Predicting the Volume of Mosquito Borne Diseases Via the AMEA Sensor/Radio Network

Miguel Jose Bueno, James Ervin, Anushka Jain, Owen Luo, Aarnav Tendulkar
Mentors: Dr. Rusty Low, Om Shastri, Cassie Soeffing

Abstract

The A.M.E.A. (AI-Powered Mosquito Environmental Analyzer) Sensor/Radio Network is an effective tool towards predicting volumes of mosquito borne diseases. In this study, data based on light, temperature, and precipitation is fed into a ML model that then processes said data, producing Mosquito Borne Disease, (MBD), predictions for that specific area and climate. Will the data received from the A.M.E.A. Network be sufficient enough in creating accurate predictions in a given location? Using a pair of computers and a transmitter (Raspberry Pi Speaker/Kenwood TS-440) and receiver (Computer microphone/Kenwood R600) unit, each connected to their respective radios, we can demonstrate the ability to transmit data using a novel technique with Slow Scan Television (SSTV) images that can be fed into an ML Model that processes and decodes the data and images to create a prediction on the volume of MBDs in a given area. These MBD predictions may then be compared to GLOBE data in order to assess if our MBD predictions are accurate based on a given location and climate. Our decision forest model had an accuracy of 64.6%, and the neural network Google Teachable Machine image-recognition ML model is able to predict the color of an SSTV image with 100% certainty. Since our AMEA sensor/radio network proved to be extremely accurate, in the future, scientists and public health professionals may better prepare for outbreaks in advance, limiting the amount of cases and casualties from mosquito-borne diseases in communities.

Research Questions

Question 1: Will the A.M.E.A. Network be able to collect quality data in a given location?

- This question focuses on the hardware aspect of the project, ensuring that the sensors and network are able to collect real time climate data in a local environment.

Question 2: Can the data provided by the A.M.E.A. Network be used to accurately predict the volume of mosquito borne disease human infection in a given location?

- This question hones in on the software aspect of the project; it will rely on the data collection sources, radio transmission, and the machine learning models in order to run properly.

Introduction

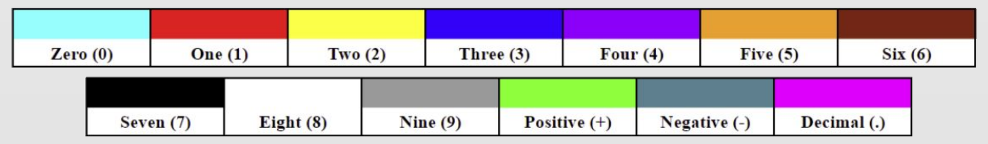
Mosquitos are a type of insect that can be found all over the world, with over 3,000 different species existing on several continents and living in various environments. Mosquito-borne diseases are diseases carried by mosquitoes that can be transmitted through the bite of an affected mosquito. These diseases can negatively affect a variety of species, predominantly humans. Mosquito populations, and thus, mosquito-borne virus rates vary due to many factors, including climate, which vary based on location. Therefore, a location's climate factors are indicative of its volume of mosquito-borne virus infections in humans. Previous projects have utilized sensor networks in order to collect massive amounts of data. In this study, we tested our idea of using sensor networks to input more locally-sourced data into our Decision Forests data to increase its accuracy. We were able to prove that such an idea could work by predicting the color of SSTV images with background noise using a Google Teachable image-recognition neural network ML model. Data was collected from the National Oceanic and Atmospheric Administration (NOAA), which described the temperature (F) and precipitation (in) in relation to the amount of Chikungunya, Dengue, and West Nile Virus cases each month in Austin, Texas. We proceeded with a Decision Forests ML model to predict the amount of total cases given the provided features.

Methodology

The process of data collection and transmission can be broken into five parts: Data Collection/Encoding, Data Transmission, Data Reception, Data Decoding, and Machine Learning (ML) Processing and Analysis.

Collection: During the encoding process, a Raspberry Pi computer captures a picture with its camera and runs a brief script to analyze the amount of brightness in the picture. After processing all pixels of the picture into sounds, it converts the sounds into .wav files, ready to be transmitted into the airwaves.

The number that comes from the analysis of the light levels is also similarly processed but instead of using a webcam picture, each digit/character of the number is converted into a solid color picture that's then converted to a .wav file and stored for transmission.

