International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

A survey on Heart Disease Detection Using Machine Learning

Pruthvirajsinh Puvar[1],Neel Patel[2] ,Akshay Shah[3] ,Ruturaj Solanki[4] ,Prof.Dhaval Rana[5]
Computer Science and Engineering dept, R. N. G. Patel Institute of Technology, India

Abstract
With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to identify the symptoms of a heart stroke at an early stage and thus prevent it. It is difficult for a common man to frequently undergo expensive tests such as ECG and thus there needs to be a system in place which is realistic and at the same time reliable, in predicting the risk of a heart disease. Thus, we propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc. The machine learning algorithms and neural networks most reliable techniques for predicting results.

Keywords: Machine Learning, Neural Networks, Heart Disease, Data Processing, CHAD, Deep Learning.

1.INTRODUCTION

Cardiovascular diseases or Heart diseases are most dangerous disease and a leading reason of human deaths. According to World Health Organization (WHO), millions of people die every year because of arteriosclerosis only [1]. In excess of 1,000,000 Americans are determined to have cardiovascular breakdown in the United States every year, and an expected 6.2 million people in the United States at present live with cardiovascular breakdown. [11]. 1 out of 4 deaths occurs because of CVDs in India, ischemic coronary illness and stroke are the reason >80% of the times [13]. Basically, heart sickness is the disorder of heart/in heart, where Blood flow to the heart, brain or body is reduced because of a blood clot (thrombosis) development of greasy stores inside a conduit, prompting solidifying and narrowing of the supply route (atherosclerosis). Estimation shows that 30 million people will die due to the heart disease before 2040. This disease tends to affect persons most productive years of their life and makes ruinous impact on the social and economic being [14]. The physiological, social furthermore, psychological effect of these cardiovascular diseases varies across populations and individuals. Fortunately, there is an array of treatment options available. But to know about it at earlier stage there are various strategies which can be helpful to us for predicting the possibilities.AI and ML have benefited from sophisticated technology and improving data processing speeds, leading to the development of algorithms for autonomous driving image and facial recognition, automated recommendation software, and processing of natural language.[12] The techniques include, Machine Learning, Neural Networks and algorithms like K-Nearest Neighbor, Linear Regression, Naïve Bayes, Multilayer Perceptron, Support Vector Machine etc. All these techniques have been verified as cost effective, highly efficient, and less painful than conventional medical techniques. However, in many computer vision problems, it becomes undeniable that both CNNs and deep learning are the technique of choice [2].

2.BACKGROUND

2.1 Heart Disease

Heart disease describes a range of conditions from which your heart can be damaged. The diseases under heart disease umbrella includes small vessel diseases, such as coronary artery heart disease (CHAD), Problems of heart rhythm (arrhythmias); and defect in heart you were born with (congenital heart defects). The word heart disease is often used as the term cardiovascular disease (CVD).

Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels which will result in a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affecting your heart's muscle, valves or rhythm, also are often considered forms of heart disease.[3] Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year. CVDs are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions. Four out of 5CVD deaths are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age [1]. Population-based studies show that atherosclerosis, the major precursor of cardiovascular disease, begins in childhood. The Pathobiological Determinants of Atherosclerosis in Youth (PDAY) study demonstrated that intimal lesions appear in all the aortas and more than half of the right coronary arteries of youths aged 7–9 years [15].

The CHAD is one of the most common reasons of mortality in heart disease, there are several features of CAHD disease that can affect the structure or function of the heart. The physicians and doctors face many problems to detect heart disease correctly and rapidly. So, it is significant to make an intelligence CAHD prediction model to predict the heart disease in an initial state with a low cost. Without any previous symptoms, twenty-five percent of people die suddenly who are suffered by CAHD [4]. CAHD is one of the most significant types of diseases that can affect the heart badly and causes heart attack. Timely treatment and being aware of diseases symptoms can reduce CAHD.

2.2 Problem Statement

According to WHO (World Health Organization) Heart Disease is one of the most dangerous disease and leading reason of mortality, Twenty-five percent of people die because of heart stroke without having any symptoms of CVD (Cardiovascular Disease) [1]. Patients must go to some clinic and test whether he/she is diagnosed with CVD. Even for prediction of heart disease patient must pay some of their money. Patients don't just go through tests they also go through mental and physical pressure depending upon the situation and place. Tests are not handy and easy to access, and some are time consuming because of the flow of system.

