PREDICTION OF PRE-CARDIAC DISEASE USING ML & DL TECHNIQUES IN T-KINTER

[1]Dr.G.Vishnupriya (Assistant Professor), [2]Pradeep T, [3]Mehanethra M, [4]Ajay Mitra VJ

[1]Department of Computer Science and Engineering Easwari Engineering College Ramapuram, Chennai, India. vishnupriya.g@eec.srmrmp.edu.in

[2]Department of Computer Science and Engineering Easwari Engineering College Ramapuram, Chennai, India. Pradeeptamil6@gmail.com

[3]Department of Computer Science and Engineering Easwari Engineering College Ramapuram, Chennai, India. mehamaran9530@gmail.com

[4]Department of Computer Science and Engineering Easwari Engineering College Ramapuram, Chennai, India. vjajaymitra@gmail.com

Abstract— One of the biggest issues facing the globe today is heart disease. The prediction of cardiovascular illness presents a significant problem for clinical data analysis. Hybrid machine learning (ML) has demonstrated its ability to effectively support decision-making and prediction from the vast amounts of data generated by the healthcare sector and hospitals. Additionally, we have observed the employment of ML approaches in recent advancements across several IoT domains (IoT). Only a few research have used machine learning to predict cardiac disease. The narrative approach we suggest in this study tries to identify relevant features by utilising Deep learning techniques, which is the accuracy of cardiovascular disease prediction. Classification Techniques used in the prediction model include Logistic Regression, Naive Bayes, Support Vector Machine, K Nearest Neighbors, Decision Tree, Random Forest, and XGBoost Artificial Neural Network with 1 Hidden layer. By using a hybrid random forest and DL technique Neural network to create a prediction model for heart disease, we are able to improve performance with a 92% accuracy rate. Each algorithm's accuracy is calculated along with the model's accuracy. The next step is to select the one with good accuracy. The suggested methodology combines all the data into a classification algorithm and use the Tkinter programme to compare current healthcare data with information from that particular reference distribution.

Keywords— Deep leaning; Machine Learning; Classification; Data sets; Features;Neural network; Random Forest; SVM;

I. INTRODUCTION

Today's hereditary and lifestyle factors are driving up the expense of healthcare. A tonne of data is generated over time. As a result, survey results or health-related data are meaningless. Yet, as data analytics become more prevalent. The data are being used by hospitals and Organizations to produce information that is beneficial. The worst hazard to the modern world is cardiovascular disease. This condition affects a person in a way that makes difficult to cure the patients. As a result, the hardest labour in the medical industry is diagnosing patients at the appropriate moment. Poor diagnosis and misdiagnosis by the hospital contribute to its poor reputation. India disputes the efficacy of the disease's therapy most patients can't get there because it's pretty difficult. Each person's There are variations in pulse rate, blood pressure, and cholesterol. Blood pressure, cholesterol, and pulse rate,

however, should be within normal bounds of 120/80, 72, and correspondingly. World Health Organization data indicate that cardiovascular illnesses are believed to be the cause.of more than 12 million fatalities each year. It is a terrible illness in India that causes more catastrophes.

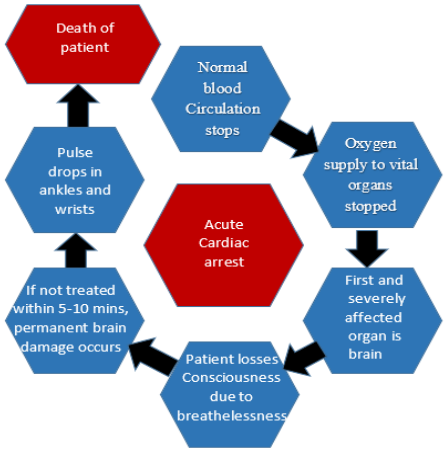


Fig. 1. Sudden cardiac attack process

Fig. 2.

