Critical Analysis of Heart Attack Disease Prediction Using Supervised Learning Algorithms

Dr .S. Jansi

Assistant Professor, Department of Computer Applications
Madanapalle Institute of Technology & Science,
Angallu, Madanapalle-517326
drjansis@mits.ac.in

K. Jaswanth

Department of Computer Applications,
Madanapalle Institute of Technology & Science,
Angallu, Madanapalle-517326
21691F0049@mits.ac.in

ABSTRACT

One of the rising areas of artificial intelligence is machine learning, which has a wide range of uses. It is a technology whose fields of use include data and artificial intelligence. The basic goal of developing machine learning algorithms is to build a model that comprehends, evaluates, and aids in prediction of the given data. Methods of machine learning can be used in many different fields. The Covid outbreak-related lockdown recently saw a sharp rise in internet sales.

Heart attacks happen when anything prevents your heart from getting the oxygen it needs from the blood flowing to it. Recent studies demonstrate a significant increase in the number of heart attack-related deaths, with sudden heart attacks accounting for the majority of these deaths.

The fundamental goal of this work is to foretell heart attacks before they occur so that we can lower the risk of death and lead happy lives. The goal is to create an algorithm that can foresee cardiac attacks. of this endeavour. that uses machine learning algorithm to find the underlying causes of heart attacks.

KEYWORDS

Logistic Regression, Heart Attack, Random Forest Algorithm, K-Nearest Neighbour Algorithm, Decision Tree Algorithm, Support Vector Machine Algorithm, Disease Prediction, Machine Learning, Data Analysis.

1.INTRODUCTION

numerous risk factors, including irregular heartbeat attacks, high blood pressure, diabetes, and high cholesterol. tempo, tension, etc. The patient's symptoms and signs will mostly be used to diagnose heart disease. Heart conditions are complex in nature. The illness must be effectively treated in order to prevent mortality; else, the person's heart may be harmed [1].

We use logistic regression, support vector machines, K-neighbor classifiers, and decision tree machine learning methods. The main difficulties are identifying and detecting heart diseases. While there are several methods available to help They either cost a lot of money or are ineffective at predicting the possibility of heart disease. The mortality rate can be lowered by early detection. and effects of heart illness.

Nevertheless, daily patient monitoring is still not realistic. Because there is a lot of data in the current world, we will utilise a variety using techniques from machine learning for analysing the data for hidden effects [2]. You've just started working as a data scientist in a hospital where more and more patients are showing up with various heart symptoms. A cardiologist will assess vital signs and give you the information as a consequence so that you may perform data analysis and decide whether particular patients have cardiac illness. We would like to develop a machine learning algorithm to instruct our AI to learn from experience and advance. We want to classify people as either having or not having heart disease as a result.

2.RELATED WORK

A wide range of variables, including age, gender, pulse rate, and many more, can be utilised to determine heart disease. Data analysis aids in disease prognosis, improved diagnosis, symptom analysis, provision of appropriate medications, development of optimal treatment, price reduction, extension of existence, and reduction in the death rate of cardiac patients, among other medical tasks. It's crucial to use [3] methods to spot complex heart issues that pose a serious threat to people's lives. Due to a shortage of doctors and diagnostic tools that affect the medical treatment of cardiac patients, the diagnosis and management of these patients are currently very difficult. Early identification of heart disease is crucial for reducing heart-related disorders and safeguarding it from significant threats. Diagnostic methods that are intrusive are used a patient's medical history, an expert's report on their evaluation of their symptoms, and a physical laboratory report can all point to cardiac difficulties. Additionally, the diagnosis is inaccurate and delayed as a result of human interaction. For now, it is expensive, time-consuming, and computationally painful. [4].

Data analysis in the medical field aids in illness prediction, improved diagnosis, symptom analysis, providing the appropriate medications, improving the bar of care, lowering expenses, prolonging life expectancy, and reducing the death rate among cardiac patients. ECG (Electro Cardio Gramme) aids in the early diagnosis of irregular heartbeat and stroke by applying an ECG (Electro Cardio Gramme) to a patient's chest and monitoring their heartbeat. The occurrence of cardiac disease can be predicted by professionals with the aid of comprehensive clinical data. Blood arteries in the heart must function normally for human life to exist. Due to kidney failure, brain imbalance, and inactive hearts, inadequate blood flow can also cause impending mortality. Obesity, smoking, diabetes, blood pressure, cholesterol, inactivity, and a poor diet are some risk factors for heart disease [5].

