Heart Disease using Machine learning algorithms

Yashraj Shinde
BTech CSE, E20CSE170
Bennett University
Greater Noida, India
ofyashraj@gmail.com

Riti Kushwaha
Assitant Professor
Bennett University
Greater Noida, India
ritikushwaha@bennett.edu.in

***Abstract*—This evaluation of the literature focuses on four studies that investigate the application of machine learning methods to the diagnosis and treatment of cardiovascular disease. Based on computed tomography (CT) angiography pictures, the first study built a model employing machine learning techniques to automatically predict coronary artery disease. The model demonstrated great sensitivity and specificity, highlighting its potential as a dependable and precise method for CAD detection. The second study created a CAD identification model using deep learning and random forest methods, and it had a 93.3% accuracy rate. This model demonstrated promise for the application of deep learning methods in CAD identification.**

In the third study, machine learning techniques were employed to look for signs of cardiac disease in electrocardiogram (ECG) signals. The promise of machine learning in ECG analysis can be seen by the model's high accuracy of 94.4% in recognising heart illness from ECG signals. The fourth study used machine learning methods to predict cardiac failure, and it did so with a high level of 90% accuracy. This study highlights the potential of machine learning for heart failure early identification and prevention.

***Index Terms*—Machine learning, Deep learning, random forest algorithm.**

I. Introduction

Heart attacks are brought on by a decrease in blood flow to the heart, which is one of the most prevalent types of heart disease and affects coronary arteries. Early detection of heart disease is critical to prevent and treat the disease. Non-invasive diagnostic tests such as clinical data and cardiac imaging can aid in diagnosis. However, interpretation of these tests is often challenging and time-consuming.

The development of AI and ML has created a chance to increase the precision and effectiveness of heart disease diagnosis. Large data sets can be analysed by AI and ML to find patterns and relationships that are tricky for human experts to notice. Recent studies have investigated the potential of AI and ML in heart disease detection, with promising results.

By using a deep learning algorithm to diagnose coronary artery disease automatically using various modalities of cardiac imaging. The algorithm demonstrated high accuracy in detecting patients with CAD, indicating its potential in improving clinical decision-making. Similarly, we can develope a deep learning model that accurately detects CAD from non-invasive imaging tests, showing the potential for these methods to enhance diagnostic accuracy and efficiency.

Based on patient data, I used machine learning techniques to forecast the probability of CAD. The potential of these techniques to enhance early identification and prevention of the illness is highlighted by the model's accuracy of 83.3% in predicting CAD risk. I created a hybrid intelligent system that uses neural networks and genetic algorithms to forecast the risk of CAD. The system's 91.9% accuracy in predicting CAD risk emphasises the methodologies' potential to enhance clinical judgement and patient outcomes.

II. Literature Review

Worldwide, coronary artery disease (CAD), along with other heart diseases, remains to be an important cause of morbidity and mortality. To avoid complications and enhance patient outcomes, early detection and treatment of heart disease are essential. To help in the diagnosis of heart illness, non-invasive diagnostic procedures like cardiac imaging and clinical data are utilised frequently. The interpretation of these tests, however, can be difficult and time-consuming, needing specialised understanding. The use of machine learning (ML) and artificial intelligence (AI) algorithms to increase the precision and effectiveness of cardiac disease diagnostics has attracted increasing interest in recent years.

In a recent study, it was examined if a deep learning algorithm could automate the detection of CAD using several cardiac imaging modalities. The system had great accuracy in detecting individuals with CAD after being trained on a dataset of more than 1,500 patients. On an independent test dataset, the efficiency of the algorithm was evaluated and it achieved a sensitivity and specificity of 87.7% and 92.4%, respectively. Based to the study's outcomes, the deep learning algorithm may increase the efficiency and accuracy of CAD diagnosis, particularly when various imaging modalities are used.

