Abstract

Arrhythmia data is obtained from UCI Machine Learning Repository and analyzed to distinguish between the presence and absence of arrhythmia in the patients based on the attributes specific to patients. A model ensembling approach is used to combine predictions from tuned Stochastic Gradient Boosting Model, tuned Neural Networks and tuned Support Vector Machines to achieve an Area under ROC Curve equal to 0.7793, improving upon the individual AUROC achieved by Stochastic Gradient Boosting Model (0.7658), Support Vector Machines (0.7132) and Neural Networks (0.7177).

Aim

The goal of this project is to distinguish between the presence and absence of cardiac arrhythmia in the subjects tested upon, using the data collected from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Arrhythmia)

Introduction

This database contains 279 attributes, 206 of which are linear valued and the rest are nominal. The aim is to distinguish between the presence and absence of cardiac arrhythmia in the patients. The names and id numbers of the patients were removed from the database.