Imperial College London

Mapping malaria prevalence in sub-Saharan Africa with Deep Learning and Satellite Imagery

Iwona Hawryluk[1], E.Semenova[2], X.Miscouridou[1], T.A.Mellan[1], S.Bhatt[1,3]

[1]Imperial College London,[2]University of Oxford, [3]University of Copenhagen

Background

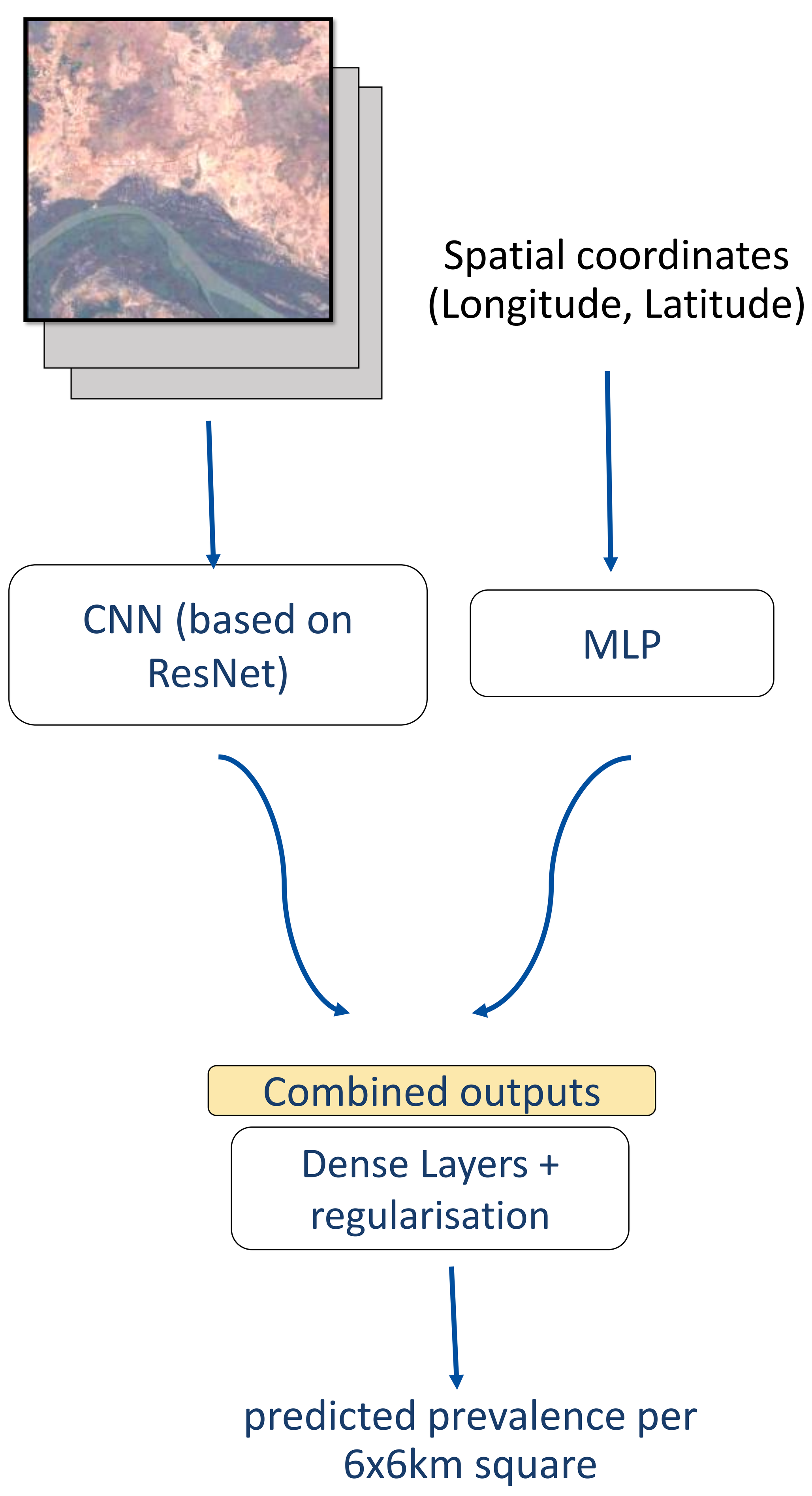
Maps of an infectious disease are integral to guiding public health policy. Typically, prevalence data is collected through time-consuming and resource-intensive household surveys, which cannot be as regular as routinely collected data from health facilities. Statistical models are therefore critical to fill in the data gaps. Although the use of satellite imagery has been a long-standing topic in epidemiological studies, practitioners rely on manually extracting the predictors such as temperature from the image and rarely consider very high resolutions (pixel sizes > 1km).