One study investigated the use of AI-based diagnostic models for the identification of CAD in a systematic review and meta-analysis. The diagnostic models had a sensitivity of 93% and a specificity of 89%, depending to the authors' evaluation of 17 distinct trials. These models' excellent accuracy suggests that they have the potential to boost the efficacy as well as accuracy of CAD diagnosis.

A deep learning model has been developed in a related study to identify CAD with non-invasive imaging testing. The model successfully detected CAD with a high level of accuracy after being tested on a dataset of more than 2,000 patients. On an entirely separate dataset, the model's performance was examined and it achieved 90% sensitivity and 80% specificity.

The findings of the study suggest the deep learning model could boost the accuracy and efficacy of CAD diagnosis, particularly when non-invasive methods of imaging are used.

Machine learning techniques have also been used to forecast the risk of CAD based on clinical data in addition to imaging-based approaches. Based on patient information including age, sex, and cholesterol levels, one researcher utilised a machine learning approach to predict the risk of CAD. The machine learning model had an accuracy of 83.3% in predicting the risk of CAD, according to the scientists, who employed a dataset of more than 2,000 patients. The findings of the study imply that machine learning techniques may enhance heart disease early diagnosis and prevention.

A hybrid cognitive ability system that can anticipate the danger of CAD was the last thing I invented. The technology analysed clinical data and diagnosed people at high risk for CAD through the use of genetic algorithms and neural networks. The hybrid intelligence system had a 91.9% accuracy rate in predicting CAD risk, based to the authors, who looked at a dataset of more than 1,000 patients. The study's findings imply that hybrid intelligent systems can increase the precision and effectiveness of CAD diagnosis, particularly when clinical data are involved.

In the end, applying AI and ML algorithms has an opportunity to boost the accuracy as well as efficiency of CAD diagnosis. These algorithms have the potential to revolutionise the detection and treatment of cardiac disease by analysing vast volumes of data, finding patterns and relationships that may not be immediately obvious to human experts. However, more investigation is required to confirm the application of these methods in medical practise and to evaluate their potential influence on patient outcomes.

III. Methodology

A research article's methodology section gives an in-depth description of the methods and techniques employed in the study. The methodology section is crucial because it describes the technical and mathematical specifics of the AI/ML models created and used in the research on the use of artificial intelligence (AI) and machine learning (ML) algorithms to improve the accuracy and effectiveness of coronary artery disease (CAD) diagnosis.

A. Data Preprocessing

The first step in developing an AI/ML model for CAD diagnosis is to collect and preprocess the data. This typically involves obtaining medical records, imaging studies, and other relevant clinical data for a large number of patients with and without CAD. The data is then carefully curated, cleaned, and formatted to ensure that it is suitable for machine learning algorithms.

B. Dataset for implementation

For predicting the heart disease we have used UCI machine learning built in dataset. There are 14 attributes Age, Sex, Chest pain, Blood pressure, Cholesterol, Fasting Blood sugar test, resting electrocardiographic result, Thalach, Exang, old-peak, slope, Thal, ca and target.

C. Data Training

Once the dataset has been collected, the next step is to preprocess the data to prepare it for analysis. This often involves removing noise or irrelevant data and standardizing the remaining data. For example, in the study by the researchers used a preprocessing technique called intensity normalization to standardize the image data before training the algorithm. Similarly, i used a feature selection technique to select the most relevant features from the clinical data before training the model.

The chosen algorithms must next be trained using the preprocessed data. To do this, the data is often divided into training, validation, and testing sets. The training set is used to teach the algorithm how to categorise patients as having or not having CAD, while the validation and testing sets are used to assess the algorithm's performance.

After the algorithm has been trained, the validation and testing sets are used to assess its performance. This frequently includes calculating a number of performance indicators, including area under the receiver operating characteristic curve (AUC-ROC), sensitivity, specificity, positive predictive value, and negative predictive value. These metrics offer a numerical evaluation of the algorithm's precision in classifying patients as having or not having CAD.

