

Finding the Best Feature Selection Method for Dengue Diagnosis Predictions

R Lathesparan^{1#}, RMKT Rathnayaka² and WU Wickramaarachchi¹

¹*Department of Computing, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*

²*Department of Physical Sciences and Technologies, Sabaragamuwa University of Sri Lanka Belihuloya, Sri Lanka*

#lathesparan@gmail.com

Dengue is a mosquito-borne viral disease that has dramatically increased around the world in recent years. The spread of Dengue depends on the tropics, rainfall, temperature, relative humidity and unplanned urbanization. Severe Dengue can lead to circulatory system failure, shock and even death. The development of an effective Dengue fever prediction model is therefore essential for better Dengue case management. Feature selection is the predominant phase in developing the Dengue diagnosis prediction model. It is required to identify the most crucial attributes, as not all attributes have notable effects on the results. Therefore, this study focuses on the feature selection methods such as Principal Component Analysis (PCA) and Wrapper feature selections method with Naïve Bayes, K-Nearest Neighbor (KNN), and J48 algorithms. Simple Artificial Neural Networks (ANN) were developed to validate the performance based on the accuracy of each feature selection method, since it can work well with the partial dataset. Myalgia and Retro-Ocular Pain are the most expressive features chosen by all wrapper feature selection methods. In addition, with PCA, the initial 22-dimensional system was reduced to an 8-dimensional system with a cumulative variance of 59%. ANN with PCA resulted in the higher accuracy of 72.47% and ANN with Wrapper feature selection (KNN) showed the lowest accuracy of 54.47%. In conclusion, PCA is identified as the best feature selection method for the given dataset in this study based on the accuracy of ANN. In future, multiple Dengue diagnosis prediction models can be developed with higher accuracy and efficiency using the most vital attributes.

Keywords: *feature selection, artificial neural networks, dengue diagnosis prediction models*