3. METHODOLOGY

To initiate with the work, we have to start collecting data in each and every aspect towards the goal of the system. At first, the research was in the direction of the main causes or the factors which have strong influence on the heart health. Some factors are unmodifiable like age, sex and family background but there are some parameters like blood pressure, heart rate etc. which can be kept in control by following certain measures. Many doctors suggest healthy diet and regular exercise to keep the heart healthy. Following are the parameters which are considered for the study in designing the system which have major risk percentage with respect to CAD

1. Age
2. Sex
3. Blood Pressure
4. Heart Rate
5. Diabetes
6. Hyper cholesterol
7. Body Mass Index

The next step was to collect dataset. For this there are many datasets which can be used. i.e. Framingham heart study dataset. The dataset contains over 4000 people's records with over 14 features related to them. It has the target attribute which suggest the probability of the person to get diagnosed by CVD in coming 10 years [16], Cleveland dataset from UCI library. The datasets may contain as many as 76 parameters describing the complete health status of heart. These parameters are obtained by expensive clinical tests like ECG, CT scan etc. Out of these, the traditional heart disease prediction system uses 13 major parameters [6][7]. Since these parameters require expensive lab tests to find ECG, chest pain type, ST depression etc. To avoid these and to make system less complex partial datasets can be used by measuring the impacts of the attributes. The following research work briefly explains the following existing methods/algorithms are used while prediction of disease.

4. RELATED WORK

The use of computer algorithms that can learn complex patterns from data has significant potential to impact cardiology due to the number of diagnostic and management decisions that depend on digitized, patient-specific details such as , echocardiograms, and more, and due to the growing amount and sophistication of medical expertise the complexity of medicine now exceeds the ability of the human mind.ML algorithms can help in acquiring, interpreting, and synthesizing health care data from disparate sources and putting it at our fingertips. Machine Learning is mainly used for prediction purpose. In which, the machine gets trained using older dataset and the trained machine gets tested on a different dataset from the dataset it was trained on. Below are some algorithms used to train model.

4.1 K-Nearest neighbour

K-Nearest Neighbor has been commonly used to mine comprehensive medical database information [8]. KNN algorithm is a method of classification which is based on the similarity of one case to other cases. At a particular point, when a case is new, its distance from every one of the cases in the model is determined. The output of the technique indicates the case just like the closest neighbor, which is the most comparable. In this manner, it puts the case into the output that contains the closest neighbors. K-NN has 2 steps:

(1) Find the K training occasions which are closet to the unidentified occurrence.

(2) Pick the most frequently occurring classifications for these K occasions.

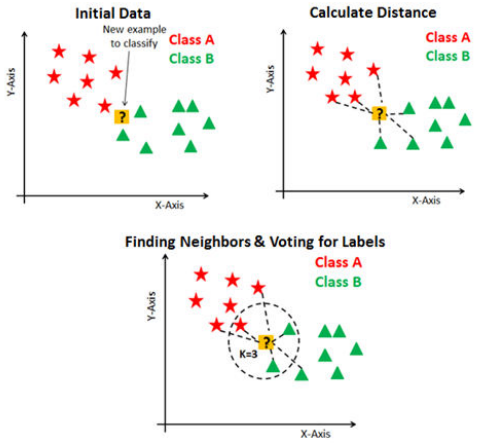


Fig- K-Nearest Neighbor [8][17]

Here in above figure suppose that R1 is the point, for which label needs to predict. First, find the K closest points to the R1. These distant points are measured using distance measure like Euclidean distance, Manhattan distance, Hamming distance etc. Then each object votes for the class they belong to and the class which has the most votes is selected as the predicted value [17].