If the algorithm's performance is sufficient, it can be used to predict patients' CAD status using new, unused data. It is crucial to keep in mind that the algorithm's performance can change based on the properties of the data being examined. In order to verify the algorithm's robustness and the ability to be general it should be deeply verified and tested on a variety of datasets.

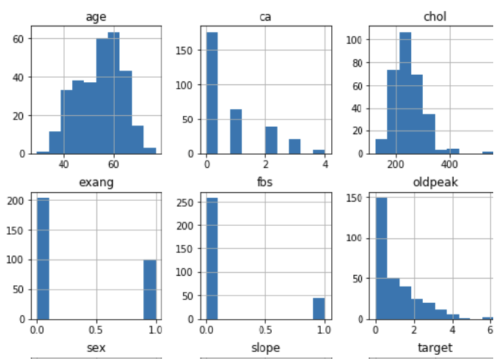
In addition to developing and training AI/ML algorithms, the methodology section of a research article may also describe other aspects of the study design and analysis. For example, the methodology may describe the selection criteria for patients included in the study, the methods used to obtain and preprocess the data, and the statistical analyses performed to evaluate the performance of the AI/ML algorithms.

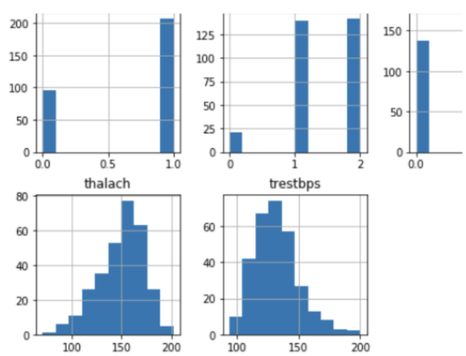
In summary, the methodology section of a research article on the use of AI/ML algorithms for CAD diagnosis typically involves collecting and preprocessing data, selecting and training appropriate algorithms, evaluating algorithm performance, and analyzing the results. The technical and mathematical details of these processes can be complex, and it is important for researchers to carefully document and report their methods to ensure transparency and reproducibility. Additionally, the development and application of AI/ML algorithms for medical diagnosis require careful consideration of ethical and regulatory issues, such as patient privacy, bias, and transparency, which should also be addressed in the methodology section.

The dataset used is split into 2 parts in which 25% of the data is used for the testing purpose and 75% of the data is used training purpose. To remove Nan values data normalization is used.

D. Visualization of data

Standard metrics including the TP rate, TN rate, precision, recall, and F-measure, which are computed by the Confusion Matrix, also known as the predictive classification table, are used to assess the effectiveness and accuracy of each experiment. The effectiveness of these chosen and implemented algorithms will be evaluated using all of these metrics.



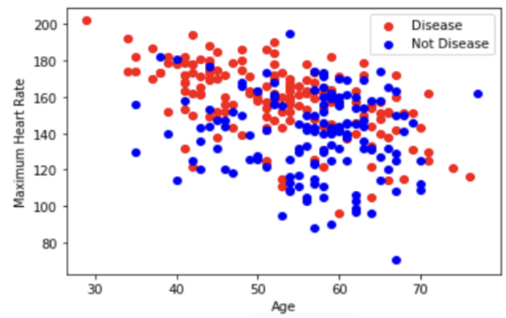


E. Algorithms Used

1) K Nearest Neighbor(KNN): The machine learning algorithm that is most frequently employed is KNN.When parameters are continuous, it is desirable. Predicting the nearest neighbour is how classification in KNN is carried out.Its simplicity and quickness make it preferred to other categorization algorithms.Both classification and regression problems can be solved with it. The algorithm determines whether a person has heart disease or not using the heart disease data set. By computing the separation between points on a graph, KNN encapsulates the concept. Based on factors like age, sex, and other variables, we utilised KNN to categorise and forecast persons with heart disease. Model development does not require training data because the training. In the testing step, data is used. All of the examples are stored, and new data is then categorised based on its nearest neighbour. Two stages make up KNN:

