International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

# *Prediction and Classification of Heart Disease through Artificial Neural Networks and Decision Trees*

**Abhinay Singh, Arifa khan**
Computer Science and engineering
Lucknow institute of technology, Lucknow
abhinaysinghcs@gmail.com

**Abstract: Numerous studies have been conducted on the prediction and progression analysis of heart disease using Artificial Intelligence (AI) and machine learning techniques. These approaches assist researchers and healthcare professionals in identifying cardiac abnormalities at an early stage, thereby improving diagnosis and treatment planning. To develop and evaluate AI-based prediction models, researchers commonly utilize benchmark medical datasets available in the University of California, Irvine Machine Learning Repository. Among these datasets, the Cleveland Heart Disease dataset is the most frequently used and widely accepted for heart disease prediction research due to its reliability and comprehensive clinical attributes. The Cleveland dataset contains 303 patient records along with several clinical and physiological parameters associated with cardiovascular health. Although the original dataset includes 76 attributes, researchers generally focus on 14 significant features that are considered highly relevant for heart disease diagnosis and classification. These attributes include age, sex, chest pain type, resting blood pressure, serum cholesterol level measured in mg/dl, fasting blood sugar level greater than 120 mg/dl, resting electrocardiographic (ECG) results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, and the number of major vessels colored by fluoroscopy. The final target attribute indicates the presence or absence of heart disease in a patient. These selected features play a crucial role in training machine learning and deep learning models such as Decision Trees, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Random Forests, and other predictive algorithms. By analyzing patterns within these attributes, AI models can effectively classify patients into different risk categories and support clinicians in making accurate and timely medical decisions.**

**Keywords: Cardiovascular Disease, Decision Support System, Data Mining, Hybrid Intelligent System**

## 1. Introduction:

Heart disease is one of the leading causes of death worldwide and poses a significant challenge to modern healthcare systems. Cardiovascular diseases, including coronary artery disease, heart failure, and arrhythmias, affect millions of people every year and often result in severe health complications if not diagnosed at an early stage. The increasing prevalence of unhealthy lifestyles, obesity, diabetes, hypertension, smoking, and stress has contributed substantially to the rapid growth of heart-related disorders across both developed and developing countries. Early detection and accurate diagnosis of heart disease are therefore essential to reduce mortality rates and improve patient outcomes. However, traditional diagnostic methods often require extensive clinical examinations, expert interpretation, and considerable time, which may delay effective treatment.

In recent years, Artificial Intelligence (AI) and machine learning techniques have emerged as powerful tools in the healthcare sector for disease prediction and medical decision-making. These techniques enable computers to analyze large volumes of medical data, identify hidden patterns, and generate accurate predictions with minimal human intervention. AI-based systems can assist healthcare professionals in detecting heart disease at an early stage, improving diagnostic accuracy, reducing human errors, and supporting timely treatment planning. Machine learning algorithms such as Decision Trees, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Random Forests, and Logistic Regression have been widely applied in heart disease prediction studies due to their capability to handle complex medical datasets efficiently.

Among the publicly available datasets used for heart disease research, the Cleveland Heart Disease dataset from the University of California, Irvine repository is the most widely utilized benchmark dataset. This dataset contains 303 patient records with multiple clinical attributes related to cardiovascular health. Although the original dataset includes 76 attributes, researchers commonly select 14 important features such as age, sex, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar, electrocardiographic results, maximum heart rate, exercise-induced angina, and the number of major vessels. These attributes provide valuable information for training intelligent prediction models capable of distinguishing between healthy individuals and patients suffering from heart disease.

Decision Tree and Artificial Neural Network (ANN) models are among the most effective machine learning approaches

used for heart disease prediction and classification. Decision Tree algorithms are simple, interpretable, and capable of generating rule-based predictions, making them suitable for medical decision support systems. On the other hand, ANN models mimic the functioning of the human brain and are highly efficient in identifying complex nonlinear relationships within medical data. By combining these intelligent techniques, researchers can develop robust and accurate predictive systems that support clinicians in diagnosing heart disease more effectively.

