ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

A NOVEL FRAMEWORK FOR CLASSIFICATION OF CARDIOVASCULAR DISEASE USING LOCAL LINEAR EMBEDDING AS FEATURE SELECTION METHOD

Nabila Kausar[1], Dr. Hamid Ghous[2]
Institute of Southern Punjab (ISP), Multan, Pakistan
[1]milkywaymn555@gmail.com
[2]Hamidghous@isp.edu.pk

ABSTRACT— A huge amount of biological data has been collected which is related to cardiovascular disease in healthcare industry and it's increasing day by day. This enormous amount of data is in irregular form. So it is difficult to extract useful information in limited time and within affordable cost range. That's why we need some dimensionality reduction feature methods to process this data with combination of data mining techniques to extract useful information from this huge amount of data. Data mining techniques such as Decision tree, Naïve Bayes, Neural network, K-Nearest neighbor and Random Forest has been previously used by many researchers for classification of cardiovascular patients. In this work, LLE used as feature selection method on two datasets; Cleveland and Statlog before applying classification methods, Decision Tree (DTree), Random Forest (RF), Support vector Machine (SVM) and Neural Network (NN) to predict the cardiovascular disease in patients. The results show that Random Forest is the best classifier for the prediction of cardiovascular disease with highest prediction accuracy for both Cleveland and Statlog heart disease datasets. It shows AUC-ROC as 1 with 80% training datasets. This framework will also be time effective and cost effective.

Keywords— Cardiovascular disease, data mining techniques, dimensionality feature reduction methods, local linear embedding (LLE).

I. INTRODUCTION

Cardiovascular Disease: Cardiovascular Disease (CVD) is one of the fatal diseases. Cardiovascular disease contains many types of heart diseases. Cardiovascular disease caused a large number of people die every year. According to World Health Organization (WHO), twelve million people died by cardiovascular disease every year in the whole world. In almost every 34 seconds one person died due to heart problem throughout the world, Kumar(2018).

Cardio refers to as heart, so cardiovascular disease refers to various types of heart problem in which coronary artery disease (CAD), angina pectoris, arrhythmias, myocarditis, cardiomyopathy, congential heart disease and congestive heart failure are included.

- Coronary Artery Disease: It is the disease of coronary artery which supplies blood and oxygen toward the whole body. It occurs when a plaque or clot deposits in artery and causes the insufficient supply of blood and oxygen.
- Angina pectoris: It is chest pain also called angina. It is the condition of insufficient supply of blood to the heart. It is the warning of heart attack.
- Arrhythmias: Arrhythmias is disorder of heart movement. Heartbeat can be fast, slow and irregular in this condition.
- Myocarditis: It is uncommon disease of heart. It is the inflammation of heart muscle. It is caused by viral, fungal and bacterial infection.
- Cardiomyopathy: It is the weakness of heart muscle or change of heart structure in pumping blood.
- Congential heart disease: It is the formation of abnormal heart due to defect in structure of heart or its functioning. It is the disease of heart in which new born baby are suffered.
- Congestive heart failure: It is the disease in which heart failed to supply the sufficient blood in the whole body. It is also known as heart failure, (Sudhakar & Manimekalai, 2014; Karthiga et al, 2017).

❖ **Factors Causing Cardiovascular Disease**

There are many factors which are causing cardiovascular disease. Some factors are controllable factors which can be

ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

easily controlled by humans and some are uncontrollable which cannot be controlled by humans.

a) Controllable factors

The factors causing heart disease which can be easily controlled by humans are called controllable factors. These factors are as follows

- **Hypertension:**

Hypertension is controllable factor caused cardiovascular disease in humans. It increases the blood pressure that can damage the muscles of heart.

- **Smoking:**

Smoke addicts have more chance of heart attack because tobacco has such dangerous chemicals which cause to damage artery walls. It is also controllable factor causing cardiovascular disease.

- **Cholesterol:**

Cholesterol closes and narrows the arteries of heart. It deposits in arteries in form of plaques. Plaques may block and narrow the arteries which cause heart failure.

- **Blood Sugar:**

High blood sugar is also controllable factor that can cause injury in blood arteries. It also promotes blood clots in arteries that can cause high blood pressure and high cholesterol. It can be major factor of causing cardiovascular disease.

- **Obesity:**

Obesity is also risk factor that causing cardiovascular disease. When fat is growing fast, it increased the obesity which cause of high cholesterol in man's body. So it is the cause of heart attack in fatty person more than smart persons.

b) Uncontrollable factors

The factors that are fixed and cannot be controlled by humans are called uncontrollable factors which cause cardiovascular diseases in humans. These factors are as follows

- **Age:**

Age is uncontrollable risk factor of heart disease. Heart diseases mostly affect the persons above age 40 years and mostly die the people having age above 65 years.

- **Sex:**

Heart diseases mostly attack the men rather than women. The men's death ratio is high than women due to cardiovascular disease. The sex is uncontrollable risk factor of CVD, Devi et al (2016).

This study handles the both types of factors.

❖ **Dimensionality feature reduction method**

The methods that are used to reduce the dimensions and features of a dataset in data mining are termed as dimensionality feature reduction methods. These methods are very important and play a vital role in data mining to tackle the huge amount of data. In this study, local linear embedding is used as dimensionality feature reduction method to reduce the features.

- **Local Linear Embedding (LLE):**

LLE is a local linear embedding technique for dimensionality reduction. It makes graph representation of the data points. It preserves only local properties of the data that makes LLE less sensitive to short circuiting. Successful embedding of non convex manifolds is allowed by preservation of local properties. It describes the local properties of data point x_i as a linear combination W_i of its k-nearest neighbors x_{ij}. LLE can be written in mathematical form as;

$$\emptyset(Y) = \sum_i (y_i - \sum_{j=1}^{k} w_{ij}\, y_{ij})^2$$, Maaten (2007).

After reducing the attributes of a dataset, this study used some data mining classification techniques to classify the data into two categories as; healthy or unhealthy.

❖ **Classification Methods**

Classification methods are data mining techniques which used for the classification of data. They are used to categorize the data into healthy and unhealthy classes in healthcare sector. This project implemented four classification methods on heart disease datasets to classify the data as:

i. Support Vector Machine (SVM):

ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

Vapnik and Cortes proposed SVM which have been applied for gender classification by many researchers. This is a linear classifier which can identify two classes. There is a line which is used to separate the dataset is called separating hyperplane (Patil et al, 2019; Kanikar & Shah, 2016). The advantage of SVM is that it is easy to interpret results and computationally inexpensive. But on the other hand disadvantage is that it can handles only binary classification, Kanikar & Shah (2016). Mathematically, SVM can be written as

If $Y_i = +1$; $wx_i + b \geq 1$ → (1)

If $Y_i = -1$; $wx_i + b \leq 1$ → (2)

For all i; $Y_i (w_i + b) \geq 1$, Ayon et al (2020).

ii. Decision Tree (DTree):

Decision tree is tree like structure having roots and leaves. Root node represents the top node, internal node represents the testing of attribute, branch node represents the output of testing and leaf node represents the class (Subhadra & Vikas, 2019; Shylaja & Muralidharan, 2019). Both numerical and categorical data can be handled by decision trees. In this algorithm, first information gain of the attributes should be found. By using the below equation (1), information gain of the attributes can be identified.

$$E(S) = -P(P)\log_2 P(P) - P(N)\log_2 P(N) \quad (1)$$

David & Beley (2018).

iii. Random Forest (RF):

The group of different classifiers of decision trees is called random forest. In this technique different decision trees with randomly selected attributes are combined, so it is called random forest. These classifiers used ensemble algorithms, Subhadra & Vikas (2019). Random forest accomplished in different steps as:

a) First it selects a bootstrap sample from training set
b) Then it generates an unpruned tree on this bootstrap sample
c) Then it selects each interior node randomly and also determines the best split
d) There is no need to pruning if each tree is completely developed

Using majority voting method from all trees, final output is taken, Ayon et al (2020).

iv. Neural Network (NN):

Neural network consists of many neurons which are interconnected with each other. It consists of three layers, input layer, hidden layer and output layer. Neural network always works like a brain and consists of neurons, so it is called neural network (Subhadra & Vikas, 2019; Shylaja & Muralidharan, 2019). Neural network works in different steps as follows;

a) First processed data given to input layer
b) Then this data is transferred to hidden layer with some random weight value
c) From hidden layer, this data transferred to output layer
d) Each value at output layer compare with original information
e) If information is incorrect then modify the weight value and again process the information
f) If information is correct then no alternation is done in weight value

Finally output layer gives the output after processing information, Maji & Arora (2019).

