JOURNAL OF APPLIED BIOANALYSIS, November 2025 p. 912-919
http://doi.org/ 10.53555/jab.v11si12.2379 (ISSN 2405-710X)
Vol. 11, No. 12s

OPEN ACCESS

Critical Analysis of Heart Attack Disease Prediction Using Supervised Learning Algorithms

Dr.R.Kumar[1], Dr.N.Magendiran[2], R.Menagadevi[3], Dr.K.Baskar[4*], Dr.R.Prabu[5]

[1]*Associate Professor, Department of Biomedical Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, India, Email: pbcevp@gmail.com ORCID ID- 0000-0002-8274-9001*

[2]*Associate Professor, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, India, Email: maheindia76@gmail.com ORCID ID- 0009-0000-8236-5561*

[3]*Assistant Professor, Department of Information Technology, Sengunthar Engineering College (Autonomous), Tiruchengode, Tamilnadu, India, Email: meena.gopalan@gmail.com ORCID ID- 0009-0003-9585-1149*

[4*]*Associate Professor, Department of Biomedical Engineering, Vivekanandha College of Engineering for Women, Tiruchengode, Tamilnadu, India, Email: baskarresearch17@gmail.com ORCID ID- 0000-0001-9362-7547*

[5]*Associate Professor, Department of Biomedical Engineering, K.S.R College of Engineering, Tiruchengode, Tamilnadu, India, Email: prathinam11@gmail.com ORCID ID- 0000-0002-3963-3548*

***Corresponding Author:** Dr.K.Baskar*

baskarresearch17@gmail.com

ABSTRACT

Heart attack occurs when something stops the blood flow to your heart therefore it can't receive the oxygen it needs. Recent studies show that the number of deaths due to heart attacks have increased seriously, and majority of the deaths are mainly due to sudden heart attacks. This study's primary goal is to anticipate heart attacks before they happen, in order to lower the risk of death and lead happier lives. The goal of this work is to create a system that can forecast heart attacks and, using machine learning techniques, pinpoint their underlying causes. Identification and detection of heart disorders are the key challenges. While there are several methods available to help predict heart disease, they are either expensive or inefficient at determining the likelihood of heart disease. Early detection can lower the mortality rate and effects of heart illness. Nevertheless, daily patient monitoring is still not realistic. Since there is a lot of data in the current world, we will utilise a range of machine learning techniques to examine the data for hidden impacts. In the proposed study, we can clearly see that Logistic Regression outperformed KNN, SVM ,Decision Tree, and Random Forest Algorithms in terms of precision, f1 score ,recall, support and accuracy.

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

1.INTRODUCTION

Numerous risk factors, such as diabetes, high cholesterol, high blood pressure, an irregular pulse attack. Rate, stress, etc. Heart disease will primarily be identified based on the patient's signs and symptoms. Heart disorders are intricate in nature. In order to prevent mortality, the sickness must be treated effectively or else it may harm the person's heart [1].

We employ decision tree algorithms of machine learning, support vector machines, K-neighbors classifiers, and logistic regression.

Identification and detection of heart disorders are the key challenges. There are numerous tools that can assist forecast heart disease, but they are either expensive or ineffective at estimating the likelihood of heart disease. The mortality rate and consequences of cardiac disease can be reduced by early identification. However, it is still impractical to check

on patients every day. We will utilise several machine learning techniques to analyse the data for hidden effects because we have a lot of data in the modern world [2].

In a hospital where an increasing number of patients are arriving with varied cardiac complaints, you have just been hired as a data scientist. To do data analysis and determine if specific patients have cardiac disease, a cardiologist takes vital signs and provides you with the resulting information. In order to educate our AI to learn from experience and improve, we would like to create a machine learning algorithm. As a result, we would like to categorise patients as either having or not having cardiac disease.

2.RELATED WORK

Around the world, heart disease is regarded as the condition that kills people the fastest. The heart is unable to supply enough blood to the remaining bodily components to allow them to perform their usual activities, especially when this type of illness is present. Heart disease symptoms can include physical weakness, breathing problems, swelling feet, etc [3].

