Project ID: 201

Senior Division

Computational Biology and Bioinformatics

Medha Aravind
Del Norte High School
Gr. 9

Developing a Machine Learning Solution for Discovering Novel MiRNA Biomarkers for Gastric Cancer

Gastric cancer presents a significant global health challenge, often diagnosed in advanced stages, resulting in poor prognosis and high mortality rates. As a potential solution, my project aims to identify microRNA (miRNA) biomarkers for gastric cancer diagnosis, offering a non-invasive approach to improve early detection. By comparing gastric cancer miRNAs to control miRNAs, I sought to pinpoint specific diagnostic markers. Utilizing machine learning techniques, including WEKA and TensorFlow Keras, various classifiers were employed to distinguish gastric cancer-associated miRNAs. Promising results demonstrate the potential of miRNA-based diagnostics, although further validation and clinical exploration are needed to fully assess their utility in improving gastric cancer management and outcomes.

Project ID: 202

Senior Division

Computational Biology and Bioinformatics

Prakruti Bhatt
Del Norte High School
Gr. 10

E-venomics: Revolutionizing Drug Discovery with In Silico Exploration of Snake Disintegrins for alpha-v beta-3 Integrin Inhibition in Cancer Cells

AWARDS:

CSEF Qualified

Breast cancer stands as one of the most prevalent and deadly malignancies globally, posing a significant healthcare challenge. Novel therapeutic approaches are urgently needed. Among the various molecular targets implicated in breast cancer progression, the integrin αvβ3 is a promising target due to its direct correlation with cancer cell proliferation and metastasis. This integrin is abundant in breast cancer cells and contributes to their aggressive phenotype. This project proposes harnessing the inhibitory potential of snake venom disintegrins to target the αvβ3 integrin. By specifically binding to αvβ3, these disintegrins effectively inhibit its activity, thus halting cancer cell growth and metastatic spread. Project results indicate that the disintegrin derived from Vipera anatolica senliki's venom, a snake species native to Turkey exhibits the best inhibitory potency. The calculated E-TOTAL/E-SHAPE value for this disintegrin was -698.90, suggesting a strong binding affinity and interaction with the target integrin.

Project ID: 203

Senior Division

Computational Biology and Bioinformatics

Aarushi Garg
Canyon Crest Academy
Gr. 11

A Novel Deep Learning Technique to Predict Rhegmatogenous Retinal Detachment Using Ultrawide Fundus Images

AWARDS:

Association for Women in Science - Winner
Society of Women Engineers - San Diego County Section - Senior Division 2nd Place
CSEF Qualified

Rhegmatogenous retinal detachment (RRD) is among the main emergency indications in ophthalmology that can result in permanent blindness if the treatment is delayed by even a few hours. Because early detection is crucial, this study developed two different AI models to identify RRD in ultrawide fundus images (EfficientNet, CNN). A dataset of 1693 images (1017 RRD, 676 non-RRD) was utilized, with 75% (1270 images) for training and 25% (423 images) for testing. The EfficientNet model had 99.62% accuracy, 100% precision, 99.74% sensitivity (95% CI: 0.99-1.00), 99.49% specificity (95% CI: 0.99-1.00), and 100% area under the curve. The CNN model had 96.94% accuracy, 97% precision, 96.26% sensitivity (95% CI: 0.93-1.00), 97.75% specificity (95% CI: 0.95-1.00), and 99.23% area under the curve. The EfficientNet, being more computationally efficient and accurate, offers a potential solution to address the shortage of ophthalmologists, particularly in regions like South Asia, burdened with a high prevalence of visual impairment but lacking adequate healthcare personnel. Even in the South Asian country with the most ophthalmologists, only one ophthalmologist serves 250,000 individuals. This novel study demonstrates that an EfficientNet can outperform ophthalmologists with five years of experience in accurately identifying RRD, providing a valuable and environmentally friendly tool for regions with limited access to ophthalmologists.