Table 1: Heart Attack data analysis and Prediction Literature Survey is as shown in Table:1

S. No	Title	Proposed Work	Limitations
1.	Heart Disease Prediction using Evolutionary Rule Learning [6]	On the patient's dataset, they used frequent pattern growth association mining. This will assist (help) in reducing the amount of services and demonstrate how the vast majority of regulations help in the best cardiovascular illness prognosis.	Direct information (data) extraction from electronic records cannot be done properly.
2.	Improved Heart	The suggested learning technique can	less effective than the

	Disease Detection [7]	improve medical staff members' ability to detect heart failure in patients.	current system
3.	Using Heart Disease to Predict Data analytics and machine learning Approach [8]	To determine their own personal forecast of heart disease, the user inserts their specific medical facts into the suggested method. The algorithm will decide whether cardiac disease is likely to be present.	The creation of a prediction model has only sometimes been successful.
4.	Heart attack prediction system [9]	The suggested learning system serves as a decision support system and will help doctors make diagnoses.	An earlier stage does not allow for detection.
5.	Chronic Disease Prediction by mining the data. [10]	Through data mining, efficient methods for predicting chronic diseases have been developed. historical health records	It takes time to employ diverse acquired data in practise.

2.1 Random Forest Algorithm

Classification and regression issues are dealt with using the Random Wooded Area Method, a well-known supervised machine learning algorithm. A woods is made up of multiple different species of bushes, and the more diverse the bushes are, the healthier the woodland will likely be. Similar to this, as the range of bushes increases, a Random forest algorithm gains accuracy and capacity to solve problems. A classifier known as Random forest utilises A few of decision nodes concerning various subsets of the input data to improve the dataset's projected utmost accuracy. relies solely on aggregate learning research, which mixes different classifiers for dealing with tough challenges and enhance model functionality. algorithm for random forests of trees is as shown below Fig:1
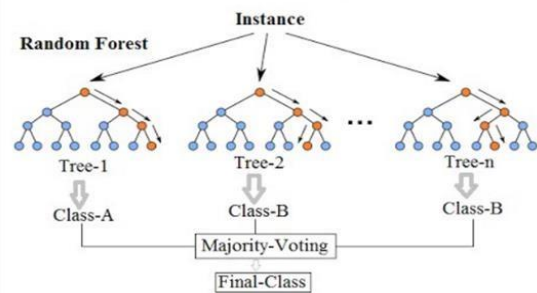
 Fig 1: Random Forest Algorithm

2.2 Logistic Regression Algorithm

It falls under the domain of supervised learning and is one of the most used ML algorithms. Predicting results for categorical variables that are dependent and resolving classification puzzles are some of its main applications. The development of a classifier can help stop fraudulent credit card transactions. Logistic regression can be used to predict the outcome from a binary dependent variable. The conclusion must therefore be neither discrete or categorical in structure. Instead of precise integers between 0 and 1, It delivers probabilistic values that vary from 0 to 1. True, False, 0 or 1, as well as Yes or No, are additional outcomes that can occur. A crucial machine learning methodology is logistic regression because it can classify and assign probability to new data. Below is a graph of logistic regression. Fig:2
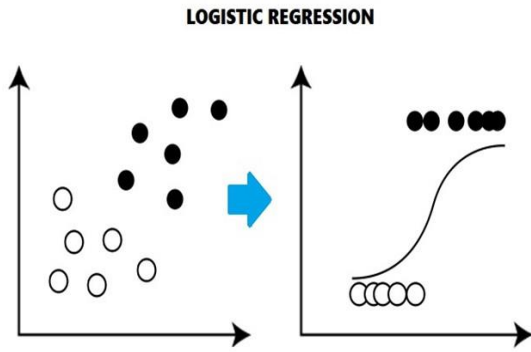
LOGISTIC REGRESSION

Fig 2: Logistic Regression

Limitations

Logistic regression cannot forecast a continuous outcome. The dependent (predicted) variable and the predictor (independent) variables are believed to have a linear relationship in a logistic regression. Logistic regression may produce erroneous findings if the sample size is too small.

