iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IV Month of publication: April 2023

DOI: https://doi.org/10.22214/ijraset.2023.51250

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Prediction of Coronary Heart Disease using Deep Learning Models

Gengadevi. A
Department of Computer Science and Information Technology, Kalasalingam Academy of Research and Education, 626126, Tamil Nadu, India

Abstract:* More over 80% of deaths from heart disease, including those in Nigeria, are caused by coronary artery disease (CAD), the most prevalent type. The victims were mostly less than 70 years old. More than 17 million people died in 2015 from CVD-related causes, accounting for more than 30% of all deaths worldwide. It develops over time and progresses through several stages. The stages of CAD include Fatty Streaks, Mild atherosclerosis, Moderate atherosclerosis, Sever atherosclerosis. In this paper, a diagnostic CAD dataset that was collected from the Kaggle website was used to construct a machine learning predictive model for CAD. The dataset was used to create prediction models using machine learning techniques like Naive Bayes, Support Vector Machine, Random Forest, and Gradient Boosting and DNN are applied to compare the results and analysis of the CAD Disease. The models' accuracy, precision, recall, and f1score scores were assessed using performance evaluation methodologies in order to determine the CAD that would have or had already occurred based on a person's physical condition and data from medical records. Result shows that compared to ML algothims and DL technique, DNN gives more accuracy in less time for the prediction. The prediction accuracy obtained by DNN algorithm is 93.56% and the prediction accuracy obtained by SVM is 83.34%.

Keywords: Coronary Artery, Random Forest, Naive Bayes, SVM, XG Boosting, DNN.

I. INTRODUCTION

The most common type of heart disease in the US is coronary artery disease (CAD), which affects heart blood flow. Reduced blood flow may cause a heart attack. When a buildup of fatty substances in the coronary arteries inhibits or interferes with your heart's blood flow, it is referred to as coronary heart disease.[3] Fatty deposits may eventually form on the walls of your arteries

These now comprise the right coronary artery, posterior descending coronary artery, left major coronary artery, left anterior descending coronary artery, and left circumflex coronary Artery. Coronary heart disease, Cerebrovascular disease, rheumatic heart disease, among other illnesses, are among the heart and blood vessel disorders collectively referred to as CVDs.[6]

The main behavioural risk factors for heart disease[8] and stroke include poor diet, inactivity, smoking, and problematic alcohol use. Only a few symptoms of a person's exposure to behavioural risk factors include high blood pressure, high blood sugar, high blood lipids, overweight, and obesity. These "intermediate risk factors" indicate an increased risk of heart attack, stroke, heart failure, and related effects and can be examined in primary care settings. It is possible to swiftly study these datasets, which are too big for human brains to absorb, using a number of machine learning techniques.

As a result, these algorithms have become much better over the past few years at predicting the existence or absence of heart-related illnesses. The main objective of this research is to develop and use a reliable disease prediction model. Using a variety of algorithms, including Navie Bayes, XGB, SVM, Random Forests, and DNN, the fast evolving field of artificial intelligence can draw conclusions and forecasts from the enormous volumes of data produced by the healthcare industry. Based on the given issue, ML and DL offer a number of classification techniques to estimate the probability that a patient would develop heart disease.[7]

A variety of illnesses that affect your heart are referred to as heart disease. According to estimates from the World Health Organization, cardiovascular[1]illnesses are currently the top cause of mortality worldwide, accounting for 17.9 million fatalities per year. According to World Health Organization, the risk of heart disease [2] has increased as a result of several harmful behaviours, including excessive cholesterol, obesity, an increase in triglyceride levels, hypertension, etc. A few symptoms are listed by the American Heart Association, including trouble sleeping, an erratic heartbeat, swollen legs, and, in some cases, weight gain that happens quickly—up to 1-2 kilogrammes per day. All of these symptoms are similar to various illnesses that often affect young people, making a proper diagnosis challenging. This leads to fatal cities shortly. Exercise stress tests, chest X-rays, heart scans (CT), cardiac magnetic resonance imaging (MRI), coronary angiograms, and electrocardiograms (EKG) are currently utilized to determine the degree of heart disease in patients.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

