# A Hutchinson-Based Approach to Mosquito Modeling: Predicting Mosquito Threat with Machine Learning and Edge Computing

**Mosquito Mappers, NASA SEES 2021 | By Avi Bagchi, Govind Gnanakumar, Shyam Polineni, Sujay Rasamsetti, Om Shastri, Gianna Yan, Spencer Burke**

ArcGIS Dashboard

## INTRODUCTION

Mosquitoes are one of the world's most dangerous organisms, spreading deadly diseases like malaria, Dengue, and Zika. They've spread to nearly every continent, and are only further increasing their range as a result of the extreme weather conditions caused by climate change. The ability to identify mosquito hotspots (areas of high mosquito density) can be especially valuable in preventing the spread of mosquitoes and the diseases they carry. Species distribution models (SDMs) that use climate variables to make binary predictions are effective tools for niche prediction in current and future climate scenarios. Thus, we set out to find an answer to our research question:

*How can we use climate and citizen science mosquito data to develop a machine learning algorithm that can predict mosquito hotspots?*

## DATA COLLECTION & MACHINE LEARNING MODEL

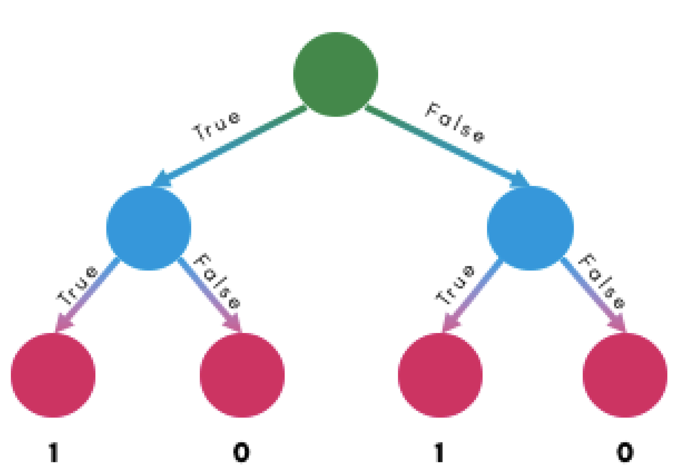
Fundamental ideas provided by **Hutchinson's Niche** postulates that a mosquito's niche is defined solely by environmental variables.

$$A = \begin{bmatrix} a_1 & b_1 & c_1 & d_1 & e_1 \\ a_2 & b_2 & c_2 & d_2 & e_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_n & b_n & c_n & d_n & e_n \end{bmatrix}$$

We defined a Hutchinson hypervolume with temperature, humidity, air pressure, precipitation, and cloud cover climate vectors collected from the National Oceanic and Atmospheric Administration that we matched to mosquito presence and absence points extracted from GLOBE Observer and the National Ecological Observatory Network.

$\overrightarrow{Y_1} = \{a_1, b_1, c_1, d_1, e_1\}$ *defines one mosquito observation*

$\overrightarrow{C_1} = \{a_1, a_2, a_3 \cdots a_n\}$ *defines temperature*
$\overrightarrow{C_2} = \{b_1, b_2, b_3 \cdots b_n\}$ *defines humidity*
$\overrightarrow{C_3} = \{c_1, c_2, c_3 \cdots c_n\}$ *defines air pressure*
$\overrightarrow{C_4} = \{d_1, d_2, d_3 \cdots d_n\}$ *defines precipitation*
$\overrightarrow{C_5} = \{e_1, e_2, e_3 \cdots e_n\}$ *defines cloud cover*




We used a **Random Forest** model, an ensemble machine learning algorithm that utilizes decision trees. Because our output is determined via binary classification, the result is given as a number between 0 (mosquito absence) and 1 (presence).