This study focuses on the prediction and classification of heart disease using Decision Tree and ANN models. The primary objective is to analyze the performance of these machine learning techniques in terms of prediction accuracy, efficiency, and reliability. The proposed approach aims to assist healthcare professionals by providing an intelligent and automated system for early heart disease detection, ultimately contributing to improved patient care and reduced healthcare burden.

## 2. Related Work:

Artificial intelligence (AI) computations using relapse tree, Several machine learning techniques, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have been extensively studied for evaluating and predicting the severity of cardiovascular diseases. Researchers have reported that SVM provides superior performance compared to many other classification algorithms in heart disease diagnosis. Various studies demonstrated that SVM achieved a prediction accuracy of approximately 94.60%, while also reducing classification errors in disease prediction. Due to its high precision and reliability, SVM has been considered one of the most effective methodologies for coronary heart disease diagnosis and clinical decision support systems [8–10].

In another study, seven machine learning algorithms—namely Naïve Bayes, Decision Trees, K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), OneR classifier, and SVM—were evaluated for heart disease forecasting using a dataset containing 302 patient cases. After comparing the performance of these techniques under different conditions, the researchers concluded that the SVM algorithm produced the most accurate and effective results for heart disease prediction [11]. SVM-based approaches have also been applied successfully in predicting cardiovascular complications among diabetic patients, further highlighting their effectiveness in medical diagnosis systems [12].

Recent advancements in mobile healthcare technologies have also contributed significantly to heart disease monitoring and prediction. A specially designed mobile-based machine learning model was developed for monitoring coronary heart disease through smartphone applications. Clinical data collected from 200 patients were analyzed using intelligent algorithms, and the system achieved a prediction accuracy of 90.5% [13]. Additionally, several studies have focused on evaluating the execution and performance analysis of different machine learning algorithms for heart disease prediction [14]. Researchers have also proposed cloud-based e-healthcare frameworks where machine learning models predict the risk level of heart disease and allow both patients and healthcare professionals to access medical information remotely through integrated wellness systems [15].

Decision Tree analysis has also been widely used to develop predictive models for detecting cardiovascular disease. In one investigation, clinical information was collected from 1159 healthy individuals and 1187 patients who had undergone coronary angiography. The developed Decision Tree-based model demonstrated excellent performance, achieving specificity, sensitivity, and accuracy values of 87%, 96%, and 94%, respectively. Tree-based approaches were found to improve the effectiveness and interpretability of heart disease prediction systems [16]. Furthermore, comparative studies between traditional statistical techniques and advanced machine learning methods revealed that logistic regression models were also capable of producing highly accurate predictions for cardiovascular disease diagnosis, sometimes outperforming certain data mining and AI-based approaches depending on the dataset characteristics and clinical conditions [17].

## 3. Methodology:

Deep learning has emerged as one of the most influential and widely discussed areas in the field of Artificial Intelligence (AI). In recent years, it has gained tremendous attention from researchers, industries, and the media due to its remarkable success in solving complex real-world problems. Deep learning can be described as a subset of machine learning that utilizes multi-layered Artificial Neural Networks (ANNs) to learn patterns and representations from large amounts of data. These neural networks consist of interconnected artificial neurons that imitate the functioning of biological neurons in the human brain. The primary objective of deep learning is to enable machines to automatically learn hierarchical features and make intelligent decisions without extensive manual intervention.

The concept of Artificial Neural Networks originated from studies on the human nervous system and the way the brain processes information. The foundation of neural network research dates back to the 1940s when Warren McCulloch and Walter Pitts introduced the first mathematical model of an artificial neuron. Their work attempted to explain how biological neurons interact and transmit electrical and chemical signals in the human brain. Later, in the 1950s, Frank Rosenblatt proposed the perceptron model, which became one of the earliest neural network architectures capable of performing simple pattern recognition tasks. Despite these developments, research interest in neural networks gradually declined because scientists faced significant challenges in training neural networks with multiple hidden layers.