Acute or Sudden Cardiac Arrest occurs suddenly and deprives the life in less than 10 minutes. Hence the survival rate of this disease is less than 7% where only 2 out of 25 people survive according to World Heart Foundation Survey. This disorder is initiated with the electrical imbalance of heart rhythms or malfunction in pumping blood even though there is no blockage in the heart or any problem with the brain. This can also happen due to the irregular heart beat or genetic disorder. It happens even in children.

Acute Cardiac Arrest can be identified as an unanticipated loss in the functioning of the heart, loss of breath and ultimately leading to loss of consciousness in no time. Such problems arise due to various reasons. The chief one being the previous history of heart attack and arrhythmias problems with irregular electro cardio gram (ECG) result. The Acute Cardiac Arrest is expected to have the following stages when it affects the human being.

Acute or Sudden Cardiac Arrest surfaces without proper symptoms. Even for skilled medical personnel, it is exceedingly difficult to foresee issues beforehand.. Hence the methods to identify the heart problems are too identified to enhance predictions of this dreadful disease at the earliest possible.

II. RELATED WORKS

The work proposed in this paper focuses primarily on various data mining techniques used in the prediction of heart disease. The human heart is the most important organ in the body. It regulates the flow of blood throughout our bodies. Any abnormality in the heart might aggravate discomfort in other parts of the body. Heart problems refers to any condition that impairs the heart's regular operation.

In [1], they suggested using machine learning to predict heart disease. They employed a multilayer perceptron for their prediction method, and the MLP provides a trust worthy based on the user's input, the output.This model is predictive and the accuracy is effective.

In [2], they put forth a framework for a hybrid intelligent system that uses methods for machine learning to forecast heart disease.Additionally, they used an ANN (MLP) with 16 hidden layers, and their feature selection is excellent. The model accuracy percentage was 94%.

In [3], they suggested utilizing Machine Learning heart disease prediction algorithms and its model Decision Tree performed better than the other classifier Nave Bayes. The increase in quality is what causes the accuracy to improve. In [4], they proposed utilizing neural networks to predict heart disease. Additionally, a neural network with

a 95 percent accuracy was developed using the back propagation approach and more precise in terms of heart disease.

In [5], they suggested methods for machine learning to forecast heart disease. They employed the Decision Tree model, and their accuracy levels were 91 percent for the Decision Tree model and 87 percent for the Naive Bayes classifier. The best algorithm for handling medical datasets is the decision tree algorithm. In [6], they proposed a strategy to predict cardiac diseases using data mining techniques. We use data mining, decision trees, neural networks and Naive Bayes. This study's objective was to develop a prediction model for identifying heart illness utiliing data extraction techniques from the Transthoracic Echocardiography Report dataset, with the goal of improving the accuracy of echocardiography-based cardiac disease diagnosis. On the transthoracic echocardiography dataset, models were generated utilizing the machine learning tool Weka 3.6.4 and three different supervised ML approaches, namely J48 Classifier, Naive Bayes, and Multilayer Perception.

In [7], they proposed to use Naive Bayes for the heart disease prediction system's decision support. Data mining was utilized., Naive Bayes, and decision support. This model could react to complex queries, each with a distinct benefit for the ease of model interpretation, availability of complete information and accuracy. They suggested Using classification algorithms, heart disease may be predicted. In [8], In this study, heart attacks are predicted using data mining techniques like J48, Naive Bayes, REPTREE, CART, and Bayes Net. The research's findings indicate a 95% forecast accuracy rate Dataset trends may now be predicted by the health industry.

In [9], This study's key contribution is to aid non-specialized clinicians in making informed decisions about the likelihood of developing heart disease. In that order, the produced rules are categorized as There are five types of rules: original, pruned, duplicate-free, sorted, and polished. Results demonstrate the framework. has exceptional potential for more precisely predicting the cardiac sickness risk level when the framework shows how to evaluate employing data mining tools to reduce the risk of heart disease. Various technologies provide varying degrees of precision depending on the attributes taken into account. Heart disease risk was calculated based on KNN Moreover, the ID3 performance is assessed in terms of of arrangement precision. They proposed a data mining based human heart disease prediction system. In [10], This paper's primary goal is to provide light on algorithm, and the accuracy level was also provided for various numbers of attributes.

