Using Deep Learning Frameworks For Mosquito Oriented Land Classification

Ananth Madan
NASA SEES Internship: Mosquito Mapper Project

Kuleen Sasse
NASA SEES Internship: Mosquito Mapper Project

Lawrence Wu
NASA SEES Internship: Mosquito Mapper Project

Saravana Polisetti
NASA SEES Internship: Mosquito Mapper Project

Lucy Xie
NASA SEES Internship: Mosquito Mapper Project

Abstract

Mosquito-borne illnesses are a growing risk as the climate changes. We need better ways to deal with this ever-increasing problem with better quality models describing mosquito breeding sites on the landscape. This project used a 10 x 10-meter square to classify land into 13 mosquito-oriented land classes. These classes focus on different water types and vegetation where mosquitoes are known to oviposit. We utilized a custom program to take screenshots of the plot data from Collect Earth Online, an open-source land cover classification tool and processed the images using the OpenCV library to filter out the plots' lines and dots. Our data set was 8000 pictures divided into a 75, 15, 10 training, validation, and test split. VGG16, InceptionV3, Xception AlexNet models were employed to classify our data and compared the results. VGG16 had an accuracy of 20 percent, AlexNet 11 percent, Xception 31 percent, and InceptionV3 with 21 percent. The results from these experiments had low accuracy. However, the work suggests a compelling proof-of-concept that Deep Learning tools have a role in processing land cover data for inclusion in risk models of mosquito vector-borne disease.

Introduction

Mosquitoes are a growing problem in the world. As temperatures rise, mosquito seasons increase which increases the risks of going outside. The only way to combat this problem is to try and predict the areas of great risk and take action. One of the great factors in determining mosquito risk is the land cover (Vanwambeke et al., 2007). Deep learning provides a way to quickly classify land cover. We chose to approach this problem as an image classification problem where the computer classifies the entire picture.

Methods

Gathering Data

For our land classification, we divided our classes into 13 distinct classes. This classification system is more focused on mosquitoes as it divides water and vegetation into different types as mosquitoes have higher chances to breed some but not others.

Our data was collected through a team of interns who chose a center point for a 3.5km by 3.5km area of interest. That large area of interest was sampled with 37 times. 36 of these squares were in an even square formation inside the larger area. The 37th square lied in the center of the larger area. Each sample was a 100 meter square which was further divided into 121 evenly spaced points where the classification is based on the 10 meter by 10 meter area around that point. The interns labelled the points using a website called Collect Earth Online. Collect Earth Online is "a next generation of web-based, crowd-sourcing technology for Earth Science analyses" (Saah et al., 2019).

The data points were captured using a custom web scraper that took photos of the individual points from Collect Earth Online and documented them. The photos were saved as 64 pixel by 64 pixel PNGs. Open source satellite photos were too low resolution to use for our purposes.

Data Overview

A total of 88,000 usable data points were distributed into 13 classes. We removed a few classes as labeling issues were rampant in a few specific classes. We only used about 13,000 of those photos for our dataset as that many would be too computationally intensive for our resources.

The data taken from Collect Earth Online still had lines and the dots from the screenshots. We replaced them with the Gaussian average of the surrounding pixels and resized our photos to 128 pixels by 128 pixels.

Model Overview

We will be testing four models on our task: VGG16, InceptionV3, Xception, and AlexNet (Chollet, 2017; Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014; Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016). All these models are image classification models, perfect for our situation. We wanted to see how well these baselines worked on a new problem. These models do pretty well on classifying large swaths of land. However, we wanted to see if a smaller classification could lead to better results. For our split of the data, we are giving 75% to training,15% to development, and 10% to testing. The models were trained first for 25 epochs on a smaller dataset and then 25 more on a larger, augmented dataset. The InceptionV3, Xception, and AlexNet were trained with 20 batch size using the resources of Google Colaboratory. VGG16 was trained on a GTX 1070 with a 20 batch size.

Table 1
Train, Development, and Test Data Split Total

	Train	**Development**	**Test**
Count	10366	2069	891

Measurements

The model will be evaluated on accuracy which is defined in Equation 1 as

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} * 100 \tag{1}$$

where TP is True Positives, TN is True Negatives, FN is False Negatives and FP is False Positives

Results

Table 2
Model Accuracy on Data Splits

Model	**Accuracy on Train Data**	**Accuracy on Development Data**	**Accuracy on Test Data**
VGG16	91%	86%	20%
AlexNet	87%	85%	11%
Xception	**99%**	**90%**	**31%**
InceptionV3	80%	70%	21%

Discussion

We can see that the models did well on all the data sets except during the testing phase. The testing set shows that each of the models have grossly overfit, but this will be mitigated by a larger, more varied testing set and a stronger early stopping procedure in order to help the model generalize.

Despite not matching to any of the accuracies of previous studies, we find this was a good proof of concept. Our most glaring error was removing the land classifications because of lack of clear labelling guidelines. In addition, we should have used a more robust way to remove the lines and dots. We think that with more time and resources we could push our numbers significantly. We feel that this is a stepping stone towards better land cover data for mosquito modelling.

Acknowledgements

We would like to thank Russanne Lowe, Peder Nelson, Cassie Soeffing, Celena Miller, Margaret Baguio for giving us the opportunity to make a real impact on the world.

The STEM Enhancement in Earth Science (SEES) program is a partnership with NASA Cooperative Agreement NNH15ZDA004C between NASA, The University of Texas at Austin's Center for Space Research, and Texas Space Grant Consortium.

Support for the NASA 2020 SEES Virtual High School Internship program: Mosquito Mappers was sponsored by NESEC, through an award to the Institute for Global Environmental Strategies, Arlington VA. by NASA under Award NNX6AE28A.

References

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1251–1258).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., … San Aung, K., et al. (2019). Collect earth: An online tool for systematic reference data collection in land cover and use applications. *Environmental Modelling & Software*, *118*, 166–171.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2818–2826).

Vanwambeke, S. O., Somboon, P., Harbach, R. E., Isenstadt, M., Lambin, E. F., Walton, C., & Butlin, R. K. (2007). Landscape and land cover factors influence the presence of aedes and anopheles larvae. *Journal of medical entomology*, *44*(1), 133–144.

NASA

TEXAS space grant consortium

CSR

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

Basic Level Coding

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

Intermediate Level Coding

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

Advanced Level Coding

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

CLIMATE SCIENCE BACKGROUND

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Program Manager, Education and Outreach

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

EXPLORING THE EARTH

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach

STEM Enhancement in Earth Science

SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti

For completing the lesson module on

SPACE EXPLORATION, SCIENCE AND ENGINEERING

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach

STEM Enhancement in Earth Science

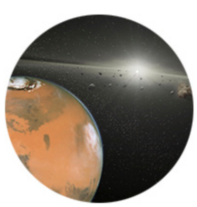
SEES

SUMMER HIGH SCHOOL INTERN PROGRAM

CERTIFICATE OF COMPLETION

is awarded to

Saravana Polisetti



For completing the lesson module on

MARS AND BEYOND

Summer 2020

Dr. Timothy J. Urban
Director, Texas Space Grant Consortium

Margaret Baguio
Senior Program Manager, Education and Outreach