The revival of neural network research occurred in 1986 when David E. Rumelhart, Geoffrey Hinton, and Ronald J. Williams popularized the backpropagation algorithm, which enabled efficient training of multilayer neural networks. This breakthrough significantly improved the learning capability of neural networks and laid the foundation for modern deep learning techniques. Over the past decade, rapid advancements in computational power, graphics processing units (GPUs), and the availability of large-scale datasets have further accelerated the growth of deep learning. These developments have enabled deep neural networks to achieve exceptional performance in tasks such as image classification, speech recognition, natural language processing, medical diagnosis, and autonomous systems.

Today, deep learning plays a vital role in numerous real-world applications and is heavily utilized by major technology companies such as Google
, Microsoft
, and Meta
for developing intelligent systems and services. Applications powered by deep learning include image search engines, voice assistants, recommendation systems, facial recognition, language translation, and healthcare analytics. For example, Google Lens and image recognition systems can automatically identify objects and translate text from images into multiple languages in real time. Due to its capability to process complex and unstructured data efficiently, deep learning is now considered a state-of-the-art approach in AI research and practical implementation.

Before exploring advanced neural network models and deep learning algorithms, it is important to understand the early foundations of Artificial Intelligence and neural computation. The initial efforts by researchers such as McCulloch and Pitts to model artificial neurons established the basis for modern neural network architectures. These early concepts eventually evolved into sophisticated deep learning frameworks that continue to transform various scientific, industrial, and healthcare domains today.

McCullock and Pitts described such a nerve cell as a simple rational entryway with two results: first, different signs arrive at the dendrites, are then incorporated into the phone body, and, if the total number of signs exceeds a certain threshold, a result signal is generated and transmitted by the axon. Candid Rosenblatt published the primary idea of the perceptron learning rule in light of the MCP neuron model within a few years after it was first proposed (F. Rosenblatt, The Perceptron, a Seeing and Perceiving Robot. Cornell Aeronautical Research facility, 1957). Rosenblatt's perceptron rule is a calculation that would logically become familiar with the optimum weight coefficients, which are then replicated with the information highlights to decide whether or not a neuron fires. Such a formula might therefore be used to predict whether an example belonged with one class or the other in terms of controlled learning and grouping. To describe this problem more formally, we may frame it as a parallel characterisation task and refer to our two classes as 1 (positive class) and - 1 (negative class) for simplicity. Then, where z is the alleged net info, we can define an enactment capability (z) that accepts a direct blend of particular data values x and a comparing weight vector w. The perceptron gathers the contributions of an example x and joins them with the loads w to analyse the network information, as shown in figure 2. The initiation capability (in this case, the unit step capability), after receiving the net information, generates a double result of 1 or +1, which corresponds to the example's expected class grade. This outcome is used to correct the expectation error and update the weights throughout the learning phase. A quantizer, which is similar to the unit step capability we have previously seen, can then be used to predict the class names while the direct initiation capacity is being used to learn the loads.

**Proposed Work:**

In machine learning problems, Confusion Matrix is generally used for statistical analysis of data. It is also known as the error matrix. A confusion matrix [15] basically denotes the relationship between the actual and the predicted results of a classifier. Classifier performance can be viewed based on the values obtained in the confusion matrix.As the name suggests it denotes how much the model is confused when it makes predictions on the test set. If we have a binary classifier the confusion matrix will be a 2X2 matrix which can be represented as follows. Assume that there are two classes positive and negative.