Project ID: 204

Senior Division

Computational Biology and Bioinformatics

Diego (Toa) Hernandez
Sweetwater High School
Gr. 11

AI to the Rescue! New Technology Sees What Doctors Cannot

In our study focusing on predicting treatment response in GBM patients, we analyzed a comprehensive dataset containing clinical records (Through the Cancer Archives), imaging scans, genomic profiles, and treatment outcomes. Through applying artificial intelligence (AI) models, we aimed to assess the correlation between AI-predicted treatment response scores and actual treatment outcomes. Our data analysis revealed a statistically significant correlation between the treatment response scores predicted by AI models and the actual treatment outcomes observed in GBM patients. Specifically, we found that AI-predicted scores accurately reflected the efficacy of standard-of-care treatments, including chemotherapy and radiation therapy, in managing GBM. This significant correlation suggests that AI-driven approaches hold promise in emphasizing personalized treatment planning and prognostic assessment for GBM patients, therefore improving clinical decision-making and patient care in the management of this challenging disease. These results underscore the potential of AI technology to completely revolutionize the field of oncology by providing clinicians with valuable insights into treatment response and prognosis. By weighing out AI-driven approaches, healthcare providers can tailor treatment strategies more effectively to individual patients, ultimately enhancing overall patient outcomes and quality of life. Continued research in this rapidly evolving field is essential to refine AI models further and maximize their utility in clinical practice for patients like my mother to get the proper diagnosis bright and early for less aggressive treatment alternatives.

Project ID: 205

Senior Division

Computational Biology and Bioinformatics

Henry Hou

The Bishop's School

Gr. 9

One-Drug-Fits-All: Identifying Multi-Target Drugs with Machine Learning for Rheumatoid Arthritis

Rheumatoid Arthritis (RA) is an autoimmune disease, with the main cause being the immune system identifying the joint as a potential intruder, using cytokines as messengers to initiate the immune response. To combat this, the medical community employed a cocktail of DMARDs (disease-modifying antirheumatic drugs), with each DMARD treating a specific cytokine, necessitating the identification and repurposing FDA approved drug able to combat all major cytokines at once. In this study to inhibit multiple types of cytokines, I elect to target three of the main contributors to RA symptoms: TNF-alpha, IL-1 beta, and IL-6. Then, I locate the raw data from publicly available databases, convert the data into descriptors, and denoise the data using Principal Component Analysis. Afterwards, I used this dataset to train machine learning models for each cytokine in order to predict if a molecule can inhibit the cytokine. I then use these models to run through the testing dataset using each cytokine's machine learning model to create a set of predictions. Lastly, I can look for commonalities in each set of predictions to check if a drug can inhibit all three cytokines, successfully obtaining two qualified drugs which are predicted to inhibit all three identified cytokines, which were Mitomycin C, a treatment for cancer, and Romidepsin. I also obtained a varying scope of drugs that can inhibit two out of the three cytokines, including Orfadin, Glimepiride, Deflazacort, etc. This machine learning method can be also viable for repurposing the FDA approved drugs to treat other diseases and viruses.

Project ID: 206

Senior Division

Computational Biology and Bioinformatics

Arunraj Jeyaprakash
Canyon Crest Academy
Gr. 11

A Prediction Diagnostic Model of Glioblastoma Using Exosome Based miRNA Gene Expression in Glioma Types

No abstract submitted in application

Project ID: 207

Senior Division

Computational Biology and Bioinformatics

Emma Kao
The Bishop's School
Gr. 11

Utilizing an Autoencoder to Denoise Confocal Microscopic Images

Many areas of science and biology depend on microscopic images to determine molecular properties and the existence of microorganisms invisible to the human eye. By the nature of a microscopic lens, pixel sizes, and other factors, unwanted noise often finds its way into the microscopic images. Fluorescence microscopy, in particular, suffers from excess noise due to the bleaching of fluorophores. Thus, denoising such images is of great interest. In the past decade, artificial intelligence (AI) denoising algorithms experienced great success in image denoising; one of the most popular algorithms is the convolutional neural network (CNN), due to its ability to robustly process spatial information present in images. However, CNN autoencoders have been untested in microscopy imaging denoising tasks; further, many microscopy denoising algorithms are validated on unrealistic noise mechanisms (strictly Gaussian). Here, we use a CNN autoencoder to denoise confocal microscopic images with a realistic noise mechanism and apply the encoder layers to extract specific features in different types of cell images.

We begin by cleaning and reorganizing a dataset with images containing Poisson-Gaussian noise, mimicking error that might be encountered in a real-life noisy image. Our final dataset contains (noisy, clean) pairs that the model can train on. We then construct a CNN autoencoder capable of denoising our microscopy images, yielding competitive PSNR and SSIM with other benchmarks. Next, we use the encoder half of the model to condense significant features of our images into two dimensions (and analyze biologically meaningful patterns from these encodings). Finally, we reflect on future uses and potential challenges for further research.

