Predicting heart disease using machine learning and IoT techniques

N Divya[1*], Md Riyazuddin[2], Abdul Ahad[3], Sridhar Reddy Vulapula[4], A Manjula[5], Mohd Sirajuddin[6]

[1]Department of Data Science, Sreenidhi Institute of Science and Technology, Hyderabad,Telangana, India

[2]Department of Information Technology, Anurag University, Hyderabad, Telangana, India

[3]Department of Artificial Intelligence, Anurag University, Hyderabad, India

[4]Department of Information Technology, Vignana Bharathi Institute of Technology, Hyderabad, Telangana, India

[5]Department of CSE, Jyothishmathi Institute of Technology and Science, Karimnagar, Telangana, India

[6]Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India

ABSTRACT

Nowadays, there is an increase in heart disease in all age groups in society. It is therefore necessary to set up a machine learning system in order to be able to detect and prevent indications of heart disease at an early stage. There must also be a mechanism in place that is handy and at the same time capable of being trusted. Thus, we propose to create an application that can predict the potential for heart disease given basic symptoms such as age, gender, ECG, heart rate, chest pain, cholesterol, blood pressure, blood sugar. The method will use various models trained using machine learning algorithms such as the support vector machine, the Naïve Bayes classifier and the decision tree. The accuracy of the method will be measured and distinguished in order to select the best model for estimating heart disease. Latest advances in online healthcare can be used for IoT and sensing applications. In the medical field, IoT devices and cloud computing techniques are used to manage the massive amount of data.

Keywords: Machine learning; Heart disease; ECG; Chest pain

Address for correspondence:

N Divya
Department of Data Science,
Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India;
E-mail: naademdivya@sreenidhi.edu.in

INTRODUCTION

Heart disease can be effectively controlled with a combination of habits changing medicine in some cases by surgery. With accurate treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The expected outcome can be used to stop and thus reduce the burden of surgical treatment. The overall objective of my work will be able to predict the existence of heart disease precisely with fewer tests and attributes. The attributes considered form the initial basis for the tests and provide accurate results. Various input attributes may be used, but our goal is to predict with few attributes to identify the risk of heart disease [1].

The greatest loss of life both in India and abroad is due to heart disease. Nowadays, doctors are supporting many scientific technologies and approaches for both identification and diagnosis not only of a known disease, but also of many deadly diseases.

Successful treatment is always associated with a correct and accurate diagnosis. Doctors may occasionally neglect to make accurate decisions when analyzing a patient's heart disease, therefore heart disease prediction systems that use machine learning algorithms help to achieve accurate results in similar cases.

People have a regular and stressful lifestyle that produces stress and anxiety in our everyday life. Besides that, the proportion of people obese and smoking is still quite high. This leads to cardiovascular diseases, tumors, etc. The prediction of these diseases is high challenge. The pulse rate and blood pressure values for every person are different. The heart rate must be between 60 and 100 beats per minute and blood pressure between 120/80 and 140/90, which has been confirmed medically. Heart disease is one of the world's most significant causes of death. Other conditions such as sex, diabetes, and BMI, however, also lead to the disease. We tried to predict and evaluate heart disease by considering age, sex, heart rate, blood pressure, diabetes, etc. [2].

Many facets of cardiovascular disease are involved, and the prediction for this disease is difficult. Four of the major symptoms of a cardiac attack are:

- Breathe shortening, chest power.

- Nausea, cardiovascular disease, indigestion, tumor.

- Tiredness and sweat.

- Upper back pain stretching through the shoulder.

The category of cardiovascular condition below is cardiovascular. Thus, all coronary diseases are linked to cardiovascular diseases. Various heart disease types:

- Cardiovascular diseases.

- Angina pectoris; angina pectoris.

- Angina pectoris; angina pectoris.

- Heart failure with congestion.

- Carthritis.

- Heart disease congenital.

The pinching of the coronary arteries is heart disease. Blood and oxygen pass *via* coronary arteries. It makes a lot of people sick or kills them. It is one of the major cardiac diseases. High blood glucose can damage the heart and blood vessels and the nervous system. People with diabetes are at higher risk of developing heart disease in the future. Some factors contribute to diabetes heart disease. They smoke which can make it difficult for the heart to pump blood, which can strain the heart and damage the blood vessels. Being susceptible to heart disease, one's family history plays a part [3,4].

Age, gender, unhealthy diet and stress are another risk factor. When a person grows older, his or her risk of heart disease increases. The risk of heart disease is high for men. However, menopause is also the same for women. Stressful existence may also damage the arteries and increase the risk of heart condition. Thus, we attempt to estimate the chance of heart attack in the following conditions. A lot has been achieved on the cardiovascular system by several writers using different techniques of automated learning and algorithms. These techniques may be based on machine learning, profound intelligence, data mining, etc. All these papers are planned to achieve precision and enhance the performance of the method in order to predict the possibility of heart disease. Age, sex, stress and an unhealthy diet are all risk factors.

When a person is elderly, the risk of a heart attack increases. The risk of heart disease is higher for men. Women have the same chance of menopause, however. Stressful life can also damage the arteries and increase the risk of cardiac disease. Therefore, we try to predict the risk of heart disease in these factors described above. Many writers used different methods and algorithms for the heart prediction method. Much work was done. These techniques can be based on deep learning, computer education, data collection, etc. All these documents are designed to increase the precision and reliability of the system in order to predict heart disease risk.