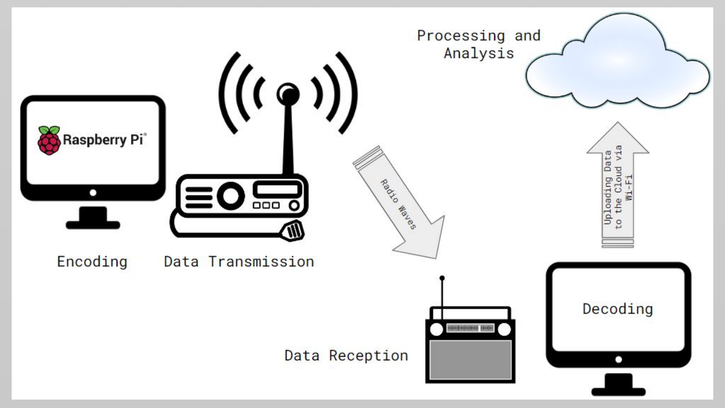


During the transmission process, the stored .wav files are transferred to the transceiver via a sound card. During the reception process, the converted radio waves can be analyzed, processed, and used.

During the decoding process, the decoder software separates the recording into six different chunks. The first chunk is the encoded SSTV image. The second through sixth chunks are the encoded characters of the data. The decoder then runs each one of these files through colaclanth's SSTV Decoder and reverses the process that the encoder did to prepare the images for transmission resulting in usable images. Once all the images have been decoded and stored, the computer then uploads the data to the cloud where our ML model can process it and make predictions of the volume of Mosquito-Borne Diseases in the area where the A.M.E.A. Network has been deployed.

Processing Data from the A.M.E.A. Network
Following image transmission, the images are run through the neural network Google Teachable Machine image-recognition ML model which works to identify colors. As discussed before, a unique color code was set up in which colors correspond to specific digits, as seen below.

The images started off as solid blocks of color before transmission. However, transmission through the airwaves gets obstructed by interference which causes discoloration and choppiness to appear in the images. Because the images are not exact replicas after transmission, a Google Teachable Machine image-recognition machine learning model was created to recognize and distinguish the transmitted color.

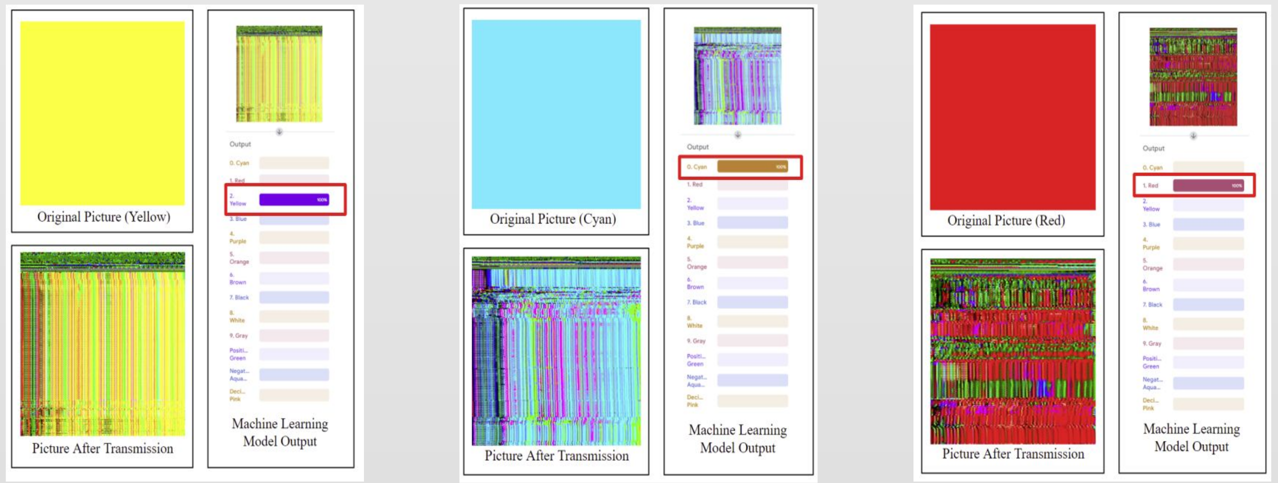


Results

SSTV Image Processing Results

Following image transmission, the images are run through the neural network Google Teachable Machine image-recognition ML model which works to identify colors. As discussed before, a unique color code was set up in which colors correspond to specific digits, as seen previously.

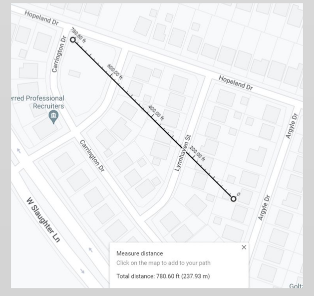
The images started off as solid blocks of color before transmission. However, transmission through the airwaves gets obstructed by interference which causes discoloration and choppiness to appear in the images. Because the images are not exact replicas after transmission, a Google Teachable Machine image-recognition machine learning model was created to recognize and distinguish the transmitted color. In order to train this model, many images for each color had to be uploaded to the training dataset. To create these images, they were transmitted through the radio network multiple times, each time with a different distance between the transmitter and the receiver with varying levels of volume. The resulting images all had unique amounts of interference and were varying examples of what transmitted images could look like in the final project, which is why they made up a good training dataset. These images were then uploaded into the Google Teachable Machine program and linked up to their corresponding digits. For example, the transmitted cyan pictures were uploaded together under the "0" digit. After the model was trained with this dataset, it was able to predict the color of a transmitted image, and effectively convert the images back into a number code.