1.1 Problem Statement: The data about heart patients is increasing continuously and factors causing heart disease also rising simultaneously. So to predict the heart disease is very difficult task and also time consuming and very costly. That's why; this study introduced a framework that consist of dimensionality feature reduction method, to reduce the factors of heart disease data and then used the classification methods to predict the heart disease. In this way, this study will reduced the time and cost as well as improve the classification methods.

1.2 Identify Research Questions: There are some questions related to research,

- Is this possible to help biologists in early diagnose of cardiovascular disease using data mining?

ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

- Can we use linear or non linear dimensionality reduction techniques as feature reduction methods for cardiovascular disease data?

- Is this model can be time effective and cost effective also?

1.3 Significance/Objective/Scope: Cardiovascular Disease (CVD) is one of the fatal diseases. According to World Health Organization (WHO) twelve million people die every year by CVD. In almost every 34 seconds one person die due to heart problem throughout the world, Kumar(2018). In this project, we want to predict cardiovascular disease in patients using biological data. We are using two authentic CVD datasets from UCI machine learning repository which already used in (El-Bialy et al, 2015; Ayon et al, 2020). This framework will help biologist in early diagnosis/prognosis of cardiovascular diseases.

All the above discussed methods are widely used in classification and prediction of cardiovascular disease. In next section, we reviewed the previous work using data mining techniques on heart disease datasets.

II. LITERATURE REVIEW

Many researchers in past have worked on early diagnosis of cardiovascular disease. As cardiovascular disease is one of the fatal diseases and very vastly cause death among people. There are many factors which cause heart disease. So data about heart disease is increased day by day. That's why many researchers have been trying to handle this enormous data. For this purpose, different studies have been carried out by using data mining techniques like Naïve Bayes, Neural Network, Decision Tree, Random Forest, Support Vector Machine and different hybrid models on different heart disease datasets, their results shows the performance of these models for the prediction of cardiovascular disease.

Verma et al. (2016) used multinomial logistic regression (MLR), multilayer perceptron (MLP), fuzzy unordered rule induction algorithm (FURIA) and C4.5 on clinical data of 26 features and 335 instances to predict accuracy and incorrectly classified instances in prediction of coronary artery disease. They found that MLR has highest prediction accuracy which is 83.5%. After that they proposed hybrid method with correlation based feature subset selection (CFS) with particle swam optimization (PSO) search method to reduce the features. After applying CFS and PSO five features are selected. In this way accuracy of MLR is increased 0.67%. After feature selection, K mean clustering is applied so accuracy of MLR is increased to 88.4%. They also applied this proposed method on Cleveland data set with 14 features and 303 instances. After applying hybrid model, features are reduced to seven and accuracy is increased to 92.8% [12].

Verma & Srivastava (2016) collected the dataset from UCI machine repository contributed by Cleveland. They used 70% data for training the model and 30% data for testing the model. They used artificial neural network (ANN) based model to predict the coronary artery disease. They used Probabilistic Neural Network (PNN), alternating decision tree (ADTree) and RBF network to predict the CAD with more accuracy. They evaluated the performance of diagnostic model by measuring the difference between actual values and predicting values. They found that the prediction accuracy of PNN is higher than ADTree and RBFN which is 96% and misclassification rate is 4%. They also compared their model with other researchers work and found PNN has highest accuracy [13].

Kumar et al, (2018) used four classification algorithms, Naïve Bayes, Multilayer Perceptron, Random Forest and Decision Table to classify a patient. Patient is tested positive or negative for heart diseases based on some measurements included into the dataset. They also compare these four algorithms and found that Naïve Bayes has better accuracy for classification of heart disease. They found the maximum accuracy of Naïve Bayes is 87.20% and minimum accuracy of Random Forest is 83.72% using confusion matrix [14].

Maji & Arora (2019) used two techniques C4.5 and ANN for prediction of heart disease and develop a hybrid DT

ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

by combining ANN with C4.5. Hybrid DT implemented on same dataset and found that its accuracy is 78% which is better than other two techniques. They also found the sensitivity and specificity of C4.5, ANN and hybrid DT [15].

Sharma & Parmar (2020) deployed a model to improve the prediction accuracy of heart disease using Cleveland heart disease dataset. They proposed an optimized Deep Neural Network (DNN) model using Talos. They also applied some other classification models such as Logistic Regression (LR), KNN, SVM NB, RF and then applied proposed model Hyper-parameter Optimization using Talos . The results showed that hyper parameter optimization using Talos performed better with accuracy of 90.78% [16].

Ayon et al (2020) implemented the seven classification techniques Logistic regression (LR), Support vector machine (SVM), Deep neural network (DNN), Decision tree (DT), Naïve bayes (NB), Random forest (RF) and K-nearest neighbor (K-NN) on two heart disease datasets Statlog and Cleveland. These datasets were collected from UCI machine learning repository to early diagnose the Coronary heart disease. They implemented these techniques using Python 3.0. They calculated accuracy, sensitivity, specificity, precision, NPV, F1 score and MCC by using five-fold and ten-fold cross validation. They found that DNN shows better accuracy 98.15% with Statlog dataset and SVM shows better accuracy 97.36% with Cleveland dataset using five-fold cross validation [11].

Ricciardi et al. (2020) proposed a method to classify the patients into two groups, healthy and unhealthy. They used clinical dataset collected from the Department of Advanced Biomedical Sciences, University Hospital Federico II of Naples, Italy. They first applied Linear Discriminant Analysis (LDA) on dataset and classify patients into two groups, healthy and unhealthy. They calculated accuracy 84.5%, precision 94.2%, sensitivity 62.8% and specificity 97.7%. After that applied LDA with combination of PCA (Principal Component Analysis), extracted 22 features. Finally, they calculated accuracy 86.0%, precision 96.2%, sensitivity 65.4% and specificity 98.4%. All these experiments were performed using Knime and R programming languages [17].

Joloudari et al. (2020) proposed a method for the improvement of accuracy of coronary heart disease diagnosis. They implemented 10-fold cross validation on Z-Alizadeh Sani dataset to classify into ten subsets as 90% training dataset and 10% testing dataset. Then they implemented the Decision trees of Chi_squared automatic interaction detection (CHAID), Decision trees of C5.0, support vector machine (SVM) and random trees (RTs) classification methods on dataset. They found that RT has accuracy 91.47%, SVM 69.77%, CHAID 80.62% and C5.0 has 82.17%. They showed that RT is the best classifier than other classifiers [18].

Kolukisa et al. (2019) used two datasets; Cleveland and Alizadeh Sani heart disease datasets. They proposed an adaptive ensemble machine learning algorithm. They implemented K-nearest neighbor (KNN), logistic regression (LR), linear discriminant analysis (LDA), naïve bayes (NB), support vector machine (SVM) and ensemble method on two datasets. They found that ensemble method shows highest accuracy of 83.43% on Cleveland dataset and 88.38 % on Alizadeh Sani dataset [19].

Bhaskura & Devi (2019) proposed a new automated system for accurate diagnose of heart disease, named as Hybrid Differential Evaluation based Fuzzy Neural Network (HDEFNN). The Cleveland heart disease dataset collected from UCI machine learning repository. First they normalized the dataset then applied the proposed method algorithms on normalized dataset. The simulation results performed in Matlab K-fold cross-validation. After applying HDEFNN, their results compared with J48, NB and RF with reference to accuracy. They found that HDEFNN shows better accuracy than other algorithms [20].

