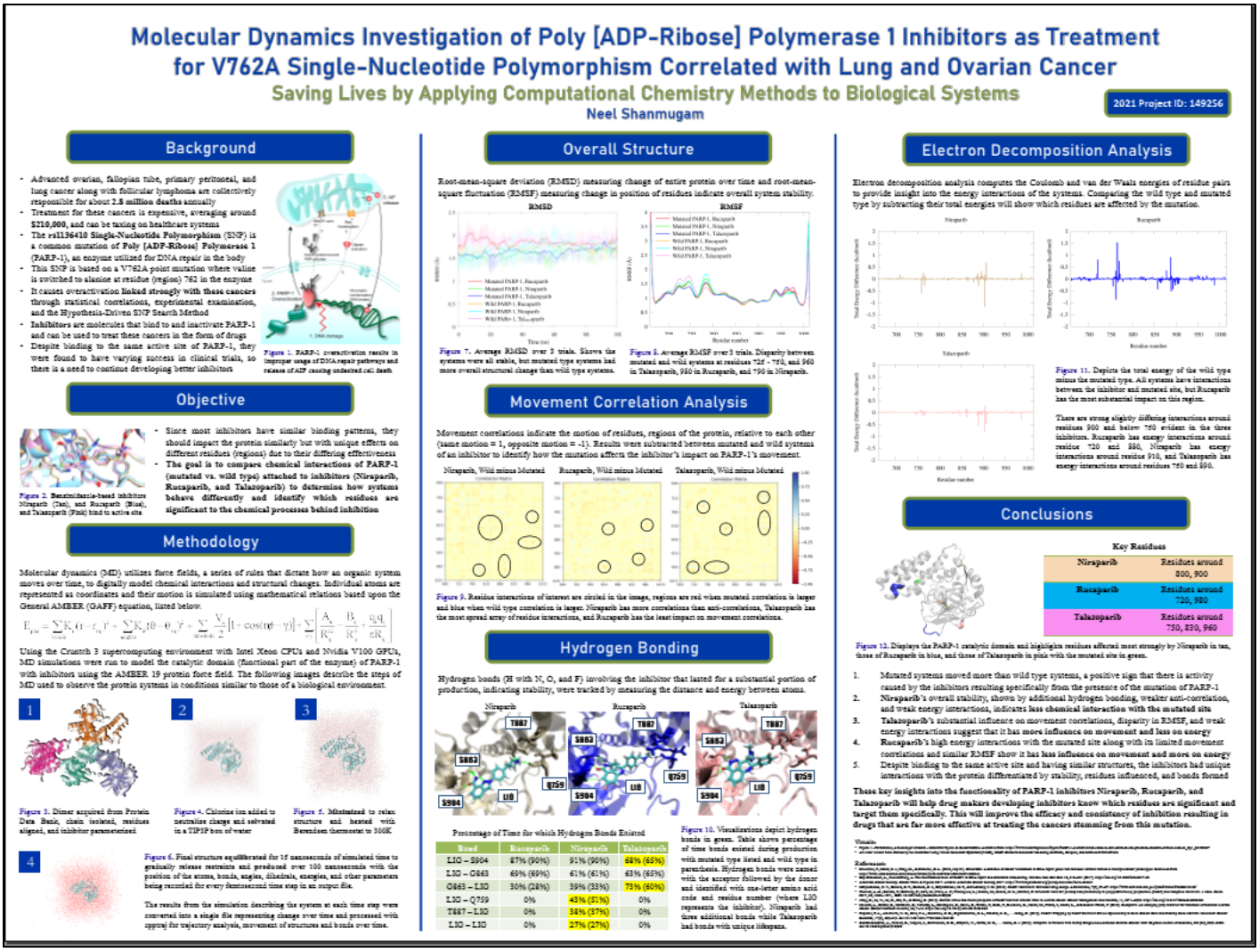
Fort Worth Regional Science and Engineering Fair – 2021

Judging Criteria:

Attributes or Variables?	More credit for using variable data
Sample Size ?	More credit for larger sample sizes
Experiment or Study?	More credit for experimentation vs static data analysis/ research data
Statistical Analysis ?	More credit for various types: mean, std dev, statistical tests, ANOVA, etc.
Graphical Analysis ?	More credit for various types: bar, line, box plots, histograms, etc.
Presentation Quality ?	Completeness to standard content from Hypothesis to Conclusions
Level of Creativity ?	Creation of measurements, apparatus, uncommon topics, analysis methods
Amount of Effort ?	More credit for duration of personal time over weeks/months