Transmission Range Results

In order to test the transmission of data and the reception quality, we ran 100 watts of power into a center-fed, five meter dipole using Single Sideband Modulation, or SSB. Testing the reception at 780 ft, we used a Kenwood R600 Receiver with a RadioShack 20-161B Telescoping Scanner Antenna and determined the reception quality to be significantly higher than a 5/9. For reference, the RST is a signal classification system developed by Arthur W. Braaten in 1938. RST stands for Readability, Strength, and Tone and serves as a sort of rubric that Amateur Radio Operators use to grade each other's signals and reception. Most of the time, RST is shortened to RS due to tone being only used by Morse Code operators. Readability is on a scale from one, being unreadable, to five, being perfectly readable. Strength is measured on a scale from one, being faint—signals barely perceptible, to nine, being extremely strong signals. So our result that was deemed to be higher than a 5/9, is a signal that is easily readable, where a computer could easily distinguish the different tones being transmitted, and with an extremely strong signal, where a computer can easily make a high quality image with very little to no interference appearing in its result.

No.	Readability	Signal Strength	Tone (Morse Code)
1	Unreadable	Faint	Harsh tone with hum
2	Barely readable	Very weak	Harsh tone with modulation
3	Readable with difficulty	Weak	Rough tone with hum
4	Almost perfectly Readable	Fair	Rough tone with modulation
5	Perfectly Readable	Fairly good	Wavering tone, strong hum
6	-	Good	Wavering tone, strong modulation
7	-	Moderately strong	Good tone, slight hum
8	-	Strong	Good tone, slight modulation
9	-	Very strong	Perfect tone, no hum or modulation



Machine Learning and Prediction Model Results

We were able to plug in data from the National Oceanic and Atmospheric Administration from Austin, Texas into two ML models designed to predict the volume of mosquito-borne diseases.We used two different ML models. The first one being a linear regression model which is the process of creating a line that best fits the data points given to the trained model and this is the most simple form of a machine learning model. The second model we made is a random forest model that creates a forest full of decision trees that creates different predictions and shows us which decision tree is the most accurate. The Decision Forests model had an accuracy of 64.6%, while the linear regression model only had an accuracy of 37.5%. Decision Forests is far more accurate, so we kept the Decision Forests model and left the linear regression model out. In the future, we would like to add a column for light level from our sensor network. We would like for our own sensor network to feed additional features and respective data values into our current Decision Forests ML model. For now, this portion is unutilized, but it has lots of potential seeing that the neural network Google Teachable Machine image-recognition ML model is able to predict the color of an SSTV image with 100% certainty.

Discussion

Our project was able to track and predict the amount of Chikungunya, Dengue, and West Nile Virus cases. These cases are vector diseases, which means that they are carried by mosquitoes. This can be a big problem for tropical regions as well as areas in the world with poor health infrastructure. In fact, such developing regions have less access to air conditioning, which prompts people to seek the coolness of shaded, well-ventilated areas, precisely the sites where *Ae. aegypti* prefers to feed, directly causing the increase of Dengue cases. [4]

A 100%-certain image-recognition ML model proves that our sensor network would be able to acquire accurate data in the future to predict the amount of mosquito-borne diseases. With sensor-created data, we would be able to obtain more localized data opposed to the ones in city-wide datasets. Moreover, the readability of data at an elevation of 780 ft proves that our sensor network device works at all altitudes. Our 64.6% accuracy level for the NOAA Decision Forests ML Model is still not ideal though, as we don't have much data in the datasource for Austin, Texas. Some possible sources for the low accuracy has to do with the ML model being overfitted with data which made the model accurate only to that data.

Overall, our A.M.E.A sensor network as well as an improved ML model would give scientists and citizens around the world a heads up before a major mosquito-borne outbreak occurs. Based on these findings, in the future, scientists and public health professionals will be better prepared for outbreaks in advance, limiting the amount of cases and casualties from mosquito-borne diseases in communities.

