

Development of Predictive Tools for
Anti-cancer peptide candidates using
Generative Machine Learning Models

p. 13

Mapping Agricultural Tillage Practices
Using Kernel Extreme Learning
Machine and Airborne Remote
Sensing Images

p. 21

The Gut Microbiome and the
Immune System: Investigating Links
between Diet and Antimicrobial
Peptide Sublancin 168 Production

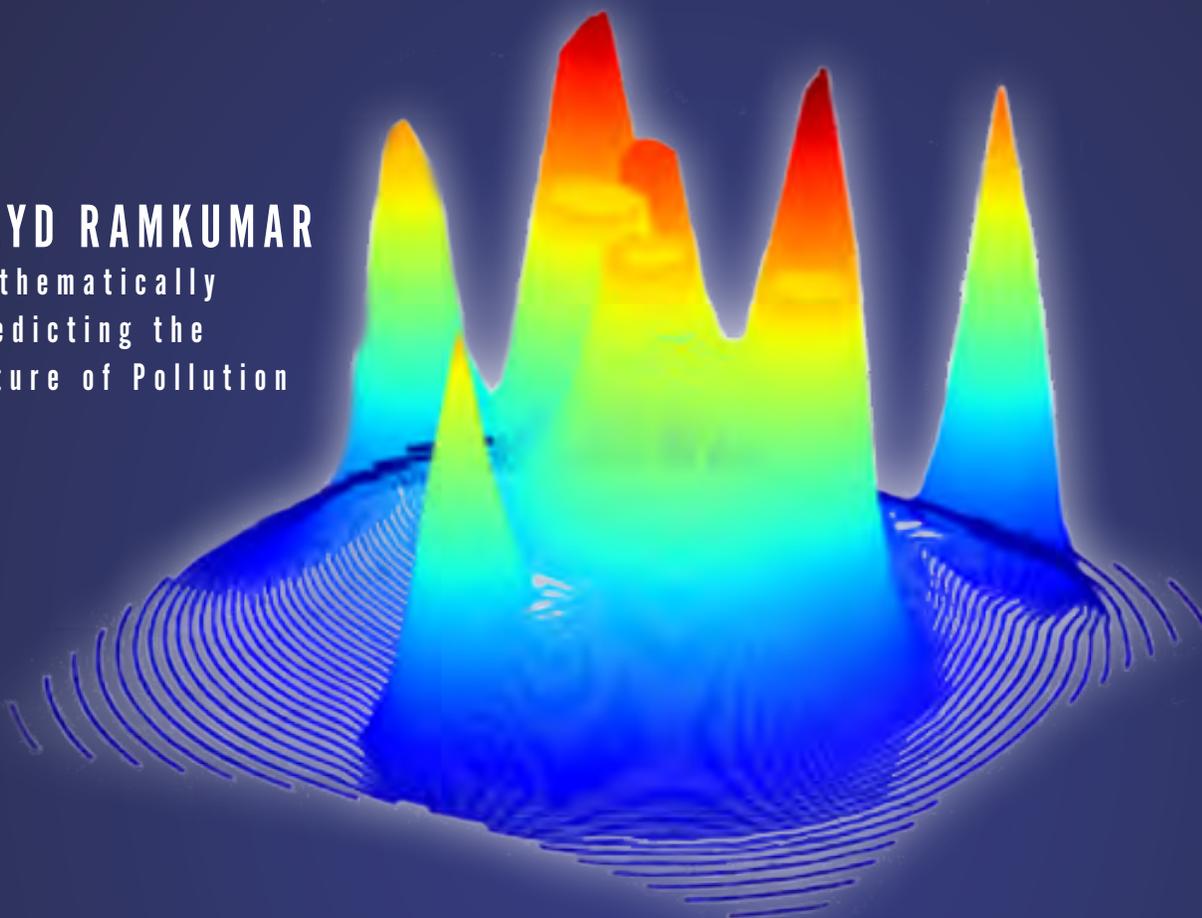
p. 30

M S M S

S C I E N C E J O U R N A L

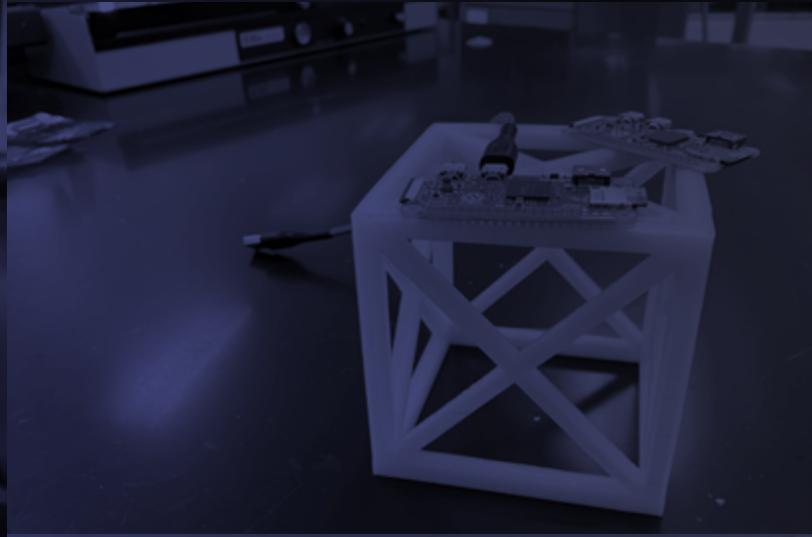
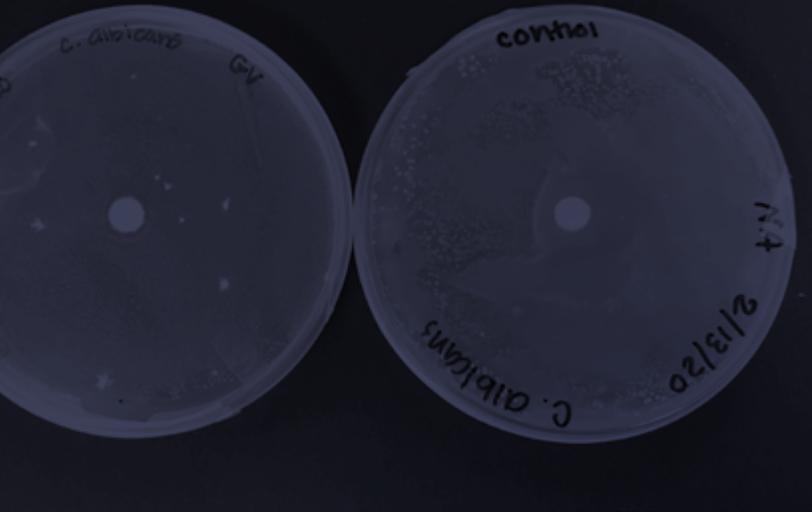
VAYD RAMKUMAR

Mathematically
Predicting the
Future of Pollution



WAVES OF SCIENCE

JUNE 2020 | VOL. 1 | ISSUE 1 | EST. 2020



Cover Diagrams: Vayd Ramkumar (2020)

Faculty Advisor: Dr. Tina Gibson

Cover Design: Skylar Nguyen

Photography: Gina Nguyen (2020)

Editors: Dennis Lee (2020) and Skylar Nguyen (2021)



CONTENTS

pg 04 **Forecasting Influenza Outbreaks Using Machine Learning Models**

by Aaron Wan

pg 06 **World Food Prize: Research that supports Increased Life Sustainability through Increasing Access to Education in the Democratic Republic of the Congo**

by Skylar Nguyen

pg 07 **Analysis and Modeling of Charge Density in Lab-Simulated Lightning Strikes Streamers**

by David C. Heson and Nathaniel Lee

pg 09 **Creating an Algorithm for Autonomous Vehicles to Calculate Routes**

by Guillermo Hoffmann Meyer and David Barber

pg 10 **Developmental Effects of Nicotine Exposure on *Drosophila Melanogaster***

by Catherine Min and Alicia Argrett

pg 11 **Analysis of the Effectiveness and Antimicrobial Capabilities Against Bacterial Pathogens of *Lactobacillus sp.* and Other Strains Found in Probiotics**

by Violet Jira and Gina Nguyen

pg 12 **World Food Prize: Education in Haiti**

by Victoria Callahan

pg 13 **Development of Predictive Tools for Anti-cancer peptide candidates using Generative Machine Learning Models**

by Michael Lu

pg 15 **World Food Prize: China: Removing Roadblocks to the Belt and Road Initiative**

by Aaron Wan

pg 16 **Modular CubeSat**

by Blake Cheater

pg 18 **Evaluating the Efficacy of Green Synthesized Silver Nanoparticles with Aminopenicillins, Carbapenems, Third/Fourth Generation Cephalosporins, and Intracellular Antibiotics Against Multi-drug Resistant Bacteria**

by Shanay Desai

pg 20 **World Food Prize: Combating rampant malnutrition in Pakistan**

by Muneebah Umar

pg 21 **Mapping Agricultural Tillage Practices Using Kernel Extreme Learning Machine and Airborne Remote Sensing Images**

by Dennis Lee

pg 23 **Solar Water Disinfection and The Advanced Oxidation Process: Effects of Organic Matter, pH, and Ions on the Disinfection of *E. Coli*, and the Environmental Impact**

by Hong “Reggie” Zheng and Colin Gordy

pg 24 **Assimilation of Global Pollution Data: A Novel Spatiotemporal Model for Widespread Classification of Atmospheric Particulate Clusters (Year 2)**

by Vayd Ramkumar and Jackson Flowers

pg 27 **Comparative Study between the Antimicrobial Resistance of Lichen from Mississippi and Alabama against *Staphylococcus epidermidis* and *Escherichia coli***

by Skylar Nguyen

pg 29 **The Gut Microbiome and the Immune System: Investigating Links between Diet and Antimicrobial Peptide Sublancin 168 Production**

by Rachel Zheng and Bertha Mireles

pg 30 **Development of a Vaccine to Enhance Mucosal Immunity**

by Fiona Dawe and Clare Seo



Aaron Wan

Forecasting Influenza Outbreaks Using Machine Learning Models

Abstract

Accurate flu forecasts are needed to limit the spread of outbreaks. Current flu forecasts have not incorporated several environmental, demographic, and transport factors. This study looks at temperature, precipitation, wind speed, air quality, population density, underage population, elderly population, median household income, vaccination rate, and enplanements to determine which factors influence the spread of influenza. It was hypothesized that using these factors alongside machine learning models can pinpoint the severity of the monthly “Influenza-Like-Illness” (ILI) rates in all fifty states. The data was fit to a LASSO regression model to determine which factors most impacted the spread of the flu. The correlation between each of the factors and ILI rate was calculated. A p-test was performed to determine whether the correlation was statistically significant. The data was incorporated into a K-Nearest Neighbors Classifier, a Random Forest Classifier, a Gaussian Naïve Bayes Classifier, and a Decision Tree Classifier. The K-NN Classifier, the most accurate model, was 99.25% accurate on training data and 89.39% accurate on test data. This model was able to predict the severity of the ILI rates in each of the fifty states in February 2019 with 90% accuracy. This study sheds light on the effect of the studied factors to on flu outbreaks. A statistically significant, negative correlation was found between air quality and ILI rate, meaning that poor air quality should be addressed when tackling flu outbreaks. The ten data sources in this study should be incorporated into other models to improve the accuracy of flu forecasts.

Introduction

In the United States, flu season occurs during the fall and winter months and typically peaks during December, January, or February. Influenza infects up to 35 million people in the United States every year and causes between 12,000 and 56,000 deaths per year. [1] Research has shown that several environmental, demographic, and transport factors can influence the spread of the flu. [2] In order to limit the spread of outbreaks and reduce morbidity in those already infected, accurate and timely flu forecasts are needed to aid in deploying necessary public health measures to combat the spread of the flu. While many statistic and machine learning models have been developed to predict the spread of the flu, current flu forecasting models tend to use public health surveillance data, electronic health records, online search trends, and social media data to predict the spread of the flu. [3] These models have yet to incorporate several factors such as climate factors, demographic factors, and transportation factors which have been shown to have the potential to influence the spread of the flu. This study looks at ten of these factors (Temperature, Precipitation, Wind Speed, Air Quality, Population Density, Underage Population, Elderly Population, Median Household Income, Vaccination Rate, and Enplanements) to determine which of these factors influence the spread of the flu and then incorporates those factors into four machine learning models to predict the spread of the flu. As such, the objectives of this study as follows: First, to develop a better understanding of relationships between these ten factors and the spread of the flu, and second, to incorporate these ten factors into four machine learning models to predict the spread of the flu. It was hypothesized that using monthly temperature, precipitation, wind speed, air quality, population density, underage population, elderly population, median household income, vaccination rate, and enplanements data, a machine learning model can be developed to accurately predict the severity of the monthly “Influenza-Like Illness” rates in each of the fifty states.

Methods

Data: Monthly data was taken from each state from the 2014-2015 flu season to the 2017-2018 flu season for a total of 1,550 data points. The weather data (Temperature, Precipitation, Wind Speed) was gathered from Weather Spark. The air quality data was taken from the U.S. Environmental Protection Agency Database. The demographic data (Population Density,

Underage Population, Elderly Population, and Median Household Income) was obtained from 2010 U.S. Census data. The enplanements data was obtained from the Federal Aviation Administration Database. The flu data as well as the vaccine rates data was obtained from the Center for Disease Control Database. The ILI rates were sorted into severity levels based on national baselines

Influenza Like Illness Rates: The ILI rates were sorted into 4 severity levels based on national baselines

ILI Rate	Severity Level
≤2.4	Minimal
>2.4 and ≤5.4	Low
>5.4 and ≤9.4	Moderate
>9.4%	High

Machine learning models: The data was first inputted into a Lasso Regression algorithm to determine which of the factors have a significant impact on ILI rates. The data was also graphed to determine the correlation coefficient (R value) between each of the factors and ILI rate. A p-test ($p < 0.05$) was performed on the R values to determine whether the correlations were statistically significant. The data was then incorporated into four different machine learning classification models, namely a K-Nearest Neighbors Classifier, a Random Forest Classifier, a Gaussian Naive Bayes Classifier, and a Decision Tree Classifier.

Parameter tuning and cross-validation: The training data was randomly selected and consisted of 90 percent of the dataset, while the test data consisted of the remaining 10 percent of the dataset. The data was also scaled using MinMax Scaling. For Lasso regression, cross validation was used to find the optimal β value. The accuracy of the models on both training and test data were recorded to determine the most accurate machine learning model

Application in forecasting influenza outbreaks: After the most accurate machine learning classification model was determined, the model was used with the historical data to predict the ILI rates of each of the fifty states during February 2019, the peak flu month of the 2018-2019 flu season. To further apply the model, the model was used to forecast the severity of the ILI rates in each of the fifty states for February 2020.

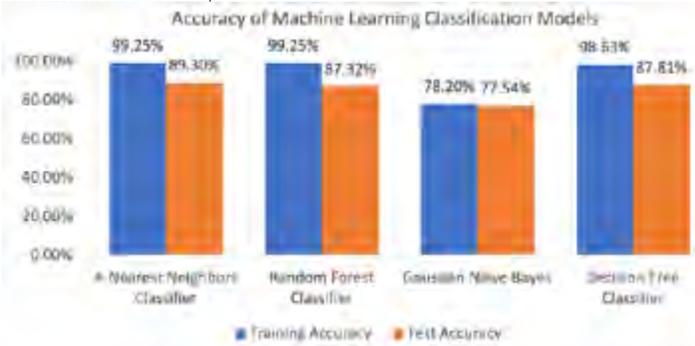
Results

Relationships Between Influenza-Like Illness Rate and the Studied

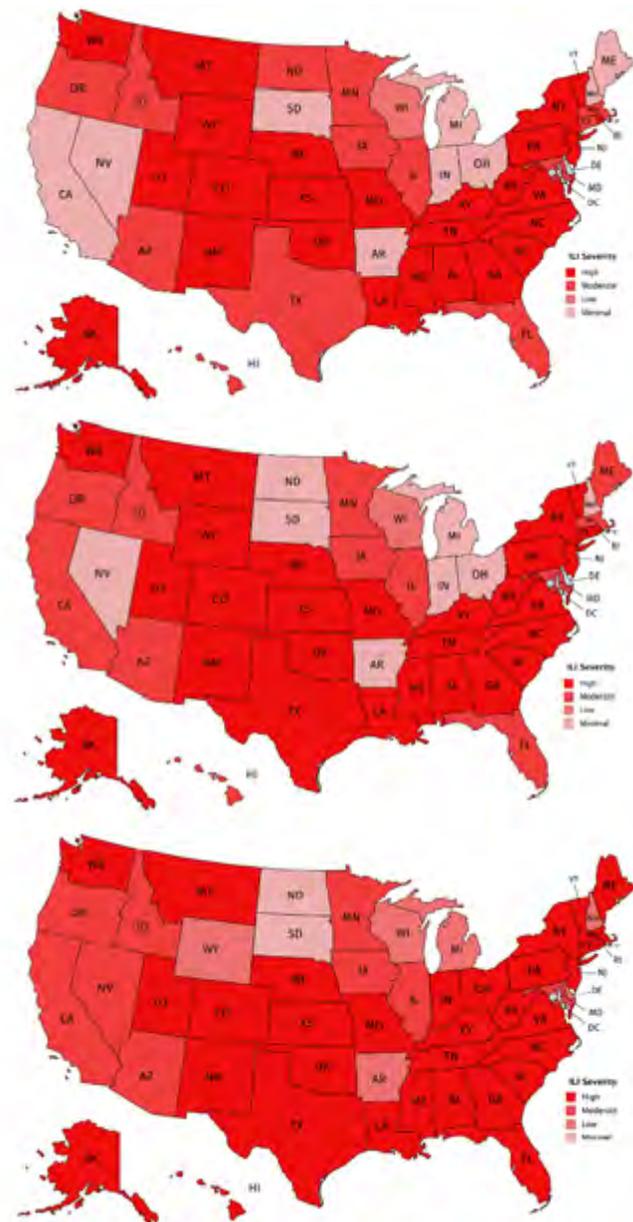
Factors: LASSO Regression revealed which of the factors in this study had the most significant influence on the spread of the flu. The R values quantified the strength of the correlation between ILI rate and the various factors. The p-test confirmed whether the correlations were statistically significant. A negative correlation was found between temperature and ILI rate, wind speed and ILI rate, elderly population and ILI rate, median household income and ILI rate, and vaccination rate and ILI rate. A positive correlation was found between precipitation and ILI rate, underage population and ILI rate, enplanements and ILI rate, and sulfur dioxide concentration and ILI rate. Performing a p-test on the correlation coefficient for population concentration revealed that the R value was not statistically significantly different from zero, meaning that there was no statistically significant correlation between population concentration and ILI rate. The most influential factor found was wind speed, which had a strong, negative correlation with ILI rate. The least influential factor found was enplanements, which had a weak, positive correlation.

Features	Weight	R-Value	p-Value
Temperature	-1.609	-0.2126	< .00001
Precipitation	3.532	0.5739	< .00001
Wind Speed	-6.117	-0.7607	< .00001
Population Concentration	-0.8151	-0.0173	.503624
Percent of Population Over 65	-2.269	-0.1170	< .00001
Percent of Population Under 18	0.7036	0.1277	< .00001
Median Household Income	-1.343	-0.1623	< .00001
Enplanements	0	0.0787	.00193
Vaccination Rate	-0.9487	-0.4133	< .00001
Air Quality (Sulfur Dioxide Concentration)	3.917	0.3809	< .00001

Modeling: The K-Nearest Neighbors model was the most accurate model, with 99.25% accuracy on training data and 89.30% accuracy on test data. For the K-Nearest Neighbors Classifier, the optimal “n_neighbors” value found was 1. For the remaining three machine learning models, the default parameters were found to be the optimal parameters for maximizing classification accuracy.



Applying the Model: The K-Nearest Neighbors Classifier was able to predict the severity of the ILI rates in each of the fifty states during February 2019 with 90% accuracy. The classifier was only inaccurate in predicting the ILI severity in five states. The model inaccurately predicted the severity for California, North Dakota, Maine, Connecticut, and Texas. For each of these states, the model predicted a severity level one level below the actual severity level. Using projected climate data and air quality data for February 2020 and keeping other factors consistent, a map was also created of the model’s projections for the severity of the ILI rates in each of the fifty states for February 2020.



Forecasting the Flu for February 2020: Prediction of the ILI Severity in each of the 50 states for February 2020 using the K-Nearest Neighbors Classifier

Discussion

The results of this study provide a deeper analysis of the relationships between the ten factors studied (Temperature, Precipitation, Wind Speed, Air Quality, Population Density, Underage Population, Elderly Population, Median Household Income, Vaccination Rate, and Enplanements) and ILI rate.

Interestingly, a negative correlation was found between percent elderly population and ILI rate. It would be expected that states with higher percent elderly populations would have higher ILI rates, but the opposite relationship was found. A possible explanation for this relationship is that a higher percent elderly population inherently means a lower percent underage population, and the data suggests that a higher percent underage population correlates with higher ILI rates.

The results also show that with just ten data sources and 1,550 data points, it is possible to accurately predict influenza outbreaks. The K-Nearest Neighbors Classifier model was not only 89.30 percent accurate on test data, but it was also able to predict the severity of the ILI rates in each of the fifty states for February 2019 with 90 percent accuracy. This verifies the hypothesis that

using environmental, demographic, and transport data alone can accurately predict the spread of the flu

Implications and Recommendations

The findings in this study implicate two significant recommendations for health and public policy officials to better deal with influenza outbreaks:

First, through LASSO regression and finding correlation coefficients, this study was able to give us a new, better understanding of which factors can impact the spread of the flu. Thus, health and public policy officials should strongly consider using these findings to better address vulnerabilities that contribute to the spread of the flu. More specifically, this study resulted in four notable findings:

- A statistically significant, negative correlation was found between the median household income and ILI rates. This logically suggests that poorer regions tend to have lower quality healthcare infrastructure, so health and public policy officials may want to focus on poorer areas to address the spread of the flu.
- A statistically significant, positive correlation was found between sulfur dioxide concentration and ILI rate. A potential explanation for this relationship is that higher sulfur dioxide concentrations result in poorer air quality, and poorer air quality often causes in respiratory issues, which could increase a person's vulnerability to becoming infected by the flu. Health and public policy officials should begin focusing on addressing poor air quality as a part of efforts to reduce the spread of the flu.
- A statistically significant negative correlation was found between vaccination rate and ILI rate. The data show that vaccinations are an effective way to mitigate flu outbreaks, so officials should urge the public to get vaccinated.
- A statistically significant, positive correlation was found between percent underage population and ILI rate, suggesting that officials should focus on preventing the spread of the flu among children. The data shows that children are at a high risk of being infected.

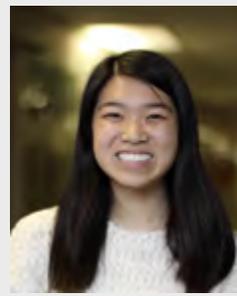
Second, the K-Nearest Neighbors model used in this study was accurate 89.30% on test data, and the model was 90% accurate in predicting the severity of the ILI rates in each of the fifty states for February 2019, the peak of the 2018-2019 flu season. This is significant, as current flu forecasting models use hundreds, or even thousands of data sources and millions of data points, while the model used in this study only used five data sources and 1,550 data points and was still 90% accurate. Thus, flu forecasters should strongly consider incorporating the five data sources identified in this study into their models to improve the accuracy of their flu forecasts.

