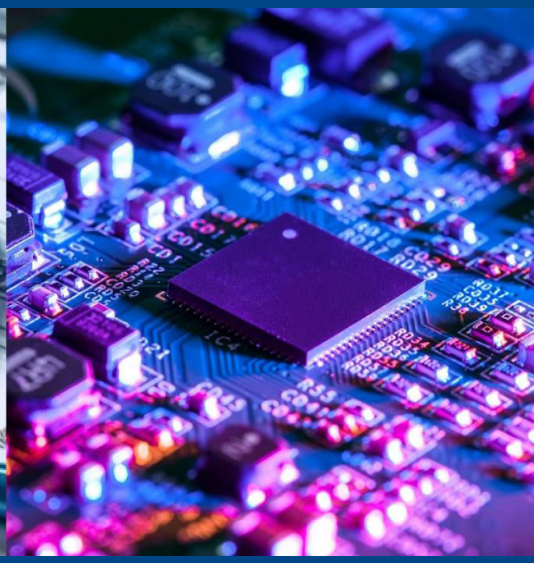
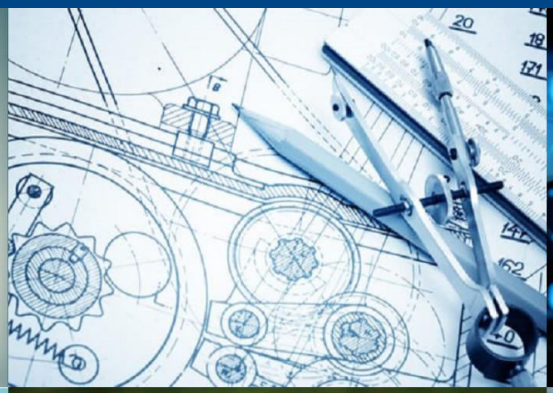
ISSN: 2582-7219

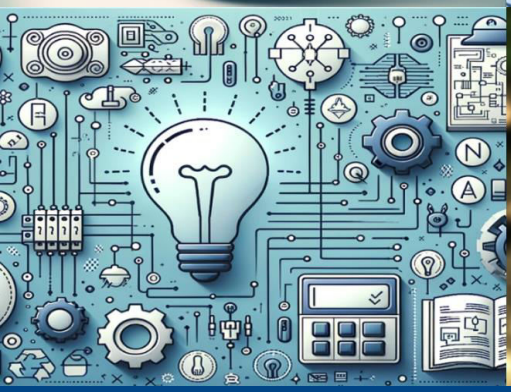




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 8, Issue 6, June 2025**

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806014

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Comparative Analysis of Heart Disease Prediction using Machine Learning Algorithms

Anushree M R

Department of Computer Applications, St Joseph Engineering (Autonomous)

College, Vamanjoor, Mangalore, India

ABSTRACT: Day by day the cases of heart diseases are increasing and it's very Important and concerning to predict any such diseases beforehand. So, there is a need for accurate, and feasible system to diagnose such diseases in time for proper treatment. In the last ten years, heart disease has become a leading cause of death worldwide. Heart Disease is an illness that affects many lives, is severely life-threatening, and can weaken a person's ability to live a conventional life. The reasons for heart disease are many, predicting makes it complicated. Here we have used Machine Learning algorithms to predict the accuracy of heart disease. Our objective is to use Machine Learning to ease the prediction of heart disease. Here several sets of data sets have been used on many machine learning algorithms. Here are some of the proposed algorithms that have been tested: Logistic Regression, Random Forest, Support Vector, KNN and Naive Bayes. After rigorous testing Logistic Regression, Random Forest and Naive Bayes stayed dominant in most of the testing achieving accuracies of 85.25% and KNN, Support Vector with the accuracy rate of 63.93% and 68.85%. As shown in this paper, Machine Learning is a prominent tool to predict Heart Disease, and results can be further improved with the help of medical professionals and more research.

KEYWORDS: heart disease, Machine Learning, Logistic Regression, Logistic Regression, Naive Bayes, KNN, Support Vector

I. INTRODUCTION

"Machine Learning is a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data". Machine Learning (ML) is a very powerful tool for the manipulating and known or unknown important information from data. Heart diseases have become a leading cause of death. Lifestyle factors such as smoking, alcohol consumption, physical inactivity, mental health, age, gender, sleeping patterns etc. causes heart disease.

Addressing this multifaceted challenge requires innovative approaches, and machine learning emerges as a promising avenue for predictive modelling and risk assessment. The expansive field of machine learning encompasses a diverse array of classifiers falling under the categories of Supervised, Unsupervised, and Ensemble Learning.