Where TP represents True Positive It depicts the number of tuples of positive that have been correctly predicted.FN represents False Negative: It depicts the number of tuples of Class 1 that have been incorrectly predicted to be in negative.FP represents False Positive: It depicts the number of tuples of negative that have been predicted to be in positive.TN represents True Negative: It depicts the number of tuples of positive that have been correctly predicted to be in negative.

**4. Result and Discussion:**

The exploratory data analysis is performed on the data set to find the variance of the data set .the exploratry data analysis is the most important phase of data analysis by using which the data is interprited and gives a better understanding for the type of model to be used with the data . the vaious features of the feature set is plotted by using count plots and histograms so that we can check the kind of data and data distribution availabe.the below image shows the analysis of each columns of the data set. The data is skewd for the removal of skewness of data we can perform either over sampling or under sampling but in our case we have only 310 rows in our data hence we will prefer the over sampling by using the ADASYN over sampler. The data is preproccesd using the min max scalar, the the data is divided into 80:20 ratio as the training set and testing set. The train data set is firstly treated for the

data skewness and then the data set is used to traing a decision tree so that the important features of the data set can be obtained by building the decision tree. The clolumns used as the nodes of the tree is used as the important features . for the optimization of the model we have used the random search cv as the hyper parameter tuning model the ANN model is optimized for the activation layers and the loss model. The ANN model used contains three different layers the first layer is take the input of 13 columns and it is transformed to 10 by uising the relu activation function,the second layer converts the 10 columns into 8 columns by using the relu activation function then finally the 8 input is usedto genrate one output by using the activation function of sigmoid. The values predicted is between 0 to 1 then by round function we have converted the values back to either 0 or 1.

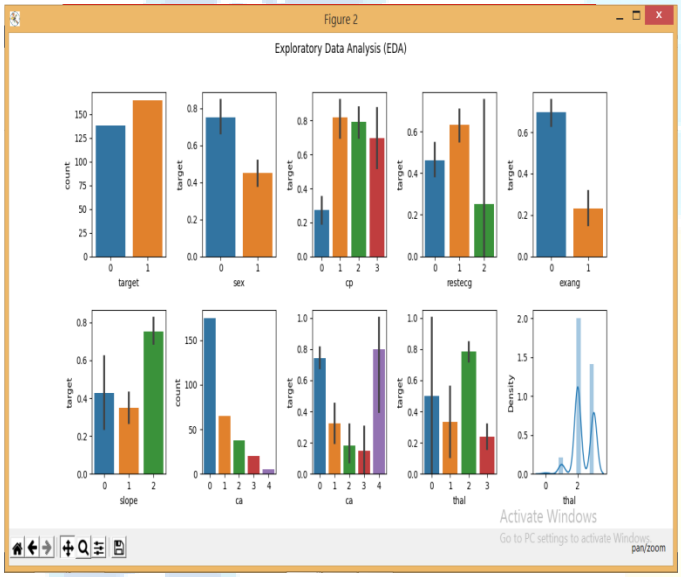


**Fig 3: exploratory data analysis of the data set**

Then by using the different metrics system we have calculated the precision, accuracy, recall, f1-score. The same data is used to train the various other machine learning models which were previously to build the models the result of metrics of each model is shown in the below figure. The machine learning models are also hyper tuned for different parameters in the different models like k neaighbors is tuned for the n neighbours concept, the optimal value for k neighbors is found to be seven. The decsion tree and the random forest are hyper tuned for the concept of random_state , the optimal value for random state in decision tree is 11 and in case of random forest it is 323. The logistic regression model, suppportvector machine algoritms are tuned for the parameters of the learning rate. For the implementation of the ANN model wehave used the tensorflow and the keras modules and for other models the sklearn module is used. The result of each model is shown in the below figure. The results clearly shows that the ANN model has the highest accuracy as compared to other models. and the k neighbours has the least accuracy. The Ann model resulted a accuracy of 92.16% accuracy, 94.12 % precision, 88.89% recall, 91.43% f1-score.