A. Architecure

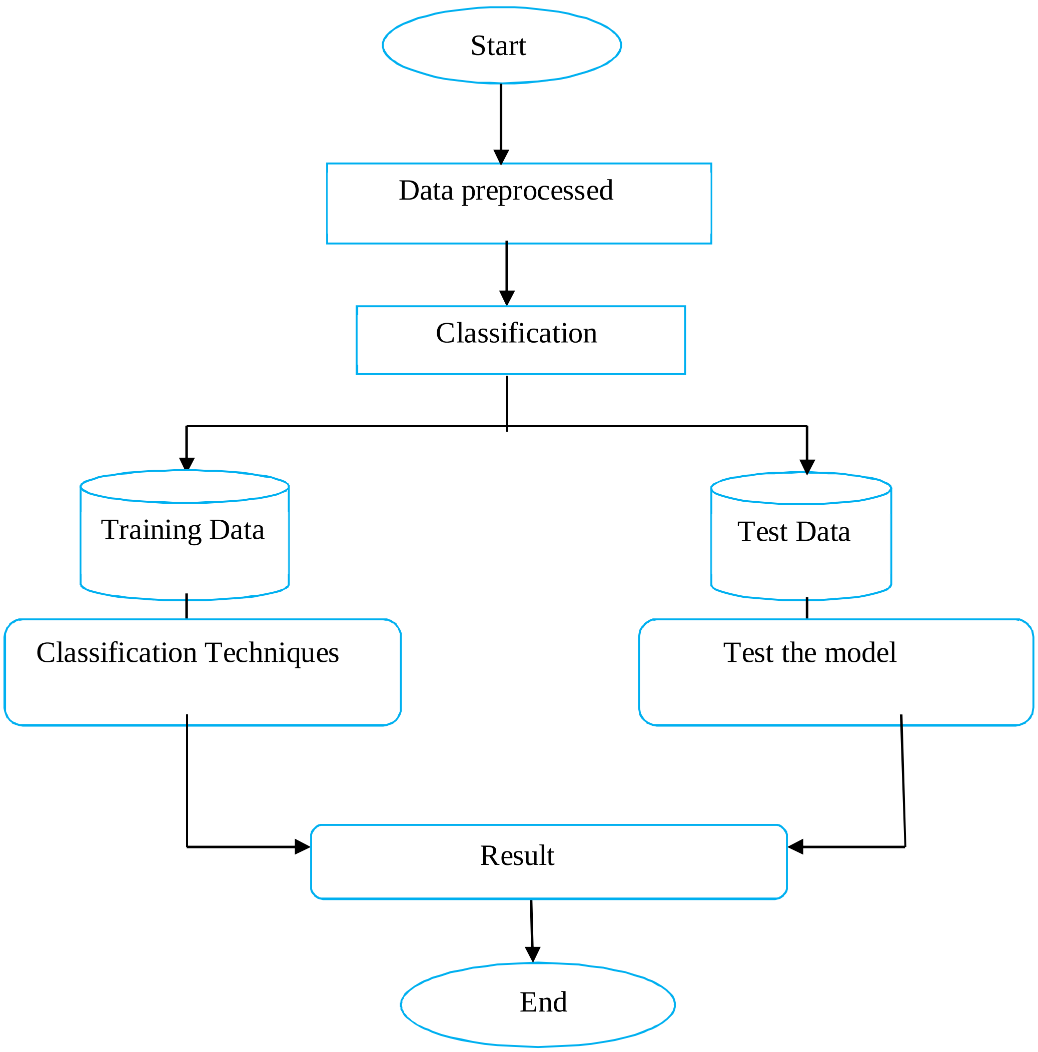

Fig 1: Architecture of the proposed model

II. LITERATURE REVIEW

Using the Kaggle Machine Learning dataset, extensive research has been done to predict cardiac disease. Several data mining techniques have resulted in varied accuracy levels.

Approaches that are described here.

A. K. Grate-Escamila et al. [4] Proposed by The development of machine learning predictive models for CAD using diagnostic CAD datasets from the two general hospitals in Kano State, Nigeria, is one of the major contributions. The dataset was used to build predictive models using machine learning algorithms such as logistic regression, support vector machines, K nearest neighbours, random trees, Naive Bayes, gradient boosting, and naive Bayes. The models were assessed based on their accuracy, specificity, sensitivity, and receiver operating curve (ROC) performance evaluation methods.