Shylaja & Muralidharan (2019) developed a hybrid classifier by hybridizing of Support Vector Machine (SVM) and Artificial Neural Network (ANN) classifiers for the prediction of heart disease. The Cleveland heart disease dataset was collected from UCI Repository and implemented the data mining techniques on this dataset as ANN, SVM, RIPPER, Decision Support, NB and hybrid SVM-ANN

ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

classifier. They performed all these experiments using MATLAB and calculated their accuracy, sensitivity and specificity. They showed that hybrid SVM-ANN is the best classifier with accuracy 88.54%, sensitivity 91.47% and specificity 82.11% [9].

Latha & Jeeva (2019) proposed an ensemble technique for increasing the accuracy of some weak classifiers for the prediction of heart disease. They collected the Cleveland heart disease dataset from UCI machine learning Repository. They were performed some classification methods like Bayes Net, Naïve Bayes, Random Forest, C4.5, Multilayer Perceptron and PART on this dataset using WEKA tool and calculated their accuracy. After that they were used some ensemble classifiers as bagging, boosting, stacking and majority voting with weak classifiers on this dataset and calculated their accuracies again. They were improved the accuracy with majority voting classifier. After that they were implemented majority voting classifier with feature selection and found that highest accuracy was obtained by using this ensemble classifier [21].

Jan et al; Pushkala et al; Patra & Khuntia (2019) applied machine learning algorithms on Cleveland dataset to predict the heart disease. They implemented SVM, NN and other classification techniques on same dataset but achieved different results. They calculated the accuracies of SVM as 0.90%, 61%, 63.15% and NB as 0.98%, 91%, 65.01% respectively [22][23][24].

Research Gap: This section describes the previous work of different researchers and also describes the novelty of this project that how different this model from previous studies.

Previous Work: This chapter contains the comparative analysis of the previous study. In previous researches, different feature selection methods are used on different heart disease datasets to select some distinct features to reduce the datasets. They also processed heart disease datasets by removing noise and unwanted values. After preprocessing the datasets, they applied different data mining techniques to predict the heart disease and also improve the accuracies of different prediction methods.

Novelty of this Project: In contrast to previous researches, this study used two heart disease datasets; Cleveland dataset and Statlog dataset. As we know that features of causing heart disease increasing from some decades. That's why; prediction of cardiovascular disease becomes so difficult task. This task is costly and also time consuming.

So in this study, we first reduce the features of these datasets by applying the local linear embedding (LLE). After the preprocessing of datasets by reducing features, applied the classification methods; decision tree (DTree), random forest (RF), support vector machine (SVM) and neural network (NN) to predict the cardiovascular disease. This model leads to predict the CVD in very short time and within effective cost range. This model also improves the performance of the prediction model by increasing the classification accuracy. So this model helps the biologist to use dimensionality feature reduction methods for largest datasets to early diagnose the cardiovascular disease.

III. METHODOLOGY

Flow of the Study: In this work, two heart disease datasets are used for the early diagnosis of cardiovascular disease. Then we applied dimensionality feature reduction methods to reduce the dimensions. Finally, we applied classification methods on preprocessed datasets to classify the CVD. Then this study evaluated these methods by AUC-ROC graphs.

Data Collection: In this work, two heart disease datasets are used for the early diagnosis of cardiovascular disease. Both datasets are taken from UCI machine learning repository contributed by Cleveland *[https://archive.ics.uci.edu/ml/datasets/Heart+Disease]* and Statlog dataset from the site https://github.com/renatopp/arff-datasets/blob/master/classification/heart.statlog.arff. These two datasets: Cleveland heart disease dataset and Statlog heart disease dataset are used by many researchers for the prognosis of heart disease instances. In 14 attributes, first two attributes 'age' and 'sex' are non-clinical features, one attribute is class and rest 11 attributes are clinical features.

ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

Statlog dataset consists of 14 attributes in which one attribute is target value and 13 are input values and 270 instances. In 14 attributes, first two attributes 'age' and 'sex' are non-clinical features, one attribute is class and rest 11 attributes are clinical features.

Data Preprocessing: In this study, datasets that are used for the classification of heart disease, first preprocessed by using Rattle package of R studio. Data preprocessing is very essential for the accurate classification. This study preprocessed the data by reducing dimensions by applying dimensionality feature reduction methods.

3.3 Model Diagram:

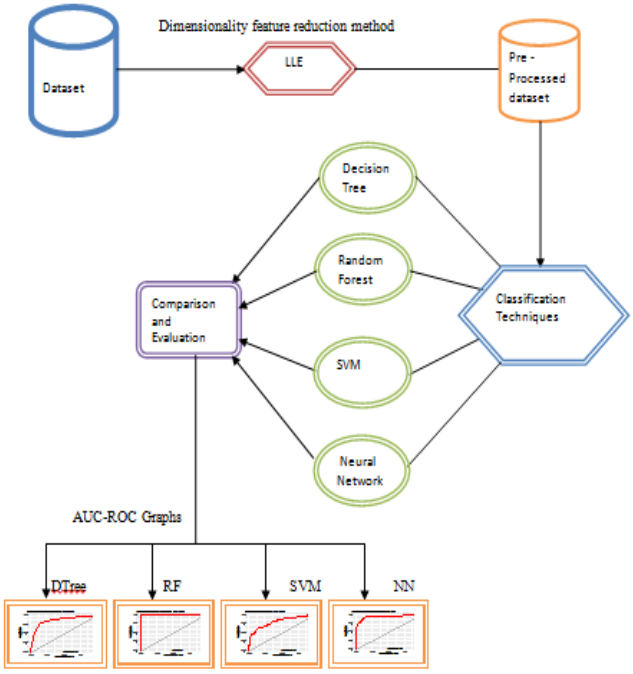


FIGURE 1. FRAMEWORK OF RESEARCH

IV. DATA ANALYSIS & RESULTS

Tool: We used R Studio for the simulation of datasets. R Studio is an integrated development environment for R. R is a programming language for statistical computing and visualization of graphs. We also used Rattle package in R. Rattle is graphical user interface for data miningExplain what software were used for interpretation and analysis of data, what analysis was carried out to explain descriptive statistics. What steps were used in inferential statistics including details regarding selection of appropriate statistical test and evaluation of their assumptions?

4.1 Simulation of Cleveland dataset without testing

First, loaded Cleveland heart disease dataset after processing it and then applied these four classification methods Decision tree (DTree), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) on Cleveland dataset without testing and training after implementing LLE. This study calculated Area Under the Reciever Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods. Random Forest in these classification methods shows highest AU-ROC which is 1, Neural Network shows 0.93, Support Vector Machine shows 0.82 and Decision tree shows 0.85 given in table 1.

TABLE 1. SIMULATION OF CLEVELAND DATASET WITHOUT TESTING

MODEL	ROC-ACCURACY	ERROR RATE %	TIME IN SECONDS
DTREE	**0.854**	**19.25**	**0.02**
RF	**1**	**0**	**0.38**
SVM	**0.824**	**26.1**	**0.09**
NN	**0.929**	**15.8**	**0.16**

4.2 Simulation of 20% testing Cleveland dataset

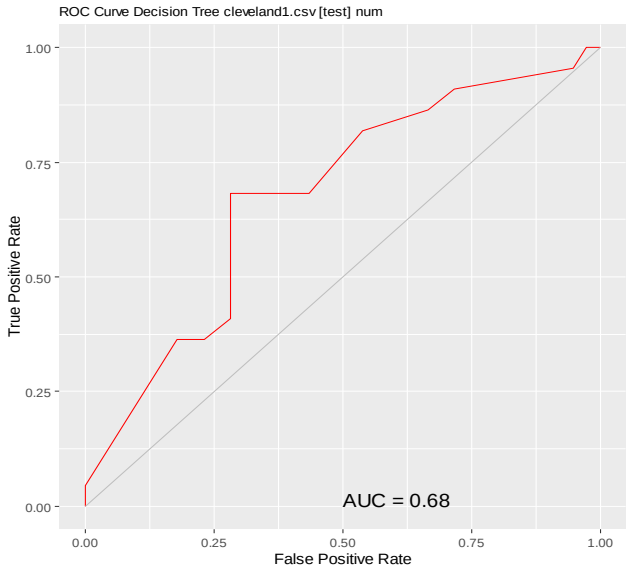
These four classification methods Decision tree (DTree), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) are also applied on 20% testing and 80% training preprocessed Cleveland dataset. Then calculated Area Under the Reciever Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods on the basis of testing dataset. Random Forest in these classification methods shows AU-ROC which is 0.77, Neural Network shows 0.64, Support Vector Machine shows 0.79 and Decision tree shows 0.68 given in table 2.