These classifiers are harnessed to predict and evaluate the accuracy of datasets related to heart disease. The models generated through machine learning techniques analyze new input data, filling any gaps in the dataset and providing accurate predictions. This facilitates the assessment of heart disease probability in individuals based on their unique health profiles. The massive datasets collected for heart disease prediction are rich in information but often plagued by noise. Machine learning algorithms prove invaluable in effectively processing such extensive and noisy data. The integration of machine learning into healthcare systems aids professionals in disease management, medication planning, and the discovery of intricate patterns and relationships within diagnostic data. Machine learning, with its ability to comprehend complex interactions between diverse risk factors, presents an opportunity to revolutionize cardiovascular risk prediction. By tapping into the intricate web of lifestyle choices, genetic predispositions, and health indicators, machine learning models have the potential to significantly enhance the accuracy of predicting an individual's susceptibility to heart diseases. In this comprehensive exploration, we delve into the various facets of machine learning applied to heart disease prediction. From understanding the risk factors to evaluating the performance of machine learning algorithms, this study aims to shed light on the transformative potential of ML in the realm of cardiovascular health.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE REVIEW

The escalating prevalence of heart disease has spurred a significant surge in research dedicated to enhancing diagnostic capabilities. This paper presents a comprehensive literature survey on the application of machine learning algorithms for efficient heart disease prediction. Various algorithms, including Logistic Regression, Random Forest, Support Vector Machines (SVM), KNearest Neighbors (KNN), and Naive Bayes, have been explored, each demonstrating distinct strengths in achieving predefined objectives. Harshit Jindal [1] conducted a study employing diverse machine learning techniques. KNN,

Logistic Regression,
Random Forest Classifier emerged as top performers, where KNN has achieved an impressive accuracy score of 87.5%. This research lays the foundation for understanding the comparative strengths of different algorithms in heart disease prediction.

Chaimaa Boukhatem [2] employed the methodologies like Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB). The SVM algorithm demonstrated its prowess in achieving accurate predictions for heart diseases of 91.67%.

T. Chithambaram [3] employed the methodologies like KNN, SVM, Random classier, decision tree classier given accurate result for Heart Disease Prediction System. The prediction was made better accuracy of 98.83% by decision tree machine learning method than other methods.

K. Karthick [4] here Support vector machine (SVM), Gaussian Naive Bayes, logistic regression, LightGBM, XGBoost, and random forest algorithm have been employed for developing heart disease risk prediction model and obtained the accuracy as 80.32%, 78.68%, 80.32%, 77.04%, 73.77%, and88.5%, respectively. The data visualization has been generated to illustrate the relationship between the features. According to the findings of the experiments, the random forest algorithm achieves 88.5% accuracy during validation for 303 data instances with 13 selected features of the Cleveland HD dataset.

Sonam Nikhar [5] proposed a hybrid approach, combining Decision Tree algorithm, Naive Bayes where Naive Bayes has a high level of accuracy. Krishna Battula [6] The model using several methods such as Logistic Regression, k Nearest Neighbors(kNN), Decision Trees, and Random Forest in order to predict heart disease. The performance of the Random Forest algorithm is found good compared to the remaining three algorithms. Random forest algorithm best fits the data with an accuracy of 88.16%.

P. Sujatha [7] In this research paper, the presence of heart disease is predicted by employing Decision Tree, Naïve Bayes, Random Forest, Support Vector Machine,K-Nearest Neighborand logistic Regression algorithms. From the experimental result, it is found that the Random Forest is more accurate for predicting the heart disease with accuracy of 83.52% compared with other supervised machine learning algorithms. The precision score of Random Forest classifiers are 84.21%, 88.24% and 88.89% respectively. Anish Gopal Pemmaraju [8] focused on cardiac disease prediction, with algorithms like Random Forest, Decision Tree, K- Nearest Neighbor, Support

Vector Machine, Gaussian Naive Bayes and Logistic Regression, are applied on two distinct heart disease datasets from Kaggle. The Random Forest classification algorithm has achieved the highest accuracy in both the datasets with accuracy on the first dataset is 77.25% and accuracy on the second dataset is 93.25%. Reldean Williams [9] The art Machine Learning Techniques Namely Artificial Neural Networks, Decision Trees and Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machines and XG Boost, were implemented at various evaluation stages to predict heart diseases. Results show that Random Forest technique has achieved a prediction accuracy of 95%.

Suraj Raut [10] the most effective algorithm was Decision Trees, Logistic Regression, Naive Bayes, SVM were implemented at various evaluation stages to predict heart diseases. The SVM had a highest level of accuracy of around 81.57%.

Malavika G [11] Simulation based experiments were conducted using six methodologies named Naive Bayes Classifier, Logistic Regression, Random Forest, SVM, Decision Tree Classifier and KNN. From the result it's been seen that the

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

random forest gives more accuracy of 91.80%.

Md. Julker Nayeem Nayeem [12] In order to calculate prediction accuracy, K-Nearest Neighbors (KNN), Naive Bayes and Random Forest are applied to the heart disease dataset. Random Forest among all has given the best classification accuracy which is 95.63%.

