HEART DISEASE PREDICTION USING DEEP LEARNING AND MACHINE LEARNING

Deepak Pippiri, M.Tech Student, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, Email: *pippirideepak@gmail.com*
Dr. Rajiv Kode, Associate Professor, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, Email: *rajiv.research31@gmail.com*

Abstract: Heart disease is a serious threat to human survival as a result of considerable lifestyle changes. Accurate prediction and diagnosis become more critical for early prevention, detection, and treatment. To forecast this lethal disease, I have used Deep Learning models with data obtained from a few people, data containing their previous medical records and family background details like their family members previously affected with any heart disease will be helpful in accuracy prediction. I am going to use Deep Learning techniques namely Artificial Neural Network (ANN) for prediction of Heart disease. We are going to predict whether the person will be affected by heart disease or not in this project using data.

Keywords: Artificial Neural network, Navie Bayes, Logistic Regression, and Support Vector Machine

INTRODUCTION

Risk identification of a disease is an important stage in the medical sector that may be utilized to avoid an illness. Early detection can also be used to increase the effectiveness of therapy. According to a 2018 WHO research, heart disease is the leading cause of death in developing nations like Indonesia, accounting for 35% of the proportionate mortality rate. Given the high risk of heart disease in Indonesia, it is crucial to look for remedies to stop the condition before it gets out of hand. Short hypotheses, a debate, and a conclusion are covered in this project. Predictive analysis and the assessment of the model used in this project will both be explained in the discussion. Structure-based algorithms are the main focus of the machine learning discipline known as deep learning. Additionally, convolutional neural networks (CNNs) and long-term memory networks are examples of artificial neural networks (ANNs), which perform similar activities to the human brain (LSTMs).

RELATED WORK

Machine learning algorithms were employed in prediction studies of cardiovascular illness by Reldyn Williams, TokozaniShongwe, Ali N. Hasan, and Vikash Rameshar. This investigation has the benefit of allowing you to apply a machine learning algorithm to your dataset. This is what happened. Received 67.21 % KNN, 85.25 % NB, 81.97 % decision tree, 81.97 % SVM, 85.25 % logistic regression, 95.08 % random forest, and 85.25 % XG boost. [1].

In ECG data analysis and heart disease prediction using machine learning algorithms research,S. R. Tithi, A. Aktar, F. Aleem and A. Chakrabarty used Machine Learning algorithms, advantages in this research paper they have predicted various heart disease and the best algorithm to predict disease, in their research work they predicted which algorithm is efficiently work for a particular disease, here is the result follows Disease Name, BestAlgorithm, ScoreRight, for disease Bundle

Branch Block Logistic Regression algorithm is working efficiently and the score is 96%, for disease Myocardial Infarction Decision Tree algorithm is working efficiently and the score is 96% ,for disease Sinus Bradycardia Decision Tree algorithm is working efficiently and the score is 95% ,for disease CADNaive Bayes algorithm is working efficiently and the score is 94%[2].
In her research on technical and predictive analytics for heart disease using machine learning and deep learning, IndrajaniSuteja used machine learning and deep learning algorithms. The advantage of this study is that the accuracy obtained in this study is 86% logistic regression. 88%, Naive Bayes is 86%, but LSTM is 84% accurate, RNN is 90%, CNN is 84%. The conclusion of this study is that RNN models will propagate with up to 90% accuracy and are ideal for predicting whether a person has heart disease. [3].

J. Thomas and Theresa Princy employed machine learning and data mining approaches in their study on the prediction of human heart disease utilizing these techniques, taking use of the NN, KNN algorithm, DT, and NB. Utilizing data mining tools, scientists were able to identify the risk factor for heart disease. Due to its primary concentration on data mining techniques, this sort of research may not have much to add to the understanding of cardiac disease. [4].