heart disease prediction using ANN

| ALGORITHMS | ACCURACY | PRECISION | RECALL | F1-SCORE |
|---|---|---|---|---|
| Logistic Regression | 85.25 | 88.24 | 85.71 | 86.96 |
| Naive Bayes | 85.25 | 91.18 | 83.78 | 87.32 |
| Support Vector Machine | 81.97 | 88.24 | 81.08 | 84.51 |
| K-Nearest Neighbors | 67.21 | 67.65 | 71.88 | 69.7 |
| Decision Tree | 81.97 | 82.35 | 84.85 | 83.58 |
| Random Forest | 88.24 | 87.41 | 85.85 | 90.0 |
| ANN | 92.16 | 94.12 | 88.89 | 91.43 |

**Fig 4: results of the ANN model and other models.**

The results obtained are graphicaly shown in the below figure for the better result analysyis for implementing the graphical representation of the results we have used the matplotlib and the seaborn library.

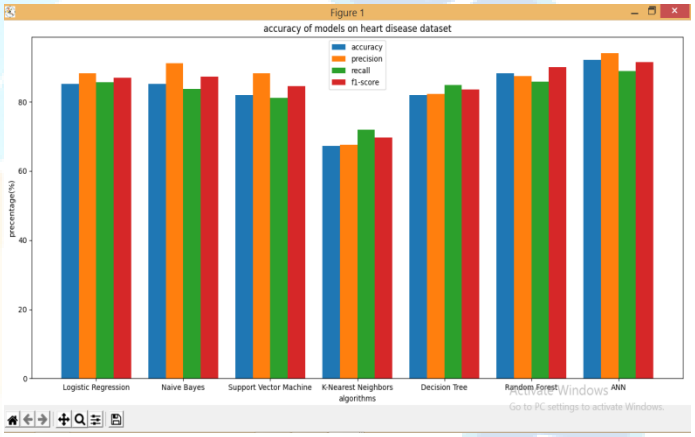


**Fig 5: results analysis**

## 5. Conclusion:

In our work we have proposed an ANN-DT (Artificial neural network decision tree) based model for predicting the heart disease. Previously various Machine learning, Datamining and Neural Network models were employed for the same. Machine learning and the Datamining did not have a good method for the selection of the important features. We have used the Decision tree as the model for the feature selection. And the selected features are used to train the ANN model. The ANN model was built by using the relu action function, sigmoid function, Adams solver and the binary entropy loss function. The ANN approach has outperform the existing techniques for the heart disease prediction. We have done the proper pre-processing of data by using the scalers and the skewness is removed by using the ADASYN over sampler.

## References:

[1] Nidhi Bhatla, and Kiran Jyoti, Oct. 2012,"An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 1, Issue 8, ISSN: 2278-0181, pp. 1-4.

[2] Sumitra Sangwan, and Tazeem Ahmad Khan, Mar. 2015,"Review Paper Automatic Console for Disease

