International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Heart Attack Prediction Using Machine Learning

[1]Harshitha K S

Department of Computer Science and Engineering
VTU CPGS, Mysuru,
Karnataka, India

Dr.G F Ali Ahammed, Associate professor
Department of Computer Science and Engineering
VTU CPGS, mysuru,
Karnataka, India

Abstract - In living things, the heart is important. The greatest percentage of deaths worldwide each year are caused by heart disease. detection and It's crucial to predict heart ailments, also referred to as cardiovascular disorders. Therefore, the prediction in question must be extremely precise and accurate. The accurate diagnosis of heart conditions can save lives. The dataset in question makes use of 13 key attributes for analysis. The dataset contains a few meaningless values and features, which have been dealt with using normalisation. Algorithms for machine learning are utilised in prediction and diagnosis. K-Nearest Neighbour (KNN), Support Vector Classifier, Logistic Regression, Naive Bayes, Random Forest Classifier, etc. are some of these algorithms. With the aid of an accuracy and confusion matrix, many findings are confirmed. We got the most accurate results.

Keywords – Heart Disease Prediction, Logistic Regression

1. Introduction

One of the most important human organs is the heart. therefore, the heart's care and upkeep are crucial. Theterm "heart disease" refers to a number of different illnesses that harm the heart. According to estimates bythe World Health Organisation, heart disease or cardiovascular disorders have been the largest cause of death over the past ten years. According to theircalculations, around 17.9 million people die each year as a result of cardiovascular disease. They are primarily cerebral stroke and coronary artery disease. Early detection and diagnosis of heart conditions are crucial. Heart disease risk is increased by a number of harmful behaviours. High cholesterol, obesity, hypertension, smoking, and other factors may be among them. Symptoms displayed by individuals can vary. They frequently experience chest discomfort, back pain, jaw pain, neck pain, and other types of pain.

Even though heart disease has recently been the leading cause of death worldwide, it may be properly managed and controlled. The accuracy in this case depends on when the sickness was really detected. Medical specialists have produced a significant collection of records that arein this available for study and extraction. Data mining techniques can be used to extract data, which can then be efficiently handled by Machine Learning (ML) algorithms for further use in diagnosis, detection, and prediction. We can analyse the given data by using machine learning to find hidden discrete patterns in massive datasets. The dataset was subjected to many algorithms, and the maximum accuracy of 86.89% was obtained using Logistic Regression and Naive Bayes.

2.Literature Survey

This section reviews the research done to date in this topic as well as the previously applied models for heartdisease prediction.

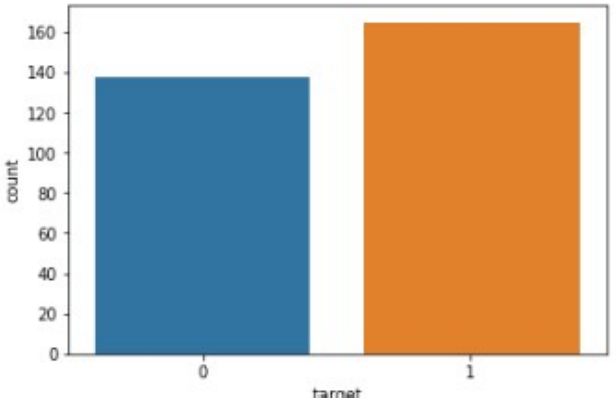
Paper[1] presents a model based on supervised learning techniques such as Naive Bayes, decision trees, K-nearest neighbours, and Random Forest algorithm for theprediction of various heart disease-related variables. The KNN method was used in this research study to get the maximum accuracy. To forecast cardiovascular illnesses, the paper[2] combines deep learning and machine learning techniques. The implementation of ML models for algorithms like Random Forest, KNN, SVM, LR, and XGBoost. Three dense layers were used in the deep learning architecture. All ML algorithms, it was found, performed better in predicting cardiac disorders. Paper[3] also discusses a comparison of SVM, LR, KNN, and Decision Tree, four alternative machine learning algorithms, for prediction. various cardiac conditions.SVM and KNN algorithms each achieved the best accuracy in this case. The performance of severalmachine learning algorithms is analysed in the paper[4], which gives a comparison study. The testing results showed that, when compared to other ML algorithms used, the Random Forest approach had the best accuracy (90.16%).

A thorough analysis of the prior research in this area has been effectively presented in paper [5]. Additionally, a ata classification system for the prediction and classification of cardiovascular disorders has been given study employing the DT, KNN, KMeans, and ADABoost algorithms. An artificial neural network algorithm was created in paper[6] to categorise heart disease based on specific clinical characteristics and traits. Nearly 80% accuracy was attained in this study. The paper's proposed model[7] makes use of voting using a weighted combination of the methods for Logistic Regression, Random Forest, K Nearest Neighbour,Gaussian Naive Bayes, and Artificial Neural Networks with ReLU function. To arrange the dataset, the proposed work in

Neighbour algorithm, and Logistic Regression algorithm. The ensemble model's accuracy with logistic regression is 95.06%, whereas its accuracy without it is 98.77%.

