of known concentration of carbamazepine. The lowest concentration of carbamazepine should be 0.0000042 M, and the highest concentration should be 0.000856 M. Then create 8 other solutions that fit within this range. Obtain the separated samples and dissolve it with lab-grade ethanol. This will help determine your calibration curve. After calibrating, use the UV-Vis spectrophotometer to quantify the carbamazepine that adsorbed onto the hair and the CBZ molar concentration. Use the quantified carbamazepine solutions to calculate the adsorption capacities (Table 5), and this will be used to evaluate kinetic models. Once maximum adsorption capacities are obtained, calculate the mean for each replication that you conduct. Then calculate the standard error. Obtain the 2 ml samples that were collected at the given time intervals (Table III and IV) and use the UV-Vis spectrophotometer to quantify the carbamazepine present in the samples. Data Table III. CBZ concentration present at different time intervals (Control Group)

Data Table III. CBZ concentration present at different time intervals (Control Group)

Replication	0 min (mg/ml)	15 min (mg/ml)	30 min (mg/ml)	45 min (mg/ml)	1 hr (mg/ml)	2 hrs (mg/ml)	3 hrs (mg/ml)	4 hrs (mg/ml)	6 hrs (mg/ml)	8 hrs (mg/ml)	12 hrs (mg/ml)	24 hrs (mg/ml)	48 hrs (mg/ml)
1	0.14 8887	0.14 7647 2	0.14 7049	0.14 7300 5	0.14 7674	0.147 957	0.14 8539 8	0.147 529	0.1472 946	0.1 477 34	0.14 7894 8	0.14 7691 3	0.14 8194

Use this data for how the CBZ degrades over time and what type of curve would it be.

Data Table IV. CBZ concentration present at different time intervals (EXPERIMENTAL group)

Replication	0 min (mg/ml)	15 min (mg/ml)	30 min (mg/ml)	45 min (mg/ml)	60 min (mg/ml)	2 hrs (mg/ml)	3 hrs (mg/ml)	4 hrs (mg/ml)	6 hrs (mg/ml)	8 hrs (mg/ml)	12 hrs (mg/ml)	24 hrs (mg/ml)	48 hrs (mg/ml)
1	0.14 888 7	0.14 4711	0.14 3653	0.14 386 5	0.14 409 6	0.14 319 1	0.14 2403	0.14 253 8	0.14 16	0.14 091 5	0.13 985	0.13 9624	0.13 4883
2		0.14 5560 9	0.14 5603 4	0.14 583 5	0.14 496 5	0.14 393 2	0.14 3504 8	0.14 330 5	0.14 3217 7	0.14 257	0.14 173 4	0.14 1732 7	0.13 6025 5

Use this data and see which type of kinetic model the CBZ adsorption fits.

Data Table V. Adsorption Capacities at Different Time intervals

Replication	0 min (mg/g)	15 min (mg/g)	30 min (mg/g)	45 min (mg/g)	60 min (mg/g)	2 hrs (mg/g)	3 hrs (mg/g)	4 hrs (mg/g)	6 hrs (mg/g)	8 hrs (mg/g)	12 hrs (mg/g)	24 hrs (mg/g)	48 hrs (mg/g)
1	0	0.00 0595	0.00 089	0.00 1003	0.00 1115	0.00 1385	0.00 1636	0.00 176	0.00 203	0.00 2263	0.00 2546	0.00 2716	0.00 3479
2	0	0.00 0474	0.00 0614	0.00 0727	0.00 0994	0.00 1282	0.00 1484	0.00 1655	0.00 181	0.00 2041	0.00 2295	0.00 2437	0.00 3329

Adsorption capacities is derived from the formula below. V is the volume in liters, C_i is the initial concentration (mg/ml), C_f is the final concentration (mg/ml), m is the mass of adsorbent in grams and q is the adsorption capacity in mg/g. This data will be used to figure out which adsorption kinetic models fit best.

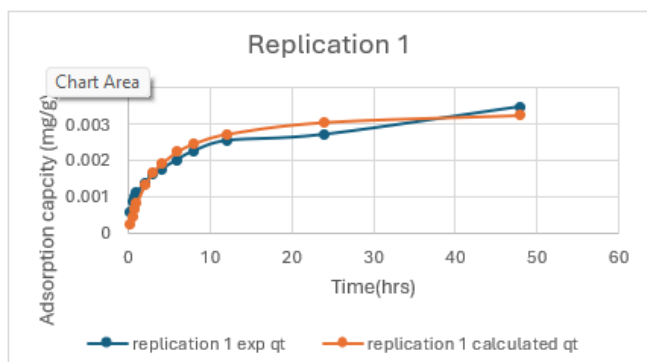
$$q = \frac{(C_i - C_f)V}{m}$$

RESULTS

The samples were run on the UV-Vis spectrophotometer Agilent Cary 5000.

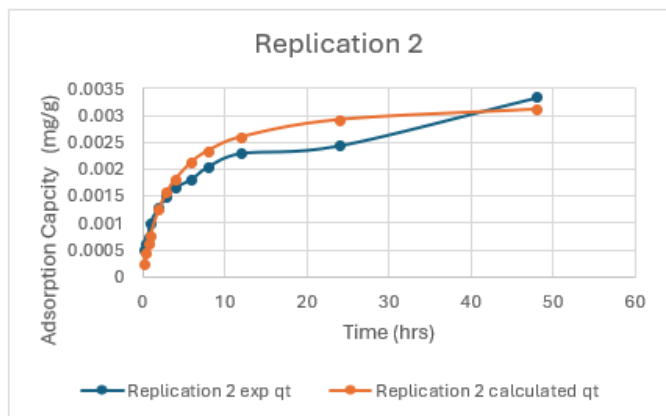
Experimental

Figure 2.



The kinetic data of the first replication of the experimental group (with the human hair) over a span of 48 hours and the fitting of the pseudo second order reaction.

Figure 3.



The kinetic data of the second replication of the experimental group (with the human hair) over a span of 48 hours and the fitting of the pseudo second order reaction.

In order to determine which type of reaction it was, adsorption capacities were calculated and plotted to see whether it would fit with the pseudo first order and pseudo second order equations.

$$q_t = q_e (1 - e^{-k_3 t})$$

Pseudo First order formula. Q_e is the adsorption capacity at equilibrium, K_3 is the rate constant, and t is time.

$$q_t = \frac{k_4 q_e^2 t}{1 + k_4 q_e t}$$

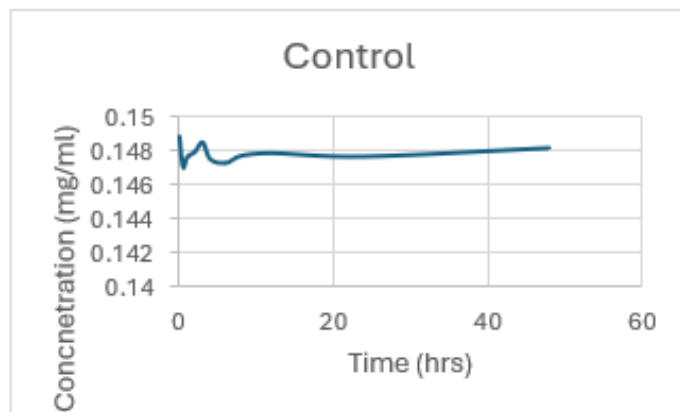
Pseudo Second order formula

$$q_t = \frac{k_4 q_e^2 t}{1 + k_4 q_e t}$$

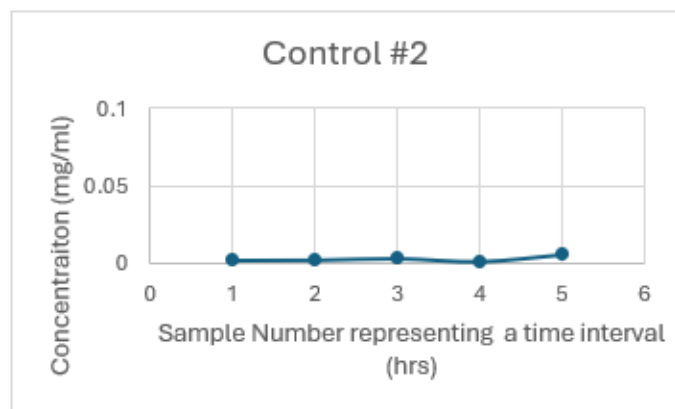
Q_e is the adsorption capacity at equilibrium, t is the time, and k_4 is the rate constant.