2.3 K-Nearest Neighbour Algorithm

A supervised class strategy is the k-nearest Neighbours approach. It is classified entirely based on who is the object's nearest neighbour. It uses an entirely instance-based method of learning. A property's distance from another one is measured using the Euclidean formula neighbours. It makes use of a number of defined factors and makes a clear suggestion using them. k-NN may be used to fill within the blanks once the records are taken care of according to how comparable they may be. Following the completion of the lacking values, the facts set is put to some of prediction techniques. Accuracy can be improved by using combining these algorithms in unique approaches. k-Nearest Neighbour is as
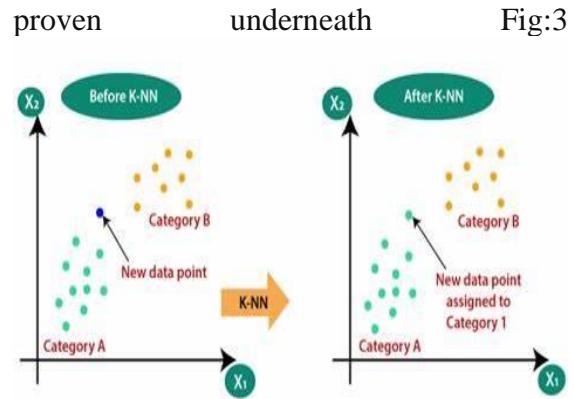
proven   underneath   Fig:3

Fig 3: K-Nearest Neighbour

2.4 Decision Tree

both categorical and numerical facts can be categorised the usage of the decision tree technique. decision trees are used to create tree-like systems. A easy and nicely-liked approach for dealing with medical datasets is a choice tree. A tree-fashioned graph's facts is simple to enforce and determine. The have a look at of the choice tree model is based on three nodes. Root node: The principal node on which the relaxation of the community is constructed. The inner node manages numerous properties. Leaf node: display the outcomes of every take a look at. choice Tree is as shown beneath Fig:4
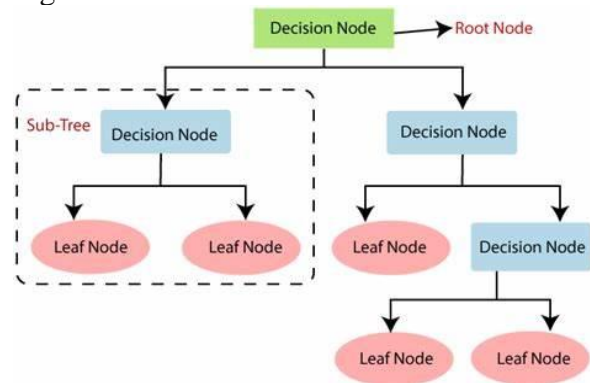

Fig 4: Decision Tree

2.5 Support Vector Machine

one of the most popular supervised mastering strategies is called the support Vector gadget, or SVM, and Regression and type issues are resolved with it. additionally, it is utilised, although, when gadget studying class affords a undertaking.

The SVM approach appears for the most useful selection boundary or line which can divide n-dimensional area into instructions. which will quickly categorise fresh facts factors inside the future. The name of this satisfactory-case choice boundary is a hyperplane. SVM is used to choose the hyperplane's maximum severe factors and vectors. The SVM method is built on the inspiration of assist vectors, which are utilised for representing those difficult instances. help Vector gadget is as proven under Fig:5
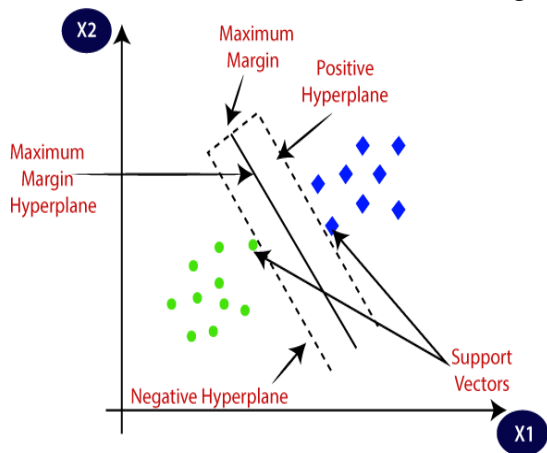