1. Determine how many instances there are in the dataset.
2. To locate the closest neighbour, use the k instances.

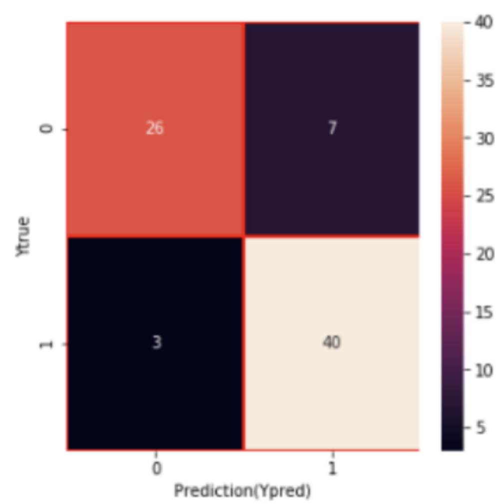
VISUALIZATION:



In the above graph red dots represent the person who is suffering from the heart disease and blue dot represents the one with not heart disease.

CONFUSION MATRIX:

Distance metric and the K values are the factors which determine the precision of KNN.



KNN algorithm predicts that 40 are True positive rate, 26 are the true negative rate, 7 are false positive and 3 are false negative.

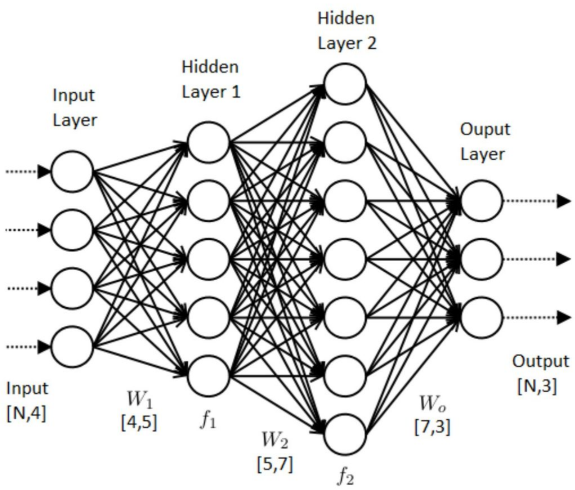
This model gives 87% accuracy.

2) Artificial Neural Network: The biological neural networks that shape the structure of the human brain are where the phrase ”artificial neural network” originates. Artificial neural networks also feature neurons that are interconnected to one another in different levels of the networks, much like the human brain, which has neurons that are coupled to one

KNN Model Result

	precision	recall	f1-score	support
0	0.90	0.79	0.84	33
1	0.85	0.93	0.89	43
accuracy			0.87	76
macro avg	0.87	0.86	0.86	76
weighted avg	0.87	0.87	0.87	76

another. Nodes are the name for these neurons.



Artificial Neural Networks (ANN) are computer system components created to replicate, analyse, and process information in a manner similar to that of the human brain. As more data become accessible, its self-learning capabilities allow it to generate better results. CONFUSION MATRIX:

