Heart Attack Prediction Using Machine Learning Algorithms

Manjula P
Department of Computer Science and Engineering
Jain Institute of Technology
Davanagere, India

Aravind U R
Department of Computer Science and Engineering
Jain Institute of Technology
Davanagere, India

Darshan M V
Department of Computer Science and Engineering
Jain Institute of Technology
Davanagere, India

Halaswamy M H
Department of Computer Science and Engineering
Jain Institute of Technology
Davanagere, India

Hemanth E
Department of Computer Science and Engineering
Jain Institute of Technology
Davanagere, India

***Abstract* — The heart is one of the vital organ in the humans. It aids in purification and circulation of blood throughout body. Heart Attack is the one, that causes the death in worldwide. Some symptoms included chest pain, a quicker heartbeat, and difficulty breathing. This information was examined on a regular basis. A general overview of heart attacks and current techniques was established in this paper. Moreover, a review of the significant machine learning techniques for heart attack prediction accessible in the literature is briefly given. Decision Tree, Logistic Regression, SVM, Naive Bayes, Random Forest, KNN, and XGBoost Classifier are the machine learning methods mentioned. On the basis of the braced of features, the algorithms are compared.**

Keywords—Machine Learning, Heart Attack Prediction, KNN, Logistic Regression,SVM,Decision Tree, Naïve Bayes, Random Forest,XGBoost.

I. INTRODUCTION

Heart is the expansive and plays an critical part in human body. In order to maintain that to take care of heart is necessary for every individual. Many of diseases are related together with the heart so that's why the prediction of heart attack is required and for this a study is made in this field comparatively, And also at present most of patients are dying due to the heart attack and is recognized at the last stage. This is happening because of lack of instruments inorder to provide better accuracy in efficient way by using algorithms for the prediction of heart attack.

The difficulties facing by the healthcare industries in today's reality is initial prediction of heart attack whether the individual is predicted or not. The statistics of health history is complex and the statistics in physical world be incompatible, incapable and inconsistent able.

Researchers strained to train prototype, which is accomplished in predicting of the heart attack in the initial period, and they are not capable to build a suitable prototype. All the structure has its own advantages and disadvantages. Machine learning systems were educated to understand how to process and utilise data. "Machine intelligence" refers to the confluence of both technologies. According to the explanation of machine learning, it acquires from the usual occurrences and usual thing. In this, everybody are using natural constraints like analysis records, such as cholesterol, blood pressure, sex, age, and so on, and established on these, you all compared the accuracy of algorithms, such as Decision Tree, Logistic Regression, K-Nearest Neighbour, SVM, Naive Bayes, Random forest, and XGBoost. In this exploration, they calculate the accuracy of seven different machine learning algorithms thus regulates which one is the top based on the outcomes[1] [2].

During the testing stage, unevenly 80% accuracy on the testing set is attained. It takes time to place data from previous records to practical use. Low rate of accuracy. As a result, they've used the Random Forest method to produce more outcomes that are accurate in less time.

II. LITERATURE REVIEW

Many investigations have been undertaken in medical midpoint on heart attack prediction systems that use various Machine Learning algorithms.

Santhana Krishnan. J[3] Heart attack prediction using a machine learning algorithm This paper used classification methods to predict heart attacks in patients. This document provides efficient details about heart attacks, containing verity, frequent types, and risk components. Heart attack is predicted in this system using Naive Bayes and Decision Tree. Decision Trees, such as ID3 Algorithms, and Naive Bayes Techniques, are the most repeatedly used techniques for prediction. Among these Naïve Bayes classifier achieves more accuracy.

Avinash Golande [4] suggested Heart Attack Prediction using Effective Machine Learning Techniques, in that experts use a few data mining policies to help people in charge or doctors identify between heart attacks. Decision trees, k-nearest neighbor, and Naive Bayes are common methodologies. The upcoming portion clearly describes the methods that were used in the test. Among these Decision trees achieves more accuracy.

V.V. Ramalingam[5] suggested Heart attack prediction using machine learning approaches, in that Machine Learning algorithms and methods were utilized to a variety

of medical raw data to mechanize the interpretation of huge and complicated data. This study inspects the presentation of several models built on such approaches and methodologies. Decision trees, Support vector machines, Naive Bayes, K-nearest neighbor, Random Forest, and cooperative copies stay common among scientists, and they need to stay to a variety of medicinal statistical records to mechanize the study of large and composite records. Scientists favor copies established on supervised learning techniques including Support vector machines, Naive Bayes, K-nearest neighbor, Random Forest, Decision trees and cooperative models. Methods to help the health-care business and specialists in the study of heart-related disorders.