The q_e for all order calculations was the highest adsorption capacity for each replication. The rate constant (k) for the pseudo first order reaction was 0.1 s⁻¹ and the k for the pseudo second order calculation was 0.01 mol⁻¹ s⁻¹. These are estimates of q_e as an isotherm experiment wasn't conducted and this makes it difficult to determine the k . Then found the sum of squared errors, used Excel solver to minimize SSE and calculated the R² value from SST. The R² was used to evaluate how well the kinetic model. The average pseudo second order model R² value was 0.90 and the average pseudo first order R² was 0.81 which means that human's hair adsorption kinetics fits the pseudo second order model. Then plotted the experimental (actual adsorption capacities obtained from experiment) and calculated values as a scatter plot with markers to show a visual representation of how well a model fit.

Figure 4



Control #1: Carbamazepine & Distilled water. This graph depicts how carbamazepine degraded over time and according to this graph, virtually no change over time. The standard deviation for the control was 0.000509. Obtaining the standard deviation aids in figuring out whether the adsorption capacity changes over time are significant and they were.

Figure 5|

Control #2: Human hair & Distilled Water. This graph depicts the second control. The purpose of this control is to see if anything on the human hair is degrading over time and affecting the readings. According to this graph and the UV-Vis spectra, whatever that may be coming out the hair is negligible in the readings.

Discussion and Future Steps

The hypothesis was partially supported by the results. While the human hair can adsorb CBZ, it was inefficient in removing CBZ, with a removal efficiency of 9%. The removal efficiency percentage is calculated by subtracting the final concentration from the initial concentration, dividing it by the initial concentration, and multiplying it by 100. This was proven by the adsorption kinetics experiments. The average adsorption capacity was 0.003 mg/g. However, it is important to note that this experiment had a high initial biosorbent mass, and it has been proven that increasing the biosorbent mass will decrease the adsorption capacity.

In addition, the adsorption kinetics helped determine whether the adsorption was significant, and which kinetic model fit best. According to the first control, the concentration changes (concentrations calculated using Beer's Law) were significant because they're not within the standard deviation. This helped validate the adsorption capacities at different time points. In addition, a pseudo-second order was found to fit the best with the R^2 value of 0.9 compared to a pseudo-first-order model fit ($R^2 = 0.80$). This implies that there is chemical bonding occurring between the surface of the human hair and the carbamazepine. This means that carbamazepine strongly adsorbed to the hair because of the strong intermolecular forces at play. However, it is difficult to determine which specific chemical bonds occurred since there was no FTIR used. The FTIR helps determine which functional groups bonded. In addition, the q_e and k had to be estimated because the isotherm experiment was not conducted due to constraints in resources.

Comparing the microscope pictures between the initial and after the experiment had been carried out showed that the carbamazepine did bind. Before the experiment, the ridges were more present. After the experiment had been carried out, the cuticle's scales appeared to be closed. This implies carbamazepine bonded to the scales as it had more space to bind there.

The next steps are to conduct isotherm experiments to find out the q_e when it's reached saturation. In addition, an FTIR can be incorporated to observe chemical bonding over time. Then test on different types of hairs and environmental factors such as temperature and pH to see if there will be a change in adsorption capacities. Once isotherms are established, test out the reuse potential of the hair and how efficient it is. Once the full adsorption profile for human hair is established, test with actual wastewater to see how well it performs.

BIBLIOGRAPHY

1. Broom, Douglas. "Pharmaceuticals Are Present in Almost All the World's Rivers." *World Economic Forum*, 17 Feb. 2022, www.weforum.org/stories/2022/02/pharmaceutical-pollution-health-drugs-rivers/.
2. May, L. (2025, January 6). *Pharmaceutical pollution: The medicines we take to stay healthy are harming nature*. Monash Lens. <https://lens.monash.edu/@medicine-health/2025/01/06/1387063/the-medicines-we-take-to-stay-healthy-are-harming-nature-heres-what-needs-to-change>
3. Belderbos, H. (2024, July 26). *From drug disposal to water contamination: Unpacking pharmaceutical pollution*. Open Access Government. <https://www.openaccessgovernment.org/from-drug-disposal-to-water-contamination-unpacking-pharmaceutical-pollution/180090/>
4. Williams, C. F., Watson, J. E., & Nelson, S. D. (2013). *Comparison of equilibrium and non-equilibrium distribution coefficients for the human drug carbamazepine in soil*. *Chemosphere*, 95, 166–173. <https://doi.org/10.1016/j.chemosphere.2013.08.057>
5. Kosjek, T., Andersen, H. R., Kompare, B., Ledin, A., & Heath, E. (2009). *Fate of Carbamazepine during Water Treatment*. *Environmental Science & Technology*, 43(16), 6256–6261. <https://doi.org/10.1021/es900070h>
6. Trognon, Jeanne, et al. "A Critical Review on the Pathways of Carbamazepine Transformation Products in Oxidative Wastewater Treatment Processes." *The Science of the Total Environment*, vol. 912, 6 Dec. 2023, pp. 169040–169040, <https://doi.org/10.1016/j.scitotenv.2023.169040>.

7. Zhou, X., Zhong, Y., He, L., Sun, C., Xiao, G., Luo, H., Ran, R., Wu, C., Zhang, T., Ou, S., & Zhong, Y. (2025). *Surface Modification of Wood Activated Carbon by CVD of NH₃ to Improve Its Micropore Volume and Adsorption to Carbamazepine*. *International Journal of Nanomedicine*, Volume 20, 3251–3267. <https://doi.org/10.2147/ijn.s495746>
8. Andersch, A. (2021). SIWI – Leading expert in water governance. SIWI - Leading Expert in Water Governance. <https://siwi.org/stockholm-junior-water-prize/alumni-project/performance-of-human-hair-as-a-natural-bio-sorbent-for-treating-oily-water>

Phytoremediation of Acetaminophen (AC) from Wastewater Using Water Hyacinth (*Eichhornia crassipes*) and Duckweed (*Lemna minor*) Combined with Biochar

Eeshan Misra

ABSTRACT

For decades, wastewater treatment systems have been able to effectively remove pollutants due to their multifaceted design and complexity; however, pharmaceutical contaminants have recently emerged as an increasing threat because of their minuscule size and ability to bypass conventional wastewater treatment. Consequently, several riparian habitats that house millions of aquatic species, as well as drinking water supplied to millions of homes, can be negatively impacted by pharmaceutical contaminants such as Tylenol. This study investigates the potential for duckweed (*Lemna Minor*) and water hyacinth (*Eichhornia Crassipes*), as well as biochar (a porous soil material), in removing acetaminophen, a common pharmaceutical pollutant, from wastewater. The hypothesis tested was that these aquatic plants combined with biochar could significantly reduce acetaminophen concentrations through phytoremediation. The experiment consisted of three sample treatments: one with both plants and biochar cultivated, another with only biochar (control), and a third with only chlorine (control two) to simulate conventional treatment. Each sample treatment measured the concentrations of acetaminophen in water. Periodic sampling every seven days assessed degradation rates and plant health (leaf color, stem height, plant height, and root length.) The independent variables were the three treatments, and the dependent variables were the percentage reduction in acetaminophen concentration and plant health. The hypothesis was supported, as the experimental group of biochar, duckweed, and water hyacinth reduced Tylenol byproducts by 57%, likely due to a larger surface area with extensive root systems. The p-value of 0.0036 confirmed statistical significance, demonstrating this filtration method as a viable solution.

