2024-06

Feature Selection Approach to Improve Malaria Prediction Model's Performance for High- and Low-Endemic Areas of Tanzania

Mariki, Martina

Springer Link

Provided with love from The Nelson Mandela African Institution of Science and Technology

Feature Selection Approach to Improve Malaria Prediction Model's Performance for High- and Low-Endemic Areas of Tanzania

Martina Mariki, Neema Mduma, Elizabeth Mkoba

To download complete text, click that link.

Abstract

Malaria remains a significant cause of death, especially in sub-Saharan Africa, with about 228 million malaria cases worldwide. Parasitological tests, like microscopic and rapid diagnostic tests (RDT), are the recommended and standard tools for diagnosing malaria. However, clinical diagnosis is advised in areas where parasitological tests for malaria are not readily available. This method is the least expensive and most widely practiced. A clinical diagnosis called presumptive treatment is based on the patient's signs and symptoms and physical findings at the examination. A malaria diagnosis dataset was extracted from patients' files from four (4) identified health facilities in Kilimanjaro and Morogoro. These regions were selected to represent the country's high- (Morogoro) and low-endemic areas (Kilimanjaro). The dataset contained 2556 instances and 36 variables. The random forest classifier, a tree-based, was used to select the most important features for malaria prediction since this classifier was selected for feature selection because it was robust and had high performance. Regional-based features were obtained to facilitate accurate prediction. The feature ranking indicated that fever is universally the most noteworthy feature for predicting malaria, followed by general body malaise, vomiting, and headache. However, these features are ranked differently across the regional datasets. Subsequently, six predictive models, using important features selected by the feature selection method, were used to evaluate the performance of the features. The identified features comply with the malaria diagnosis and treatment guidelines WHO and Tanzania Mainland provided. The compliance is observed to produce a prediction model that will fit in the current healthcare provision system.