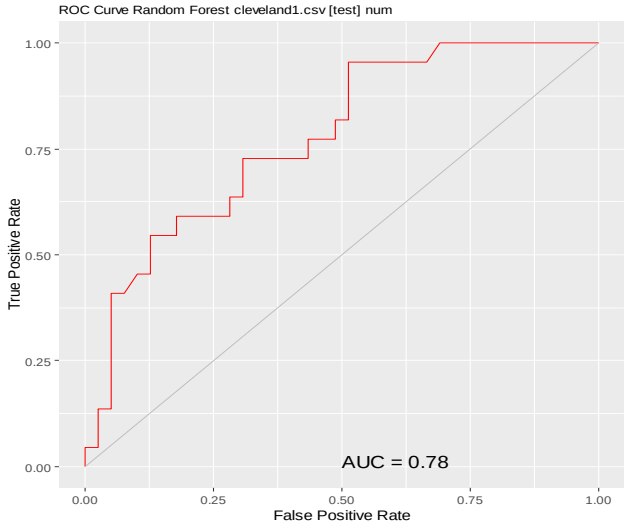
TABLE 4.2: SIMULATION OF CLEVELAND DATASET WITH 20% TESTING

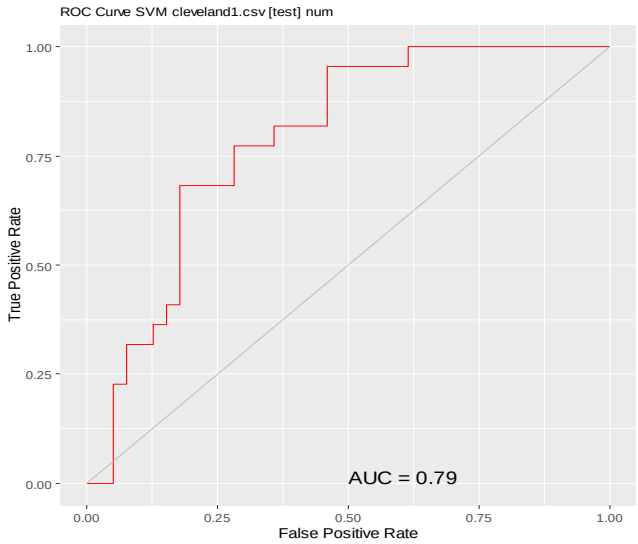
MODEL	ROC-ACCURACY	ERROR RATE%	TIME IN SECONDS
DTREE	0.676	30	0.02
RF	0.776	29.05	0.24
SVM	0.786	28.7	0.07
NN	0.644	39.4	0.09

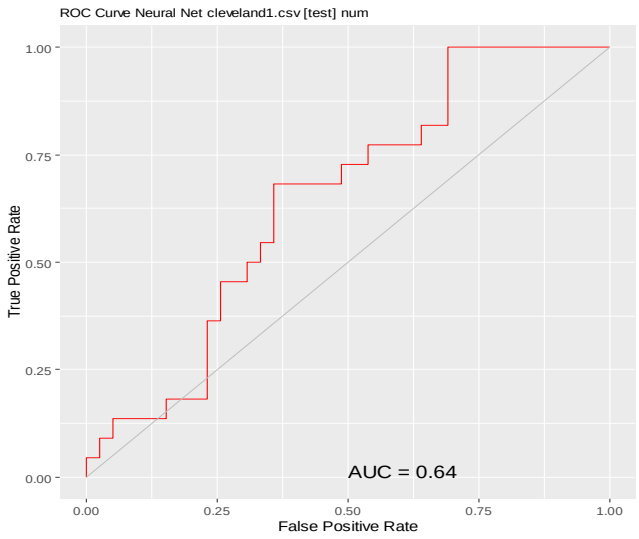
ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

ROC Curves









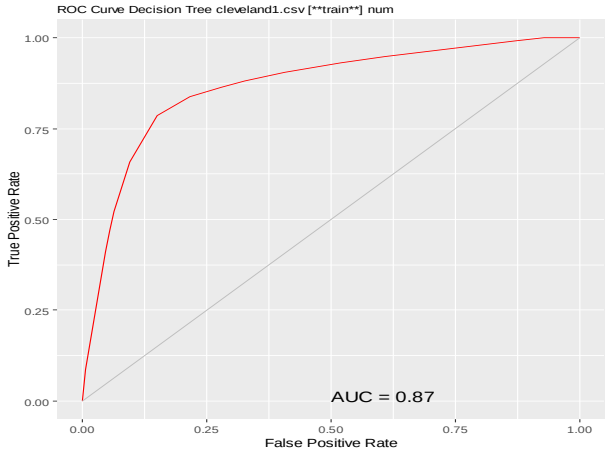
4.3 Simulation of 80% training Cleveland dataset

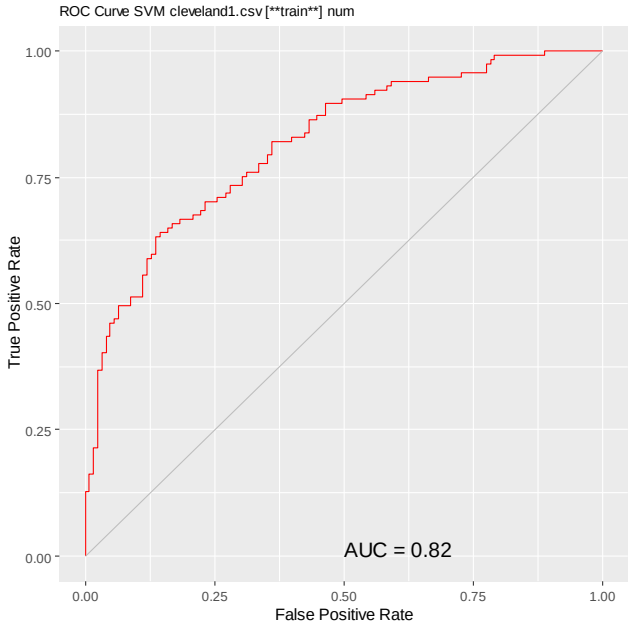
Then applied these four classification methods Decision tree (DTree), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) on Cleveland dataset with 20% testing and 80% training after implementing LLE. Then calculated Area Under the Reciever Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods on the basis of training. Random Forest in these classification methods shows highest AU-ROC which is 1, Neural Network shows 0.96, Support Vector Machine shows 0.82 and Decision tree shows 0.87 given in table 3.

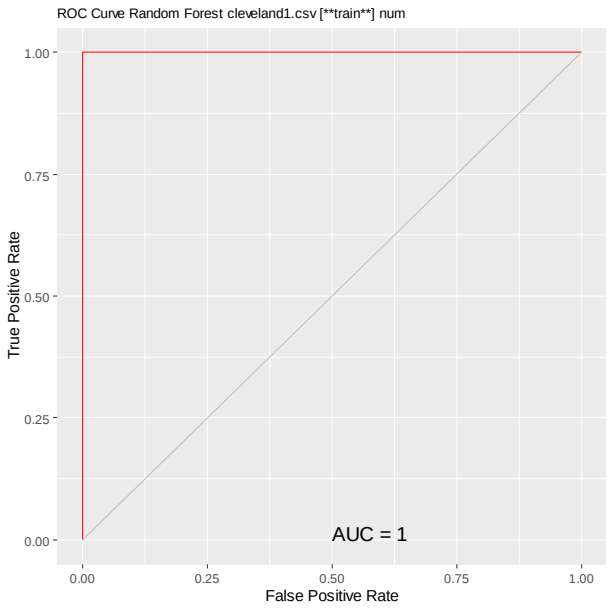
TABLE 4.3: SIMULATION OF CLEVELAND DATASET WITH 80% TRAINING