Senior Division – Grades 10-12

First Place Winner - $250 Cash Award



Senior Division – Grades 10-12

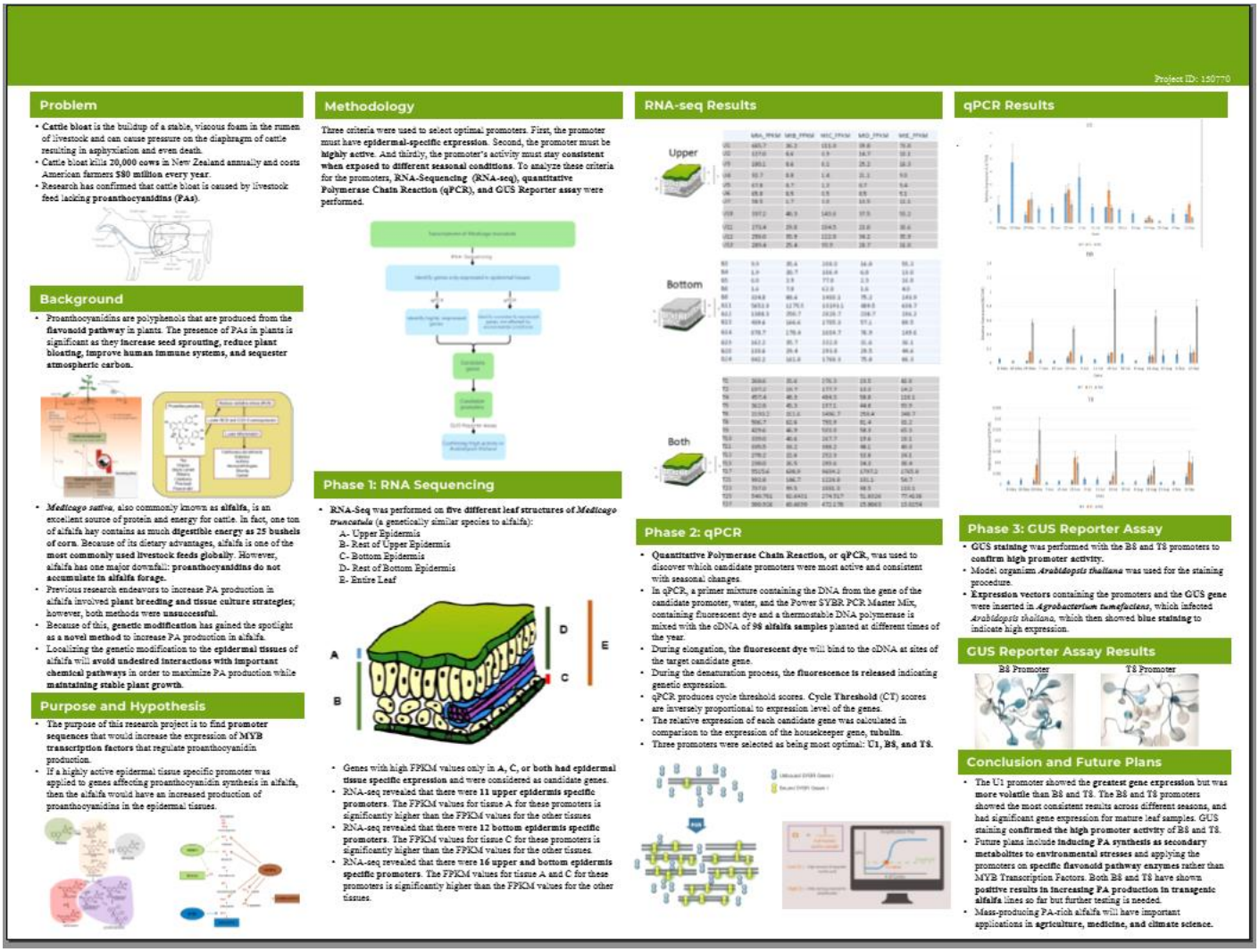
Second Place Winner - $100 Cash Award

2 Methods to Control the Pore Sizes and Shapes

Senior Division – Grades 10-12

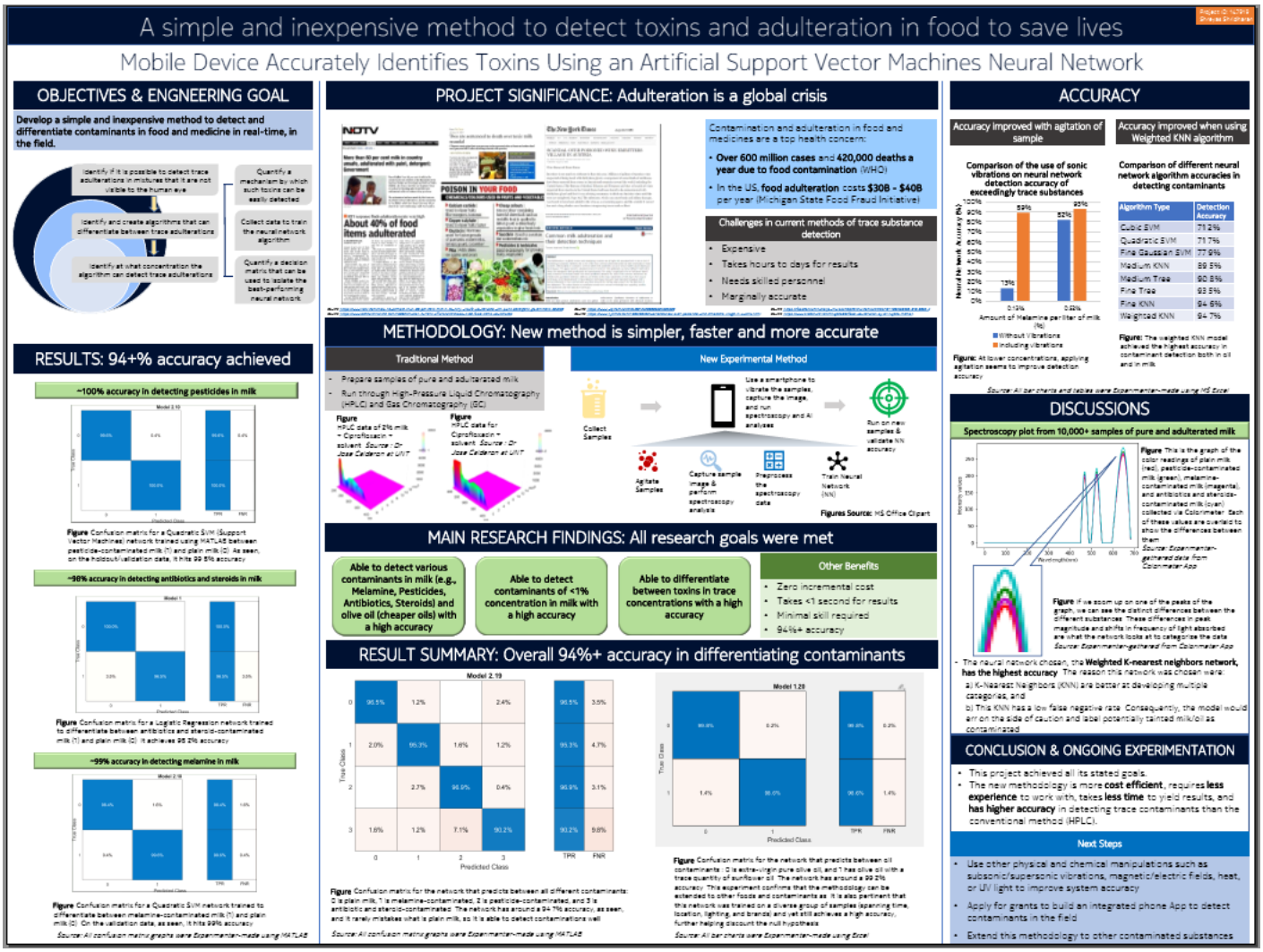
Third Place Winner - $50 Cash Award

3 Genetic Modification of Medicago



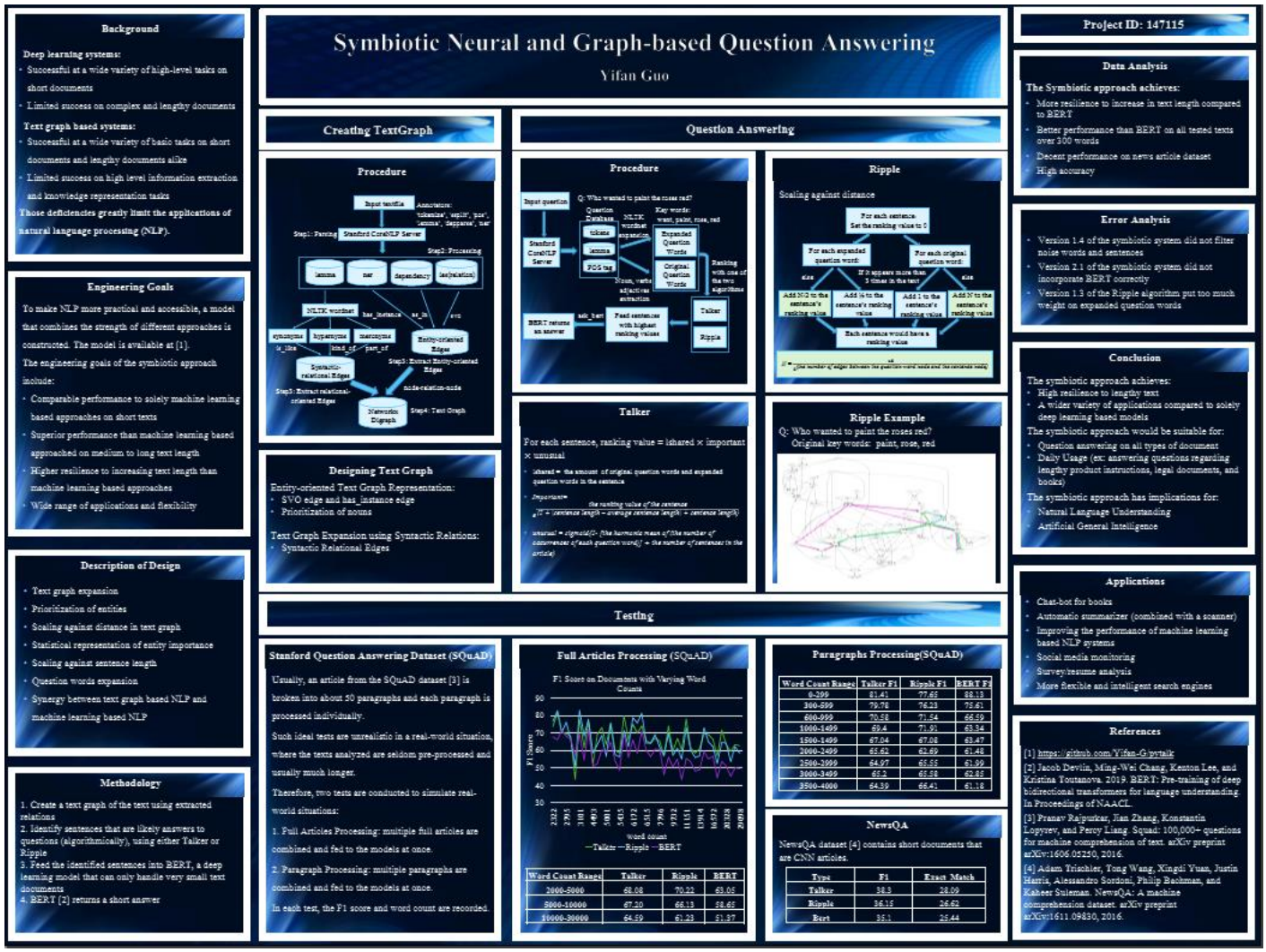
Senior Division – Grades 10-12

Special Award - $50 Cash Award – Field Application



Senior Division – Grades 10-12

Special Award - $50 Cash Award – Business Application



Senior Division – Grades 10-12

Special Award - $50 Cash Award – Learn from Mistakes, Not a Failure

Project ID:139225

CrossNET: A Cross-Species Deep Learning Model to Predict Human H3K4me3/H3K27Ac Peaks

Anay Gupta

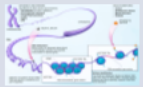
Introduction

Non-Coding DNA

- Over 90% of DNA within the human genome is non-coding, which accounts for gene regulation/expression, transcriptional regulation, translational regulation, etc.
- **Currently, there is no way in which geneticists can predict the effect of mutations present in non-coding DNA.**
- Granted this knowledge, it is necessary to then develop a medium in which non-coding DNA markers can first be predicted in an accurate manner.