TABLE I. COMPARISON OF LITERATURE SURVEY

REF NO	METHODOLOGY	PERFORMANCE METRICS	MERITS
1	Machine learning for the prediction of heart disease	This model predictive accuracy is effective.	MLP produces consistent results based onuser input.
2	A framework for a hybrid intelligence system to predict heart disease using machine learning algorithms.	The best sensitivity was 100 percentage of classifier ANN (MLP) with 16 hidden layers.	Feature selection is excellent.
3	Machine learning	Decision tree out	The

III. PROPOSED PLAN

The collected data is pre-processed and using a random forest algorithm and CNN is saved in a knowledge base, which helps in the structural construction of the model. This project is useful for many people, just enter the details given in the

	algorithms are being used to predict heart disease.	performs.	improvement in accuracy arises from increased attributes
4	Neural network prediction of heart disease	NN was trained with a backpropagation algorithm accuracy of 95 percent.	More accurate diagnosis for heart disease.
5.	Using machine learning algorithms to predict heart disease.	The decision tree model Accuracy level is 91 percent. Naïve Bayes classifier accuracy is 87 percent.	The decision tree algorithm is best for handling medical datasets.
6	Cardiovascular disease prediction techniques through data mining	The J48 classifier was used on selected attributes and achieved a classification accuracy of 95.56 percent.	The three data mining objectives were met with the chosen model, which was built using the J48 classifier.
7	Naive Bayes decision support in a cardiac disease prediction system	Decision Tree is the most accurate model for predicting patients with heart disease.	The decision tree works better.
10	Heart disease was predicted using data mining techniques.	Naïve Bayes. KNN. Decision Tree.	Using ID3 and KNN algorithm, heart disease was detected and accuracy level was also promising.
9	Efficiency Heart Disease Prediction System.	The system was trained and tested using the 10fold approach, with an accuracy of 86.3 & 87.3 respectively.	It provides early patient safety and results in smart performance even without retraining.

text box and press submit. The following project has been built using machine learning concept using random forest algorithm and neural network algorithm. After the classification process we implemented the concept of t-kinter application to display the output in web application.