III. PROPOSED SYSTEM

The system is proposed with the common heart attack prediction based on some parameters.

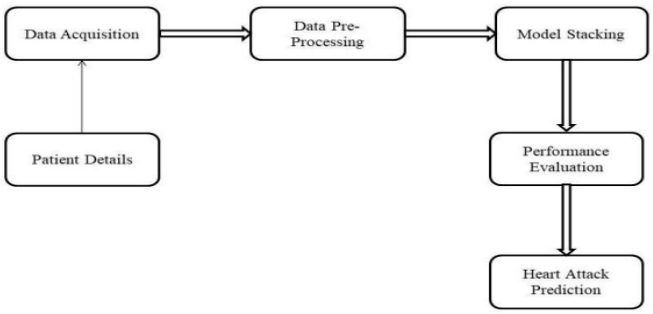

Fig 1. Methodology of proposed system.

The proposed contains different blocks of as show in the fig above i.e fig 1.

3.1 Data Acquisition:

The data acquisition is the process that calculate real world physical circumstances and convert the data into numerical values that computer can control is called data acquisition.

3.2 Data Pre-Processing:

Putting together raw data for use in a machine learning model is called data pre processing. It plays a very important role in creating machine learning model while working on this project they cant have access to clean and prepare the data. When do not always have access to clean and prepare data. When you don't always have access to data that has been cleaned and prepared. Before doing any operation on data, it must first be cleaned and unwanted data must be deleted so that everyone use pre-processing service.

3.3 Model Stacking:

Model stacking is a process of collecting all the regression classification models that can be used in two layer estimaters. To make the first layer on the test data set the base line models are used to forecast the outcomes. From the repressors or meta –classifier in layer two they use the input as base line model prediction and generate new output.

The algorithims used in this model are Desion tree, Logistic Regression, Navie Bayes, KNN, Random forest, XGBoost, SVM.

A. Logistic Regression

Logistic regression is the form of statistical model that is used to for predictive analysis and it is used for classification it estimates the chances of an occurring an event based on independent variable on the given data set since the output of a probability between the dependent variable leaps between 1 and 0. In this regression the odds are applied from the logit transformation that is the chances of success/chances of failure . It is known as log odds.

The logistic function is of the form:

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

where s is a scale parameter and μ is a location parameter (the curve's midpoint, where p(mu)=1/22)

B. K-Nearest Neighbor Classifier

K-Nearest Neighbor is a supervised learning method in which everyone predict using basic machine learning algorithm presumes equivalence between the available cases and the new data and set the fresh case to the grouping that is common to the obtainable groupings. In KNN algorithim copies the current all the current data and groups and new data based on the similar data it means that current data appears that can be simply categorize from a well applied category from using k nearest classifier.

Euclidean distance formula is used to find the interval between the data points.

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

C. Decision Tree

A decision tree is a supporting tools that can be used as a tree similar to decision making models and their viable consequences and it includes utility, event outcomes resource costs. Decision tree is one of the way to show an decision tree algorithm that can contain control statements that are condition.

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

Information Gain:

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

Fig 2. Decision Tree Formula.

D. Naïve Bayes

Navie Bayes algorithim is used to find classification problems and it is one of the machine learning technique and it is derived from bias theorem. It's a simplest potential machine learning algoritim used in numerous industries to find applications.

$$\mathrm{P(A|B)} = \frac{P(B|A) * P(A)}{\mathrm{P(B)}}$$

Fig 3. Naïve Bayes Formula.

E. SVM(Support Vector Machine)

The Support Vector Machine is one of the best exception Supervised Learning model (SVM). The main objective of the SVM model remains to discover the finest line or resolution borderline for finding hyperplane in N-

dimensional space into different modules so that extra or additional statistics facts can be readily places in the exact type in the upcoming.

F. Random Forest

Random forest is used as a supervised machine learning technique and it is well defines model. In this model everyone use this technique for both regression and classification problem. Ensemble learning idea is used in random forest model. It is one of the classifier that accommodate the number various subset of given data set in decision tree and finalise the predicted accuracy of the given data set based on the given label.