- When determining which markers to identify the non-coding regions, **H3K4me3/H3K27Ac** markers were chosen, respectively.

Background Information

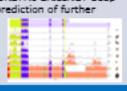
- **H3K4me3** is a marker located on the H3 histone on the 4th lysine residue signifying the tri-methylation, known to mark for enhancers. **H3K27Ac** similarly occurs on the same histone, but on the 27th lysine residue highlighting acetylation and is near to promoter regions.



- **These markers occur more frequently and are more "important" in conserved regions of the human genome.**

Cross-Species Integration

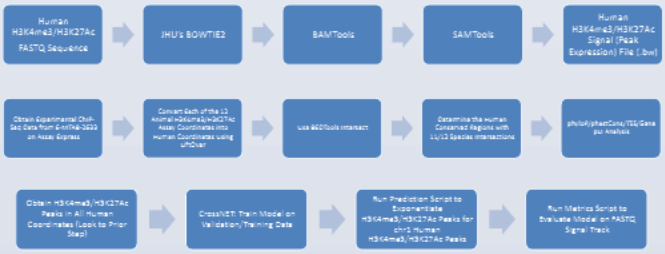
- In order to determine conserved regions in the human genome, LiftOver was used to convert the respective animal's H3K4me3/H3K27Ac ChIP-Seq assay coordinates of the 12 animals used into human coordinates. Then, BEDTools was used to determine the intersection between these coordinates for find the conserved regions.
- The cross-species assay data in human coordinates also enables an increase of data for the H3K4me3/H3K27Ac CrossNET deep-learning model for prediction of further peaks



References

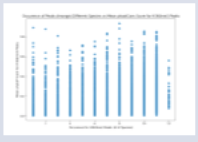
