Heart Attack Prediction Using Artificial Neural Networks

N. Nava Bharathi
Department of Electronics and Instrumentation Engineering,
V. R. Siddhartha Engineering College Vijayawada, India
navabharathi2001@gmail.com

Mahammad Firose Shaik
Department of Electronics and Instrumentation Engineering,
V. R. Siddhartha Engineering College Vijayawada, India
firose@vrsiddhartha.ac.in

T. Poojita
Department of Electronics and Instrumentation Engineering,
V. R. Siddhartha Engineering College Vijayawada, India
poojithatamminana@gmail.com

T.Sravanthi
Department of Electronics and Instrumentation Engineering,
V. R. Siddhartha Engineering College Vijayawada, India
sravanthithorlikonda@gmail.com

Mohammad Rafi
Department of Electronics and Instrumentation Engineering,
V. R. Siddhartha Engineering College Vijayawada, India
mohammad67rafi@gmail.com

Inakoti Ramesh raja
Department of Electronics and communucation Engineering
Aditya College of Engineering and Technology,
Surampalem, India
rameshraja.inakoti@acet.ac.in

2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) | 979-8-3503-9737-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACCS57279.2023.10112664

***Abstract*—Heart disease is responsible for 12 Million deaths yearly according to the World Health Organization. The National Crime Records Bureau estimated that the number of deaths caused by heart attack has been increasing since 2014. In that year 18,309 deaths occurred due to heart attack disease, year after year the number of deaths are growing and stated 28,005 lives in 2019, up by 53% in 5 years. With growing population diagnosing and starting treatment at the early stage is difficult. The annual deaths from cardiovascular disease in India is from 2.26 to 4.77 million during 1990 to 2020. The occurrence of heart disease in India is around 1.6% to 7.4% in rural and 1% to 13.2% in urban population. The intention of the study is to identify the features of heart disease and predict the risk using Artificial Neural Networks(ANN). Data analysis is done in Python using Jupyter Notebook to verify the Artificial Neural Networks.**

Keywords—ANN, Heart attack prediction, Jupiter Lab.

I. INTRODUCTION

The Cardiovascular disease is one of the crucial problem which causes majority of the deaths. It is very hard to identify heart disorders with a naked eye. So we can use conventional methods, which are based on the data of patient. The most significant advantage of predicting heart disease is that we can take precautions in its early stage. Heart disease detection in the initial stage can reduce its seriousness.

This system also helps in better decision-making regarding a patient's health condition, and various researchers have included risks of different features.

The most prevalent features are the 13 features [1]. Here, we are using artificial neural networks for higher prediction accuracy. [4]It is a computing system that performs similar tasks to the working of the brain. ANN solves the problem like a human brain.

ANN is a multilayer network consisting of an input, hidden, and output layer. By using ANN, we can predict the disease earlier and take precautions.

II. LITERATURE SURVEY

In 2016, A.Shetty and others, [1] proposed a heart attack predicting system using patient's data. To train the system they have taken 13 factors into consideration. This prediction system provides an accuracy of 84%.

SY Huang, et al. [2] The proposed system uses one of the Artificial Neural Network techniques that is Learning Vector Quantization Algorithm. This system involves three steps. Firstly, 13 features like chest pain, cholesterol, old peak, etc. Secondly, the model is trained using Artificial Neural Networks. At last, the prediction model is designed with an accuracy around 80%.

P. U. Anitha, et al. [3] proposed Machine Learning Algorithm based heart attack prediction model. SVM provides higher accuracy than Neural Networks in reference with this paper. The accuracy of the system using SVM is 87.5%.

Jayshril S. et al. [4] suggested heart disease prediction using Artificial Neural Network methodology. Using Vector Quantization Algorithm the network is trained. It consists of input,hidden and output layers.13 neurons are present in the input layer which represents the parameters. The hidden layer is utilized to achieve high accuracy and minimum errors. The output layer shows the result. System's performance can be increased by variable number of neurons. 86.6% accuracy is achieved.

Megha Shahi and others, [5] advised Data Mining methods for predicting heart disease. Waikato Environment for Knowledge Analysis software is utilized to diagnosis automatically and provide excellent healthcare. The paper consists of several algorithms and SVM is more accurate.

C. Beyene and others [6] suggested for prediction of heart attack in a short time using data mining approaches. This system is helpful for people who has less knowledge and skill in this area and for diagnosing the disease at an early stage. Here, we use WEKA software for analysis of the dataset.

The proposed paper introduced [7] a cardiovascular disease prediction system with the help of data mining methods. This helps in decision making by using various features. 87.3% and 86.3% is achieved during training and testing phase. Ajam [8] used neural network methodology for diagnosing heart attack. The model is tested to determine its effectiveness using feed-forward neural and backward propagation learning algorithms. ANN provides 88% accuracy in the hidden layer with appropriate functions.