Aim: Predict the prevalence of malaria across sub-Saharan Africa on a fine spatial scale using raw satellite images and numeric covariates

Data

- Landsat 8 satellite images covering 6x6km squares, resolution 30m per pixel [1]
- Prevalence of P.falciparum infections across over 43k locations in sub-Saharan Africa [2] (Fig. 2)
- Set of climate, socioeconomic, geographical and other covariates of varied resolution [3]

Multi-input model



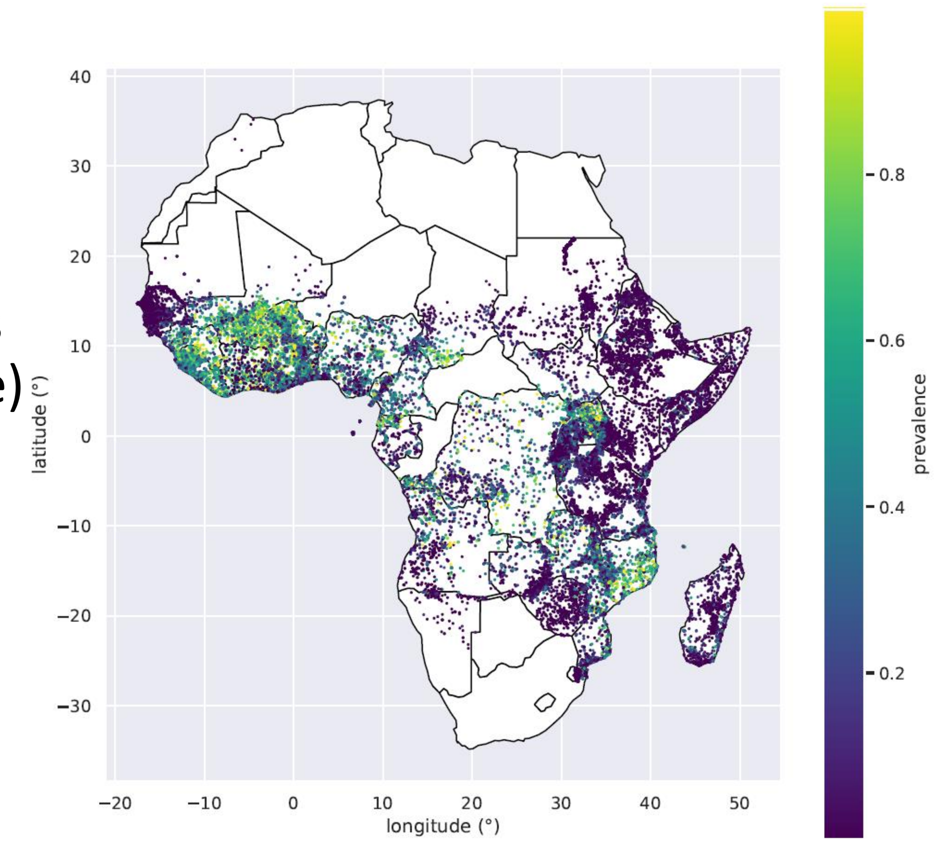


Fig.2 P.falciparum prevalence, used as training data for the model.

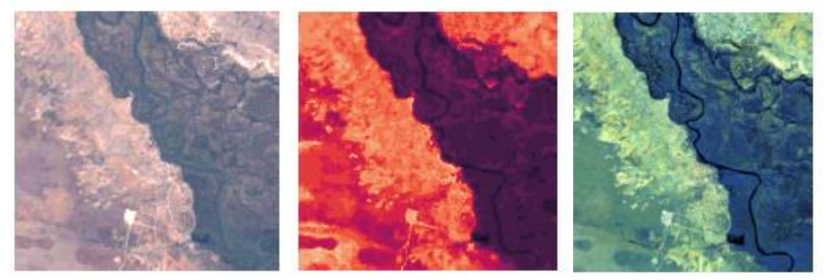
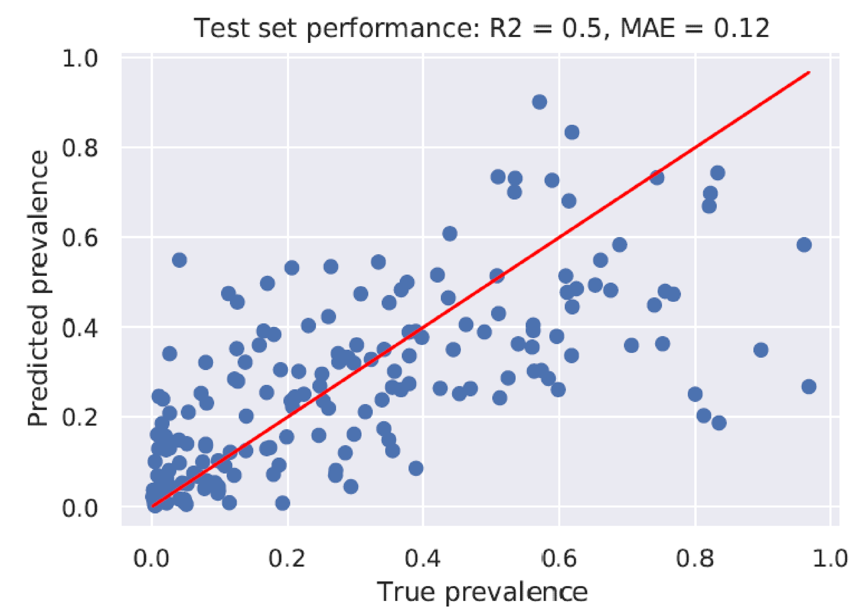


Fig.3 Example of different Landsat 8 bands from one location. From left to right: RGB bands, Infrared bands, Thermal band.