Future Research

For future research, the data sources used in this study could be incorporated into other machine learning models. For instance, the dataset could be used to train a neural network that could be potentially more accurate than the machine learning classifiers used in this study. In addition, the dataset used in this study only consisted of 1,550 data points; current flu forecasting models use datasets that are hundreds of times larger than this one, so more data points can be added to the dataset used in this study to increase the accuracy of the machine learning models.

References

1. Chakraborty P, Lewis B, Eubank S, Brownstein JS, Marathe M, Ramakrishnan N (2018) What to know before forecasting the flu. *PLoS Comput Biol* 14(10): e1005964. <https://doi.org/10.1371/journal.pcbi.1005964>
2. May P. S. Yeung, Frank L.Y. Lam, Richard Coker, Factors associated with the uptake of seasonal influenza vaccination in adults: a systematic review, *Journal of Public Health*, Volume 38, Issue 4, December 2016, Pages 746–753, <https://doi.org/10.1093/pubmed/fdv194>
3. "The future of influenza forecasts" Cécile Viboud, Alessandro Vespignani. *Proceedings of the National Academy of Sciences* Feb 2019, 116 (8) 2802 2804; <https://www.pnas.org/content/116/8/2802>



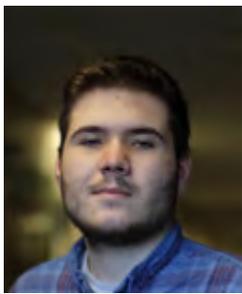
Skylar Nguyen

World Food Prize: Research that supports Increased Life Sustainability through Increasing Access to Education in the Democratic Republic of the Congo

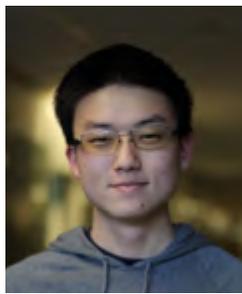
Hunger and conflict have plagued the Democratic Republic of the Congo (DRC) for the past twenty-five years. The DRC was ranked 5th of 178 countries on the 2019 Fragile States Index (fragilestateindex.org, 2019). To put this into perspective, the DRC has a population of almost 80 million people with 80% of its citizens under the international poverty line, living off of \$1.25 per day (Keysa, 2018). Studies have shown that the lack of human development indicators (health, education, and access to resources) is widespread within the DRC, validating the fact that poverty is a mass phenomenon. This common situation of extreme poverty within the DRC does not discriminate against age or sex. The repeated cycle of destitution from generation to generation continues to keep children out of school. Currently, 7 million children from the ages of 5 to 17 are not receiving primary education (unicef.org, 2020). Being that the DRC is the second-largest country in Africa and has access to an abundance of natural resources, it has potential to improve its country, more specifically its education system and policies. Policies would aim to outline the importance of education and decrease the number of students out of school in hopes of increasing reading outcomes in the DRC. To have maximum effectiveness, the new policies would have to be coupled with the implementation of free schooling with trained administration, a peace centered curriculum, and a safe and conducive learning environment. The success of these implications is validated by similar education programs already placed in the DRC by UNICEF, USAID, and DfID. These specific implementations would be used to combat the DRC's main barriers against education: poverty, conflict and displacement, sexual violence, and expensive schooling (educateachild.org, 2013). Through the joint implication of inexpensive school, a peace centered curriculum, and an ideal learning environment in conjunction with policies advocating for students to stay in school, education gives a holistic approach to tackling the most pressing problems within the DRC. Because of the plethora of opportunities education offers, its application satisfies sustainable development goal number four on the United Nations (2019). Education can lead the way towards economic stability, establishing the skills needed to work a higher paying job, the knowledge needed to escape difficult situations, and the values needed to avoid conflict. These adversities that education can overcome shows that education is not only life-sustaining, but it is also life-saving.

Works Cited:

- Educate A Child. (2013). Democratic Republic of Congo. Retrieved from <https://educateachild.org/our-partners-projects/country/democratic-republic-congo>
- Keysa, K. (2018). Facts about Poverty in the Democratic Republic of the Congo. Retrieved from <https://borgenproject.org/facts-about-poverty-in-the-democratic-republic-of-the-congo/>
- The Fund for Peace. (2019). Fragile States Index. Retrieved from <https://fragilestatesindex.org/wp-content/uploads/2019/03/9511904fragilestatesindex.pdf>
- Unicef DRC Oatway. (2020). Education- Every child has a right to go to school and learn. Retrieved from <https://www.unicef.org/drcngo/en/what-we-do/education>
- United Nations. (2019). Sustainable Development Goals Knowledge Platform. Retrieved from <https://sustainabledevelopment.un.org/sdg4>



David C. Heson



Nathaniel Lee

Analysis and Modeling of Charge Density in Lab-Simulated Lightning Strikes Streamers

Abstract

Thunder strikes and their mechanisms are still problematic to researchers, physicists, and engineers alike, due to the unpredictability and elusive nature. Using the facilities of Mississippi State University’s High Voltage Lab, lightning strikes were simulated at 80 kilovolts, and their electrical fields were recorded from E-Field sensors placed radially 20 feet away from the center of the aluminum and iron mesh cloud. Using the computer program Octave with a few of its packages and math derived from Maxwell’s equations, the average charge density over the luminous channel could be established and then be fitted to a model dependent on the length, time, and source voltage from the strike. After data-conditioning was performed to eliminate any potential “noise” and create a window of interest, trimming the data to contain only the region of interest (when the lightning strike actually happens), the necessary math was applied to derive the charge density. From the data obtained, it would seem like the charge density resembles noise, having functionally zero correlation between its data points, hence hinting towards the fact that for a single-channel lightning strike, the average charge density does not follow any clear pattern.

Introduction

Experimentation in the field has recently picked up in pace due to advancements in technology and procedure which allow for accurate data collection/interpretation and the set-up of the experiment itself. While it is desired for data to be collected for lightning strikes naturally produced, laboratory settings can be used to replicate the desired elements found from the wild and produce a controlled lightning strike which ought to accurately represent the characteristics of a natural one. As Rubin Jiang et al (2014) [1] have demonstrated, laboratory-obtained data can be fitted to models established off natural lightning strike behaviors, hence confirming, or alternating, currently existing hypotheses. These studies are extremely important as they can help progress the realization of more efficient lightning protection systems for structures such as buildings, cruise ships, and airplanes. Additionally, they can help investigate for the cause of failures in different apparatus and destruction caused by lightning strikes. In 2018 alone, lightning strikes have done 909 million dollars in damages in the U.S., as per the Insurance Information Institute.

When attempting to create a simulated lightning bolt, a proper set up has to be set such that a desired shot will be produced while also preventing flashover. Appropriate cloud distances will need to be used (which will depend on the source voltage and atmospheric conditions in the lab), and a low potential area will need to be set up to create a predictable area to where the lightning will strike. There are multiple factors that could potentially affect how the charge density will change with time, such as the initial voltage (which is expected to increase from the beginning due to increasing the current), atmospheric conditions such as humidity, temperature, and air pressure, the electrical potential differences between the source “cloud” and the ground, and the distance over which the luminous channel forms.

This project also seeks to follow the sustainability goals set by the United Nations, namely the goals of Affordable and Clean Energy and the goals of Industry, Innovation, and Infrastructure. While this project will not only facilitate lightning protection, it can also pave the way towards a rather futuristic endeavor, and that is the generation of electrical power through lightning strikes, which has been attempted in the past but unsuccessful due to the technology and academia of the time, achieving very low levels of efficiency.

Methods

The data for this project has been collected with the help of the Mississippi State High-Voltage Lab (HVL). With the help of Dr. David Wallace and

several of the graduate students, the “cloud” and other apparatus was constructed. After the data collection, with the usage of Octave, the data will be conditioned and then analyzed to obtain the average charge density over the bolt.

The 3 MV Nissan Impulse Test Set of the lab was used to generate the charge, while a Haefely 2.4 MV Voltage Divider was used to measure the voltage. The current was transmitted from the tower to the “cloud” through 4” wide aluminum strips, while the “cloud” itself was formed from 4” aluminum pipes and an iron mesh. To create the desired low potential area, an elevated copper plate with aluminum pipes was used. To record the electrical field generated by the lightning, four electrical field sensors were placed equally distanced from each other and twenty feet away radially from the center of the “cloud”. These sensors were connected with the HVL’s hardware for recording, using one hundred decibel attenuators. Fifteen lightning strikes were created in total, five for each of the three different voltage levels used, which were seventy kilovolts, eighty kilovolts, and ninety kilovolts. Figure 1 is a picture of the set-up,

Figure 1.



After the data was collected, an algorithm will be written in Octave which will perform two different functions. First, it will serve to condition the data to where it is easier to analyze and interpret. It will trim the data to where it will only contain the region of interest, which is when the dielectric breakdown itself happens. Afterwards, a Soviatsky-Golay filter will be applied to smoothen out any noise. For this, the signal package for Octave will be imported. The second part of the algorithm will consist of the data analysis itself.

The current density will be derived from the electrical field using Maxwell’s equations, as shown below.

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad \nabla \times E = -\frac{\partial B}{\partial t}$$
$$\nabla \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}$$

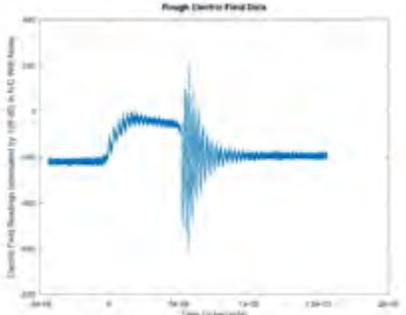
Where B is the magnetic field, J is the charge density, E is the electric field, t is time, μ_0 is the permeability of air, and ϵ_0 is the permittivity of air. It is assumed the lab uses the standard permittivity and permeability constants, for the atmospheric conditions differ minimally from the ones under which those constants were derived. Once the current density is known, it can be simply integrated over it to derive charge density.

The algorithm will be run over all different runs, and if a relationship is identified, the charge density versus time obtained for each different data set will be attempted to be fitted to different models of non-linear equations, with time as the variable.

Results

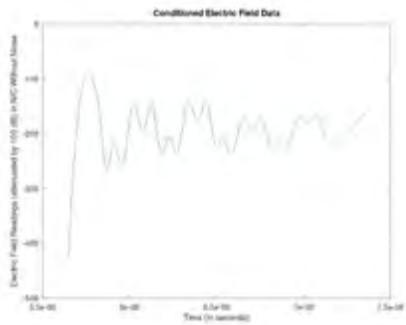
The data collection went smoothly, with all fifteen shots being recorded at the High Voltage Lab. Figure 2 is a graph of one of the eighty kilovolt shots as recorded.

Figure 2



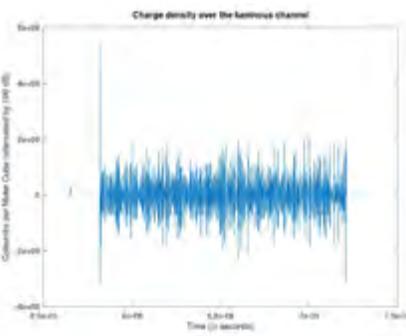
After this, using the constructed trimming algorithm which tried to isolate only the middle portion (where the dielectric breakdown itself happens). The algorithm looked at 30 points at the time and checked if there was any significant difference between any of them (significant here being set at 5% or more of a difference). Once trimmed, a Savitzky-Golay fit with a third degree was used to clear out the significant amount of noise that existed in the data. Figure 3 shows the conditioned data of the same eighty kilovolt shot.

Figure 3



Afterwards, code was written to apply the math shown in the Methods section. The results were unexpected, with there being no clear relationship present between charge density and time, the graph simply resembling what would be called statistical noise. Figure 4 shows the graph for the obtained charge density for the same eighty kilovolt shot.

Figure 4



All of the data collected follows this same pattern more or less, with no clear pattern being identifiable thus. Also, all the values in the graph are with the 100 decibel attenuation which was used when collecting data.

Conclusion

There does not seem to be any pattern for how the average charge density over the simulated lightning strikes change with time. This would indicate that either the charge density over lightning strikes is chaotic and does not follow any clear pattern, or there is always the chance that the data collected is flawed itself or there is a gimmick within the analysis. There are worries to be had regarding the data collection due to the sheer amount of noise which is present.

A noteworthy thing to consider is the laboratory's environment itself, for there is an extremely good chance for it to have produced a lot of the noise in the data. Before testing occurred, the lightning tended to strike into the concrete next to the copper plate, and not on the grounded copper plate itself. An insulating board had to be placed to prevent such an event to happen again, and even with this it can be seen from the picture of the set-up that streamers still attempted to hit areas outside of the copper plate.

Future Research

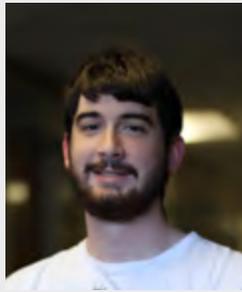
For future research, the effect of branching could be further analyzed as per Shi, W., & Li, Q. (2014) [2], and investigate how charge density changes with different branches of a lightning strike. This would help the model be more relevant in real life scenario, since most natural lightning strikes tend to branch. The main difficulty for such an analysis would be the distance needed to generate a lightning which would chaotically generate branching. A cross-analyzation from the data from all channels while using high-speed cameras could be also performed. This way, a 3D view of the lightning strike can be had and the way the luminous channel forms in a particular direction can be accounted for within the electric field, hence more accurate calculations be conducted. If a relationship is then found between time and charge density, lightning shots at different voltages could be compared to determine the weight that the voltage has on the charge distribution inside of the luminous channel. Initially, this experiment was supposed to be carried with a high-speed camera but due to the available resources, it was not possible to obtain one.

Works Cited

Jiang, Rubin et al. (2014). High speed video observation on random stepping and branching of negative leader. 2014 International Conference on Lightning Protection (ICLP). doi: 10.1109/iclp.2014.6973357
 Shi, W., & Li, Q. (2014). Research on the charge distribution of lightning stepped leader considering the branches. 2014 International Conference on Lightning Protection (ICLP). doi: 10.1109/iclp.2014.6973342



Guillermo Hoffmann Meyer



David Barber

Creating an Algorithm for Autonomous Vehicles to Calculate Routes

Abstract

The purpose of this experiment was to simulate an automatic vehicle’s delivery route and examine how it recalculates if a road is blocked or slowed. To do this, Dijkstra’s algorithm, which calculates the shortest route between points, was used. First, a python program that employed Dijkstra’s algorithm was found. After that, a grid was arbitrarily chosen, and values were assigned to the lengths between nodes randomly between 1 and 8. These lengths represent time to travel rather than actual distance. Then, a function to simulate a car moving on the grid was added to the program. After that, the program calculated the optimal path from the top left node to the bottom right node. Then, a function to add random accidents was added; this would render certain distances unfavorable to cross. The accidents happened at intersections and were manifested in an increase in travel time across a street. When the random accident happened, the program would adjust the node distances, and recalculate its path if the accident happened on a path it was going to take. The conclusion was that even if an accident occurs on the path ahead of a vehicle, it could still be the optimal path. If an accident happens on a series of shorter paths, then waiting out the shorter paths could be preferred over taking other streets. Future uses of this research could include observing changes in traffic multiple nodes away from the accident and optimizing paths based on other cars’ movement as well.

Introduction

The use of electric or automatically driven vehicles is becoming increasingly popular and efficient in the modern world. Vehicles that can transport things to humans without chance for human error provide much business advantage as opposed to paying people to drive vehicles. These autonomous vehicles rely on Light Detection and Ranging (LiDAR) in addition to ultrasonic sensors to detect both still and moving objects around them along with traffic signals. They also use a GPS to keep track of where their position is anywhere in the world, and coordinate with high definition maps, that have every road and traffic light in cities, with constant updates that will alert an autonomous vehicle to a traffic jam or roadblock ahead. This allows the autonomous vehicle to calculate a new and faster route easily. Dijkstra’s Algorithm can be used to show how both them together will provide the safest and fastest route possible.

Research Objectives

- The primary objective of this exploration was to simulate what a vehicle moving along a GPS-provided path in the event of an accident and observe how its path might change. The model was made on a few base assumptions:
1. Any accident that occurred does not completely remove a road from transit. – this could create situations in which the Dijkstra algorithm has no connections to its destination node, making the path unsolvable.
 2. The path is always from the top left of the grid to the bottom right of the grid. – as opposed to random spots in the grid, this makes the path lengthy and thus able to be changed by an accident. Although the accident node is random, it has a higher chance of affecting a longer path.
 3. The GPS instructions and navigations are correct. – This removes the possibility of the GPS miscalculating routes or giving separate routes with similar estimated times of arrival.
 4. The accident does not render the vehicle unable to move. – If the accident, such as a tree falling, leaves the vehicle unable to move, the model cannot calculate routes.
 5. The accident happens at an intersection. – This simplifies the model, as an accident in the middle of a road would require the creation of multiple new paths and closing paths off from one direction. Additionally, the Federal Highway Administration states that more than fifty percent of car crashes occur at intersections.[1]

Methods

The first step was to procure previously created and open-sourced code for

Dijkstra’s algorithm in python [2]. From there, the program was modified to account for accidents and larger grids.

In its initial state, the program has six nodes, “a” through “f.” These were replaced with double letter names in favor of easy recognition of nodes. The first step in making a program to simulate a city situation was to make the amount of nodes easily expandable. To do this, the use of many lists of vertices were used. Each row was comprised of a list of vertices, or nodes. From there, each row was added to a list of rows. Then, for each node in each row, there was a height variable assigned randomly to be the distance between two nodes in the same positions in different rows. Similarly, the width variables were used between two nodes of the same row. (Figure 1) The numbers for height and width were taken as random integers between 1 and 8, inclusive. These numbers aren’t distances, but rather represent the difficulty or time necessary to cross certain paths.

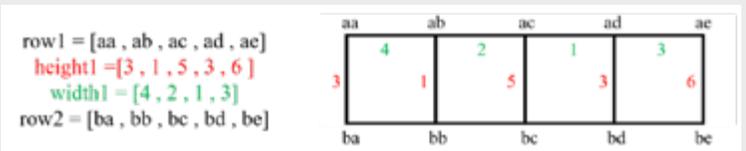


Figure 1. The lists of nodes, widths, and heights, and what they correspond to on a grid.

The next step was to make sure that all the lists of heights, nodes, and widths were able to be made into a list of paths. A path consisted of a start node, an end node, and a length. For example in figure 1, the path would appear as (“aa”, “ba”, 3). The list of rows was looped through to find all total nodes, and then for each node, its height and width (if available) were used to make a path. After that, all the paths were added to a list of paths, which is the input that the Dijkstra function took.

Once the Dijkstra program outputted a list of nodes to visit in order, the path distance between the nodes was found by checking through the list of paths. From there, the total amount of time it would take to travel nodes was calculated. Using this total travel time, a random variable was created to simulate the time at which an accident occurred.

The time of travel was then tracked as the program followed its path until the accident happened. Then, it finished its current road and recalculated its path, with adjustments made to the path: every node that was connected to the accident node gained difficulty or travel time between 3 and 9; if the random number was 9, it was changed to 50 (which made the road not optimal,

simulating the closing of the road for a traffic jam or bad traffic.

Results and Discussion

First, the program optimized a route between the two nodes. (Figure 2) In this example, a 5x5 node grid was used. The program outputted multiple different paths across multiple simulations, and there were three possible cases. One where the accident happened on the path and the path changed, one where the accident happened on the path and the path didn't change, and a third where the accident occurs off the path and thus does not affect anything. (Figures 3, 4, and 5, respectively)

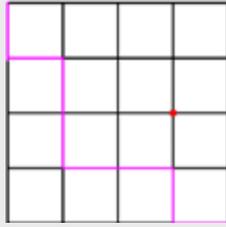


Figure 2. A path from the point aa to the point ee

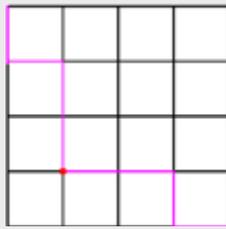


Figure 3. The path has changed after an accident at point db.

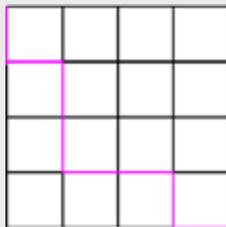


Figure 4. The path was not affected, as the accident at point cd does not affect any nodes travelled through

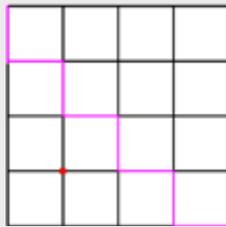


Figure 5. The accident happens on the path at db, but the path does not change. The first and third cases are expected of a GPS, as when an accident occurs cars are rerouted to faster paths. However, the second case shows that sometimes it is possible that waiting out slow traffic near an accident is faster than attempting to take a different route. The second case could be a product of the random variable assigned to road lengths and delay size; if the roads where the accident happens appears to be short, the path could remain optimal.

Conclusion

The use of Dijkstra's Algorithm can result in unexpected outcomes, such as continuing along a path with an accident-related delay on it. The presence of the first of the three cases confirms that the vehicle can and will reroute itself if the accident is in its way and proves to be a sub-optimal path. The second case shows that even if an accident delays its future path, it may continue to be the optimal path. The third case is the most common one; an accident occurs but does not affect the car or any path the car takes.

This model's strengths lie in being able to generalize a city into nodes and paths and observe the procedure a GPS would take to reroute a vehicle after being informed of an accident. However, its weaknesses include that it cannot account for crashes or accidents on streets rather than intersections,

and it cannot account for a road being completely blocked off from all traffic. Future research could observe the impact of accidents on nearby nodes rather than just the path taken: the notion that cars are rerouted to optimal paths could be used to check if the accident at one node could affect traffic at a different node multiple nodes away. If an accident happens in one spot and all vehicles are rerouted to the same other path, that other path will become slower and thus not optimal.

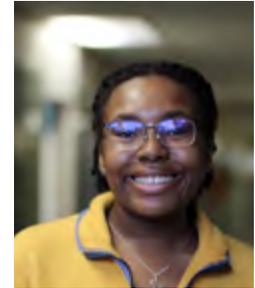
The project is made with sustainable development goals 9 and 11 in mind. The increase in presence of electric and self-driving vehicles will overhaul the infrastructure system and reduce air pollution. Goal 11 in particular deals with public transport, and the use of self-driving public transport will save cities money and emissions in addition to less traffic jams.