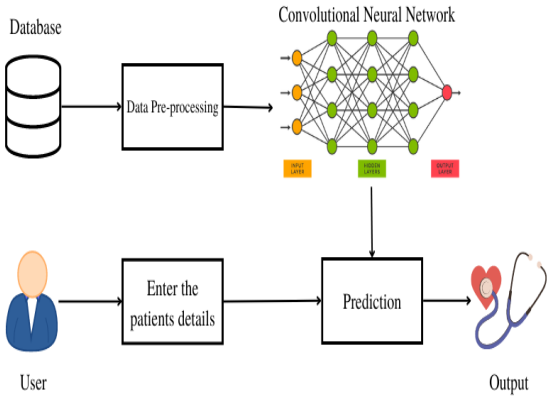


Fig 2: Flowchart of proposed Model

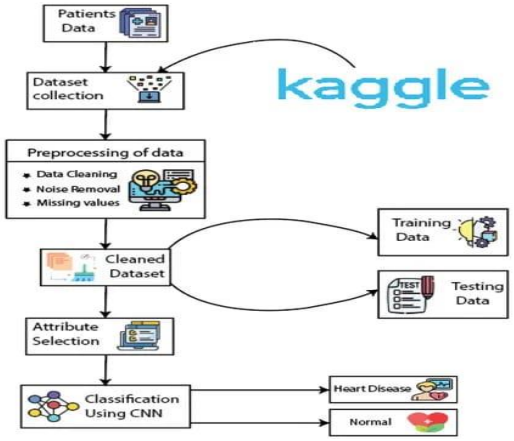


Fig 3: Architecture of proposed Model

III. WORK FLOW

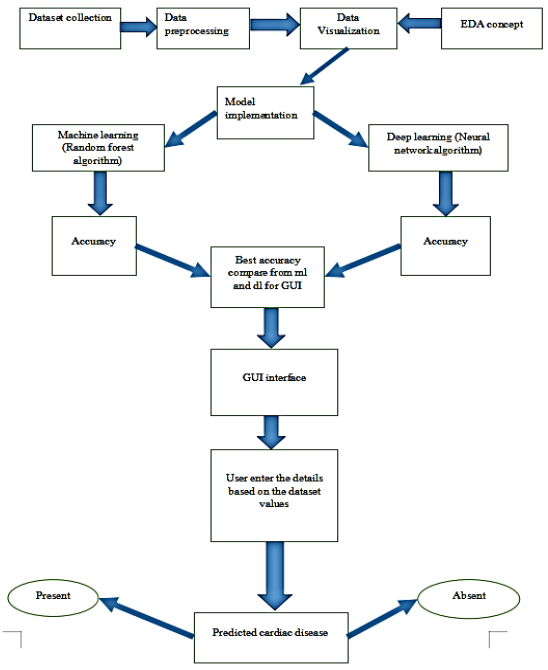


Fig 4: Workflow of the project

V. DATASET & PRE-PROCESSING

Generally the datasets are collected from the public sources such as the Kaggle, Phish tank etc. The raw dataset is pre-processed in order to remove the redundant and irregular data. Exploratory Data Analysis (EDA) technique was applied on the dataset after the pre-processing. The data visualization method was employed to analyse, explore and summarize the dataset Data preparation is a component of analysis and data mining. It handles unstructured data and converts it into a format which computers can quickly utilize. In fact, raw data may include inaccuracies, missing information or a lack of regularity In contrast, all machines & algorithms are programmed to process structured, identical, and organized data. This data pre-processing module is therefore vital to the process. Here, the raw data in the cardiovascular disease data is sanitized and the metadata is inserted by eliminating the integer-conversions. Thus, the data facilitates training. Consider all the info.

During this pre-processing, we initially load the metadata into it and then connect it to the data, replacing the transformed data with the metadata. The data will then be moved forward, unnecessary data will be removed from the list, and the data will be separated into trained and tested data. The pre-processed data will be helped in splitting into train and test groups depending on the weights supplied in the code by importing train test split from Scikit-Learn..The divide between test and train is 0.2% and 0.8%, or 20% and 80%, respectively.

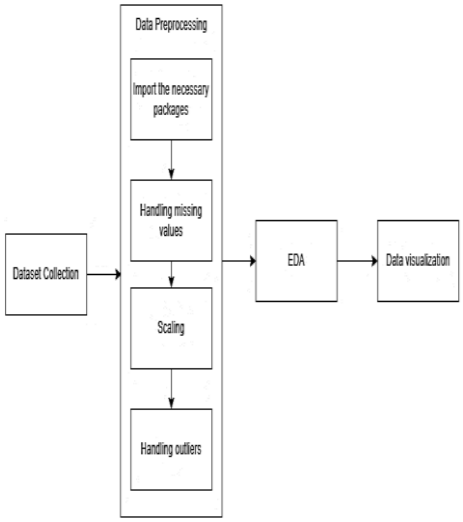


Fig 5: Flowchart for data preprocessing

V. IMPLEMENTATION

A.) DATASET:

The first and most important stage in utilising machine learning algorithms to achieve trustworthy results is data collection. The dataset utilised was produced. from the "UCI machine learning repository," a renowned data source. Heart illness in Cleveland, Switzerland, VA Long Beach, Hungary, and Statlog datasets are among the five separate datasets. For this study, we pooled them together to produce more accurate results. From their database, 14 unique features and more than 1190 examples are compiled in a text file. The 13 properties of these combined datasets are used as diagnostic inputs, with the attribute "num" chosen as output.All or the majority of the data had six characteristics that are recognised as pertinent in the medical literature: age (years), and sex (sex), findings for resting electrocardiography, fasting blood sugar, kind of chest discomfort, and resting blood pressure (restecg). The various qualities and their range of values are described in Table 1.