4.2 Naive bayes algorithm

The Naive Bayes classifier is a probabilistic classifier based on the application of the Bayes theorem with simple assumptions of (naive) independence between characteristics. It is easy to create a Naive Bayesian model without complicated iterative parameter estimation, which makes it particularly useful for diagnosing heart patients in the field of medical science. Despite its simplicity, The Naive Bayesian classifier is often surprisingly good and is commonly used because it often outperforms more advanced methods of classification. How to calculate the posterior probability, P(c|x), from P(c), P(x), and P(x|c) is provided by the Bayes theorem. The Naive Bayes classifier assumes that the impact on a given class (c) of the value of a predictor (x) is independent of the values of other predictors. This presumption is called the conditional independence of the class

3.2 Equations:

• The Bayes theorem is:

$$P(A|B) = P(B|A)\ P(A) / P(B)$$

• P(A|B) is the posterior probability of the predictor class (target) given (attribute).

• P(A) is the prior probability of class

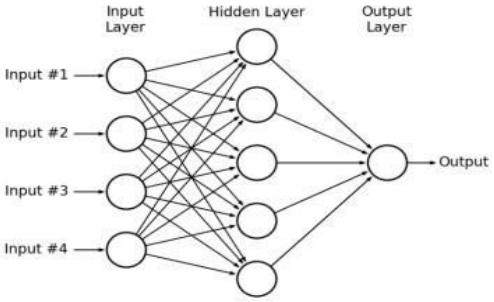
• P(B|A) is the probability of the predictor class given.

•P(B) is the prior probability of predictor

Where the two events are C and X (e.g. the probability that the train will arrive on time given that the weather is rainy). The probability theory is used by such Naïve Bayes classifiers to find the most likely classification of an unseen (unclassified) instance. If we have numerical data in the training set, the algorithm performs positively with categorical data but poorly. [7][18].

4.3MLP

As its name indicates, the multilayer perceptron neural network is composed of multiple layers. Only linearly separable problems are solved by the single layer perceptron, but many of the complex problems are not linearly separable, so one or more layers are added to single layer perceptron to solve such problems, so it is known as multilayer perceptron. The multilayer perceptron network, as shown in Fig.1, is known as a feed-forward neural network with one or more hidden layers. In general, they are used for pattern recognition, input pattern classification, prediction based on input information and approximation. It is independent of the orientation and size of the image. Color histogram is the most popular method for extracting the color information from the image. It gives the information about distribution of different colors that are present in an image.



i

Fig- MLP[9]

In case of digital images, histogram is that the number of pixels that have colors values in each of a fixed list of color ranges that cover the image's color space. The color histogram can be designed for any form of color space, but it is mostly used for three-dimensional spaces like RGB or HSV. In case of monochromatic images, the term intensity histogram may be used instead. Above network have an input layer with three neurons, one hidden layer(at the center) with three neurons and an output layer with three neurons [9]. Input Layer - The input layer accepts the input vector (x1...xp) and standardizes the values of every variable within the of -1 to 1. Then the distribution of those standardized values together with constant input called bias valuable 1 is given to every hidden layer neurons by input layer this bias value is then multiplied by a weight and added to the sum that's going into the neuron. Hidden Layer - At each neuron within the hidden layer, a weight (wji) is multiplied to the worth from each input neuron. Hidden Layer - At each neuron within the hidden layer, a weight (wji) is multiplied to the value from each input neuron. Hidden Layer - At each neuron within the hidden layer, a weight is multiplied to the worth from each input neuron. Then a combined value ui is produced by adding the resulting weighted values from each hidden layer neuron. This weighted sum (uj) is then given to the a transfer function, Generation of useful outputs hj. The combined outputs obtained from the hidden layer's neurons are then given to the neurons in the output layer. Output layer - The neuron weight (wkj) of each output layer is multiplied by the value of each hidden layer of the neuron, and then, by adding the resulting weighted values, the combined value vj is formed.

4.4 SVM

Support vector machines have shown excellent performance for disease prediction in the medicine sector in recent years[8]. SVM is a supervised learning system and the main aim is to design it for tasks of regression and classification, as well as minimizing errors in generalization. SVM classifies the data over a hyperplane into two classes. In spaces with large dimensions, SVM is very effective even the dimensions are greater than the number of samples.