Techniques are necessary to spot complex heart conditions that pose a high danger of harming people's lives. Due to the lack of doctors and diagnostic tools that affect the treatment of cardiac patients, the diagnosis and treatment processes are currently quite difficult. Early detection of heart disease is crucial to reducing heart-related problems and safeguarding it from serious threats. Based on a patient's medical history, an expert's report on their study of their symptoms, and a physical laboratory report, invasive techniques are used to identify cardiac disorders. Furthermore, because of human intervention, it delays and results in inaccurate diagnosis. At the moment of assessment, it is time-consuming, computationally demanding, and expensive [4].

Many signs, including age, gender, pulse rate, and others, can be used to indicate the presence of heart disease. In the field of medicine, data analysis helps with disease prediction, improved diagnosis, symptom analysis, supplying the right medications, raising the standard of care, cutting costs, extending life expectancy, and lowering the death rate among cardiac patients. By placing an ECG (Electro Cardio Gram) on a patient's chest and monitoring their heartbeat, ECG (Electro Cardio Gram) aids in the early detection of irregular heartbeat and stroke. With the help of thorough clinical data, experts can forecast the occurrence of heart disease. The heart's blood vessels must operate properly for human life to exist. Insufficient blood flow can also result in imminent death due to heart inactivity, kidney failure, and brain imbalance. Obesity, smoking, diabetes, blood pressure, cholesterol, inactivity, and poor diet are a few of the risk factors that might result in 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. The number of services will be reduced as a result, and it will be demonstrated that the vast majority of the laws contribute to the most accurate prognosis of cardiac illness.	Direct information (data) extraction from electronic records cannot be done manually.
2.	Improved Heart Disease Detection [7]	The suggested learning approach can aid medical professionals in better heart failure identification.	Less efficiency than existing system
3.	Heart Disease Prediction using Machine Learning and Data Analytics Approach [8]	In the proposed system the user inputs its specific medical details to get the prediction of heart disease for that user. The algorithm will calculate the probability of presence of heart disease	The development of a predictive model has had only sporadic success.
4.	Heart attack prediction system [9]	The proposed learning system act as decision support system and will prove to be an aid for the physicians with the diagnosis	Detection is not possible at an earlier stage.
5.	Chronic Disease Prediction by mining the data. [10]	Effective mechanisms have been used for chronic disease prediction by mining the data containing historical health records	Practical use of various collected data is time consuming

2.1 Random Forest Algorithm

To address classification and regression problems, the Random Forest methodology, a very well-liked supervised machine learning method, is used. A forest is made up of numerous different species of trees, and the forest will be more vigorous the more trees there are. Similar to this, the number of trees in a Random Forest Algorithm increases the algorithm's accuracy and ability to solve problems. A classifier known as Random Forest employs numerous decision trees on various subsets of the input data to boost the predicted accuracy of the dataset. It is built on the concept of ensemble learning, which is the practise of combining different classifiers to solve a difficult problem and improve the performance of the model. Random Forest Algorithm is as shown below Fig:1

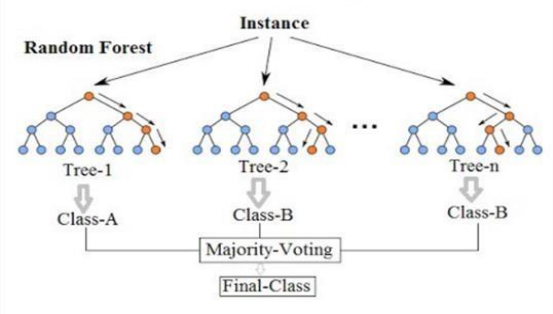


Fig 1: Random Forest Algorithm

2.2 Logistic Regression Algorithm

It is one of the most used ML algorithms and belongs to the category of Supervised learning. Its primary applications involve resolving classification problems and forecasting the results of categorical dependent variables. The creation of a classifier can assist in preventing credit card transaction fraud. Logistic regression can be used to forecast the outcome of a dependent variable that is categorical. The outcome must therefore be a discrete or categorical value. It provides probabilistic values that fall between 0 and 1 rather than the exact numbers between 0 and 1. Possible results include True, False, 0 or 1, as well as Yes or No. Since it can classify and assign probabilities to fresh data, logistic regression is an essential machine learning technique. Logistic Regression is as shown below Fig:2