**Figure 4: Gini Impurity**

$$G = 1 - \sum_{i=1}^{C} p(i)^2$$

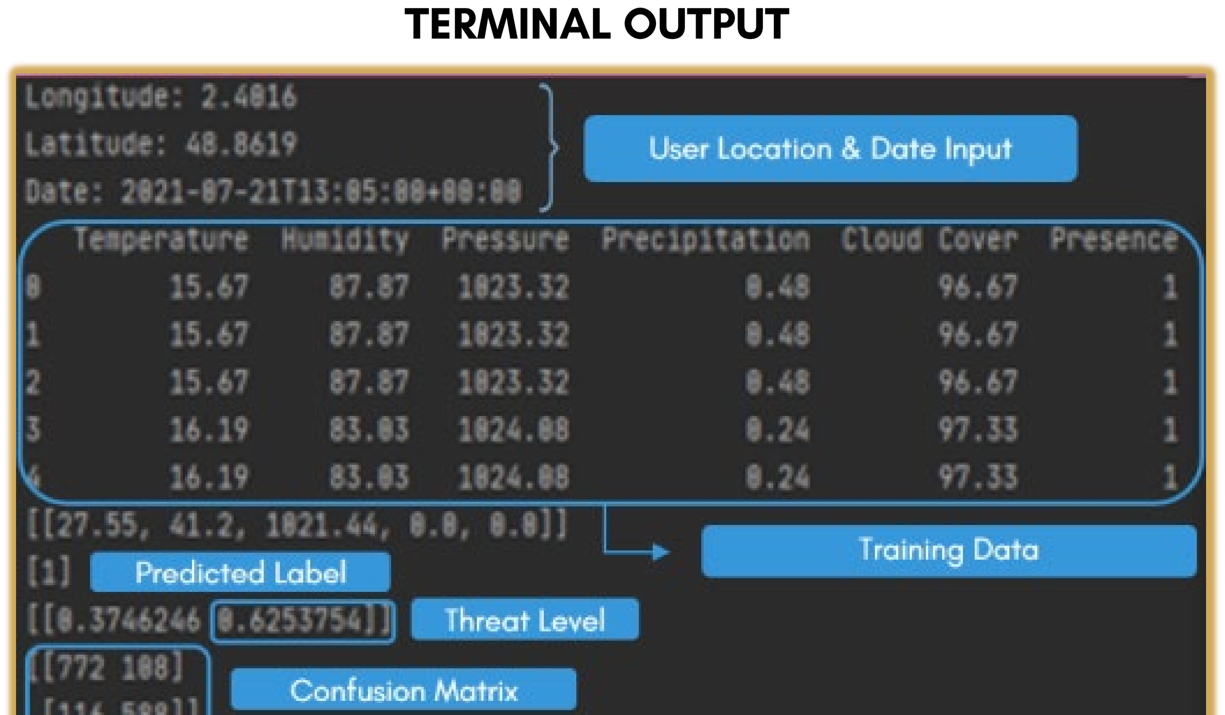
**Gini Impurity** can be defined more simply as the probability of incorrectly classifying an element. The model seeks to minimize Gini impurity and achieve a more homogeneous subset at the end node.

$C = Number\ of\ classes$

$p(i) = Probability\ of\ selecting\ class\ i$

$G = gini\ impurity$

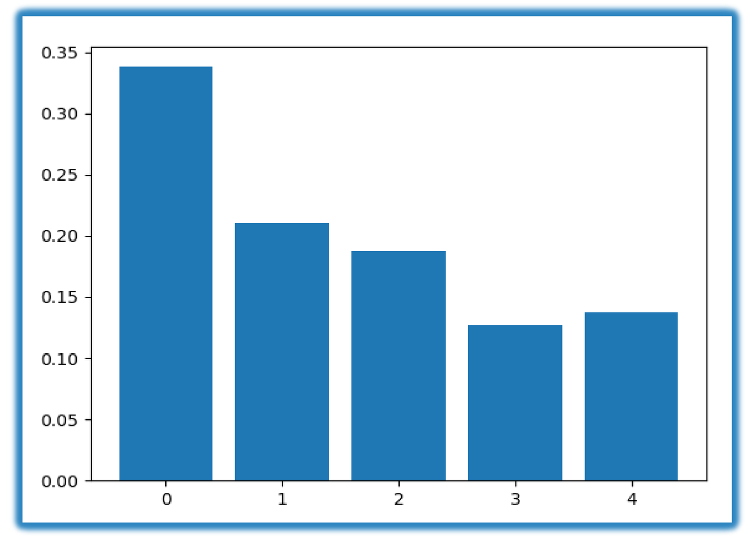
### TERMINAL OUTPUT




## RESULTS

Our model has an **86% accuracy.** This is highly accurate considering the vast number of variables accounted for in the ecological niche model. Given a location and date input, the model produces a threat level based on the number of decision trees that vote for a presence label.