AshirJaveed et al. [10] Discussed by The donations, The suggested diagnostic system employs the random forest model for heart failure prediction and the random search algorithm (RSA) for feature selection. Using the grid search technique, the proposed diagnostic system is optimised. To assess the accuracy of the suggested method, two different types of experiments are conducted. The proposed RSA-based random forest model is developed in the second experiment, whereas the first experiment just develops a random forest model. The Cleveland dataset, an online collection of heart failure data, is used in experiments.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Dengqing Zhang et al. [5] Proposed by . major contributions, By a variety of bodily indications, our research aims to precisely and promptly anticipate cardiac disease. An innovative model for predicting cardiac disease is presented in this research. We provide an approach for predicting cardiac disease that incorporates deep neural networks and the embedded feature selection method. The L1 norm is used as a penalty item in the embedded feature selection approach, which is based on the LinearSVM algorithm, to choose a subset of features significantly linked with heart disease. The deep neural network we created is fed with these properties. In order to improve the performance of the predictor, gradient varnishing or explosion is prevented by initialising the network's weight with the He initializer. Anna Karen Garate-Escamilla et al. [9] Discussed by The work's major contributions are the suggestion of a dimensionality reduction method and the use of a feature selection strategy to identify heart disease features. The Heart Disease section of the UCI Machine Learning Repository provided the data for this analysis. Chi-Square Selector used the study to derive anatomical and physiologically significant aspects, such as cholesterol, the greatest heart rate, chest pain, traits associated with depression, and cardiac vessels. The experiment's findings demonstrated that most classifiers perform better when chi-square and PCA are combined. Yet the precision is poor. Ahmed AI Ahdal et al. [1] to create a variety of machine learning techniques based on the medical characteristics of people and the UCI set of data to help with the early diagnosis of cardiovascular disease.This will make it easier for the doctor to act in the right way. Only the presence of a heart condition will be able to be determined by the proposed technique. This approach cannot assess the degree of cardiac disease.
As mentioned above to improve the accuracy, the most crucial stage of the software development process is literature research. Determine the time factor, profitability, and company strengths before building a tool.

III. METHODOLOGY

A. Proposed Methology

The medical system is employed to anticipate disease at an early stage. Only after that can the death rate be decreased. This study's goal is to accurately determine whether a patient has heart disease. The healthcare provider inputs the data from the patient's health report. The proposed work predicts heart disease by Comparison of Machine learning and Deep learning techniques. There are machine learning algorithm used namely SVM,Random forest,XGB,Gaussian Navie Bayes and compared to DNN gives more accuracy in less time for the prediction. which displays the accuracy rate of guesses of heart disease. Three modules make up the suggested system.
There are

1) Data Collection

Coronary artery disease dataset with 16 parameters as RID, age, sex, family history, smoking, chest pain, diabetes, etc. For the attributes of the given dataset, multiclass variables and binary classification are introduced.

.No	Feature	Units	Range
1	Age	Years	24-97
2	Sex	Male(1),Female(0)	0,1
3	Family History	Present(1), Absent(0)	0,1
4	Smoking	Yes(1), No(0)	0,1
5	Diabetes	Yes(1), No(0)	0,1
6	Hypertension	Yes(1), No(0)	0,1
7	Blood Pressure	Yes(1), No(0)	0,1
8	Anaemia	Yes(1), No(0)	0,1
9	Chest pain	Asymptomatic(0),Non angina l pain(1),Atypical pain(2),Typical angina(3)	0-3
10	Glucose	mg/dl	36-480
11	Cholesterol	mg/dl	141-564
12	IDL	mg/dl	1.87-15.33
13	Creatinine	mg/dl	0.3-10.66
14	BMI	Kg/m	10-50
15	Heart Rate	Bpm	48-120