Heart disease has caused significant concerns among researchers; the proper diagnosis of the existence of cardiac conditions in an individual is one of the key challenges in cardiovascular disease. Early methods did not demonstrate that even medical professionals are not as effective as possible to predict cardiac disease. There are various medical devices available for predictive cardiac disease on the market, with two major problems: First, very costly, and second, that the risk of heart disease is not adequately measured by people. Medical personnel could only predict 67 percent reliably of heart disease, according to the latest WHO study and a number of investigations are therefore being carried out in the field of estimation of individual heart diseases.

Heart disease is one of the greatest obstacles of medical research, as an accurate prediction of this disease entails many restrictions and technicalities. Machine learning can be a better option for estimating accurate not only heart disease, but also other diseases, since this variety of equipment uses a vector and various types of data for heart disease assessment, algorithms like Naive Bayes, Decision Tree and SVM are used for preaching heart disease possibilities. Each algorithm has its own unique specialization, such as SVM and Naive Bayes, which use the likelihood of heart disease prediction, while the decision tree provides a categorized heart disease study. Both these approaches use old patient records to find new patients'

predictions. This cardiac disease approach helps doctors predict early-stage cardiovascular disease which has saved millions of lives.

Decision is often made by the doctor's understanding and awareness of the rich data concealed in the database instead of by knowledge of the sector. This practice leads to unintended prejudices, mistakes and excessive medical expenses impacting patients' quality of care. There are different ways of describing a medical mistake. A misdiagnosis of a severe disease may have particularly risky and devastating consequences when a physician is a defect or hospital worker. 42 percent of patients have been diagnosed with a medical error or have been overlooked by the reliving national patient care fund. Often the patient safety is due to inaction in the back seat due to other issues, such as the cost of medical tests, medications and operations. Detection at an early stage is not feasible because it takes time for several data to be practically used.

The internet of things is essential to humans. Things such as education are within its sphere of influence. Social media, business networking, healthcare technology etc. The Health care industry has been adapting new technology through innovations. promoting better healthcare services [1]. The Internet of remote monitoring is made practical at this time by technological innovation. In this way, it unleashes the potential to monitor continuous governance. Health and medication can assist physicians in providing advice, testing and treatment. We should ensure prompt treatment as an entire population. If people are suffering from coronary illness, they should be properly diagnosed. Start emergency development to save people's life.

The main objective of the research is for three engineering techniques, such as SVM, Decision Trees and Naive Bayes, to build the prototype heart disease prediction method. The system proposed has data to assess whether patients have heart disease or not, based on the system characteristics. This proposed method would attempt to use this data in order to construct a model to determine whether a patient has this disease (read data and data exploration). In this proposed method, use the classification algorithms. To obtain reliable performance, implement the SVM, Decision Tree and the Naive Bayes algorithm. From the data we supply, a range of structured data dependent on the heart characteristics of the patient should be classified.

From the availability of the data we must create a model that uses the SVM, Decision Tree and Naive Bayes algorithm to predict disease for the patient. We have to import data sets first of all. Data should include age, gender, ECG, heart rate, chest pain, cholesterol, blood pressure, blood sugar, etc., and the data in the read datasets. The data should be tested to verify the facts. Creates and builds a temporary variable. The precision is improved by the use of the SVM, Decision Tree and the Naive Bayes algorithm [5-7].

LITTERATURE REVIEW

Review of the research paper

Heart disease survey: Causes, prevention and empirical research: Mohammed Abdul Khaleel has written a paper to the local finds of frequent diseases survey on medical data mining. The emphasis of this paper is on dissecting information mining procedures necessary for the mining of medicinal information, particularly for local visits to diseases like heart disease, lung malignancy, pneumonia, etc. [8-10]. Information mining can be used to extract the information required to find inactive examples, to evaluate and diagnose cardiovascular disease by Vembandasamy etc. The Naive Bayes algorithm was used here. In the Naive Bayes algorithm, they

used theorem Bayes. Naive Bayes however possesses very significant power independently to make conclusions. The used dataset is extracted from the leading diabetic research institute in Tamilnadu, Chennai. The dataset includes more than 500 patients. Weka is the tool used, and 70% of percentage split is graded. Naive Bayes gives 86.419% of the precision [1].

Survey on remote health watch outcome success prediction victimization baseline and initial month intervention data Costas Sideris, Nabil Alshurafa, HaickKalantarian and Mohammed Pourhomayoun provided a paper called Remote Health Watch Outcome Success Prediction Victimization Initial Month and Base Intervention Information [11-15]. RHS square systems measure the effectiveness of saving prices and the reduction of health problems. During this paper, A portrays the hierarchical RHM framework, Wanda-CVD, which is mostly cell phone based and which is designed to allow remote instruction and social facilitation for members. CVD countervailing action measure square measure perceived as a key focus of social welfare associations around the world [2].

Monika Gandhi, et al., used Naive Bayes, named tree algorithms, and analysed the medical dataset. There are several choices you should be concerned about. So, there is a need to limit choices. A choice of features will be selected. The argument is minimized as this is done. They also introduced call trees and neural networks [16-19].

J Thomas, R Mother Teresa Princy established the use of K nearest neighbor algorithmic rule, predictive Bayesian neural network, the Naive Thomas Bayes classifier and the call tree for cardiopathy risk detection prediction. By mining data on patient's health records, they have started to search for patients at risk of such diseases [20].