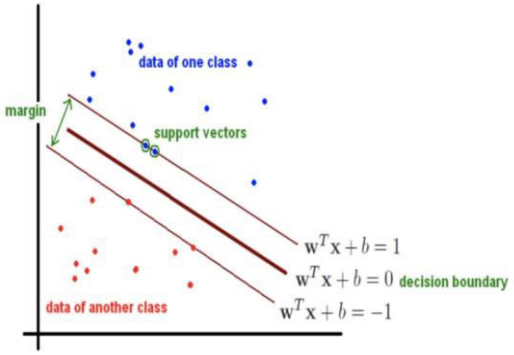


Fig- SVM[19]

Mathematically, SVM represented as follows

If $Yi = +1; wxi + b \geq 1$ (1)

If $Yi = -1; wxi + b \leq 1$ (2)

For all i; $yi(wi + b) \geq 1$(3)

x is a vector point in the equation and w is a weight and a vector. Therefore, the data in Equation (1) should always be greater than zero to separate and the data in Equation (2) should always be less than zero to separate. SVM selects the one among all possible hyperplanes where the hyperplane distance is as wide as possible.

4.5 Neural Network

Neural networks are a set of algorithms that are designed to recognize patterns, modelled loosely after the human brain. Through a kind of machine perception, labelling or clustering raw input, they interpret sensory data. The patterns they recognize are numerical, contained in vectors, into which all real-world information must be translated, be it images, sound, text or time series .[21]

Neural networks help us cluster and classify. They help to group unlabeled data according to similarities among the example inputs, and they classify data when they have a labelled dataset to train on.[21]

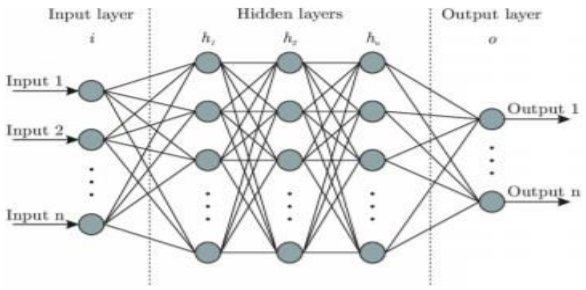


Fig- Neural networks[21]

Deep neural networks contribute to medicine, pathology, and other medical sectors[7]. A DNN is a complex neural network structure where, between the input and output layers, there is a neural network with several hidden layers. The input data is transformed into non-linearity or activation functions in order to output one or more linearly dividable classes. The intermediate layers are known as layers that are hidden. A deep neural network with a hidden layer is a f: RA → RB function, where A and B are respectively the size of the input vector and output vector. The relation between the vectors of input and output is expressed as follows: f(x) = Ø(b(2) + w(2)

(ϕ(b(1) + w(1)))) (5)

With bias vectors b(1) and b(2) ,

weight matrices w(1) and w(2) ,

and activation functions φ and ϕ.

4.6 Decision Tree

Decision tree is the graphical representation of the data and it is also the kind of supervised machine learning algorithms.

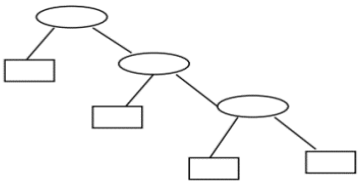


Fig. Decision Tree[22][24]

For the tree construction we use entropy of the data attributes and on the basis of attribute root and other nodes are drawn.

Entropy= $-\sum P_{ij} \log P_{ij}$

Pij is the probability of the node in the above entropy equation and the entropy of each node is calculated according to it. As the root node, the node that has the highest entropy calculation is selected and this process is repeated until all the tree nodes are calculated or until the tree is built.[22] When the number of nodes are imbalanced then tree is create the over fitting problem which is not good for the calculation and this is one of reason why decision tree have less accuracy as compare to linear regression.[24]

We first use each descriptive feature and divide the dataset along the values of these descriptive features to find the best feature that serves as a root node in terms of information gain, and then calculate the dataset entropy. Once we have split the dataset according to the function values, this gives us the remaining entropy. Then we subtract this value from the dataset's originally calculated entropy to see how much this splitting of

features reduces the original entropy that gives a feature's information gain and is calculated as:

InformationGain(Feature) = Entropy(Dataset) - Entropy(Feature)

The feature with the largest information gain should be used as the root node to start building the decision tree. ID3 algorithm uses information gain for constructing the decision tree.