https://docs.google.com/document/d/1Gb6q3SrdfmfS6OVPFqlm4kLS1cfOLEZ0IFYnkWSI6Tc/edit?usp=sharing

Methodology

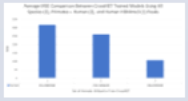


Results

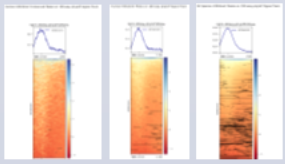
Analysis of the phyloP Score vs Nearest Transcriptional Start Site (TSS) of the Conserved H3K4me3 Peaks: On average, the TSSs were around 1kb in distance from the nearest conserved human H3K4me3 peaks. In relation with the phyloP score, there were rarely ever any "false-positives". However, there is no clear distinct linear correlation between the phyloP score and TSS distance as expected.



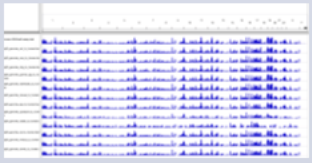
Analysis of the Functional Conservation Score vs the Mean phastCons Score of Conserved H3K4me3 Peaks: The graph displays a clear linear association between the functional conservation score (based on the number of species intersected among each other w/ BEDTools Intersect) and the mean phastCons Score between the chromosomal intervals.



Analysis of MSE Comparison Between Different Sets of Animals: The MSE of the All Species Trained Model was the lowest and had a Spearman Rank Coefficient Score of 0.86, which indicates extremely high correlation between the predictions of the CrossNET model and the FASTQ generated signal of the H3K4me3 Peaks.