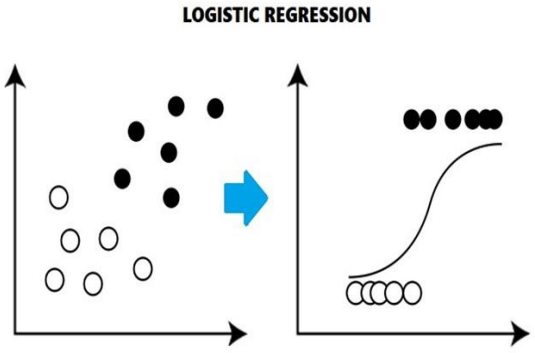


Fig 2: Logistic Regression

2.3 K-Nearest Neighbor (K-NN)

A supervised classification approach is the K-nearest neighbours' algorithm. Objects are categorised based on their closest neighbour. It is an instance-based learning approach. Euclidean distance is used to calculate how far apart a property is from its neighbours. It makes use of a set of named points and applies them to the marking of another point. The data are grouped based on how similar they are, and K-NN can be used to fill in the missing values. Several prediction approaches are applied to the data set after the missing values have been filled in. By combining these algorithms in different ways, it is feasible to improve accuracy. K-Nearest Neighbour is as shown below Fig:3

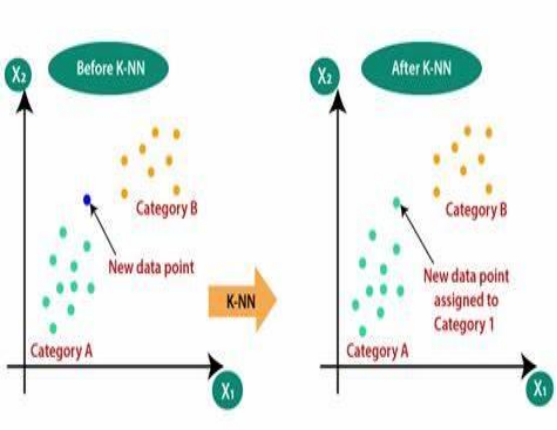


Fig 3: K-Nearest Neighbour

2.4 Decision Tree

A classification system known as a decision tree can be used with both category and numerical data. Tree-like structures are made using decision trees. A decision tree is a straightforward and popular tool for managing medical datasets. The data in a tree-shaped graph can be easily implemented and analysed. Three nodes serve as the basis for the decision tree model's analysis. Root node: The primary node, upon which all other nodes are founded. Various properties are handled by the interior node. Leaf node: Display each test's outcome. Decision Tree is as shown below Fig:4

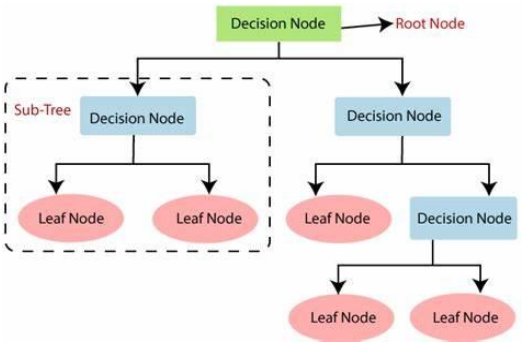


Fig 4: Decision Tree

2.5 Support Vector Machine

Support Vector Machine, or SVM, is one of the most used supervised learning approaches, and it is used to tackle problems with classification and regression. It is primarily utilized, nevertheless, when machine learning classification presents a challenge. In order to swiftly classify new data points in the future, the SVM approach aims to find the best line or decision boundary that may divide n-dimensional space into classes. A hyperplane is the name for this best-case decision boundary. The extreme vectors and points that make up the hyperplane are chosen using SVM. The SVM method is built on support vectors, which are utilised to represent these extreme situations. Support Vector Machine is as shown below Fig:5