Sanaa Bharti, Shailendra Narayan Singh generated the use of Particle Swarm optimization, artificial neural network, and genetic algorithm rules of the game for prediction. We assume that associative classification, a new technique, can be a highly economical technique that combines association rule mining and classification to a model for prediction, which is accomplished at a sensible accuracy [21].

In Purushottam, et al., Machine-driven system will has many improvements in treatment and also this can scale back price. The system's goal was to operate on concepts that would predict a patient's health. The purpose of the principles was to prioritise the user demands. The results of the method show that it is reliable in providing predicted the probability of organ related injury.

Sellappan Palaniyappan, Rafiah Awang has developed the use of a calling tree called Naive Thomas Bayes, call tree and artificial networks for the development of intelligent heart pathology prediction systems. It jointly helps to minimise the price of therapy by delivering successful therapies. Hidden trends and relationships have usually been revealed. This case was remedied by sophisticated data processing techniques.

Call tree use, help vector-machine, deep Himanshu SharmaK, neighbourhood algorithms most similar to learning. Because the datasets contain noise, they attempted to reduce the noise through the cleaning and pre-processing of the dataset and collectively tried to reduce the data set room. Responsive accuracy of neural networks has been demonstrated.

The condition and entirely different symptoms of attack were thoroughly described by Animesh Hazra, et al., Similar classification varieties and cluster algorithms and methods have been used [22].

AN victimisation data processing provided by V. Krishnaiah, G-Narsimha, N. Subhash Chandra. The study has shown that similar/completely (totally different/totally different/entirely different) methods are victims and that different characteristics are taken for heart disease prediction.

Ramandeep Kaur, Er. Prabhsharn Kaur, has been shown to contain unnecessary duplicate data on good sick information. This should be done in advance. They also state that the data set for achieving higher results should be based on the feature collection.

Data treatment was used by J. Vijayashree and N. Ch. Sriman Narayana Iyengar. Daily is developed a large amount of information. It cannot be manually understood as such. Data treatment predicts diseases from such databases effectively are accustomed. Completely different square measurements examined for cardiopathy info during this article. Finally, this paper analyses and contrasts the work on cardiopathy knowledge however with entirely different classification algorithms.

Their square tests seven primary heart disease factors, including smoking, physical inactivity, diet, obese, steroidal alcohol, diabetes and high power per unit area, according to Benjamin EJ, et al. They also addressed cardiopathy figures along with stroke and disease.

Abhay Kishore, et al., have shown that replicated neural networks have sensitive accuracy in comparison to alternative algorithms such as CNN, Naïve Thomas Bayes and SVM. Th erefore in cardiopathy prediction, neural networks perform well. Together they have achieved a device that can anticipate quiet heart attacks and warn the user as soon as possible.

M. Nikhil Kumar, et al., used the random forest, Naïve Thomas Bayes, KNN, vector support, model tree supply algorithm law. The algorithmic rule of Naive Thomas Bayes yielded important results in relation to alternative algorithms. They used the cardiopathy dataset UCI repository. In comparison, the J48 algorithm rule took less time to develop and achieved significant results [23].

Kau, et al., compared a variety of different algorithms, such as Artificial N etwork N eural, K-nearest neighbour, Naive Thomas Bayes, Cardiopathy support vector machine.

Writer F Weng, et al., [15] used four machine learning algorithms like supplying regression, random forest, gradient boosting machines and neural networks. They showed that machine learning algorithms perform well at predicting the guts malady cases properly. They assert that this is often the primary experimentation victimization machine learning techniques to routine patient information in electronic records. The supply of the dataset is that the Clinical observe analysis Datalink (CPRD). These square measures the electronic medical records that contains all the medical connected information like statistics of human population, case history, specialists. It conjointly contains details of medication intake, outcomes, and details of hospital admissions.

Sahaya Arthy, et al., [16] analyse the present works on cardiopathy prediction that uses data processing. The information mining techniques square measure usually utilized in cardiopathy prediction. They conjointly discuss the databases used like the guts malady dataset from UCI repository, tools used like Maori hen, fast manual labourer, Data melt, Apache driver, Rattle, KEEL, R data processing and then on. They conclude that use of single algorithmic rule leads to higher accuracy in prediction. However, use of cross of 2 or a lot of algorithms will enhance and improve the guts malady prediction with sensible accuracy.

Heart disease prediction system uisng machine learning techniques

Machine learning is the application of Artificial Intelligence (AI) which provides systems with learning abilities without being explicitly programmed. UM focuses on a situation of computer programmes that access data and use it to be stated for their own.

Machine learning begins with observing data like examples, direct experience, or instruction, so that appear to be for patterns in data and make better choices for future. First of all, the requirement is to let the computer detect the information without human interference.

Heart disease prediction

Coronary ill health depicts the range of conditions that affect your heart. Diseases below the coronary ill health under the umbrella includes venous disease, such as coronary artery disease, heart beat problems (arrhythmias), and heart loss to the globe [24-27]. The term "coronary illness" is often used conversely with the term "cardiovascular malady" this refers to arteries that are blocked or limited and can lead to a coronary episode, chest pain (angina), or stroke. These various heart conditions can result in heart arrhythmias, *i.e.*, an irregular rate or rhythm of the heartbeat. Coronary ill health is surely one of the major reasons for sickness and death among many people on the globe. Expectation of vessel illness can be viewed as principal subject within the area of clinical study. There is so much to learn within this industry.