S.H.Mujawar [9] recommended Naïve Bayes and k-means for cardiac attack prediction. Classification and clustering methods are used on 13 risk factors that are essential for building the model. From Cleveland heart database, 300 cases are taken into consideration for training the model. Detection accuracy is around 93% and non-detecting accuracy is about 89%. the result of this paper is in between diseased and not diseased.

Bayes Net, j48, K star, multilayer perception and SMO techniques are proposed in this paper [10]. Here, we use WEKA software. In this paper, SMO achieved higher accuracy than others. However, there is still room for improvement in the accuracy performances of those algorithms. In order to provide a better illness diagnosis, the accuracy performance is enhanced. The accuracy of K Star, J48, SMO, Bayes Net, and multilayer perception is respectively 75%, 86%, 88%, 89%, and 86%.

ANN algorithm is suggested by P. S. Chandrasekhar Reddy and others [11]. Since, the diagnosis of the illness is costlier this provided a need for developing a new system which is economical. Clinical parameters like age, maximum heart rate, old peak, etc. are used for building the model. This model predicts if the person is ill or not. The accuracy is displayed in Java.

This paper [12] proposes using a classification algorithm for heart attack prediction. It uses techniques such as Hadoop Distributed File System (HDFS), SVM, and Map Reduce to detect cardiac attack with attributes set. This paper compared different data mining techniques for predicting heart attack. HDFS is used for data storing in large amounts and SVM is used as predicting algorithm.SVM is applied in parallel, which improves computation speed compared to sequential SVM. SVM offers accuracy levels of 85% and 82.35% that are better and more effective. Parallel SVM outperforms sequential SVM in terms of accuracy.

Senthilkumar, [13] Mohan proposed Hybrid Machine Learning Techniques for heart attack detection. Linear Method (LM) & Random Forest (RF) are combined for achieving higher accuracy. It has given an accuracy of 88.7%.

Polaraju and others [14] recommended the Prediction of Heart attack using Multiple Regression algorithm, Multiple Linear Regression is suitable for predicting cardiovascular chance. This work consists of 3000 records with 13 risk factors. The data is partitioned as 7:3 where 30% is the testing data and 70% is the training data. It is obvious from the findings that the regression technique has a higher classification accuracy than any other algorithm. [16, 17]

Shah. D., and others [15] suggested using machine learning and data mining algorithm to detect heart attack. The intention of this research paper is to discover hidden patterns using data mining approaches. When compared to LMT, J48 has the highest accuracy. Sharan monica and sathees kumar [19] suggested Analysis of CardioVasular Disease Prediction using Data Mining approaches. This paper includes different techniques like j48, naive Bayes and simple cart which obtained 91%, 82% and 92 % accuracy respectively. R.kasabe and G.Narang [18] introduced cardiovascular ailments prediction and analysis based on deep learning techniques. This research paper contains different machine learning techniques which includes naive Bayes and decision tree. It concluded that Random Forest is the best classifier among them.

III. PROPOSED SOLUTION

The Artificial Neural Network (ANN) is used for complicated and challenging tasks. It is a data mining tool which is used for clustering & classification. A neural network [6] is made up of neuron-like processing units and there are weighted connections between them. In Artificial Neural Networks, there are two modes. One of them is called as activation transfer mode and the other one is learning mode. When the activation is sent over the network it is known as activation transfer mode. When the network typically arranges itself depending on the latest recent activation transfer. This is called as learning mode.

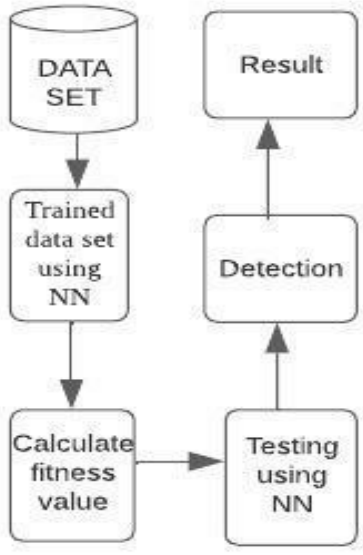

Fig. 1. Flow Diagram of the Artificial Neural Networks

Fig. 1 shows the flow diagram of the solution. This diagram contains various steps, including collecting the data, training, and testing the model. The output is whether the person is diseased or not.

A. Collecting Data

The heart attack data set consists of 13 features of 1024 instances like age, gender, chest pain, etc. [5] collected from UCI. Table 1 shows the attributes and their description. It shows the 14 features.