Table-1: Description of the dataset features

2) Data Preprocessing

A Coronary artery Disease dataset was gathered for this study via the Kaggle website. There are 1498 patient records in the dataset, The "target" field alludes to the patient's having heart illness.0 means there is no disease, while 1 means there is a disease. When building a model, data preprocessing is essential because it removes undesired noise and outliers from the dataset that could cause the model to diverge from the training it was intended for. This stage addresses anything impeding the model from performing more effectively. Data preprocessing for CAD involves gathering of numerous records and analyzing the data to identify patterns and relationship that can predict and prevent the disease. The relevant dataset must be acquired, cleaned, and ready before modelling can begin. The dataset used has 16 features, as was previously mentioned. First, the RID is ignored because it has no relevance to how the model is built. For data preprocessing, a number of Python libraries were employed, including Pandas, Numpy, Scikit Learn, Seaborn, and Matplotlib. A variety of functions are offered by these libraries for processing, transforming, analysing, and displaying the phishing dataset.

3) Classification and Prediction

a) Support Vector Machine

SVM can be used to predict the presence or absence of coronary artery disease(CAD) in the condtext of CAD based on a variety of risk variables. The SVM algorithm's objective is to provide the optimal line or decision border that can categorize n-dimensional space, allowing us to quickly assign additional data points to the appropriate category in the future are,

- Specify the problem, the input characteristics, and the goal variable.
- Divide the dataset into training and verification sets.
- Normalise or standardise the features' scales.
- Use the training set to train the SVM model.
- Utilise cross-validation to adjust the hyperparameters.
- Utilise performance indicators like accuracy, precision, recall, and F1-score to assess the model on the testing set.

b) Random Forest

Both classification and regression are accomplished using Random Forest algorithms. On huge databases, it operates effectively. The goal of random forests is to convert a set of decision trees with high variation and low bias into a model with low variance and low bias. It is employed in ML to address Arrangement and Relapse challenges.

- Define the problem, the attributes of the input, and the goal variable.
- Create training and verification sets from the dataset.
- Use the training set to train the Random Forest model.
- Apply cross-validation to the hyperparameter adjustments.
- To evaluate the model on the testing set, use performance metrics like as accuracy, precision, recall, and F1-score.

c) XGBoost

The gradient-boosting framework is used by the Xgboost ensemble machine learning algorithm, which is based on decision trees. The goal function XGBoost used to forecast numerical values. In prediction problems involving unstructured data, artificial neural networks usually outperform all existing algorithms or frameworks (pictures, text, etc).

- Define the desired variable, the problem, and the characteristics of the input.
- From the dataset, produce training and verification sets.
- Use the training set to train the XGB model.
- The tweaks to the hyperparameters should be cross-validated.
- Use performance indicators like accuracy, precision, recall, and F1-score to assess the model on the testing set.

d) Gaussian Navie Bayes

The performance goal of this supervised Naive Bayes classifier is to properly predict an incoming test instance using the class label of the training instance. Identify the problem, the target variable, and the characteristics of the input.

- Create training and verification sets from the dataset.
- Use the training set to train the Gaussian Naive Bayes model.
- Use performance metrics like accuracy, precision, recall, and F1-score to assess the model on the testing set.

e) DNN (Deep Neural Network)

The DNN model differs from a traditional multilayer perceptron neural network classification model. One of the key differences is regarding the network depth, which depends on the number of hidden layers in the network. The DNN classification model usually utilizes a regularization algorithm, which would decrease the complexity of the DNN model while maintaining the same number of large parameters.

Input: Defining the input data is the initial stage in any machine learning method.

Split: It is customary to divide input data into training and validation sets after it has been defined.

For loop: The model is iteratively trained using the training data and a for loop.

Train: By modifying the neural network's weights to reduce the difference between the projected output and the actual output, the model is trained on training data.

Predict: The model can be used to anticipate the existence of coronary disease in new individuals once it has been trained.

Return: Finally, the model and its weights are returned to be used for future predictions.

IV. EXPERIMENTAL RESULT AND DISCUSSION

The purpose of the study was to evaluate the performance of an ensemble machine learning technique for heart disease prediction. Several machine learning models, including Random Forests, Support Vector Machines, Naviebayes, XGBoost Classifier, and DNN were trained and tested on the dataset. . Result shows that compared to ML algothims and DL technique, DNN gives more accuracy in less time for the prediction The findings demonstrated that the suggested strategy outperformed other methods and individual models, achieving high accuracy, precision, and recall, as well as a high F1 score in prediction phishing websites.