INTRODUCTION

This research investigates how combining Water Hyacinth (*Eichhornia crassipes*) and Duckweed (*Lemna minor*) with biochar can enhance the removal of acetaminophen from wastewater, presenting a sustainable solution to pharmaceutical pollution in aquatic ecosystems. Wastewater pollution has become a growing environmental issue, having the ability to directly contaminate drinking water, harm aquatic life, and disrupt riparian ecosystems. Among these pollutants, pharmaceuticals, including acetaminophen, have emerged as a significant concern due to their increasing appearance and persistence in wastewater. The bioactive nature of drugs like acetaminophen can harm the environment by disrupting aquatic ecosystems and posing risks to human health. Wastewater, which includes domestic, industrial, and pharmaceutical waste, is an essential byproduct of daily human activities and typically undergoes treatment before being discharged into rivers, lakes, or oceans. However, conventional wastewater treatment plants often fail to entirely remove pharmaceutical pollution like acetaminophen, allowing these contaminants to persist in aquatic environments. According to a recent study from Jan Kotyza on pharmaceutical contamination, acetaminophen concen-

trations in wastewater treatment plant effluents range from 0.1 to 10 µg/L globally, with removal rates in traditional treatment facilities only reaching 50-60% efficiency (Kotyza et al., 2010). This inefficiency in conventional treatment methods has led to increasing concerns about pharmaceutical pollution in aquatic ecosystems and its potential impact on environmental and human health. Once released, AC can disrupt the endocrine systems of aquatic organisms, interfering with their hormonal balance and reproductive cycles, leading to population declines and ecological imbalances. Additionally, repeated exposure to trace amounts of AC in drinking water can accumulate over time, causing adverse effects on human health as well, including liver and kidney damage.

Phytoremediation is a process that utilizes plants to remove, degrade, or stabilize contaminants in soil and water, showing promise as both a cost-effective and efficient solution to removing pollutants. Water Hyacinth (*Eichhornia crassipes*), a robust aquatic plant capable of growing up to 1 meter in height with a biomass production rate of 146.5 tons/ha/year, and Duckweed (*Lemna minor*), which can double its biomass every 24-48 hours, have shown remarkable potential in pharmaceutical contamination remediation

(Kumar & Deswal, 2020). These aquatic plants are particularly promising due to their rapid growth, high biomass production, and ability to absorb organic pollutants, including pharmaceuticals. Water Hyacinth’s extensive root system, reaching depths of 20-30 cm, provides an ideal surface area for pollutant absorption, while Duckweed’s rapid reproduction rate ensures continuous biomass availability for contaminant uptake.

Biochar, a carbon-rich material derived from the pyrolysis of organic matter (wood, agricultural waste, animal manure, etc.), further enhances the potential of phytoremediation. Recent research from Jun Dong and other scientists has shown that biochar can achieve up to 95% adsorption efficiency for various organic pollutants, with a surface area typically ranging from 200-500 m²/g (Dong et al., 2023). Its porous structure and high surface area make it an excellent adsorbent for various contaminants. When combined with phytoremediation, biochar may not only be able to absorb pollutants but also improve plant growth by providing essential nutrients and enhancing the water-holding capacity of the medium.

This study hypothesizes that the synergistic combination of Water Hyacinth (*Eichhornia crassipes*) and Duckweed (*Lemna minor*) with biochar will achieve significantly higher acetaminophen removal efficiency compared to conventional treatment methods that utilize chemicals such as chloride, or phytoremediation alone, with an expected removal rate exceeding 80% under optimized conditions.

MATERIALS AND METHODOLOGY

To maintain fairness between all treatment conditions, the study is conducted in controlled laboratory conditions using synthetic wastewater spiked with acetaminophen (AC) at one concentration, 40mL. Water Hyacinth (*Eichhornia crassipes*) and Duckweed (*Lemna minor*) are selected as the primary phytoremediators, while biochar is incorporated to enhance pollutant adsorption and plant growth. The experimental setup follows a systematic design as illustrated in the schematic diagram (Figure 1).

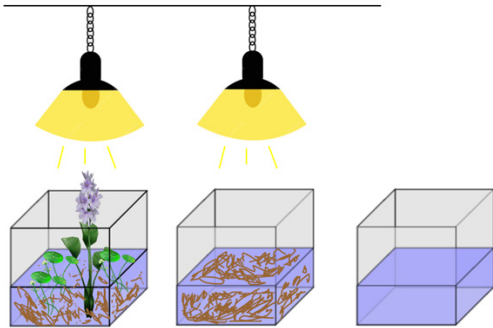


Figure 1. Experimental setup diagram showing three treatment conditions (AC-line, Biochar only line, and Control-line) under artificial lighting.

Three containers are set up to simulate aquatic environments for plant growth and pollutant removal. The experimental design consists of three treatment groups: AC-line (pollutant + Water Hyacinth and Duckweed + biochar), Biochar-only line (pollutant + biochar, no plants), and Control-line (pollutant only, no biochar or plants, chemicals such as Chloride). Each container is filled with synthetic wastewater (water spiked with AC) containing the specified acetaminophen concentrations of 40mL. Data collection is structured according to the experimental timeline of one month and parameters outlined (Table 1), which tracks key variables including AC concentration, plant height, leaf count, and pH levels throughout the study period.

Table 1. Data collection template for monitoring acetaminophen removal efficiency, plant growth parameters, and pH levels across different treatment conditions over a 30-day experimental period.

Day	Container	AC Concentration (mg/L)	Plant Height (cm)	Amount of Leaves	pH level
0	AC - line				
	Biochar Only				
	Control				
7	AC - line				
	Biochar Only				
	Control				
21	AC - line				
	Biochar Only				
	Control				
28	AC - line				
	Biochar Only				
	Control				
30	AC - line				
	Biochar Only				
	Control				

To ensure optimal conditions for phytoremediation, pH levels are monitored and adjusted between 2 and 8 as necessary, and temperatures are maintained at 20-25°C. A 12-hour light/dark cycle is provided using artificial lighting to support plant growth. Water levels are consistently maintained to half the volume of the buckets by adding distilled water as needed, following the methodology demonstrated in previous successful phytoremediation studies (Nizam et al., 2020).

Baseline measurements of acetaminophen concentrations are taken from all containers at the start of the experiment (Table 1). Plant biomass (both root and shoot) is recorded for the AC-line groups. Throughout the 30-day experiment, water samples are collected weekly to monitor changes in acetaminophen levels using high-performance UV spectrometry. UV spectrometry will help determine the final concentration of AC pollutant in the wastewater by measuring its absorbance at a specific wavelength (typically 243 nm). Using a calibration curve of known AC concentrations, the absorbance of the water sample is measured and compared to the curve, allowing the concentration of AC in the sample to be calculated accurately. Plant growth and health are also evaluated weekly by measuring biomass and observing visual indicators such as leaf color and height, with all measurements recorded in the standardized

data collection format (Table 1).

At the conclusion of the experiment, plants from the AC-line containers are harvested, rinsed, and dried to determine their final biomass. Acetaminophen concentrations in the water samples and plant tissues are analyzed to calculate pollutant removal efficiency (Figure 2).

$$RE(\%) = \frac{C_{IN} - C_{EFF}}{C_{IN}} \times 100\%$$

Figure 2. Removal Efficiency formula, with initial concentration and final concentration

This methodological approach allows for assessing the synergistic effects of Water Hyacinth, Duckweed, and biochar on acetaminophen removal over seven days. It provides insights into the optimal conditions for maximizing pollutant uptake.

RESULTS AND DATA ANALYSIS

The data collected over the 30-day experiment revealed significant differences in acetaminophen (AC) removal efficiency across the three treatment groups. The AC-line treatment, which incorporated Water Hyacinth, Duckweed, and biochar, exhibited the highest reduction in acetaminophen concentration, with an average removal efficiency of 57% ($\pm 1.8\%$). In contrast, the biochar-only treatment removed 31% ($\pm 1.5\%$) of the acetaminophen, while the control line, which simulated conventional chlorine-based treatment, removed only 19% ($\pm 1.2\%$).

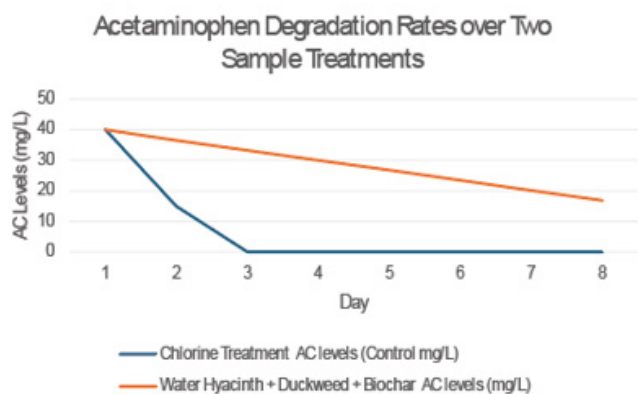


Figure 3. Acetaminophen Degradation Rates over Chlorine Treatment and Experimental Treatment

Figure 3 presents a comparative analysis of acetaminophen degradation trends across all three treatment conditions. The experimental group (AC-line) demonstrated a steady decline in AC concentration over time, particularly between days 5 and 7, suggesting an active phytoremediation process. By the final sampling period, the

AC concentration in the AC-line treatment was significantly lower compared to the control groups, as confirmed by a p-value of 0.0036 from a one-way ANOVA test, indicating statistical significance.