Fig 5: Support Vector Machine

3.PROPOSED WORK

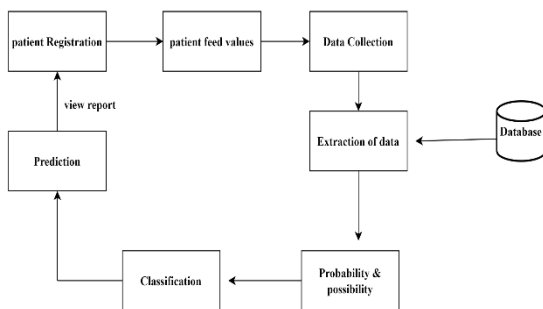Design of Heart Attack Data Analysis and Prediction is as shown below Fig:6

Fig 6: Design of Heart Attack Data Analysis and Prediction

A potential system for heart attack data analysis and prediction could include the following steps:

Registration of patients: Details about the patient's personal and medical history are collected during registration. The next stage will come after this one.

Values for the patient at this stage: information about the causes and consequences of heart attack disease.

data gathering Compile relevant data on heart attack victims, including their racial and ethnic backgrounds, medical histories, lifestyle preferences, and test results (such as blood and ECG results).

To handle missing or inconsistent numbers and ensure that the data is in a format that can be used for analysis, data cleaning and pre-processing are required.

Engineering of Feature: Take relevant characteristics that can be utilised to predict heart attacks from the data. Calculating summary statistics like mean and standard deviation or creating new features based on data that is previously known (such age, blood pressure, or cholesterol levels) are two examples of how to do this.

selecting a model Based on the quantity, kind, and desired accuracy of the forecast data, select the right artificial intelligence algorithm for heart attack prediction. Common algorithms include random forests, random decision trees, and support vector algorithms, among others. selecting variables on the dataset, as is seen below. Fig:7
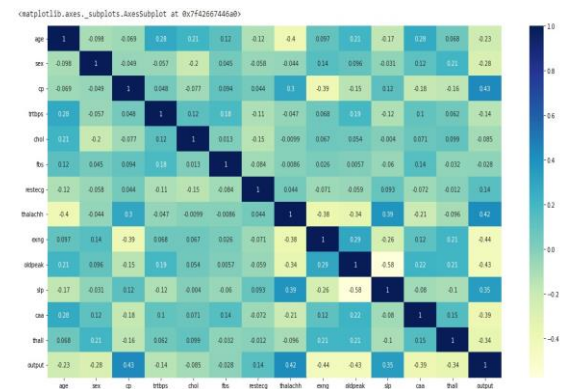

Fig 7: Selecting attributes from the dataset.

Model Training: To improve performance, adjust the model's parameters while practising the chosen machine learning algorithm on the pre-processed data.

We must balance the data to train the system more effectively. The dataset's balancing graph is displayed as follows. Fig:8
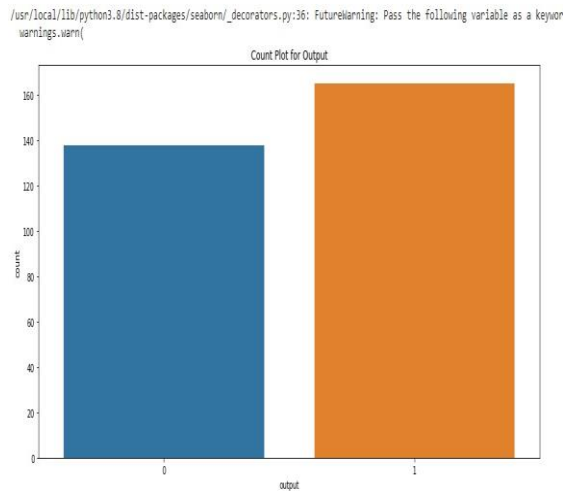

Fig 8: Balancing graph of dataset

Testing the precision, recall, and accuracy of the trained model using a different test dataset is known as model validation.