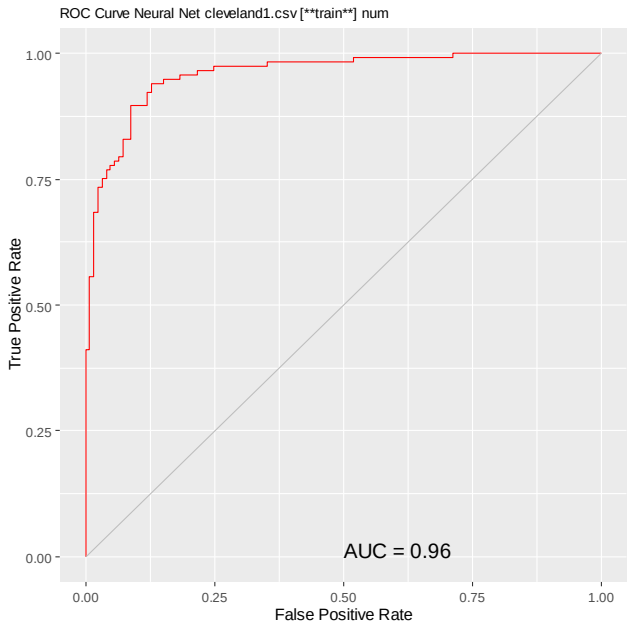
MODEL	ROC-ACCURACY	ERROR RATE %	TIME IN SECONDS
DTREE	0.866	18.3	0.02
RF	1	0	0.24
SVM	0.818	27	0.07
NN	0.958	9.55	0.09

ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150









In this research work, implement the dimensionality feature reduction methods, with the combination of classification methods to diagnose the cardiovascular disease. This study concludes that by reducing features, the performance of prediction model is improved. This project calculated the highest AUC-ROC of 1 with 0 error rate by using random forest (RF) on 80% training Cleveland dataset.

4.5 Simulation of Statlog dataset without testing:

Then loaded Statlog dataset and calculated Area Under the Receiver Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods of Statlog dataset. Random Forest in these classification methods shows highest AU-ROC which is 1, Neural Network shows 0.99, Support Vector Machine shows 0.85 and Decision tree shows 0.83 given in table 4.

Table 4.4: Simulation of Statlog dataset without testing

MODEL	ACCURACY %	ERROR RATE %	TIME IN SECONDS
DTREE	0.834	22.6	0.02
RF	1	0	0.31
SVM	0.853	21.4	0.08
NN	0.986	4.7	0.10

4.6 Simulation of 20% testing Statlog dataset:

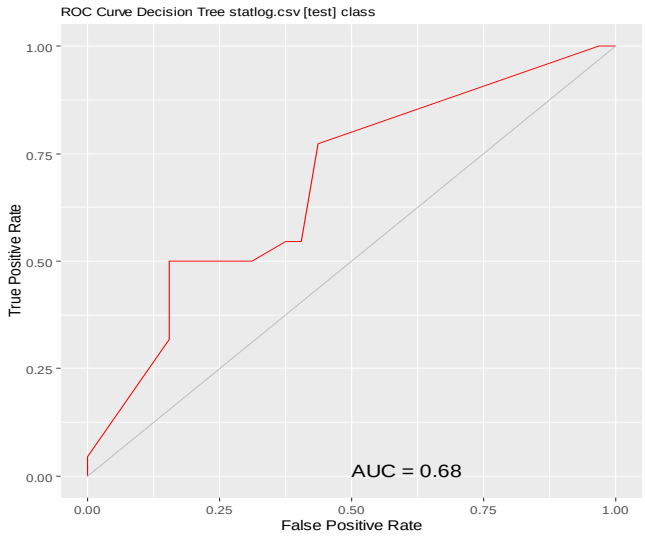
Then applied these four classification methods Decision tree (DTree), Random Forest (RF), Support Vector Machine

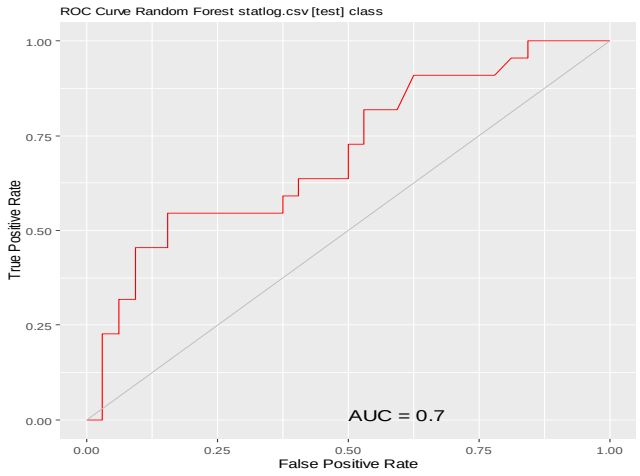
ISSN: 2708-7123 | Volume-01, Issue Number-03 | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

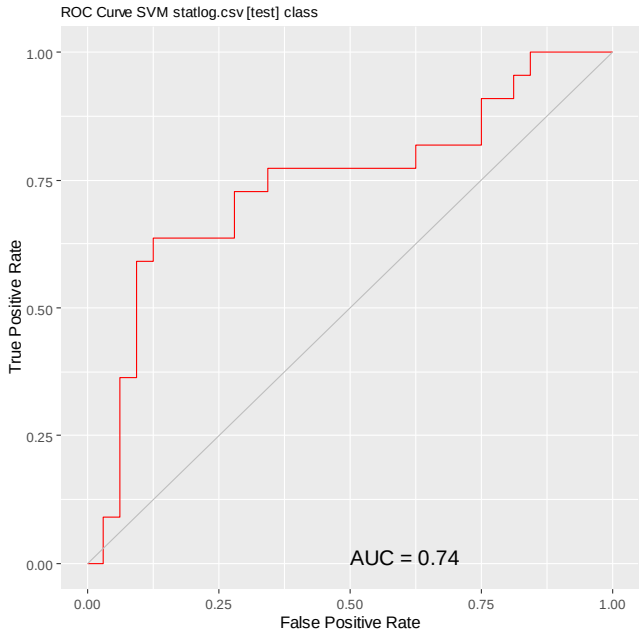
(SVM) and Neural Network (NN) on Statlog dataset with 20% testing and 80% training after implementing LLE. Then calculated Area Under the Reciever Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods on the basis of testing. Random Forest in these classification methods shows AU-ROC which is 0.70, Neural Network shows 0.63, Support Vector Machine shows 0.74 and Decision tree shows 0.68 shown in table 5.

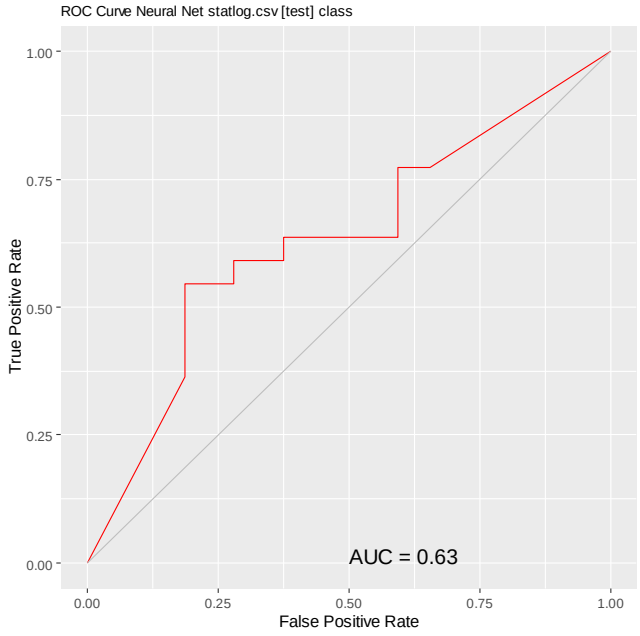
Table 4.5: Simulation of Statlog dataset with 20% testing

Model	Accuracy%	Error Rate %	Time in seconds
Dtree	**0.679**	**40.6**	**0.02**
RF	**0.701**	**32.15**	**0.32**
SVM	**0.740**	**27.6**	**0.09**
NN	**0.632**	**36.95**	**0.11**