The 'num' attribute can have a value of 0, 1, 2, 3, or 4. The range of values between 1 and 4corresponds to the different phases of chronic heart disease, with the projected value of '0' signifying that a patient is free of heart disease.

a summary of the total patient count in the merged dataset for every num attribute value.

.

We change all values in the range of 1 to 4 to a 1 because the goal of this study is to assess whether or not a patient has heart illness. As a result, the property presently includes a variety of (0, 1).

S. No	Feature	Characteristic representation
1	Age	Age
2	Sex	Sex
3	Chest pain	Cp
4	Rest blood pressure	trestbps
5	Serum cholesterol	Chol
6	Fasting blood sugar	Fbs
7	Rest electrocardiograph	Restecg
8	MaxHeart rate	Thalach
9	Exercise – induced angina	exang
10	ST depression	oldpeak
11	Slope	slope
12	No. of vessels	Ca
13	Thalassemia	thal
14	Num (class attribute)	Class

TABLE II. DATASET ATTRIBUTES

B.) FEATURE SELECTION TECHNIQUES:

In order to extract the optimal qualities for categorization, feature The machine learning process depends on selection approaches. The execution time is also shortened by this. The Relief Feature Selection approach, as well as the Least Absolute Shrinkage and Selection Operator, have been chosen.

C.) RANDOM FOREST :

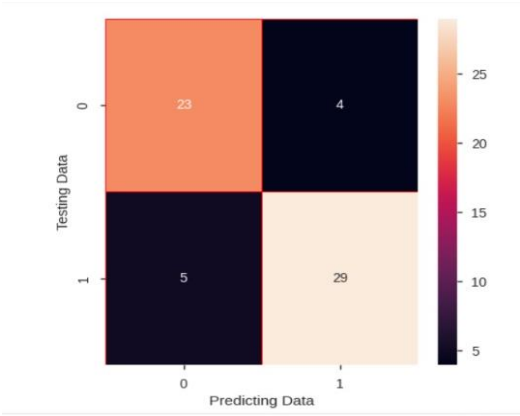
It is the most widely used machine learning approach for the purposes of classification and regression. It makes use of the ensemble learning approach., where predictions are based on the averaged output of a number of different individual models. Data that has been divided into training and testing is used with the Random Forest Classifier. For categorization and prediction purposes, the model is trained. The accuracy rating of the model is evaluated.

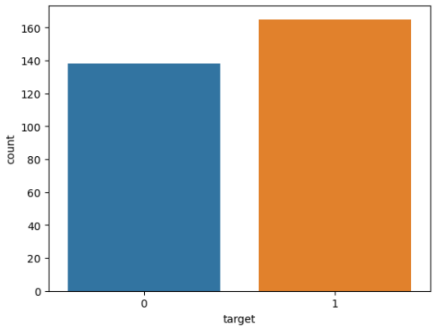
D.) CONVOLUTIONAL NEURAL NETWORK:

To build the Convolutional neural network, the input is split into training and testing data, and further layers are added (CNN).The model is developed and evaluated for accuracy. For processing organized arrays of data, such as images, convolutional neural networks, CNNs, or deep learning neural networks, are a form of deep learning neural network. Convolutional neural networks are widely utilized in computer vision and are the cutting-edge for many visual applications, including picture categorization. Also, they have had success with text categorization using nlp.

E.) DATA VISUALIZATION:

Data visualization makes it simple to understand the data through illustrative graphs or maps. To find the link between the attributes, heat maps are employed. Also, it demonstrates which characteristics are crucial for producing the most precise predictions. The Data visualisation is the representation of data using standard visuals such as infographics, charts, and even animations. These informative visual representations make complicated data interactions and data-driven insights easier to understand.





