# Disposition Masterarbeit MAS Data Science: How can multi-modal data fusion inform and enhance the prediction and reporting of malaria outbreaks?

## Context

Traditionally underserved communities in Sub-Saharan Africa and Southeast Asia are disproportionately burdened by tropical diseases such as malaria, dengue, zika, and others. These diseases often thrive in regions with limited healthcare infrastructure and resources, leading to significant morbidity and mortality. What’s more, the timely allocation of adequate healthcare resources, especially during outbreaks in these regions often represents a significant challenge for local healthcare systems.

This paper would like to explore how multi-modal approaches to machine learning might be used to enhance the prediction and reporting of outbreaks of these diseases. By integrating diverse data sources such as incident reports and environmental data – weather data in particular – multi-modal ML approaches could potentially provide a means to better predict regional outbreaks of a disease weeks or months in advance, facilitating timely interventions and improving the planning and allocation of healthcare resources. This technological advancement could ultimately serve to better inform public health strategies and significantly mitigate the impact of tropical diseases in these vulnerable regions.

The idea of predicting outbreaks of tropical disease based on weather patterns is not new. However, with the recent advances in data science there is new potential in vastly improving the accuracy of such predictions, potentially allowing healthcare systems to act earlier and, ultimately, save more lives. In their research, Martineau et al. (2022) established a model which takes into account climate and Malaria incidence data to “show that tropical climatic variability and associated sea surface temperature over the Pacific and Indian Oceans are valuable for predicting malaria in Limpopo, South Africa, up to three seasons ahead”.

This paper will build on the techniques put forth by Martineau et al. and evaluate whether prediction accuracy can be further improved by incoporating additional data (predictors) e.g. satelite imagery, precipitation, or others. As a secondary objective, it would be interesting to expand on this work geographically, i.e. applying the model to other regions in order to assess whether the fundamental approach proves effective globally, or whether the selection of effective predictors is region-specific.

## Method

### Datasets

The following is a non-exhaustive list of datasets which could be valuable in the context of this research.

**Climate data**

* NCEP-NCAR Reanalysis 1 (as used in Martineau et al.): <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>
* PRISM or CPC Global Temperature and Precipitation Datasets & satelite imagery

**Malaria data**

Datasets on Malaria outbreaks is not easy to come by in the public domain. Potential sources to reach out to:

* Malaria case counts in Limpopo province (1998 to 2020) provided by the Malaria Institute in Tzaneen, South Africa (not publicly available, however, may be obtainable by seeking contact)
* Swiss TPH: The Malaria Epidemiology and Control and [Analytics and Intervention Modelling](https://www.swisstph.ch/en/about/eph/health-interventions/analytics-and-intervention-modelling) groups at the TPH may have historical data on malaria outbreaks or be able to provide guidance on where to find it.

**Public Health & Epidemiological Reports**

ProMED-mail

**Additional datasets**

The WHO and CDC publish large datasets on tropical disease incidence at regular intervals.

### Feature Engineering

**Climate Data/Satelite Image Data**

TBD

**Malaria Incidence Data**

TBD

**Modeling & Evaluation**

TBD

### Model Development & Evaluation

TBD

## References

Shenoy S, Rajan AK, Rashid M, Chandran VP, Poojari PG, Kunhikatta V, Acharya D, Nair S, Varma M, Thunga G. Artificial intelligence in differentiating tropical infections: A step ahead. PLoS Negl Trop Dis. 2022 Jun 30;16(6):e0010455. doi: 10.1371/journal.pntd.0010455. PMID: 35771774; PMCID: PMC9246149.

Attai K, Amannejad Y, Vahdat Pour M, Obot O, Uzoka F-M. A Systematic Review of Applications of Machine Learning and Other Soft Computing Techniques for the Diagnosis of Tropical Diseases. Tropical Medicine and Infectious Disease. 2022; 7(12):398. <https://doi.org/10.3390/tropicalmed7120398>

Martineau, Patrick, Swadhin K. Behera, Masami Nonaka, Ratnam Jayanthi, Takayoshi Ikeda, Noboru Minakawa, Philip Kruger, and Qavanisi E. Mabunda. “Predicting Malaria Outbreaks from Sea Surface Temperature Variability up to 9 Months Ahead in Limpopo, South Africa, Using Machine Learning.” *Frontiers in Public Health* 10 (August 25, 2022). <https://doi.org/10.3389/fpubh.2022.962377>.

Nazir, Usman, Muhammad Talha Quddoos, Momin Uppal, and Sara Khalid. “Predicting Malaria Outbreaks Using Earth Observation Measurements and Spatiotemporal Deep Learning Modelling: A South Asian Case Study from 2000 to 2017.” The Lancet Planetary Health 8 (April 1, 2024): S17. <https://doi.org/10.1016/S2542-5196(24)00082-2>.