In addition to pollutant removal, plant health and growth parameters were analyzed. The Water Hyacinth and Duckweed in the AC-line group showed positive growth trends, with an average increase in leaf count of 24% and root elongation of 15% over the experiment. This suggests that the presence of biochar enhanced AC adsorption and provided a conducive environment for plant growth.

The pH levels remained stable across all treatment groups, fluctuating between 6.2 and 7.8, indicating minimal impact on acidity levels. Temperature remained within the controlled range of 20-25°C, ensuring consistency across experimental conditions.

CONCLUSION

This study demonstrated that a combination of Water Hyacinth (*Eichhornia crassipes*), Duckweed (*Lemna minor*), and biochar is a viable and effective method for removing acetaminophen from synthetic wastewater. The AC-line treatment significantly outperformed the biochar-only and control treatments, highlighting the synergistic effect of phytoremediation and biochar adsorption in enhancing pollutant removal. The statistical analysis confirmed that the results were significant ($p = 0.0036$), reinforcing the reliability of this method as an alternative wastewater treatment strategy. The ability of these aquatic plants to uptake and degrade acetaminophen, coupled with biochar's adsorptive properties, suggests a sustainable and cost-effective approach to mitigating pharmaceutical pollution in water systems.

Future research could focus on optimizing plant density and biochar composition to maximize pollutant removal efficiency. Testing this method in real wastewater environments over a longer period with mixed pharmaceutical contaminants could further validate its applicability on a larger scale. Additionally, by integrating phytoremediation techniques into existing wastewater treatment frameworks, such as in wastewater stabilization ponds that are built to treat wastewater further through organic processes outside the treatment centers, there is potential to significantly reduce pharmaceutical pollution and its associated ecological risks. This process can also be implemented into constructed wetlands in rural villages with limited access to clean water, further lowering the risk of contaminated water entering the local supply.

BIBLIOGRAPHY

1. Dong, J., Jiang, P., Wang, H., Lu, R., Liu, Y., Li, Y., Gan, Y., Bolan, N. (2023, September 11). Influence of biomass feedstocks on magnetic biochar preparation for efficient pb(ii) removal. *Environmental Technology & Innovation*. <https://www.sciencedirect.com/science/article/pii/S2352186423003590>
2. Kumar, S., & Deswal, S. (2020). Phytoremediation capabilities of *Salvinia molesta*, water hyacinth, water lettuce, and duckweed to reduce phosphorus in rice mill wastewater. *International Journal of Phytoremediation*, 22(11), 1097–1109. <https://doi.org/10.1080/15226514.2020.1731729>
3. Banerjee, A., Cardozo, S., Khuntia, S. P., & Upadhyay, A. K. (2022). Chapter 19 - Phytoremediation of pharmaceutical wastes. In *Bioremediation of Emerging Contaminants from Soils: Soil Health Conservation for Improved Ecology and Food Security*. Elsevier. <https://doi.org/10.1016/B978-0-323-89874-4.00003-0>
4. Kotyza, J., Soudek, P., Kafka, Z., & Vaněk, T. (2010). Phytoremediation of pharmaceuticals—Preliminary study. *International Journal of Phytoremediation*, 12(3), 306–316. <https://doi.org/10.1080/15226510903563900>
5. Singh, S., Pant, A., Dutta, K., Rani, R., Vithanage, M., & Daverey, A. (2024). Phytoremediation of pharmaceuticals and personal care products using the constructed wetland. *Environmental Chemistry and Ecotoxicology*, 7(2), Article 100175. <https://doi.org/10.1016/j.enceco.2024.04.001>
6. Wang, W., Wu, S., Sui, X., & Cheng, S. (2023). Phytoremediation of contaminated sediment combined with biochar: Feasibility, challenges and perspectives. *Journal of Hazardous Materials*, 452, Article 133135. <https://doi.org/10.1016/j.jhazmat.2023.133135>
7. Mohammed, A. A., Mutar, Z. H., & Al-Baldawi, I. A. (2021). *Alternanthera* spp. based-phytoremediation for the removal of acetaminophen and methylparaben at mesocosm-scale constructed wetlands. *Heliyon*, 7(10), Article e08403. <https://doi.org/10.1016/j.heliyon.2021.e08403>

AWARDS

**These awards capture the work of the 24-25 academic calendar year
for Nikola Tesla STEM High School students and staff**

**AMERICAN ASSOCIATION OF UNIVERSITY
WOMEN (AAUW) AWARDS: April 2025
High School Scholar Recognition**

Madeleine Vanderheiden
Layla Kaim
Chloe Bissonnette

**ANATOMAGE NATIONAL TOURNAMENT
Las Vegas, Nevada: May 2025**

Ranking in the top 0.76%
Shuhui Yang, Kunrong Wang, Adhrith Vutukuri

**AMERICAN INVITATIONAL MATHEMATICS
EXAMINATION (AIME) QUALIFIERS: Feb 2025**

Ekaansh Agrawal , Jai Bindlish, Pranav Kannepalli, Reese Long, Alexander Peev, Max Xie

**CHESS
WASHINGTON STATE HIGH SCHOOL TEAM
CHESS CHAMPIONSHIPS
Stanwood, WA: March 2025**

2A division
5th Place Vivaan Bansal, Anish Bharadwaj, Ujwal Garine, Devansh Sharma, and Stephanie Velea

Singles Tournament
2nd Place Ender Ramsby

Tournament All-Star
Devansh Sharma

**ENVIROTHON, KING COUNTY
REGIONALS: April 2025**

2nd Place Saharsh Bhargava, Sky Chen, Aarav Jain, Khush Kothari, Saketh Madiraju, Ayaan Oberoi, Nikhil Patil, Aditya Patnaik, Shipra Prajapati, Sree Sarvepalli, Madeleine Vanderheiden

EXPLORAVISION, TOSHIBA / NSTA

Honorable Mentions

Brenden Lee, Felicity Souza, Casey Uyehara, “AmyloTrak: A Diagnostic Test for Alzheimer’s Via Beta Amyloid Detection in Urine”

Saahiti Mamidipaka, Harshita Raja, “Buggin’ Cancer with TherRIA: A Novel Approach to Targeting Tumors”

EXPLORAVISION, TOSHIBA / NSTA

Honorable Mentions Cont.
Gautham Ayer, Danyaal Hasan, Arjun Rao, Marco Ruiz, “Tumor Necrosis Factor Attacker (TNFA): The Next-Gen TNF- α Blocker “

Marta Jayantilal, Gracie Li, Jessica Schweizer, Will Stillwell, “VasLink: The Reversible Vasectomy of the Future”

**FUTURE BUSINESS LEADERS OF AMERICA
(FBLA)**

**Future Business Leaders of America (FBLA) State
Bellevue, Washington: April 2025**

OBJECTIVE TESTING:

Computer Applications
5th Place Elaine Jiang
6th Place Rhea

Computer Problem Solving
10th Place Alexander Nyczkowski

Economics
7th Place Max Xie

Financial Math
4th Place Max Xie

Introduction to Business Concepts
3rd Place Shaurya
9th Place Ihita Gupta

Introduction to Financial Math
5th Place Ihita Gupta
6th Place Bryan Li

Introduction to Information Technology
6th Place Hem Akarapu

Introduction to Marketing Concepts
7th Place Vivaan Bansal

Networking Infrastructures
5th Place Alexander Nyczkowski
8th Place Hem Akarapu

Personal Finance
5th Place Himanshu Dasari

Securities and Investments
10th Place Scott Alexander

UX Design

4th Place JD Kim

Word Processing

6th Place Tanvi Patil

PRESENTATION EVENTS:*Banking and Financial Systems*

1st Place Varun Hari, Sriram Marasanapalle-Kalle

Business Ethics

7th Place Adya Chandar, Diya Kowshik

Business Management

1st Place Nitya Bhartiya, Saahiti Mamidipaka

Business Plan

1st Place Safwan Hasan, Anirudh Narasimhan,
Rafael Reyes

Coding and Programming

4th Place Amogh Atreya, Aarush Kapoor,
Arnav Singh

7th Place Alisha Goyal, Siddhesh Nikam,
Cindy Yang

Computer Game and Simulation

6th Place Jai Bindlish, Aashrith Pisupati,
Pragyan Singh

Data Analysis

3rd Place Shravan Mahesh Kumar, Keerthi Rajesh

Digital Animation

8th Place Anastasiya Netchaeva

9th Place Adeline Wheeler

Digital Video Production

4th Place Harshini Kondaparthi, Kavya Shenoy

5th Place Sanjay Nagesh, Swayam Pharate,
Nino Saeki

Electronic Career Portfolio

3rd Place Luis Padilla

4th Place Pallavi Thiruveedu

5th Place Rhea Singh

6th Place Anika Joshi

Entrepreneurship

10th Place Samyuktha Krishnan

Financial Statement Analysis

5th Place

Richa Badi, Veda Hiremath, Shikha Singh

Future Business Educator

2nd Place Diya Balaji

Graphic Design

1st Place Iris Dey, Tejas Panja, Shreyansi Swain

International Business

3rd Place Diya Balaji

4th Place Harshita Raja, Sahaana Sridhar

Introduction to Business

4th Place Samanyu Ezhilarasan, Zakary Tran

Introduction to Social Media

7th Place Ananiah Gashaw, Nethaniah Gashaw

Local Chapter Annual Business

7th Place Marta Jayantilal, Elaine Jiang

Management Information Systems

2nd Place Alekhya Nidadavolu, Rhea Singh

Network Design

7th Place Siddharth Mulupuru, Krish Senthil

On-Campus Commerce

1st Place Bhrugu Mehta, Kanishk Yadav

Parliamentary Procedure

8th Place Eva Jain, Aahana Mittal,
Megha Ramachandran, Riya Singh

Public Service Announcement

1st Place Adya Chandar, Diya Kowshik

Public Speaking

2nd Place Shravan Mahesh Kumar

Sales Presentation

2nd Place Rehet Hans, Sriram Marasanapalle-Kalle

5th Place Saahiti Mamidipaka, Harshita Raja

Social Media Strategies

1st Place Varshini Hari, Layla Kaim

Sports and Entertainment Management

4th Place Raehan Baliga, Vihaan Fadnavis,
Rajvir Mishra

6th Place Shaurya Gupta, Pragyan Singh

Visual Design

1st Place Lauren Ouyang,

2nd Place Nimeshikaa Saravanakumar
Pia Dave, Kaavya Trivedi

Website Coding and Development

3rd Place Aakarsh Balla, Sacchin Saravanan,
Abhinav Vallabhaneni
4th Place Iris Dey, Safwan Hasan,
Rishabh Menasinakayi

Website Design

2nd Place Marta Jayantilal, Elaine Jiang,
Sahaana Sridhar
6th Place Vihaan Fadnavis, Rajvir Mishra

Future Business Leaders of America (FBLA) Nationals Anaheim, California: June-July 2025

Banking and Financial Systems

7th Place Sriram Marasanapalle-Kalle, Varun Hari

Business Plan

2nd Place Safwan Hasan, Jaskirat Singh,
Rafael Reyes

Coding and Programming

7th Place Arnav Singh, Amogh Atreya

Data Analysis

9th Place Keerthi Rajesh

Future Business Educator

3rd Place Diya Balaji

Introduction to Business Presentation

5th Place Samanyu Ezhilarasan, Zakary Tran

Network Infrastructures

5th Place Alex Nyczkowski

Personal Finance

6th Place Himanshu Dasari

Public Speaking

5th Place Shravan Mahesh Kumar

Social Media Strategies

10th Place Layla Kaim, Varshini Hari

Sports and Entertainment Management

7th Place Rajvir Mishra, Raehan Baliga,
Vihaan Fadnavis

Website Design

1st Place Elaine Jiang, Alisha Goyal

Word Processing

1st Place Tanvi Patil

HOLOCAUST FOR HUMANITY AWARDS

2025 Dee Simon Teacher of Excellence Jenai Sheffels

Art Contest

2nd Place Arina Pastukhova
Honorable Mention Elizabeth James

JOURNAL OF HIGH SCHOOL SCIENCE

Defne Dingiloglu, "The Untapped Potential of the Endophytic Actinomyces, Streptomyces Scabiei, in Producing Antibiotic and Antitumor Secondary Metabolites", published in The Journal of High School Science.

KEY CLUB FUNDRAISER: April 2025

Club Advisor Mrs. Jenai Sheffels.

Nikola Tesla STEM Key Club raised \$1,824 for The Sophia Way, a women's shelter and outreach center serving Seattle and the Eastside. Board members: , Edbert Chenjp, Zita Cheung, Anoushka Gharat, Laura Shang, Kat Xie, Grace Zhao

MADEPOSSIBLEBYFERTILIZER

DIGITAL CONTEST: October 2024

1st Place Taein Kim, "Fertilizer: A Key to Ending Hunger" – An engaging animated video connects the fight against global hunger to the power of fertilizer. Through the story of a struggling tomato plant, viewers learn how nutrients like nitrogen, phosphorus, and potassium can help grow more food and move us closer to Zero Hunger.

MODEL UNITED NATIONS (MUN) CLUB

Annual Seattle MUN Conference

Best Delegate Alex Zhang

Outstanding Position Paper for exemplary writing and research preparation in the Ad-hoc committee:

Marilyn Zhao

**WASHINGTON MUSIC EDUCATORS
ASSOCIATION: SOLO AND ENSEMBLE
COMPETITION**

Solo Flute- Piccolo
3rd Place Minami Chiba

NATIONAL MERIT SEMI FINALISTS

Achintya Agrawal
Rajiv Arvind
Bhanu Atmakuri
Saharsh Bhargava
Ryan Chung
Keshav Iyer
Elizabeth James
Pranav Kannepalli
Aishika Koppolu
Keying Li
Sarthak Lodha
Reese Long
Nicole Mihaylova
Megha Ramachandran
Sacchin Saravanan
Rohit Sundaram
Kaavya Trivedi
Anisha Vaish
Edison Wheeler
Shuhui Yang
Marilyn Zhao

**ORIENTEERING
WASHINGTON INTERSCHOLASTIC
ORIENTEERING LEAGUE (WIOL)**

2024-2025 Winter League #3: Dec 7, 2024

Pierce College Fort Steilacoom Park, Lakewood, WA

JV Female

1st Place Anastasiya Netchaeva

JV Male

14th Place Paul Denton
26th Place Tyler Vandevoort

**2025 MATE REGIONAL ROV COMPETITION
PACIFIC NORTHWEST**

The MATE ROV Competition uses underwater robotics (a.k.a. remotely operated vehicles or ROVs) to inspire and challenge students to learn and creatively apply scientific, engineering, and technical skills to solve real-world problems. This competition challenges students from all over the world to tackle missions based on scenarios from the workplace, encouraging students to work together, network, and learn from technical professionals and each other.

