Using Machine Learning Regressors for the Discovery of *Culex* Mosquito Habitats and Breeding Patterns in Washington D.C.

Iona Xia[1], Neha Singirikonda[2], Landon Hellman[3], Jasmine Watson[4], Marvel Hanna[5] and Dr. Russanne Low[6#]

[1]Monta Vista High School, USA
[2]CHIREC International School, India
[3]Dos Pueblos High School, USA
[4]Brewer High School, USA
[5]Huntington Beach High School, USA
[6]Institute for Global Environmental Strategies, USA
[#]Advisor

ABSTRACT

Culex mosquitoes pose a significant threat to humans and other species due to their ability to carry deadly viruses such as the West Nile and Zika. Washington D.C., in particular, has a humid subtropical climate that is ideal as a habitat for mosquito breeding. Thus, tracking mosquitoes' habitats and breeding patterns in Washington D.C. is crucial for addressing local public health concerns. Although fieldwork techniques have improved over the years, monitoring and analyzing mosquitoes is difficult, dangerous, and time-consuming. In this work, we propose a solution by creating a Culex mosquito abundance predictor using machine learning techniques to determine under which conditions Culex mosquitoes thrive and reproduce. We used four environmental variables to conduct this experiment: precipitation, specific humidity, enhanced vegetation index (EVI), and surface skin temperature. We obtained sample data of these variables in the Washington D.C. areas from the NASA Giovanni Earth Science Data system, as well as mosquito abundance data collected by the D.C. government. Using these data, we created and compared four machine learning regression models: Random Forest, Decision Tree, Support Vector Machine, and Multi-Layer Perceptron. We searched for the optimal configurations for each model to get the best fitting possible. Random Forest Regressor produced the most accurate prediction of mosquito abundance in an area with the four environment variables, achieving a mean average error of 3.3. EVI was the most significant factor in determining mosquito abundance. Models and findings from this research can be utilized by public health programs for mosquito-related disease observations and predictions.

1. Introduction

Culex mosquitoes are some of the most common species of mosquitoes worldwide and can carry many types of diseases, including the West Nile and Zika Viruses (Omodior et al. 2018). At present, there are no licensed vaccines or medicines for such diseases. Culex sp. are native to Africa, Europe, and Asia, and are found worldwide (Soh and Aik 2021). Therefore, a large number of societies are prone to these deadly diseases. Past studies have been limited in their noticeable effects on human populations; however, very little has been done due to the lack of understanding of how these viruses affect the human body (Center of Disease Control 2020). Furthermore, there are currently no established systems that predict mosquito-borne disease outbreaks.

Washington D.C. is known to have a humid subtropical climate which tends to be ideal for mosquito breeding habitats. With an annual rainfall average of 42 inches per year, Washington D.C.'s rainfall is 12 inches higher than the nationwide yearly average of 30.2 inches. Thus, Washington D.C. is prone to abundant mosquito populations because they thrive in relatively wet places (National Centers for Environmental Information 2022). Washington D.C. has a unique environment, with the Potomac River, a freshwater stream running through Washington D.C., serving this area and the areas around it as habitable mosquito microenvironments. Washington D.C.'s natural topography varies, as the areas to the east of the Potomac River range from 100-140m above sea level, while the west of the Potomac River ranges from 40-80m (Topographic-Maps 2022, Figure 1). The first human case of West Nile Virus in the United States was reported in the District of Columbia (DC Health 2022). In 2018, there were 13 reported human cases, and there were 11 cases in 2019, in a mere 68.34 mi^2 area (Center of Disease Control 2020). This is evidence that Washington D.C. has a significant public health threat from mosquitos and the West Nile virus, as well as other mosquito-transmitted diseases, which can become a severe problem in small geographic areas.

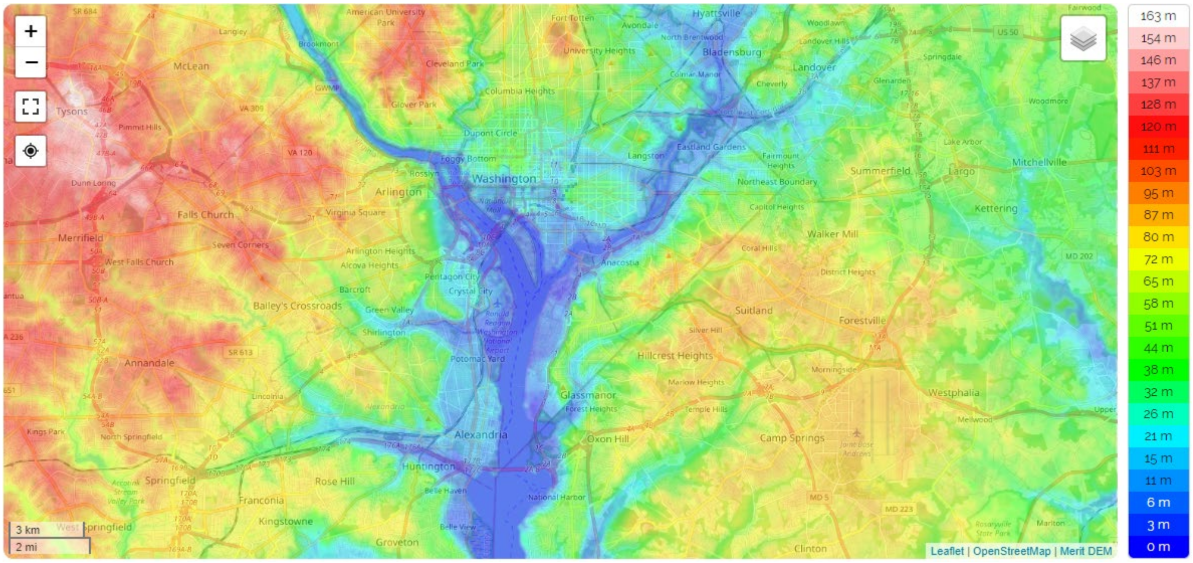

Figure 1. A representation of Washington D.C.'s natural topography. The geographic center of Washington D.C. is near the intersection of 4th and L Streets NW. The highest natural elevation in D.C. is 409 feet (125 m) above sea level at Fort Reno Park in upper northwest Washington. The lowest point is at sea level at the Potomac River.
Scale Unknown, (CC-BY-SA 3.0), Washington D.C., 2022.