Gini index is calculated by subtracting the sum of squared probabilities of each class from one. It favors larger partitions and easy to implement whereas information gain favors smaller partitions with distinct values.[23]

Gini index = $1 - \sum (p_i)^2$

A feature with a lower Gini index is chosen for a split. The classic CART algorithm uses the Gini Index for constructing the decision tree.

4.7 Logistic Regression

Logistic regression is basically an extended version of linear regression. Logistic regression is a statistical model which uses logistic function in its basic form to model a binary dependent variable. It is mostly used when the dependent variable is categorical [19]. Logistic regression helps in various ways for disease prognosis and diagnosis [20]. Mathematically,

Logistic regression estimates multiple linear regression functions defined as.

$\log p\,(y = 1) / 1 - (p = 1)$

$= \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta kXk;$

where k = 1, 2, ..., n

Sigmoid is one of the basic functions which are used in logistic regression which is bound to binary values, i.e. 0 and 1. The sigmoid curve ranges between 0 and 1 so the classification of two classes can be done using it. The curve shows the likelihood of data.

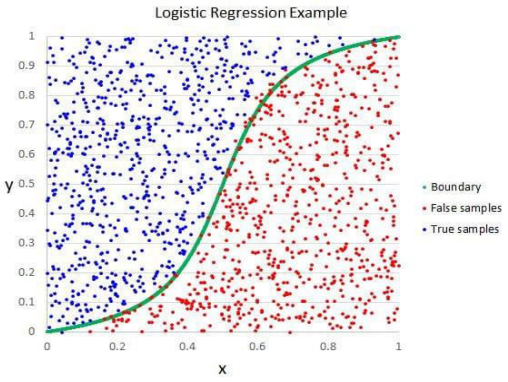


Fig-Logistic regression[19][20]

5. Paper Summery

[1] Prediction of Heart Disease Using Machine Learning Algorithm.[25]

In this Paper Authors Have Discussed about prediction of heart disease using Naïve Bayes and Decision tree. Firstly, by using data preprocessing the system will preprocess the past data of patients and if the data is small than the system will use Naïve Bayes Algorithm for predicting the risk of disease. If the data is big system will use Decision tree algorithm to predict the risk. Using K-means to cluster the similar data will help the system in predicting risk in better way.

[2] Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques.[26]

They compared several computer intelligence methods for the prediction of coronary artery heart disease in this paper. Seven Logistic Regression (LR) computer intelligence techniques, Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN) were used and a comparative study was drawn up. Using Statlog and Cleveland heart disease datasets, which are collected from the UCI machine learning repository database with several evaluation techniques, the performance of each technique is assessed.

[3] Intelligent Heart Disease Prediction System Using Data Mining Techniques.[27]

In this paper, they have used Data Mining method Decision tree which can handle high dimensional categorical data. Decision Trees can also handle continuous data (as in regression) but they must be converted to categorical data. Using Naïve Bayes, the system can find new ways of understanding and exploring data it learns from the target and other data provided to the algorithm. Neural Network consists 3 layers. input layer, output layer and hidden layer. Neural Network

algorithms use Linear and Sigmoid transfer functions. Neural Networks are suitable for training large amounts of data with few inputs. Using this techniques system will provide predicted result of heart disease.

[4] Prediction of Heart Disease Using Machine Learning Algorithms.[28]

There is a vast amount of information in the health care field, certain techniques are used for processing those data. One of the techniques often used is data mining. The leading cause of death worldwide is heart disease. The heart is the human body's vital organ. If a heart does not perform its function properly, it will affect the human-like kidney, brain, etc. According to WHO statistical data, one-third of the world's population has died from heart disease. Risk factors for heart disease include: high cholesterol, high BP, diabetics, smoking, family history of coronary disease, too much alcohol consumption, etc. Symptoms of Heart attack: They compared several computer intelligence methods for the prediction of coronary artery heart disease in this paper. Seven Logistic Regression (LR) computer intelligence techniques, Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbor (K-NN) were used and a comparative study was drawn up. Using Statlog and Cleveland heart disease datasets, which are collected from the UCI machine learning repository database with several evaluation techniques, the performance of each technique is assessed.

