
Predicting the Risk of Transmitting Dengue Based on Individual Lifestyle Data Using Predictive Data Mining

Final Year Project Thesis By

K.K.I.S de Alwis – W1473544

Supervisor: Mr. Nishan de Silva

Date: 2nd May 2018

Department: Computer Science

University of Westminster

Key Words: Data Mining, Neural Network, Dengue

This report is submitted in partial fulfillment of the requirements for the BEng. (Hons) in Software Engineering Degree at the University of Westminster.

© The copyright for this project and all its associated products resides with Informatics Institute of Technology.

Abstract

The purpose of this research is to develop an early warning system to predict the possibility of infecting dengue fever to an individual. To achieve this purpose, a literature survey performed by studying various researches and similar work related to disease prediction. As the solution to dengue risk prediction problem, a disease prediction model that is capable to find a relationship between human lifestyle factors and human immunization to dengue is determined to build. Several human lifestyles are identified as input parameters to the risk prediction model. The main aim of this research is to develop a disease prediction model that is capable to make accurate predictions about the probability of infecting dengue to an individual by analyzing the lifestyles of that individual.

Dengue, which is a global issue, is a mosquito-borne viral disease of the tropics and sub tropics. Dengue cause sudden fever, severe pain in the joints and mortality. Dengue virus consists of four distinct stereotypes. Bite of an infected mosquito transmit the disease to humans. Since there is no specific antiviral treatment for Dengue, prevention and control programs take important place. One of the main purposes of Dengue risk prediction system proposed by this project is to integrate it to current ongoing Dengue Prevention programs.

The predictive data mining techniques are selected to develop the proposed dengue risk prediction model. The data set for this research is collected from dengue patients & non dengue patients around Sri Lanka. Decision trees, random forests, neural networks and logistic regression are selected to apply to the data set. This project focus on developing ensemble model by stacking individual models. This disease prediction model will analyze the factors that will influence the human lifestyle that are collected through various researches and identify which factors that make an individual more susceptible to Dengue virus.