Historically, machine learning models have proven to be valuable tools for predicting trends and operational patterns. Machine learning is significantly helpful in predicting breeding patterns of various animals, including mosquitoes. In this work, we chose four primary machine learning models: Random Forest, Decision Tree, Support Vector Machine, and Multilayer Perceptron. Given its several capabilities, random forest regression models were recently used in predicting West Nile Virus positivity rates and abundance in the city of Chicago (Schneider et al. 2021). Random Forest algorithms can construct a collection of decision trees to perform classification assignments, allowing users to make highly accurate discrete predictions and solutions for their data sets. Moreover, the decision trees constructed by the Random Forest models can be used to create regression tasks that further help users predict continuous outputs for nonlinear inputs (Schonlau and Zou 2020). Because Random Forest is heavily used specifically in geology and earth science, we predict that it would also do the best in our tests. Decision Tree regression, on the other hand, features only one decision tree analysis feature by partitioning data and fitting a simple model for each partition and exists as a more straightforward method of Random Forest (Lou 2011). The Support Vector regression model is based on a linear regression that fits its data by a hyperplane in a higher dimension, allowing it to recognize subtle patterns

(Basak and Pal 2007). In addition, Multilayer Perceptron regression is a neural network algorithm that consists of various nodes and can learn non-linear models, which could prove to be helpful (Murtagh 1991).

A previous study by Drakou et al. (2020) displayed that the amount of Culex pipiens, Aedes detritus, and Aedes caspius mosquitoes grew due to precipitation increase. Francisco et al. (2021) found connections between the dengue virus, another disease transmitted by mosquitoes, and environmental and landscape factors, including precipitation, land surface temperature, the normalized difference of vegetation index, and impervious surfaces, including roads. Other studies, such as Madewell et al. (2019), have also discovered connections of mosquito habitat with urbanization. In this study, we found that studying landcover GLOBE data was useful with the various environmental factors available from the NASA Giovanni Earth datasets; and our quest was to use machine learning techniques to predict Culex mosquito breeding patterns in Washington D.C. with GLOBE and open-sourced data.

2. Methods

We chose Washington D.C. as our primary area of interest (AOI) due to its high accessibility of open-sourced mosquito data and environmental data, as well as its historical abundance of mosquitoes. We obtained data from the NASA Giovanni Earth science remote sensing data system, the GLOBE database, and the Washington D.C. government data. The data we collected spans from April 2016 to October 2018. Data sets included the daily data of Average Surface Temperature, Specific Humidity, Precipitation, and EVI (Giovanni 2022). The Washington D.C. government provided open-sourced quantitative mosquito data in our respective AOI (Open Data DC 2021).

Initially, when trying to extract data from GLOBE, we ran into the issue of getting inconsistent data, specifically because of the opportunistic nature of this database. Although we originally had a hard time finding several data points in one area, we eventually found a way around this problem by discovering the data from a different perspective. Instead of investigating the GLOBE Observer Mosquito Habitat Mapper data, we started looking at GLOBE's land cover data. This helped us immensely in our quest to answer our initial research question by giving us perspective on how the land cover pictures around Washington D.C. correlate to the environmental factors we used in our experiment. The availability of GLOBE citizen science data in Washington D.C. indicates the location as an area of interest for health programs and government efforts. We used the GLOBE data to analyze specific land cover observations in Washington D.C., which allowed us to determine which habitats correlate with specific environmental factors and, therefore, the mosquito breeding patterns. GLOBE data was collected using land cover measurements that were updated daily, including green-down, green-up, tree height, and land cover classifications. The qualitative GLOBE open-source data determined that the land cover points were generally concentrated in Ward 3 of Washington D.C., with many GLOBE citizen science users. Using GLOBE and Washington D.C. open data, Ward 3 displayed relatively low container removals (Figure 2). Therefore, we speculate that Ward 3 will result in somewhat greater mosquito breeding as there are lower container removals which are an ideal breeding habitat for Culex mosquitoes.

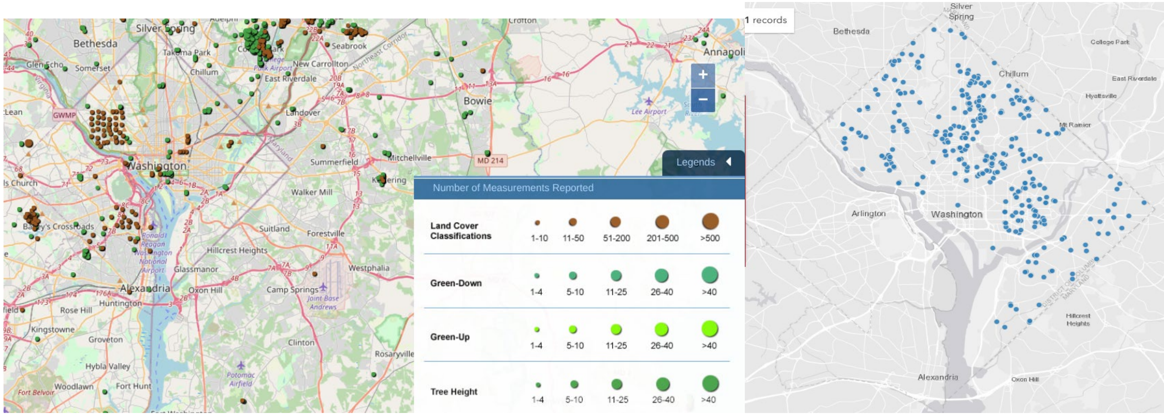

Figure 2. A representation of our data collected from both the government and citizen science.