[5] Diagnosis of heart disease for diabetic patients using naïve bayes method[29]

In medical databases, the discovery of knowledge is a well-defined process and data mining is an essential step. In a large database, data mining involves the use of techniques to find underlying structures and relationships. We are applying the Naïve Bayes data mining classifier technique in this study that generates an optimal prediction model using a minimum training set. The proposed system uses diabetic diagnosis to predict characteristics such as age, sex, blood pressure and blood sugar, and the chances of a diabetic patient getting heart disease. Diabetes mellitus is a chronic disease that causes serious health problems, including renal failure, stroke, blindness, and heart disease, most notably. People with diabetes are two to four times more likely to get cardio vascular diseases, so the Naïve Bayes method is used to help diagnose diabetic patients with heart disease for this purpose. Naïve Bayes classifier is a term that deals with a simple probabilistic classifier based on the application of strong independence assumptions of the Bayes theorem. The data set used for this work is a set of clinical data. The specification of the clinical data set provides a concise, unequivocal definition of diabetes-related products. For data mining, the tool WEKA ('Waikato environment for knowledge analysis') is used. WEKA's main features are an open source and independent platform. It provides many distinct data mining and machine learning algorithms. The proposed naïve model of Bayes was able to properly classify the input instances.

[6] Analysis and prediction of cardiovascular disease using machine learning classifiers[30]

A general term for conditions that affect the heart or blood vessels is cardiovascular disease (CVD). It is usually associated with an accumulation of fatty deposits within the arteries (atherosclerosis) and an increased risk of blood clots, referring to conditions that include limited or blocked veins that can cause a heart attack, torment of the chest (angina) or stroke. The classifier for machine learning predicts the ailment depending on the patient's side effect state. This paper intends to look at the presentation of the machine learning tree classifiers in anticipating cardiovascular disease(CVD). Dataset from the UCI (University of California at Irvine) repository; 10 attributes such as age, gender, cp, trestbps, cho, fbs, restecg, thalach, ca, and target were included in the data set. The cleaned data is divided into 60 percent training and 40 percent test based on the split criterion, then the dataset is subjected to five classifiers of machine learning such as Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN). Using the confusion matrix, the precision of the classifiers was calculated. It is possible to determine the classifier that offers the highest accuracy as the best classifier.

[7] Prediction of Cardiovascular Disease Using Machine Learning Algorithms[31]

In this Paper Authors Have Discussed about data pre-processing uses techniques like the removal of noisy data, removal of missing data, filling default values if applicable and classification of attributes for prediction and decision making at different levels. This is done by comparing the accuracies of applying rules to the individual results of Support Vector Machine, Gradient Boosting, Random forest, Naive Bayes classifier and logistic regression on the dataset taken in a region to present an accurate model of predicting cardiovascular disease.

[8] Predicting Heart Disease at Early Stages using Machine Learning.[32]

In this Paper Authors Have discussed about predicting heart disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. Several types of heart disease are there in the world; Coronary artery disease (CAD), and heart failure (HF) are the most common heart diseases that are present. Heart disease is a significant

problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise. Some of the supervised machine learning techniques used in this prediction of heart disease are artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes) (NB) and k-nearest neighbor algorithm. There are some common attributes which are used to predict

the heart diseases are: Gender, Age, resting blood pressure, Types of chest pain, Serum cholesterol in mg/dl, Fasting blood sugar, ECG results, Heart rate, Thalassemia, Old peak. Heart disease is a very critical issue in the present growing world. So, there is a need for an automated system to predict heart disease at earlier stages.

[9] Heart Attack Risk Prediction Using Machine Learning.[33]

The major cause of morbidity and mortality globally is heart disease: it accounts for more deaths than any other cause annually. The authors discussed the development of a screening tool in this paper to predict whether a patient has a 10-year risk of developing coronary heart disease (CHD) using various Framingham dataset machine learning techniques. The objective of the classification is to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). provides the patient's information. It includes over 4,000 records and 15 attributes. Each attribute is a potential risk factor. There are both demographic, behavioral and medical risk factors. The attributes are Demographic, Education, Behavioral, Information on medical history and Information on current medical condition. This model can then be used as a simple screening tool and all that we need to do is to input ones: age, BMI, systolic and diastolic blood pressures, heart rate and blood glucose levels after which the model can be run and it outputs a prediction.