M. Gandhi and S. N. Singh employed machine learning and data mining approaches in their research on heart disease predictions. A medical data set was evaluated for this study, and decision tree (DT), naive bayes (NB), and neural network (NN) techniques were utilized. It was discovered that many characteristics are influenced, thus the number of features must be constrained. According to their analysis, reducing the amount of characteristics can cut down on processing time. They used the neural networks and decision tree techniques. The research was strong in terms of accuracy in predicting cardiac disease using decision trees and neural networks. But rushing might compromise the precision of the results. [5].

PROPOSED WORK

Architecture

The architecture of my project is shown below in fig, In this project firstly we are going to acquire the data, then the acquired data is preprocessed like replacing the empty spaces with mean, median or mode of the total attribute, we will be getting clean data, the cleaned data will be divided into two parts training dataset training contains 80% of the data and the testing data set will be 20% of the datasets, on training dataset will be training the deep learning logarithm to obtain test model, here we are using ANN (Artificial Neural Network), A shaft input, output layer, and one or more hidden layers are used for creating artificial neural networks. The network predicts in an attempt to learn from the data provided. ANN is the simplest type of neural network (artificial neural network). The resulting model has been tested on the testing dataset and the resulting test precision can be determined. ANN does not have a defined structure. There is only one neural layer used in guessing.

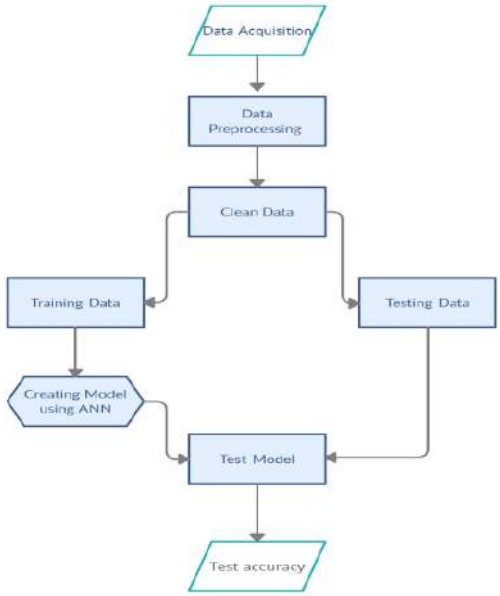

FIGURE 1. Architecture

User Interaction

In this research the user will be giving their information according to the attributes, here the model will be predicting the user will be effected with heart disease, the user can know they having any heart disease or may be in future they will be effect with other kind of heart diseases.

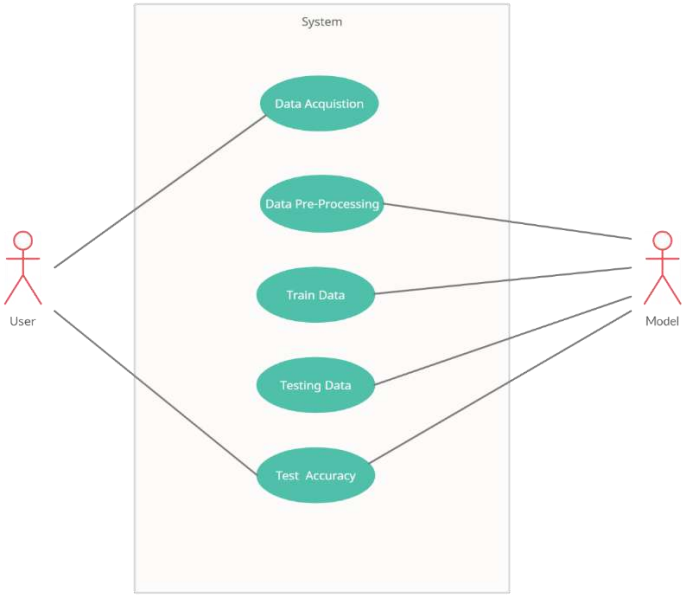

FIGURE2. Use Case Diagram

The data is collected from several hospitals containing both person effected by heart disease and normal person data, it contains 14 attributes which are shown intable below **FIGURE 3**.