Left: Washington D.C. land cover data updated daily made with GLOBE Observer. The legend includes land cover classifications, green-down, green-up, and tree height. Global Learning and Observations to Benefit the Environment (GLOBE) Program, July 21, 2022, https://globe.gov

Right: Washington D.C. open access data from ArcGIS Online showing container removal updated monthly.
Scale Unknown, Open Data DC in the Office of the Chief Technology Officer, Washington D.C., 2022

2.1 Data Features

We calculated the correlation between mosquito breeding patterns and the environmental variables that have been shown to influence mosquito breeding patterns in the past (i.e., precipitation, EVI, specific humidity, surface skin temperature). We chose these four environmental factors as we hypothesized that they would significantly affect mosquito breeding patterns. We decided to calculate this relationship because it allowed our machine learning models to predict Culex mosquito breeding patterns given certain environmental conditions in our AOI location. The visual derivatives of the mosquito abundance in our AOI location showed high variability over our selected time and were utilized in our models to identify the specific relationships further. The means and ranges of all the data used are shown in Table 1.

Table 1. The mean and ranges of inputs from NASA Giovanni Data and Washington D.C. Open Data.

Factor	Mean	Range
Average Mosquitoes	8.3600	37.720
Precipitation (mm/day)	2.5939	34.623
Specific Humidity (kg/kg)	0.0113	0.0150
Average Skin Surface Temperature (℃)	22.352	25.639
EVI (Spectral Index)	0.3279	0.1882

2.2 Analyzing Trends and Lag

We aligned our ecological variables with the daily mosquito data taken from the government of Washington D.C. We displayed the graphical relationships between each environmental variable and mosquito abundance in Figure 3. The p-value was calculated by using the Ordinary Least Square Regression model. For us to consider a factor statistically significant, it needed to have a p-value < 0.05. We had previously hypothesized that there would be a lag in the data due to the lengthy incubation period of mosquito eggs, which would lead ecological variables to only show an effect in the population several days after (Environmental Protection Agency 2022).

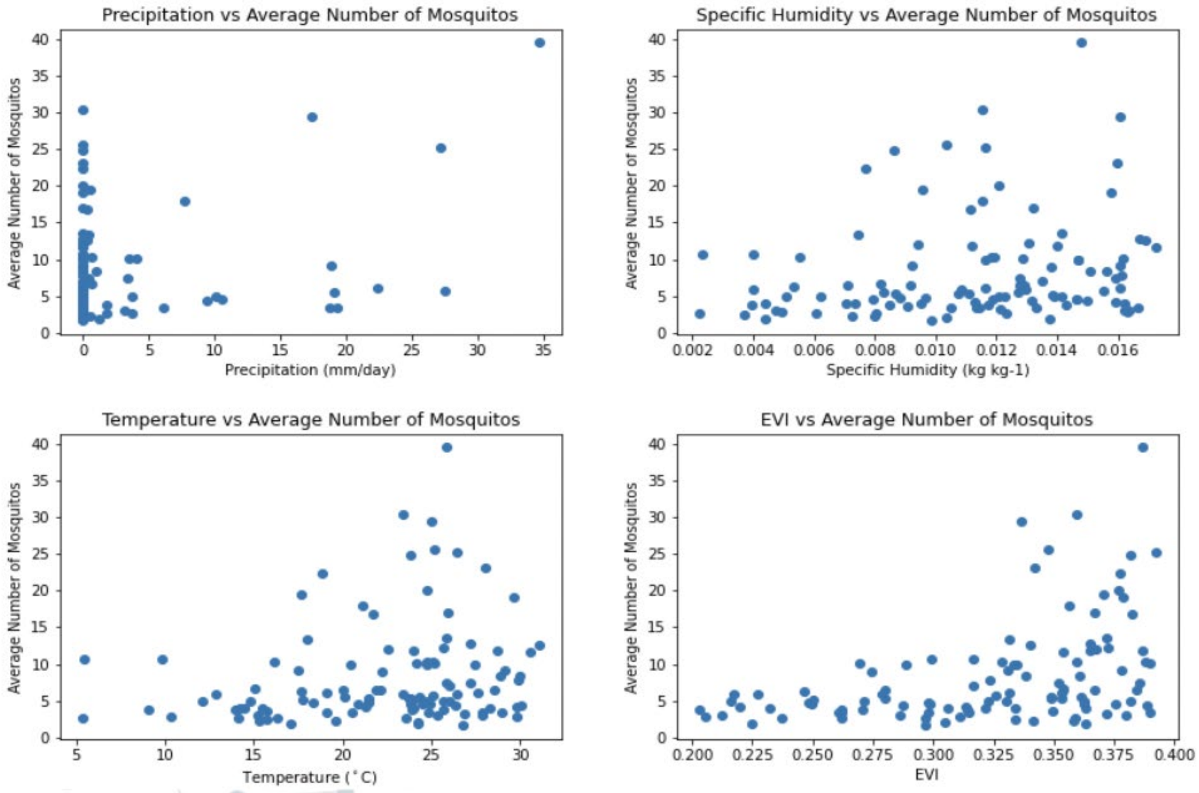

Figure 3. Relation between the mosquito abundance vs. each environmental variable.

