ZzApp Malaria: Learning from satellite images to fight malaria

Summary

With satellite imagery, computer vision is commonly used to identify visible objects, with Convolutional Neural Networks (CNNs) most often being the default choice. Even when the objects themselves are not clearly visible, predictive models using CNNs can sometimes infer the presence of objects based on other characteristics e.g., the region surrounding the object in question. However, as the experience of ZzApp Malaria shows, this is not always the case, and in such cases, traditional techniques such as linear regression may be good enough.

Opportunity: Preventing malaria through treatment of standing water

In 2019, Malaria caused an estimated 409,000 deaths, 94% of which were in Africa and 67% of which were children[1]. Vector control is the main way to prevent transmission (i.e., through bites from malaria-carrying Anopheles mosquitos), and the primary focus of this thus far has been on commodities (e.g., bed-nets) or indoor spraying of insecticides, but these are only partially effective and don't work outdoors. An alternative approach is to treat standing water within the community (where mosquitos breed and multiply) directly, but such programs have not been systematic or comprehensive enough to be effective at scale.

Challenge: Small water bodies not visible in satellite imagery

The difficulty with water treatment lies in identifying standing water so it can be managed. Larger water bodies are easily visible in satellite imagery, and computer vision algorithms have been developed to identify them automatically. Smaller bodies, however, are difficult to detect, even with sophisticated satellite imaging techniques (e.g., infrared), and they can be covered, or only present season to season. If this this standing water could be better located then it would be possible to better direct spraying, and better control the mosquito population.

Solution: Infer presence of standing water through topography

ZzApp Malaria[2] was founded to address this problem and began by looking at malaria hotspots in Sao Tome. They created an app that enabled on-the-ground inspectors to log the location of water bodies encountered, and to track water treatments over time, to establish a training set of positive examples (i.e., locations where standing water was present). They also collected satellite imagery (e.g., photographs, infrared and radar), and used this to train a Convolutional

[1] https://www.who.int/news-room/fact-sheets/detail/malaria

[2] https://www.zzappmalaria.com

Neural Network (CNN)-based object detection algorithm, which performed well for large water bodies, but poorly for small ones, particularly when they are obscured.

As an alternative, the team extracted 50 topographical and image-based features from the images and used these in a traditional linear regression-based approach to determine the likelihood of standing water in each segment of a map. At 75% accuracy, the performance was equivalent to that of the CNN a but enabled much greater transparency into which factors were driving the prediction. The team also found that topographical determinants were highly dependent on locale, and that the linear regression was more transferable to other locales than Neural Network approaches.

Outcome: A transparent and transferable approach to other locales

The relatively high performance of the regression models may have been because it was able to take advantage of characteristics about how water pools, as a function of topological features of the data, as opposed to having to infer them with CNN. In any case, the additional transparency and transferability of the regression model will be essential to ZzApp's ambition to expanding the approach beyond Sao Tome and to other locales (e.g., Ghana, Zanzibar and beyond) where the terrain can differ significantly.