References

- [1] Federal Highway Administration, Intersection Safety <https://cms7.fhwa.dot.gov/research/research-programs/safety/intersection-safety>
- [2] Boldyreva, Maria "Dijkstra's algorithm in python: algorithms for beginners", 2018, <https://dev.to/mxl/dijkstra-algorithm-in-python-algorithms-for-beginners-dkc>



Catherine Min



Alicia Argrett

Developmental Effects of Nicotine Exposure on *Drosophila melanogaster*

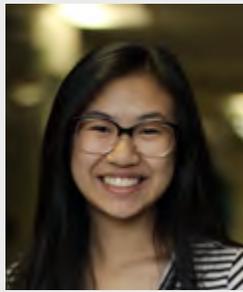
The purpose of this experiment was to subject the *Drosophila melanogaster* under different dilutions of nicotine: 5 μL , 15 μL , and 30 μL . Each dilution of nicotine was mixed into the 3 tubes of food, causing the flies to consume the mixture. Using the genomes of the flies, the study focused on the effects of nicotine of the flies by measuring oxygen intake and the developmental physiques of the flies before and after nicotine exposure. The null hypothesis is that there will be no difference of oxygen gas levels between the different dilutions of nicotine and control tubes of *Drosophila melanogaster*. The alternative hypothesis is that there will be a difference of oxygen gas levels between the different dilutions of nicotine and control tubes of *Drosophila melanogaster*. It is hypothesized that oxygen intake will vary among the different types of flies with the non-nicotine flies intaking the most oxygen. Oxygen input directly correlates with carbon dioxide output in the flies, reflecting metabolic levels. The rise of adverse effects of nicotine consumption surging through the masses has amplified health concerns relating to pulmonary responses (Yatsenko, 2014). The effects of developmental nicotine exposure have been documented in *Drosophila melanogaster*, epitomizing known effects of exposure in humans: lower birth weight, delayed development, and alterations of the cholinergic system. Prenatal nicotine exposure also leads to structural changes including differences in brain size (Morris, 2018). This experiment found that control flies, on average, were 75% larger than the 5 μL flies. Future implications would be to determine the pulmonary effects of consistent nicotine inhalation over time. Combating this mystery would put science closer to solving the third Sustainable Development goal: ensuring healthy lives and promote well-being for all at all ages. A *Drosophila melanogaster* model for these effects of nicotine exposure allows for easier focused on determining changes to oxygen intake and physique at two developmental stages of *Drosophila melanogaster* due to nicotine exposure. The aim of this experiment is to provide a realistic model of nicotine effects on humans.

Works Cited

- Morris, M. et al. Developmental nicotine exposure affects larval brain size and the adult dopaminergic system of *Drosophila melanogaster*. *BMC Developmental Biology* 18,(2018). Yatsenko, A. S., Marrone, A. K., Kucherenko, M. M. & Shcherbata, H. R. Measurement of Metabolic Rate in *Drosophila* using Respirometry. *Journal of Visualized Experiments*(2014). doi:10.3791/51681. The effects of developmental nicotine exposure have been documented in *Drosophila melanogaster*, epitomizing known effects of exposure in humans.



Violet Jira



Gina Nguyen

Analysis of the Effectiveness and Antimicrobial Capabilities Against Bacterial Pathogens of *Lactobacillus* sp. and Other Strains Found in Probiotics

Abstract

In an attempt to improve gut flora, many individuals purchase and self-administer various brands of probiotics. The manufacturers of these goods claim that the bacteria in these probiotics are better than the ones humans have naturally--that they have the ability to promote and maintain the growth of "good bacteria," aid in the eradication of "bad bacteria," and assist in the achievement of overall good health. This research sets out to prove if there is any truth to these claims. To achieve this, three commercially available brands of probiotics were cultured in Nutrient Broth and placed in a competition study with *Escherichia coli*, *Micrococcus luteus*, and *Candida albicans*. *E. coli*, *M. luteus*, and *C. albicans* are all opportunistic pathogens, *C. albicans* being a common member of the human gut flora. To facilitate this competition study, the three bacteria were acquired and diluted in sterile nutrient broth to a turbidity 0.5 McFarland (1.5 x 10⁸CFU/ml). Nutrient Agar plates were inoculated with the bacteria. Yeast Malt Agar plates were also inoculated with the *C. albicans*; this was done because Yeast Malt Agar is the optimum growth medium for yeast. Filter paper discs were soaked in a nutrient broth containing one of the three probiotics tested. Control plates contained discs soaked in only the nutrient broth. The plates were then allowed to grow overnight, and the zones of inhibition (if any) were measured. Zones of inhibition are indicative of the probiotics having antimicrobial capabilities against the yeast and bacteria. A lack indicates the opposite. These findings, collectively, are indicative of several things, such as the truth to claims offered by the manufacturers, our bodies capability to take up the bacteria within the probiotics, and the viability of the cultures within the capsules.

Introduction

In an attempt to improve gut flora, many individuals purchase and self-administer various brands of probiotics. The manufacturers of these goods claim that the bacteria in these probiotics are better than the ones humans have naturally--that they have the ability to promote and maintain the growth of "good bacteria," aid in the eradication of "bad bacteria," and assist in the achievement of overall good health. Some probiotics claim to offer significant health benefits, offering "aid in maintaining healthy intestinal flora" and "participation in the detoxification process." Another claims it can "help to maintain friendly bacteria colonies which can be depleted by stress, antibiotics, and gastrointestinal disturbances."

A majority of the bacteria found in these capsules belong to the genus *Lactobacillus*. "*Lactobacillus* has demonstrated its ability to inhibit several bacterial pathogens, including *Clostridium difficile* (McFarland, 2015), *Escherichia coli* (Kumar et al., 2016), *Shigella* spp. (Mirnejad et al., 2013), *Streptococcus mutans* (Ahn et al., 2018), *Pseudomonas aeruginosa* (Jamalifar et al., 2011), and *Staphylococcus aureus* (Kang et al., 2017)" (Chen, C. C., et al, 2019).

With this in mind, this research sets out to prove if there is any truth to these claims. To achieve this, three commercially available brands of probiotics were cultured in Nutrient Broth and placed in a competition study with *Escherichia coli*, *Micrococcus luteus*, and *Candida albicans*. *E. coli*, *M. luteus*, and *C. albicans* are all opportunistic pathogens, *C. albicans* being a common member of the human gut flora.

Objective and Hypothesis

The objective of this experiment is to determine whether the cultures found in commercially available probiotics are viable, and whether they show signs of antimicrobial activity when placed in a competition study with *E. coli*, *M. luteus*, and *C. albicans*.

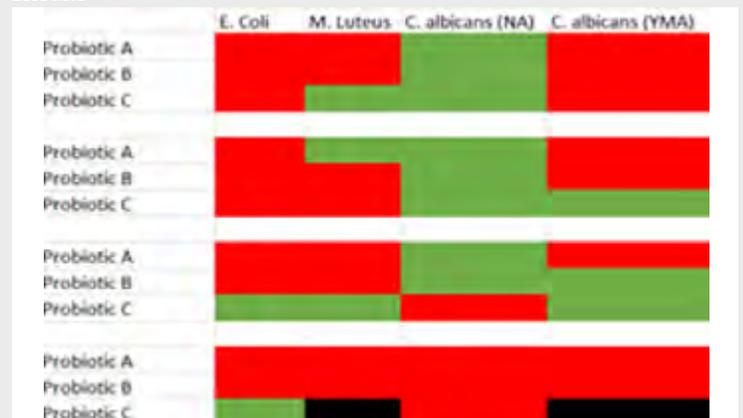
The null hypothesis is that none of the probiotics contain viable cultures, and that the probiotics do not have antimicrobial capabilities against pathogens. The alternate hypothesis is that at least some of the probiotics contain viable cultures and that the probiotics do have antimicrobial capabilities against pathogens.

Materials and Methodology

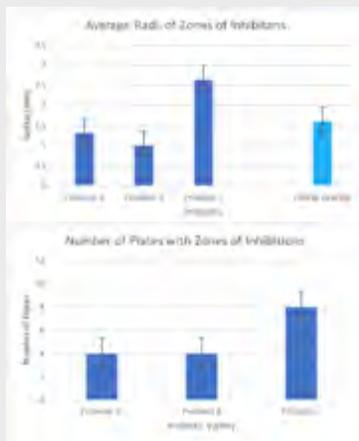
In addition to lab equipment, many materials were used throughout this experiment including Nutrient Agar, Nutrient Broth, Yeast Malt Agar, filter paper disks, *Escherichia coli*, *Micrococcus luteus*, *Candida albicans*, and Probiotics A, B, and C.

To carry out the experiment, the bacteria and yeast were subcultured to acquire fresh cultures for experimentation. Flasks of sterile Nutrient Broth (NB), Nutrient Agar (NA) plates, and Yeast Malt Agar (YMA) plates were prepared. After inoculating the NB flasks, each with a different brand of probiotics. After placing the flasks on a shaker for approximately 24 hours, a spectrophotometer was used to monitor any growth from the probiotics. Then, each of the bacteria and yeast were diluted with NB using a vortex before adding them to its respective agar plates. Sterile filter paper discs were coated with the NB-probiotic mixtures and placed in the center of the respective plates; a control was designed for the filter paper discs to be coated with just NB. After sitting in an incubator for approximately 24 hours, the plates were observed, and any zones of inhibition were measured and recorded. The plates were parafilm and stored in a refrigerator for future reference.

Results



Key
 Red- No zones of inhibition
 Green- Zones of inhibition present
 Black- No data



Discussion

Based on the results, Probiotic C proved to be most effective of the three brands, exhibiting the most and largest zones of inhibitions on the bacteria plates with an average of 2.643 mm. Probiotic A had an average of 1.313 mm, and Probiotic B had an average of 1.000 mm. One factor that could have led to this is the different bacterial strains found in each of the probiotics. There are five strains found in Probiotic C that Probiotics A and B do not have: *Lactobacillus paracasei*, *Streptococcus thermophilus*, *L. reuteri*, *L. fermentum*, and *Saccharomyces boulardii*. Specifically, *S. boulardii* is a yeast that introduces beneficial microbes to the gut flora and protects it against pathogens.

Additionally, the most resistant bacteria was *E. coli*, followed by *M. luteus* and *C. albicans* on Yeast Malt Agar, leaving *C. albicans* on Nutrient Agar as the least resistant. Because Yeast Malt Agar is the optimum medium for yeasts, the *C. albicans* had a competitive advantage. *C. albicans*, however, is a common member of the gut flora, so most probiotics will be designed to prioritize attacking and replacing it.

Conclusion

All three probiotics exhibited at least four zones of inhibition. Probiotic C was the most effective, and the bacteria that was the most resistant to the probiotics was *E. coli*. After the probiotic/Nutrient Broth mixture had been on the shaker for 24 hours, there was a significant change in turbidity within all three flasks. From this research it can be concluded that the cultures within the capsules were indeed viable.

Future Research

There is more research that could be done to discern exactly which strain of bacteria/yeast is the most effective, individually, in a competition study with pathogens. In this research, Probiotic C had the most and largest zones of inhibition. Similar research could be conducted with the strains of yeast and bacteria that were specific to Probiotic C. These four bacteria and yeast are in Probiotic C but not the other two probiotics. Research could also be done with probiotics that are more similar to Probiotic C to gather data on what the most effective bacteria and yeast are across the board.

With regards to sustainable development, research like this is important to maintain the health and well-being of those who purchase and use probiotics across the globe, or achieving sustainable development goal three. In the United States some probiotics are designated as dietary supplements, which don't require FDA approval before they are marketed. Research like this is important to keep consumers informed about what they are putting into their bodies, and to hold companies accountable for the claims that they make.

References

- Chen, C.-C., Lai, C.-C., Huang, H.-L., Huang, W.-Y., Toh, H.-S., Weng, T.-C., ... Tang, H.-J. (2019, April 18). Antimicrobial Activity of *Lactobacillus* Species Against Carbapenem-Resistant Enterobacteriaceae. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/31057508>
- National Center for Complementary and Integrative Health. (2019, August). Probiotics: What You Need to Know. Retrieved from <https://nccih.nih.gov/health/probiotics/introduction.htm>
- Jürgen Schrezenmeir, Michael de Vrese, Probiotics, prebiotics, and synbiotics—approaching a definition, *The American Journal of Clinical Nutrition*, Volume 73, Issue 2, February 2001, Pages 361s–364s, <https://doi.org/10.1093/ajcn/73.2.361s>
- Seppo J. Salminen, Miguel Gueimonde, Erika Isolauri, Probiotics That Modify Disease Risk, *The Journal of Nutrition*, Volume 135, Issue 5, May 2005, Pages 1294–1298, <https://doi.org/10.1093/jn/135.5.1294>
- Kent L. Erickson, Neil E. Hubbard, Probiotic Immunomodulation in Health and Disease, *The Journal of Nutrition*, Volume 130, Issue 2, February 2000, Pages 403S–409S, <https://doi.org/10.1093/jn/130.2.403S>



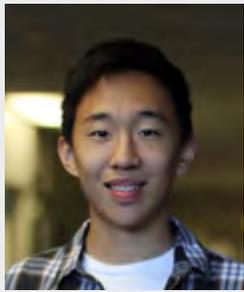
Victoria Callahan

World Food Prize: Education in Haiti

Haiti is a country in the Caribbean Sea that has two official languages: Haitian Creole and French with French being the main language of the education system despite being spoken by fewer people. Haiti has compulsory education for children between the ages of 6-11 with the curriculum based off that of the French (Lawless & MacLeod, 2019). Many people in Haiti do not have access to quality education (Girault & Lawless, 2019). Many programs have attempted to aid Haiti in improving its education system such as the USAID sponsorship of the ToTAL program. This program aimed to improve the literacy rate in both Creole and French through a new curriculum and better teacher training; focusing on children in grades one through three (USAID, 2015). Haiti's government also attempted to improve its education system passing the Benard law in the 1970s (Rea, 2014). This law could now be used as a basis to bring about education reform and to secure native funding outside of foreign aid. Programs that could be implemented would be similar to those seen in the ToTAL program. These programs would include beginning with instruction in Haitian Creole and then adding French instruction beginning in grade two. This method would allow for students to become knowledgeable in both Haitian Creole and French. The curricula for years one through three which were created previously during the ToTAL program could be used again leaving grades four through six needing a curriculum at the primary level. Implementation of programs should begin in a small, localized area for a period of two to three years. This would be similar to how the ToTAL program only operated in three areas of Haiti. After this period, the programs that had been at work could be re-evaluated and adjusted before being slowly expanded out over a period of ten to fifteen years. Ideally, over time, this concept would encompass the whole country. This method of expansion would put the timeline at anywhere between twelve to eighteen years depending on results from the original trials and expansion times. Although this time would not meet the 2030 goal of the United Nation's Sustainable Development goals, it would put Haiti on track to achieving the goal sometime within that decade. Haiti improving its education system is an important part of obtaining Sustainable Development Goal number four which is to provide a quality education to everyone.

Works Cited:

- Girault, C. A., & Lawless, R. (2019, October 30). Housing. Retrieved from <https://www.britannica.com/place/Haiti/Housing>
- Lawless, R. J., & MacLeod, M. J. (2019, October 30). Climate. Retrieved from <https://www.britannica.com/place/Haiti/Climate>
- Rea, Patrick Michael, "The Historic Inability of the Haitian Education System to Create Human Development and its Consequences" (2014). CUNY Academic Works
- USAID. (2015, January). USAID- Haiti ToTAL. Retrieved from https://pdf.usaid.gov/pdf_docs/PA00K911.pdf



Michael Lu

Development of Predictive Tools for Anti-cancer peptide candidates using Generative Machine Learning Models

Abstract

Cancer represents a leading cause of high mortality rates around the world. Many scientists have explored anticancer peptides (ACP's), which are peptides with anti-tumor activity, and they can be safer than conventional drugs due to high activity coupled with high selectivity and delivery control. However, current in vitro methods of discovery are both time-consuming and expensive. This study aims to utilize modern machine learning tools to formulate new ACP candidates. First, a support vector machine, a random forest model, a K-nearest neighbor model, and a multilayer perceptron model were trained on the DRAMP, PPD, and ADP databases totaling 584 known ACP's, and 584 random peptides from the Swiss Protein database were used as a failset. The SVM model achieved an accuracy of 90.4% with an f1-score of .89. A generation model was also developed through a long short-term memory recurrent network. The model created 40 preliminary sequences, and the sequences were checked through multi-layer verification in SVM model evaluation, literature-based properties, and web server predictors, concluding that around 90% of the generated sequences carry anticancer properties. Current literature is focused on predicting whether a peptide is an ACP, without generating drug candidates, so this study may be the first to generate new anticancer peptides. These generation techniques can be developed to facilitate the discovery of new potential anti-cancer peptides.

Introduction

Cancer is one of the leading causes of high mortality rates around the world. Drugs are used for cancer treatments, and although they are beneficial, they can also bring many harmful side effects to patients. This calls for different types of solutions to cancer that can avoid the harms of conventional drugs. Anticancer peptides (ACP's) are potentially a key solution. In fact, they are safer and have many comparative advantages. ACP's have high levels of activity, specificity, affinity, and they are less immunogenic and better delivery control (Gholibeikian et al., 2019). ACP's are peptides that typically range from 10-30 residues in length. There are three main types of ACP's. Pore-forming peptides target the cell membrane of cancer cells. Cell penetration peptides (CPP's) enter the cell, and tumor-targeting peptides bind to the receptors on cancer cells (Marqus et al., 2017).

Because the discovery of the potential of anti-cancer peptides is relatively recent, there are only 10 anticancer peptides that are currently being developed as drugs (Shoombuatong et al, 2018). More research is needed in ACP's as Yin et al. (2019) explains that the great potential of anti-cancer drugs hasn't been fully researched. Furthermore, current in vitro methods of discovery are both time-consuming and expensive (Manavalan et al, 2017), so this calls for more effective ways to create anticancer peptide candidates. Machine learning has the potential to fill this role. Machine learning models are characterized by statistical models that train from datasets to identify patterns, and these models are used to create predictors and classifiers.

As it stands, machine learning applications in this field has focused on creating only predictors that classify whether given peptides are ACP's. These classifiers have been able to predict peptides successively, without having the ability to generate new ACP candidates. Generative machine learning is a type of machine learning that analyzes existing datasets to generate new instances of data. One type of generative model is the long short-term memory (LSTM) recurrent neural network, which generates new sequences of data. This type of model has been used in the generation of antimicrobial peptides, a similar research area, and has successfully generated new antimicrobial drugs. Thus, an LSTM recurrent neural network can likely be applied to anticancer peptides. This study aims to be one of the first to utilize generative machine learning models to create new anticancer peptides.

Methods

Predictive Models

There were four classification models tested. The first model is the support vector machine model, which is a kernel-based classifier with hyperplane

divisors and margins. The second model is the random forest model, which is an ensemble learning method that utilizes a group of weak predictors known as decision trees. The third model is the K-nearest neighbor classifier, which predicts a class based on the similarity of the features of a sample using distance calculations. The fourth model is a multi-layer perceptron, which is a multi-layered, feedforward neural network.

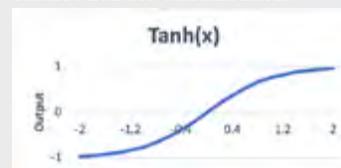
Cross Categorical Cross Entropy is the loss function for the multilayer perceptron, and it is defined by:

$$CCE = -(y \times \log(p)) + (1 - y) \times \log(1 - p)$$

where $p = \text{probability}, y = \text{true value}$

Generative Models

A long short-term memory (LSTM) recurrent neural network (RNN) was used as a generative model to generate new sequences of anti-cancer peptides. The long short-term memory component of the network is composed of a cell, an input gate, an output gate, and a forget gate, and it is used to retain long-term information in sequential data to use in conjunction with short term data to make predictions. The generative models use tanh functions and sigmoid functions for the gates listed above. The output gate is fed to dense layers, a dropout layer, and a softmax activation layer to estimate the probability distribution for classification.



Data Collection and Processing

Datasets of verified anticancer peptides were collected from three databases. The first is the DRAMP (Data Repository of Antimicrobial Peptides) database (Kang et al., 2019) (Liu et al., 2018) (Liu et al., 2017) (Fan et al., 2016), consisting of 74 tested anticancer peptides. The second database is the APD (Antimicrobial Peptide Database) (Wang et al., 2016) (Wang et al., 2009) (Wang et al., 2004), which includes 219 verified anticancer peptides. The third database is the CancerPPD database of 374 experimentally verified anticancer peptides.

Sequences shorter than 12 residues were removed, and the datasets were combined. 584 unique anticancer peptide sequences were used for data processing. 584 unique sequences between the lengths of 10-25 residues were randomly sampled from the Swiss Protein Database. This database

includes 561,568 annotated and reviewed peptide sequences found in nature. The peptide sequences were processed into 11 types of physicochemical features through the PydPi library. These sequences were converted to a total of 2049 features (like AA composition, charge, hydrophobicity, etc.), and normalized in the range from 0 to 1 for each feature given by:

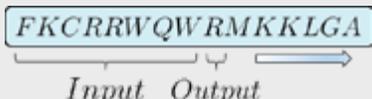
$$X_{i,k} = \frac{X_{i,k} - \min(X_k)}{\max(X_k) - \min(X_k)}$$

Then, two feature selection methods were tested. The chi-squared test was tested to find the features with the least variance for each class. The equation is given by

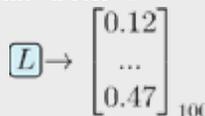
$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

The second feature selection method was through a random forest feature importance ranking of features.

There were also many data processing techniques used for the generator model. A shifting window technique was used to generate training and testing instances. For instance, the first 8 amino acids in a sequence can be used to predict the following amino acid, and this “window” is shifted to the right to create more data:



There are two-word embedding techniques tested. The first is one-hot encoding, which expresses each amino acid as a vector the size of the alphabet, with the location of the number 1 identifying the letter. For instance, if there are three letters in the alphabet [“a”, “b”, “c”], then the sequence “abca” would be represented as [1, 0, 0], [0, 1, 0], [0, 0, 1], [1, 0, 0]. The second word embedding technique is word2vec, which expresses each letter as a vector with the size 100, and it uses a neural network to extract contextual meaning in each letter. An example is shown below:



Metrics for evaluation

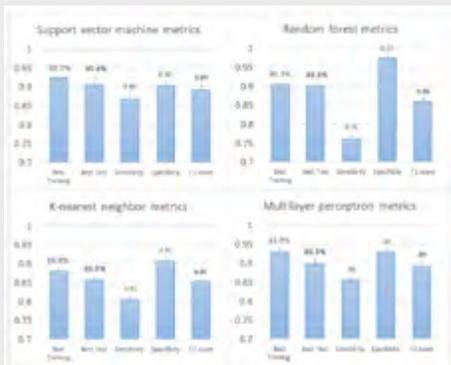
Many metrics were used to validate the effectiveness of the models. A simple accuracy metric was used to find the percentage of test and training samples the model could identify, specificity and sensitivity were found, and the F1 score was also calculated. These four metrics can be defined using true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The following metrics are used:

$$\left\{ \begin{array}{l} \text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \\ \text{Sensitivity} = \frac{TP}{TP+FN} \\ \text{Specificity} = \frac{TN}{TN+FP} \\ \text{F1 score} = 2 \times \frac{TP}{2TP+FP+FN} \end{array} \right.$$

Results

Predictor model final evaluation



To verify the different types of models, the initial dataset was split to 60% training data and 40% testing data. After the models are optimized for their parameters and trained, the average training and test accuracies are recorded

over 25 iterations. Then, the top-performing model’s training accuracy, testing accuracy, sensitivity, specificity, and F1-score is recorded.

SVM: The support vector machine achieved the highest accuracy over the other tested models. It achieved an average training and testing accuracy of 93.4% and 89.1% respectively, and the top-performing model scored a 92.3% training accuracy, a 90.4% testing accuracy, a sensitivity of 0.87, a specificity of 0.90, and an F1-score of 0.89.

Random Forest: The random forest model an average training and testing accuracy of 90.3% and 86.7% respectively, and the top-performing model scored a 90.4% training accuracy, an 89.9% testing accuracy, a sensitivity of 0.76, a specificity of 0.97, and an F1-score of 0.86.

K-nearest neighbors: The k-nearest neighbors algorithm had a training and testing accuracy of 88.0% and 85.7% respectively, and it scored a sensitivity of 0.81, a specificity of 0.91, and an F1-score of 0.85.