Project ID: 208

Senior Division

Computational Biology and Bioinformatics

Gautam Kathiravan
Canyon Crest Academy
Gr. 11

Ensemble of Deep Learning Models for Multi-Class Classification of COVID-19, Pneumonia, Tuberculosis, and Lung Opacity in Chest X-Ray Images

The rapid evolution of artificial intelligence (AI) has significantly impacted the medical diagnostics landscape, particularly in radiology, where deep learning models play a pivotal role in aiding experts to achieve precise diagnoses. As the continuous emergence of mutated viruses introduce new complexities and diverse symptoms affecting global populations, there exists an opportunity to leverage computer vision advancements for improved accuracy and early detection.

This project focuses on the classification of respiratory conditions, including COVID-19, Pneumonia, Tuberculosis, and Lung Opacity, utilizing chest X-ray images. The proposed methodology adopts an ensemble approach, amalgamating multiple deep learning models to enhance the reliability and robustness of the diagnostic system. Various neural network models, including Convolutional Neural Networks (CNNs) and Vision Transformers have undergone training for the classification of patient X-ray images into distinct categories, encompassing COVID-19, Pneumonia, Tuberculosis, Lung Opacity, or Normal. This training utilized a dataset sourced from the publicly available Harvard Dataverse, comprising a labeled collection of X-ray images corresponding to different diseases.

This dataset was employed for both testing and training purposes within the scope of this project. The dataset contains approximately 32,687 images with 5 classes and split into training, validation and test groups. The DenseNet CNN model had validation accuracy of 94% and Vision transformer with 85.41%. The ensemble approach with CNN and Vision Transformer produced a test accuracy of 94.15% for random forest and 93.02% for logistic regression method.

Project ID: 209

Senior Division

Computational Biology and Bioinformatics

Ronit Khushu
Francis Parker School
Gr. 10

Novel Method of Using 3D View for Personalized Health Monitoring and Insights

AWARDS:

Kaiser Permanente Blue Ribbon Award

Traditional healthcare for neuromuscular disorders relies on subjective data like patient self-reporting and physical examinations, which can miss subtle changes. This was sadly true for my grandmother who experienced early gait changes in Parkinson's disease (PD) that went unnoticed during regular check-ups, leading to a late diagnosis and significant challenges. Inspired by my grandmother's experience, I explored how emerging technologies such as LiDARs, cameras, computer vision, and artificial intelligence (AI) can transform how we observe and monitor changes to the human body.

In this project I developed a method that utilizes LiDAR, camera and computer vision to capture 3D point cloud scans of a human body at different times, and then compare them to visualize the changes. Additionally, I used 3D pose estimation AI models and analyzed differences in poses, providing insights into the human body and changes over time.

I applied this to a person undergoing a 30-day fitness program, revealing valuable insights about changes in his body, such as reduction in abdominal fat, increase in quadriceps muscles and no changes to arms. Additionally, I used a 90%-accurate deep learning AI model, to estimate and analyze 3D poses, highlighting changes in Parkinson's patients, such as reduced arm movements by 40° and shorter strides by 15".

This detailed and accurate 3D monitoring opens exciting possibilities for both healthcare and fitness. By making such comprehensive data readily available, we empower individuals and healthcare providers to make informed decisions, leading to personalized and improved care.

Did Not Attend Judging

Project ID: 210

Senior Division

Computational Biology and Bioinformatics

Saanvi Rao
Canyon Crest Academy
Gr. 11

In-silico Cancer Drug Development with Predictive Models Using Cancer Cell Line Datasets

Cancer is a disease characterized by the uncontrolled growth and spread of abnormal cells. It begins with mutations in the DNA sequence of cells, disrupting normal cell regulation mechanisms. These mutations can occur spontaneously or be induced by various external factors such as exposure to carcinogens, radiation, or viruses. While the human body possesses elaborate mechanisms to repair damaged DNA, errors can still occur, leading to the accumulation of genetic abnormalities that disrupt the normal regulation of cell growth and division emphasizing the need for continued efforts in prevention, diagnosis, and therapy development.
The use of data science techniques to determine the best drug treatment options shows promise. Several institutions such as the NCI (National Cancer Institute) make available datasets mapping Cancer cell lines to gene mutations, and altered signaling pathways. Also available are datasets that map drug sensitivity to cancer cell lines. Combining the drug sensitivity datasets with the mutation and signaling pathways datasets allows the training of computational models to help predict the best drug option to treat a certain mutation, allowing for a more personalized treatment based on the specific set of mutations in a particular cancer occurrence.