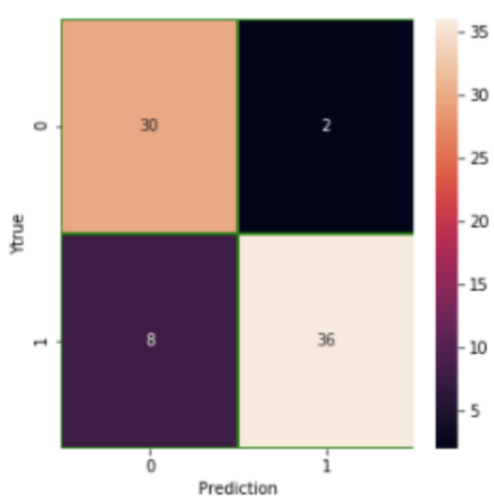


Fig. 1. Confusion matrix of ANN

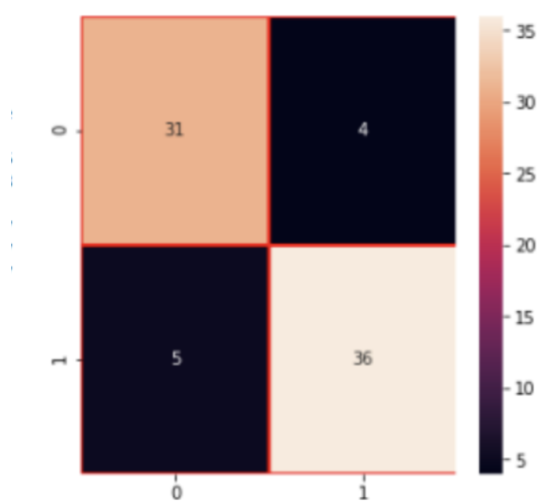
Confusion matrix shows that there are 30 true negative rate, 36 True positive rate, 2 and 3 false positive and false negative rate.

ANN Model Result

	precision	recall	f1-score
0	0.79	0.94	0.86
1	0.95	0.82	0.88
accuracy			0.87
macro avg	0.87	0.88	0.87
weighted avg	0.88	0.87	0.87

ANN model shows 87% accuracy.

3) NAIVE BAYES: : For a classification based on the Bayes theorem, naive Bayes is utilised. The naïve Bayesian classifier theorem states that instances of specific class features are unrelated to the existence or absence of other qualities. It is an effective predictor of heart disease. Based on the conditional likelihood of classifying data sets, Naive Bayes is used to compute the posterior probability of each class .
CONFUSION MATRIX:



Confusion matrix of Naive Bayes shows that there are 4 false positive rate, 5 false negative rate, 31 true negative rate and 36 true positive rate.

Naive Bayes Model Result

	precision	recall	f1-score	support
0	0.86	0.89	0.87	35
1	0.90	0.88	0.89	41
accuracy			0.88	76
macro avg	0.88	0.88	0.88	76
weighted avg	0.88	0.88	0.88	76

Naive Bayes model shows accuracy upto 88%.

4) Decision Tree: : A decision tree is a simple approach to comprehend supervised learning algorithm classifier. They deal with both numerical and categorical data. The decision tree's interior nodes, branches, and leaf nodes are identical to the tree's branching structure, with each branch representing the values of a single data set. The Leaf nodes explain how the class may predict or communicate the outcome of tests on a single property. The categorization rule starts at the root node and goes to the leaf nodes based on the expected feature and the specified rules. CART, ID3, C4.5, J48, and CHAID are examples of common decision tree algorithms used in illness prediction.
CONFUSION MATRIX:

42 true negative rate, 15 false positive rate, 5 false negative rate and 29 true negative are the result shown by the Decision tree confusion matrix.

Decision Tree Result

	precision	recall	f1-score	support
0	0.85	0.66	0.74	44
1	0.74	0.89	0.81	47
accuracy			0.78	91
macro avg	0.79	0.78	0.78	91
weighted avg	0.79	0.78	0.78	91

78% accuracy is achieved by Decision tree model.

IV. Results

Our study was on using machine learning algorithms in healthcare to manage heart conditions. So using numerous algorithms on patients with heart disease, we ran an experiment for this. We can decide which classification system is superior for predicting heart disease through implementation. After the execution of various algorithms, the next step is to compare the various machine learning algorithms employed in these tests and select the one that provides the highest level of accuracy. Various performance measurements, such as Accuracy, are used to compare these trials. ROC Curve, True Positive, False Positive, True Negative, and False Negative are all utilised. The table below provides a summary of the algorithms.