Multilayer Perceptron: The random forest model an average training and testing accuracy of 92.5% and 88.9% respectively, and the top-performing model scored a 93.0% training accuracy, an 89.5% testing accuracy, a sensitivity of 0.85, a specificity of 0.93, and an F1-score of 0.89.

Verifying generated peptides

First, the 40 generated peptides were tested against the top-performing SVM model:

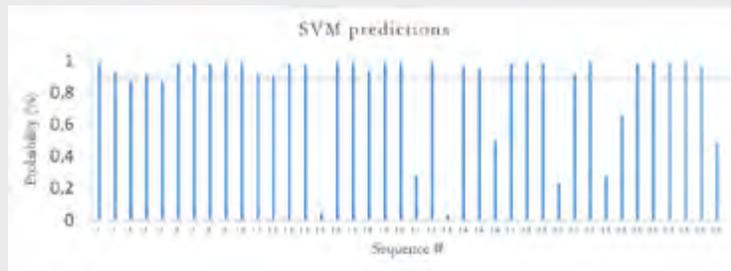


Figure 1: The initial SVM predictor’s probabilities for the 40 generated sequences with a line signifying the 90% confidence threshold.

Then, the 40 generated were tested against the ACPPred (Schaduangrat et al., 2019), ACPPred-FL (Wei et al., 2019), and Tyagi (Tyagi A . et. al., 2013) web predictors:

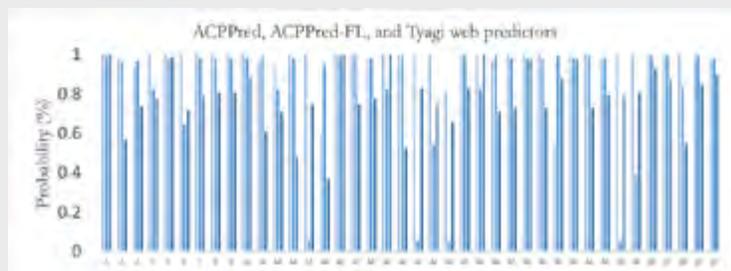


Figure 2: The ACPPred, ACPPred-FL, and Tyagi web server predictors’ probabilities for the 40 generated sequences with a line signifying the 90% confidence threshold.

Table 1: Indicators for ACP’s

Indicators	Description	Average Tested Values
1	Positive Charge	+5.08
2	High Hydrophobicity	41.30
3	Optimal Peptide Length = 21-30	22.22 residues
4	G (10.88 %), K (10.25 %) and I (11.23 %) are the most predominant.	K (21.73%), I (12.95%) are the most predominant, with G (7.54%) in 5th
5	ACP’s starts with G, I, A, F	60% of candidates follow this
6	ACP’s end with V, C, L, K	55% of candidates follow this

The 40 peptides were then tested against 6 literature-based indicators of anticancer peptides. First of all, anticancer peptides are known to exhibit a net positive charge and a high hydrophobicity (Shoombuatong et al., 2018). The forty preliminary sequences were calculated to have a +5.08 positive charge and a 41.30 hydrophobicity value. The optimal length of an anticancer peptide is found to be between the lengths of 21-30 residues long, and the

average length of the forty sequences is 22.22 residues long. The three most prevalent amino acids in anticancer peptides are Glycine (10.88%), Lysine (10.25%), and Leucine (11.23%) (Shoombuatong et al., 2018). These amino acids are also seen to be common amino acids in the generated sequences. Lysine (21.73%) and Leucine (12.95%) are the most predominant with Glycine (7.54%) as the 5th most common. Then, ACP's typically begin with a Glycine, Leucine, Alanine, and Phenylalanine, and 60% of the generated ACP candidates follow this pattern. Furthermore, ACP's usually end with a Valine, Cysteine, Leucine, and Lysine, and 53% of the generated ACP candidates follow this pattern. The 40 generated ACP drug candidates exhibit many known properties of anticancer peptides.

Discussion

Four types of machine learning models were used to create predictors to predict whether a peptide had anticancer peptides. Although there is plenty of literature on the creation of predictors for anticancer peptides, this study uses the largest dataset of anticancer peptides (584 tested anti-cancer peptides) to the knowledge of the researcher. This means that the trained models can be more generalized to generate a broader subset of anticancer peptides. The top-performing SVM model achieved a 90.4% testing accuracy, which is comparable to other predictive models.

These predictive models can be used in a few ways. First of all, these models can be used to predict favorable mutations for known ACP's. This allows known ACP's to be optimized for effectiveness. Second of all, these models can be used to predict new ACP's found in nature. Because ACP's have been found in nature in the past, then it is likely that many peptides in nature have undiscovered anticancer properties. The TrEMBL dataset includes over 177 million unreviewed peptides found in nature, and these predictors can make preliminary predictions for ACP candidates. Third of all, these predictors can be published online as web server predictors to allow scientists to verify their own anticancer peptides through statistical methods.

Furthermore, this study may be one of the first to generate new anticancer peptides using machine learning. Current literature is focused on predicting whether a peptide is an ACP, without generating drug candidates. In this project, a long short term recurrent neural network was trained on 584 known anticancer peptides to generate 40 new anticancer peptide candidates. These candidates were verified through self-created predictors, web server predictors, and known properties of ACP's. These peptides can be synthesized and tested on existing cancer lines like MCF7 and A549 to determine their actual efficacy. Overall, this study has introduced ACP predictors and novel generative machine learning models that can be used to promote the discovery of new anticancer peptides.

References

- Fan L.; Sun J.; Zhou M.; Zhou J.; Lao X.; Zheng H.; Xu H. DRAMP: a comprehensive data repository of antimicrobial peptides. *Sci Rep.* 2016 Apr 14;6:24482. PMID: 27075512
- Gholibeikian, M., Bamoniri, A., HoushdarTehrani, M. H., Mirjalili, B. B. F., & Bijanzadeh, H. R. (2019). Structure-activity relationship studies of Longicalcynin A analogues, as anticancer cyclopeptides. *Chemico-biological interactions*, 108902.
- Kang X.; Dong F.; Shi C.; Zheng H. et al. DRAMP 2.0, an updated data repository of antimicrobial peptides. *Scientific Data.* 2019; 6(1): 148. PMID:31409791
- Liu S.; Bao J.; Lao X.; Zheng H. Novel 3D Structure Based Model for Activity Prediction and Design of Antimicrobial Peptides. *Sci Rep.* 2018 Jul 25;8(1):11189. PMID:30046138
- Liu S.; Fan L.; Sun J.; Lao X.; Zheng H. Computational resources and tools for antimicrobial peptides. *J Pept Sci.* 2017 Jan;23(1):4-12. PMID:27966278



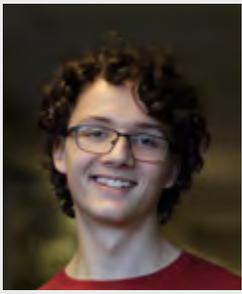
Aaron Wan

World Food Prize: China: Removing Roadblocks to the Belt and Road Initiative

China's Belt and Road Initiative, or BRI, is an initiative that seeks to boost trade and connectivity throughout its 125 member countries by investing trillions of dollars into connectivity infrastructure such as roads, bridges, and ports (Council on Foreign Relations, 2020). However, in recent years, China's BRI has run into several obstacles. Due to allegations of predatory investment, wasteful spending, and funding issues, several nations have backed away from Belt and Road projects (South China Morning Post, 2019). Since the start of the initiative, critics have considered the BRI as a plan for China to spread its geopolitical influence. Through massive BRI loans, critics say, China attempts to "trap" these countries in debt so that they will fall under the influence of China. In addition to the issue of predatory investment, the BRI also faces problems of wasteful spending. Over half of Chinese infrastructure projects analyzed were not beneficial, showing that they are economically unproductive (U.S.-China Perception Monitor, 2017). Another barrier that the BRI faces is a lack of funding; there is an annual deficit of \$819 billion (Asia Development Bank, 2018). There are several steps that China needs to take to address these issues. China should work with international institutions to improve its lending practices and help member countries pay off their debts (Center for Strategic and International Studies, 2018). To address concerns of wasteful spending, China should increase cooperation with governments of BRI member nations. Failing to effectively cooperate with other BRI member nations has allowed China to sign off on projects without pre-project viability analysis (South China Morning Post, 2019). The most critical step that must be taken by China is convincing the European Union to join the BRI. Convincing the EU to join the BRI would not only increase funding, allowing BRI projects to finally come into fruition (European Institute of Romania, 2018), but it would also increase the legitimacy for the initiative. The EU member countries are the world's strongest economic force, so the EU joining the BRI would increase the legitimacy of the BRI. As BRI member nations are increasingly backing away from projects, the EU must join the initiative in order to lend legitimacy to BRI projects. If the BRI is successful, it will lift millions of people out of poverty on a global level and help China address its domestic issues of food security and uneven development. (World Bank, 2019)

Works Cited:

- China's Massive Belt and Road Initiative. Council on Foreign Relations, Council on Foreign Relations, www.cfr.org/backgrounder/chinas-massive-belt-and-road-initiative.
- China's Belt and Road Criticised for 'Wasteful' Spending. South China Morning Post, 12 June 2019, www.scmp.com/news/china/diplomacy/article/3014214/chinas-belt-and-road-initiative-criticised-poor-standards-and.
- Ruka, and Michael B. Cerny. "China's Addiction to Unproductive Infrastructure - US-China Perception Monitor." US, 7 July 2017, uscnpm.org/2017/07/07/chinas-addiction-unproductive-infrastructure/.
- Funding the Belt and Road Initiative. Funding the Belt and Road Initiative | Global Commercial Banking HSBC, Asia Development Bank, www.business.hsbc.com/belt-and-road/funding-the-bri-initiative.
- It's a (Debt) Trap! Managing China-IMF Cooperation Across the Belt and Road. It's a (Debt) Trap! Managing China-IMF Cooperation Across the Belt and Road | Center for Strategic and International Studies, www.csis.org/npfp/its-debt-trap-managing-china-imf-cooperation-across-belt-and-road.
- Ciurtin, Horia. A Pivot to Europe. European Institute of Romania, ier.gov.ro/wp-content/uploads/publicatii/Final_Policy-Brief-5_Horia-Ciurtin-A-Pivot-to-Europe_web.pdf.
- Belt and Road Economics: Opportunities and Risks of Transport Corridors. World Bank, www.worldbank.org/en/topic/regional-integration/publication/belt-and-road-economics-opportunities-and-risks-of-transport-corridors.



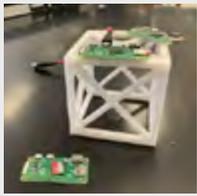
Blake Cheater

Modular CubeSat

The construction of a CubeSat starts with the assembly of the frame. Twelve aluminum L-beams nine centimeters (9cm) in length are bolted into a cube frame using eight aluminum attachment points. Aluminum is used because it is lightweight compared to other metals and relatively cheap. Similar to the Stanford SAPHIRE picosatellite, the outer panels and frame will be used for mounting various devices (Heidt, Puig-Suari, Moore, Nakasuka, & Twigg, 2000).

The CubeSat will have a miniature gyroscope which will help control the movement and orientation of the craft. It will be in the middle of the craft and receive signals from the computer network. Sensors around the satellite will gather data and send it to the computers where they will compute the numbers to determine the correct orientations. Alternatively, a signal from the ground could be received via the transmitter, and the gyroscope will orient the craft according to the commands given. The gyroscope will be bought instead of built, to ensure professional grade quality.

In the middle of the CubeSat, also modeled after the SAPHIRE picosatellite, a computer network will control the modules and satellite itself. It will be controlled by an array of three Raspberry Pi zero minicomputers. The systems



will run ubuntu Linux as the operating system and the flight controller will be a program written in python. Each Raspberry Pi zero will be mirrored and run the exact same program and calculations at every moment. They will also intake data separately from every sensor simultaneously for the calculations. All three computers will calculate a recommended course of action to be taken

at every second and then send their decisions to a central microcontroller. The microcontroller will compare all three of the inputs. If all three inputs are the same, the microcontroller will confirm the course of action. If only two of the inputs match the microcontroller will confirm the action of the two Raspberry Pi's that matched and send a reset command to the Raspberry Pi that did not match in order to fix any potential issues. This will be used as a preventative measure for bit flips caused by cosmic radiation. If a bit in a system is flipped by radiation, it will produce a different answer than the other two systems. The flight data from the affected system will be copied to one of the other systems before it is immediately rebooted. When the system comes back online, its data will be restored from one of the other systems and it will rejoin the other systems in flight control calculations. This will ensure correct calculations and commands are used and executed and is a cheap alternative for preventing the negative effects of cosmic radiation. The Pi's will receive signals from the various sensors around the satellite and tell the correlated modules what to do. For example, as previously mentioned, sensors monitoring orientations will send signals to the computer, which will then interpret the data, and tell the gyroscope what to do to ensure proper orientation (Addaim, Kherras, & Zantou).

The power source for the satellite will be a battery system. The battery will be charged by an array of solar cells on all six faces of the structure. The solar panels will measure 3.75 by 2.5 inches, have a voltage of 6 volts each, and have a capacity of 4800 mah. The power generation capacity of the panels will be at least sixteen watts. The panels will directly store their energy in the battery array which will be comprised of Tesla energy cells. Five Tesla 2170 lithium-ion cells will be wired in an array to act as a single battery unit. The satellite will target a power draw of less than sixteen watts to ensure that power consumption never exceeds that of the power production.

The first possible module will be a telecommunications system. The telecommunications system will serve as a link and relay station between other

satellites and will ensure signal does not get disrupted if another satellite goes offline. It will be broadcasting signals at a specific frequency to other satellites and the ground. Ideally, the system will slide into the module docking section of the CubeSat and will be secured by using bolts or other devices (Klofas, 2008). Various bits of information will travel through the system, either commercial signals or possible government communications frequencies will travel through as well. It will use the Microhard P2400 system to broadcast and receive signals. A separate antenna will have to be added to increase the broadcasting radius from 50 miles to the desired distance.

The second possible module is a camera. The camera would slide into the module bay, being secured by bolts or other methods. It would be oriented towards Earth by using the gyroscope. Proper orientation may require multiple gyroscopes, which may be a problem. The camera will be used for taking pictures of Earth which can aid in wildfire fighting or government applications. A thermal or infrared camera may be used for other government applications as well (Rønning, 2012). The information from the camera would be processed through the raspberry pi's and then transmitted through the antenna and P2400 to other satellites or Earth. The SD card in the raspberry pi's will be able to store multiple pictures, depending on the size of the SD card inserted into the CubeSat during assembly. The camera will also be quite large, so room may have to be made for the camera (Rønning, 2012). Many government and philanthropic organizations have uses for this, such as NASA, fire fighters, search and rescue, environmental organizations, and much more.

The third possible module will be an internet transmitter. Much of the world does not have access to the internet because of infrastructure limitations, so the best way to combat this is to bypass the problem, by using satellites. Having large satellites would be too costly, so having multiple, smaller, and cheaper satellites is the best solution to this problem. By giving impoverished countries and continents like Africa access to the internet, it could help them improve their standard of living. Plans for satellite constellations like Elon Musk's Starlink will focus on further shrinking the technology and reducing costs. At the moment, internet service satellites require large antennae and receivers as well as increased computational resources to process the data. These systems dramatically increase the power consumption of the satellite as a whole and may require larger solar panels to supply the energy needed.

The satellite, in theory, will have its own propulsion system as well. The possible propulsion systems include hydrazine propulsion, vacuum arc thrusters, and electrospray thrusters. Hydrazine propulsion uses hydrazine (H₄N₂), a highly volatile substance around certain metal catalysts. A tank full of hydrazine will connect to the different thrusters around the craft. Once it receives a signal, a valve will open allowing the compressed hydrazine to flow out the desired nozzle, either turning or propelling the craft in its desired direction. The CHAMPS CubeSat uses four monopropellant thrusters for control over the craft (Schmuland, Masse, & Sota, 2011). A similar set up would be used with four thrusters controlling the pitch, yaw, and roll. Aerojet's thrusters consist of CNC machined titanium structures to withstand the pressure of the contents of the hydrazine (Schmuland, Masse, & Sota, 2011). The benefits of the hydrazine thrusters include three axis control and ΔV . The hydrazine thrusters provide much more ΔV than its other cold gas counterparts (Schmuland, Masse, & Sota, 2011).

Vacuum arc thrusters are relatively new and provide unique benefits especially during long term space travel. It uses solid metal fuel and cathode material to create a plasma. It requires a sizable amount of electricity and uses DC power. Essentially, as voltage is applied between two electrodes, and an arc is formed. (Kuehn, Kuehn-Kauffeldt, Seipp, & Schein, 2017). Vacuum arc thrusters have

many advantages. The most notable of them being the fact that they have very few moving parts. This cuts down the likelihood of something failing and thus causing the mission to fail. However, the downsides include the cathode ray thruster head eroding after some amount of time, rendering them useless (Pietzka, et al., 2013). The insulation required to prevent this from happening is heavy and can weigh down the satellite too much and also hinders the arc forming process.

Electrospray thrusters use electrostatics to provide momentum for spacecraft. In the thruster, liquid propellant flows through a high voltage electric field. It then becomes atomized into droplets because of the electricity, and turns away from the beam as it is oppositely charged (Roy, Hrubby, Rosenbald, Rostler, & Spence, 2009). Electrospray thrusters offer a sizable amount of ΔV and control over the spacecraft. There are many advantages to electrospray propulsion. These include the fact that there are very little moving parts, reducing the probability of a part breaking, causing the craft to be unable to move, thus compromising the mission. Also, the thrusters are quite compact with no valve or electronics, allowing them to be relatively easily placed in a CubeSat, where size is very important to mission success. Electrospray thrusters also are not a massive drain on power as other methods of propulsion can be; they draw a low amount of power compared to the thrust they produce (Roy, Hrubby, Rosenbald, Rostler, & Spence, 2009).

An important aspect of CubeSats to consider is their method of reaching space. Normally, CubeSats are launched as secondary payloads in rockets (Puig-Suari & Zohar). They are relatively easy to add onto an existing rocket as they are very small and weigh very little, so the extra mass will not have a massive

effect on performance. One of the largest problems for CubeSats is launching and maneuvering them accurately. Getting to space is not hard but maneuvering from there to the proper orbit and orientation can prove to be a challenge. When launching multiple CubeSats in one rocket, the clustering of the spacecraft becomes

a problem. The deployment mechanisms are often large springs, which deploy the CubeSats away from the larger craft. The standard procedure is to launch a satellite, wait a certain amount of time depending on the conditions and desired spread of the craft, and then launch another. However, according to Puig-Suari and Zohar, this is not efficient and does not always work well. Intervals of 45 days seem to be the best. More recently, Puig-Suari has proposed a new way of transporting and launching picosatellites. His new method is encompassing them in an aluminum cage, as aluminum is light weight and very durable, and it also provides a Faraday cage to help protect the satellite against space's harmful electromagnetic radiation (Puig-suari, 2001).

Thus, for all of these parts and modules to work, the CubeSat needs to have a solid frame. As previously mentioned, an aluminum frame is the most logical as it is very light, strong and relatively cheap. The best way to create the frame is to cast twelve 10cm long rods and connect them with small bolts or ties. It is possible to TIG weld them together, but that should not be necessary. It is vital that the frame is strong and able to withstand various pressures and stresses. Aluminum is perfect for this task. Another important factor is how to construct the frame. A 3-D model was created

to demonstrate a possible configuration of the frame, providing a visual and physical model to manipulate. With the model of the frame, it was easy to see where the different modules would potentially go, and the spacing problems that inevitably arose.

In addition to providing a visual model to manipulate, the 3-D model was created. The Mk. 1 frame consistently displays a larger area of significant impact compared to the

Mk. 2 frame in respect to displacement magnitude, or how much the frame moves under pressure. However, on the Mk. 2 frame with 1000 newtons of force, most of it is concentrated on a singular point, which could lead to a critical failure. For the Mk. 1 frame under stress, it is relatively evenly spread on the bottom legs, with the most amount of force being felt at the bottom corners. For the Mk. 2 satellite under stress, it is more evenly spread than the Mk. 1 satellite. The legs feel little pressure compared to the joints at the top and bottom, which feel a considerable amount. This could potentially lead to problems.

The Mk. 1 satellite under 4000 newtons of displacement has a maximum displacement of .0130 mm compared to the Mk. 2 displacement of .00457

mm. However, the Mk. 2 displacement is significantly more concentrated than spread out. The Mk. 1 stress is 1.405×10^7 pa compared to the Mk. 2's 2.824×10^7 pa. While the Mk. 2 satellite's value is larger, it is also in a much smaller (concentrated) area. Thus,



the Mk. 2 satellite is the best design, with the most force being concentrated rather than completely spread out. The maximum force it encountered would most likely not break the frame, and the extra side paneling helps protect the

insides. More study will have to be done.

The United Nations established the sustainable development goals in order to help advance the interests of humankind and ensure the survivability of the planet. The goals are very diverse, and span from gender equality to economic productivity. The CubeSat initiative, and the modular CubeSat in particular, correlates to at least three of these United Nations goals. Goal 9 which says,

“Almost all people around the world now live within range of a mobile-cellular network signal, with 90 per cent living within range of a 3G-quality or higher network. This evolution of the mobile network, however, is growing more rapidly than the percentage of the population using the Internet”

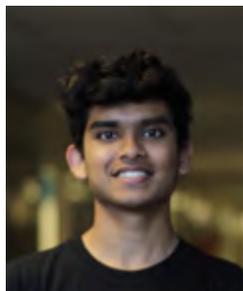
correlates to the internet module of the CubeSat. With the internet module, internet will be accessible everywhere on Earth with a blanket of picosatellites. Some of the greatest limitations today is lack of infrastructure, and that will be overcome as it is now CubeSat dependent. With the evolution of cube satellites, they become cheaper and thus internet around the world becomes less of a dream and more of a reality. Global internet access will also help achieve goal 4, quality education. With internet available, students who would not have had access to important resources would now have the ability to access any piece of information on the internet. This allows for countless possibilities and newfound opportunities for students to excel and help make their communities better and more educated.

Goal 13 depicts climate change and how it threatens modern society. CubeSats equipped with a camera could help with monitoring climate change and melting polar icecaps. They would be able to pinpoint heavily affected areas and help researchers develop ways to counter the rising temperatures. This would help curtail the effects of climate change by allowing environmentalists to develop specialized plans to thwart the harmful effects of climate change in specific areas and see where heavy carbon dioxide emissions are coming from.