ROV Class: Ranger Class

6th Place

Abhinav Nadupalli, Alexander Nyczkowski,
Arina Pastukhova, Bhruhu Mehta, Chloe Bissonnette, Daniel Zhou Ekaterina Rukhlinskaia, Evan Galuska,
Feodor Berov, Harshveer Hooda, Isahar Jimenez,
Julien Wang, Kunrong Wang, Leila Burg, Pratul Saini,
Rachel Cai, Rhea Singh, Rory Shing, Samuel Yang,
Siddhesh Nikam, Taein Kim, Tyler Chen, Vsevolod Titov,
Winson Wan

**REGENERON SCIENCE TALENT SEARCH
January 2025**

Scholar

Anson Chen, “*DATAtect: Forecasting Domoic Acid Levels From Harmful Algal Blooms Along the Pacific Northwest Coast*”

Teacher of Merit: Mr. Nigel Crowe

SCIENCE AND ENGINEERING FAIRS

CENTRAL SOUND REGIONAL SCIENCE AND ENGINEERING FAIR (CSRSEF): March 2025

Behavioral and Social Sciences

4th place Aditri Sanghubhatla, “The Impact of Wave Frequencies and Tactile Stimulation on Alzheimer’s Disease Memory Performance”

Robotics and Emerging Technology

3rd place Sarah Youn and Aditi Kandiah, “Classification of ADHD through machine learning models in children using traditional statistical and CNN based feature extraction of raw EEG data”

Computational Biology & Bioinformatics

3rd place Keerthi Rajesh, “Identification and Validation of Potential Biomarkers and Pathways for Idiopathic Pulmonary Fibrosis using In Silico Methods”

4th place Abbas Shakir, “Synthesizing Novel CD40x-CD137 Bispecific Antibodies for B-Cell Malignancies and Solid Tumors Through Deep Learning and Experimental Methods”

Chemistry, Material Science and Energy

2nd place Madeleine Vanderheiden, “Applicability of diazirine-based dynamic crosslinking for long-term plastics”

Environmental Science and Engineering

3rd place Saranya Voruganti, “Investigating the Adsorption Capabilities of Human Hair for Mitigating Carbamazepine Pollution”

Biomedical Science and Engineering

2nd place Aanchal Batwara, “A Machine Learning-Driven, Voice & Vision Controlled 3D-Printed Transhumeral Prosthesis: A Low-Cost Solution Powered by Raspberry Pi”

3rd place Joseph Wang, “Development of a Low-Cost, EMG-Guided Prosthetic Arm for Accessible Myoelectric Control”

Life and Health Sciences

3rd place Kunrong Wang, “Developing a High-Efficiency Device to Induce Traumatic Brain Injury in Larval Zebrafish (*Danio Rerio*)”

Embedded Systems and Software

1st place Enzo Van Thorne, “Tracing the Source: Identifying AI-Generated Code and Its Origins Using AST-Based Techniques Bias in Large Language Models”

4th place Ryan Chhabra, “A Novel Communication Protocol for Real-time Vehicle Synchronization and Safer Adaptive Cruise Control”

SPECIAL CATEGORY AWARDS:

Association of Women Geoscientists Award

Kavana Inamdar, “Identifying Impact of Fishing Activity on Coral Bleaching using Remote Sensing Data”

NOAA Pulse of the Planet Award

Sathkrith Segu, “The Impact of Credibility When Hitting ‘Purchase Now’”

US Air Force Award

Shreeyaa Muhunthan, “Beta Brain Wave Activity in the Frontal Lobe of Musicians: An EEG Study”

Ishaan Kothari, “Optimizing Robot Paths Through Parametric Curve Analysis”

Alexander Helfman, “An Evaluation of Hypo and Hyper Methylation within Primary and Metastatic Prostate Cancer”

Varshini Hari, “American Sign Language to Speech Conversion Model Using Machine Learning Techniques”

SWE Society of Women Engineers Excellence in Engineering Award

Sanskriti Sinha, “Optimizing Coastal Defense: A Machine Learning solution for erosion control and sea level rise adaptation”

Stockholm Junior Water Prize

Rudra Malatpure, “Comparing Bacterial Efficiency in Microbial Fuel Cells for Dairy Wastewater Treatment”

WASHINGTON STATE SCIENCE AND ENGINEERING FAIR (WSSEF): March 2025

SPECIAL CATEGORY AWARDS:

Dharmon Ghanbari ~ Excellence in the Chemical Sciences Award

Madeleine Vanderheiden

NASA Earth System Science Award

Sree Sarvepalli

NOAA ~ Taking the Pulse of the Planet Award

Sanskriti Sinha

Norman E. Borlaug Excellence in Sustainability Award

Rudra Malatpure

Ohio Wesleyan University ~ Outstanding Achievement in Science Scholarship

Kaya Hurmali

Plant Sciences Achievement Award

Eeshan Misra

Regeneron Biomedical Science Award

Mehuli Nath

Ricoh Sustainable Development Award

Eeshan Misra

Society for in Vitro Biology Outstanding Achievement for Ability and Creativity Award

Anyu Kumar

Karma Ibrahim

St. Michaels Medical Center Biomedical Engineering and Health Science Award

Joseph Wang

Kaya Hurmali

Statistical Thinking Award

Kunrong Wang

U.S. Army Awards for Outstanding STEM Accomplishments

Varshini Hari

Washington State University ~ College of Arts and Sciences Scholarship

Abbas Shakir

Water Environment Federation Stockholm Junior Water Prize

Eeshan Misra

Rudra Malatpure

Saranya Voruganti

Sree Sarvepalli

2025 WSSEF FIRST PLACE TROPHY GRADES 9 – 12

Aanchal Batwara, “A Machine Learning-Driven, Voice & Vision Controlled 3D-Printed Transhumeral Prosthesis: A Low-Cost Solution Powered by Raspberry Pi”

Ryan Chhabra, “A novel communication protocol for real-time vehicle synchronization and safer adaptive cruise control”

Basudha Ghose, “Antibacterial Activity of Stinging Nettle, Devil’s Club, and Oregon Grape against Escherichia Coli and Bacillus Cereus”

Sachita Ghosh, “Predicting Diabetes Using Machine Learning Algorithm with EEG Data”

Alexander Helfman, “An Evaluation of Hypomethylation within primary prostate cancer”

Vaidehi Hinduja, “Endocrine-Disrupting Chemicals and Carcinogenesis in Mammary Tissue: Analyzing Environmental Contributions to Breast Carcinoma Risk”

Kaya Hurmali, “Wireless, affordable, and durable oto-acoustic measurement device for hearing screening”
Kavana Inamdar, “Identifying Impact of Fishing Activity on Coral Bleaching using Remote Sensing Data”

Aditi Kandiah, “Classification of ADHD through machine learning models in children using traditional statistical and CNN based feature extraction of raw EEG data”

Shriya Kapila, “NanoDesigner – An AI-Driven Design of Blood Brain Barrier-Penetrating Nanobodies for Neurodegenerative Diseases”

Ishaan Kothari, “A Novel Approach to Autonomous Pathing for Holonomic Robots via a Parametric Curve Algorithm”

Shriya Krishnan, “Using Behavioral Methods to measure the effects of Joint Attention learning in male and female infants with a familial history of autism”

Prisha Kudikala, “Music and Memory: The Effect of Music on your Short-Term Memory”

Aravindh Kumar, “The Effect of Low and High Frequency Sounds on Mental Cognition in Adults”

Xintian Li, “Investigating the Effects of Scent on Self Perceived Confidence and Personality”

Eeshan Misra, “A Green Clean: Analyzing Plant Biology in Phytoremediation of Acetaminophen from Wastewater Using Water Hyacinth, Duckweed, and Biochar”

Shreeyaa Muhunthan, “Beta Brain Wave Activity in the Frontal Lobe of Musicians: An EEG Study”

Akchara Mukunthu, “Structural Analysis of the Giardia intestinalis 20S Proteasome $\beta 5$ Subunit and Drug Repurposing for Giardiasis”

Mehuli Nath, “*Expanding Accessibility: Improving Speech Recognition for Neurological and Motor Speech Disorders*”

Suvan Pal, “*A Generative Approach to Designing Novel Materials for Solid-State Electrolytes*”

Dhaani Pohuja, “*Investigating the Impact of Mental Fatigue on Male Teenagers compared to Female Teenagers*”

Lea Prunk, “*Investigating the Impact of Self-Directed Learning on Adolescent Motivation According to Self-Determination Theory*”

Rewa Rathee, “*Morning vs. Evening: Cognitive Training Techniques for Enhanced Memory Retention*”

Aditri Sanghubhatla, “*The Impact of Wave Frequencies and Tactile Stimulation on Memory Performance in Stressed Teenage Students*”

Aadithya Sarvajith, “*Designing a Polymeric Heart Valve: Analyzing the Role of Chitosan, Polylactic Acid, and Polyethylene Blends in Enhancing Tensile Strength and Flexibility*”

Sree Sarvepalli, “*Investigating the effect of PET microplastics on the heat-shock response of aquatic cyanobacteria*”

Sathkrith Segu, “*Scorching Summers: The Link Between Heatwaves and Apis mellifera Mortality*”

Abbas Shakir, “*Synthesizing Novel CD40xCD137 Bispecific Antibodies for B-Cell Malignancies and Solid Tumors*”

