Method for Effective Mosquito Data Classification to Identify Potential Hosts of Malaria with AI Implications

AJ Caesar, Walker Gaines, Rahul Gajendran, Tejas Ram, Aaron Lee, Obumneme Nwosu, Angelina Richter

Abstract

The majority of Earth's mosquito-borne illnesses are transmitted by mosquitoes in one of three genera: *Anopheles*, *Aedes*, and *Culex*. However, a special concern is reserved for *Anopheles* mosquitoes for their unique ability to carry and transmit Malaria, a disease that, according to WHO, infects more than 200 million and kills over 500,000 humans annually (*Malaria*, 2022). While it is most prevalent in Africa, Southeast Asia, and Central America, Malaria could soon spread to northern and southern latitudes with a changing global climate. Therefore, it is important to track the extent of the *Anopheles* range and identify any changes that could have detrimental consequences for public health. One way this can be done is using the GLOBE Observer Mosquito Habitat Mapper (MHM) app which allows global users free access to photograph mosquito larvae, attempt to identify their genus, and upload said images to a global database which records the location at which they were taken. While such citizen science is extremely helpful for mosquito research, it can be difficult for citizens with minimal training to properly classify the genus of their discovered larva. A large portion of mosquito photos uploaded to the GLOBE MHM database are either unidentified or misidentified. Therefore, the goal of this research paper is to devise and assess a manner in which the MHM database can be properly classified to create an accurate dataset with all *Anopheles* larvae photos classified by their proper genus. Besides being a vector of Malaria, another unique characteristic of the *Anopheles* mosquito is the absence of a siphon, so by scanning for this trait among MHM larvae photographs and noting positive matches, researchers created a dataset of mosquito larvae that could become vectors of Malaria as adults (Image Reference #1). This data set could then be used to train AI models utilizing Convolutional Neural Networks (CNN) or Vision Transformers (ViT) to classify the MHM database autonomously in the near future.

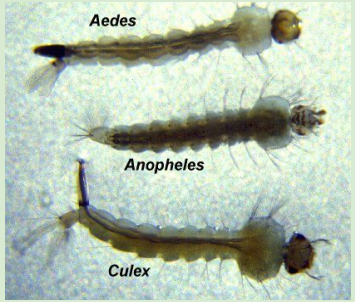


Figure 1. Larval shapes for *Aedes*, *Anopheles*, and *Culex* genus mosquitoes

Figure 2. Left: siphon-less mosquito larva; Right: siphoned mosquito larva

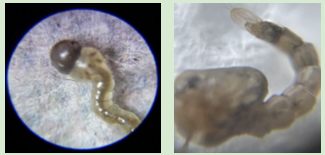
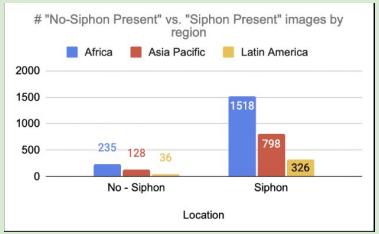
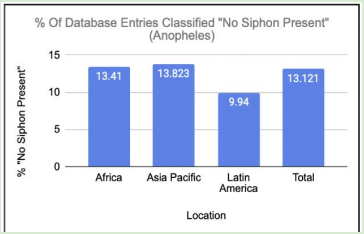


Figure 3. Indeterminate larvae photos (Left: no siphon visible; Right: mosquito in pupa phase





Results

A total of 3041 classified "No Siphon Present" and "Siphon Present" images were uploaded to final dataset, which means that over 10,000 GLOBE MHM photos from within these regions were discarded because they did not show tail section, presented ambiguous evidence , had poor image quality, were duplicate images, or were pupae.