References

- Addaim, A., Kherras, A., & Zantou, E. (n.d.). Design of Low-Cost Telecommunications CubeSat-class Spacecraft. Centre for Space Research Studies, EMI, 23. Retrieved October 2019
- Heidt, H., Puig-Suari, J., Moore, S. A., Nakasuka, S., & Twigg, J. R. (2000). CubeSat: A new Generation of Picosatellite for Education and Industry Low-Cost Space Experimentation. 14TH Annual/USU Conference on Small Satellites, 19. Retrieved October 2019
- Klofas, B. (2008, April 10). A Survey of CubeSat Communications Systems. Cal Poly San Luis Obispo, 12. Retrieved October 2019
- Kuehn, M., Kuehn-Kauffeldt, M., Seipp, L., & Schein, J. (2017). Solid Fuel Vacuum Arc Thrusters - New Concepts For Space Propulsion. University of The Federal Armed Forces Germany, Munich, Germany, 8. Retrieved December 2019
- Pietzka, M., Kuhn-Keuffeldt, M., Schein, J., Kronhaus, I., Schilling, K., Mai, T., & Lebeda, A. (2013, October 6). Innovative Vacuum Arc Thruster for CubeSat Constellations. 33rd International Electric Propulsion Conference, 7. Retrieved December 2019
- Puig-suari, J. (2001). Development of the Standard CubeSat Deployer and a CubeSat Class PicoSatellite. 7. Retrieved November 2019
- Puig-Suari, J., & Zohar, G. (n.d.). Development of CubeSat Constellations Utilizing Current Launch Opportunities. SRI International, 7. Retrieved December 2019
- Rønning, S. S. (2012, December 01). Optimizing an Infrared Camera for. Norwegian University of Science and Technology, 68. Retrieved November 2019
- Roy, T., Hrubby, V., Rosenbald, N., Rostler, P., & Spence, D. (2009). CubeSat Propulsion Using Electrospray Thrusters. Busek Co., 6. Retrieved November 2019
- Schein, J., Gerhan, A., Rysanek, F., & Krishnan, M. (n.d.). Vacuum Arc Thruster for CubeSat Propulsion. Alameda Applied Sciences Corporation, 8. Retrieved November 2019
- Schmuland, D. T., Masse, R. K., & Sota, G. C. (2011). Hydrazine Propulsion Module for CubeSats. 6. Retrieved October 2019
- Straub, J., Korvald, C., Nervold, A., Mohammad, A., Root, N., Long, N., & Torgerson, D. (2013). OpenOrbiter: A Low-Cost, Educational Prototype CubeSat Mission Architecture. MDPI, 32. Retrieved November 2019
- United Nations. (n.d.). Sustainable Development Goals Knowledge Platform. Retrieved December 2019, from Sustainable Development U.N.: <https://sustainabledevelopment.un.org/?menu=1300>





Shanay Desai

Evaluating the Efficacy of Green Synthesized Silver Nanoparticles with Aminopenicillins, Carbapenems, Third/Fourth Generation Cephalosporins, and Intracellular Antibiotics Against Multi-drug Resistant Bacteria

Abstract

The green synthesis approach to creating silver nanoparticles (AgNPs) is a novel procedure that is more environmentally friendly than other chemical. Other methods of creating AgNPs involve chemical experiments that are detrimental to the environment. The individual and combined antibacterial activities of five conventional antibiotics (Ampicillin, Imipenemon, Cefepime, Ceftazidime, and Bacitracin) with and without AgNPs were researched and tested on three multi-drug resistant bacterial strains, including *Escherichia coli*, *Staphylococcus epidermidis*, and *Bacillus megaterium*. This was carried out using the traditional Kirby-Bauer disk-diffusion method. After 24 hours, the zones of inhibition (rings formed around the antibiotic) were measured with respect to their diameter size. The critical inhibitory concentrations (CIC) were calculated from a derivation of Fick's second law of diffusion. The results show that addition of antibiotics with silver nanoparticles have a better antibacterial activity than those without AgNPs. Moreover, the addition of AgNPs with the antimicrobial agents showed a lower CIC, correlating to a lower medicinal dosage (concentration) needed to inhibit the growth of pathogenic bacteria. In conclusion, this goes to show that the AgNPs improve the efficacy of resistance against the bacterial strains. This novel method can be applied to the medical sciences, where resistance towards diseases is growing, as seen in the West Nile virus and Coronavirus (COVID – 19). Using these greener methods, antibiotic resistance can be found more effectively without harm to our environment.

Keywords: green synthesis, critical inhibitory concentration, Imipenemon, AgNPs

Introduction

Nanotechnology is a new field that is currently emerging in the twenty-first century. This field includes a variety of disciplines, such as electronics, aeronautics, biomaterials, cosmetics, and nanomedicine. Nanoparticles, fine microscopic particles that consist of organic molecules, inorganic molecules, and ions, can exhibit metallic properties as antimicrobials, as therapeutic/diagnostic sensors, and as gene/drug delivery carriers. These fine particles range in sizes from 1 nanometer (nm) to 100 nm. AgNPs have been studied thoroughly for their distinctive characteristics of antibiotic resistance to bacterial strains.

AgNPs were previously made using chemical processes, some of which were time extensive and produced chemicals that were toxic to the environment. A new opportunity with silver nanoparticles (AgNPs) includes a green synthesis approach where these particles are synthesized with fungi, bacteria, and plants. Another method in the creation of AgNPs involves dissolving ground coffee powder in a solution of silver nitrate. This process of producing AgNPs using greener methods results in safer scientific testing and less harmful effects on the environment.

In the past decade, a new discovery was made in which silver nanoparticles were used to fight pathogenic bacteria. This new discovery has led scientists and researchers to find new methods to control infections. Introducing these silver particles into cells is the new process of pathogenic resistance. In the DNA of the cells, enzymes and mutations can change the bacterial mechanism for fighting off the infection (proving resistance). This is done through altering the flow of silver composites through the cells, decreasing the susceptibility to the bacterial strain overall. This has led new areas of nanotechnology to apply these AgNPs into certain microbes as an effective way to treat infectious diseases. In this research, AgNPs created by more environmentally friendly methods than its predecessor were tested for their antibacterial efficacy separately and in combination with five widely used antibiotics (ampicillin, imipenemon, cefepime, ceftazidime, and bacitracin) against three different multi-drug resistant bacterial isolates (*Escherichia coli* (K-12), *Bacillus megaterium*, and *Staphylococcus epidermidis*), proving that resistance against pathogenic bacterial strains can be accomplished through greener, safer, and more efficient methods.

Purpose and Hypothesis

Due to the nature of AgNPs possessing well developed surface chemistry, defined size in a solution, and appropriate chemical stability, AgNPs should be a good choice to use as an organic nanosilver in combination with conventional antibiotics to decrease the susceptibility on multi-drug resistant bacterial strains.

The synergistic property of the AgNP and antibiotic pair can potentially lower the amount of both of these agents in a clinic's medicinal dosage. This would therefore increase its effectiveness overall and decrease the noxiousness on a patient.

The following antibiotics fall under the broader category of types of antimicrobial agents.

1. Ampicillin → Aminopenicillin
2. Imipenemon → Carbapenem
3. Cefepime → Fourth generation cephalosporin
4. Ceftazidime → Third generation cephalosporin
5. Bacitracin → intracellular

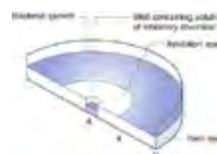
Since each of the five antibiotics exhibit qualities that make them fall under the category of the broad antimicrobial agents, patterns of resistance on other antibiotics within the same category can be extrapolated from this study and tested for similarities.

Kirby-Bauer Antibiotic Test

The process of testing the antibacterial resistance involves the Kirby-Bauer disk-diffusion process (also known as KB testing or disk-diffusion antibiotic sensitivity testing). Small wafer disks containing antibiotics are used to test whether the bacterial strain is susceptible or resistant to the specific antibiotic. To do this, the bacterial culture is grown overnight on either Mueller Hinton Agar (MHA) or Nutrient Agar (NA) with the wafer disks (containing antibiotics) slightly pressed into the agar.

Consider Figure 1 shown below

Figure 1. Antibiotic activity by agar diffusion



After the 24-hour incubation period, the antibiotic will diffuse through the agar from A to B. The concentration of the antibiotic will fall as it moves outward (toward B). Moreover, between X and B, antibiotic concentration is known as sub-inhibitory, where bacterial growth will occur. After the 24-hour incubation time period, the agar between A and X is very clear, showing a ring called the zone of inhibition, where bacterial is unable to grow around the paper disk. This occurs when the bacterial strain is resistant to the antibiotic. A larger zone of inhibition indicated that the bacterial strain was more resistant to the antibiotic in the disk. Finally, the concentration at "X" at the time the ring's edge is formed is called the critical inhibitory concentration (CIC).

Methodology

To begin, create 500 mL Nutrient Agar (NA) & 500 mL Mueller Hinton Agar (MHA) and refrigerate for 24 hours. Autoclave the NA and MHA beakers for 90 minutes at 121°C at 17 psi and pour the NA and MHA into petri dishes (100 mm x 15 mm). After 5 minutes of solidification, place the lids on the petri dishes and place in fridge. To create the silver nanoparticles (AgNPs), combine 200 mg of Folgers coffee powder with 20 mL distilled water. Mix 2 mL 0.1 N liquid silver nitrate (AgNO₃) with the 20 mL coffee/water solution on low heat. Obtain the bacterial strains (*Escherichia coli* (K-12), *Bacillus megatherium*, and *Staphylococcus epidermidis*) and antibiotic disks (ampicillin, imipenem, cefepime, ceftazidime, and bacitracin). Inoculate the *Bacillus megatherium* in vitro and place in incubator at 30°C for 24 hours. Next, inoculate the MHA (12 plates) with 3mL of reactivated *Bacillus megatherium* and spread on petri dish using a disposable spreader. Place one of each antibiotic disk on one petri dish (total of 5 disks). Repeat for another petri dish.

Dip 10 of each type of antibiotic in the AgNO₃ + coffee + water solution and soak on plate. Using forceps, place 5 of these disks on one petri dish and 5 on the other petri dish. Make sure to light the forceps with the bunsen burner each time a different antibiotic is soaked (sterilization protocol). Repeat this step for the other 4 antibiotics and incubate all the dishes at 30°C for 24 hours.

Collect data by measuring the zone of inhibition (ring formed around each disk). Compare this data to the Antimicrobial Usage Chart (Table 1) to determine which antibiotics were resistant, intermediate, and susceptible.

Repeat steps 2 – 4 for the two other bacterial strains. However, place the *Escherichia coli* (K-12) and *Staphylococcus epidermidis* in the incubator at 37°C, instead of the 30°C previously used. Also, use NA instead of MHA for these two bacterial strains.

Results & Analysis

Table 1 is used to determine which antibiotics from the study are resistant, intermediate, or susceptible. The categories are determined by the sizes of the zones of inhibition given above.

Table 1. Antimicrobial usage chart for five tested antibiotics.

Antimicrobial Agent	Disk Code	Potency	Minimum Inhibitory Concentration (MIC) (zone diameters in mm)		
			Resistant (R)	Intermediate (I)	Susceptible (S)
Ampicillin	AMP	10 µg	≥ 25	24 – 20	≤ 19
Cefepime	FEP	30 µg	≥ 18	17 – 15	≤ 14
Bacitracin	B	15 µg	≥ 13	12 – 9	≤ 8
Imipenem	IPM	10 µg	≥ 16	15 – 14	≤ 13
Ceftazidime	CAZ	30 µg	≥ 21	20 – 18	≤ 17

Table 2 gives the average zone of inhibition of each antibiotic with and without AgNPs. It also shows whether each antibiotic was resistant, intermediate, or susceptible to the bacterial strain. This was determined by the average range of zones in Table 1. From this, the effectiveness of the AgNPs can be seen.

Table 2. Combined and individual efficacy of five antibiotics and AgNPs against three multi-drug resistant bacteria

<i>Escherichia coli</i> (K-12)		<i>Bacillus megatherium</i>		<i>Staphylococcus epidermidis</i>	
Antibiotic	Average Zone of Inhibition (in mm)	Sensitivity Analysis	Antibiotic	Average Zone of Inhibition (in mm)	Sensitivity Analysis
Ampicillin	22.5	Intermediate	Ampicillin	22.5	Resistant
Ampicillin + AgNPs	25.6	Intermediate	Ampicillin + AgNPs	27	Resistant
Cefepime	20	Resistant	Cefepime	27.5	Resistant
Cefepime + AgNPs	26	Resistant	Cefepime + AgNPs	32	Resistant
Bacitracin	8	Susceptible	Bacitracin	24	Resistant
Bacitracin + AgNPs	10	Intermediate	Bacitracin + AgNPs	29.7	Resistant
Imipenem	29.5	Resistant	Imipenem	41	Resistant
Imipenem + AgNPs	30	Resistant	Imipenem + AgNPs	45	Resistant
Ceftazidime	20	Intermediate	Ceftazidime	31.5	Resistant
Ceftazidime + AgNPs	27.9	Resistant	Ceftazidime + AgNPs	43.5	Resistant
Ampicillin	14.5	Susceptible	Ampicillin	14.5	Susceptible
Ampicillin + AgNPs	20	Intermediate	Ampicillin + AgNPs	20	Intermediate
Cefepime	26.7	Resistant	Cefepime	26.7	Resistant
Cefepime + AgNPs	27.5	Resistant	Cefepime + AgNPs	27.5	Resistant
Bacitracin	15	Resistant	Bacitracin	15	Resistant
Bacitracin + AgNPs	17.5	Resistant	Bacitracin + AgNPs	17.5	Resistant
Imipenem	44.7	Resistant	Imipenem	44.7	Resistant
Imipenem + AgNPs	45	Resistant	Imipenem + AgNPs	45	Resistant
Ceftazidime	20.7	Intermediate	Ceftazidime	20.7	Intermediate
Ceftazidime + AgNPs	22.1	Resistant	Ceftazidime + AgNPs	22.1	Resistant

Each antibiotic is then graphed separately to visualize the effectiveness on the three bacterial strains. The lightest blue region is the area of resistance, followed by the darker blue as the intermediate values of resistance, and the darkest blue as the susceptible regions. Table 3 summarizes these results.

Table 3. Antibiotic Mapping Data

Antibiotic Map	Results
	Figure 2. Ampicillin, with the addition of AgNPs, was most effective (most resistant) against <i>B. megatherium</i> and least effective (most susceptible) to <i>S. epidermidis</i> . <i>E. coli</i> fall in between, with most of its zones of inhibition falling in the intermediate category, showing neither resistance nor susceptibility.
	Figure 3. Most data for imipenem and AgNPs could not be recorded because the zones of inhibition were too large for the petri dish (very resistant). Most diameters overlapped each other, so it was harder to measure. From the data taken, all the disks proved to be effective in fighting <i>B. megatherium</i> , <i>E. coli</i> , and <i>S. epidermidis</i> .
	Figure 4. On average, <i>B. megatherium</i> proved to be the most resistant to the cefepime in combination with AgNPs. For disk 5 of <i>E. coli</i> , the diameter was too large to measure from the petri dish. For disks 4-6 of <i>S. epidermidis</i> , the disks shifted on the agar, so the zone of inhibition did not form. However, of all successful trials, all bacterial strains were resistant to cefepime and AgNPs.
	Figure 5. Similarly, zones of inhibition of disks 4-10 of <i>B. megatherium</i> , disks 9-10 of <i>E. coli</i> , and disks 8 and 10 of <i>S. epidermidis</i> were not recorded because zones of inhibition were too large for petri dish. <i>B. megatherium</i> and <i>E. coli</i> were both consistent in that both were resistant to the antibiotic in every trial tested. <i>S. epidermidis</i> , however, had one susceptible case.
	Figure 6. In the end, <i>B. megatherium</i> and <i>S. epidermidis</i> was consistent in that it tested resistant for every disk. <i>B. megatherium</i> , however, produced larger zones of inhibition than <i>S. epidermidis</i> . On the other hand, <i>E. coli</i> proved to be mostly susceptible, with a few intermediate cases of inhibition. This shows that the bacitracin and Ag NPs worked best on <i>B. meg</i> and worked less on <i>E. coli</i> .

Critical Inhibitory Concentration

From Fick's second law of diffusion, diffusion across an agar medium can cause the concentration to change with respect to time in the equation

$$\frac{\partial c}{\partial t} = D \left(\frac{\partial^2 c}{\partial x^2} \right), \text{ where } C = \text{concentration in dimensions}$$

of [(amount of substance) length⁻³], t = time, D = diffusion coefficient in [length² x time⁻¹], and x = current position [length]. (<https://www.ncbi.nlm.nih.gov/pubmed/18339637>)

A derivation of Fick's second law of diffusion can be made where neither the diffusion coefficient nor the critical time data apply. This equation reads:

$$C_c = \frac{C_n}{\pi B h} e^{-\frac{A + B \ln(C_n)}{B}}$$

(<https://www.ncbi.nlm.nih.gov/pubmed/18339637>)

, where C_c = overall critical concentration (mg/mL), C_n = antibiotic critical concentration, B" = slope (m)", h = "agar depth" of 11.25 mm, and A = y-intercept.

Using this equation above, it is true for any positive values of C_n, so 1 can be used as the antibiotic critical concentration. The new equation now allows calculations of critical concentration values with the regression analysis of the zone of inhibition diameters. This equation reads:

$$C_c = \frac{1}{\pi B h} e^{-\frac{A}{B}}$$

Using Figure 7, the critical concentrations of each antibiotic can be calculated using the agar depth, slope, and y-intercept of each antibiotic. Taking the zone diameter and the square of the zone diameter, each antibiotic is plotted alone and in combination with AgNPs, which have a potency around 5 µg. The critical concentrations of each antibiotic are calculated with Fick's derivation equation and summarized in Table 3. These values are important because they allow assessment in the clinical practices, while simplifying the attempt to get a better correlation between the minimum inhibitory concentration (lowest concentration of an antimicrobial that will inhibit the growth of a microorganism) and the critical concentration.

Figure 7. Antimicrobial Critical Inhibitory Concentrations (CIC) Graph

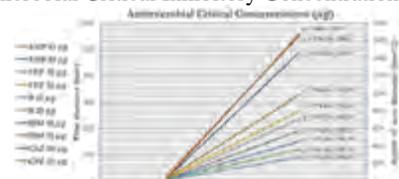


Figure 7. The slopes and y-intercepts of the antimicrobial agents with and without AgNPs are used to calculate the critical inhibitory concentration using Fick's second law derivation, as shown in Table 3.

Table 3. Calculation of Critical Inhibitory Concentration (CIC)

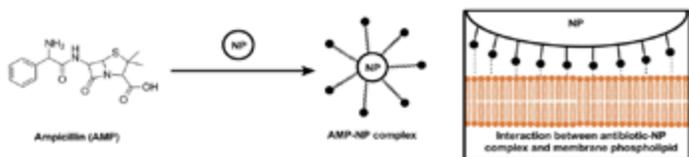
Antimicrobial Agent	Critical Inhibitory Concentration (CIC)	Antimicrobial Agent (with AgNPs)	Critical Inhibitory Concentration (CIC)
Ampicillin	.000167 mg/mL	Ampicillin	.000139 mg/mL
Cefepime	.000108 mg/mL	Cefepime	.0000951 mg/mL
Bacitracin	.000276 mg/mL	Bacitracin	.000221 mg/mL
Imipenemon	.0000521 mg/mL	Imipenemon	.0000478 mg/mL
Ceftazidime	.000133 mg/mL	Ceftazidime	.0000778 mg/mL

Table 3 compares the critical inhibitory concentrations of the antimicrobials with and without AgNPs, showing that the correlation of lowest concentrations is the amount of dosage needed to inhibit the growth of the bacterial strains. The antimicrobial agents with AgNPs showed lower CIC values than those without AgNPs, proving that AgNPs are effective in inhibiting the growth of the bacteria with a lower medicinal dosage.

Proposed Mechanism

A mechanism explaining how ampicillin is affected by the cells of the bacteria is shown below.

Figure 8. Mechanism of the AMP-NP complex and the phospholipid layer



A potential cause of the synergistic effect of ampicillin with AgNPs in vitro may happen as a result of the nanoparticles acting as specific drug carriers. The cells membrane consists of glycoproteins and phospholipids. Both of these are hydrophobic regions, regions that tend to repel water. On the other hand, the AgNPs (without the ampicillin) is likely to approach the cell membrane of the bacteria cells since the ampicillin molecules are attracted to water (hydrophilic regions). Therefore, the antimicrobial agents help to facilitate the transport of the ampicillin to the cell membrane through extracellular and intracellular biological processes.

Conclusion and Future Research

Antimicrobial resistance is currently a global problem. From the West Nile viruses to the Coronavirus pandemic, the emergence of pathogenic resistance has limited options for therapeutics. On a local scale, as well as a national and global one, monitoring the antimicrobial resistance can help to minimize the effects of the current pathogenic drug resistance problems what we are currently facing.

In comparison to all the antimicrobial agents tested on the bacterial strains, most antibiotics showed resistance to the bacterial strain. This proves that resistance to pathogenic bacterial strains can be tested easily with a novel, greener method of experimenting. The addition of nanoparticles into the agent increased the resistance to all three multi-drug resistant bacteria. As shown in Table 3, the Antimicrobial agents with AgNPs had a lower CIC than those without AgNPs. This shows that the AgNPs were, in fact, the determining factor in lowering the medicinal dosage (concentration) needed to inhibit the growth of the bacteria. With this data, zones of inhibition and CIC values can be extrapolated for other antibiotics within the same category.

In the future, data regarding whether the size of the AgNPs effects the zone of inhibition linearly or exponentially would be important for this study. Also, testing the effectiveness of different potencies of antibiotics on the bacteria and considering the mechanism involved in each AgNP – antibiotic conjugate will be useful.

References

- Awerbuch, Tamara E., and Levi Lustman. "A Mathematical Model for Determining Minimal Inhibitory Concentrations (MICs) via Diffusion Assays." *Journal of Theoretical Biology*, vol. 129, no. 2, 30 June 1987, pp. 219–230., doi:10.1016/s0022-5193(87)80014-0.
- Hwang, I. S., Hwang, J. H., Choi, H., Kim, K. J., & Lee, D. G. (2012). Synergistic effects between silver nanoparticles and antibiotics and the mechanisms involved. *Journal of medical microbiology*, 61(12), 1719-1726.
- Nadagouda, M. N., & Varma, R. S. (2008). Green synthesis of silver and palladium nanoparticles at room temperature using coffee and tea extract. *Green Chemistry*, 10(8), 859-862.



Muneebah Umar

World Food Prize: Combating rampant malnutrition in Pakistan

The South Asian nation of Pakistan's impoverished citizens have experienced some of the worst rates of food insecurity and malnutrition in the world. In the last decade, an increase of wasting, stunting, and other growth defects have occurred, especially among the most vulnerable portion of the population: women and children living in poverty (UNICEF, 2017). Having a portion of the population malnourished does not have any benefits and acts as an economic disadvantage for the Pakistani government. Physical handicaps that occur as a result of malnourishment, are considered inter-generational, meaning if a child's mother is suffering food insecurity, and therefore malnourishment, their child will most likely experience the same thing and the pattern will continue. The difficulty in accessing food is a major reason as to why so many families don't eat enough, and it will take decades before enough adequate infrastructure is built in order to reach the most remote communities. However, progress cannot be achieved at all until the main barrier to it is addressed: the strict long-held beliefs and misconceptions among the Pakistani population regarding nutrition and health. Not enough outreach to poor areas has been implemented by the government. Poor families tend to stay in the same community their parents and grandparents grew up in and hardly anyone, especially females, go elsewhere. There has long been a taboo associated with health relating to females in any way, and in turn, damaging practices are followed once a female reaches reproductive age. If a mother's diet during pregnancy, for example, is not given enough priority her fetus' size can be restricted, causing potential lifetime defects (Bhutta, 2016). A child's health is directly related to the health of their mother, but most of the urban and rural poor population is not aware of this because the knowledge their grandmothers had is all that they know for the most part. The best way to target this issue is by going to the most influential community members in these areas since the most isolated regions tend to be very tight knit. The improvement of women and children's health is a key in combating malnutrition and achieving a plethora of sustainable development goals, especially the third one, aimed at ensuring good health and wellbeing. If government workers were to educate and enlist the help of religious leaders, teachers, and high-ranking local politicians it would, no doubt, make people more willing to listen and change their dinner-table habits.