NO	ATTRIBUTE	TYPE	INFORMATION
1.	age	Integer	Age in year
2.	sex	Integer	Gender (number 1 = male; 0 = female)
3.	cp	Integer	Chest pain type
4.	restbpS	Integer	Blood pressure(mm unit Hg in the initial stage)
5.	Chol	Integer	Cholesterol with mg unit/dl
6.	Fbs	Integer	Blood sugar, if the blood sugar is > 120 mg/dl, then (1 = true; 0 = false)
7.	restecg	Integer	electrocardiography
8.	thalach	Integer	Maximum heart beat
9.	exang	Integer	angina which is induced by exercise (1 = yes; 0 = no)
10.	oldpeak	Float	ST depression caused by too much exercise than taking a rest
11.	slope	Integer	Segment slope of peak ST
12.	ca	Integer	Number of major vessels
13.	thal	Integer	Thalium Stress Test result ~ (0,3)
14.	target	Integer	having previous disease or not (1 = yes; 0 = no)

FIGURE 3. Dataset

The data set contains around 1024 values and the data set will be divided into two parts which contains of 75 percent of dataset for the training of the model and 25 percent of the dataset for the testing of the model.

IMPLEMENTATION

Deep Learning models namely Artificial Neural network, Navie Bayes, Logistic Regression, Support Vector Machine models are using in this research work to detect heart disease and we are going to compare the results from model with each other.

Artificial Neural Network

In that it is composed of several layers of neurons, exactly like the human brain, deep learning is a technology that replicates the functioning of the brain. There are connections between an input layer, an output layer, and one or more hidden layers. Review the network's given data before making any predictions. ANN is the most common form of neural network (artificial neural network). Only a portion of the neural layer utilized for prediction contains ANN, which lacks a defined structure. Make the model shown below, The methods, training, and assessment to generate RNA is the first RNA, regardless of whether a person uses RNA or has a cardiac disease, Compile the ANN, train it on the training set, generate predictions, and assess the model as described. Add an input layer, a first hidden layer, a second hidden layer, and an output layer, to do. **FIGURE 4, FIGURE 4.1 FIGURE 4.2, FIGURE 4.3.**