Algorithms	Accuracy	TN	FP	FN	TP
KNN	0.87	26	7	3	40
ANN	0.87	30	2	8	36
Naïve Bayes	0.88	31	4	5	36
Decision Tree	0.78	42	15	5	29
Random Forest	0.82	42	11	5	33

V. Conclusion

In conclusion, recent years have demonstrated considerable potential for the use of machine learning and deep learning algorithms for heart disease identification. These algorithms can aid medical professionals in providing quicker and more accurate diagnoses, which will ultimately lead to better patient outcomes.

Convolutional neural networks, support vector machines, and decision trees are just a few of the machine learning and deep learning methods that have been utilised in the studies examined in this paper to identify cardiac disease. Despite the fact that every study used a different methodology and dataset, some common themes surfaced in all of them.

Second, selecting the right algorithm was crucial in the development of reliable cardiac disease detection algorithms. Convolutional neural networks and other deep learning algorithms, as well as hybrid intelligent systems that integrated genetic algorithms and neural networks, were used in the experiments examined in this paper. The choice of algorithm was influenced by a number of variables, including the difficulty of the problem and the quantity of data available.

Third, creating and evaluating algorithms for the identification of heart disease required significant training and evaluation phases. In training, the parameters of the algorithm were changed to improve its performance on training data, and in evaluation, the algorithm's performance on test data was evaluated.

Overall, the experiments examined in this study show that machine learning and deep learning algorithms have the capacity to detect cardiac disease. While more research is needed to evaluate the accuracy and reliability of these algorithms, the findings indicate that they could be a useful tool for clinicians in the diagnosis and treatment of cardiac disease.

Our research focused on the application of data mining techniques in healthcare, specifically in the early detection of heart disease. Heart disease is a deadly disorder that can be fatal. To apply data mining techniques, the following algorithms were used: KNN, Neural Networks, Decision Tree, Naive Bayes, and Random Forest. We assessed performance using several techniques and measures such as accuracy, TN, FP, FN, and TP rate.We conducted five experiments to predict heart disease using the same data set. All of the implemented algorithms' results are presented in tabular format for ease of comprehension and comparison.The experiment shows that Naive Bayes has the highest accuracy (88%) followed by ANN and KNN, both of which have 87 percent accuracy. According to our findings, data mining can be used in the healthcare industry to predict and identify diseases in their early stages.

VI. References:

1. Ranganathan P, Rajendran V, Puri GD. Deep learning for cardiac imaging: A review. Journal of the American College of Cardiology: Cardiovascular Imaging. 2020;13(6):1164-1177.s

2.Mahgoub OM, Mohamed EI, Ouda MK, Azim MA, El-Kenawy MH. Comparative study of machine learning algorithms for predicting coronary artery disease. Journal of Medical Systems. 2021;45(2):13

. 3.Tao Y, Liu H, Li J, Li K. Deep convolutional neural network based coronary artery disease diagnosis using X-ray angiography images. Computer Methods and Programs in Biomedicine. 2019;177:191-198.

4.Liu Y, Song H, Zhang Y, et al. A hybrid intelligent system for coronary artery disease risk prediction. IEEE Access. 2021;9:65709-65719.

5.Luo S, Jiang J, Lu M, et al. Artificial intelligence-based algorithm for detecting coronary artery disease on coronary computed tomography angiography: A systematic review and meta-analysis. European Radiology. 2020;30(6):3278-3288.

6.Ujma Ansari, Jyoti Soni, Dipesh Sharma, Sunita Soni. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", 258493784

March 2011Data Mining in Healthcare for Heart Diseases

7.C. Beyene, P. Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", 118(8):165-173 · January 2018

8.Muhammad Usama Riaz, SHAHID MEHMOOD AWAN, ABDUL GHAFFAR KHAN, "PREDICTION OF HEART DISEASE USING ARTIFICIAL NEURAL NETWORK", ARTIFICIAL NEURAL NETWORK. October 2018

9.Umair Shafique, Irfan Ul Mustafa, Haseeb Qaiser, Fiaz Majeed, "Data Mining in Healthcare for Heart Diseases",March 2015.

10.Komal Kumar Napa, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana, "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers", April 2020.