Discussion

- MHM data is idealistic
- Low percentage of no-siphon presenting mosquitos may not accurately match the actual genera breakdown within regions
- Training team should choose selective entries from within dataset to eliminate bias from AI Model
- More equal breakdown to minimize bias towards either classification

Conclusion

Asia-Pacific and Africa MHM data demonstrate a higher concentration of Anopheles larvae, though the concentration in all three regions is still <15%. GLOBE MHM data is idealistic so a low percentage of no-siphon presenting mosquitos may not accurately match the actual genera breakdown within these regions. Overall, this dataset can be used to accurately train an AI model to identify mosquitos with and without siphons, though a training team should choose selective entries from within the dataset to create their train data with a more equal breakdown to minimize bias Additionally, since all mosquito pupae do not have a siphon, simply determining siphon vs no siphon from a GLOBE MHM photograph may not accurately determine if the larvae is a Malaria vector or not, so pupae vs larvae must also be investigated

Methods

- 15,000 GLOBE MHM Photographs sourced from Asia Pacific, Africa, and Latin America regions
- Absence of a siphon = Anopheles mosquito larvae
- Researchers reviewed photo
- Indeterminate photos not used
- No siphon present and siphon present photos uploaded to dataset
- Classifications of other researchers double-checked
- Photos presenting siphon (non-Anopheles) and not presenting siphon (Anopheles) counted
- Relative quantity within each region determined using summation, fractional, and percentage models

Acknowledgements

SEES Earth System Explorer mentors; Dr. Rusanne Low, Ms. Cassie Soeffing, Mr. Peder Nelson, Dr. Erika Podest, Andrew Clark, Matteo Kimura, Kellen Meymarian. The material contained in this poster is based upon work supported by the National Aeronautics and Space Administration (NASA) cooperative agreements NNX16AE28A to the Institute for Global Environmental Strategies (IGES) for the NASA Earth Science Education Collaborative (NESEC) and NNX16AB89A to the University of Texas Austin for the STEM Enhancement in Earth Science (SEES). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Research Questions

Question 1: How can scientists manually classify photos of *Anopheles* mosquito larva within the GLOBE MHM database in order to create a dataset with AI capabilities?

Question 2: What will our classification results show about the relative quantity of *Anopheles* mosquitoes in the Africa, Asia-Pacific, and Latin America regions within this data set?

Bibliography

Climate Change and Malaria - A Complex Relationship | United Nations. (n.d.). the United Nations. Retrieved July 21, 2022, from https://www.un.org/en/chronicle/article/climate-change-and-malaria-complex-relationship

Identify specimens. (n.d.). GLOBE Observer. Retrieved July 21, 2022, from https://observer.globe.gov/documents/19589576/8e9ea31-ac81-48ff-82aa-8d07a86babc2

Lindsey, R., & Dahlman, L. (2022, June 28). *Climate Change: Global Temperature*. NOAA Climate.gov. Retrieved July 21, 2022, from https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

Malaria. (2022, April 6). WHO | World Health Organization. Retrieved July 21, 2022, from https://www.who.int/news-room/fact-sheets/detail/malaria

Mosquito Life Cycle | US EPA. (2022, March 8). Environmental Protection Agency. Retrieved July 24, 2022, from https://www.epa.gov/mosquitocontrol/mosquito-life-cycle

Reiter, P. (n.d.). *Climate change and mosquito-borne disease. - PMC*. NCBI. Retrieved July 21, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240549/

Shahinfar, S., Meek, P., & Falzon, G. (2020, May). "How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Science Direct*, *57*. https://www.sciencedirect.com/science/article/abs/pii/S1574954120300352

Terezinha, S. (2017). *How to identify Culex, Anopheles and Aedes mosquitoes and their larvae?* Researchgate. Retrieved July 23, 2022, from https://www.researchgate.net/post/How_to_identify_Culex_Anopheles_and_Aedes_mosquitoes_and_their_larvae/58a41eedf7b67e73463c4a89

Tusting, L. (2014, 2 1). Larval Source Management: A supplementary Measure for Malaria Control. *Outlooks on Pest Management*, *25*. https://www.researchgate.net/publication/263420029_Larval_Source_Management_A_Supplementary_Measure_for_Malaria_Control