In the above figure, the top left scatter plot shows the relationship between the precipitation and the average mosquito populations. The precipitation has a p-value of 0.0023. The top right scatter plot shows the relationship between the specific humidity measured in (kg/kg) and the average mosquito populations. The specific humidity has a p-value of 0.0487. The bottom left scatter plot shows the relationship between the average surface skin temperature measured in ℃ and the average mosquito populations. The average surface skin temperature has a p-value of 0.0393. The bottom right scatter plot shows the relationship between the enhanced vegetation index (EVI) measured in Spectral Index (Band Ratio) and the average mosquito populations. The EVI has a p-value of 5e-6.

However, when we compared the difference between with a one-week lag and with no lag, we found that all of the data were statistically significant (with smaller p-values) with no lag. Temperature was the only one that was more statistically significant with a one-week lag. Peaks in the data also matched better with no lag (Figure 4). Certain factors such as humidity and precipitation would not have been statistically significant with the one-week lag. We compared the differences between the p-values for each ecological variable with vs. without the data lag in Table 2.

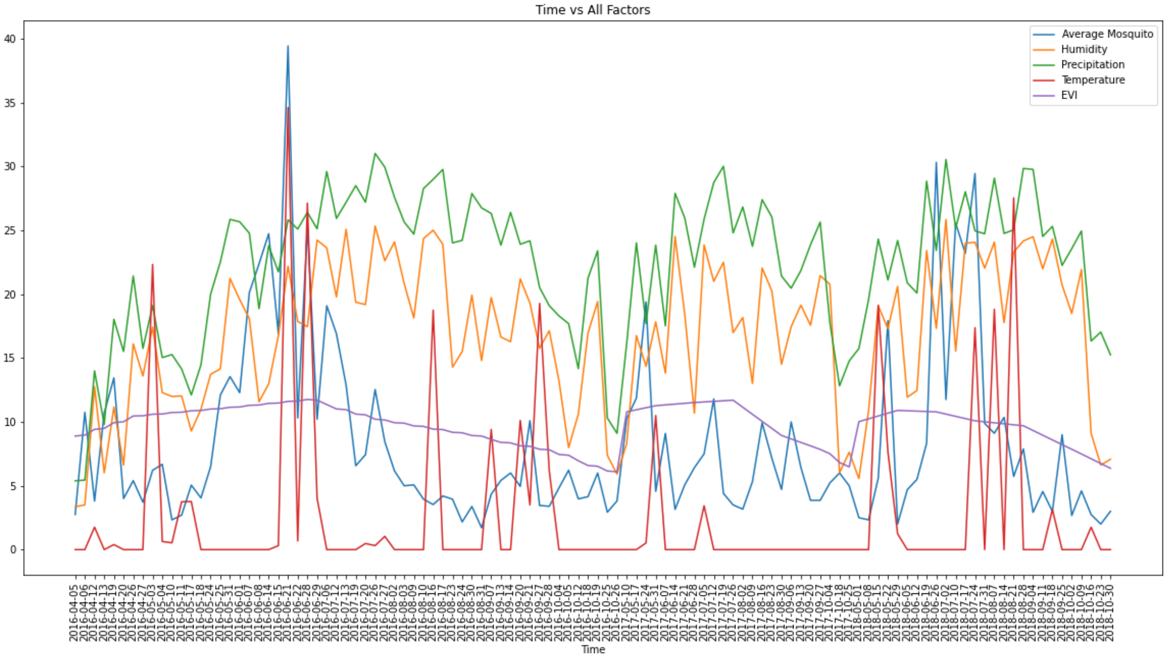

Figure 4. The line graphs show the relationship between time and precipitation, temperature, humidity, and enhanced vegetation index (EVI) with no lag time. EVI values have been multiplied by 30, and humidity values have been divided by 5 for scaling purposes.

Table 2. The differences between the p-values of precipitation, temperature, humidity, and enhanced vegetation index (EVI) when with vs. without the data lag. A statistical significance threshold is 0.05.

Variables	1 Week Lag	No Lag
Precipitation	0.349566	0.002316
Temperature	0.021255	0.039277
Humidity	0.581798	0.048652
EVI	0.000020	0.000005

As shown in Table 2, the p-value is less than 0.05 for every environmental variable, verifying that they all significantly affect mosquito populations. This rejects the null hypothesis stating that none of the environmental variables has a statistically significant relationship with the mosquito populations.

2.3 Data Preprocessing

Data used in this study came from different sources and various satellites, thus it was necessary to do plenty of data cleaning. The Washington D.C. government-collected mosquito data was measured twice a week for most of 2016 and once a week for the rest of the time. The data contained the number of both females and males of various types of mosquitoes. Because we focused on predicting mosquito population growth, we only kept data for female Culex mosquitoes. Due to the varying amounts of mosquito traps per day, we took the average number of mosquitoes per trap per day for the dates on which mosquito traps were set. We had 108 days from May to October in the years 2016-2018. For each of these days, we collected data for EVI, average surface skin temperature, specific humidity, and precipitation. We found a couple of holes in some of the environmental factor data. For these, we used SciPy's interpolation method interp1d to fill in the gaps, which we found to be the most accurate.