As per an article, coronary ill health finally ends up being the most supply of death for the 2 girls and men. The article expresses the incidental Around 610,000 people pass away of coronary ill health within the US every year-that is one in every four passing's [28-30]. Coronary ill health is that the main supply of death for the 2 folks. The larger a part of the passing's because of coronary ill health in 2009 were in men. Coronary Heart Disease (CHD) is that the most well-known quite coronary ill health, death penalty quite 370,000 people yearly. Systematically around 735,000 Americans features a metastasis failure. Of these, 525,000 are a primary heart disease and 210,000 occur in quite a whereas WHO have simply had a vessel failure. This makes coronary ill health a big worry to be managed.

Build with IoT prototype

The system is designed to work with sensors and microcontroller. The components needed for setting up have portable system are:

- Dht11 thermistor
- Heartbeat sensor
- Arduino microcontroller Electrocardiogram (ECG) sensor
- Arduino board

These 3 sensors are attached to the Arduino microcontroller to collect heart rate, basal metabolism ECG signals.

However, it's onerous to differentiate coronary ill health visible of many conducive danger factors, for instance, diabetes, high blood pressure, elevated cholesterol, uncommon heartbeat rate, and diverse completely different variables. Because of such needs, researchers have turned towards gift day approaches like data processing and machine learning for anticipating the ill [31].

The potential of use of internet of things in healthcare is pointed out. Several pioneering works on healthcare related IoT solutions are discussed. In the future, IoT will be being used for improving healthcare systems. Research in the related fields showed that remote health monitoring was feasible, but perhaps even more important is its ability to provide these benefits in different contexts. Remote monitoring has proven to replace cost of medical devices.

Module

- Support Vector Machine (SVM)
- Naive Bayes Algorithm
- Decision Tree Algorithm

Datasets

- The Kaggle heart disease databases from the UCI repository as carried. It consists of 303 records, 8 attributes.
- The data set consists of 3 different kinds of attributes.
- Input attribute
- Key attribute
- Predictable attribute

Input attributes

- Age in year
- Gender (male: 1, female: 0)
- ECG (Electro Cardio Graphic results)
- Heart rate (range measure)
- Chest pain (no of major vessels coloured by fluoroscopy (value(0-3))
- Cholesterol (mg/dl)
- Blood pressure (mmHg on admission to the hospital)
- Blood sugar (value 1: >120 mg/dl, value: 0<120 mg/dl)

Predictable attribute

Heart disease.xls

- A list of actions carried out by the user has stated above, here the user is Admin.
- Admin logins into the system, where he can execute the operations. One is the prediction and the second is classification.
- Admin imports the dataset and prediction is performed (using 3 different machine learning algorithms-SVM, Naive Bayes, Decision tree).
- The next classification is being performed where the user gets to know which algorithm gives the highest possible accuracy in the form of a Table 1, Figures 1 and 2 [19].

Tab. 1. Dataset	Age	Gender	cp	trestbps	Chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
	44	1	1	120	263	0	1	173	0	0	2	0	3	1
	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
	58	0	3	150	283	1	0	162	0	1	2	0	2	1
	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
	58	0	2	120	340	0	1	172	0	0	2	0	2	1
	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
	42	1	0	140	226	0	1	178	0	0	2	0	2	1
	61	1	2	150	243	1	1	137	1	1	1	0	2	1
	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
	59	1	2	150	212	1	1	157	0	1.6	2	0	2	1