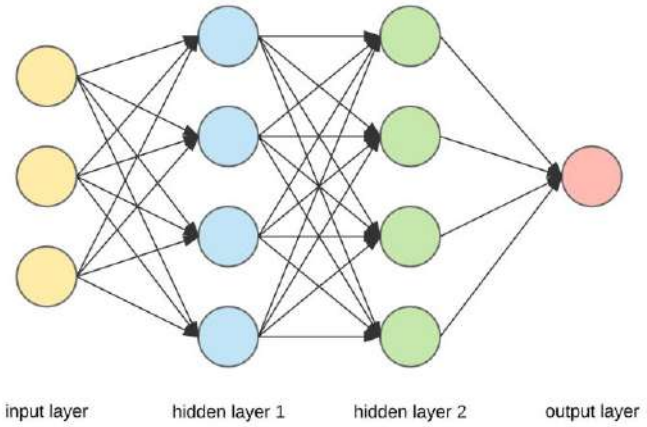

FIGURE 4. Architecture of ANN model

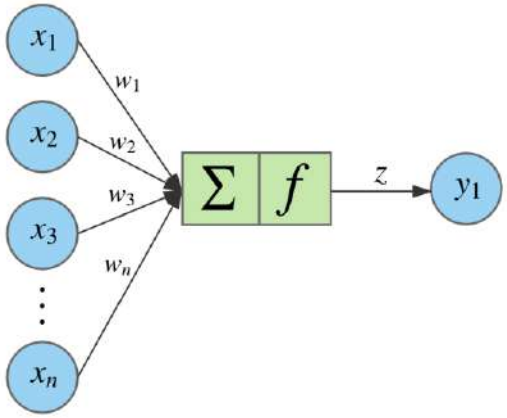

FIGURE 4.1

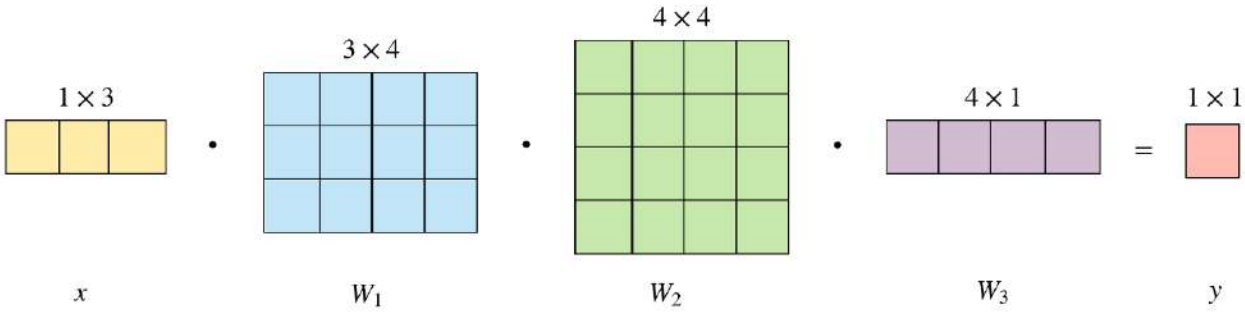

FIGURE 4.2

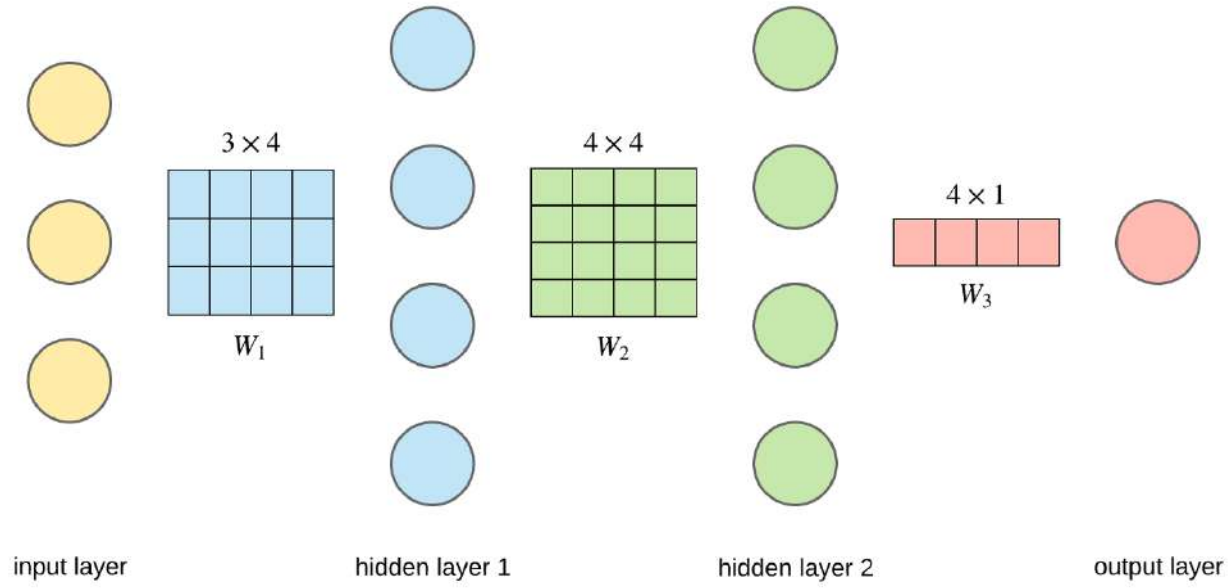

FIGURE 4.3

The output of a certain node is then used as the input of a different node in the subsequent layer after being subjected to the weighted sum of its inputs and a non-linear activation function. This process is carried out for each node, and the ultimate output the signal flowing from left to right is determined. Learning the weights corresponding to each edge in this deep neural network requires training.