Thea Sima, “*A Diffusion-Based Approach to Restoring Ancient Oracle Bone Characters*”

Sanskriti Sinha, “*CoRA: Machine Learning for Coastal Risk Assessment and Defense Optimization*”

Shreyansi Swain, “*Identifying the Genetic Association of Cervical Cancer and Insulin Resistance Using Single Sample Gene Set Enrichment Analysis*”

Enzo Van Thorre, “*Tracing the Source: Identifying AI-Generated Code and Its Origins Using AST-Based Techniques*”

Madeleine Vanderheiden, “*Applicability of diazirine-based dynamic crosslinking for long-term plastics recycling*”

Adhrith Vutukuri, “*RADAR – Radiology Assistant for Detecting Abnormalities and Rechecks: A Novel AI-based Radiology Companion System for Identifying and Correcting Visual Misses in Chest X-Ray Scans*”

Joseph Wang, “*Development of a Low-Cost, EMG-Guided Prosthetic Arm for Accessible Myoelectric Control*”

Kunrong Wang, “*Developing a High-Efficiency Device to Induce Traumatic Brain Injury in Larval Zebrafish*”

Shuhui Yang, “*Identification of Drug Candidates Targeting Insulin-like Growth Factor 1 Receptor (IGF1R) for More Effective Alzheimer’s Disease Treatment in Women*”

Sarah Youn, “*Classification of ADHD through machine learning models in children using traditional statistical and CNN based feature extraction of raw EEG data*”

Daniel Zhou, “*Association study of single nucleotide polymorphisms in X chromosome inactivation escape genes with susceptibility to autoimmune diseases in female populations*”

SCIENCE OLYMPIAD

Camas: Division C Invitational
Camas, Washington: Dec. 2024

1st Place	Team Orion
7th Place	Team Cygnus
10th Place	Team Aquila
20th Place	Team Scorpius
21st Place	Team Hydra
23rd Place	Team Ursa

Air Trajectory

2nd Place	Sree Sarvepalli, Sriram Marasanapalle-Kalle
-----------	--

Anatomy and Physiology

2nd Place	Shuhui Yang, Kunrong Wang
4th Place	Adhrith Vutukuri, Shreya Santhosh

Astronomy

1st Place	Shipra Prajapati, Suvan Pal
-----------	-----------------------------

Codebusters

2nd Place	Tyler Chen, Rohan Viswanathan, Suvan Pal
-----------	---

Disease Detectives

2nd Place	Adhrith Vutukuri, Sachita Ghosh
-----------	---------------------------------

3rd Place Grace Zhao, Varun Hari

Ecology

2nd Place Madeleine Vanderheiden, Sree Sarvepalli

Electric Vehicle

2nd Place Sachita Ghosh, Mehuli Nath

Entomology

2nd Place Kunrong Wang, Aadya Bansal

4th Place Theo Petchenlarp, Arka Samatham

Experimental Design

1st Place Aadya Bansal, Varun Hari

Forensics

1st Place Saranya Voruganti, Mehuli Nath

Fossils

1st Place Grace Zhao, Shipra Prajapati

Geologic Mapping

1st Place Shipra Prajapati, Saharsh Bhargava

Helicopter

2nd Place Sriram Marasanapalle-kalle,
Varun Hari

Materials Science

1st Place Sachita Ghosh, Alexander Peev

Microbe Mission

1st Place Kunrong Wang, Shuhui Yang

3rd Place Neal Kedia, Shreya Santhosh

Optics

4th Place Atreya Iyer, Sahana Kaliyaperumal

Robot Tour

1st Place Rohan Viswanathan, Saharsh Bhargava

Tower

2nd Place Sriram Marasanapalle-Kalle,
Rohan Viswanathan

Write It Do It

1st Place Grace Zhao, Shuhui Yang

Quantum Quandaries

2nd Place Jai Bindlish, Novyanna Bajracharya

Science Olympiad
Raisbeck: Division C Invitational
Tukwila, Washington: Jan. 2025

4th Place Team Cygnus

7th Place Team Aquila

11th Place Team Scorpius

23rd Place Team Ursa

Anatomy and Physiology

1st Place Adhrith Vutukuri, Shreya Santhosh

2nd Place Varnika Elenthendral
Hareshwaran Thiyagarajan

3rd Place Abhinav Swain, Shaashwat Gupta

Chemistry Lab

2nd Place Saranya Voruganti, Alexander Peev

Dynamic Planet

1st Place Diya Kowshik, Aravindh Kumar

Materials Science

3rd Place Zhao, Tanvi Patil

Microbe Mission

3rd Place Evan Zhao, Aakash Ghosh

Optics

1st Place Shivank Mohanty, Sahana Kaliyaperumal

2nd Place Aditya Patnaik, Sia Gupta

Quantum Quandaries

2nd Place Aarav Jain, Jai Bindlish

Science Olympiad Regional Competition
Bothell, Washington: Mar. 2025

1st Place Team Orion

5th Place Team Cygnus

9th Place Team Aquila

15th Place Team Scorpius

20th Place Team Ursa

28th Place Team Hydra

Air Trajectory

2nd Place Sriram Marasanapalle-Kalle,
Sree Sarvepalli

Anatomy and Physiology

3rd Place Shuhui Yang, Kunrong Wang

Astronomy

2nd Place Shipra Prajapati, Suvan Pal

Codebusters

2nd Place Tyler Chen, Rohan Viswanathan,
Suvan Pal

Dynamic Planet

3rd Place Saharsh Bhargava, Shipra Prajapati

Ecology

1st Place Ananya Rabindranath, Aravindh Kumar
2nd Place Sree Sarvepalli, Aadya Bansal
3rd Place Neal Kedia, Anisha Upasani

Electric Vehicle

2nd Place Sriram Marasanapalle-Kalle, Mehuli Nath

Entomology

2nd Place Aadya Bansal, Kunrong Wang

Experimental Design

2nd Place Aadya Bansal, Varun Hari
3rd Place Aravindh Kumar, Diya Kowshik

Forensics

2nd Place Shreyansi Swain, Aadya Bansal

Fossils

1st Place Saharsh Bhargava, Shipra Prajapati

Geologic Mapping

3rd Place Saharsh Bhargava, Shipra Prajapati

Helicopter

2nd Place Sriram Marasanapalle-Kalle, Varun Hari

Microbe Mission

2nd Place Neal Kedia, Shreya Santhosh

Robot Tour

1st Place Saharsh Bhargava, Rohan Viswanathan

Tower

2nd Place Kairav Sapru, Dhaani Pohuja
3rd Place Sriram Marasanapalle-Kalle,
Rohan Viswanathan

Write It Do It

1st Place Grace Zhao, Shuhui Yang

**Science Olympiad State Competition
Ellensburg, Washington: April. 2025**

2nd Place Team Orion
8th Place Team Cygnus

Air Trajectory

2nd Place Sree Sarvepalli,
Sriram Marasanapalle-Kalle

Astronomy

2nd Place Shipra Prajapati, Suvan Pal

Codebusters

2nd Place Tyler Chen, Rohan Viswanathan,
Suvan Pal

Disease Detectives

3rd Place Sachita Ghosh, Adrith Vutukuri

Dynamic Planet

1st Place Saharsh Bhargava, Shipra Prajapati

Ecology

1st Place Madeleine Vanderheiden, Sree Sarvepalli

Electric Vehicle

3rd Place Mehuli Nath, Sriram Marasanapalle-Kalle

Entomology

3rd Place Aadya Bansal, Kunrong Wang

Experimental Design

4th Place Aadya Bansal and Varun Hari

Forensics

3rd Place Shreyansi Swain, Aadya Bansal

Fossils

2nd Place Shipra Prajapati, Grace Zhao

Geologic Mapping

1st Place Saharsh Bhargava, Shipra Prajapati

Microbe Mission

2nd Place Kunrong Wang, Shuhui Yang
3rd Place Shreya Santhosh, Neal Kedia

Optics

3rd Place Sia Gupta and Aditya Patnaik

Robot Tour

1st Place Rohan Viswanathan, Saharsh Bhargava

Write It Do It

1st Place Grace Zhao, Shuhui Yang

STEM EXPO: May 2025

STARTUP FINALISTS

Campus CookUp

Kousthubh Durgaraju
Jeff Gao
Ahanya John
Raghav Kalyana Sundaram

StudyLoop

Aadya Bansal
Marta Jayantilal
Alekhya Nidadavolu

ClearSign

Siddhesh Nikam
Sanjay Nagesh

Sixers

Pranav Kannepalli
Pranav Takrani
Sarathak Lodha

Pawsi

Stephanie Velea

STARTUP AWARDS

Best Overall Startup

Sixers

Entrepreneurship Award

Pawsi

Audience Favorite

ClearSign

Gallery Walk Audience Favorite

Clarity Media

RESEARCH AWARDS

Aadithya Sarvajith, *Designing a Polymeric Heart Valve: Analyzing the Role of Chitosan, Polylactic Acid, and Polyethylene Blends in Enhancing Tensile Strength and Flexibility*