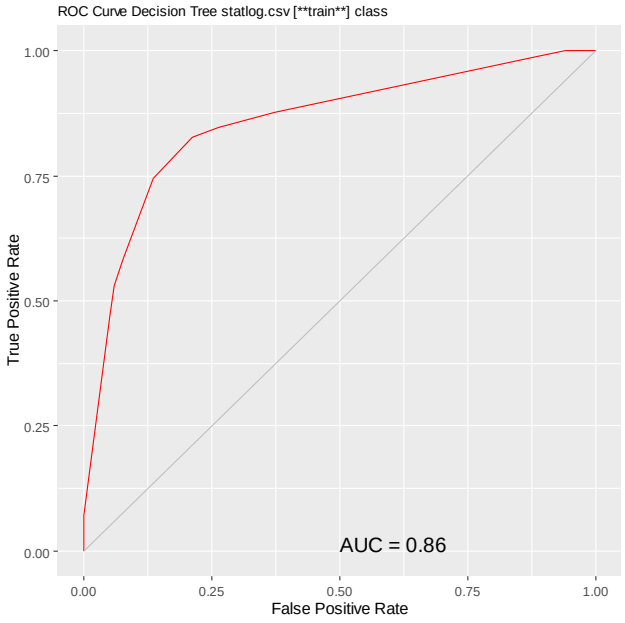
4.7 Simulation of 80% training Statlog dataset:

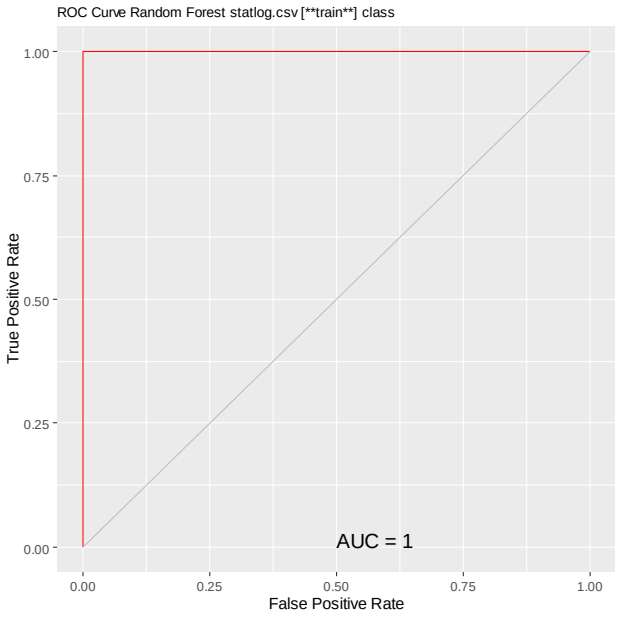
Then applied four classification methods Decision tree (DTree), Random Forest (RF), Support Vector Machine (SVM) and Neural Network (NN) on Statlog dataset with 20% testing and 80% training after implementing LLE. Then calculated Area Under the Reciever Operating Curve (AU-ROC), Error rate and time to built in seconds of these four classification methods on the basis of training. Random Forest in these classification methods shows highest AU-ROC which is 1, Neural Network shows 0.997, Support Vector Machine shows 0.85 and Decision tree shows 0.86 given in table 6.

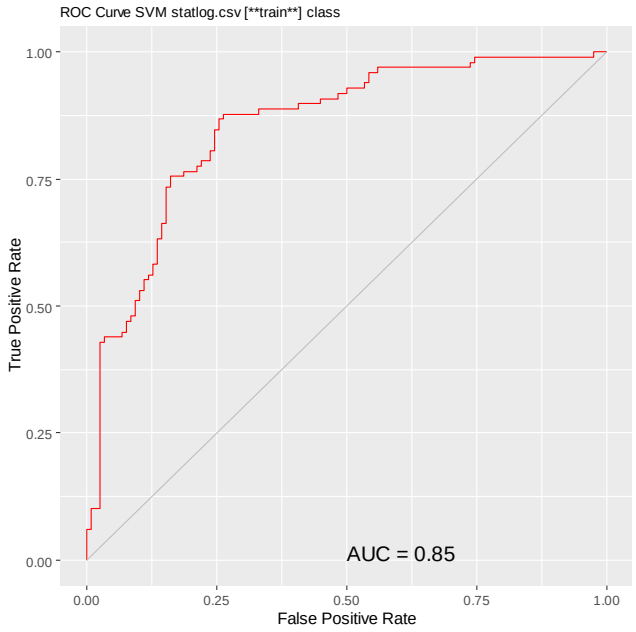
ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

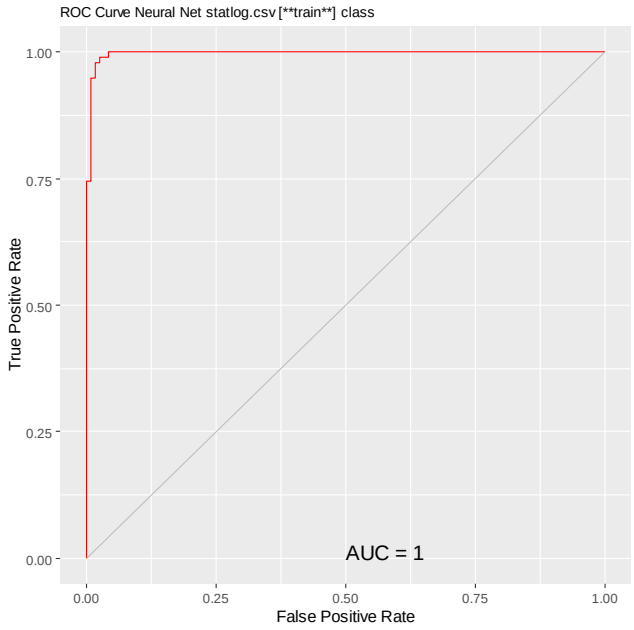
Table 6. Simulation of Statlog dataset with 80% training

MODEL	ACCURACY%	ERROR RATE %	TIME IN SECONDS
DTREE	0.856	19.55	0.02
RF	1	0	0.32
SVM	0.852	21.3	0.09
NN	0.997	2.4	0.11









In this work, first we implement LLE as the dimensionality feature reduction methods, with the combination of classification methods to diagnose the cardiovascular disease. This study concludes that by reducing features, the performance of prediction model is improved. This model calculated the highest AUC-ROC of 1(100% accuracy) with 0 error rate by using random forest (RF) on 80% training Statlog dataset.

In this section, we run two heart disease datasets; Cleveland and Statlog in R Studio to predict the cardiovascular disease. First we applied LLE as

ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

dimensionality feature reduction method and then applied the classification methods on both datasets. We divided the datasets into testing and training datasets with the ratio of 20% and 80% respectively. Finally, we conclude the results of classification methods. The result shows that Random forest is the best classifier in this model for the early diagnosis of cardiovascular disease.

V. DISCUSSION

The number of heart patient ever-increasing day by day and also the factors describing the cardiovascular disease are growing. That's why; it's so difficult to find the useful and valid information from such type of large raw data.

So, to handle out this type of difficulty, first we used local linear embedding as dimensionality feature reduction method to reduce the features. After preprocessing the datasets, in this model we implemented classification methods; DTree, RF, SVM and NN for the classification of heart patients.

Then this study calculated AUC-ROC, error rate and time in seconds of these four classification methods on datasets without testing and training. After that calculated accuracy with 20% testing and 80% training and found that Random Forest shows highest AUC-ROC as 1 (accuracy = 100%) without testing and with 80% training for both datasets and error rate is 0. The results also show that Random Forest is best classifier in this study for the prediction of cardiovascular disease with highest prediction accuracy for both Cleveland and Statlog heart disease datasets. It also shows AUC-ROC as 1 (100% accuracy) for 80% training datasets.