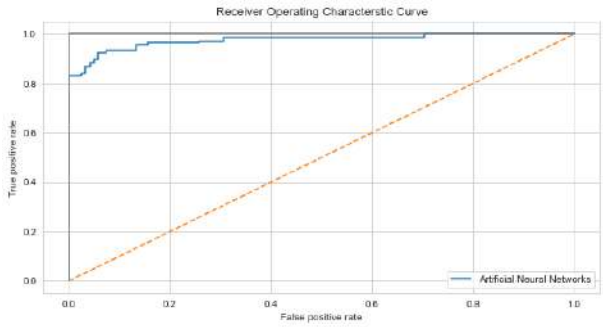

FIGURE5.ROC curve obtained by validating ANN model

Navie Bayes

An exclusive subset of NB algorithms is the Gaussian Naive Bayes algorithm. This is particularly useful when the function's values are continuous. Additionally, it is presumed that all functions have a normal distribution, or a Gaussian distribution.

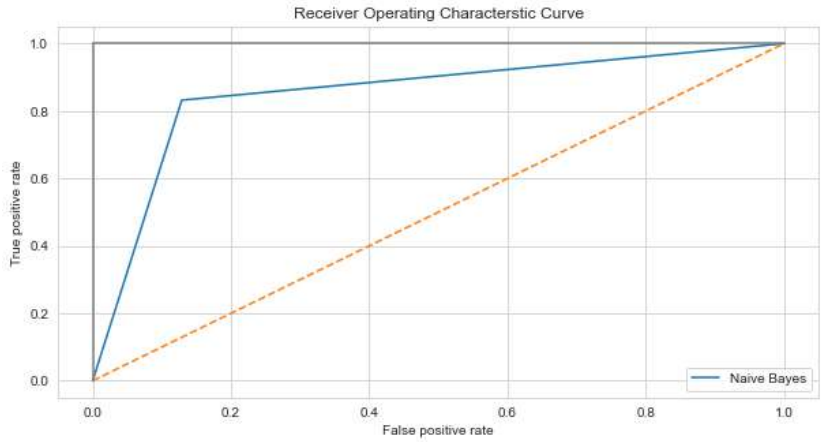

FIGURE6. ROC curve obtained by validating Navie Bayes model

Logistic Regression

Logistic regression (LR), despite its name, is a binary classification method. It is the most often used method of 0/1 categorization. The phrase "linear model" refers to LR's attempt to create a straight line to divide the classes on a two-dimensional (2D) data set. However, LR is not limited to two dimensions; it may function with any number of dimensions. It will attempt to create a 2D plane to divide the classes in 3D data. This applies generally to data in N dimensions and a hyperplane separator in N-1 dimensions. Given an input data set with numerous columns and a binary 0/1 conclusion, if you have a supervised binary classification issue

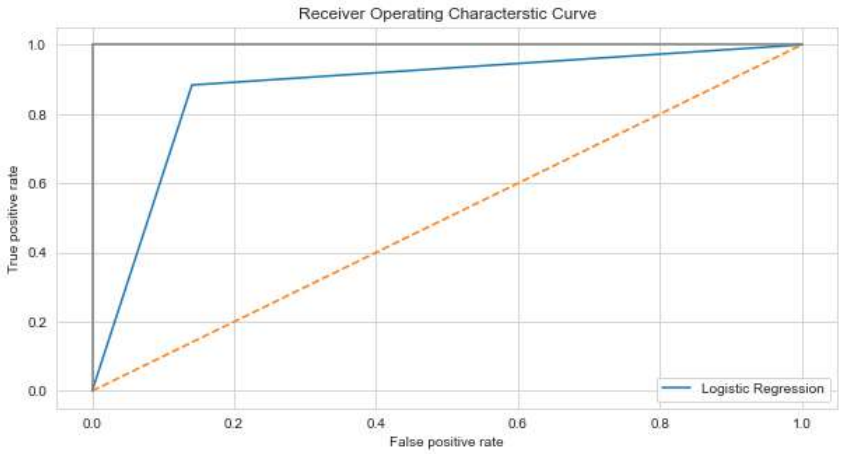

FIGURE7. ROC curve obtained by validating Logistic Regression model

Support Vector Machine

One of the most popular supervised learning techniques for addressing classification and regression issues is the use of support vector machines (SVMs). However, machine learning classification issues frequently employ it. The SVM algorithm's objective is to produce the best judgement or boundary lines for categorising an n-dimensional space.allowing for rapid future classification of additional data points. The term hyperplane is used to describe this ideal