4.RESULTS AND DISCUSSION

We obtained the results by utilising all five algorithms: Random Forest, Logistic Regression, Decision Tree, SVM, KNN Algorithms. The accuracy of each algorithm's results is depicted in Fig. 9 below.

	ML	Accuracy
0	SVM_Accuracy	0.868852
1	Decision_Tree_Accuracy	0.819672
2	Random_Forest_Accuracy	0.819672
3	KNN_Accuracy	0.688525
4	Logistic_Regression_Accuracy	0.885246

Fig 9: Accuracy analysis

Support vector machine algorithm was used to resolve classification and regression problems, and the accuracy was 0.868852. Decision tree algorithm was used to compare data in the form of decision trees, and the accuracy was 0.819672. Random forest algorithm was based on decision trees and was applied to various subsets of the input data, and the accuracy was 0. We obtained a precision of 0. 688525.And the graph for the above mentioned data is depicted in Fig. 10 below.
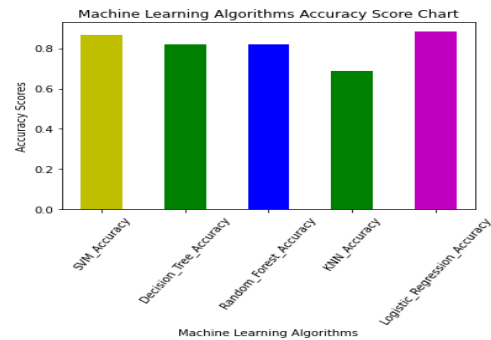

Fig 10: Area of Accuracy between the algorithms

were successful in determining the precise value of Heart Attack Data Analysis and Prediction, which was 0.8852459016393442 (88.52%). Compared to existing modules, the one that is being proposed provides more precise results and is suitable for a bigger dataset. The performance of the Logistic Regression technique will improve with more training data, but testing and application speed will remain slower. There may be a need for further pre-processing techniques.

5.CONCLUSION

A system that accurately forecasts cardiac disorders must be developed given the rise in heart disease-related mortality. The goal of the study was to identify the strongest machine learning (ML) method for heart attack identification. objective. The accuracy ratings of the Decision Tree,

Logistic Regression, Random Forest, KNN, and SVM techniques to predict heart attack are investigated in this study using the UCI machine learning repository dataset. According to the study's conclusions, the Random Forest algorithm, which has a The most accurate algorithm for predicting heart disease has an accuracy rating of 90.16 percent.

By creating a website application using the Random Forest approach and utilising a bigger dataset than the one employed here, the project can be enhanced in the future. I only take 14 important aspects into account in my study. I used methods to conduct data mining classification. as K-nearest Neighbour, SVM, decision trees, logistic regression, and random forests. Due to this study's limitations, more complex For the reason of increasing the precision of heart attack early prediction, and connected systems have to be employed. In future we can get above 90% of accuracy by overcoming or by using advanced supervised learning algorithms.

REFERENCES

[1] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[2] T. Nagamani, S. Logeswari, B. Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin," Design and Implementation Heart Disease Prediction Using Naives Bayesian", International Conference on Trends in Electronics and Information (ICOEI 2019).

[5] Theresa Princy R, J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies, Bangalore,2016.

[6] Nagaraj M Lutimath, Chethan C, Basavaraj S Pol.,'Prediction of Heart Disease using Machine Learning', International journal Of Recent Technology and Engineering,8, (2S10), pp 474-477, 2019.

[7] Sayali Ambekar, Rashmi Phalnikar, "Disease Risk Prediction by Using Convolutional Neural

[8] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, ―Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field, ‖ in Machine Learning Paradigms, 2019, pp. 71–99.

[9] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018

[10] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018