Prediction using Integrated Module of A-priori and k-mean through ECG Signal",International Journal For Technological Research In Engineering, Vol. 2, Issue 7, ISSN(Online): 2347-4718, pp. 1368-1372.
[3] Rishi Dubey, and Santosh Chandrakar, Aug. 2015,"Review on Hybrid Data Mining Techniques for The Diagnosis of Heart Diseases in Medical Ground" ,Vol. 5, Issue 8, ISSN: 2249-555X, pp. 715-718.
[4] Ashish Chhabbi,Lakhan Ahuja,Sahil Ahir, and Y. K. Sharma,19 March 2016,"Heart Disease Prediction Using Data Mining Techniques", International Journal of Research in Advent Technology,E-ISSN:2321-9637,Special Issue National Conference "NCPC-2016", pp. 104-106.
[5] Shaikh Abdul Hannan, A.V. Mane, R. R. Manza,and R. J. Ramteke, Dec 2010,"Prediction of Heart Disease Medical Prescription using Radial Basis Function",IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), DOI: 10.1109/ICCIC.2010.5705900 ,28-29 .
[6] AH Chen, SY Huang, PS Hong, CH Cheng, and EJ Lin,2011, "HDPS: Heart Disease Prediction System",Computing in Cardiology, ISSN: 0276-6574, pp.557-560. [7] Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010,"Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network",International Conference on Computer and Communication Technology (ICCCT),DOI:10.1109/ICCCT.2010.5640377, 17-19.
[8] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis, and Vijay K. Mago,2016, "Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map",IEEE International Conference on Fuzzy Systems (FUZZ),pp. 1377-1382.
[9] Carlos Ordonez,2006, "Association Rule Discovery With the Train and Test Approach for Heart Disease Prediction", IEEE Transactions on Information Technology in Biomedicine (TITB), pp. 334-343, vol. 10, no. 2.
[10] Prajakta Ghadge,Vrushali Girme, Kajal Kokane, and Prajakta Deshmukh, 2016,"Intelligent Heart Attack Prediction System Using Big Data", International Journal of Recent Research in Mathematics Computer Science and Information Technology,Vol. 2, Issue 2, pp.73-77, October 2015–March.
[11] Asha Rajkumar, and Mrs G. Sophia Reena, 2010, "Diagnosis of Heart Disease using Data Mining Algorithms",Global Journal of Computer Science and Technology,Vol. 10,Issue 10, pp.38-43, September.
[12] K. S. Kavitha, K. V. Ramakrishnan, and Manoj Kumar Singh, September 2010,"Modelling and Design of Evolutionary Neural Network for Heart Disease Detection" ,International Journal of Computer Science Issues (IJCSI), Vol. 7, Issue 5, pp. 272-283.
[13] K.Sudhakar, and Dr. M. Manimekalai, January 2014, "Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157-1160.
[14] Shantakumar B. Patil,and Dr. Y. S. Kumaraswamy, February 2009, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction", IJCSNS International Journal of Computer Science and Network Security, Vol. 9 , No. 2,pp. 228-235.
[15] Sairabi H. Mujawar, and P. R. Devale, October 2015,"Prediction of Heart Disease using Modified k-means and by using Naive Bayes", International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.
[16] S. Suganya, and P.Tamije Selvy, January 2016, "A Proficient Heart Disease Prediction Method using Fuzzy-Cart Algorithm", International Journal of Scientific Engineering and Applied Science (IJSEAS), Vol. 2, Issue1,ISSN: 2395-3470.
[17] Ashwini Shetty A, and Chandra Naik, May 2016,"Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative Research in Science,Engineering and Technology(An ISO 3297: 2007 Certified Organization), Vol. 5, Special Issue 9, pp. 277-281.
[18] K Cinetha, and Dr. P. Uma Maheswari, Mar.-Apr. 2014,"Decision Support System for Precluding Coronary Heart Disease using Fuzzy Logic.", International Journal of Computer Science Trends and Technology (IJCST), Vol. 2, Issue 2, pp. 102-107.
[19] Indira S. Fal Dessai,2013,"Intelligent Heart Disease Prediction System Using Probabilistic Neural Network", International Journal on Advanced Computer Theory and Engineering (IJACTE),Vol. 2, Issue 3, pp. 38-44.
[20] Mai Shouman, Tim Turner, and Rob Stocker, June 2012,"Applying k-Nearest Neighbors in Diagnosing HeartDiseasePatients",International Journal of Information and Education Technology, Vol. 2, No. 3,pp. 220-223.
[21] Serdar AYDIN, Meysam Ahanpanjeh,and Sogol Mohabbatiyan,February 2016, "Comparison And Evaluation of Data Mining Techniques in the Diagnosis of Heart Disease",International Journal on Computational Science & Applications (IJCSA), Vol. 6,No.1, pp. 1-15.