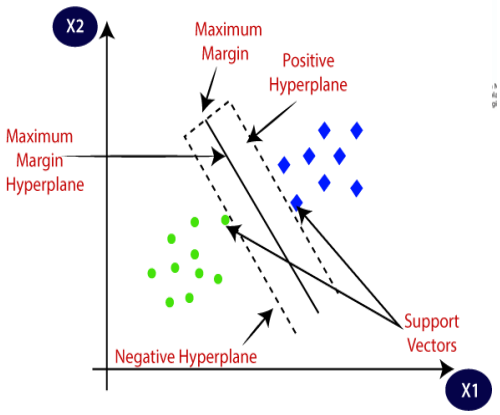


Fig 5: Support Vector Machine

3.PROPOSED WORK

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

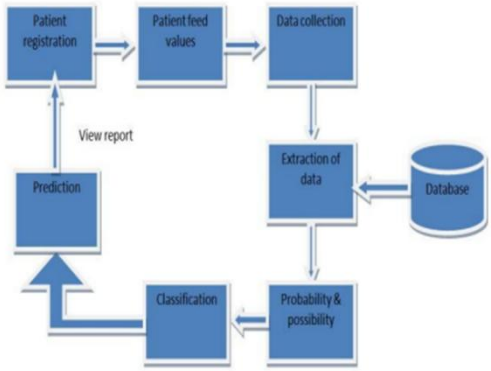


Fig 6: Design of Heart Attack Data Analysis and Prediction

The following steps could be part of a suggested system for heart attack data analysis and prediction:

Patient Registration: Patient registration process include their personal and health related details. This step will lead to proceed further process.

Patient feed values: this stage will provide information about the causes and effects of heart attack disease.

Data collection: Compile pertinent information about heart attack victims, including their demographics, medical histories, lifestyle choices, and test findings (such as ECG and blood tests).

Data cleaning and pre-processing is necessary to handle missing or inconsistent values and make sure the data is in a format that can be used for analysis.

Engineering of Feature: Extract from the data useful traits that can be used to forecast cardiac attacks. This can entail figuring out summary statistics like mean and standard deviation or developing new features based on already-existing information (such age, blood pressure, or cholesterol levels).

Model selection: Based on the volume, kind, and desired accuracy of the prediction data, select a suitable machine learning method for the purpose of predicting heart attacks. Decision trees, random forests, and support vector machines are a few examples of prevalent algorithms. Selecting attributes from the dataset is as shown below Fig:7

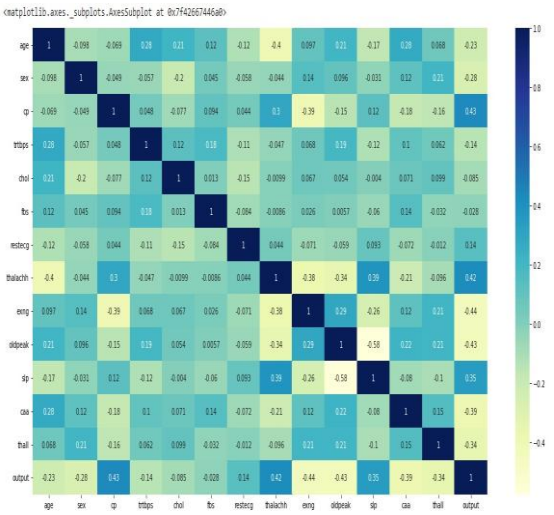


Fig 7: Selecting attributes from the dataset

Model Training: On the pre-processed data, practise the selected machine learning algorithm while modifying the model's parameters to enhance performance.