VI. EXPERIMENTAL RESULTS

THIS SECTION ELABORATES ON THE EXPERIMENTAL SETUP, UNDERLYING
platform, along with experimental parameters with hardware
and software platforms. We tried to correlate the features
found in the dataset to predict the heart health of the patients.
Interesting facts are found during the experimentation that is
discussed in the subsections
This section elaborates on the experimental setup, underlying
platform, along with experimental parameters with hardware
and software platforms. We tried to correlate the features
found in the dataset to predict the heart health of the patients.
Interesting facts are found during the experimentation that is
discussed in the subsections

This section elaborates on the experimental setup, underlying
platform, along with experimental parameters with hardware
and software platforms. We tried to correlate the features
found in the dataset to predict the heart health of the patients.
Interesting facts are found during the experimentation that is
discussed in the subsections

The experimental setup, supporting platform, and experimental parameters with hardware and software platforms are all further described in this part. We attempted to link the dataset's properties in order to forecast the patients' heart health.
During the experimentation that is covered in the subsections, interesting discoveries were discovered.

The method was put into place to increase the accuracy and improve the dynamics of existing ML algorithms. The experimental outcomes of the algorithms, whose benefits and drawbacks are discussed above, are disabled in various datasets.

Using some performance measures, one may assess the classifier's effectiveness. In machine learning, a classifier's performance can be assessed using a variety of metrics. In machine learning, accuracy is a popular

metric to assess the how effective a classification model is. It is derived by dividing the total number of predictions by the number of correct predictions. The accuracy score() function from the sklearn.metrics module is used to compute the accuracy of the Logistic Regression model.and the results are printed using the print command () function at the end of the training and testing process. If we want to calculate the accuracy of the model again after making predictions on the testing data, you can use the accuracy score() function again and print out the result. Accuracy is a commonly used metric in machine learning to measure the performance of a classification model

VII. PERFORMANCE METRICS

A few performance measures may be used to assess the classifier's performance. There are several criteria available in machine learning and deep learning to assess the

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

success of the classifier. These are some of the explanations behind these requirements.

The accuracy gauge used to assess a classifier's performance is called precision. There are fewer false positives if the accuracy

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

is high. More numbers of false positives are seen in a model with lower precision means.

Recall It serves as a gauge for the classifier's thoroughness. Less false negatives occur when recollection is higher, and more false negatives occur when recall is lower. Increasing recollection frequently causes a drop in accuracy.

$$F1\ \text{Score} = \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

F-Score is the result of combining accuracy and recall, and it may be computed using the above formula.

