Literature Survey on Early Stage Heart Disease Prediction using Machine Learning Technique

Abstract

Machine Learning is used across many spheres around the world. Machine Learning can play an essential role in predicting presence/absence of locomotor disorders, Heart diseases and more. In this modern era people are very busy and working hard in order to satisfying their materialistic needs and not able to spend time for themselves which leads to physical stress and mental disorder. Thus, heart disease is very common nowadays. Particularly in urban areas because of excess mental stress. As a result, Heart disease has become one of the most important factors for death of men and women. It has emerged as the top killer that has affected both urban and rural population. In the medical field, predicting the heart disease has become a very complicated and challenging task. It requires patient previous health records and in some cases they even need Genetic information as well. So, in this contemporary life style there is an urgent need of a system which will predict accurately the possibility of getting heart disease. Predicting a Heart Disease in early stage will save many people's Life. The main objective of this paper is to design a robust system which works efficiently and will able to predict the possibility of heart failure accurately. Machine learning (ML) has been showing an effective assistance in making decisions and predictions from the large quantity of data produced by the healthcare industries and hospitals. The prediction model is proposed with combinations of different features and several classification techniques. We will be predicting potential Heart Diseases in people using Machine Learning algorithms. The algorithms included are K Nearest Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier and Random Forest Classifier. We will analyse prediction systems for Heart disease using a greater number of input attributes. The system uses medical terms such as Sex, Age, blood pressure, cholesterol attributes to predict the likelihood of patient getting a Heart disease.

Literature Survey

[1] A Comparison Based Study of Supervised Machine Learning Algorithms for Prediction of Heart Disease.

Published in 1st International Conference on Computational Intelligence and Sustainable Engineering Solution (CISES-2022).

Authors: Deepak Kumar Chohan, Dinesh C Dobhal

Algorithm Used: This paper has analysed prediction systems for Heart disease using Supervised Machine learning Algorithms. In this work, they used the Logistic Regression, Decision Tree, SVM, Naive Bayes, Random Forest, and KNN algorithms to predict heart disease, with the purpose of determining which method is the most reliable.

Advantage: Logistic Regression gives the measure of how importance of a predictor both in positive or negative direction. We got 78.53% accurate results with Logistic Regression. Naive Bayes is easy and fast to predict class of test data set. It also performs well in multi class prediction. We got 80.0% accurate with Naive Bayes. K-Nearest Neighbors(KNN) is the most simple algorithm to implement with just one parameter. KNN gives 67.80% accuracy. Support Vector Machine works really well with a clear

margin of separation. We got 80.48% accurate with Support Vector Machine (SVM). Random Forest is robust to outliers. It lowers risk of overfitting. We got 87.80% accurate with Random Forest Algorithm. Decision Tree is easy to understand and interpret, perfect for visual representation. It gives better accuracy than other classification algorithms. We got 98.53% accurate with Decision Tree Algorithm.

Drawback of Algorithm: Logistic Regression Does not support non-linear relationship between the predictor and the outcome. KNN is a distance based-approach hence the model can be badly affected by outliers, in other words, it's prone to overfitting. In Naive Bayes, there is the assumption of independence in predictors. In real life, it is almost impossible that we get predictors which are completely independent. SVM doesn't perform well when we have large data set because the training becomes time-consuming. Random forests are found to be biased while dealing with categorical variables and also it has slow training. Decision tree is very sensitive. Small change in the data can affect prediction greatly.

[2] A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms.

Published in 2021 Asian Conference on Innovation in Technology (ASIANCON) Pune, India. Aug 28-29, 2021.

Authors: Akanksha Kumari, Ashok Kumar Mehta

In this paper, authors have tried to predict heart disease using seven machine learning algorithms and attempted to improve the accuracy of weak performing algorithms using ensemble methods like AdaBoost and voting ensemble method. The authors have evaluated and compared seven types of machine learning algorithms in this paper, these are LR-Logistic Regression, LDA-Linear Discriminate Analysis, KNN-K Neighbors Classifier, CART-Decision Tree Classifier, GNB-Gaussian Naïve Bayes, SVM-Support Vector Machine and RF-Random Forest classifier. The performance of Linear Discriminate Analysis is good among other algorithms, accuracy is somehow coming 80% which is less if compared with Logistic Regression. LDA is simple, fast and portable algorithm. Only disadvantage of using LDA is that it requires normal distribution assumption on features/predictors.