Fig.1 Multi-input Deep Learning framework for predicting malaria prevalence. The left branch of the model takes in a multi-band satellite image and passes it through Convolutional Neural Network (CNN). The right branch takes in a set of numeric covariates and passes it through Multilayer Perceptron (MLP). The outputs of both branches are then combined and passed through additional Dense and Regularising layers. The final layer returns predicted prevalence of malaria.

Results

Fig.4 Performance of the model on the test set (20% of the available data).



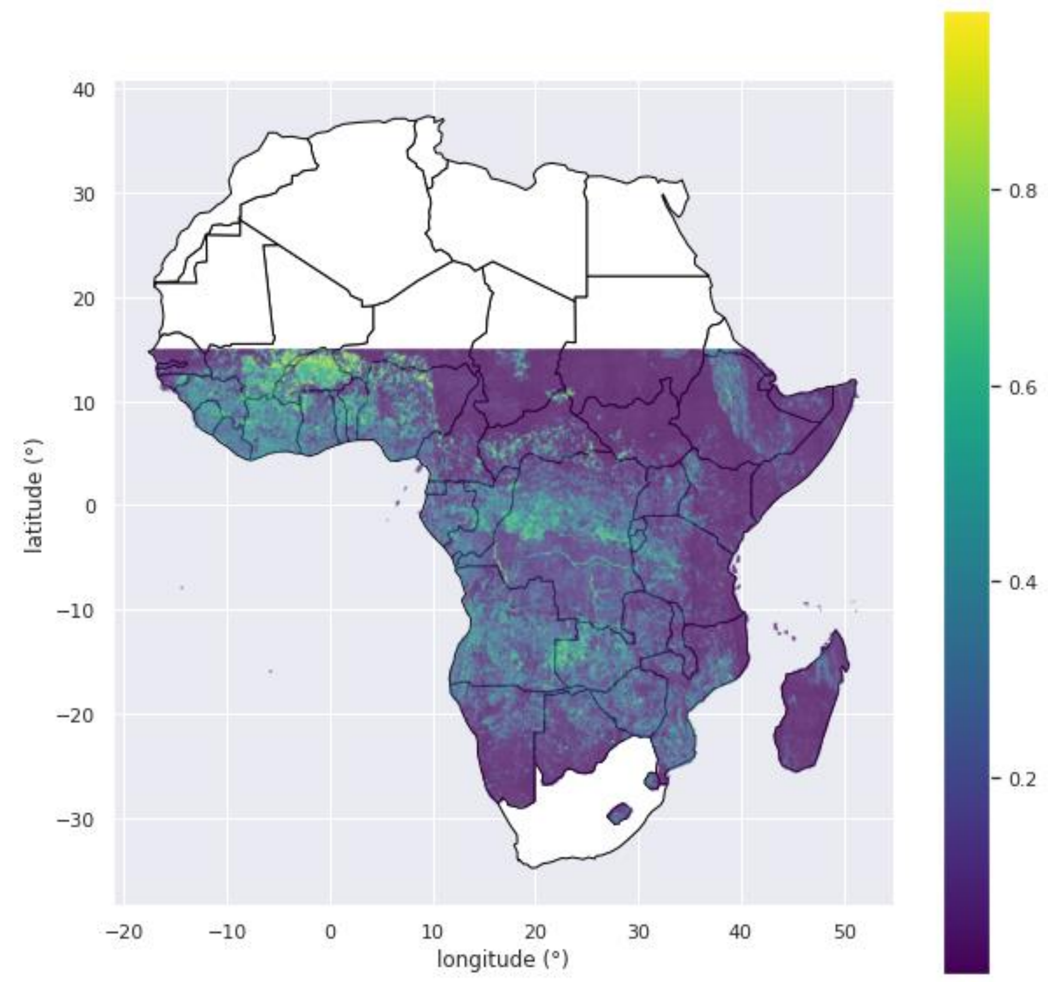


Fig.5 Prevalence of P.falciparum in sub-Saharan Africa predicted by the multi-input Deep Learning Model. The resolution of the predictions is 6x6km.

Discussion

- We presented a novel multi-input framework for predicting malaria, leveraging satellite imagery and numeric covariates and Deep Learning.
- Next steps will involve refining the predictions, adding uncertainty through ensemble models and adding Gaussian Processes to smooth out the predictions.

References

[1] https://landsat.gsfc.nasa.gov/
[2] https://malariaatlas.org/
[3] Bhatt et al. 2015, Nature

Acknowledgements

This project was funded by the Imperial College MRC Centre.

i.hawryluk19@imperial.ac.uk