Autonomous Mosquito Habitat Detection using Satellite Imagery and Convolutional Neural Networks for Disease Risk Mapping

Sriram Elango and Nandini Ramachandran
NASA STEM Enhancement in Earth and Space Science (SEES) Intern Program

Abstract

Mosquitoes are known vectors for disease transmission that cause over one million deaths globally each year. The majority of natural mosquito habitats are areas containing standing water such as ponds, lakes, and marshes. These habitats are challenging to detect using conventional ground-based technology on a macro scale. Contemporary approaches, such as drones, UAVs, and other aerial imaging technology are costly when implemented. Multispectral imaging technology such as Lidar is most accurate on a finer spatial scale whereas the proposed convolutional neural network(CNN) approach can be applied for disease risk mapping and further guide preventative efforts on a more global scale. By assessing the performance of autonomous mosquito habitat detection technology, the transmission of mosquito borne diseases can be prevented in a cost-effective manner. This approach aims to identify the spatiotemporal distribution of mosquito habitats in extensive areas that are difficult to survey using ground-based technology by employing computer vision on satellite imagery for proof of concept. The research presents an evaluation and the results of 3 different CNN models to determine their accuracy of predicting large-scale mosquito habitats. For this approach, a dataset was constructed utilizing Google Earth satellite imagery containing a variety of geographical features in residential neighborhoods as well as cities across the world. Larger land cover variables such as ponds/lakes, inlets, and rivers were utilized to classify mosquito habitats while minute sites such as puddles, footprints, and additional human-produced mosquito habitats were omitted for higher accuracy on a larger scale. Using the dataset, multiple CNN networks were trained and evaluated for accuracy of habitat prediction. Utilizing a CNN-based approach on readily available satellite imagery is cost-effective and scalable, unlike most aerial imaging technology. Testing revealed that YOLOv4 obtained greater accuracy in mosquito habitat detection than YOLOR or YOLOv5 for identifying large-scale mosquito habitats. YOLOv4 is found to be a viable method for global mosquito habitat detection and surveillance.

Intro

Vector-borne diseases, such as malaria, dengue fever, and Zika virus, are a major global health concern. These diseases are spread by mosquitoes, and they cause over 1 million deaths each year. The majority of these deaths occur in underdeveloped nations with poor medical infrastructure. Modern approaches to mapping mosquito habitats are extremely costly. For example, drones and UAVs can be used to map the distribution of mosquitoes, but this approach is very expensive. As a result, these nations are even more vulnerable to deadly diseases.Data derived from multispectral imaging technology from satellites also takes time to further interpret and analyze. This can delay the response to a possible epidemic in a region. Due to these approaches being the only method of mosquito habitat identification, there are no current efficient methods available for tracking and controlling epidemics caused by mosquitoes. As a result, we propose and utilize the approach of artificial intelligent convolutional neural networks. Our AI network is trained on satellite imagery to identify where mosquito habitats are when given a satellite image. By doing this, we can predict where photos of mosquitoes will appear within milliseconds. This allows us to map vast areas quickly and efficiently.

Methodology

Our study analyzed the results of mosquito habitat prediction data of 3 different CNN models (YOLOv4, YOLOv5, YOLOR). There are 4 steps involved in this process: Prepare, Train, Validate, and Test.