Works Cited:

- Bhutta, Z. A. (2016, December 2). SPECIAL REPORT: Malnutrition and stunting in Pakistan: Losing a generation: the impact of malnutrition. Retrieved from <https://www.dawn.com/news/1299917>
- UNICEF. (2017). Situation Analysis of Children in Pakistan [PDF File]. Retrieved from <https://www.unicef.org/pakistan/reports/situation-analysis-children-pakistan>



Dennis Lee

Mapping Agricultural Tillage Practices Using Kernel Extreme Learning Machine and Airborne Remote Sensing Images

Abstract

An efficient classifier based on extreme learning machine (ELM) is proposed to use for mapping agricultural tillage practices from airborne hyperspectral remote sensing imagery. The kernel version, called kernel ELM (KELM), is implemented due to its powerfulness. To utilize spatial information of an image, a spatial convolution filter is adopted to generate spatial-spectral features of a hyperspectral pixel by incorporating its surrounding pixels, which are the actual inputs to the KELM. The KELM can be adaptively modified when new training samples are added without completely re-training the model. Experimental results using airborne hyperspectral images demonstrate that the KELM can outperform other classic methods, such as support vector machine and random forest with lower computational cost, a larger window size such as 7×7 can be used for large homogeneous agricultural fields, and the adaptive KELM can further improve its efficiency.

Introduction

Contemporary agriculture faces enormous challenge to meet the growing food demand while simultaneously reduce its environmental harm [1]. In order to reach a more balanced agricultural practice, it is important to manage the existing land to be more productive [2]. This requires new approaches, including reforming conventional agriculture and adopting lessons from precision agriculture, where stable and high accuracy information on crop and soil characteristics need to be obtained to best utilize the agricultural resources [3].

Conservation tillage, including no-till, ridge-till, and mulch-till, leaves 30% or more crop residue cover after tillage and planting. Non-conservation tillage includes reduced-till (15%-30% crop residue remaining) and conventional-till (15% or less crop residue remaining) [4]. Conservational tillage can increase crop production while greatly reducing unsustainable uses of water, nutrients, and agricultural chemicals by reducing on-field losses. It has also gained increasing interest among farmers because of its time, labor, and fuel savings as well as high economic returns [5].

The current method to map crop tillage practices is mainly with field investigations, which is labor-costing, time-consuming, subjective, and difficult to generate widely distributed survey data [4]. Remote sensing technology, on the other hand, provides a more rapid, accurate, and objective solution. Although tillage practice mapping using hyperspectral imagery has been studied in the literature (e.g., [6]), the employed classification frameworks are often complicated with several pre-processing and feature extraction steps, which may not be realistic in practical applications requiring fast data analytics. As a result of the limitations of the state of the art, Extreme Learning Machine (ELM) [7], a simple machine learning algorithm, is introduced for mapping tillage practices. The ELM is highly efficient compared to other machine learning algorithms. The performance of the ELM can be further improved by using the kernel method (KELM) [8], which solves a linearly non-separable problem by mapping it to a very high dimensional space. Both ELM and KELM can be adaptively trained when new training samples are added without completely re-training the model. Such properties make them suitable to analyze large-scale remote sensing data when training samples being limited and gradually added.

The KELM model has been developed for mapping agricultural tillage practices by Lee in [9]. The paper aims to evaluate its performance with different spatial filtering size and computational cost and to develop and investigate the adaptive version of KELM.

Method

Extreme Learning Machine

ELM is a single-hidden layer feed forward neural network as illustrated in Fig. 1. The hidden layer is nonlinear due to the use of a nonlinear activation function $G(a,b,\mathbf{x})$ such as a sigmoid function with parameters a and b , but the

output layer is linear without an activation function. The weights \mathbf{W} between the input and hidden layer is a random matrix. Let \mathbf{x} represent a training sample, $f(\mathbf{x})$ be its neural network output, $\mathbf{B} = [\beta_1, \beta_2, \dots, \beta_m]$ of size $k \times m$ be the weighting matrix between the hidden and output layers, k be the number of hidden neurons, and m be the number of output neurons. The output in Fig. 1 can be written as:

$$f(\mathbf{x}) = \mathbf{B}^T \bullet G(a, b, \mathbf{x}) \quad (1)$$

For an ELM with n training samples and d input neurons (i.e., the number of features), Eq. (1) becomes

$$\mathbf{t}_i = \mathbf{B}^T \bullet g(\langle \mathbf{w}_j, \mathbf{x}_i \rangle) \quad , \quad i = 1, 2, \dots, n. \quad (2)$$

where \mathbf{t}_i is the m -dimensional desired output vector for the i -th training sample \mathbf{x}_i , the d -dimensional \mathbf{w} represents the j -th weight vector from the input layer to the j -th hidden neuron. Here, $\langle \mathbf{w}_j, \mathbf{x}_i \rangle$ denotes the inner product of \mathbf{w}_j and \mathbf{x}_i . The sigmoid function g is an activation function. For the j -th hidden neuron, $g(\langle \mathbf{w}_j, \mathbf{x}_i \rangle) = a / (1 + \exp(-\mathbf{w}_j^T \mathbf{x}_i / b))$ is its output, where b is the steepness parameter, and $a = 1$ here.

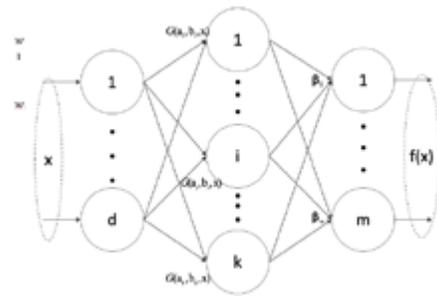


Fig. 1. A feedforward network structure with a nonlinear hidden layer and a linear output layer. Using a complete matrix notation, the compact form of model (2) can be rearranged as

$$\mathbf{H}\mathbf{B} = \mathbf{T} \quad (3)$$

where $\mathbf{h}_i = g(\mathbf{W}\mathbf{x}_i)$, $\mathbf{T} \in R^{n \times m}$, $\mathbf{B} \in R^{k \times m}$, and \mathbf{H} denotes the hidden layer output matrix. For a given \mathbf{W} which is randomly chosen, \mathbf{B} can be estimated by a least squares solution:

$$\hat{\mathbf{B}} = \mathbf{H}^+ \mathbf{T} = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T)^{-1} \mathbf{T} \quad (4)$$

where \mathbf{H}^+ denotes the pseudoinverse matrix of \mathbf{H} . To enhance stability, a regularization parameter δ is added as

$$\hat{\mathbf{B}} = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \delta \mathbf{I})^{-1} \mathbf{T} \quad (5)$$

For testing, the learned weights are directly applied as $\mathbf{y} = \hat{\mathbf{h}} \hat{\mathbf{B}} = \hat{\mathbf{h}} \mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \delta \mathbf{I})^{-1} \mathbf{T}$, where $\hat{\mathbf{h}}$ is the hidden neuron output of a testing sample. For each testing pixel, its class label will be c if the c -th output neuron yields the maximum output.

Kernel Extreme Learning Machine

To improve the performance, the kernel method can be applied to map the original data into a very high dimension where classes become linearly separable. The ELM can be simply extended to kernel ELM (KELM) by using the kernel trick [5] as:

$$y = \mathbf{K}^T (\mathbf{K} + \delta \mathbf{I})^{-1} \mathbf{T} = \begin{bmatrix} k(\mathbf{h}_1, \mathbf{h}_1) \\ k(\mathbf{h}_1, \mathbf{h}_2) \\ \vdots \\ k(\mathbf{h}_n, \mathbf{h}_2) \end{bmatrix}^T \begin{bmatrix} k(\mathbf{h}_1, \mathbf{h}_1) + \delta & \dots & k(\mathbf{h}_1, \mathbf{h}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbf{h}_n, \mathbf{h}_1) & \dots & k(\mathbf{h}_n, \mathbf{h}_n) + \delta \end{bmatrix}^{-1} \mathbf{T} \quad (6)$$

where $k(\cdot, \cdot)$ is a kernel function, e.g., radial basis function (RBF) (or the Gaussian distance) including a width parameter to be tuned.

Spatial Information Utilization

Spatial information is very useful when analyzing hyperspectral imagery, especially for agricultural images with large homogeneous area, due to high spatial correlation [10]. To utilize spatial information in an image, a simple averaging filter can be applied to each pixel. The resulting center pixel replaced the original one for classification. Such a spatial convolution process can smooth out trivial spectral and spatial variations. To make it parameter-free, the window size can be fixed as 3×3 [9]. It is assumed that pixels in this small neighborhood belong to the same class. For homogeneous areas, a larger window size, say, 5×5 or 7×7 , can be used for better smoothing effect, which will be demonstrated in the experiments.

Adaptive Learning Capability

If new training samples are added to the KELM, the trained model can be slightly updated without completely re-training the model. Let $\tilde{\mathbf{K}} = (\mathbf{K} + \delta \mathbf{I})$ in Eq. (6). If the number of training samples is increased from n to n' , then

$$\tilde{\mathbf{K}}_{n' \times n'}^{-1} = \begin{bmatrix} \tilde{\mathbf{K}}_{n \times n} & \mathbf{K}_{n \times (n'-n)} \\ \mathbf{K}_{n \times (n'-n)}^T & \tilde{\mathbf{K}}_{(n'-n) \times (n'-n)} \end{bmatrix}^{-1} \quad (7)$$

According to the theorem on the inverse of a partitioned matrix [11],

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1} & -\mathbf{A}^{-1} \mathbf{B} (\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{C} \mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} \end{bmatrix} \quad (8)$$

Note that $\mathbf{A}^{-1} = \tilde{\mathbf{K}}_{n \times n}^{-1}$ has been computed in an early stage when n training samples being available.

With newly received $(n'-n)$ samples, only the inverse

$(\mathbf{D} - \mathbf{C} \mathbf{A}^{-1} \mathbf{B})^{-1} = (\tilde{\mathbf{K}}_{(n'-n) \times (n'-n)} - \mathbf{K}_{n \times (n'-n)}^T \tilde{\mathbf{K}}_{n \times n}^{-1} \mathbf{K}_{n \times (n'-n)})^{-1}$ needs to be calculated, whose cost is $o(n'-n)^3$ much less than $o(n')^3$ if re-computing the entire $\tilde{\mathbf{K}}_{n' \times n'}^{-1}$ using all the n' samples.

Experiments

Datasets

Two datasets used in the experiments were acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) [12].

The first dataset used in the experiments was acquired over northwest Indiana's Indian Pines test site in June 1992. The image includes 145×145 pixels in the $0.4 - 2.45 \mu\text{m}$ spectral region with a spatial resolution of 20 m. A total number of 200 bands are used after removal of water-absorption bands. Two classification problems are present: corn-no-till versus corn-min-till, soybean-no-till versus soybean-min-till.

The second dataset captured an area over Salinas Valley, California, with a spatial resolution of 3.7m. Each image contains 6×217 pixels with 204 bands after 20 water absorption bands are removed. Three classes are involved in the classification: fallow, fallow-rough-plow, and fallow-smooth.

Development of Models and Experimental Design

The performance of KELM in mapping tillage practice is evaluated by assessing it against SVM and RF.

Parameters in all the three methods are tuned according to five-fold cross validation using training samples. For the KELM, two parameters need to be tuned (b and δ) plus an additional parameter in the RBF kernel, and the number of hidden neurons is empirically fixed as 2500.

In the Indian Pines experiments, five training set sizes are considered: 10%, 20%, 30%, 40%, and 50%; after training samples are randomly selected, the rest are used for testing. For each case, 20 runs are operated, and mean accuracy and standard deviation are recorded. In the Salinas experiment, 1%, 2%, 3%, 4%, and 5% of labeled samples are randomly selected for training, and the rest for testing. Similarly, 20 runs are operated for each case. Different window sizes are used to investigate the effect of spatial information processing on the accuracy of classification. The experiments are conducted using a personal computer with Intel 2.6 GHz CPU and 16 GB

RAM and MATLAB_R2016a.

Evaluation

Two metrics were used to quantify the classification accuracy of the models. Overall accuracy (OA) is calculated by dividing the sum of correctly classified field by the total number of fields examined. Kappa value is calculated as:

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (9)$$

where p_o or "observed" is the percent correct and p_e or "expected" is an estimate of the chance agreement to the "observed." A Kappa value of +1 indicates perfect accuracy of the classification. Models with a Kappa value of 0.4 or more are considered good.

To assess the time saving in using adaptive learning, the time saving ratio η is calculated as

$$\eta = 1 - \frac{T_a}{T} \quad (10)$$

where T_a is the time required for updating the weights using newly available samples, and T is time consumed by completely retraining the model with all the samples.

Results and Analysis

For the classification of corn-no-till versus corn-min-till in the Indian Pines experiments, KELM outperformed SVM and the RF, and its advantage is more obvious when using the averaging-filtered data. The spatial-averaging filter can improve the OA by 3%-6%, and the improvement is more significant when the training size is small, which is illustrated by Fig. 2. The map produced by the KELM is the closest to the ground truth.

For the classification of soybean-no-till versus soybean-min-till in the Indian Pines experiments, the KELM still offers the accuracy higher than the SVM and the RF. Using the averaging-filtered data, the OA can be increased by around 4% as shown in Fig. 3.

For the three-class classification problem in the Salinas experiment, the spatial convolution can enhance the performance of the KELM by 1%, and using training samples more than 1% does not obviously improve the accuracy, as shown in Fig. 4. The map produced by the KELM is the most similar to the ground truth.

To investigate the impact of the size of window used in the spatial averaging filter, the change of OA with 30% training samples in the Indian Pines experiments and 3% training samples in the Salinas experiment are studied. Using a larger window can increase classification accuracy and then the accuracy may slightly decrease. The improvement is obvious in the Indian Pines Soybean experiment. Fig. 5 shows the detailed changes with the four window sizes and five training set sizes for the Soybean experiments. Considering using a larger window size also increases computational cost, the best choice to achieve the balance may be 7×7 .

For the computing time of the three experiments, using more training samples is more time consuming. The training time of the KELM is less than that of SVM and RF. The RF training is the most costly. The testing time of the three methods are basically similar. Note that the step of spatial filtering is not considered here, as its cost is the same to all the three classifiers.

To validate the adaptive learning capability of KELM, the model is initially trained by 10% label samples in the Indian Pines (IP) experiments, and then another 10% samples are added by updating the weights according to Eqs. (7-8). Such adaptive learning time T_a is compared with the time T using all 20% label samples together. Similarly, with 30% training samples in total, the model is adaptively updated based on the one from 20% training samples and the newly received additional 10% training samples. In the Salinas experiment, 1% of training samples are added in each step. The resulting values of saving ratio η are plotted in Fig. 6. Obviously, more time can be saved when the number of training samples becomes larger. It is noteworthy that the adaptive KELM offers identical results as its non-adaptive version.

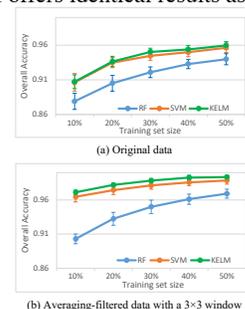


Fig. 2. Overall classification accuracy of Indian Pines data (corn-no-till vs. corn-min-till) [10].

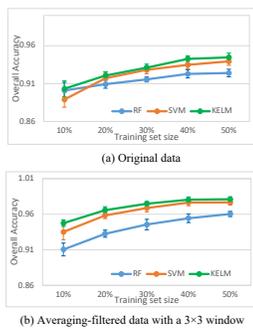


Fig. 3. Overall classification accuracy of Indian Pines (soybean-no-till vs. soybean-min-till) [10].

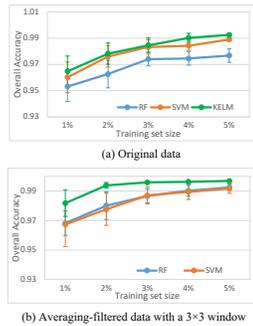


Fig. 4 Overall classification accuracy of Salinas data [10].

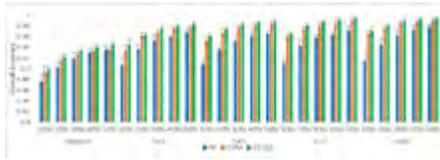


Fig. 5 Impact of window size of spatial filter in the Indian Pines Soybean experiment.

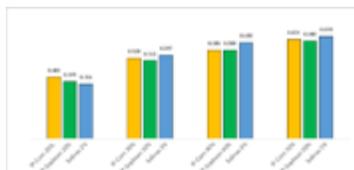


Fig. 6 Time saving ratio achieved by adaptive KELM training.

Conclusion

Conservation tillage management has been advocated for the purpose of soil preservation and sustainable crop production. The KELM can outperform the classic methods, such as the SVM and the RF, in agricultural tillage practice mapping. It offers fast training and testing with easy implementation. A very simple and parameter-free spatial averaging filter can greatly improve classification accuracy of all the classifiers, and the enhancement to the KELM is very significant. The adaptive KELM can further save more than 50% training time, providing a timely and cost-effective way to map tillage practice.

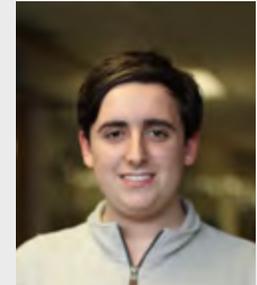
References

- [1] J. A. Foley, N. Ramankutty, K. A. Brauman, et al., "Solutions for a cultivated planet," *Nature*, vol. 478, pp. 337-342, 2011.
- [2] A. Bannari, K. Staenz, and K. S. Khurshid, "Remote sensing of crop residue using hyperion (EO-1) data," in *Proceedings of the International Geoscience and Remote Sensing Symposium, Barcelona, Spain*, pp. 2795-2799, 2007.
- [3] M. S. Galloza, M. M. Crawford, and G. C. Heathman, "Crop residue modeling and mapping using Landsat, ALI, Hyperion and Airborne remote sensing data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 446-456, 2013.
- [4] E. Choe, S. Hong, and Y. Kim, "Down-scaling of satellite hyperspectral images for monitoring croplands," in *Proceedings of the International Geoscience and Remote Sensing Symposium, Honolulu, Hawaii*, pp. 3873-3874, 2010.
- [5] R. Derpsch, T. Friedrich, A. Kassam, and H.-W. Li, "Current status of adoption of no-till farming in the world and some of its main benefits," *International Journal of Agricultural and Biological Engineering*, vol. 3, no. 1, pp. 1-25, 2010.
- [6] Q. Ran, W. Li, Q. Du, and C. Yang, "Hyperspectral image classification for mapping agricultural tillage practices," *Journal of Applied Remote Sensing*, vol. 9, 2015.
- [7] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [8] W.-Y. Deng, Y.-S. Ong, and Q.-H. Zheng, "A fast reduced kernel extreme learning machine," *Neural Networks*, vol. 76, pp. 29-38, 2016.

- [9] D. Lee, "Mapping agriculture tillage practices using extreme learning machine," *Proceedings of the 8th International Conference on Agro-Geoinformatics*, 2019.
- [10] M. Fauvel, J. Chanussot, J. A. Benediktsson, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804-3814, 2008.
- [11] H. Anton, *Elementary Linear Algebra*, 11th edition, Wiley, 2013.
- [12] http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes



Hong "Reggie" Zheng



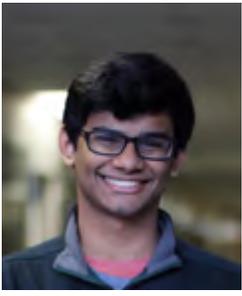
Colin Gordy

Solar Water Disinfection and The Advanced Oxidation Process: Effects of Organic Matter, pH, and Ions on the Disinfection of *E. Coli*, and the Environmental Impact

Solar Water Disinfection (SODIS) is an inexpensive and effective process used to sterilize water containing pathogens. The process utilizes plastic bottles as well as sunlight to damage the cell walls of pathogens. The Advanced Oxidation Process (AOP) enhances SODIS with photo-catalysts, such as titanium dioxide, to increase pathogen disinfection. SODIS-AOP can be used in emergency events in which clean water is not readily accessible. In this project, the effect of organic matter, pH, and inorganic ions on the disinfection of *E. Coli* in the SODIS-AOP process will be tested. The significance of plastic contamination in treated water will also be tested. To test the effects of the aforementioned variables, they were placed in titanium-dioxide-coated plastic bottles containing *E. Coli* and exposed to UV light in a sunlight simulator for 6 hours. Then, samples from the bottles were collected for data. To test plastic contamination, bottles without TiO₂ coating were reused multiple times, and the amount of plastics was measured through distillation. In general, SODIS-AOP proved to effectively decrease the amount of *E. Coli* in contaminated water, and no significant amount of plastics contaminate the treated water after repeated use. The presence of organic matter did not affect deactivation of *E. coli*, decreased pH and inorganic ions increased deactivation of *E. coli*, and hydrogen peroxide increased disinfection the most. However, more research is needed to accurately determine the efficiency of using SODIS-AOP in real-life emergency situations.

Work Cited

- Diatoms reveal freshwater pollution. (2010, May 6). Retrieved from <https://www.sciencedaily.com/releases/2010/05/100504095111.htm>
- McGuigan, K. G., Conroy, R. M., Mosler, H.-J., Preez, M. du, Ubomba-Jaswa, E., & Fernandez-Ibanez, P. (2012, August 6). Solar water disinfection (SODIS): A review from bench-top to roof-top. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0304389412007960>
- Schmid, P., Kohler, M., Meierhofer, R., Luzi, S., & Wegelin, M. (2008, October 1). Does the reuse of PET bottles during solar water disinfection pose a health risk due to the migration of plasticisers and other chemicals into the water? Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0043135408004168>
- Water Treatment Solutions. (n.d.). Retrieved from <https://www.lentech.com/processes/disinfection/chemical/disinfectants-hydrogen-peroxide.htm>



Vayd Ramkumar



Jackson Flowers

Assimilation of Global Pollution Data: A Novel Spatiotemporal Model for Widespread Classification of Atmospheric Particulate Clusters (Year 2)

Abstract

Atmospheric pollution has grown increasingly salient in decades past, and its malignant effects are projected to strictly grow, notwithstanding immediate climate reform. Thus, preemptive classification of possible pollutant hotspots is necessary in supplementing legislative action. In this study, we used multifaceted pollutant data from OpenAQ to interpolate data through a Moore-Penrose pseudoinverse (generated by previous work). We defined weights for each class of pollutant based on recommendations from the WHO and determined the malignance of particulate clusters using Von Neumann-Morgenstern (VNM) utility formalism. We classified clusters by the product of average pollution concentration and area density of data. All appropriately classified clusters were assigned a categorical datum based on the value of its associated VNM, which divulged policy recommendations to local governments (supplied by the WHO). A new dataset was used for every 6 hours in the timeframe, which was from May 2015 to April 2019, and we defined the vector field of pollution as $V = V(i+1) - V(i)$ for time i , from which we determined the flux by integrating across our domain D (all latitudes and longitudes that wrap North America in a rectangular lattice), the discrete values of $\text{div}(V)dAdz$. We assumed continuity because all $dx dy = dA < 0.00001$ units, which is roughly 11 meters, insignificant relative to $\text{dim}(D)$. After application of the model, the relative error in interpolation of data was 0.3138%, applied over 4 years, which we used to predict data from May 2024 to a 4.234% error, split over 10 different VNM recommendations and 9 pollutants.