For this study Random Forest is the best classifier for the classification of cardiovascular disease. This model can be used for the classification of any disease. Also concludes that this model can help the biologists in early diagnosis of cardiovascular disease. Local linear dimensionality reduction techniques are used as feature reduction method. It is also time effective and cost effective because this model reduces the factors that describing the heart disease. In this study, classification methods are applied to classify the patients with presence or absence of cardiovascular disease. Accuracy graph and confusion matrix are used to show the performance of this model and also used the histograms to represent the number of patients having heart disease.

In future work, this model will be tested on gene expression dataset that having more than10K or sometimes 50K features. To handle such type of data, dimensionality feature reduction methods are very useful. So this model will be very functional and helpful to diagnose cardiovascular disease in very short time by reducing features. It will be cost effective also. This model will be tested for the prediction of any disease in future.

VI. CONCLUSION

In this study, we calculated AUC-ROC, error rate and time in seconds of four classification methods on datasets without testing and training. We also calculated accuracy with 20% testing and 80% training and found that Random Forest shows highest AUC-ROC as 1 (accuracy = 100%) without testing and with 80% training for both datasets and error rate is 0. So we concluded that Random forest is the best classifier in this model to classify the heart patients and healthy persons. We also concluded that this model is very cost and time effective for the prediction of cardiovascular disease. But our study is limited to show the results for smaller datasets. In future, this model will be tested on bigger datasets and gene expression datasets that have more than 50K features. This model will be tested for the prediction of any disease at early stage.

REFERENCES

[1] Kumar, M. N., Koushik, K. V. S., & Deepak, K. (2018). Prediction of heart diseases using data mining and machine learning algorithms and tools. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(3), 887-898.

[2] Sudhakar, K., & Manimekalai, D. M. (2014). Study of heart disease prediction using data mining. International journal of advanced research in computer science and software engineering, 4(1), 1157-1160.

[3] Karthiga, A. S., Mary, M. S., & Yogasini, M. (2017), Early Prediction of Heart Disease Using Decision Tree Algorithm, International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST) Vol.3, Issue.3, March 2017, ISSN (ONLINE):2395-695X ISSN (PRINT):2395-695X.

[4] Devi, S. K., Krishnapriya, S., & Kalita, D. (2016), Prediction of Heart Disease using Data Mining Techniques, Indian Journal of Science and Technology, Vol 9(39), DOI: 10.17485/ijst/2016/v9i39/102078, October 2016, ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645.

[5] Maaten, L. J. P. (2007), An Introduction to Dimensionality Reduction Using Matlab,, Universiteit Maastricht MICC/IKAT P.O. Box 616 6200 MD Maastricht The Netherlands, Maastricht, July 2007

International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 7 Issue I, Jan 2019- Available at www.ijraset.com.

[7] Kanikar, P., & Shah, D. R. (2016), Prediction of Cardiovascular Diseases using Support Vector Machine and Bayesian Classification, International Journal of Computer Applications (0975 – 8887) Volume 156 – No 2, December 2016.

[8] Subhadra, K., & Vikas, B. (2019), Neural Network Based Intelligent System for Predicting Heart Disease, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.

[9] Shylaja, S., & Muralidharan, R. (2019), Hybrid SVM-ANN Classifier is used for Heart Disease Prediction System, © IJEDR 2019 | Volume 7, Issue 3 | ISSN: 2321-9939.

[10] El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015), Feature Analysis of Coronary Artery Heart Disease Data Sets, International Conference on Communication, Management and Information Technology (ICCMIT 2015), Procedia Computer Science 65 (2015) 459 – 468.

[11] Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020), Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques, IETE Journal of Research, DOI: 10.1080/03772063.2020.1713916, ISSN: 0377-2063 (Print) 0974-780X (Online).

[12] Verma, L., Srivastava, S., & Negi, P. C. (2016), A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data, # Springer Science+Business Media New York 2016, J Med Syst (2016) 40:178 DOI 10.1007/s10916-016-0536-z.

[13] Verma, L., & Srivastava, S. (2016), A Data Mining Model for Coronary Artery Disease Detection using Noninvasive Clinical Parameters, Indian Journal of Science and Technology, Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/105707, December 2016, ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645.

[14] Kumar, M., Shambhu, S., & Sharma, A. (2018), Classification of Heart Diseases Patients using Data Mining Techniques, IJRECE VOL. 6 ISSUE 3 (JULY - SEPTEMBER 2018) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).

[15] Maji, S., & Arora, S. (2019), Decision Tree Algorithms for Prediction of Heart Disease, © Springer Nature Singapore Pte Ltd. 2019

[16] Sharma, S., & Parmar, M. (2020), Heart Diseases Prediction using Deep Learning Neural Network Model, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020.

[17] Ricciardi, C., Valente, A. S., Edmund, K., Cantoni, V., Green, R., Fiorillo, A., Picone, I., Santini, S., & Cesarelli, M. (2020), Linear discriminant analysis and principal component analysis to predict coronary artery disease, Health Informatics Journal, 1–12, © The Author(s) 2020.

[18] Joloudari, J. H., Joloudari, E. H., Saadatfar, H., Ghasemigol, M., Razavi, S. M., Mosavi, A., Nabipour, N., Shamshirband, S., & Nadai, L. (2020), Coronary Artery Disease Diagnosis; Ranking the Significant Features Using a Random Trees Model, Int. J. Environ. Res. Public Health 2020, 17, 731.

[19] Kolukisa, B., Yavuz, L., Soran, A., Bakir-Gungor, B., Tuncer, D., Onen, A., & Gungor, V. C. (2019), Coronary Artery Disease Diagnosis Using Optimized Adaptive Ensemble Machine Learning Algorithm, International Journal of Bioscience, Biochemistry and Bioinformatics, Volume 10, Number 1, January 2020.

[20] Bhaskaru, O., & Devi, M. S. (2019), Accurate and Fast Diagnosis of Heart Disease using Hybrid Differential Neural Network Algorithm, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019.

[21] Latha, C. B. C., & Jeeva, C. (2019), Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, Informatics in Medicine Unlocked 16 (2019) 100203, journal homepage: www.elsevier.com/locate/imu.

[22] Jan, M., Awan, A. A., Khalid, M. S., & Nisar, S. (2019), Ensemble approach for developing a smart heart disease prediction system using classification algorithms, Research Reports in Clinical Cardiology downloaded from https://www.dovepress.com/ by 85.202.195.147 on 10-Jan-2019.

[23] Pushkala, V., Agalya, T., & Angayarkanni, S. A. (2019), Comparative Study of Heart Disease Prediction Using Machine Learning Algorithms, International Journal of Innovations in Engineering and Technology (IJIET) http://dx.doi.org/10.21172/ijiet.124.10, Volume 12 Issue 4 March 2019, ISSN: 2319-1058.

[24] Patra, R., & Khuntia, B. (2019), Predictive Analysis of Rapid Spread of Heart Disease with Data Mining, 978-1-5386-8158-9/19/$31.00©2019IEEE.

[25] David, H. B. F., & Beley, S. A. (2018), heart disease prediction using data mining techniques, ISSN: 2229-6956 (ONLINE) ICTACT JOURNAL ON SOFT COMPUTING, OCTOBER 2018, VOLUME: 09, ISSUE: 01 DOI: 10.21917/ijsc.2018.0253

[26] Taneja, A. (2013). Heart disease prediction system using data mining techniques. Oriental Journal of Computer science and technology, 6(4), 457-466.

[27] Chaurasia, V., & Pal, S. (2014). Data mining approach to detect heart diseases. International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol, 2, 56-66.

[28] Aziz, A., & Rehman, A. U. (2017). Detection of Cardiac Disease using Data Mining Classification Techniques. Int. J. Adv. Comput. Sci. Appl, 8(7), 256-259.

[29] Masethe, H. D., & Masethe, M. A. (2014, October). Prediction of heart disease using classification algorithms. In Proceedings of the world Congress on Engineering and computer Science (Vol. 2, pp. 22-24).

[30] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019, January). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 619-623). IEEE.

[31] Shaji, S. P. (2019, April). Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0848-0852). IEEE.

[32] Mirmozaffari, M., Alinezhad, A., & Gilanpour, A. (2017). Heart disease prediction with data mining clustering algorithms. Int'l Journal of Computing, Communications & Instrumentation Engg, 4(1), 16-19.