### FEATURE IMPORTANCE CHART




Most influence: **temperature**

- Suggests mosquito presence/absence is strongly correlated with high/low temperatures.

Least influence: **precipitation**

- Consistent with ecological literature, suggests that mosquitoes are attracted to standing water rather than precipitation itself. Mosquitoes prefer humid environments.

The feature importance chart and regression shows a positive, linear correlation between humidity and mosquito threat, as well as between temperature and threat below a threshold of 28°C. In accordance with the aforementioned statistical analysis, we found high threat clusters in warm, humid regions.

### RELATIONSHIP BETWEEN THREAT LEVEL & CLIMATE VARIABLES


1. Some association between *temperature and model output.* As temperature increases, threat level increases as well. Threat level rapidly decreases past a threshold of 28°C.

2. Moderate, linear association between *humidity and threat.* Consistent with feature importance chart and ecological literature, suggesting that mosquitoes prefer humid climates.

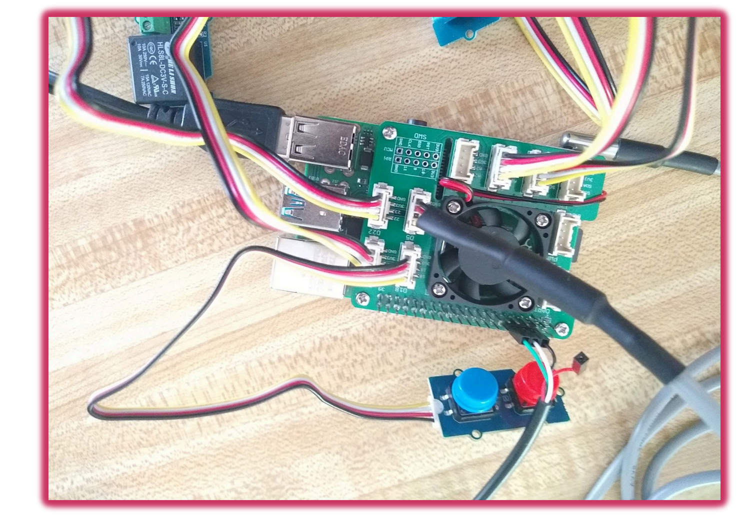
3. Weak, linear association between *pressure and threat.* This finding may be by chance since there's no ecological support for why mosquitoes would prefer regions with high air pressure.

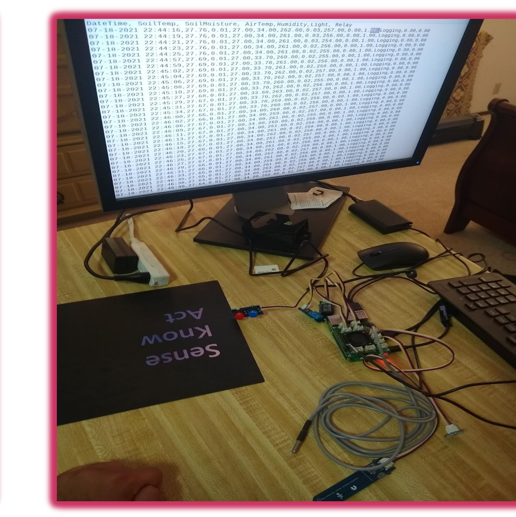
4. No association between *precipitation and threat* Supports findings that mosquitoes are attracted to standing water rather than precipitation itself.

5. No association between *cloud cover and threat.* Despite cloud cover being somewhat related to temperature and precipitation, there's no relationship between cloud cover and threat level.

## EDGE COMPUTING

We built a device leveraging GPS smartphone technology and the IoT to collect and analyze data on the edge, which allows for users to obtain a real time threat level in remote areas without cloud connectivity.