B. Preprocessing

The data cleaning is done, and checking whether the date set contains missing values is done. The imbalance of the date set is checked since it significantly affects the model's accuracy. To nullify this problem [4], we use the 'mean of column' method. The process includes the identification of noisy and inconsistent data. The data set contains numerical and categorical data. Some of the dataset's attributes are multiclass variables with double classification characteristics. The data set includes 14 attributes. Fig. 2 shows a picture of the dataset in the code.

TABLE 1. FEATURES AND THEIR DESCRIPTION

1	**Sex**	Person's sex (0=female and 1=male)
2	**Age**	Age of a person in years
3	**fbs**	Fasting blood sugar of a patient (>120mg/dl,1=true;0=false)
4	**thalach**	Maximum heart rate of a patient
5	**Resting ECG**	Resting electrocardiogram results Value2 means having ST-T wave abnormality Value 1 represents normal Value 0 represents Showing probable or definite left ventricular hypertrophy
6	**CP**	Chest pain types: Value 0 represents Asymptomatic Value 1 represents Atypical angina Value 2 represents Non-angina pain Value 3 represents Typical Angina
7	**Chol : cholesterol**	Cholesterol level in a person
8	**Restbps**	Blood pressure of a person
9	**Slope**	Represents slope of the peak exercise ST part. Value 0:down sloping; value 1:flat; value 2:up sloping
10	**Thal**	Thalassemia is a blood disorder. Value 0 represents Null Value1 represents fixed defect Value 2 represents Normal blood flow Value 3 represent Reversible defect
11	**Old peak**	ST depression induced by exercise relative to rest(ST relates to positions on the ECG plot)
12	**ca**	Number of major vessels(0-3)
13	**Exang**	Angina due to exercise
14	**target**	Heart disease(0=yes and 1=no)

C. Splitting Dataset

For estimating the performance of an desired algorithm, we use train-test split method. Firstly, we have to check if there are any missing values in the dataset. If any, find ways to reduce them. The next step is to differentiate between numerical and categorical data. The data is partitioned into two parts. One of them is the training part, in which the data is utilized to train the model. The other one is the test dataset in which the data is used to check the performance of the model. The data is given to the input layer and algorithm is performed. Here, the data is divided into 8:2 ratios.

D. Evaluation Metrics

There is a need to measure the performance of the model. So that we will know the need for improvement. Confusion Matrix results was employed to display an overview of prediction results for classification problems, including both correct and erroneous outcomes.[12] Additionally, it is used to identify type of errors and errors. The Confusion matrix components are as follows:

True Positives are the occurrences that are detected true and have the illness.

True Negatives are the occurrences that are detected false and do not have the illness.

False Positives are the occurrences that are detected false and have the illness.

False Negatives are the occurrences that are detected true and do not have the illness.

Various calculations performance for the model. Such as model's accuracy, using this matrix. The table 2 shows the parameters and formulas for this:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
5	58	0	0	100	248	0	0	122	0	1.0	1	0	2	1
6	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
7	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
8	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
9	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
10	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
11	43	0	0	132	341	1	0	136	1	3.0	1	0	3	0
12	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
13	51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
14	52	1	0	128	204	1	1	156	1	1.0	1	0	0	0
15	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
16	51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
17	54	1	0	124	266	0	0	109	1	2.2	1	1	3	0

Fig. 2. Dataset in the code

TABLE 2. CALCULATION FORMULAS FROM CONFUSION MATRIX

Terms	Formula
Accuracy of the model	(TN+TP)/(TN+TP+FN+FP)
Misclassification	(FN+FP)/(TN+TP+FN+FP)
Rate of True Negative	TN/(TN+FP)
Sensitivity	TP/(FN+TP)

Fig. 3 shows the confusion matrix. There are 100,102, 1, and 2 TP, TN, FP, FN, respectively.

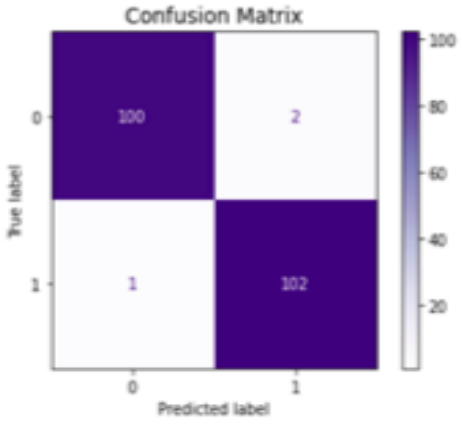

Fig. 3. Confusion Matrix

IV. RESULT

Data is obtained from UCI that contains 1024 samples. The dataset is partitioned into two halves for training and testing. ANN is implemented and the accuracy obtained is 98.33%.