Note: Curable (value 0: (No cardiovascular disease); value 1: (Has cardiovascular disease)).
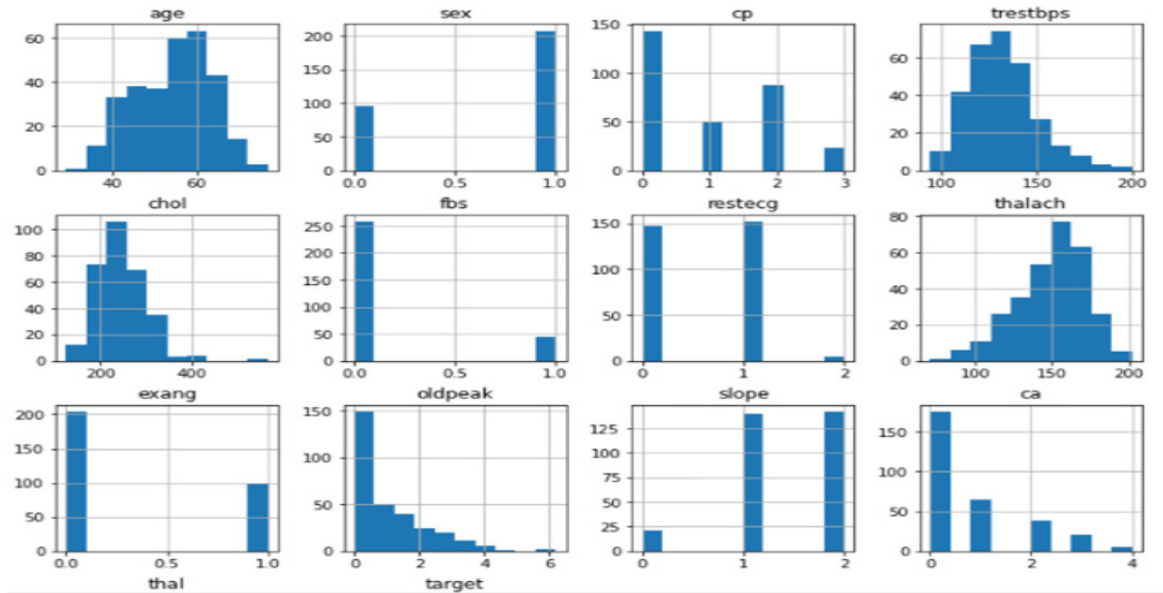
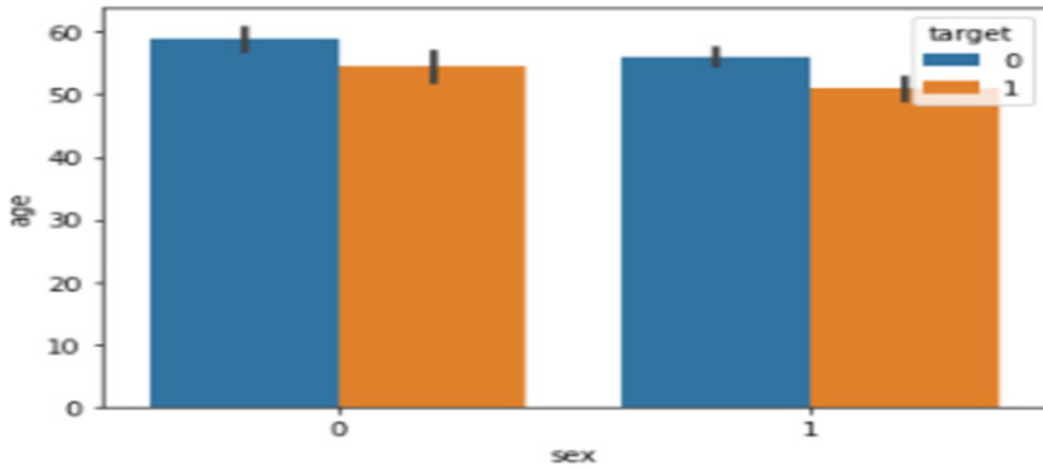**Fig. 1.** Dataset preprocessing.

Fig. 2. Probability of getting heart disease age *vs.* sex.

System architecture

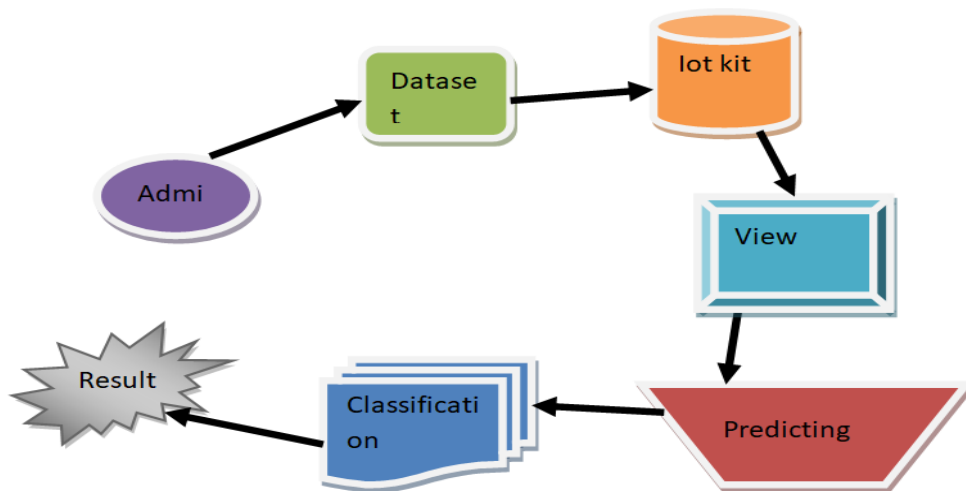The architecture diagram shows in Figure 3.

Fig. 3. Architecture diagram.

Defined how this system works step by step

- List of databases containing patient details.

- The method for selecting attributes selects the useful attributes for predicting heart disease.

- Once the information services available have been established, they are additionally chosen, cleaned and rendered in the form requested.

- Various classification methods, as indicated, will be applied to pre-processed data to predict the accuracy of heart disease.

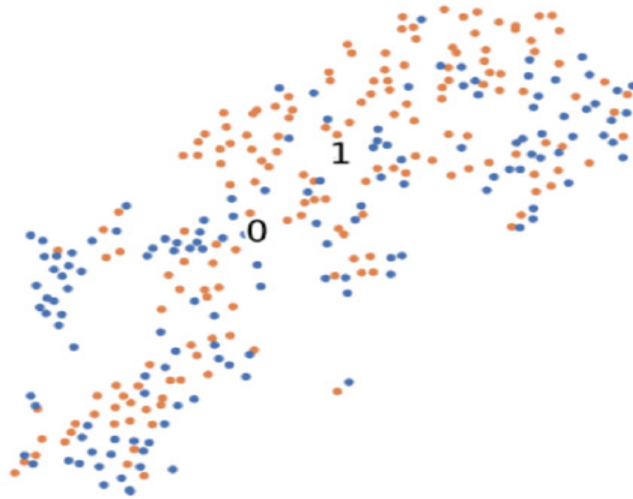- Accuracy measuring compares the different classifier's accuracy (Figure 4).