3. DATASET DESCRIPTION

The collecting of pertinent data, along with its processing and cleaning, was the first and most important phase in this project. The Heart Disease prediction UCI dataset [9] was the source of the data used in this study. It provides information about a number of characteristics of a heart disease patient. It has various components that influence a person's cardiac issues both alone and together. 14 attributes and 303 items make up this dataset. 165 items in the dataset's 303 entries were for male patients, whereas 138 entries were for female patients. This is depicted in Fig. 1 below. '0' signifies the number of females and '1' the number of males.



]

From the aforementioned list of features, just a few critical ones were picked for the model's training andtesting.
After that, the data was divided in half, 80:20, for training and testing the model.

1. METHODOLOGY

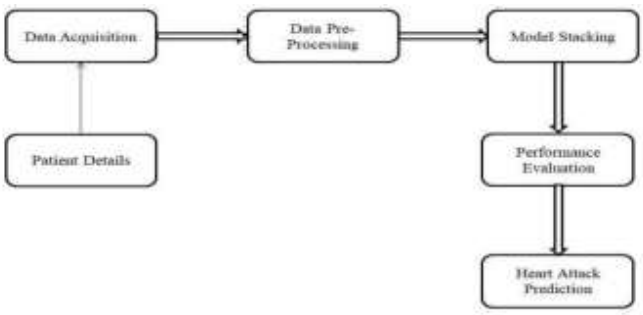


Fig 1. Methodology of proposed system.

The proposed, as seen in Fig. 1 above, consists of numerous components.

3.1 Data Acquisition:

Data acquisition is the process of computing actual physical conditions and converting the results into controllable numerical numbers for computers. Data

Pre-Processing:

It is necessary to first group the raw data before using it in a machine learning model. They are unable to access the data that must be prepared and cleaned when working on this project, which is necessary for the creation of machine learning models. Access to clean, ready data is not always possible. because you don't always havecleaned-up, prepared data at your disposal. Before anyone may use pre-processing services, data needs to be cleanedup and unnecessary data deleted.

Model Stacking:

Model stacking is the procedure of gathering every regression classification model that may be applied to two layer estimators. On the test data set, create the first layer as follows:

To forecast the results, one uses baseline models. They obtain new output from the layer two repressors or meta-classifier by using the input as the baseline model prediction.

The techniques used in this model are Deterministic Tree, Logistic Regression, Navie Bayes, KNN, Random forest, XGBoost, and SVM.

Logistic Regression

Logistic regression is a type of statistical model used in classification and prediction analysis because the result of a probability between the dependent variable swings between 1 and 0. It establishes the probability that an event will occur based on an independent variable and the available data set. The odds, which in this regression represent the likelihood of success vs. the likelihood of failure, are applied using the logit transformation. It's referred to as log odds.

The logistic function is of the form:

$$p(x)=\{1/1+e^{-(x-mu)/s}\}$$

where p(mu)=1/22 is the midpoint of the curve, and s and are location parameters.

K-Nearest Neighbor Classifier

A fundamental machine learning technique is utilized to forecast utilizing the previous instances and the new data in the K-Nearest Neighbor supervised learning approach. The new case is placed in the group that is shared by the available groups. As a result, current data from a properly

applied category that can be quickly classified using the k nearest classifier is obtained. To group new data using existing data, which the KNN algorithm completely replicates, this is done.

The Euclidean distance formula is used to determine the separation between the data points.

$$\text{A and B} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \qquad (1)$$

Decision Tree

In addition to utility, event outcomes, and resource costs, a d ecision tree is a supporting tool that can be utilised as a tree c omparable to decision making models and their workable ou tputs. A decision tree is one of the ways to illustrate a decisio n tree algorithm that can have conditional control statements
.

$$\text{Entropy: } H(S) = -\sum_{i=1}^{n} p_i(S) * log_2 p_i(S)$$

Information Gain:

$$IG(S, A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{S} H(S_v)$$

Fig 2. Decision Tree Formula.

3.4 Performance Evaluation: One of the most crucial aspects of the machine learning proc ess is performance evaluation, which requires meticulous ex ecution. Dataresampling, performance measurement, and r esults data with statistical significancearethe threemainsubtasks that are being evaluated.

3.5 Heart Attack Prediction: After completing the aforementioned steps, users receive th e prediction for their own input, and as a result, the project's a nticipated conclusion will be the prediction of an accuracy s core for a certain dataset and whether the patient should be di agnosed with a heart attack or not.

RESULT ANALYSIS

The major goal of this study is to lessen the probability that someone will have a heart attack. and make recommendations for the following actions. High accuracy rates are possible when using the Random Forest technique. They gave the example of the following data set:

Table 1. Data set

age	49	64	43	69
cp	1	3	2	0
trestbps	120	150	172	135
chol	239	219	283	233
fbs	0	1	0	1
thalach	178	163	174	114
exang	0	1	0	1
old peak	1.4	0.6	1.8	0.8
thal	1	2	1	2
target	0	1	0	1

The information in Table 1 can be used to calculate a person's risk of having a heart attack. Each statistic in thedata set is a result of the heart functions.