TABLE 3. COMPARISON TABLE

Dataset obtained from	Year	Author	Technique used	Accuracy of the model(in %)
UCI	2016	[1]	ANN	84
UCI	2022	[3]	ANN	86.6
UCI	2015	[8]	ANN	88
UCI	2018	[12]	SVM	85
UCI	2019	[13]	LM and RF	88.7
UCI	Proposed		ANN	98.33

The Table 3 shows the comparison of previous paper and the current solution. By the end of the implementation part, [6]. The ANN classifier achieved 98.33% of accuracy with training 80% and testing 20%. The results outperformed compared to previous research work. Fig. 4 shows the classification report of the model. It includes precision, recall, f1 score and support.

```
In [190]: print(classification_report(y_predicted_tf,y_test))
              precision    recall  f1-score   support

           0       1.00      0.97      0.99       105
           1       0.97      1.00      0.99       100

    accuracy                           0.99       205
   macro avg       0.99      0.99      0.99       205
weighted avg       0.99      0.99      0.99       205
```

Fig. 4. Classification report

Fig. 5. Experiment result

Fig. 5 demonstrations the accuracy result and the accuracy obtained is 98.8%.

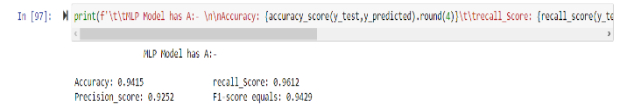

Fig. 6. Multilayer perceptron report

Fig. 6 demonstrations the classification report of multilayer perceptron. The report consists of accuracy, recall score, precision score and f1 score.

V. CONCLUSION

The early diagnosis of cardiovascular diseases can help high-risk patients in making decisions on lifestyle modifications and reduces problems, which can be a significant milestone in medicine. In addition, this model can be helpful to the medical practitioners at their clinic as a decision support system. Through this research, we have attempted to analyze the Artificial neural network technique and anticipate whether someone in particular, given different attributes and indications, will get the illness. The primary thought was to build a high-accuracy model and help in decision-making.

REFERENCES

[1] A. Shetty A, C. Naik. "Different Data Mining Approaches forPredicting Heart Disease", *International Journal of Innovative Research in Science,Engineering and Technology*, Vol. 5, Special Issue 9, May 2016

[2] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," *2011 Computing in Cardiology*, 2011, pp. 557-560.

[3] P. U. Anitha, BV P Kumar, Ch. Prudvini.,"Prediction of Heart Disease using Machine Learning Algorithm: Support Vector Machine", *International Journal of Advanced Research in Science, Communication and Technology* (IJARSCT),Volume 2, Issue 1, August 2022.

[4] Sonawane, Jayshril & Patil, Dharmaraj. (2015). Prediction of heart disease using multilayer perceptron neural network. 2014 *International Conference on Information Communication and Embedded Systems*, ICICES 2014. 10.1109/ICICES.2014. 7033860.

[5] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient *J. Computer Science Technology*, vol.6 2017, pp.457-466.

[6] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining

Techniques", *International Journal of Pure and Applied Mathematics*, 2018

[7] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", *Procedia Computer Science*, 2016, pp.962-969

[8] Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Network", *The International Institute of Science, Technology and Education*, vol.5, No.4, 2015, pp.7-11.

[9] Sairabi H.Mujawar, P.R.Devale, "Prediction of Heart Disease using Modified K-means and by using Naïve Bayes", *International Journal of Innovative research in Computer and Communication Engineering*, vol.3, October 2015, pp.10265-10273

[10] Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold cross-validation", *The Institute of Electrical and Electronics Engineers*, March 2017

[11] Mr.P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, "Heart Disease Prediction using ANN Algorithm in Data Mining", *International Journal of Computer Science and Mobile Computing*, April 2017, pp.168-172.

[12] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", I*nternational Journal of Computer Science and Engineering,* May 2018

[13] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques." *IEEE Access* 7 (2019): 81542-81554.

[14] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", *International Journal of Engineering Development and Research Development*, ISSN:2321-9939, 2017.

[15] Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. *SN COMPUT. SCI.* **1**, 345 (2020). https://doi.org/10.1007/s42979-020-00365-y

[16] Riddhi kasabe, Geetika Narang , "Cardio Vascular ailments prediction and analysis based on deep learning techniques", International Journal of Engineering and Applied Physics, vol.1, May 2021.

[17] S Monica, Sathees k ,"Analysis of cardiovascular disease prediction using data mining techniques",International Journal of Modern Computer Science,vol.4.February 2016,pp.55-58.

[18] Riddhi kasabe, Geetika Narang , "Cardio Vascular ailments prediction and analysis based on deep learning techniques", International Journal of Engineering and Applied Physics, vol.1, May 2021.