

# **Predicting dengue transmission rates by comparing different machine learning models with vector indices and meteorological data**

## **ABSTRACT**

Machine learning algorithms (ML) are receiving a lot of attention in the development of predictive models for monitoring dengue transmission rates. Previous work has focused only on specific weather variables and algorithms, and there is still a need for a model that uses more variables and algorithms that have higher performance. In this study, we use vector indices and meteorological data as predictors to develop the ML models. We trained and validated seven ML algorithms, including an ensemble ML method, and compared their performance using the receiver operating characteristic (ROC) with the area under the curve (AUC), accuracy and F1 score. Our results show that an ensemble ML such as XG Boost, AdaBoost and Random Forest perform better than the logistics regression, Naïve Bayens, decision tree, and support vector machine (SVM), with XGBoost having the highest AUC, accuracy and F1 score. Analysis of the importance of the variables showed that the container index was the least important. By removing this variable, the ML models improved their performance by at least 6% in AUC and F1 score. Our result provides a framework for future studies on the use of predictive models in the development of an early warning system.