Analysis of the phyloP Signal Heatmaps: In a side-by-side comparison of the conserved H3K4me3 Peak heatmap, human H3k4me3 assay heatmap, and a cross-species heatmap, it is seen that the conserved peaks generally have a higher phyloP score based on the line graph associating with the phyloP signal.



Analysis of LiftOver Effectiveness on IGV Genome Browser: The 12 species below the human assay were converted from their respective assay. The similarity of viewing the peak expression from a broad perspective shows that LiftOver proved to be effective and similar to the original human H3K4me3 ChIP-Seq Experimental Data.

Hypothesis/ Research Question

Would it be possible to effectively fabricate a deep learning model (convolutional neural network) with the input of a particular DNA sequence (1000 bp), genetic-disease causing mutation variant, with evolutionary conservation scores (phyloP/phastCons) and functional conservation scores to successfully train a neural network using cross-species genomic DNA sequences to predict the effect of a genetic disease variant?

Error Analysis

- CrossNET's training method in how it is integrated with Keras/Tensorflow does not work the way it is supposed to with a multitude of species (conglomerates all the peaks)
- Combining the cross-species dataset together added too much biological noise to the data resulting in less accurate results than expected
- Functional conservation did not correlate well enough with the evolutionary scores due to the largeness of the intersections to determine the conserved regions with BEDTools

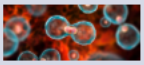
Conclusion

In this project, the CrossNET had to pass/answer the hypothesis/research question.

- CrossNET's Cross-Species Model proved to be the most effective in comparison with the other animal sets
- Functional Conservation Scores (intersection between species) and the Evolutionary Scores have little correlation
- Mutation Prediction soon is possible through this model

Real-World Application

- The idea of predicting the effects of genetic variants has a tremendous implication for a plethora of diseases. This enables people to control their risk factors if they have a higher probability of becoming diseased based on their specific genome. Some examples of these diseases include
  - Alzheimer's
  - Cancer
  - Heart Disease
  - Diabetes



Future Research

- Enabling the input to be a 1000bp DNA Sequence to predict the effect of the mutation in conjugation with the functional conservation score
- Reducing the Biological noise within the Cross-Species Training Data
- Determining why the functional conservation score and the phyloP/phastCons score did not have as a high of a correlation as intended
- Try training the model through a LSTM approach rather than an 8-layer CNN

Junior Division – Grades 7-9

First Place Winner - $125 Cash Award

Don't Get Burned

By: Kyle Alferink, Jerry Knight STEM Academy 7th Grade

ABSTRACT

This project, Don't Get Burned, will look at how different types of fabrics block UV radiation. The project will test that all of the fabrics will reduce UV transmission and darker fabrics will reduce UV levels the most. After testing all the materials and comparing the data results, it was determined that the initial hypothesis was partially correct. All fabrics reduce UV transmission but thicker fabrics, like beach towels and sweatshirts, do protect better. In addition, fabrics containing nylon provide more protection than just cotton. Cotton mixed with polyester in similar fabrics appears to provide less protection, but more research is needed.

PURPOSE

This project will look at how different types of fabrics block UV radiation from your skin.

HYPOTHESIS

All fabrics will reduce UV transmission and darker fabrics will reduce UV levels the most.

MATERIALS

UVAB Light Meter (General UV 513AB)
Digital Camera
Logbook
Pencil or pen
Access to WeatherBug website or app for UV Index
Frame to hold uniform sizes of fabric
Fabric scraps from a variety of clothing, swim, and poolside items

Fabric Used in this experiment:

White t-shirt: 100% Cotton
Black t-shirt: 100% Cotton
Gray t-shirt: 90% Cotton/10% Polyester
Beach Towel: 100% Cotton
Swimsuit: 80% Nylon/20% Spandex
Hawaiian Shirt: 100% Polyester
Gray Sweatshirt: 50% Cotton/50% Polyester
Navy UV Rashguard: 82% Nylon/18% Spandex

EXPERIMENT IN PROGRESS

Project Data

Project Setup

Date/Time: 11/3/20 at 10:35AM
UV Index: 4
Weather Conditions: Sunny

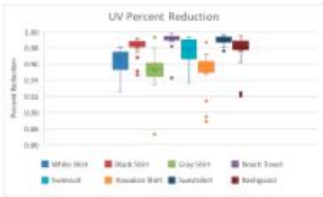
Date/Time: 11/19/20 at 11:04AM
UV Index: 3
Weather Conditions: Partly Sunny

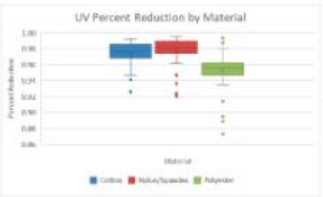
Date/Time: 12/[illegible]/20 at 12:36PM
UV Index: 2
Weather Conditions: Cloudy

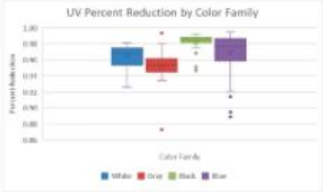
Date/Time: 1/22/21 at 10:35AM
UV Index: 1
Weather Conditions: Cloudy w/ rain

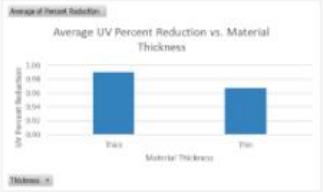
RESULTS

Graphs detail the percent reduction in UV for each type of fabric and provide comparisons for fabric color, fabric type, and thickness.