2.4 Training the Models

We trained a variety of models to find the best possible solution. All models were from the SciKit-Learn python package, and we tested model hyperparameters using the Grid Search Cross Validation tool in SciKit-Learn. We tested four models in particular: the Random Forest Regressor model, the Decision Tree model, the Multilayer Perceptron model, and the Support Vector Regression model. The hyperparameters tested and chosen for Random Forest Regressor and Decision Tree Regressor are listed in Tables 3 and 4, respectively. We decided to use a 70/30 training-testing split, using SciKit-Learn's function "train_test_split" to split them up randomly.

Table 3. Hyperparameters for the Random Forest regressor including the values tested and the values chosen by GridSearchCV.

Hyperparameter	Values Tested	Value Chosen
'bootstrap'	True, False	True
'max_depth'	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, None	60
'max_features'	'auto', 'sqrt'	'sqrt'
'n_estimators'	100, 300, 500, 700, 900, 1100, 1300, 1500, 1700, 1900	300
'min_samples_leaf'	1, 2, 4	1
'min_samples_split'	2, 5, 10	2

Table 4. Hyperparameters for the Decision Tree regressor including the values tested and the values chosen by GridSearchCV.

Hyperparameters	Tested Values	Chosen Value
'splitter'	'best', 'random'	'random'
'max_depth'	1, 3, 5, 7, 9	9
'min_samples_leaf'	1, 2, 3, 4, 5, 6, 7	4
'min_weight_fraction_leaf'	0.1, 0.2, 0.3, 0.4, 0.5	0.1
'max_features'	'auto', 'log2', 'sqrt', None	'auto'
'max_leaf_nodes'	10, 20, 30, 40, 50, 60, None	60

3. Results

When testing on the four models, we measured the two metrics: the mean absolute error (MAE) and the root mean square error (RMSE). These two measurements are both critical in different ways. MAE measures the average magnitude of the error (i.e., the difference between the predicted value and the true value) without caring about the direction. RMSE, on the other hand, measures the square root of the average of the squared differences. MAE is more often used, primarily for comparing model statistics; it has been found that RMSE is better used to represent the model performance when the error is in Gaussian distribution. We found it necessary to measure both metrics to understand which model has the best performance (Chai and Draxler 2014). When comparing all four models, we discovered that the Random Forest regressor performed better in both MAE and RMSE than other models. Support Vector Machine performed the worst in RMSE, and Multi-Layer Perceptron performed the worst in MAE. Table 5 details the scores for each model in MAE and RMSE. However, all models did perform similarly.

Table 5. The performance metrics for the Random Forest, Decision Tree, Support Vector Machine, and Multi-Layer Perceptron models.

Model	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)
Random Forest Regressor	3.27046	5.14630
Decision Tree Regressor	3.40613	5.30988
Support Vector Regressor	3.51837	6.08155
Multi-Layer Perceptron Regressor	3.92544	5.40554

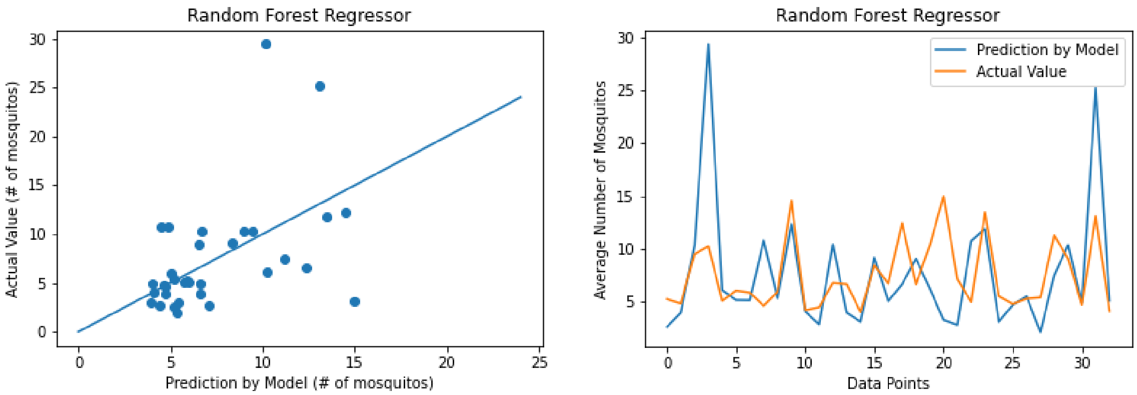

Figure 5. The relationship between the predicted number of mosquitoes by the model and the actual number of mosquitoes for the Random Forest Regressor. Left: A plot of the model's prediction vs the actual number of mosquitoes. Right: Another representation with the prediction by the model being plotted in blue and the actual value plotted in orange.

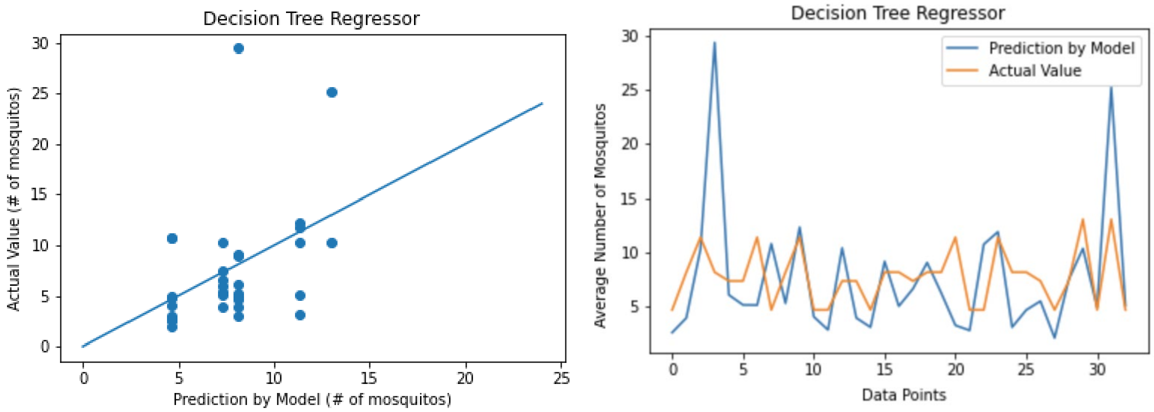

Figure 6. The relationship between the predicted number of mosquitoes by the model and the actual number of mosquitoes for the Decision Tree Regressor. Left: A plot of the model's prediction vs the actual number of mosquitoes. Right: Another representation with the prediction by the model being plotted in blue and the actual value plotted in orange.