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

Fig 4.Random Forest Equation

G. XGBoost

XGBoost is adaptable and highly accurate executable of gradient boosting that pushes the bound of computing power for boosted tree algorithms, it is mainly used for fast computing and also energize performance.

$$\text{obj}(\theta) = \sum_{\text{i}}^{\text{n}} \text{l}\,(y_i - \hat{y_i}) + \sum_{j=1}^{j} \Omega(f_i)$$

Fig 5.XGBoost Formula

3.4 Performance Evaluation: Performance evaluation is one of the most important feature of machine learning process it needs to be carefully conducted. The evaluation of 3 main sub tasks are data resampling, performance measurement and results data having statistical significance.

3.5 Heart Attack Prediction: After accomplishing all the above procedure they get the prediction for own input and hence the anticipated outcome of these project will be prediction of an accuracy score of particular dataset and whether the patient should be diagnosed with heart attack or not.

IV. Result Analysis

This project's major goal is to regulate whether or not a person is predicted with heart attack. And make recommendations about how to proceed. It is possible to get excellent accuracy rates with the Random Forest algorithm. The following is the data set that they used: (sample):

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

Table-1 has adequate information to determine whether a person is predicted from heart attack or not. Each feature in the statistics set is the consequence of cardiac functionalities.

For eg, cp- The type of chest pain categorized into 4 tenets. (1. Characteristic angina 2.Uncharacteristic angina 3.Non-angical pain 4. Asymptomatic) The statistics set represents features are listed in table-1

- 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

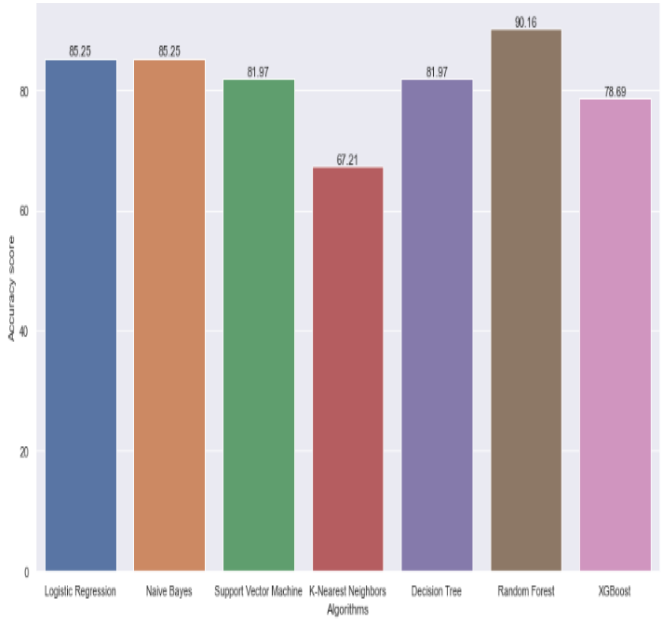

Fig-5: Graph of comparing accuracy of different algorithms

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

According to fig. 5, the application constructed utilising the Random forest algorithm has a higher precision level than supplementary algorithms.

V. CONCLUSION AND FUTURE SCOPE

The Random Forest algorithm is a powerful collaborative learning system for regression and classification procedures. The procedure generates N Decision trees and returns the session that is the middling of all the Decision trees outputs. As a result, early-stage prediction accuracy is efficiently accomplished. The handling of healthcare records, specifically records connected to the heart, will aid in the early recognition of heart attacks or aberrant heart conditions, saving lives in the long run.

In today's world, predicting heart attacks is a huge challenge. If the patient or user is unable to contact a surgeon, he or she can utilize this application to anticipate a heart attack simply by putting the report standards. And can decide whether or not towards seek medical advice.

FUTURE SCOPE:

This application can be enhanced in the future by addition of new functionalities, such as directing a message to all of the patient's family members if a heart attack is predicted. The info must also be forwarded to local hospital. Additional option that would be provided is online doctor discussion with the other doctor.

It's worth noting that ML applications based on numerous efficient algorithms are used not only in the field of heart attack prediction and analysis, but similarly in radiology, bioinformatics, and medicinal imaging analysis.

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) 684-687.

[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.

[16] V. Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.