Fig. 4. Data visualization using scatterplot.

METHODOLOGY

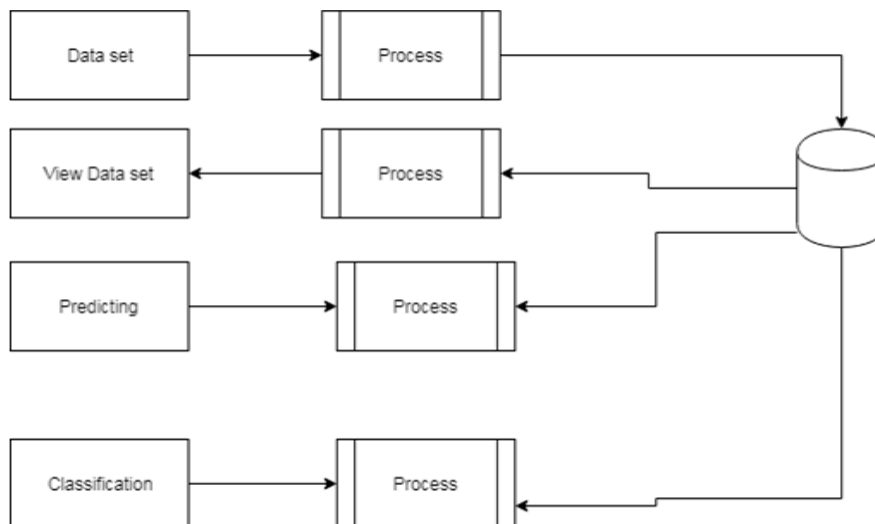### Data flow

The data flow shows in Figure 5.

Fig. 5. Data flow.

Support vector machine

SVM is a supervised ML algorithm that could be used for classification or regression problems. It uses a method known as the kernel trick to convert your data and then based upon these changes it finds the best possible boundary between the possible outputs. Various SVM algorithms use different kinds of kernel functions. These features can be for different types. For example, polynomial, nonlinear, linear, Radial Basis Function (RBF), and sigmoid.

Advantages of support vector machine

- Accuracy
- Works well with the limited datasets
- Kernel SVM includes a non-linear conversion function to convertthe complex non-linearly separate data into linearly separate data.

Disadvantages of support vector machine

- It does not work well with bigger datasets.

- Sometimes, SVM training time can be high.

Model building classification steps for SVM in Python

- **Step 1**: Import the Pandas library and dataset using Pandas.
- **Step 2**: Describe the features and the objective.
- **Step 3**: Before building the SVM algorithm model divide the dataset into train and test using sklearn.
- **Step 4**: Import SVC function from Sklearn. Build the SVM model with the help of the SVC function.
- **Step 5**: Predicting the values using the SVM model.
- **Step 6**: Evaluate the performance of SVM model.
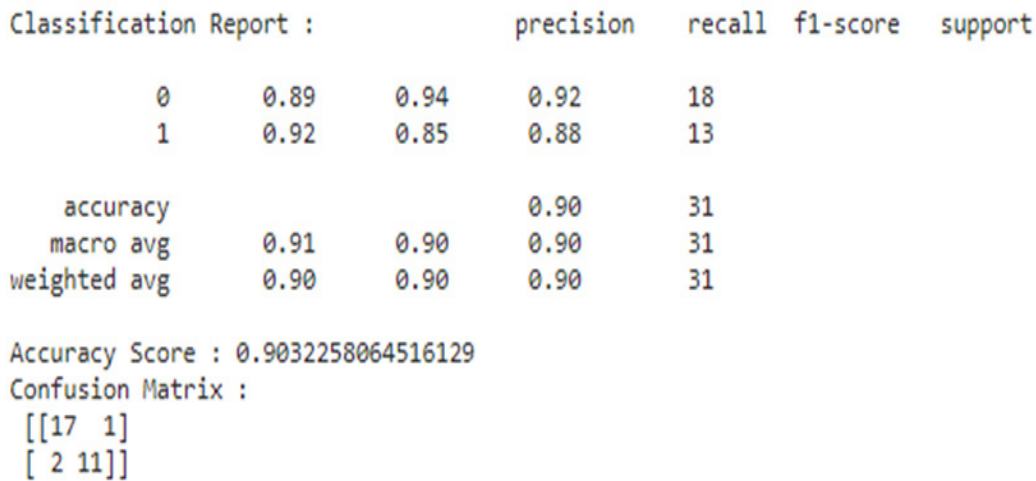
SVM Results shows in below Figure 6.

```
Classification Report :          precision  recall  f1-score  support

              0       0.89      0.94      0.92        18
              1       0.92      0.85      0.88        13

       accuracy                           0.90        31
      macro avg       0.91      0.90      0.90        31
   weighted avg       0.90      0.90      0.90        31

Accuracy Score : 0.9032258064516129
Confusion Matrix :
 [[17  1]
 [ 2 11]]
```

Fig. 6. SVM Results.

Naive Bayes classifier

The Naive Bayes algorithm is based on Bayes theorem. Time-series clustering is one of the simplest algorithms to track. Naive Bayes is an accurate and easy model for classifying probabilities. Naive Bayes classification is highly reliable and quick on large datasets [32].

Naive Bayes classifier is based on the premise that the influence of a particular feature within a class is independent of the effects of other features.