Conclusions

Our project was able to track and predict the amount of Chikungunya, Dengue, and West Nile Virus cases. These cases being vector diseases, meaning they were carried by mosquitos, can be a big problem for tropical regions as well as areas in the world with poor health infrastructure. Our model and sensor network gives scientists and citizens around the world a heads up before a major mosquito-borne outbreak occurs. This enables citizens around the world to protect themselves from possible mosquito-borne viruses by knowing general data trends around their area.

Based on the current 100% certainty of the color-identifying model and the 64.6% accuracy of the prediction machine learning model, we are confident that we can continue working on this project to make it more accurate and precise. Next steps include adding more data to the training datasets and adding more climate factors such as humidity. If these are implemented, the A.M.E.A. network will be much more functional and advantageous. Based on these findings, scientists and public health professionals in the future may better prepare for outbreaks in advance, limiting the amount of cases and casualties from mosquito-borne diseases in communities.

Working with project mentors has allowed us to have multiple points of view regarding our research, authorizing us to make sensible and required changes to our research methods and experiment protocols.

Bibliography

[1] "How to capture a image from webcam in python?" GeeksforGeeks. (2021, December 21). Retrieved July 25, 2022, from https://www.geeksforgeeks.org/how-to-capture-a-image-from-webcam-in-python/

[2] "Kenwood TS-440S Rear Panel." Kenwood TS-440s Rear Panel, Universal Radio Inc., https://www.universal-radio.com/catalog/hamhf/ts440sr.html.

[3] Polineni S, Shastri O, Bagchi A, Gnanakumar G, Rasamsetti S, Sundaravadivel P. MOSQUITO EDGE: An Edge-Intelligent Real-Time Mosquito Threat Prediction Using an IoT-Enabled Hardware System. Sensors. 2022; 22(2):695. https://doi.org/10.3390/s22020695

[4] Reiter P. Climate change and mosquito-borne disease. Environ Health Perspect. 2001 Mar;109 Suppl 1(Suppl 1):141-61. doi: 10.1289/ehp.01109s1141. PMID: 11250812; PMCID: PMC1240549.

[5] Texas Department of State Health Services. "DSHS Arbovirus Weekly Activity Reports." *Texas Department of State Health Services*, https://dshs.texas.gov/idcu/disease/arboviral/westNile/reports/weekly/.

[6] Hsu, Angel, et al. "Next-Generation Digital Ecosystem for Climate Data Mining and Knowledge Discovery: A Review of Digital Data Collection Technologies." *Frontiers*, Frontiers, 1 Jan. 1AD, https://www.frontiersin.org/articles/10.3389/fdata.2020.00029/full.

[7] Ghafouri-Fard, Soudeh, et al. "Application of Machine Learning in the Prediction of COVID-19 Daily New Cases: A Scoping Review." *Heliyon*, Elsevier, Oct. 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8503968/.

[8] "Electromagnetic Frequency, Wavelength and Energy Ultra Calculator." *1728.Org*, https://www.1728.org/freqwave.htm.

[9] "HF Frequency Band." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., https://www.britannica.com/technology/HF.

[10] Poole, I. (1999, November). Radio waves and the ionosphere - ARRL. ARRL. Retrieved July 28, 2022, from https://www.arrl.org/files/file/Technology/pdf/119962.pdf

[11] Mbaabu, Onesmus. "Introduction to Random Forest in Machine Learning." Section.io, https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

[12] US Department of Commerce, NOAA. *Climate*, NOAA's National Weather Service, 3 Mar. 2022, https://www.weather.gov/wrh/climate?wfo=ewx.

IVSS Badges

I am an Engineer:

Engineering is defined as being the design, construction, and modification of the end product with the end goal of meeting the expectations and requirements for a task. In this project, we combined RF, Electrical, Software, AI, and Computer Engineering in order to create a network of computers connected with radio transmitters and receivers that feeds the data collected in the field into a predictive model designed to prevent the contraction of harmful, mosquito-borne diseases.

I Make an Impact:

Our project was able to track and predict the amount of Chikungunya, Dengue, and West Nile Virus cases. These cases are vector diseases, which means that they are carried by mosquitoes. This can be a big problem for tropical regions as well as areas in the world with poor health infrastructure. Our model and sensor network gives scientists and citizens around the world a heads up before a major mosquito-borne outbreak occurs. Based on these findings, in the future, scientists and public health professionals may better prepare for outbreaks in advance, limiting the amount of cases and casualties from mosquito-borne diseases in communities.

I am a Data Scientist:

We input reputable data from the National Oceanic and Atmospheric Administration into a dataset. We realized that the Decision Forests model was far more accurate than the linear regression model, so we decided to use this. We realize that we are limited in the amount of data that we currently have, but the fact that our model is constantly improving shows that our model works in the long run. We also created a neural network ML model to predict the color of an SSTV image, showing that our sensor network works.