Project ID: 211

Senior Division

Computational Biology and Bioinformatics

Adham Sadek
Canyon Crest Academy
Gr. 10

Using AI to Improve Type 1 Diabetes Management

Type 1 Diabetes impacts 8.4 million individuals worldwide, arising from the loss of pancreatic function. Patients navigate the demanding task of monitoring blood glucose levels and administering insulin. The typical challenges often include the hurdle of predicting blood glucose levels 15 to 30 minutes in the future, and estimating carbohydrate intake, particularly with unlabeled foods. In response to these challenges, a proposed solution integrates two innovative approaches that use state of the art AI technology.

The first solution involves a Python-trained LSTM deep learning model, which forecasts blood glucose levels 30 minutes ahead with high accuracy. Leveraging the DiaTrend dataset which includes data for 54 subjects collected using continuous glucose monitoring, this model achieves a 95% F1 score for hyperglycemia and 82% for hypoglycemia, providing invaluable insights for proactive management. Secondly, utilizing a ChatGPT 4 generative AI model that takes as input both an image and a text prompt, food images are analyzed to calculate calorie and carbohydrate information. The results for simple fast-food images were accurate, but more complex food images were less accurate.

The proposed solutions present promising methods for improving Type 1 Diabetes management. By harnessing cutting-edge AI technologies, it aims to alleviate the burden of constant monitoring and insulin administration. In the future, I intend to work on the refinement and integration of these models into a cohesive application running on a mobile phone. Training the generative AI model with more food images will help improve the results for a wider range of cuisine.

Project ID: 212

Senior Division

Computational Biology and Bioinformatics

Anne Sun

San Marcos High School

Gr. 11

Alcohol and Anxiety Comorbidity through Epigenetics

AWARDS:

BD "Advancing the World of Health" - Senior Division 2nd Place

Alcohol use disorder (AUD) and anxiety disorders are prevalent public health issues with substantial comorbidity. This study investigates the comorbidity of alcohol use and anxiety disorders through an epigenetic lense by analyzing the similarity of epigenetic changes caused by both these conditions. This study utilized data from the National Institutes of Health's Gene Expression Omnibus (GEO) database, gene expression profiles from brain regions associated with AUD and anxiety disorders were analyzed using DESeq2 and DAVID Functional Annotation Clustering tools. Differential gene expression analysis revealed shared pathways and functions impacted by both conditions, including cell signaling, GTPase activity, membrane activity, ER signaling and stress, and DNA-binding. Moreover, the analysis revealed consistent impacts of alcohol and anxiety on various brain regions, including the nucleus accumbens, amygdala, and hippocampus. These regions exhibited similar patterns of gene expression alterations, with both up and downregulated genes sharing overlapping functional roles across all three brain regions. This study illuminates the interconnected nature of alcohol use and anxiety disorders, with a focus on affected genetic pathways and gene expression patterns through an epigenetic lenses that underscore the obscure relationship between these conditions. Keywords: alcohol, amygdala, stress, anxiety, hippocampus, brain, nucleus accumbens

Project ID: 213

Senior Division

Computational Biology and Bioinformatics

James Sun

Canyon Crest Academy

Gr. 11

Identifying Gene Targets for Biocontainment of C. ljungdahlii Using Deep-learning Tuned Metabolic Models

Clostridium ljungdahlii is a bacteria that can be used to produce biofuels, and as usage of this industrial strain increases, it is important to develop secure biocontainment methods to prevent escape of *C.ljungdahlii* from contained environments. While previous research has already been done on mechanisms to activate biocontainment strategies, no research has been done specifically on identifying the most effective essential genes for biocontainment, mainly because it is prohibitively difficult to screen thousands of genes in search of better targets. This study takes a systems biology approach to this problem by applying metabolic models and omics data to identify gene targets for biocontainment of *C.ljungdahlii*. Simulations were done with ME-models of *C.ljungdahlii* in three different growth mediums, and an algorithm was used to identify the biocost of genes in the bacteria. Additional simulations were performed to identify non-recoverable essential genes of the bacteria. Furthermore, a deep learning method for identifying enzyme turnover rates (kcat) was employed to tune the metabolic model before simulation in order to improve model accuracy. Overall, the project was successful in identifying the top 5 non-recoverable essential gene targets in *C.ljungdahlii* based on biosynthetic cost, which can potentially be utilized when engineering biocontainment systems for other bacteria. The deep learning method of identifying kcats was also successful, showing an over 1000% increase in model accuracy.