This is a very easy assumption, and it makes estimation simpler. Class conditional freedom plays a major role in education.

The following steps are used to calculates and probability for the Naive Bayes classifier [20]:

- **Step 1:** Measure the conditional probability for each of the labels for the stated conditions.

- **Step 2:** Find the likeliness likelihood for any attribute for each class.
- **Step 3:** Now, enter these statistics into the Bayes formula and calculate the likelihood.
- **Step 4:** The next move is to classify which class has a higher likelihood.

Naive Bayes sentence

Phase 1: Setting up Naive Bayes classifier.

Phase 2: The data sets given to the classifier should be trued in two phases.

Phase 3: Predictions should be made (Figure 7).

Results shows in Figure 7.

```
Classification Report :            precision   recall  f1-score   support

               0        0.94       0.94      0.94        18
               1        0.92       0.92      0.92        13

        accuracy                             0.94        31
       macro avg        0.93       0.93      0.93        31
    weighted avg        0.94       0.94      0.94        31

Accuracy Score : 0.9354838709677419
Confusion Matrix :
 [[17  1]
 [ 1 12]]
```

Fig. 7. Naive Bayes classifier results.

Tree for decision

- Decision tree is a technique of supervised ML that can be used for any classification and regression problems, but is preferred for classification problems to be determined. It is tree-structured, wherever the characteristics of a dataset are present within the nodes, branch labels the rules of the decision and each blade node illustrates the outcome [33].

- Only two nodes, the decision node and the leaf node, exist in a decision tree. Decision nodes are used to generate decisions and have several branches while leaf nodes are the output and no other branches.

- Decisions or tests shall be taken on the basis of the particular data set characteristics.

- It is a graphical display to find all potential problem/decision solutions based on the conditions given.

- A decision tree is named, since it, like a tree, starts with a root node, spreads over other branches, forming a structure like a tree.

- We use the CART algorithm to construct a tree, which represents the tree algorithm for classification and regression decision.

- A decision tree asks a question and is based on the answer (yes/no), dividing the tree further into sub-trees.

Algorithm steps of decision tree

- **Step-1:** Tree begins with the root node, says S, which contains the entire dataset.

- **Stage-2:** Find the best attribute in the dataset by using the selection Attribute Measure (ASM).

- **Step-3:** Split the S into sub-sets that contain the best attribute values.

- **Step-4:** Build a node of decision tree with the best attribute.

- **Step-5:** A recursive decision-making bodies using sub-sets of the dataset generated by step-3. Continue this method until you reach a point in which the nodes cannot be further classified and the final node identified as a leaf node.

Results show in Figure 8.

```
Classification Report :            precision   recall  f1-score   support

               0        0.89       0.94      0.92        18
               1        0.92       0.85      0.88        13

        accuracy                             0.90        31
       macro avg        0.91       0.90      0.90        31
    weighted avg        0.90       0.90      0.90        31

Accuracy Score : 0.9032258064516129
Confusion Matrix :
 [[17  1]
 [ 2 11]]
```

Fig. 8. Tree for decision results.

RESULTS AND DISCUSSION

Results and analysis

- Accuracy of the logistic regression model on your train: 0.9714285714285714.

- Accuracy of the Decision tree model on your test dataset is: 0.983111111111111.

Random forest has slight edge over DT. Logistic regression is a very simple model compared to DT/RF

- Accuracy of the SVM model on your test dataset is: 0.7773333333333333.

- Accuracy of the Naive-Bayes model on your test dataset is: 0.8171323489313833.

Finally, random forest has given better accuracy of prediction whether the person is having heart disease based on the give parameters *i.e.*, 0.9915 [34,35].

CONCLUSION

In my study, we compare the achievements of the different classification algorithms for the prediction of a heart disease with the use of SVM, NAIVE BAYE and DECISION TREE. Eight attributes were included in the experimental data collection. However, not all the characteristics of heart disease are equally emphasised. This is why a feature-selection approach was developed to eliminate unnecessary attributes that are not closely related to the other classification features. Instead of using 14 attributes in prediction of heart disease, each classification algorithm performs dramatically when using the 8 attributes selected. The DECISION TREE algorithm is 98 percent SVM 77 percent and NAIVE BAYES 81 percent with a 100 percent precision among the classifiers tested. The alternative is to figure out whether the patient is heart failure or not. A binary dilemma is overcome the multi-class issue for the identification of heart disease is strongly recommended by separating patients into separate groups. This method is tailored not only to forecast whether or not heart disease occurs, but also to predict cardiac arrest risk factor, so that patients receive additional treatment at an early stage to avoid cardiac insufficiency. Real-time data from different hospitals can be collected to classify patients with cardiovascular disease and to measure the classificatory' efficacy to make the diagnosis of patients with cardiac disease more accurate.

FUTURE WORK

In future work, other algorithms considering different attributes and different datasets can be performed, and instead of just predicting whether the patient has disease or not the level or the stage of the disease can be predicted. And also predicts whether it is curable or not!

REFERENCES

1. Han J, Kamber M, Pei J. Data mining: Concepts and techniques, Waltham. Morgan Kaufmann Publishers. 2012.

2. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In2008 IEEE/ACS international conference on computer systems and applications. IEEE. 2008; 108-115.

3. Shanthi D, Narsimha G, Mohanthy RK. Human intelligence *vs.* artificial intelligence: Survey. Int J Electron Comm Comp Eng. 2015; 6:30-34.