METHODS & PROCEDURES

1) Pick a location outdoors with consistent sun exposure over a range of time midday.
2) Setup fabric board for consistent measurements.
3) Read manual for UVAB Light meter for proper usage.
4) Record the date, time, UV Index as reported by WeatherBug, and note anything unique about the weather.
5) Take a photo of conditions/sun coverage.
6) Take a UV measurement ($\mu W/cm^2$)with no cover and record.
7) Take a UV measurement ($\mu W/cm^2$) behind each fabric type and record.
8) Repeat over the course of several weeks during varying the weather conditions and reported UV indexes.
9) Calculate the percent reduction in UV level for each fabric and compare.
10) Compare UV index to UV reduction by fabric type.
11) Draw conclusions based on UV reduction by fabric color, thickness, and material content.

CONCLUSION

After testing all the materials and comparing the data results, I determined that my initial hypothesis was partially correct. All fabrics reduce UV transmission but thicker fabrics, like beach towels and sweatshirts, do protect better. In addition, fabrics containing nylon provide more protection than just cotton. Cotton mixed with polyester in similar fabrics appears to provide less protection, but more research is needed.

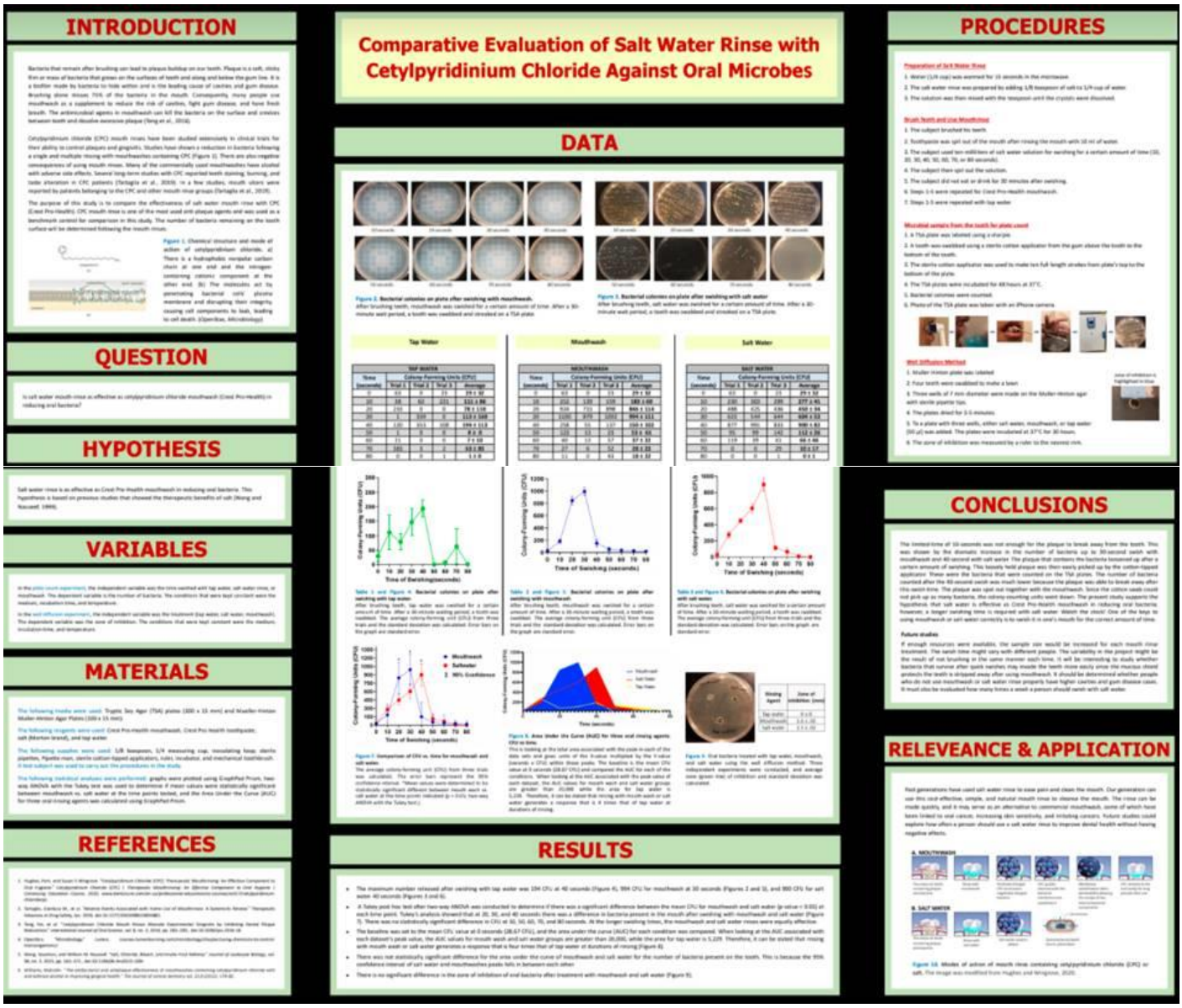
FUTURE RESEARCH

Does the color of your t-shirt impact UV reduction or is it more a function of the fabric content?