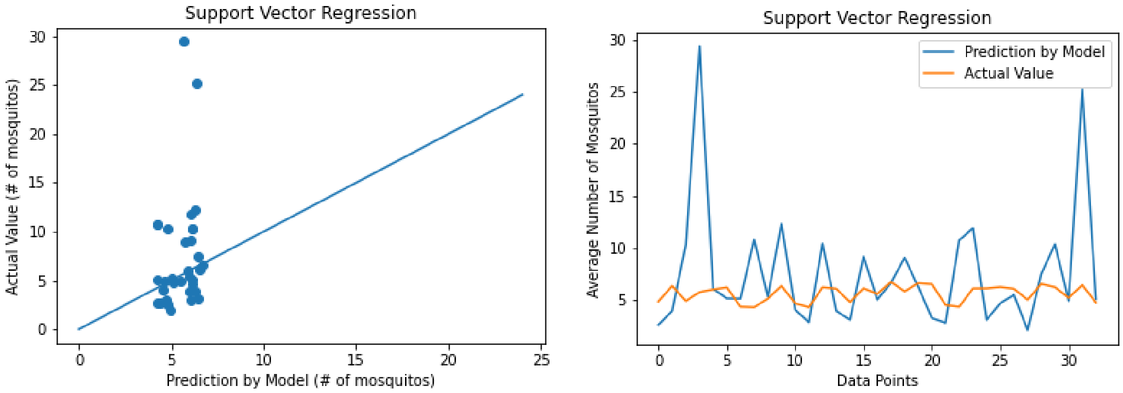

Figure 7. The relationship between the predicted number of mosquitoes by the model and the actual number of mosquitoes for the Support Vector Regressor. Left: A plot of the model's prediction vs the actual number of mosquitoes. Right: Another representation with the prediction by the model being plotted in blue and the actual value plotted in orange.

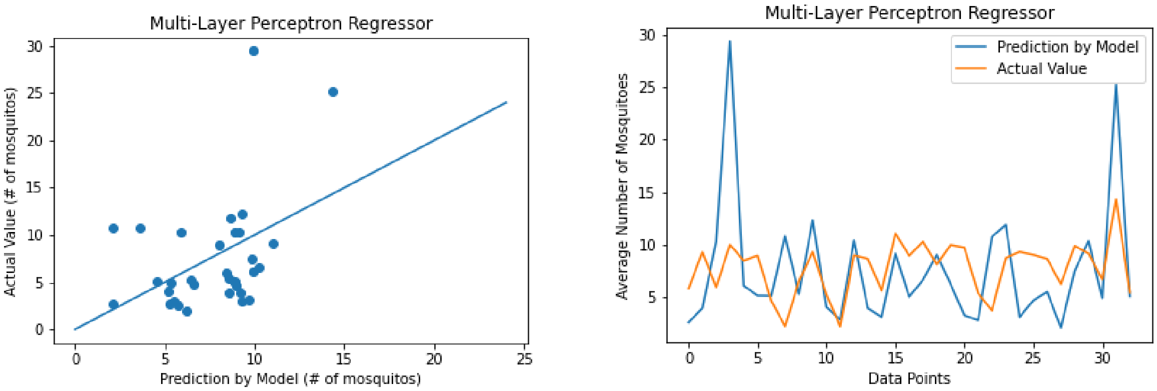

Figure 8. The relationship between the predicted number of mosquitoes by the model and the actual number of mosquitoes for the Multi-Layer Perceptron Regressor. Left: A plot of the model's prediction vs the actual number of mosquitoes. Right: Another representation with the prediction by the model being plotted in blue and the actual value plotted in orange.

We plotted the comparison between the predictions and the actual mosquito values for the four models in Figures 5, 6, 7, and 8, respectively. We noticed that all models struggled to correctly predict the values for the days July 24th, 2018, which had an average of 29.45 mosquitoes, and June 28th, 2016, which had an average of 25.27 mosquitoes. Most models predicted that July 24th, 2018, would have about an average of 10 mosquitoes, and June 28th, 2016, would have an average of 15 mosquitoes. These two points had the highest values among all data points, and are about ten mosquitoes more than the next largest. Furthermore, when these two points were removed, the Random Forest Regressor accuracy metrics significantly improved, as shown in Table 6. The MAE was able to improve by 0.65, while the RMSE was able to improve by 1.71.

Table 6. The scores for Random Forest Regressor with and without the outliers.

Condition	Mean Absolute Error	Root Mean Square Error
With Outliers	3.27046	5.14630
Without Outliers	2.62273	3.43861

When we looked at the environmental factors, as shown in Figure 9, we found that these two data points had higher values in most of the environmental factors, especially for June 28th, 2016, which had the highest value for two of the environmental factors. This may explain why while the model didn't give a close enough prediction for this point, it still predicted it to be a much higher value than the other points. However, July 24th, 2018 only had the highest specific humidity level and was in the medium range for the other factors, which is most likely why it was not predicted to have a very high value. Nevertheless, this case represents how the amount of change in environmental factors is not always comparable to the change in the average number of mosquitoes, especially when the average number of mosquitoes is extremely high. Therefore, it is necessary to have a significantly larger amount of data, especially more data to represent the high mosquito abundance cases, to train the models in order for them to handle such cases well.