1st Place Madeleine Vanderheiden, *Applicability for Diazirine-Based Dynamic Crosslinking for Long-Term Plastics Recycling*

2nd Place Abbas Shakir, *Synthesizing Novel CD40x-CD137 Bispecific Antibodies for B-Cell Malignancies and Solid Tumors*

3rd Place Aadithya Sarvajith, *Designing a Polymeric Heart Valve: Analyzing the Role of Chitosan, Polylactic Acid, and Polyethylene Blends in Enhancing Tensile Strength and Flexibility*

TECHNOLOGY STUDENT ASSOCIATION (TSA)

Technology Student Association (TSA) State
Spokane, Washington: April 2025

Animatronics

1st Place Anoushka Gharat, Sahana Kaliyaperumal, Akanksha Revuru

2nd Place Aarush Kapoor, Arnav Mittal, Naman Mutalik Desai

Audio podcasting

7th Place Evan Wang, Adeline Wheeler

10th Place Aadya Bansal, Zita Cheung, Mahika Kalra, Shreeyaa Muhunthan, Keerthi Rajesh, Pallavi Thiruveedu

Biotechnology design

1st Place Iris Dey, Raghav Kalyana Sundaram, Sanjana Katti, Saranya Voruganti

6th Place Rehet Hans, Shravan Mahesh Kumar, Sriram Marasanapalle-Kalle, Neil Parikh, Jaskirat Singh

Board Game Design

5th Place Zixuan Shang, Sanskriti Sinha, Pallavi Thiruveedu

Cad 3D, Engineering

10th Place Zifeng Gao

Career Prep

9th Place Shubh Jain

Children's Stories

6th Place Sia Gupta, Sahana Kaliyaperumal, Shriya Kapila, Tanisha Karthikeyan, Vivian Li, Anisha Upasani

Coding

5th Place Kaushal Dabbiru, Aarush Kapoor

7th Place Ryan Chhabra, Ishaan Kothari

Data Science And Analytics

2nd Place Raghav Kalyana Sundaram,
Pranav Kannepalli

Debating Technology Issues

9th Place Vivian Li, Sanskriti Sinha

Digital Video Production

4th Place Arthur Chen, Sethmi Ekanayake,
Anoushka Gharat, Mina Liu,
Manas Mishra, Anvie Sana
5th Place Cleofus Balaranjith, Soham Kulkarni,
Brian Lee, Demian Orlov, Ryan Park
6th Place Aarav Gumber
9th Place Vicky Ding, Siyona Kane,
Akshara Pathangi, Yu Tane Quah

Engineering Design

3rd Place Inesh Dey, Iris Dey, Shreetya Muhunthan,
Mehuli Nath, Pallavi Thiruveedu,
Saranya Voruganti
4th Place Cleofus Balaranjith, Anish Bharadwaj,
Brian Lee, Shivank Mohanty,
Rohan Viswanathan
6th Place Elaine Jiang, Aarush Kapoor, Vivian Li,
Sriram Marasanapalle-Kalle, Rhea Singh
7th Place Raehan Baliga, Rohan Baliga,
Vihaan Fadnavis, Arda Keskinpala,
Rajvir Mishra, Neil Parikh

Flight Endurance

6th Place Max Xie

Forensic Science

4th Place Aadya Bansal, Saranya Voruganti

Future Technology Teacher

8th Place Naman Mutalik Desai

Geospatial Technology

1st Place Hem Akarapu, Cleofus Balaranjith,
Anindita Purkayastha
4th Place Aadya Bansal, Zixuan Shang,
Pallavi Thiruveedu

Infographic

6th Place Gabriel Austin

Manufacturing Prototype

4th Place Wenxin Fang, Shriya Kapila, Vivian Li,
Sanskriti Sinha, Darren Suwirya

Music Production

2nd Place Timothy An, Arthur Chen,
Darren Suwirya, Evan Wang

On Demand Video

10th Place Akshath Dommaraju, Vihaan Fadnavis,
Varshini Hari, Arda Keskinpala,
Arlene Shergill

Promotional Design

1st Place Aadhyia Patel
3rd Place Shriya Kapila

Robotics Challenge

4th Place Akshath Dommaraju, Vihaan Fadnavis,
Alex Merceline, Shivank Mohanty,
Rohan Viswanathan
8th Place Kaushal Dabbiru, Anirudh Narasimhan,
Mayank Pandey, Neil Parikh,
Kanishk Yadav, Cindy Yang

Silent Film

2nd Place Megan Halim, Harshita Raja
3rd Place Shaan Gada, Kirit Singh

Software Development

1st Place Inesh Dey, Anirudh Narasimhan,
Enzo Van Thorre
4th Place Achintya Agrawal, Ryan Chhabra,
Kaushal Dabbiru, Pranav Kannepalli,
Ishaan Kothari, Shailja Tyagi
10th Place Paul Grigorovitch, Andrew Ma,
Paarth Nathani

Stem Mass Media

1st Place Diaa Bijoy, Sacchin Saravanan

Transportation Modeling

3rd Place Rhea Singh
4th Place Anvie Sana
7th Place Anindita Purkayastha
9th Place Aravindh Kumar

Trebuchet

4th Place Rohan Gupta, Joseph Wang

Virtual Reality Simulation (VR)

4th Place Hem Akarapu, Bhanu Atmakuri,
Vivaan Pradhan, Aryaman Rtnunjay,
Sacchin Saravanan, Kirit Singh

Webmaster

1st Place Yohaana Ardeni, Inesh Dey, Ishaan Kothari,
Rishabh Menasinakayi,

**Technology Student Association (TSA) Nationals
Nashville, Tennessee: June-July 2025**

Children's Stories

5th Place Sia Gupta, Sahana Kaliyaperumal,
Shriya Kapila, Tanisha Karthikeyan,
Vivian Li, Anisha Upasani

Data Science And Analytics

4th Place Raghav Kalyana Sundaram

Digital Video Production

10th Place Arthur Chen, Sethmi Ekanayake,
Anoushka Gharat, Mina Liu,
Manas Mishra, Anvie Sana

Engineering Design

6th Place Inesh Dey, Iris Dey, Shreetya Muhunthan,
Mehuli Nath, Pallavi Thiruveedu,
Saranya Voruganti

Technology Bowl

5th Place Inesh Dey, Ishaan Kothari, Sathkrith Segu

Webmaster

1st Place Yohaán Ardeni, Inesh Dey, Ishaan Kothari,
Rishabh Menasinkayi, Anirudh
Narasimhan, Sathkrith Segu

SEMIFINALISTS:

Animatronics

Aadya Bansal, Aarush Kapoor, Arnav Mittal

Future Technology Teacher

Sia Gupta

Robotics

Akshath Dommaraju, Alex Merceline, Shivank Mohanty,
Rohan Viswanathan

**VA VENTURES INTERNSHIP
VA Ventures (Veterans Health Administration)**

Interns: Evan Galuska, Yohaán Ardeni, Dhruthi Vellala
Internship to research and design prototypes that utilize 3D
printing to repurpose resin waste.

Health Science

Individual Division

1st Place Anisha Upasani, "Relationship Between
Household Income and Loss to Follow Up for Neonatal
Hearing Screening in Washington State"

Group Division

3rd Place Keerthi Rajesh, Aadya Bansal, "The Cor-
relation Between Obesity, Poverty, and Esophageal Can-
cer"

Science Communication

Group Division

2nd Place Inesh Dey, Iris Dey, Ishaan Kothari, "The
Opioid Epidemic"