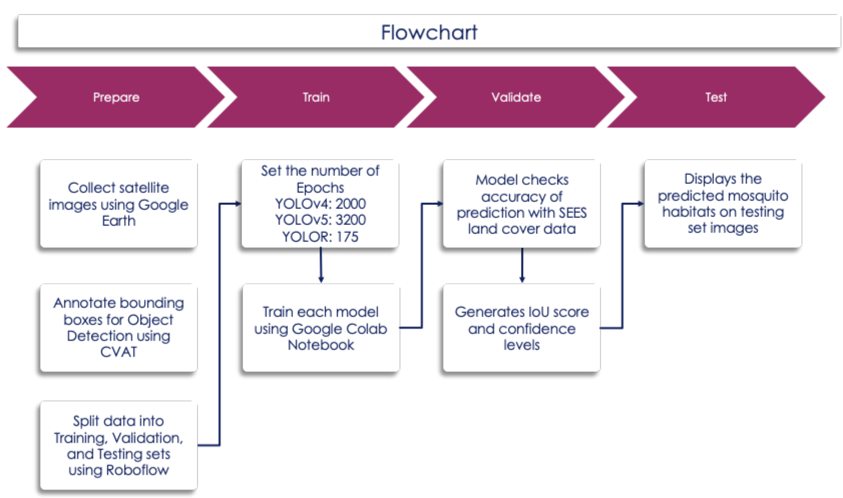
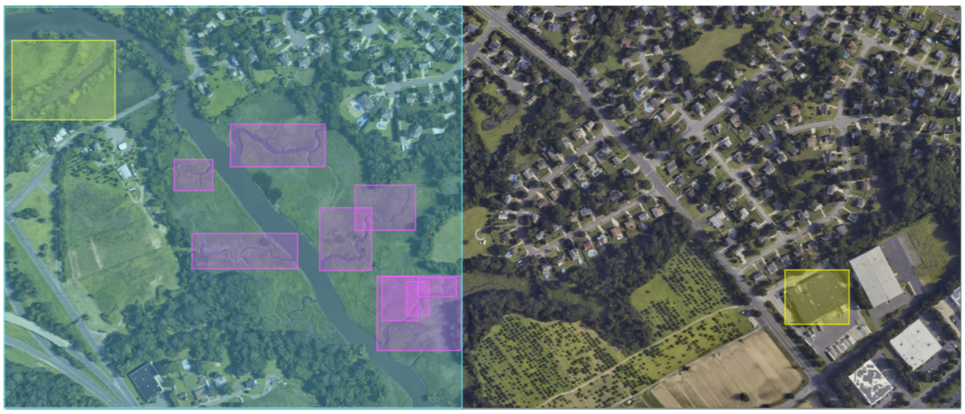
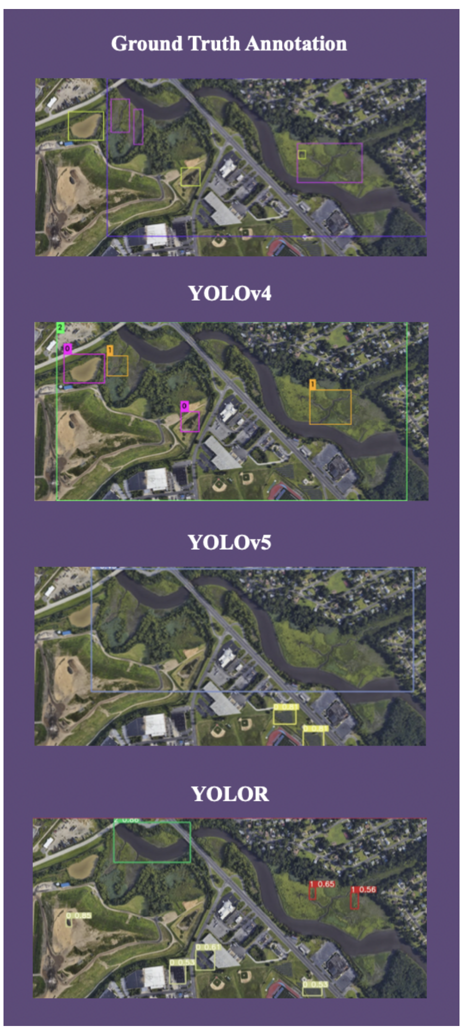


Figure 1: Model Flowchart and CNN Framework

Mosquito habitats were classified in the following groups: Ponds/Lakes, River Inlets, Rivers. The annotated dataset of about 500 images was divided into a 62-16-25 % Train/Test split ratio using Roboflow. The images were exported using each model's respective frameworks (YOLOv4 - Darknet, YOLOv5 & YOLOR - TensorFlow). The satellite output data was then analyzed for accuracy of mosquito habitat detection and compared with ground truth annotations.

Sample Output





Results

Through analysis and experimentation utilizing the different convolutional networks and the corresponding frameworks, it was found that YOLOv4 performed with extreme accuracy and precision, identifying most mosquito habitats with extreme speed.

Conclusion

Compared to other Aerial approaches for mosquito habitat detection, CNNs are the quickest, most efficient, and cost effective way to go. YOLOv4 performed with the highest accuracy among the 3 models with an average IoU score of 55.85%. The proposed solution can be implemented on a global scale using readily available satellite data. It can be used to aid in preventative measures of the global transmission of mosquito-borne vector diseases through risk mapping as well as integrated in public health policies. The spatiotemporal distribution of mosquito habitats can additionally serve to map impoverished and hard to reach areas in order to determine mosquito migration patterns. Hotspots can be linked with the effects of climate change and variables such as soil moisture, temperature, and land use. There are a multitude of uses for mosquito habitat distribution data, and CNNs are a cost- effective, autonomous approach to achieving this.

Limitations

Micro scale habitat data like smaller sites such as puddles, footprints, and tires tend to go unnoticed by this model as its scope is geared towards dealing with larger bodies of standing water. It also does not take into account nearby vegetation and weather-related data that may have an impact on breeding patterns and female oviposition. Future research could include an integration of such factors mentioned above to narrow the amount of viable mosquito habitats.

International Virtual Science Symposium Badges

Badges Applied For:

Data Science: This badge is being applied for due to the various data collected from a diverse array of geographic locations. We utilized these various data for the purposes of training our CNN models in order to predict mosquito habitat locations.

Engineer: This badge is being applied for as a result of utilizing transfer leading on the YOLOV4 model structure in order to obtain a model capable of identifying mosquito habitats in the region.

Impact: This badge is being applied for as a result of our model being the first to detect mosquito habitats using satellite imagery, thus enabling global mosquito habitat monitoring and therefore contributing to the active prevention of mosquito epidemics worldwide.