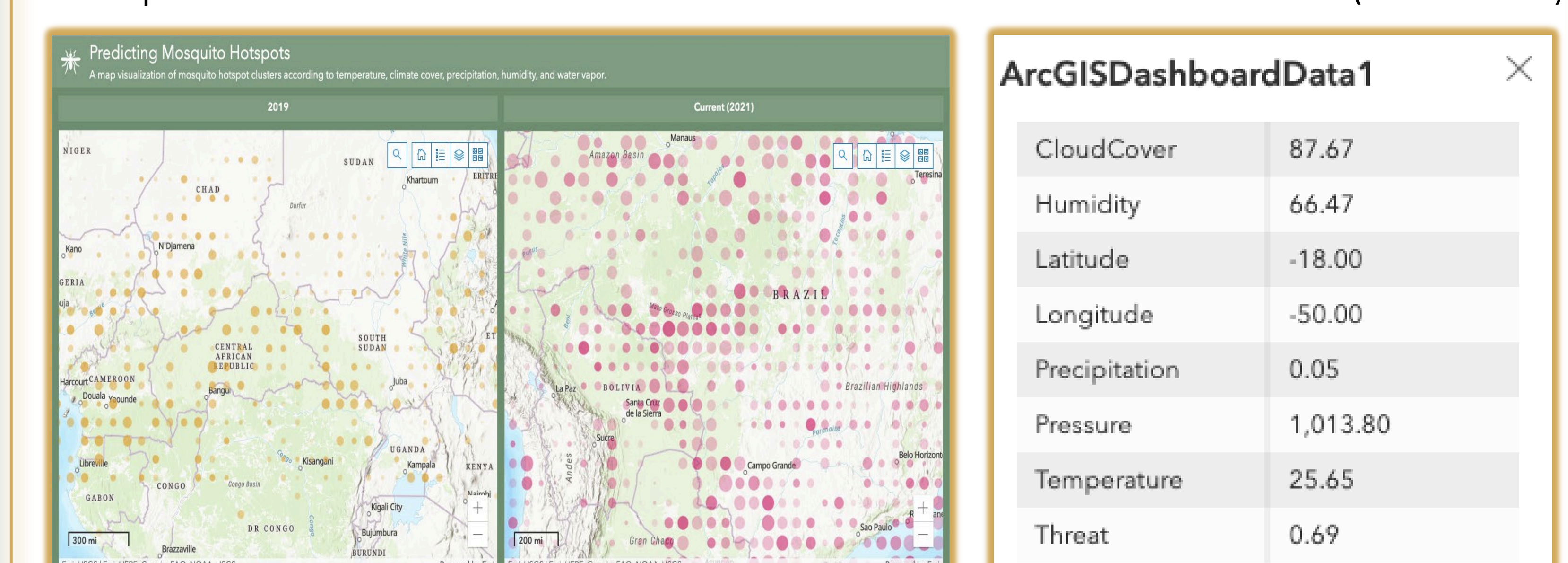


1. Utilized Raspberry Pi 4 to configure climate sensors to collect real time data
2. Analyzed and formatted data, sent data to PC over serial.
3. Ran model using and outputted a threat level.

This localization of data collection and analysis allows for increased speed and removes the need for WiFi access. Pairing these capabilities with a low device cost could be useful for deployment in developing countries! We're also currently developing a TensorFlow Lite model (a low footprint format) that removes the need for a PC entirely.

## ArcGIS DASHBOARD

Mosquito threat-level visualization on an interactive ArcGIS dashboard (2019-2021)




- Insignificant changes between 2019 and 2021 maps as climate change provides little change in mosquito demographic during a two year gap
- Low mosquito threat levels in Sahara region due to low mosquito population growth in dry regions
- High mosquito threat levels in hot and humid areas in South America and Asia

## CONCLUSION

With temperature, humidity, atmospheric pressure, precipitation, and cloud cover data, the model has 86% accuracy

- Moderate, linear association between humidity and threat
- Weak, linear association between pressure and threat
- Our findings coincide with existing theories that mosquitoes are heavily influenced by climatic conditions
- The model has made significant statistical findings paired with the utilization of thousands of data points from NEON, Storm Glass API, Globe Observer

## FUTURE APPLICATIONS

- Adding more environmental features (e.g. normalized vegetation difference index and human population density)
- Increasing model's accuracy by tailoring it to specific mosquito species