Background and Introduction:

Anthropologically-generated atmospheric pollution has grown increasingly salient in the last few decades, and currently, with inadequate environmental standards and climate reform, its malignant effects are only projected to grow. Indeed, the WHO projects that each year, atmospheric pollution can be attributed to over seven million premature deaths worldwide [1]. Most forms of pollution tend to fall short of 10-6 meters per cluster (in radius), implying that they easily infiltrate human extremities (i.e. bloodstream, alveolar ducts) and carry the potential to inflict serious damage (surface area/volume tends to be around $3.0 * 10^6$). With such high potential for risk, the ability to model pollution dispersion and act accordingly in the wake of such models is paramount. In this study, we seek a two-faceted approach to combating rising pollution levels: 1) the prediction of future potential hotspots of high pollution concentration, and classification of their potential harmful effects; 2) the development of an automatic response system that aids local governments, giving them recommendations on how to curb the potential side effects of increased pollution. We seek to develop a model that bridges the gap between activism and research, so that future activism and future government policy may be directly based on scientific inquiry.

In this project, we form a continuation from previous research, invoking research from 2019 [2] from which our chief continuity is in fine particulate matter data – which is set to encompass roughly 60% of our total data. We generated this data using a novel approach, a dynamic least-squares projection, supplemented by dynamic wind patterns. We make the preliminary justification for the continued use of this data based on the fact that it is projected to be at most 3.105% off (an upper bound determined by power series), which falls under our arbitrary (yet generally accepted) tolerance of data within an upper bound of 5% off from accepted measurements. Thus, in application and in computation, we simply must propagate error from research in previous years, which is enough to draw causal links.

Hypotheses:

In this study, we test two different hypotheses. We do not include the application of automated responses in pollution, since this is categorically proven to be possible: it is a simple assignment based on a hypothetical scalar value, yet it still is novel.

1. Von Neumann-Morgenstern Utility theory may be applied to different kind

of atmospheric pollutants.

2. A higher VNM utility (which is linked to lower happiness) can be correlated with similarly high human development index.

We construct these hypotheses so that our final result (the recommendations to local governments) is based on as accurate predictions as possible. Ideally, we are able to make predictions for data in upcoming years so that legislation may be appropriately constructed to reflect the future status of pollution.

Mathematical Model Development:

To begin, we define our method of temporal interpolation. We let V_i be the spatial distribution of data at any time i , and i increases incrementally over the natural numbers. The net step over any time is:

$$\vec{v} = V_{i+1} - V_i$$

We continue by calculating the influx of pollution, which we assume to be strictly positive for $t \in \mathbb{R}^+$. Let D be the relevant domain over which the pollutants exist. The net influx, a scalar I , can be:

$$I = k \cdot \int_0^\lambda \iint_D \nabla \cdot \vec{v} dAdz$$

In which k is some constant, and λ is the concentration of the nearest interpolated cluster.

Additionally, we choose to associate each point (over all observed times) with a utility, which is formally assigned by Von Neumann-Morgenstern Utility Theory [3]. Let a “node” be the conditions of an arbitrary point at an arbitrary time. We define a lottery $L(x,y,t)$ to be the node, and the possible outcomes of that node are guaranteed to exist by VNM. That is, we have that:

$$\mathcal{L} := \{L(x, y, t) \mid (x, y) \in C \subseteq \mathbb{R}^2, t \in \mathbb{R}^+\}$$

And we are guaranteed the existence of some function $\mu: L^*(x,y,t) \rightarrow \mathbb{R}^+$ such that all possible lotteries may be assigned a utility. Note that C is the range of possible coordinates in which nodes exist.

We now define a metric for induced stochasticity in both determination of the above lottery and nodes. Let $F(x(t),y(t),z(t))$ be a vector field. We say that the induced stochasticity of that vector field, another vector $S(t)$, can be defined as the curvature of an optimized path on F . Let this path be $r(t)$. We have that:

$$\vec{r}'(t) = \nabla F^i, \vec{S}(t) = s_k \frac{\vec{r}'(t) \times \vec{r}''(t)}{\|\vec{r}'(t)\|^3}$$

In which s_k is an arbitrary constant $\in \mathbb{R}^+$. In our calculations, we randomly assign a tolerance of $r(t)$ to all nodes in order to simulate general stochastic effects.

Finally, we formalize the notion of how we generalize wind: through the use of Lorenz attractors. We assume that atmospheric convection may be analogous to the flow of particles as per the solution to ODE's [4]:

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z. \end{aligned}$$

In which x is proportional to the rate of convection (the magnitude of the previously-defined vector v), y as the gradient of the temperature in the horizontal direction, and z the gradient of the temperature in the vertical direction. Note that this usage of Lorenz attractors is distinct from previous usage for the vector v , because we use the method of finite differences to calculate the magnitude of the convective force, which in turn is used for the vector.

We now have a four-faceted mathematical model, and our objective function is a lottery, which is influenced by the intensity and distribution of nodes (which are determined by the convective force due to wind and induced stochasticity). We observe that for some force vector H , in which the vector component form of H are the specific components above (i.e. $H = \langle x, y, z \rangle$ and $|H| = (x^2+y^2+z^2)^{1/2}$). We claim that the overall movement of some node with associated utility I can be generalized as MHH , for MH an arbitrary constant (in positive reals) to be determined through numerical analysis.

From this, we are ready to begin numerical analysis, which is less taxing given the data from Year 1 of this project: all our data is expected to be quite similar, yet we can still project a novel conclusion because of our use of Lorenz attractors to preclude the need for constant sources of wind data. Our proposed mathematical model is more autonomous and self-reliant than any other available.

Numerical Analysis:



Fig. 1 (left to right): simulated data over Boulder, CO, an example of a Lorenz attractor acting as wind over Oxford, MS (in a temporary convergence zone), and data over Starkville-Tupelo-Oxford, MS.

Table 1: Derived Growth Rates, Utility Values, and Lottery Probabilities for Pollutants

Pollutant	Rate of Increase	Utility (Relative)	Lottery Probability
PM _{2.5}	$6.027 \times 10^{-12} \text{ m day}^{-1}$	1.0000	0.5872
PM ₁₀	$1.572 \times 10^{-12} \text{ m day}^{-1}$	0.4971	0.2064
SO ₂	$1.403 \times 10^{-12} \text{ m day}^{-1}$	2.0503	0.0053
CO	$9.844 \times 10^{-13} \text{ m day}^{-1}$	0.9833	0.0200
NO _x	$2.396 \times 10^{-12} \text{ m day}^{-1}$	3.4724	0.0041
Heavy-Metal Complex	$4.151 \times 10^{-13} \text{ m day}^{-1}$	5.3503	0.0023
O ₃ (Tropospheric)	$8.108 \times 10^{-13} \text{ m day}^{-1}$	0.1834	0.0173
CH ₄	$6.013 \times 10^{-12} \text{ m day}^{-1}$	0.2713	0.1057
CFC's	$7.069 \times 10^{-12} \text{ m day}^{-1}$	4.8949	0.0517

We note that the notion of relative utility is based on how bad a pollutant is

(isolated) relative to PM2.5. Reactions between pollutants are deemed to be negligible (on the order of 10-3 percent, relative to the relative utilities).



Fig. 2: A snapshot of what collected wind data looks like, on February 24, 2020 across the United States. Note the presence of a strong Lorenz attractor in Boulder, CO.



Fig. 3: An example of collected data from February 3, 2020 at 06:00 A.M. (roughly 130,000 datapoints across the United States and Canada). This is a wider view of the depiction in in Fig. 1, rightmost. Data updates once every six hours.

We now have that each moment in time is associated with a utility map, and we can apply the method of Lorenz attractors to each moment of time, using just one dataset (instead of appending multiple for each day in question). Additionally, we know that our modeling technique that was developed in Year 1 was accurate, because of its low error and arbitrary domain. Thus, we do not re-formalize the notion of a least-squares projection over any arbitrary domain, and we can apply it without restrictions. For the pure sake of simplicity, we only consider the United States and Canada in our physical model application.

Table 2: Numerically-Determined Values of Constant and Relevant Parameters

Parameter	Description	Value and Units
k	Constant applied to net influx of all kinds of pollution	1.0000
s_k	Relative emphasis of induced stochasticity's role in model	0.4971
M_H	Movement of nodes due to wind and the convective force	2.0503
dt_x	Time step in the spatial (pollution data)	0.250 day(s)
dt_w	Time step in the wind data	0.0001 day(s)
ϕ	Constant applied to the cluster parameter	0.0001 day(s)

Cluster Classification and Recommendations:

We assign recommendations to local governments based on the calculated utility parameter for each cluster of data. To do this, we must formally define the notion of a cluster of pollutants. We define this on a parameter dependent on discrete regions, the average pollution concentration, and their relative abundance. That is, we have a cluster parameter:

$$C_{ij} = \phi n_{ij} \hat{\lambda}_{ij}$$

In which i and j are latitude/longitude in a matrix, ϕ is a constant, n_{ij} is the

number of datapoints in $x: (i, i+1)$, $y: (j, j+1)$, and hatted λ is the average of all relevant datapoints in n_{ij} . We say that each federal jurisdiction ought to have at least one cluster, and if not, we assign it one to ensure that at least some recommendation is given. Based on the value of this cluster parameter C , we categorically assign recommendations that are loosely based on WHO and other international environmental agencies.

Table 3: Recommendations to Local/Municipal Governments (District/Riding) [5,6,7]		
ID	C Range	Recommendation Description
1	[0, 1)	Maintain status quo; no climate/legislative action necessary, yet still encouraged
2	[1, 3)	Implement 1 AND use of information campaigns to raise awareness of possible household pollutant sources: encourage energy conservation
3	[3, 6)	Implement 2 AND promote locally-subsidized goals with no incentives: encourage recycling goals, ensure that core school curriculum iterates the importance of it
4	[6, 10)	Implement 3 AND pass legislation that ensures recycling bins must accompany all public waste bins, increase fine for littering, undue dumping of waste publicly
5	[10, 15)	Implement 4 AND increase tax rebates for purchase of electric/low-emissions vehicles, constrain harmful environmental practices such as fracking, drilling
6	[15, 20)	Implement 5 AND subsidize green technology, incentivizing residents to make use via local tax rebate for increased use of renewables
7	[20, 30)	Implement 6 AND impose a locally-mandated recycling and carbon footprint initiative, which progressively taxes increased use of carbon-producing products
8	[30, 45)	Implement 7 AND form a public initiative/local agency for the express creation of local green tech infrastructure/solar panels/wind turbines
9	[45, 65)	Implement 8 AND increase corporate taxes on private entities that sell pollutant-producing products (such as exhausts, motorbikes, etc.)
10	[65, ∞)	Implement 9 AND declare a state of local emergency so that state/government funds can be diverted to assist curbing of pollutant metrics, distribute face masks and treat the situation as a public health crisis

Now we have all necessary components to augment an existing mathematical model with more data and a categorical response metric. We demonstrate its usefulness by the construction of a generalized prediction for the United States and Canada.

In order to demonstrate the impacts of our model, we develop two separate forecasts: one if local governments take into consideration our recommendations as of the beginning of FY 2021; one if they maintain the status quo. Since data is fairly sparse in certain areas, the most specific that we may get parts of North America is limited to Congressional Districts/ridings.

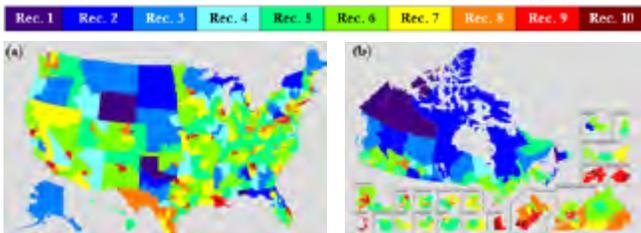


Fig. 4 (a): Data-based prediction for atmospheric pollution concentration in the United States on May 01, 2024 (b) in Canada, assuming the initial recommendations given by this project are not realized.

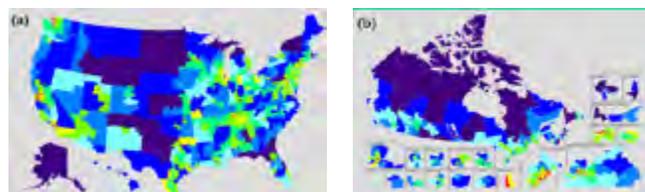


Fig. 5 (a): Data-based prediction for atmospheric pollution concentration in the United States (b) in Canada, assuming the initial recommendations given

by this project are considered.

We mathematically determine the impact of our recommendations by cross-referencing them to historical precedent and fit each recommendation to an exponential decay curve, such that the effectiveness of a specific recommendation p is $(1 - ae^{-pt})$, in which a is a constant and t is the time (in years). Figures 4 and 5, which reflect two possible different states of data on May 01, 2024, clearly show that our proposed recommendations are not only effective but are necessary to limit the inevitable future spread of pollution.

We bound the possible error on our prediction from above by 4.243% (determined through statistical analysis and propagation of error in data, and we test the second hypothesis per a correlation matrix [8], revealing that the correlation between HDI and our derived VNM parameter is 0.9961. From evaluating the HDI vs. VNM and the sources of pollution, we now have a self-updating model that can be generalized to the world in order to classify clusters/give adequate recommendations to local governments, a method by which pollution can be combated, and to a degree of accuracy that is unrivaled by any other relevant project or research endeavor to date. By giving people all around the world methods to prepare for and combat pollution 4 years in advance, we have the potential to save millions of lives.

Conclusion:

We can affirm both of our hypotheses. First, we empirically show that VNM utility is a useful means by which pollution can be first quantified and then categorized, which comes in the form of our maximum error in the model of 4.234%. A correlation matrix affirms the second. Thus, as we expected, we are able to create and generalize a multifaceted and multipurpose model for pollution in a novel way that exceeds most commonly-employed models. In the future, we intend on vastly expanding the scope of the model, so that instead of introducing the notion of generality, we formally define predictions across the globe. There are many hundreds of terabytes of data available, which implies that computational speed is a salient problem in our model: we hope to optimize speed either by fitting data to a neural network (so that we may over time fit data to expected predictions and remove the stochasticity section entirely). Additionally, by the refinement of the data collection process, we may have a more specific domain (perhaps even to the level of precinct) by which we may make recommendations.

Finally, the goal of this research is to help realize the sustainable development goals put forth by the United Nations [10]. Ultimately, this project seeks to help fulfill goals 11, 13, and 15 (Sustainable Cities and Communities, Climate Action, and Life on Land) by promoting recommendations which allow for streamlined activism, and ultimately benefit humans by alerting them of information about the atmosphere.

References:

- [1] Air pollution. (2019, September 24). Retrieved from <http://www9.who.int/airpollution/en/>
- [2] Abstract Search. (n.d.). Retrieved from <https://abstracts.societyforscience.org/Home/FullAbstract?State&ProjectId=18206>
- [3] Neumann, John von and Morgenstern, Oskar, Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press, 1953.
- [4] Hirsch, Smale & Devaney (2003), pp. 303–305
- [5] Recommendations. (n.d.). Retrieved from <https://www.ewg.org/research/particulate-air-pollution/recommendations>
- [6] Actions You Can Take to Reduce Air Pollution | Ground-level Ozone | New England | US EPA. (2018, September 27). Retrieved from <https://www3.epa.gov/region1/airquality/reducepollution.html>
- [7] Guidelines. (2017, November 7). Retrieved from <https://www.who.int/airpollution/guidelines/en/>
- [8] Human Development Reports. (n.d.). Retrieved from <http://hdr.undp.org/en/data>
- [9] OpenAQ. (n.d.). Retrieved from https://openaq.org/#/?_k=meq74v
- [10] Sustainable Development Goals. World Bank. (n.d.). Retrieved from <http://pubdocs.worldbank.org/en/787881551982397673/Accessing-the-Online-Application-2019-for-developing-country-nationals-FINAL.pdf>



Skylar Nguyen

Comparative Study between the Antimicrobial Resistance of Lichen from Mississippi and Alabama against *Staphylococcus epidermidis* and *Escherichia coli*

Abstract

Nosocomial infections are healthcare-acquired infections that can appear within 48 hours of being admitted to the hospital or within 30 days of a patient's discharge (Niger, 2012). The infections are unrelated to the illness that brought the patient into the hospital. The Center for Disease Control (CDC) notes that there are millions of people annually who acquire a nosocomial infection. Research of the antibacterial properties within lichen gives way to its potential in the commercial biopharmaceutical industry. This study compares the antimicrobial activity of lichen species *Flavoparmelia caperata* and *Ramalina farinacea* from the Gulf Coast and Black Prairie region of Mississippi and the Southeastern Gulf Coastal Plain region of Alabama and notes the possible environmental factors that contribute to their differences in antimicrobial activity against *Staphylococcus epidermidis* and *Escherichia coli* which are both bacteria linked to numerous nosocomial infections. Deduke et al. (2012) explains how the increased bioactivity of the same lichen species from different locations can be originated from environmental factors such as sun exposure, humidity, mineral deposits, and air quality which could affect the lichen's secondary metabolites. The lichen from the Gulf Coast region of Mississippi exhibited the highest antimicrobial existence against *S.epidermidis* and *E.coli*. The sun exposure and air quality are not sufficiently different between both coastal regions. The results of this study leads further research to identify the other factors within the regions that could also be affecting the lichen's antimicrobial resistance.

Introduction

In the United States, over 1.7 million people acquire nosocomial infections, and 99,000 pass away annually. Nosocomial infections are healthcare acquired infections (HAIs) that can appear within 48 hours of being admitted to the hospital or within 30 days of a patient's discharge. The infections are unrelated to the illness that initially brought the person into the hospital. Due to the increasing rates of nosocomial infections in the 21st century, it is at the forefront of research.

Specifically, *S. epidermis* and *E. coli* have been proven to be the underlying causes of some nosocomial related infections. In the United States, one hundred and eight episodes of bloodstream infections were associated with *S.epidermidis* were detected over a seven-year period from 1976-1982 (Ponce De Leon and Wenzel, 1984). Rodriguez-Bano et al. (2010) conducted a study investigating the risk factors for nosocomial bloodstream infections linked to *E. coli* in thirteen Spanish hospitals.

Testing on bacterial strains associated with high numbers of nosocomial infections through innovative biopharmaceutical drugs will offer new possibilities to irradiate the spread of these bacteria strains. Many studies have confirmed the antimicrobial resistance of lichen species against different bacterial strains, but there is not research comparing the antimicrobial resistance of the same lichen species from difference regions within the United States. A study in Turkey has confirmed significant differences in the antimicrobial resistance between the same lichen species from different regions. Through further application of this study paired with factors from collected regions to two lichen species from three regions within Mississippi and Alabama, the possible antimicrobial resistance of a lichen species could be maximized.

The objective of this study are to 1) Compare the antimicrobial resistance of two lichen species from the black prairie region of Mississippi, gulf coast region of Mississippi, and southeastern plains region of Alabama 2) Develop an understanding of what factors contribute to the difference in antimicrobial resistance from between regions. The controlled variables of this study are the species of lichen, solvent, and amount of extract. The independent variables of this study are the geographic location, micro-habitat conditions, rainfall, light, UV exposure, elevation, temperature, wind, and air quality. The dependent variable of this study is the antimicrobial resistance presented from each lichen sample. It was hypothesized both lichen species *Ramalina farinacea* and *Flavoparmelia caperata* from the Southeastern Gulf Coastal Plain region of Mobile would have the highest antimicrobial resistance.

Through comparing the inhibition zones of lichen species from the Black Prairie region of Mississippi, Gulf Coast region of Mississippi, and Southeastern Gulf Coastal Plain region of Alabama, additional factors can be an-

alyzed to see if they play a role in the bioactive compounds of lichen. The data gained from this study could allow for the maximization of the potential commercial application of lichen in the biopharmaceutical industry.

Methodology

Lichen material: Lichen was collected on November 30, 2019 in the Southeastern Gulf Coastal Plain region of Alabama; December 2, 2019 in the Gulf Coast region of Mississippi; and December 3, 2019 black prairie region of Mississippi. Determination of species was found through cross examined of three different data bases.

Extracts: Lichen samples were cleaned from foreign materials like soil and tree bark. After dissection, the samples were nearly powered in a blender. The lichen remains were weighed and transferred to sterile centrifuge tubes. The lichen samples were suspended in three solvents for two nights; the solvents were 95% ethanol, distilled water, and .1M phosphate buffer. After the two nights, the samples were centrifuged for 10 minutes at 14,000 rpm.

In vitro antibacterial assays: The bacterial suspensions were adjusted to the 0.5 Mc Farland standard and were cultivated in petri dishes with nutrient agar media. After extracts were centrifuged, sterile disks were placed in the extracts and dried overnight. Evaporated disks were placed on bacteria-fed media utilizing the Kirby- Bauer Disk diffusion method.

Statistical analysis: Values on graphs were expressed in mean, standard deviations, and standard error.

Exposure Category	UV Range	Index Values	Description
Low	<2	0-50	Good
Moderate	3 to 5	51-100	Moderate
High	6 to 7	101-150	Unhealthy for Sensitive Groups
Very High	8 to 10	151-200	Unhealthy
Extreme	11+	201-300	Very Unhealthy
		301-500	Hazardous

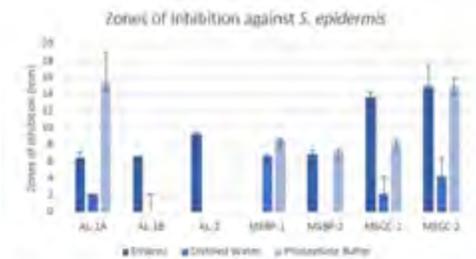
Table 1: UV Exposure chart, UV exposure is measured to be the total radiation factor or secondary metabolic output factor.

Table 2: Air Quality chart, AQI quality is measured to affect the amount of oxygen and carbon released for the right person.

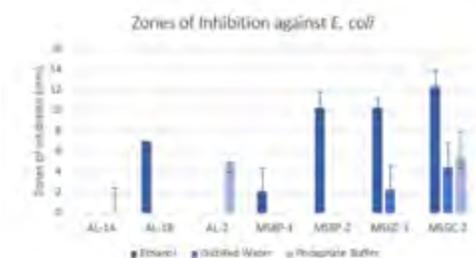
Locality Number	Species	Locality	Coordinates	Altitude (ft)
AL-1A	<i>Flavoparmelia caperata</i>	Piedmont, AL	30.7452° N, 88.0997° W	5-30 ft
AL-1B	<i>Flavoparmelia caperata</i>	Mobile, AL	30.6954° N, 88.0999° W	5-30 ft
AL-2	<i>Ramalina farinacea</i>	Mobile, AL	30.6954° N, 88.0999° W	5-30 ft
MSBP-1	<i>Flavoparmelia caperata</i>	Columbus, MS	33.4957° N, 88.4273° W	217ft
MSBP-2	<i>Ramalina farinacea</i>	Columbus, MS	33.4957° N, 88.4273° W	217ft
MSGC-1	<i>Flavoparmelia caperata</i>	Gaithers, MS	30.3858° N, 88.6117° W	7-10 ft
MSGC-2	<i>Ramalina farinacea</i>	Gaithers, MS	30.3858° N, 88.6117° W	7-10 ft

Factors	Collected Area	UV Exposure	Temp (Deg.)	Air Quality	Infrastructure/Statistical Area	Specific Location
Richard, AL	Water Oak	7-9	67°F 43°F	28	No/Yes	John Dorton Park
Mohale, AL	Live Oak	7-9	71°F 44°F	29	Yes/No	Metal of Honor, Scots park
Columbus, MS	Live Oak	1-4	57°F 33°F	32	Yes/No	MUW Campus
Greenville, MS	Live Oak	7-9	67°F 43°F	32	No/Yes	Shelby state park

Table 4: Observed and Researched factors of areas collected from



Graph 1: The bars represent the average inhibition zones of each extract in 3 different solvents against *S. epidermis* with standard error bars.