decision boundary. SVM chooses endpoints and vectors to aid in the hyperplane's construction. The SVM approach is built around the support vectors that are used to represent these extreme occurrences.Think on the illustration below. Here, we categorise two distinct groups using decision boundaries or hyperplanes.

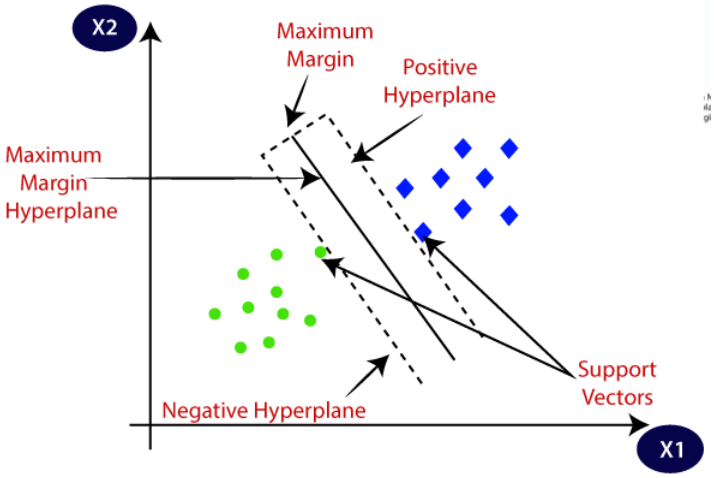

FIGURE 8. Support Vector Machine

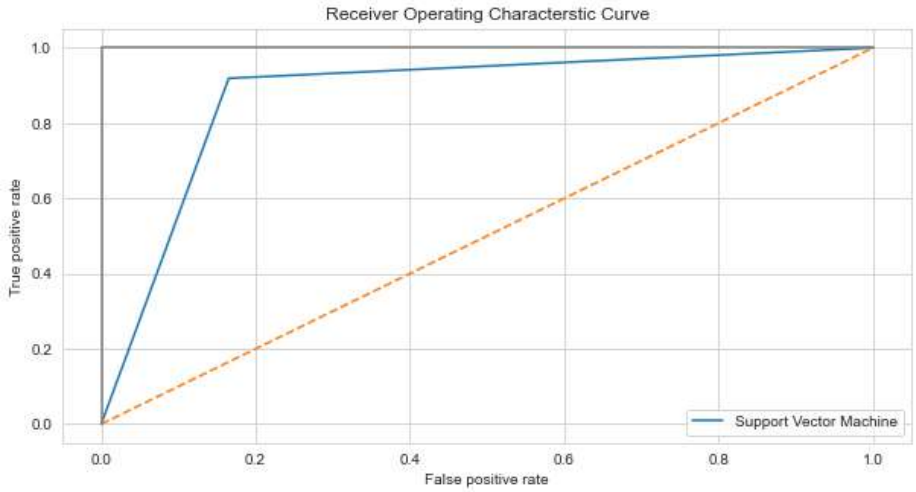

FIGURE9.ROC curve obtained by validating Support Vector Machine model

RESULTS AND COMPARISON

Below **TABLE 1** contains accuracy of every deep learning model used in the research work is mentioned, From all the four deep learning models ANN has high accuracy compared to remaining deep learning models.