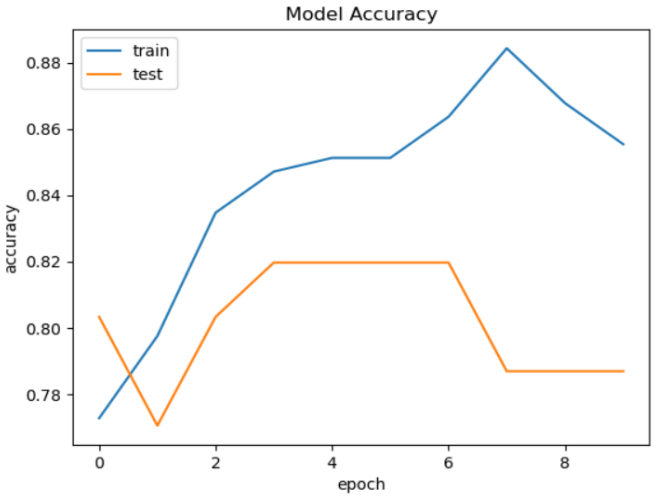


Fig 6: Graphical representation of training and validation accuracy

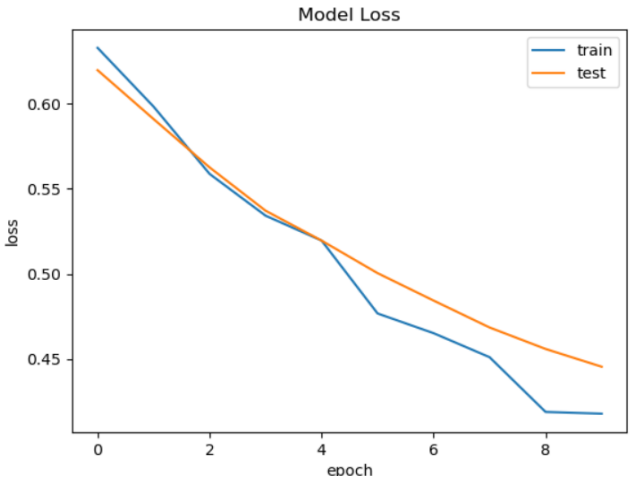


Fig 7:Graphical representation of training and validation loss

VIII. CONCLUSION AND FUTURE WORK

In order to this study provides the correlative use and analysis of machine learning by providing an instant way for the user to employ the ML and DL algorithms in Python programme for anticipating cardiovascular illnesses. algorithm and deep learning algorithm, such as the random forest

algorithm and convolutional neural network, respectively. In this study, a trustworthy multi-process strategy for developing a system to forecast the probability of cardiac arrest using both random forest and CNN is proposed. Worldwide, cardiac arrest is now the leading cause of mortality. Early symptom detection is the most efficient strategy to lower these fatalities. Such prophylactic screening is often avoided by people. As a source of records with a thorough patient history, this system can additionally hospitals assist doctors in focusing on a specific course of treatment for each patient. In order to improve the system's operations in the future, we plan to advance this deep learning study. Cardiac disease prediction using the deductive learning approach of the system.

REFERENCES

1. Jaymin Patel, Prof.Tejal Upadhyay, Dr.Samir Patel “Heart disease prediction using Machine learning and Data Mining Technique" March 2020.
2. G.Parthiban, S.K.Srivasta "Applying Machine learning methods in Diagnosing Heart disease for Diabetic Patients" August 2021
3. Matjaz Kuka, Igor Kononenko, Cyril Groselj, Katrina Kalif, Jure Fettich" Analysing and improving the diagnosis of ischaemic heart disease with machine learning" Elsevier Artificial intelligence in Medicine, Volume23, May 2021
4. Igor Kononenko “Machine learning for medical diagnosis: history, state of art& perspective” Elsevier - Artificial intelligence in Medicine, Volume23, Aug 2001
5. Gregory F. Cooper*,” An evaluation of machine learning methods for predicting pneumonia mortality”- Elsevier Feb 2020
6. Gregory F.Cooper, Constantin F.Aliferis, Richard Ambrosino” An evaluation of Machine learning methods for predicting pneumonia mortality”-Elsevier, 2020.
7. Sanjay Kumar Sen” Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms”- June 2020.
8. Abhishek Taneja” Heart Disease Prediction System Using Data Mining Techniques ”Vol.6, No(4) December 2021.
9. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee” Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review”- Advances in Computational Sciences and Technology ISSN 0973-6107, Volume10, Number7(2021).
10. Thenmozhi.K and Deepika.P, Heart Disease Prediction using classification with different decision tree techniques. Oct 2019.
11. Gavhane, A., Kokkula, G., Pandya, I. and Devadkar, K., 2018,March.
12. Prediction of heart disease using machine learning. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1275- 1278). IEEE
13. Karayilan, Tulay, and Ozkan Kilic. "Prediction of heart disease using neural network." 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, 2017.
14. Krishnan, Santhana, and S. Geetha. "Prediction of Heart Disease Using Machine Learning Algorithms." 2019 1st international conference on innovations in information and communication technology (ICIICT). IEEE, 2019.

15. Nikhar, Sonam, and A. M. Karandikar. "Prediction of heart disease using machine learning algorithms." International Journal of Advanced Engineering, Management and Science 2.6 (2016): 239484.