Title / Topic	Author	Year	Algorithm	Results / Accuracy
Prediction of Heart Disease Using Machine Learning Algorithm.	Rajesh Nichenametla, T. Maneesha, Shaik Hafeez, Hari Krishna	May-2018	K-Means, Machine Learning, Naïve Bayes, Decision Tree (ID3).	Naïve Bayes is more accurate if the input data is cleaned and well maintained even though ID3 can clean it self it cannot give accurate results every time
Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques	Safial Islam, Md.Milon Islam, Md.Rahat Hossain	Jan-2020	Decision tree, K-nearest neighbor, Logistic regression, Deep neural network, NaïveBayes, Random forest, Support vector machine	Support vector machine : 97.36% (in cleaveland dataset) Deep Neural Network : 98.15% (in Statlog dataset)
Intelligent Heart Disease Prediction System Using Data Mining Techniques	Ms.Ishtake S.H, Prof. Sanap S.A	April 2013	Decision Tree, Neural Network, Naive Bayes	Decision Tree :79.06%, Neural Network: 76.17%, Naive Bayes : 76.17%
Prediction of Heart Disease Using Machine Learning Algorithms	Mr.Santhana Krishnan.J, Dr.Geetha.S	Jan 2018	decision tree, naïve bayes	Decision tree: 91% Naïve Bayes : 87%
Diagnosis of heart disease for diabetic patients using naïve bayes method	G Parthiban, Rajesh Appusamy, Shesh Srivatsa	June 2011	Naïve bayes algorithm, data mining	Naïve bayes: 74%
Analysis and prediction of cardiovascular disease using machine learning classifiers	N.Komal Kumar, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana	Feb 2020	Random forest, decision tree, logistic regression, support vector machine(SVM), K-nearest neighbors(KNN)	RandomForest:85.71% (Highest) ROC AUC: 0.8675

Fig- COMPARISON TABLE

[10] HPPS: HEART PROBLEM PREDICTION SYSTEM USING MACHINE LEARNING.[34]

Heart is the most important organ of a human body. It circulates oxygen and other vital nutrients through blood to different parts of the body and helps in the metabolic activities. In this paper authors have analyses about the different prescribed data of 1094 patients from different parts of India. Using this data, they have built a model which gets trained using this data and tries to predict whether a new out-of-sample data has a probability of having any heart attack or not. They collected data from a survey of approximately 1000 patients from different parts of India and found a correlation between the various risk factors they collected. Family history, smoking, hypertension, dyslipidaemia, fasting glucose, obesity, life style, CABG and high blood serum are the risk variables that have been taken as an input. They have the demographic details, apart from the risk factors mentioned, as well. The main intension of this paper is to help in the decision making of a doctor for detecting the possibility or identifying the patient's suffering or going to suffer from heart problems. they try to give the doctor with the better option with the history similar data result.

Using these data, the doctor can have a transparency with the patient and the patient won't feel cheated at the end.

7. PROPOSED SOLUTION

The development of artificial intelligence and machine learning capabilities, there is shining potential to spare time and mitigate errors saving millions of lives in the long run. Machine Learning (ML) and Convolutional Neural Networks (CNNs) can automate most of the diagnosis process with equal or more accuracy than the current methods. We can make this system cost effective, easy to access and handy so that all class of people can use this system and we can reduce the level of this diseases. It can serve as a helping tool for getting confident about the detection of heart disease even without having any symptoms of heart disease.

8. CONCLUSION

The Heart Disease Detection System using Machine learning Algorithm, which is MLP provides its users with a prediction result that gives the state of a user leading to Coronary Artery Disease. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we are going to use Multi Layered Perceptron, K nearest neighbor, Naïve Bayes, Support Vector Mechanism in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known to the application and the rate of people dying due to heart diseases will reduce eventually. Due to the consequences of inadequate detection and the necessity of excellent detection accuracy, the detection of Heart disease has been deemed challenging. The implications of artificial intelligent methods and the efficient utilization of soft computing skills would negate the issues related to detection inaccuracy. Numerous techniques highlighting the classification and detection of heart disease were discussed in this study.

