Abstract

Due to big data progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Moreover, different regions exhibit unique appearances of certain regional diseases, which may result in weakening the prediction of disease outbreaks. In the proposed system, it provides machine learning algorithms for effective prediction of various heart disease occurrences in disease-frequent societies. It experiments the altered estimate models over real-life hospital data collected. The project will implement 4 linear models and one deep learning model: Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbors and Multilayer Perceptron Neural network to investigate their performance on diabetes and heart disease datasets obtained from the UCI data repository. In addition to the comparison of the algorithms, each algorithm has been integrated into a prediction engine and exposed over an API. The project will also include a web platform to facilitate collaboration among researchers and developers.