4. Discussion

Our research presents an analytical and consistent viewpoint on the relationships between mosquito abundance and environmental factors. Our machine learning models produced highly accurate predictions by correlating our selected factors with increases or decreases in mosquito abundance. Based on our results, we found that EVI has the most significant effect on mosquito abundance populations in our respective AOI. This idea can be applied virtually anywhere across the globe. However, we acknowledge that there are infinitely many anthropogenic and non-anthropogenic factors, including the variables we used, that could affect mosquito abundance in any specific area. Several examples of literature have claimed that temperature and EVI significantly affect mosquito population growth, which partially supports the results from our machine learning models.

Our research could have been improved in several ways. All of our mosquito data was taken from Washington D.C.'s public database, dating from 2016 to 2018. Due to climate change, environmental factors have changed drastically over the years, and our machine learning predictors are more suitable for analyzing 2016 trends. Also, due to the opportunistic nature of recording mosquito abundance in an AOI, we did not have complete mosquito abundance data when we ran our machine learning models, leaving us with only 108 data points — which is relatively low compared to most training and testing sets for machine learning models. Our models would have created more accurate predictions about isolated areas around Washington D.C. with even more consistent mosquito data and additional environmental variables.

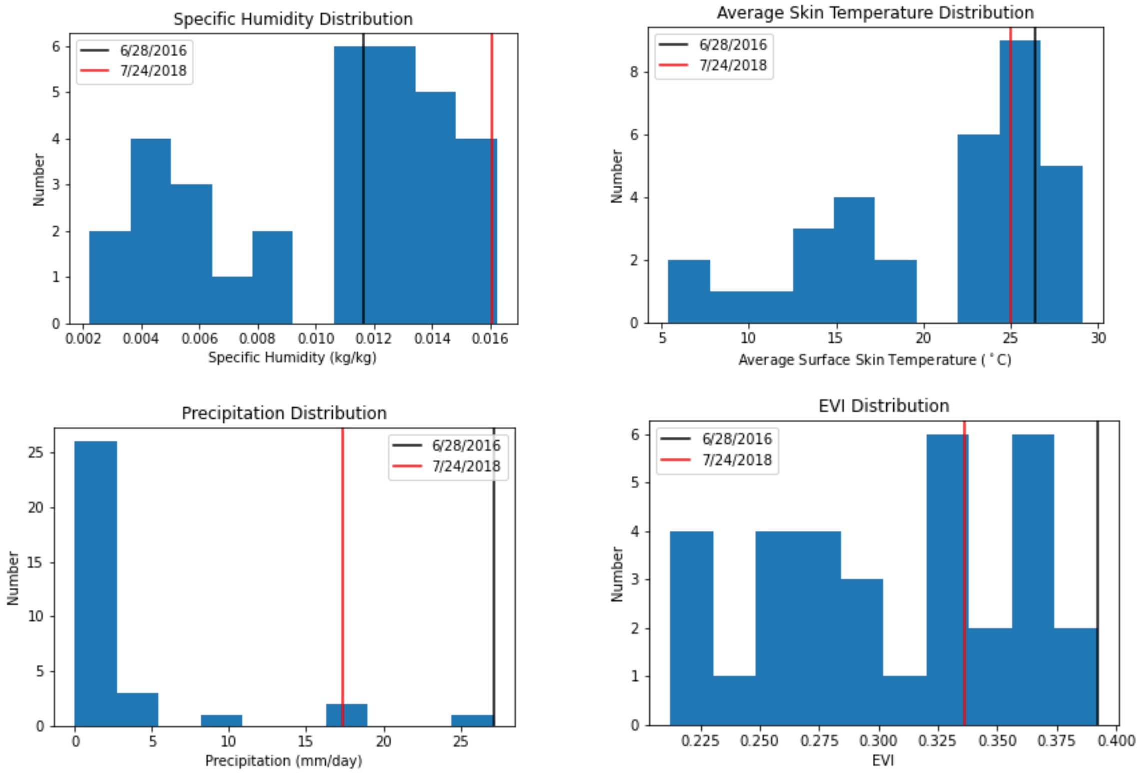

Figure 9. The distributions of environmental factors for our test set versus our outliers. Top Left: The distribution for specific humidity. July 24th, 2018 had one of the highest humidity values. Top Right: The distribution for average surface skin temperature, both July 24th, 2018 and June 28th, 2016 had high temperatures. Bottom Left: The distribution for precipitation. June 28th, 2016 had the highest precipitation. Bottom Right: The distribution for EVI. June 28th, 2016 had the highest EVI.

5. Conclusion

As we anticipated, the Random Forest regressor provided the best prediction result for mosquito abundance, with EVI as the most influential factor. However, we did find that all models had more or less similar results. All four machine learning models have proven to be useful predictors in analyzing mosquito abundance patterns. Our developed model provides a thorough, easily examinable way to understand the correlations between environmental factors and mosquito populations. These machine learning models can be used in the future for predicting mosquito breeding patterns across various areas of interest, including but not limited to Washington D.C., to help public health organizations in predicting mosquito habitats and mosquito-borne diseases.

We hope future research projects will expand their areas of interest to locations outside Washington D.C. and the United States, as training models with data in diverse areas can yield better predictors. Moreover, we hope more ecological variables are available to be used in future research projects, which may lead to more accurate results.