[33] Malav, A., & Kadam, K. (2018). A hybrid approach for heart disease prediction using artificial neural network and K-means. International Journal of Pure and Applied Mathematics, 118(8), 103-10.

[34] Varun Sapra and Madan Lal Saini, "Computational Intelligence for Detection of Coronary Artery Disease with Optimized Features", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6C, April 2019

[35] Verma, L., Srivastava, S., & Negi, P. C. (2018). An intelligent noninvasive model for coronary artery disease detection. Complex & Intelligent Systems, 4(1), 11-18.

[36] Sudha, M. (2017). Evolutionary and neural computing based decision support system for disease diagnosis from clinical data sets in medical practice. Journal of Medical Systems, 41(11), 178.

[37] El Bialy, R., Salama, M. A., & Karam, O. (2016, May). An ensemble model for heart disease data sets: a generalized model. In Proceedings of the 10th International Conference on Informatics and Systems (pp. 191-196).

[38] Joshi, S., & Nair, M. K. (2015). Prediction of heart disease using classification based data mining techniques. In Computational Intelligence in Data Mining-Volume 2 (pp. 503-511).

[39] A. T. Sayad and P. P. Halkarnikar, "Diagnosis of heart disease usnig neural network approach", International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009 Volume- 2, Issue-3, July-2014

[40] Chaithra, N., & Madhu, B. (2018). Classification models on cardiovascular disease prediction using data mining techniques. Journal of Cardiovascular Diseases and Diagnosis. doi, 10, 2329-9517.

[41] Shekar, K. C., Chandra, P., & Rao, K. V. (2019). An Ensemble Classifier Characterized by Genetic Algorithm with Decision Tree for the Prophecy of Heart Disease. In Innovations in Computer Science and Engineering (pp. 9-15). Springer, Singapore.

[42] Singh, R., & Rajesh, E. (2019). Prediction of Heart Disease by Clustering and Classification Techniques. International Journal of Computer Sciences and Engineering, 7, 861-866.

[43] Dr. S. Anitha and Dr. N. Sridevi, "Heart Disease Prediction Using data Mining Techniques" Journal of Analysis and Computation (JAC) (An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XIII, Issue II, February 2019

[44] Miao, K. H., & Miao, J. H. (2018). Coronary heart disease diagnosis using deep neural networks. Int. J. Adv. Comput. Sci. Appl., 9(10), 1-8.

[45] Ramasamy, S., & Nirmala, K. (2020). Disease prediction in data mining using association rule mining and keyword based clustering algorithms. International Journal of Computers and Applications, 42(1), 1-8.

[46] Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC medical informatics and decision making, 19(1), 211. [41] Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q., & Wang, Q. (2017). A hybrid classification system for heart disease diagnosis based on the RFRS method. Computational and mathematical methods in medicine, 2017.

[47] K.Gomathi and dr. Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification", International Journal for Research in Applied Science & Engineering Technology (IJRASET), www.ijraset.com Volume 4 Issue II, February 2016 IC Value: 13.98 ISSN: 2321-9653

[48] Kirmani, M. M., & Ansarullah, S. I. (2016). Classification models on cardiovascular disease detection using Neural Networks, Naïve Bayes and J48 Data Mining Techniques. International Journal of Advanced Research in Computer Science, 7(5).

[49] Kautkar Rohit A, "A Comprehensive Survey on Data Mining", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308 , Volume: 03 Issue: 08 | Aug-2014, Available @ http://www.ijret.org

[50] Nikookar, E., & Naderi, E. (2018). Hybrid Ensemble Framework for Heart Disease Detection and Prediction. IJACSA) International Journal of Advanced Computer Science and Applications, 9(5).

[51] KS, D., & Kamath, A. (2017). Survey on Techniques of Data Mining and its Applications.

[52] Sharma, A., Sharma, R., Sharma, V. K., & Shrivatava, V. (2014). Application of data mining–a survey paper. International Journal of Computer Science and Information Technologies, 5(2), 2023-2025.

[53] Ghorbani, R., & Ghousi, R. (2019). Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. International Journal of Data and Network Science, 3(2), 47-70.

[54] Marikani, T., & Shyamala, K. (2017). Prediction of heart disease using supervised learning algorithms. Int J Comput Appl, 165(5), 41-4.

[55] Yekkala, I., Dixit, S., & Jabbar, M. A. (2017, August). Prediction of heart disease using ensemble learning and Particle Swarm Optimization. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 691-698). IEEE.

[56] Saranya, S., & Manavalan, R. Computational Framework for Heart Disease Prediction using Deep Belief Neural Network with Fuzzy Logic. International Journal of Computer Applications, 975, 8887.

[57] Zriqat, I. A., Altamimi, A. M., & Azzeh, M. (2017). A comparative study for predicting heart diseases using data mining classification methods. arXiv preprint arXiv:1704.02799.

[58] Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684-687.

ISSN: 2708-7123 | Volume-**01**, Issue Number-**03** | September-2020
LC INTERNATIONAL JOURNAL OF STEM
Web: www.logicalcreations.org/stem | www.lcjstem.com | DOI: https://doi.org/10.47150

[59] Kemal Akyol and Ümit Atila, "A Study on Performance Improvement of Heart Disease Prediction by Attribute Selection Methods", Academic Platform Journal of Engineering and Science 7-2, 174-179, 2019

[60] Tarawneh, M., & Embarak, O. (2019, February). Hybrid approach for heart disease prediction using data mining techniques. In International

Conference on Emerging Internetworking, Data & Web Technologies (pp. 447-454). Springer, Cham.

[61] Abdar, M., Kalhori, S. R. N., Sutikno, T., Subroto, I. M. I., & Arji, G. (2015). Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. International Journal of Electrical & Computer Engineering (2088-8708), 5(6).

[62] Karayılan, T., & Kılıç, Ö. (2017, October). Prediction of heart disease using neural network. In 2017 International Conference on Computer Science and Engineering (UBMK) (pp. 719-723). IEEE.

[63] Hasan, S. M. M., Mamun, M. A., Uddin, M. P., & Hossain, M. A. (2018, February). Comparative Analysis of Classification Approaches for Heart Disease Prediction. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.

[64] Jabbar, M. A., & Samreen, S. (2016, October). Heart disease prediction system based on hidden naïve bayes classifier. In 2016 International Conference on Circuits, Controls, Communications and Computing (I4C) (pp. 1-5). IEEE.

[65] S. Mohan et al., "Effective Heart Disease Prediction Using Hybrid ML Techniques", VOLUME 7, 2923707, 2019, IEEE Access

[66] Jabbar, M. A. (2017). Prediction of heart disease using k-nearest neighbor and particle swarm optimization.

[67] Lakshmi Devasena. C, "Performance Evaluation of Memory Based Classifiers With Correlation Based Feature Selection Subset Evaluator For Smart Heart Disease Prediction", IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308

[68] Takci, H. (2018). Improvement of heart attack prediction by the feature selection methods. Turkish Journal of Electrical Engineering & Computer Sciences, 26(1), 1-10.

[69] R.Suganya et al., "A Novel Feature Selection Method for Predicting Heart Diseases with Data Mining Techniques", Asian Journal of Information Technology 15 (8): 1314-1321, 2016, ISSN: 1682-3915

[70] Pandey, A. K., Pandey, P., & Jaiswal, K. L. (2013). A novel frequent features prediction model for heart disease diagnosis. International Journal of Engineering Mathematics and Computer Sciences, 1(2).

[71] Dangare, C. S., & Apte, S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-48.

[72] Yekkala, I., & Dixit, S. (2018). Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection. International Journal of Big Data and Analytics in Healthcare (IJBDAH), 3(1), 1-12.

[73] Saboji, R. G. (2017, August). A scalable solution for heart disease prediction using classification mining technique. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1780-1785).

[74] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. Journal of Network and Innovative Computing, 4(2016), 175- 184.

[75] Reddy, N. S. C., Nee, S. S., Min, L. Z., & Ying, C. X. (2019). Classification and feature selection approaches by machine learning techniques: Heart disease prediction. International Journal of Innovative Computing, 9(1).