

Abstract

Numerous illnesses that affect the heart are categorized as heart diseases, commonly referred to as cardio vascular diseases. These can include heart rhythm issues, blood vessel issues, etc. related to congenital cardiac abnormalities. In recent years, it has been the main cause of death worldwide. Thus, it can save many people's arteries if it can detect this dangerous issue early. Accurate and reliable methods must be found quickly in order to automate the process of early disease diagnosis and achieve effective disease management. In addition, it has been concentrated on an effective strategy to raise the effectiveness. In the previously existing systems, the heart disease dataset is applied to K-Nearest Neighbors (KNN), Naive Bayes and CNN for the prediction. The proposed system uses the info-gain feature selection technique to eliminate features that are unneeded and the features to be used are extracted. One of the supervised machine learning algorithms called random forest and XG Boost is used to forecast the presence of cardiac disease in patient bodies. The optimization technique called hyper parameter tuning is to be used to tune some hyper parameters in the random forest algorithm this serves as a base algorithm for XG Boost algorithm to eliminate overfitting. The UCI heart disease dataset obtained from kaggle repository is used. To calculate the accuracy, precision, recall and F1 score methods are to be used. The optimized algorithm is then used to train and test the data thus predicts the presence of heart disease in people and the increased accuracy is obtained.