A. Predicting Heart Disease

The training set is different from the test set. In this study, we used this method to verify the universal applicability of the methods. In the k-fold cross-validation method, the whole dataset is used to train and test the classifier for coronary heart disease. We have compared with several algorithms to predict the accuracy result.

Comparison Table for all algorithms

ML Model	Accuracy	F1_Score	Recall	Precision
SVM	83.34	0.82	0.80	0.82
Random Forest	83.34	0.81	0.79	0.81
XGBoost	83.33	0.80	0.80	0.82
Gaussian Naïve Bayes	83.33	0.82	0.81	0.81
DNN	93.56	0.93	0.89	0.89

Table-2: Accuracy comparison Table for various algorithm

Where Table-2, To diagnose CAD,was used to create prediction models using machine learning techniques like Naive Bayes, Support Vector Machine, Random Forest, and Gradient Boosting and DNN are applied to compare the results and analysis of the CAD Disease. Result shows that compared to ML algothims and DL technique, DNN gives more accuracy in less time for the prediction. The prediction accuracy obtained by DNN algorithm is 93.56% and the prediction accuracy obtained by SVM is 83%.

B. Graphical Representation

The analyses of proposed systems are calculated based on the approvals and disapprovals. This can be measured with the help of graphical notations such as bar charts. The data can be given in dynamic data.

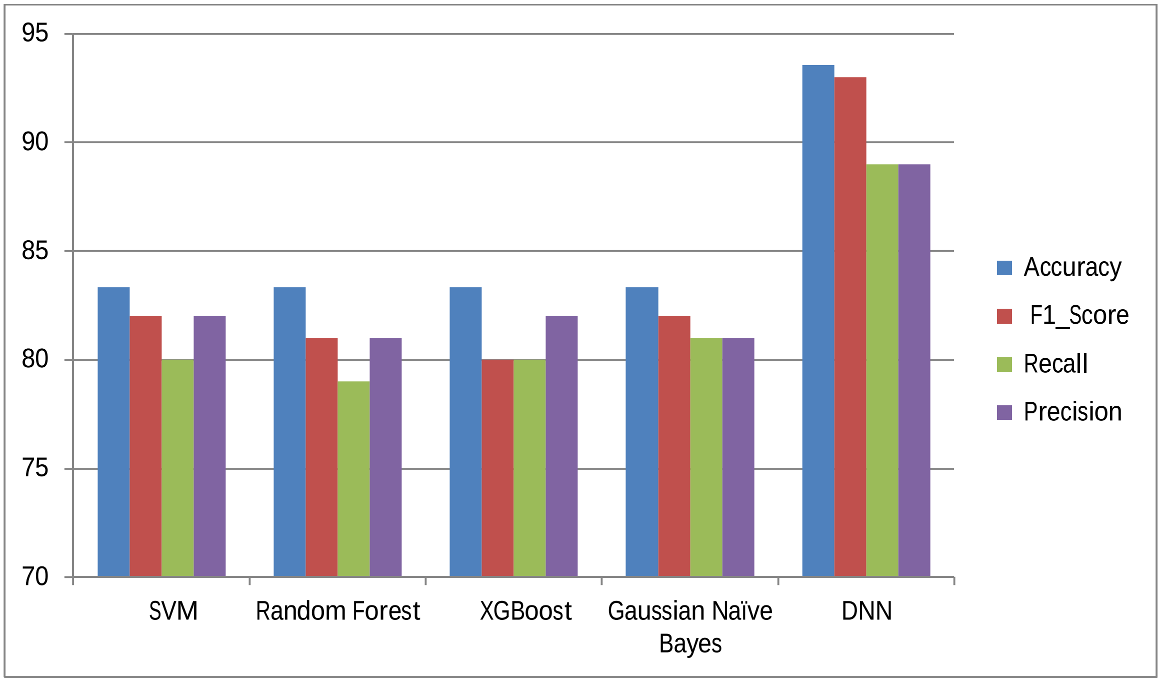

Fig 2: Algorithm Comparision

Here, Fig 2 Shows evaluation of the proposed approach and classification algorithms based on their accuracy, precision, recall, and F-measure

C. Comparison with Existing Research

The study compares existing studies in the field of heart disease that used machine learning, using a variety of data mining techniques, such as Navie Bayes, Logistic Regression, Decision Trees, and Random Forests. The results show that, when compared to other ML algorithms used, the Random forest approach has the highest accuracy (90.16). In my proposed system, they used an advanced deep learning technique.