Project ID: 215

Senior Division

Computational Biology and Bioinformatics

Nithika Vivek
Del Norte High School
Gr. 10

Deep Learning Multi-Modal Melanoma Detection

AWARDS:

Grand Award - Senior Division Life Sciences – ISEF FINALIST
CSEF Qualified

The visual similarity of melanoma and seborrheic keratosis has made it difficult for elderly patients with disabilities to know when to seek medical attention, contributing to the metastasis of melanoma. In this paper, we use deep learning multi-modal techniques to distinguish between melanoma and seborrheic keratosis. Background research and past studies that distinguished other skin disorders led to the hypothesis that the ResNet50 model combined with structured metadata would provide the best accuracy in differentiating between the two skin conditions as it mainly employs texture analysis such as border patterns to differentiate between acne and eczema with accuracy. Our strategy is three-fold: (1) utilize patient image data to train and test three prebuilt deep learning models using transfer learning (ResNet50, InceptionV3, and VGG16) and one author constructed model, (2) use patient metadata to train and test a deep learning model, and (3) combine the predictions of the image model with the best accuracy and the metadata model, using non linear squares (NLS) regression to specify weights to each model based on accuracy. The overall accuracy of the combined model was 91% on test data. Results from this experiment could be used to eliminate late diagnosis of melanoma that are harder to treat after metastasis if a patient waits too long to see a doctor. Seeking early attention is vital to prevent metastasis of melanoma and ensure safe recovery. Future experiments could combine image and metadata in a deep learning model rather than a simple regression model to identify more complex patterns. Additionally, utilizing text data (subjective data pertaining to how the patient felt over a certain period of time) can allow this model to reflect the real hospital setting to a greater extent.

Project ID: 216

Senior Division

Computational Biology and Bioinformatics

Dishti Wadhwani
Del Norte High School
Gr. 11

Early Detection of Sepsis in Hospital Patients

AWARDS:

CSEF Qualified

Sepsis is a life threatening condition that is one of the leading causes of hospital death around the country. Current diagnoses are typically made too late to intervene effectively, emphasizing the importance of a computerized detection system. It is hypothesized that common hospital data such as laboratory results and vital signs can be used to develop a scalable model that enables timely intervention and improves patient outcomes in the face of sepsis.

The PhysioNet Sepsis Prediction Dataset was cleaned and analyzed, including dropping irrelevant features, handling missing values, and preparing the data for input into a Recurrent Neural Network (RNN) architecture. The RNN model, comprising LSTM layers and dense fully connected layers, is trained using the Adam optimizer and binary cross entropy loss function to predict sepsis occurrence, demonstrating effectiveness in handling sequential data and achieving accurate classification results.

The model demonstrated strong performance metrics, including an accuracy of 0.88 and an AUROC of 0.861, validating the success of dropping irrelevant features identified through correlation analysis. Various tuning methods, such as adjusting epochs and selecting appropriate loss and activation functions, were employed to develop an effective model for early sepsis detection. Moving forward, collecting datasets without unimportant values and validating the model with external datasets are essential steps. The model's lightweight architecture and efficient interface make it easily deployable across internet-accessible devices, particularly within hospital systems, where real-time data integration can also further enhance its accuracy.

Project ID: 217

Senior Division

Computational Biology and Bioinformatics

Kailin Xuan
The Bishop's School
Gr. 10

Smart Analytics for Weight Management: A Data-Driven Approach to Addressing Obesity and Anorexia

Currently, 40% of Americans are affected by obesity, which can elevate the risks of developing diabetes, heart diseases, and cancer. Simultaneously, there has been a rise in eating disorders, particularly among teenagers. Many of them use the internet or Large Language Models (LLM) to access information or ask questions. Incorporating Retrieval Augmented Generation (RAG) with Large Language Models (LLMs) has emerged as a crucial methodology, enhancing model accuracy and credibility by integrating external knowledge. This project builds domain knowledge bases in the domain of weight management and introduces Analytical Retrieval Augmented Generation (ARAG) to further refine RAG, aiming to create a stronger framework by combining the strengths of LLMs, domain-specific knowledge, and relevant statistics and data analytics. We constructed a domain knowledge base KB_1 from professional websites, academic papers, and online sources in the domain of weight management, more specifically addressing obesity and anorexia. In addition, we created a data-driven knowledge base KB_2 with statistics and data analytics results in this field. We have subsequently implemented the ARAG strategy in a Python program, which uses ChatGPT as a subprogram and then applies it to the question-answering application. Through ARAG integration, the system mitigates hallucinations, delivers accurate and up-to-date information, and presents pertinent statistics that would otherwise be absent in a conventional LLM setup. In the experiment of question-answering related to obesity, over 80% of the 50 most frequently asked questions were answered with the latest statistics and data analytics in addition to responses by a traditional RAG approach