REFERENCES

[1] World Health Organization overview of CVDs [Online] Available: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

[2] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, ``Deep learning for visual understanding: A review,"Neurocomputing, vol. 187, pp. 27_48, Apr. 2016.

[3] Mayo clininc, Mayoclinic.org [Online] Available: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118

[4] R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby. Braunwald's heart disease: A textbook of Cardiovascular Medicine," Vol. 9, Saunders, New York, 2012.

[5] A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.

[6] Nichenametla, Rajesh & Maneesha, T. & Hafeez, Shaik & Krishna, Hari. (2018). Prediction of Heart Disease Using Machine Learning Algorithms. International Journal of Engineering and Technology(UAE). 7. 363-366. 10.14419/ijet.v7i2.32.15714.

[7] K.Vembandasamy*PR*,*R*R.Sasipriya*PP*andE.Deepa "Heart Diseases Detection Using Naive Bayes Algorithm" IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015

[8] Safial Islam Ayon, Md. Milon Islam & Md. Rahat Hossain" Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques" https://www.tandfonline.com/doi/abs/10.1080/03772063.2020.1713916

[9] Sonawane, J. S., & Patil, D. R. (2014). *Prediction of heart disease using multilayer perceptron neural network. International Conference on Information Communication and Embedded Systems (ICICES2014).*

[10] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586.

[11] Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. Circulation 2017;CIR, 0000000000000659.http://refhub.elsevier.com/S0002-8703(20)30215-5/rf0005

[12] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. Proceedings of the 25th international conference on machine learning. ACM; 2008;160-7. http://refhub.elsevier.com/S0002-8703(20)30215-5/rf0055

[13] India State-Level Disease Burden Initiative CVD Collaborators. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990–2016.Lancet Glob Health. 2018; 6:e1339–e1351. doi: 10.1016/S2214-109X(18)30407-8

[14] https://www.ahajournals.org/doi/10.1161/CIRCOUTCOMES.118.005195

[15] Vanhecke TE, Miller WM, Franklin BA, Weber JE, McCullough PA (October 2006). "Awareness, knowledge, and perception of heart disease among adolescents". European Journal of Cardiovascular Prevention and Rehabilitation. 13 (5):71823. doi:10.1097/01.hjr.0000214611.91490.5e. PMID 17001210. S2CID 36312234.

[16] Framingham heart study dataset available: https://datacatalog.med.nyu.edu/dataset/10046#__sid=js0

[17] K-Nearest Neighbor: https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn

[18] Naïve Bayes: https://towardsdatascience.com/introduction-to-na%C3%AFve-bayes-classifier-fa59e3e24aaf

[19] https://towardsdatascience.com/heart-disease-risk-assessment-using-machine-learning-83335d077dad

[20] https://www.sciencedirect.com/science/article/pii/S0735109720378943

[21] Akash, Kunder & Shashank, H & .S, Srikanth & A.M, Thejas. (2020). Prediction of Stroke Using Machine Learning.

[22] https://blog.quantinsti.com/gini-index/

[23] https://blog.clairvoyantsoft.com/entropy-information-gain-and-gini-index-the-crux-of-a-decision-tree-99d0cdc699f4

[24] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, Gorakhpur, India, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

[25] Nichenametla, Rajesh & Maneesha, T. & Hafeez, Shaik & Krishna, Hari. (2018). Prediction of Heart Disease Using Machine Learning Algorithms. International Journal of Engineering and Technology(UAE). 7. 363-366. 10.14419/ijet.v7i2.32.15714.

[26] https://www.tandfonline.com/doi/abs/10.1080/03772063.2020.1713916

[27] Ms. Ishtake S.H and Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques" *2013 International J. of Healthcare & Biomedical Research*

[28] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms.," *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICIICT1.2019.8741465.

[29] Parthiban, G & Appusamy, Rajesh & Srivatsa, Shesh. (2011). Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method. International Journal of Computer Applications. 24. 10.5120/2933-3887.

[30] N. K. Kumar, G. S. Sindhu, D. K. Prashanthi and A. S. Sulthana, "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers," *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020, pp. 15-21, doi: 10.1109/ICACCS48705.2020.9074183.

[31] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, India, 2018, pp. 1-7, doi: 10.1109/ICCTCT.2018.8550857.

[32] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586.