4. Sultana M, Haider A, Uddin MS. Analysis of data mining techniques for heart disease prediction. In2016 3rd international conference on electrical engineering and information communication technology (ICEEICT). IEEE. 2016; 1-5.

5. Xu S, Zhang Z, Wang D, Hu J, Duan X, et al. Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework. In2017 IEEE 2nd international conference on big data analysis (ICBDA). IEEE. 2017; 228-232.

6. Pouriyeh S, Vahid S, Sannino G, de Pietro G, Arabnia H, et al. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In2017 IEEE symposium on computers and communications (ISCC). IEEE. 2017; 204-207.

7. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. J Clin Epidemiol. 2013; 66:398-407.

8. Mansoor H, Elgendy IY, Segal R, Bavry AA, Bian J. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: A machine learning approach. Heart Lung 2017; 46:405-411.

9. Jabbar MA, Deekshatulu BL, Chandra P. Computational intelligence technique for early diagnosis of heart disease. In2015 IEEE International Conference on Engineering and Technology (ICETECH). 2015; 1-6.

10. UCI. UCI Machine learning repository. UCI. 2010.

11. Shanthi D, Lalitha A, Lokeshwari G. IoT based medical diagnosis expert system application. InProceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2018). Springer International Publishing. 2018; 685-692.

12. Nagamani T, Logeswari S, Gomathy B. Heart disease prediction using data mining with mapreduce algorithm. Inter J Innov Technol Exp Eng. 2019; 8:2278-3075.

13. Alotaibi FS. Implementation of machine learning model to predict heart failure disease. Int J Adv Comput Sci Appl. 2019; 10.

14. Repaka AN, Ravikanti SD, Franklin RG. Design and implementing heart disease prediction using Naives Bayesian. In2019 3rd International conference on trends in electronics and informatics (ICOEI). IEEE. 2019; 292-297.

15. Thomas J, Princy RT. Human heart disease prediction system using data mining techniques. In2016 international conference on circuit, power and computing technologies (ICCPCT). IEEE. 2016; 1-5.

16. Lutimath NM, Chethan C, Pol BS. Prediction of heart disease using machine learning. Int J Recent Technol Engg. 2019; 8:474-477.

17. UCI. Heart disease data. UCI. 2020.

18. Ambekar S, Phalnikar R. Disease risk prediction by using convolutional neural network. In2018 Fourth international conference on computing communication control and automation (ICCUBEA). IEEE. 2018; 1-5.

19. Rjeily CB, Badr G, El Hassani AH, Andres E. Medical data mining for heart diseases and the future of sequential mining in medical field. Machine learning paradigms: Advances in data analytics. 2019; 71-99.

20. Shanthi D, Kuncha P, Dhar MM, Jamshed A, Pallathadka H, et al. The blue brain technology using machine learning. In2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE. 2021; 1370-1375.

21. Alarsan FI, Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data. 2019; 6:1-5.

22. Vijiyarani S, Sudha S. An efficient classification tree technique for heart disease prediction. In International Conference on Research Trends in Computer Technologies (ICRTCT-2013) Proceedings published in International Journal of Computer Applications (IJCA). 2013; 201.

23. Rairikar A, Kulkarni V, Sabale V, Kale H, Lamgunde A. Heart disease prediction using data mining techniques. In2017 International conference on intelligent computing and control (I2C2). IEEE. 2017; 1-8.

24. Vazirani H, Kala R, Shukla A, Tiwari R. Use of modular neural network for heart disease. Int J Comput Commun Technol. 2010; 1:88-93.

25. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In2008 IEEE/ACS international conference on computer systems and applications. IEEE. 2008; 108-115.

26. Sultana M, Haider A, Uddin MS. Analysis of data mining techniques for heart disease prediction. In2016 3rd international conference on electrical engineering and information communication technology (ICEEICT). IEEE. 2016; 1-5.

27. Thomas J, Princy RT. Human heart disease prediction system using data mining techniques. In2016 international conference on circuit, power and computing technologies. IEEE. 2016; 1-5.

28. Kinge D, Gaikwad SK. Survey on data mining techniques for disease prediction. Int J Res Eng Technol. 2018; 5:630-636.

29. Shanthi D, Kiran K, Surya KL. Automatic vehicle alert system. Published in Jardcs Elsevier. 2018.

30. Sumalatha G, Muniraj NJ. IEEE. Survey on medical diagnosis using data mining techniques. In2013 International Conference on Optical Imaging Sensor and Security (ICOSS) 2013; 1-8.

31. Babu S, Vivek EM, Famina KP, Fida K, Aswathi P, et al. Heart disease diagnosis using data mining technique. In2017 international conference of electronics, communication and aerospace technology (ICECA). IEEE. 2017; 1:750-753.

32. Kaur A, Arora J. Heart disease prediction using data mining techniques: A survey. Int J Adv Comput Res. 2018; 9:569-572.

33. Parthiban L, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. Int J Biol Sci. 2008; 3.

34. Lee HG, Noh KY, Ryu KH. Mining biosignal data: Coronary artery disease diagnosis using linear and nonlinear features of HRV. InPacific-Asia conference on knowledge discovery and data mining. Springer. Heidelberg, Berlin. 2007; 218-228.

35. Guru N, Dahiya A, Rajpal N. Decision support system for heart disease diagnosis using neural network. Delhi Business Review. 2007; 8:99-101.