Crediting Contributors

Author Contributions: Conceptualization, I.X., N.S., L.H., and J.W.; data collection, I.X., N.S., L.H., and J.W.; formal analysis, I.X. and N.S.; investigation, L.H., J.W. and M.H.; methodology, I.X.; software, I.X., writing—original draft, I.X., N.S., L.H., J.W., and M.H.; writing—review and editing, I.X., N.S., L.H., J.W., and M.H. All authors have read and agreed to the submitted version of the manuscript.

Data Availability

GLOBE Observer data were obtained from NASA and the GLOBE Program and are freely available for use in research, publications, and commercial applications. GLOBE Observer data analyzed in this project are publicly available at globe.gov/globe-data (accessed on July 5, 2022). Giovanni Earth data were obtained from NASA and are publicly available at https://giovanni.gsfc.nasa.gov/giovanni/ (accessed on July 5, 2022). Washington D.C. mosquito data was provided by the government of the District of Columbia and is publicly available at https://opendata.dc.gov/datasets/DCGIS::mosquito-trap-sites/about (accessed on July 5, 2022).

Acknowledgements

The authors would like to acknowledge the support of the 2022 Earth System Explorers Research Team, NASA STEM Enhancement in the Earth Sciences (SEES) Virtual High School Internship program. Special thanks to Dr. Rusanne Low, Ms. Cassie Soeffing, Mr. Peder Nelson, Dr. Erika Podest, Andrew Clark, and Alexander Greco for their guidance when preparing this work. The NASA Earth Science Education Collaborative leads Earth Explorers through an award to the Institute for Global Environmental Strategies, Arlington, VA (NASA Award NNX6AE28A). The SEES High School Summer Intern Program is led by the Texas Space Grant Consortium at the University of Texas at Austin (NASA Award NNX16AB89A).

References

Basak, D., & Pal, S. (2007). Support Vector Regression. *Statistics and Computing*, *11*(10), 203–224.

Celestin, M. N., & Musteata, F. M. (2021). Impact of Changes in Free Concentrations and Drug-Protein Binding on Drug Dosing Regimens in Special Populations and Disease States. *Journal of pharmaceutical sciences*, *110*(10), 3331–3344. https://doi.org/10.1016/j.xphs.2021.05.018.

Centers for Disease Control and Prevention. (2021, December 17). *Final Cumulative Maps and Data*. Centers for Disease Control and Prevention. Retrieved July 22, 2022, from https://www.cdc.gov/westnile/statsmaps/cumMapsData.html.

Centers for Disease Control and Prevention. (2020, December 7). *Prevention*. Centers for Disease Control and Prevention. Retrieved July 22, 2022, from https://www.cdc.gov/westnile/prevention/index.html

Chai, T., & Draxler, R. R. (2014). Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments against Avoiding RMSE in the Literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

Environmental Protection Agency. (n.d.). EPA. Retrieved July 22, 2022, from https://www.epa.gov/mosquitocontrol/mosquito-life-cycle.

Francisco, M. E., Carvajal, T. M., Ryo, M., Nukazawa, K., Amalin, D. M., & Watanabe, K. (2021). Dengue Disease Dynamics Are Modulated by the Combined Influences of Precipitation and Landscape: A Machine Learning Approach. *Science of The Total Environment*, *792*, 148406. https://doi.org/10.1016/j.scitotenv.2021.148406.

GLOBE, *Globe Data User Guide*. (n.d.). Retrieved July 23, 2022, from https://www.globe.gov/documents/10157/2592674/GLOBE+Data+User+Guide_v1_final.pdf/863a971d-95c5-4dd9-b75c-46713f019088.

Loh, W. Y. (2011). Classification and Regression Trees. *WIREs Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8.

Murtagh, F. (1991). Multilayer Perceptrons for Classification and Regression. *Neurocomputing*, *2*(5-6), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5.

NASA. (n.d.). *Giovanni*. NASA. Retrieved July 22, 2022, from https://giovanni.gsfc.nasa.gov/giovanni/.
National Centers for Environmental Information. (2022, May 11). *Washington D.C. Precipitation*. Retrieved July 27, 2022, from https://www.weather.gov/media/lwx/climate/dcaprecip.pdf.

Open Data DC. (2021, December 8), *Mosquito Trap Sites*. Retrieved July 22, 2022, from https://opendata.dc.gov/datasets/DCGIS::mosquito-trap-sites/about.

Schneider, J., Greco, A., Chang, J., Molchanova, M., & Shao, L. (2021). Predicting West Nile Virus Mosquito Positivity Rates and Abundance: A Comparative Evaluation of Machine Learning Methods for Epidemiological Applications. https://doi.org/10.1002/essoar.10509422.1.

Schonlau, M., & Zou, R. Y. (2020). The Random Forest Algorithm for Statistical Learning. *The Stata Journal: Promoting Communications on Statistics and Stata*, *20*(1), 3–29. https://doi.org/10.1177/1536867x20909688.

Soh, S., & Aik, J. (2021). The Abundance of Culex Mosquito Vectors for West Nile Virus and Other Flaviviruses: A Time-series Analysis of Rainfall and Temperature Dependence in Singapore. *Science of The Total Environment*, *754*, 142420. https://doi.org/10.1016/j.scitotenv.2020.142420.

Washington, D.C. Topographic Map, Elevation, Relief. Topographic. (n.d.). Retrieved July 22, 2022, from https://en-nz.topographic-map.com/maps/sqll/Washington-D-C/.

West Nile Virus. West Nile Virus. (n.d.). Retrieved July 22, 2022, from https://dchealth.dc.gov/service/west-nile-virus.