The four principles of the chest pain (cp) type, for instance. Angina symptoms, number one 2. An uncommon gina case
Nodule-free pain 3. 4. No symptoms Table 1 lists the characteristics that the statistics set depicts.

- trestbps- Level of plasma pressure at relaxing mode.
- chol- Serum cholesterol in mg/dl.
- fbs -

Plasma sugar levels on fasting (if>120mg/dl represented as 1 otherwise 0)

- restingecg- Results of electrocardiogram while at rest.
- exang- Angina induced by exercise (0-No, 1-Yes)
- old peak-

Exercise induced ST depression incomparison with state of rest.

Table-2: Data set with results.

age	49	64	43	69
cp	1	3	2	0
trestbps	120	150	172	135
chol	239	219	283	233
fbs	0	1	0	1
restecg	178	163	174	114
thalch	0	1	0	1
exang	1.4	0.6	1.8	0.8
oldpeak	1	2	1	2
slope	1	64	43	69
thal	49	3	2	0
target	0	1	0	1
Heart attack	NO	YES	NO	YES

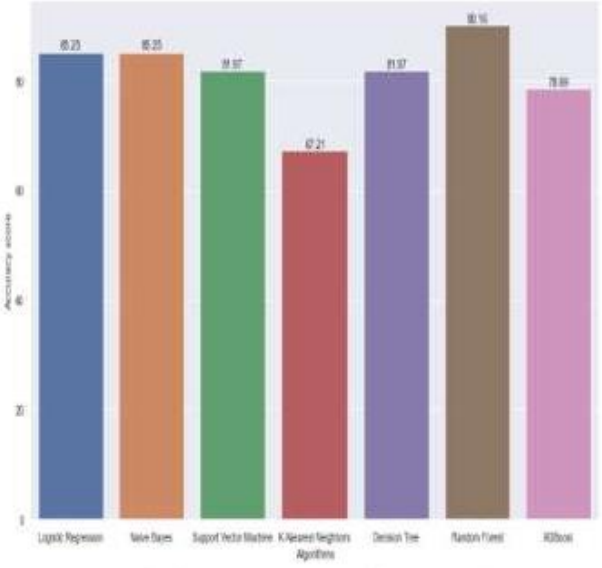


Fig-5: Graph of comparing accuracy of different algorithms

Table-3: Accuracy Results.

Algorithms	Accuracy
Logistic Regression	85.25
Naïve Bayes	85.25
SVM	81.97
KNN	67.21
Decision Tree	81.97
Random Forest	90.16
XGBoost	78.69

2. Figure demonstrates that compared to other methods, the application developed utilizing the Random Forest technique has a higher level of precision.

4.CONCLUSIONS

For classification and regression applications, the Random Forest approach provides an efficient collaborative learning system. The technique initially creates N decision trees before returning the session that reflects the average of all decision tree outputs. Thus, precise and effective early prediction is accomplished. Monitoring medical information, especially those related to the heart, can aid in preventing heart attacks and other major heart disorders by seeing them early.

It can be quite difficult to foresee heart attacks in the modern world. By entering the report standards, a patient or user who is unable to reach a surgeon can use this application to predict a heart attack. and the choice of whether to seek medical attention.

FUTURE SCOPE:

Future updates to this programme might add new capabilities including the capability to alert the patient's family as a whole in the event of a probable heart attack. The information must also be given to the neighbourhood hospital.
Online interactions amongst medical professionals are an additional choice.

It's crucial to remember that many different industries, like radiology, bioinformatics, and the study of medical imaging, use ML applications built on a variety of efficient algorithms.

REFERENCES

[1] Azizkhan F Pathan, Chetana Prakash, Attention-based position-aware framework for aspect-based opinion mining using bidirectional long short-term memory, Journal of King Saud University - Computer and Information Sciences, 2021, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.09.011.

[2] Jayaprakash, S., Nagarajan, M.D., Prado, R.P.D., Subramanian, S. and Divakarachari, P.B., 2021. A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. *Energies*, *14*(17), p.5322.

[3] Mr.Santhana Krishnan.J, Dr.Geetha.S," Prediction of Heart Disease Using Machine Learning Algorithms",2019 1st International Conference on Innovations in Information andCommunicationTechnology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.

[4] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878,Volume-8, Issue-1S4, June 2019.

[5] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja," Heart disease prediction using machine learning techniques: a survey", International Journal of Engineering & Technology, 7 (2.8) (2018)

[6] Hazra, A., Mandal, S., Gupta, A. and Mukherjee, " A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology , 2017.

[7] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science,Engineering and Information Technology ,IJSRCSEIT 2019.

[8] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation(ICCCA), 2018.

[9] Amandeep Kaur and Jyoti Arora,"Heart Diseases Prediction using Data Mining Techniques: A survey" International Journal of Advanced Research in Computer Science , IJARCS 2015-2019.

[10] Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.

[11] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[12] S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018.

[13] Abhay Kishore1, Ajay Kumar2, Karan Singh3, Maninder Punia4, Yogita Hambir5," Heart Attack Prediction Using Deep Learning",International Research Journal of Engineering and Technology(IRJET), Volume: 05 Issue: 04 | Apr-2018.

[14] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.

[15] A. S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.