Graph 2: The bars represent the average inhibition zones of each extract in 3 different solvents against *E. coli* with standard error bars.

Summary of Results

The lichen samples from the varying areas displaced significantly different antimicrobial resistance with an average range being 0mm to 12.33mm for the bacteria strain *E. coli*. Extract MSGC-2 in the ethanol extract solvent had the highest average zone of inhibition against the gram-negative bacteria strain *E. coli*. Extracts AL-1A in the ethanol, distilled water, and phosphate buffer solvent; AL-1B in the distilled water and phosphate buffer solvent; AL-2 in the ethanol and distilled water solvent; MSBP-1 in the distilled water and phosphate buffer solvent; MSBP-2 in the distilled water and phosphate buffer solvent; and MSGC-1 in the phosphate buffer did not show antimicrobial resistance against *E. coli*. When all zones of inhibition against *E. coli* were added, extract MSGC-2 had the greatest antimicrobial resistance.

The lichen samples from the varying areas displaced significantly different antimicrobial resistance with an average range being 0mm to 15.50mm for the bacteria strain *S. epidermis*. Extract AL-1A in the phosphate buffer solvent had the highest average zone of inhibition against the gram-positive bacteria strain *S. epidermis*. Extracts AL-1B in the distilled water and phosphate buffer solvent, AL-2 in the distilled water and phosphate buffer solvent, MSBP-1 in the ethanol solvent, MSBP-2 in the distilled water extract the did not show antimicrobial resistance against *S. epidermis*. When all zones of inhibition against *S. epidermis* were added, extract MSGC-2 had the greatest antimicrobial resistance.

Conclusions

The hypothesis is rejected; the lichen samples from the Southeastern Gulf Coastal Plain region did not show the most antimicrobial resistance. The lichen species *Ramalina farinosa* (MSGC-2) had the largest zones of inhibition, making its extracts the most antimicrobial resistant against *S. epidermis* and *E. coli*. The results could be explained through further research of the independent variables (factors that could affect the secondary metabolites). This is because lichen is a reflection of its environment. Specifically, air quality has shown to be a significant contributing factor that the algal partner relies on. Because of this, specific gasses in the air affect the lichen's secondary metabolite production and concentration, affecting its antimicrobial resistance. Air quality is one example of many; location (exposure to the sun) also plays a significant role in secondary metabolite production. The MSGC-2 sample could have been living in the most ideal conditions, explaining why

it had the highest antimicrobial resistance.

The specific solvents used could have pulled out specific poplar or non-polar molecules that were more resistant to the bacterial strains chosen. This conclusion is drawn because the same species lichen from very close areas geographically showed large differences in their rings of inhibition (AL-1A and AL-1B). The lichen within different locations could be absorbing different molecules that could also possess antimicrobial resistance and that are compatible with the solvents used. It is also noted that within the same sample, there are large differences in zones of inhibition between the three different solvents, and the different solvents were more effective versus gram positive bacteria than gram negative bacteria and vice versa. This supports the conclusion that certain solvents were more compatible with different lichen samples, making certain extracts more effective against the chosen bacteria strains.

Implications and Recommendations

The two implications of this study align will sustainable development goals 3 and 12. Sustainable development goal 3 is to ensure healthy lives and promote well-being for all at all ages. Lichen has been found to have secondary metabolites that have the ability to inhibit the growth of bacteria. The effectiveness of antibacterial compounds of lichen from various areas were analyzed on bacteria linked to nosocomial bloodstream infections, an infection that kills hundreds of thousands of people annually. Sustainable development goal 12 is to ensure sustainable consumption and production patterns. Lichen is an easily accessible resource that has been used in medicine, food, cosmetics, dyes, and other ethnobotanical purposes since ancient times. More knowledge on the antibacterial properties of lichen will give way to how they can be further utilized in society. Lichen-derived bioactive compounds hold potential for their use commercial use in the pharmaceutical industry.

Recommendations include increasing the utilization of lichen, specifically its bioactive compounds. The solvents used in this study, ethanol, distilled water, and phosphate buffer, are easily accessible. The reasoning behind using easily accessible solvents is so the extracts could be recreated and the use of lichen in the biopharmaceutical industry could become normalized. With the solvents being easily accessible, this study's aim was also to make its findings applicable to underdeveloped countries. If people knew what signs to look for, they could be able to pick lichen with the maximum potential in antimicrobial resistance.

Future Research

Future research aims to further analyze the additional factors that could affect the antimicrobial resistance of lichen: micro-habitat conditions, rainfall, light, UV exposure, elevation, temperature, wind, and air quality. The addition of machine learning to this study could compile thousands of data points to validate which factors most significantly contribute to the antimicrobial resistance of lichen. More testing on different species and a larger range of solvents in conjunction with the data gathered from machine learning will offer the needed information to developed more conclusive results. The research would tell which species and solvent combination are the most antibacterial resistant, furthering the potential of lichen in biopharmaceutical industry and the possible development of new medicines. The addition of nanoparticles to the experiment could also further the development of lichen in the biopharmaceutical industry and aid in the developments of new medicines. This addition of nanoparticles could increase the antimicrobial resistance of lichen and open the gateway for how how other fungus, algae, and mold could further be applied.

References

- Deduke C, Timsina B, Piercey-Normore MD. (2012). Effect of environmental change on secondary metabolite production in lichen-forming fungi. In: Young S., (Ed.) International Perspectives on Global Environmental Change. 11(1), 197 – 230
- Niger Med J. (2012). Healthcare-associated infections: A public health problem. Nigerian Medical Journal. 53(2), 59-6
- Özyiğitoğlu, G., Açıkgöz, B., Tahiroğlu, G., & Sesal (2017). Comparison of antibacterial and antibiofilm activity properties of *Hypogymnia tubulosa* (Schaer.) Hav. lichen extracts from different locations in Turkey. Mycosphere. 8(8), 994-1002
- Ponce De Leon, S. & Wenzel, R. (1984). Hospital-acquired bloodstream infections with staphylococcus epidermidis: Review of 100 cases. The American Journal of Medicine. 77(4), 639-644
- Rodríguez-Baño, J., Navarro, M., Romero, L., Muniain, M., Perea, E., Pérez-Cano, R., Hernández, J., & Pascual, A. (2010). Risk Factors and Prognosis of Nosocomial Bloodstream Infections Caused by Extended-Spectrum-β-Lactamase-Producing *Escherichia coli*. Journal of Clinical Microbiology. 48(5), 1726-1731



Rachel Zheng



Bertha Mireles

The Gut Microbiome and the Immune System: Investigating Links between Diet and Antimicrobial Peptide Sublancin 168 Production

Abstract:

With the rise of antibiotic-resistant bacteria, finding novel ways to control pathogen growth is crucial. In the case of immunodeficiencies, a deeper look into how the gut microbiome can be utilized to combat infection is promising. *B. subtilis* 168 is a common member of the gut microbiome with a distinctive ability to produce a substance to combat immunosuppression, a glycopeptide called sublancin 168. Past research has suggested a significant ability of this naturally occurring bacteriocin to fight off a variety of pathogens including the methicillin-resistant *Staphylococcus aureus* (MRSA). This project focuses on targeting the gut microbiome through a selection of grains to quantify impacts on sublancin 168 production. The production of sublancin 168 was quantified through the difference in bacterial growth of *B. subtilis* and *B. megaterium* in vitro to determine the antimicrobial response of *B. subtilis* on *B. megaterium*. Once the grain that induced the most sublancin 168 production was identified, *Caenorhabditis elegans* were utilized to simulate a pathogen-host interaction under the resultant growth conditions. The common yeast pathogen, *Candida albicans*, that colonizes the gastrointestinal tract and can create a serious infection in immunodeficient patients, resulting in death, was used. Diet can cause fluctuations in the gut microbiome that can be beneficial to aiding the immune system against infection. Through analysis of members of the human gut microbiome, *B. subtilis*, *B. megaterium*, and *C. albicans*, sublancin 168 could be an effective booster to the immune system and an alternative to better fight off antibiotic resistant pathogenic infections.

Sublancin has antimicrobial properties that inhibits activity of other bacteria. A study conducted by the University of Illinois at Urbana-Champaign showed that glucose slowed down the antimicrobial activity while other sugars did not. The sugar differences were found to be linked with the phosphotransferase system (PTS) of *B. subtilis*. Using macromolecular assays, DNA synthesis was found to be greatly inhibited by controlling ciprofloxacin; therefore, sublancin disrupts macromolecular synthesis pathways such as DNA.

Our goal is to increase sublancin concentration by varying the carbon sources and measuring the production through the pathogenic behavior, and then testing our findings in a pathogen-host model.

Hypothesis

Null: Changes in diet in *C. elegans* have no effect on the immune system or the production of sublancin 168 by *B. subtilis* to overcome *C. albicans* infection.

Alternate: Changes in diet in *C. elegans* have an effect on the production of sublancin 168 by *B. subtilis* to overcome *C. albicans* infection.

Methodology

Phase I

Rice, wheat, and corn flours were weighed in the following concentrations onto 52 nutrient and potato agar plates: 0.005 g/mL, 0.01 g/mL, 0.015 g/mL, and 0.02 g/mL. 15 mL of autoclaved agar were pipetted into each plate. 10 mL of the sterile Luria Broth was pipetted using a sterile 10 mL pipette into 10 sterile, plastic vials and shut tightly. The vials were then inoculated with *B. subtilis* and *B. megaterium* individually by adding individual colonies until the turbidity meets a 0.5 McFarland Standard. After 24-hour growth, a generous amount of bacteria on each plate were inoculated in Durham tubes, and allowed to grow for 24 hours. After this, a nitrate reduction test was performed using sulfanilic acid and alpha-naphthylamine solution as reagents, and absorbance

readings using a spectrophotometer were recorded.

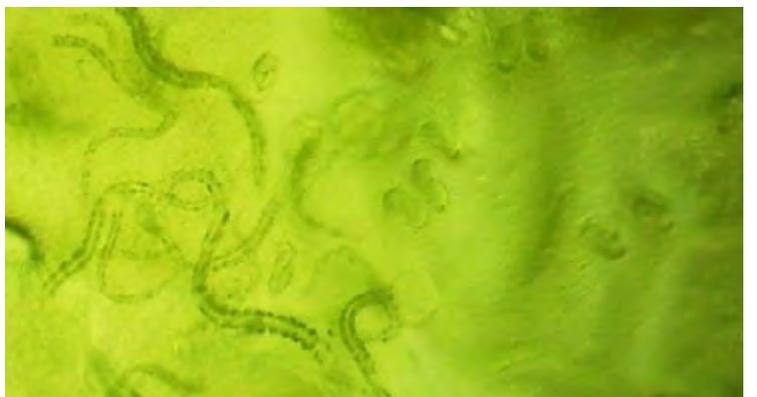
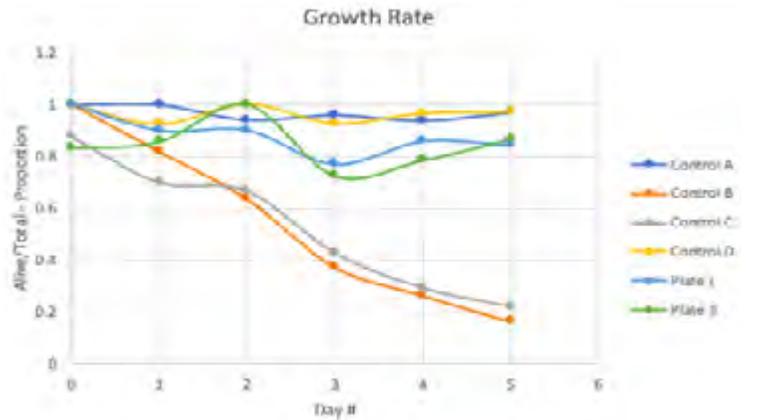
Phase II

After finding the optimal concentration of 0.02g/mL of wheat flour from analysis of absorbance readings, 6 separate nutrient agar plates were prepared with the optimal concentration:

Contents	<i>C. elegans</i>	<i>E. coli</i>	Wheat F	<i>C. albicans</i>	<i>B. megaterium</i>	<i>B. subtilis</i>
Control A	X	X	X			
Control B	X	X	X	X		
Control C	X	X	X		X	
Control D	X	X				
Plate I	X	X	X	X		X
Plate II	X	X	X		X	X

Table 1: Plates and their content

The worms were observed over the course of 5 days under a microscope with 100X magnification. The number of adult worms surviving on each plate was recorded.



Results and Conclusion

Difference of Proportions of Dead Count/Total Count of Day 5 (2 sample Z test)
 $\alpha=0.05$

Comparison	p-value	Significant?
Control A and Plate I	$p = 0.00114$	yes
Control A and Plate II	$P = 0.01390$	yes
Control B and Plate I	$P < 0.00001$	yes
Control C and Plate II	$P < 0.00001$	yes
Control A and Control D	$P=0.48803$	no

Table 2: 2-Sample Z Test Results

For Phase I, since the presence of wheat flour in the agar caused the increased growth of *B. subtilis* compared to that of *B. megaterium*, it is reasonable to conclude that the presence of wheat flour caused the optimal production of sublancin 168 and can therefore be used in Phase II.

For Phase II, since Control A and Control D have no significant difference in their death rates, the presence of flour in the agar can be factored out of the reason for the death rate discrepancies in the other plates. Since the difference in death rates for Control B and Plate I is significant with a $p < 0.00001$, it is reasonable to conclude that the death rate difference is due to the presence of sublancin 168 production by *B. subtilis* in Plate I. Since the difference in death rates for Control B and Plate II is significant with $p < 0.00001$, it is reasonable to conclude that the death rate difference is also due to the presence of sublancin 168.

The null hypothesis can be rejected as there is a relationship between the presence of flour, the production of sublancin 168, and *C. albicans* infection. Implications and Future Work

The upsurge of antibiotic resistant infections has caused a need for new, alternative antibiotics. In hospital settings and for people with compromised immune systems, altering the gut flora of patients is a promising way to counter antibiotics that can have debilitating side effects such as immunosuppression. Increasing sublancin production is a crucial method in combating these side effects. Further research into in vivo studies and clinical trials should be conducted and researched to give scientists a greater edge on natural antibiotic treatment.

References:

- Elkabti, Asmaa & Issi, Luca & Rao, Reeta. (2018). *Caenorhabditis elegans* as a Model Host to Monitor the Candida Infection Processes. *Journal of Fungi*. 4. 123. 10.3390/jof4040123 https://www.researchgate.net/publication/328794591_Caenorhabditis_elegans_as_a_Model_Host_to_Monitor_the_Candida_Infection_Processes
- Feistel, D. J., Elmstafa, R., Nguyen, N., Penley, M., Morran, L., & Hickman, M. A. (2019). A Novel Virulence Phenotype Rapidly Assesses Candida Fungal Pathogenesis in Healthy and Immunocompromised *Caenorhabditis elegans* Hosts. *MSphere*, 4(2). <https://doi.org/10.1128/mSphere.00697-18>.
- Ji, S., Li, W., Xin, H., Wang, S., & Cao, B. (2015). Improved Production of Sublancin 168 Biosynthesized by *Bacillus subtilis* 168 Using Chemometric Methodology and Statistical Experimental Designs. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4538419/>.
- Wu, C., Biswas, S., Garcia De Gonzalo, C. V., & van der Donk, W. A. (2019, March 8). Investigations into the Mechanism of Action of Sublancin. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30582697>.
- Zhang, R., & Hou, A. (2013). Host-Microbe Interactions in *Caenorhabditis elegans*. *ISRN Microbiology*. <https://www.hindawi.com/journals/isrn/2013/356451/>.



Fiona Dawe

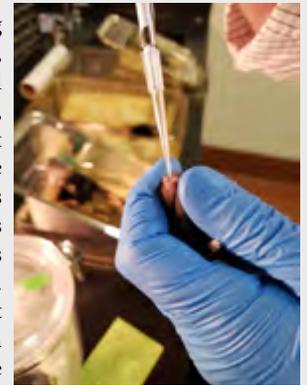


Clare Seo

Development of a Vaccine to Enhance Mucosal Immunity

As witness from the recent corona virus outbreak, infectious respiratory diseases are the significant global concerns to public health as well as world economy. Most respiratory pathogens colonize and penetrate mucosal membrane in our respiratory system. These specialized antibodies called secretory IgA (sIgA) play an important role in mucosal immunity to fight off invading pathogen (Johansen and Kaetzel, 2011). Immunoglobulin antibodies provide protection by preventing bacterial adherence to their host, neutralizing toxins, activating complement systems, forming anti-body-dependent cellular cytotoxicity (ADCC) complex, and secretory IgA is the major isotype of antibody at the mucosal membrane (Lamm, 1997). The surface immunoglobulin A (sIgA) is a key element in mucosal immunity

that provides protection against invading pathogens at the mucosal membranes, including the respiratory system, urogenital tract, and intestines (Gerbert, 1997). However, a preliminary study demonstrated that humans develop very poor IgA response to the common human pathogens such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, based on the data that was gathered at the beginning of the experiment. An adjuvant is an additive to a vaccine that can enhance immunogenicity or certain types of immune responses. Although there are many vaccine adjuvants that increase IgG responses, very few have been proven to trigger a robust IgA response. Human sera was obtained against *S. aureus* proteins and found to contain very high levels of IgA. This elicited in response to MHC II analog protein (MAP). Thus, we hypothesize that the use of MAP as a vaccine adjuvant will help to trigger an IgA response, thereby providing better protection following vaccination. To test this, the MAP gene amplification was observed by using the polymerase chain reaction. Additionally, the cloning of the protein expression plasmid vector and inducing the expression, was observed. Then, the recombinant MAP protein was purified. C56BL/6 mice were immunized with recombinant MAP protein alone or combined with other antigens by researchers at MSU. The results from this immunization were not included in this study. Following immunization, antibody titers and isotypes were quantitated from serum and tissues using an enzyme-linked immune absorbent assay (ELISA). The results of this study will contribute to improving the efficacy of vaccines against many important viral and bacterial pathogens.



As witness from the recent corona virus outbreak, infectious respiratory diseases are the significant global concerns to public health as well as world economy. Most respiratory pathogens colonize and penetrate mucosal membrane in our respiratory system. These specialized antibodies called secretory IgA (sIgA) play an important role in mucosal immunity to fight off invading pathogen (Johansen and Kaetzel, 2011). Immunoglobulin antibodies provide protection by preventing bacterial adherence to their host, neutralizing toxins, activating complement systems, forming anti-body-dependent cellular cytotoxicity (ADCC) complex, and secretory IgA is the major isotype of antibody at the mucosal membrane (Lamm, 1997). The surface immunoglobulin A (sIgA) is a key element in mucosal immunity that provides protection against invading pathogens at the mucosal membranes, including the respiratory system, urogenital tract, and intestines (Gerbert, 1997). However, a preliminary study demonstrated that humans develop very poor IgA response to the common human pathogens such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, based on the data that was gathered at the beginning of the experiment. An adjuvant is an additive to a vaccine that can enhance immunogenicity or certain types of immune responses. Although there are many vaccine adjuvants that increase IgG responses, very few have been proven to trigger a robust IgA response. Human sera was obtained against *S. aureus* proteins and found to contain very high levels of IgA. This elicited in response to MHC II analog protein (MAP). Thus, we hypothesize that the use of MAP as a vaccine adjuvant will help to trigger an IgA response, thereby providing better protection following vaccination. To test this, the MAP gene amplification was observed by using the polymerase chain reaction. Additionally, the cloning of the protein expression plasmid vector and inducing the expression, was observed. Then, the recombinant MAP protein was purified. C56BL/6 mice were immunized with recombinant MAP protein alone or combined with other antigens by researchers at MSU. The results from this immunization were not included in this study. Following immunization, antibody titers and isotypes were quantitated from serum and tissues using an enzyme-linked immune absorbent assay (ELISA). The results of this study will contribute to improving the efficacy of vaccines against many important viral and bacterial pathogens.

Works Cited

- Gerbert, A. (1997). The role of M cells in the protection of mucosal membranes. *Histochemistry and Cell Biology*, 108(6), 455–470. doi: 10.1007/s004180050186
- Johansen, F. E., & Kaetzel, C. S. (2011). Regulation of the Polymeric Immunoglobulin Receptor and IgA Transport: New Advances in Environmental Factors that Stimulate pIgR Expression and its Role in Mucosal Immunity. *Nature*.
- Lamm, M. E. (1997). Interaction of Antigens and Antibodies At Mucosal Surfaces. *Annual Review of Microbiology*, 51(1), 311–340. doi: 10.1146/annurev.micro.51.1.311

SCIENCE JOURNAL ENTRIES

Name/County	Competition
Trevor Allen/DeSoto	Region VII Science Fair, 1st Place Engineering Mechanics
Alicia Argrett/Madison	Region VII Science Fair
David Barber/Washington	Region III Science Fair, ISEF Recipient
Victoria Callahan/Madison	World Food Prize, Global Youth Institute
Blake Cheater/Lauderdale	Region IV Science Fair
Fiona Dawe/Oktibbeha	Region V Science Fair/1st Place Microbiology
Shanay Desai/Madison	Region VII Science Fair, 1st Place Microbiology
Jackson Flowers/DeSoto	Region VII Science Fair, ISEF Recipient
Colin Gordy/Lamar	Region III Science Fair, Alternate ISEF Recipient
David Heson/Lee	Region IV Science Fair, 1st Place Physics
Guillermo Hoffmann Meyer/Oktibbeha	Region III Science Fair, ISEF Recipient
Violet Jira/Bolivar	Region III Science Fair, ISEF Recipient
Dennis Lee/Oktibbeha	Region V Science Fair, ISEF Recipient
Nathaniel Lee/Rankin	Region VII Science Fair, 1st Place Physics
Michael Lu/ Oktibbeha	Region V Science Fair, 1st Place Math and Systems Software
Catherine Min/DeSoto	Region VII Science Fair
Bertha Mireles/Lowndes	Region VII Science Fair
Gina Nguyen/Harrison	Region III Science Fair, ISEF Recipient
Skylar Nguyen/Jackson	Region VII Science Fair, 2nd Place Microbiology, World Food Prize, Global Youth Institute
Vayd Ramkumar/Oktibbeha	Region VII Science Fair, ISEF Recipient
Clare Seo/Oktibbeha	Region V Science Fair, 1st Place Microbiology
Muneebah Umar/Oktibbeha	World Food Prize, Global Youth Institute
Aaron Wan/Oktibbeha	Region V Science Fair, ISEF Recipient, World Food Prize, Global Youth Institute
Hong "Reggie" Zheng/Bolivar	Region III Science Fair, Alternate ISEF Recipient
Rachel Zheng/Madison	Region VII Science Fair

MSMS Science Journal

Waves of Science

The MSMS Science Journal is a collection of research projects conducted by the students at
The Mississippi School for Mathematics and Science
1100 College Street, MUW-1627
Columbus, Mississippi 39701

The MSMS Science Journal is available to read on our website at www.themsms.org