Does fabric content (i.e., 100% cotton vs. Cotton/Poly blend) in similar items of clothing, like t-shirts, impact UV reduction?

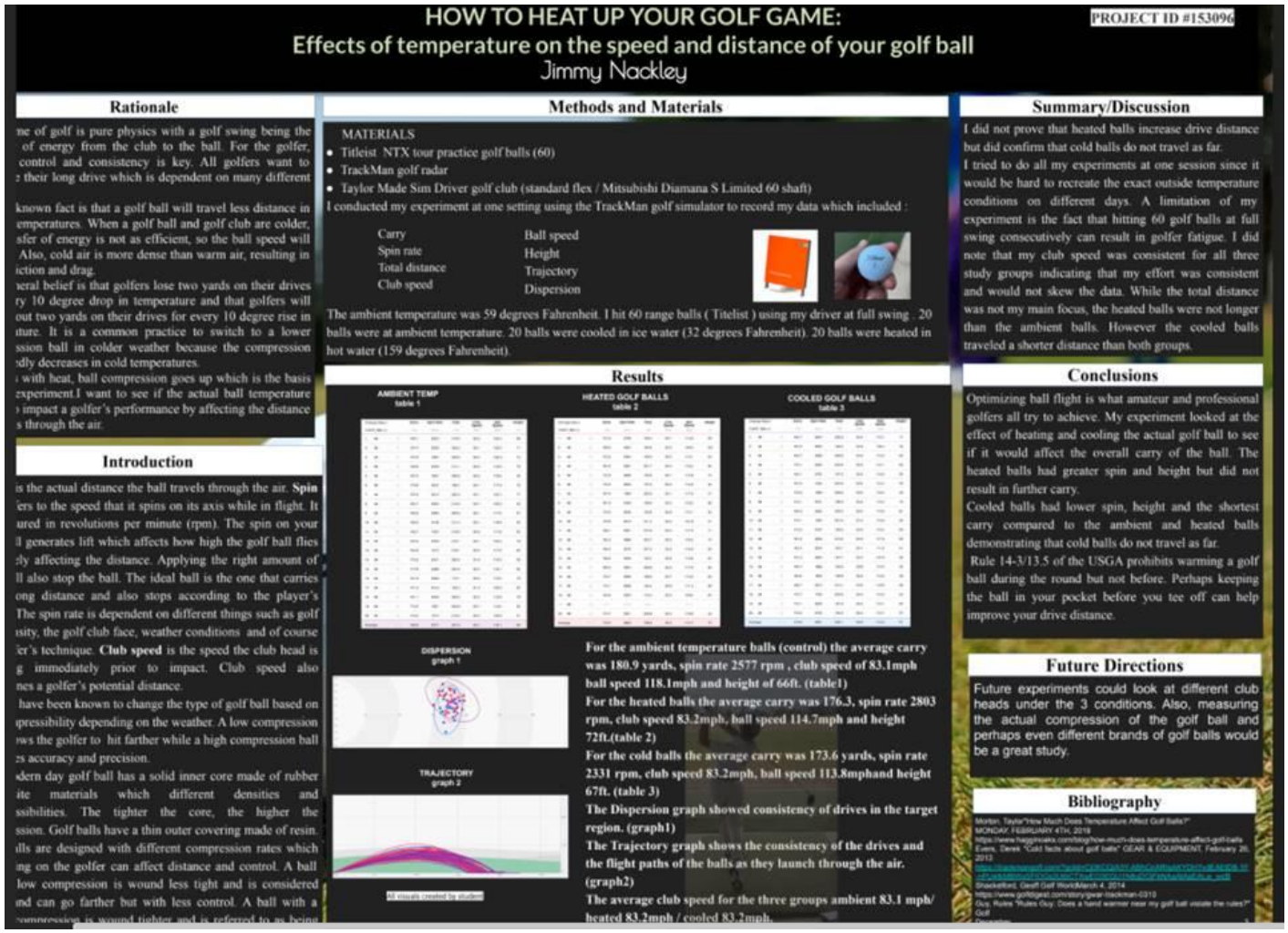
Junior Division – Grades 7-9

Second Place Winner - $150 Cash Award



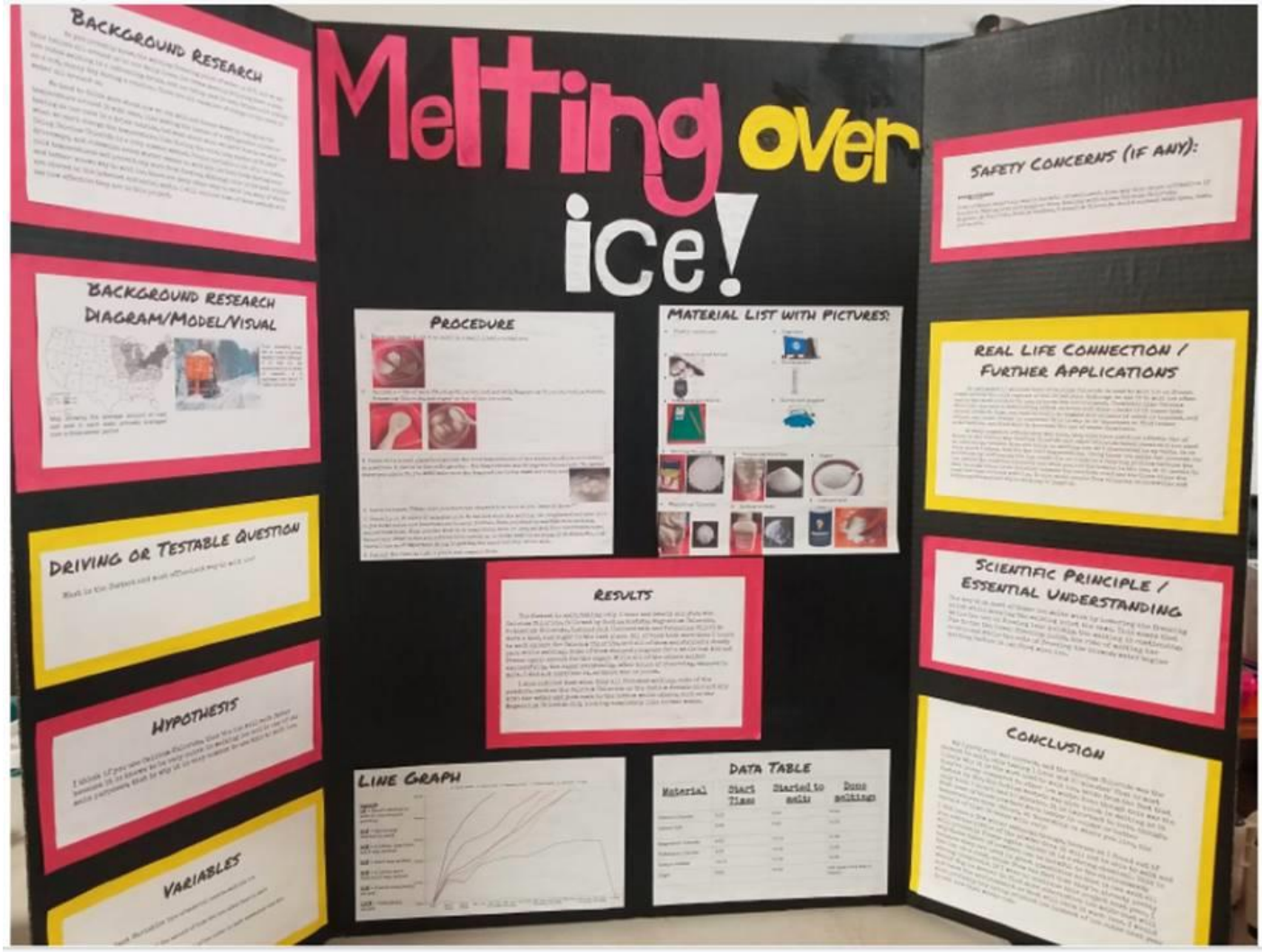
Junior Division – Grades 7-9

Third Place Winner - $25 Cash Award



Junior Division – Grades 7-9

Special Award - $25 Cash Award – Practical Application



Junior Division – Grades 7-9

Special Award - $25 Cash Award – Computer Simulation Programming

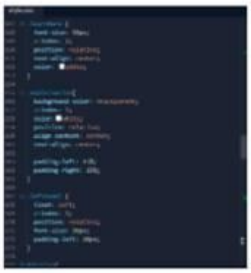
What Resources do we Have Left: An interactive 3D modeled globe.

Abstract

This project was made to better reinforce viewers about climate change. Through visual reinforcement, the viewer is often better off then reading some article. Research has show that people better understand something when they are shown images or models. This project was created in order to fix the way that people are informed about climate change. It is innovating the industry in order to convince more people. I did some research about climate change and how our world has been changing. I funneled all of this research into the website so it was informative. In conclusion, my website combines standard learning with an interactive website allowing the user to fully understand climate change.

https://www.our-only-earth.rocks/

Simulation of the earth where the user makes the decisions that decides the outcome of the earth.



500+ lines of code that was coded by me!

Conclusion

According to a survey of multiple family and friends, 90% of them said that they had a better understanding of the world's changing climate due to the website.

Rationale

Over the years, there has been many attempts to try and get the point across that climate change is a big deal. They get people to understand how its important, but don't explain ways to help. My website does and has a page that talks about ways anyone can help.

Materials

Laptop, Coding Software, Blender, free website hosting using GitHub developer pack.

Engineering Goal

Does a more interactive website better promote people to be more aware about climate and take action themselves.

Procedure

First I designed my website with a drawing and decided which features I was going to add. Then I started coding each part, bit by bit. That's how I did it!

Data and Observations

Data collected from family and friends shows that 90% of them benefited from the website and said that they understood the state of the earth better than before.

Future

Next year I plan on adding many new features, such as a more in depth simulation and more iterations of the earth.

Junior Division – Grades 7-9

Special Award - $50 Cash Award – Water Purification