ECE657 Research Paper Proposal-Group 23

Shuyue Wang 20771353
Su Sun 20862738
Yue Teng 20774328

University of Waterloo
Electrical & Computer Engineering
Astronomy & Physics
s754wang@uwaterloo.ca
s266sun@uwaterloo.ca
y9teng@uwaterloo.ca

1 Problem Statement

1.1 Brief description

Heart disease is really common nowadays. Accurate diagnosis plays an significant role in better medical treatment. As the rapid development of machine learning approaches, better tools for detection and classification of heart disease appeared. In this project, several machine learning approaches mentioned below will be investigated. These approaches will be manipulated to innovatively design a system and the performance will be evaluated by the comparison with the baseline methods. The data set of relevant patients are retrieved from the following link, which provides plenty of information formatted.

1.2 Dataset

Heart Disease Data Set*(link)*: This dataset from the UCI Machine Learning Repository contains 4 databases related to heart disease diagnosis with 920 samples in total, Cleveland, Hungarian, Switzerland, and Long Beach VA, respectively. Each of the databases contains 76 numeric-valued features, but 14 among these are the most commonly used. Each of the samples represents a patient who may have heart disease (with "goal" field 1,2,3,4 values) or not (value 0). Several approaches will be conducted with this dataset to form an ensemble of several models and we will compare its accuracy with each individual model.

2 Proposed Approach

The traditional machine learning classification algorithms, such as logistic regression, neural networks, K-nearest neighbour, SVM and decision tree, are going to be implemented in this certain data set, and their performances will be compared for this specific application. In addition, the random forest using random search algorithm as hyperparameter tuning method is going to be applied for

performance comparison. Nevertheless, this paper will design and develop an ensemble of various classifiers which aims at increasing the accuracy of heart disease classification. Each single classifier is regarded as the baseline and the performance evaluation metrics will contain confusion matrix, precision, recall and F1 score.