Table 1: Accuracy of the Deep Learning models and Machine Learning Models

Models	Accuracy
Artificial Neural Networks	93.774319

Naive Bayes	83.657588
LogisticRegression	85.214008
Support Vector Machine	85.603113

CONCLUSION

This study would be beneficial in the near future in identifying new people who potentially have heart problems. By doing so, it will be easier to take preventative measures and keep the patient from developing a heart failure risk. So, if a physician's predict that the situation is indicative of cardiac trouble. Will carefully review the condition's diagnostic data. Consider, for instance, that the patient has diabetes, which will cardiac ailment or illness, the patient will eventually develop maintain control of diabetes, which can reduce the risk of heart disease in turn.

We have validated the dataset with various Deep Learning models and obtained accuracy to predict a heart disease accurately. This paper validated the dataset using Artificial Neural network, Navie Bayes, Logistic Regression and Support Vector Machine models. Among those models, the Artificial Neural network model predicts at high accuracy, which is more efficient than other models of Deep Learning. We can see the roc curve for each and every model we validated in case of ANNthe roc curve is normal, but in the other model the curvature is not typical. For future work we can include more Deep algorithms may improve accuracy.

REFERNCES

1. R. Williams, T. Shongwe, A. N. Hasan and V. Rameshar, "Heart Disease Prediction using Machine Learning Techniques," 2021 International Conference on Data Analytics for Business and Industry (ICDABI), 2021, pp. 118-123, doi: 10.1109/ICDABI53623.2021.9655783.
2. S. R. Tithi, A. Aktar, F. Aleem and A. Chakrabarty, "ECG data analysis and heart disease prediction using machine learning algorithms," 2019 IEEE Region 10 Symposium (TENSYMP), 2019, pp. 819-824, doi: 10.1109/TENSYMP46218.2019.8971374.
3. Sutedja, "Descriptive and Predictive Analysis on Heart Disease with Machine Learning and Deep Learning," 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), 2021, pp. 1-6, doi: 10.1109/ICORIS52787.2021.9649585.
4. J. Thomas and R. T. Princy. "Human heart disease prediction system using data mining techniques", 2016.
5. M. Gandhi and S. N. Singh. "Predictions in heart disease using techniques of data mining", 2015.
6. Sutedja, "Descriptive and Predictive Analysis on Heart Disease with Machine Learning and Deep Learning," 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), 2021, pp. 1-6, doi: 10.1109/ICORIS52787.2021.9649585.
7. F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," Int. J. Adv. Computer. Sci. Appl., vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.
8. S. Bharti and S. N. Singh. "India analytical study of heart disease prediction comparing with different algorithms", May 2015.

9. eophytou, N., (2012). ECG event detection & recognition using time frequency analysis (Doctoral dissertation).
10. Guvenir, H. A., Acar, B., Demiroz, G., and Cekin, A. (1997). Supervised machine learning algorithm for arrhythmia analysis. Computers in cardiology, pages 433–436
11. V. Krishnaiah, G. Narsimha, and N. S. Chandra. "Heart disease prediction system using data mining techniques and intelligent fuzzy approach:" A review, February 2016.
12. M. Alloghani et al., "Implementation of machine learning algorithms to create diabetic patient re-admission profiles," BMC Med. Inform. Decis. Mak., vol. 19, no. Suppl 9, pp. 1–16, 2019, doi: 10.1186/s12911-019-0990-x
13. D. Baroud, A. N. Hasan, and T. Shongwe. "A study towards implementing various artificial neural networks for signal classification and noise detection in PFDM/PLC Channels," 12th IEEE international symposium on communication systems, networks and digital signal processing (CSNDSP), 2020.
14. M. Fatima and M. Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," J. Intell. Learn. Syst. Appl., vol. 09, no. 01, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.
15. J. Latif, C. Xiao, S. Tu, S. U. Rehman, A. Imran, and A. Bilal, "Implementation and Use of Disease Diagnosis Systems for Electronic Medical Records Based on Machine Learning: A Complete Review," IEEE Access, vol. 8, pp. 150489–150513, 2020, doi: 10.1109/ACCESS.2020.3016782.