Algorithm	Precision	Recall	F-measure	Accuracy
Decision Tree	0.845	0.823	0.835	81.97%
Random Forest	0.937	0.909	0.909	90.16%
Logistic Regression	0.845	0.869	0.869	85.25%
Gaussian Naïve Bayes	0.837	0.873	0.873	85.25%

Table-3: Comparison of our methodology with previous research

Here, Table-3 shows that Methodology comparsion on existing system.

V. CONCLUSION

A machine learning predictive model for the prediction of coronary artery diseases has been developed with the medical expert diagnostic dataset CAD. The dataset was partitioned into an 80% training set and 20% testing set, respectively, where the models were trained with the 80% and tested with the 20% dataset The dataset was applied to machine learning algorithms including support vector machine, random tree, Naïve Bayes, gradient boosting, deep neural network algorithms to build the predictive models and the models were evaluated based on accuracy,precision,recall,F1 score performance evaluation techniques. As a result, the created deep learning classification and prediction models can offer extremely trustworthy and accurate diagnoses for coronary heart disease and decrease the quantity of incorrect diagnoses that could potentially hurt patients. Hence, the models can be applied to help patients and healthcare professionals around the world improve public health and global health, particularly in underdeveloped nations and resource-constrained regions where there are fewer cardiac specialists available. Predicting heart disease is a difficult task in today's world. By entering the report values, the patient or user can utilise this application to anticipate disease even if they are not in close proximity to a doctor. and decide whether to seek medical advice or not before moving further.

REFERENCES

[1] Ahmed AI Ahdal.,ManikRakhra.,& Rahul Rajendran.,R.(2023) Phishing Detection: "Monitoring Cardiovascular problems in heart patient using Machine learning".2023,Journal of Health Engineering,vol 2023,Article ID 9738123.

[2] Chintan M.Bhatt.,Parth Patel.,& Tarang Ghetia.(2023) "Effective Heart Disease Prediction using Machine Learning Techniques" 2023,MDPI Journal,Vol 16,Issue 2

[3] Chaimaa Boukhatem.,Heba Yahia Youssef.,Ali Bou Nassif.(2023) "Heart Disease Prediction using Machine Learning" 2022,IEEE xplore Advances in science and Engineerning Technology International Conferences.

[4] L.J.Muhammad.,Ibrahem AI-Shourbaji.,& Ahmed Abba Haruna.(2021) "Machine learning predictive models for coronary Artery Disease".2021,Springer Nature Journal.

[5] Dengqing.,Yunyichen.,&Yuxuan chen.(2021). "Heart Disease prediction Based on the embedded feature selection Method and Deep Neural Network"2021,Journal of Healthcare Engineering

[6] Pronab Ghosh.,SamiAzam.,& Mirjam Jonkman.(2021)."Efficient Prediction of cardiovascular Disease using machine learning Algorithms with Relief and LASSO feature selection Techniques" 2021,IEEE xplore Journal,Vol 9,ISSN:2169-3536.

[7] Surai Shinde.,Juancarios.,Martinez Ovando. (2021)."Heart Disease Detection with Deep learning using a combination of Multiple Input Sources" 2021, IEEE fifth Ecuador Technical Chapters Meeting (ETCM).

[8] Apurb Rajdhan., Milan Sai.,& Poonam Ghuli.(2020)."Heart Disease prediction using machine learning" 2020,International Journal of Engineering Research & Technology(IJERT),Vol.9,Issue 04,ISSN: 2278-0181

[9] Anna Karen Garate – Escamila.,Amir Hajjiam EL. Hassani.,& Emmanuel Andres (2020). "Classification Models for heart Disease prediction using feature selection and PCA" 2020,IEEE xplore Journal.

[10] Ashir Javeed.,Shijie Zhou.,& LIAO YoungJian. (2017). "An intelligent Learning system based on Random Search Algorithm and optimized Random Forest Model for improved Heart Disease Detection".2017,IEEE Transactions and Journals.