For Training the machine in better manner we need to balance the data. Balancing graph of dataset is as shown below Fig:8

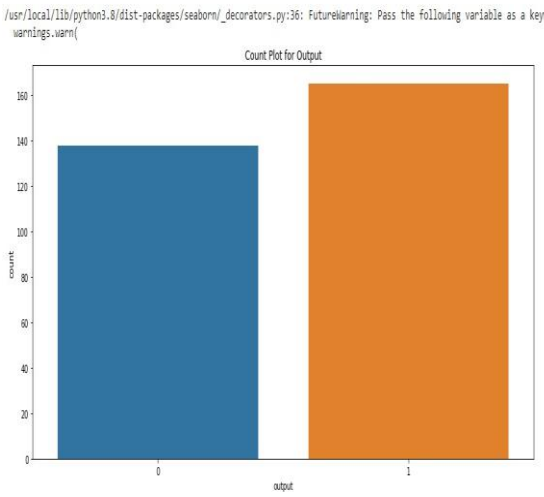


Fig 8: Balancing graph of dataset

Model validation involves testing the trained model's precision, recall, and accuracy on a different test dataset.

4.RESULTS AND DISCUSISION

By using all the five algorithms that are Logistic Regression, Random Forest, SVM, Decision Tree, KNN Algorithms we got the result of accuracy of each algorithm is as shown below Fig:9.

	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

By Using Support Vector Machine Algorithm we done the Classification and regression issues are resolved with the accuracy of 0.868852, for Decision_Tree_Algorithm in this we done the comparison by data in the form of decision trees in that we will have decision nodes and related leaf node of solution and the expected output with the accuracy of 0.819672, Random Forest Algorithm is used based on decision trees on various subsets of the input data and got the accuracy of 0.819672, for Logistic Regression we compared the dependent and independent variables and the final accuracy we got is0.885246, for KNN Objects are categorised based on their closest neighbour. We got the accuracy of 0.688525.And the graph for the above-mentioned data is as shown below Fig:10

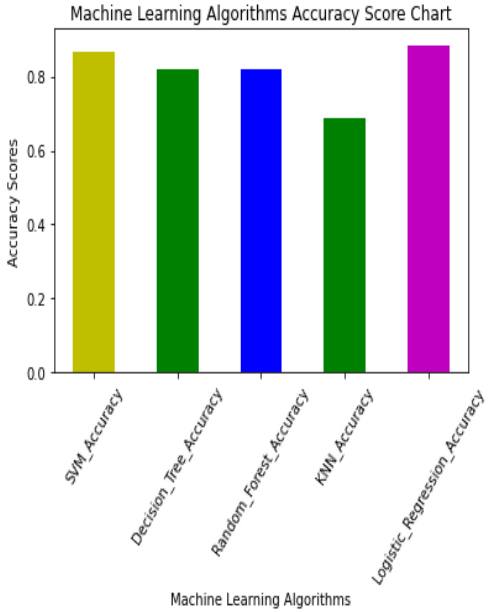


Fig 10: Area of Accuracy between the algorithms

Therefore, utilising a Logistic Regression technique with new improvements, we were able to obtain the accurate value of Heart Attack Data Analysis and Prediction, which was 0.8852459016393442 (88.52%). This proposed module offers more accurate findings and is appropriate for a larger dataset than existing modules. With lots of training data, the Logistic Regression technique will perform better, but testing and application speed will still be slower. Additional pre-processing methods might also be helpful.

5.CONCLUSION

It is crucial to create a system that can reliably anticipate cardiac diseases given the rise in heart disease-related mortality. Finding the best effective ML algorithm for heart disease detection was the study's main objective. The accuracy ratings of the Decision Tree, Logistic Regression, Random Forest, KNN, and SVM algorithms for predicting heart disease are examined in this study using the UCI machine learning repository dataset. The study's findings demonstrate that the Random Forest algorithm, which has an accuracy rating of 90.16%, is the best algorithm for predicting heart disease. By developing a web application based on the Random Forest algorithm and employing a bigger dataset than the one used here, the work can be improved in the future. I only take 14 important aspects into account in my study. I applied data mining classification techniques as K-nearest Neighbour, SVM, decision trees, logistic regression, and random forests. Due to this study's limitations, more complex and linked models must be used in order to improve the accuracy of heart disease early prediction.

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