Project ID: 218

Senior Division

Computational Biology and Bioinformatics

Anvay Yadav
Del Norte High School
Gr. 10

A Federated Learning Framework for Machine Learning Tasks with Different Omics Data Types

AWARDS:

CSEF Qualified

Machine learning models rely heavily on both the quantity and quality of data utilized. In the case of healthcare, strict privacy regulations restrict data availability, presenting potential hurdles in creating unbiased machine learning models. Federated learning presents a solution by allowing the construction of a global model that uses information from multiple clients without actual data sharing. Instead, it transfers the weights of localized models to a central one. This study aims to investigate the feasibility of training global machine-learning models using various omics data from multiple sources without compromising accuracy.

To achieve this goal, a federated learning framework is developed using Pytorch and Flower libraries. To assess the efficacy of the framework, three distinct datasets are employed: the first comprises 30 geometrical features for breast cancer classification, the second utilizes genomic data for lung cancer classification, and the third offers multi-class classification of pancreatic cancer using biomarkers. For each dataset, a baseline model is initially established through a serial process. Subsequently, a machine learning model is constructed using federated learning, with data distributed across multiple clients. The current model is tested across up to five clients, and the federated learning method is evaluated based on two parameters: (a) convergence of the model and (b) accuracy compared to the baseline.

The results indicate that the federated learning model achieves convergence within a few communication rounds (less than 10) among clients for all three cases used, and its accuracy is similar to that of the baseline model for all cases.

Project ID: 219

Senior Division

Computational Biology and Bioinformatics

Ryan Yuan
Canyon Crest Academy
Gr. 11

Epilepsy Diagnosis Using miRNA Biomarkers and Machine Learning

Epilepsy affects over 50 million people worldwide. Epileptic patients experience recurrent seizures, which can lead to loss of consciousness or death. To address this problem, my research seeks to create a miRNA-based machine-learning model that can diagnose epilepsy with an accuracy greater than 80%.

First, miRNA with known epilepsy-diagnostic properties from a specific dataset as well as normal miRNA were gathered. Each miRNA's nucleotide sequence and target genes were collected as biomarkers used to classify the miRNA. Next, a program script was written to extract nucleotide combination frequencies from the miRNA sequences and filter genes by the target score threshold. Finally, these descriptors were used to build the machine-learning models for classification.

With the Hoeffding Tree, Logistic Regression, and J48 classifiers, I obtained classification accuracies of 86.36%, 79.55%, and 63.64%, respectively. The models were then evaluated on the epilepsy miRNAs of a second independent dataset and achieved classification accuracies of 87.50%, 75.00%, and 87.50%, respectively. Using a ranking system, it was determined that the Hoeffding Tree had the highest performance with a score of 0.872.

My initial hypothesis was met, where the Hoeffding Tree classifier achieved an accuracy greater than 80% on both datasets. The model's high performance suggests a strong correlation between a miRNA's association with epilepsy and its sequence and target gene patterns. Furthermore, the performance evaluation suggests Hoeffding Tree as the most promising classifier for future epilepsy diagnosis with miRNA.

Did Not Attend Judging

Project ID: 220

Senior Division

Computational Biology and Bioinformatics

Brianna Zhang
Canyon Crest Academy
Gr. 9

The Relationship Between Covid-19 Vaccines and Outcomes

This investigation explores the complex relationship between COVID-19 outcomes and vaccination efficacy. The hypothesis is that COVID-19 vaccines can effectively prevent disease and limit transmission under controlled conditions, with varying degrees of efficacy influenced by factors such as virus variants, individual age and health status, and duration of vaccine-induced protection.

To test our hypothesis, a data analysis was conducted using SQL Database and Microsoft Excel. The dataset "COVID-19 Outcomes by Vaccination Status" was imported and processed in SQL, covering variables such as age group, vaccination rates, and specific COVID-19 outcomes (cases, hospitalizations, deaths). The analysis focused on comparing key metrics among individuals who were unvaccinated, solely vaccinated, and those who received booster shots.

Our study shows that vaccines work differently based on age. They are very effective in younger people but in older age groups, how well they work can vary, and immunity might decrease over time. This highlights the need for a more detailed understanding of how vaccines perform in different age ranges.