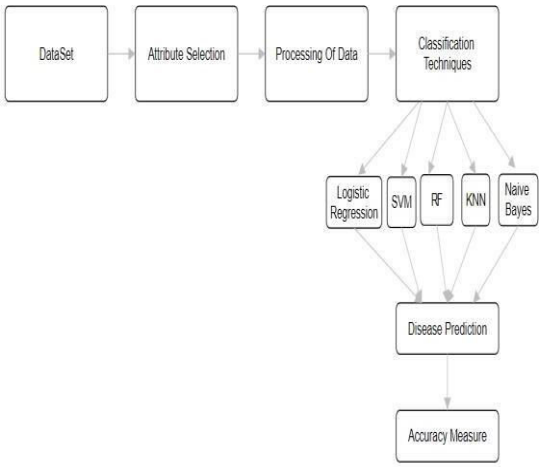
Archana Singh [13] Numerous algorithms were used, and research was done on each algorithm before training the model using them on the datasets. In this paper, we calculate accuracy of machine learning algorithms for predicting heart disease, for this algorithm are k-nearest neighbour, decision tree, linear regression and support vector machine.

Nouh Sabri Elmitwally [14] Researchers used different algorithms Naïve Byes, Neural Network, Decision Tree, and genetic algorithm for the prediction of heart disease Naïve Bayes shows good results and accuracy was about 96.6%.

Mr. VALLE HARSHA VARDHAN [15] Here the algorithms used are Decision Tree, Random Forest, Naive Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting. The extreme gradient boosting classifier has the highest accuracy (81%), when all seven are compared.

III. METHODOLOGY

This paper shows the analysis of various machine learning algorithms, the algorithms that are used in this paper are Logistic Regression, Random Forest, Support Vector, KNN and Naive Bayes. Classifiers which can be helpful for practitioners or medical analysts to accurately diagnose heart disease. This paperwork includes examining the journals, published papers and the data of cardiovascular disease of recent times. After preprocessing of data, a classifier is used to classify the pre-processed data. The classifiers used in the proposed model are KNN, Logistic Regression, Random Forest Classifier. We have evaluated our model on the basis of accuracy and performance using various performance algorithms.



Dataset: The study utilizes a heart disease dataset, encompassing crucial parameters such as sex, chest pain test results, testibps, cholesterol levels, fbs (fasting blood sugar), restecg (resting electrocardiographic results), and exang (exerciseinduced angina). Post pre-processing, the dataset is bifurcated into distinct training and testing subsets. Employing logistic regression, K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, and Random Forest models, the training data is utilized for model training. Subsequently, the models are rigorously tested using the designated testing set, providing a comprehensive evaluation of their predictive capabilities in identifying potential instances of heart disease.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
('age', (41, 63))
('sex', (2, 1))
('cp', (4, 3))
('trestbps', (49, 145))
('chol', (152, 233))
('fbs', (2, 1))
('restecg', (3, 0))
('thalach', (91, 150))
('exang', (2, 0))
('oldpeak', (40, 2.3))
('slope', (3, 0))
('ca', (5, 0))
('thal', (4, 1))
('target', (2, 1))
```

Attribute Selection: The data that is collected from the dataset of the patients may sometimes not be useful in the future. This process involves identifying and selecting the most highly used data or attributes from the dataset. This can be done using statistical methods or domain expertise.

Processing of Data: The data that is selected may be processed before they can be used in the classification step. This step may involve cleaning of the data, Handling of all the null or missing values.

Classification Techniques: Here in the Classification Techniques we choose and apply a machine learning algorithm. This may involve different machine learning techniques and from them the best will be selected. Some of the machine learning techniques that are used in this paper are logistic regression, support vector machines (SVM), random forests, K-nearest neighbors (KNN), and naive Bayes.

Disease Prediction: Once application of the machine learning algorithm is completed then the trained models will provided will the data set for prediction of the heart disease for the dataset of new patients.

Accuracy Measure: After the disease prediction process is complete, the algorithm with best best accuracy will be selected. This can be done by comparing the predictions of the model to the actual diagnoses of the patients.

IV. COMPARISION

Logistic Regression: Logistic regression, a prominent machine learning algorithm, is primarily employed for binary classification tasks. It leverages a logistic function that takes independent variables as input, yielding a probability value ranging from 0 to 1. In a scenario with two classes, say Class 0 and Class 1, if the logistic function output surpasses 0.5, the input is classified under Class 1; otherwise, it falls into Class 0. Despite the term "regression," it functions primarily for classification rather than regression. This algorithm extends from linear regression, tailoring its application to effectively address and solve classification problems by determining the likelihood of an input belonging to a specific class.

Here when the analysis was done the accuracy level was up to 85.25%. Hence Logistic regression is reliable in predicting the likelihood of a person having heart disease or not.

Support Vector Machine (SVM): SVM, a supervised machine learning algorithm applicable to classification and regression tasks, predominantly excels in classification scenarios. Its primary aim is to identify the optimal hyperplane within an Ndimensional space, effectively segregating data points across various classes in the feature space. The algorithm endeavors to maximize the margin between the nearest points of distinct classes, ensuring robust separation. The dimensionality of the hyperplane aligns with the number of input features; for instance, with two features, it forms a line, and with three features, a 2-D plane. Visualizing hyper planes becomes challenging asfeature dimensions exceed three, underscoring the algorithm's adaptability to diverse datasets.
Here when the analysis was done the accuracy level was up to 68.85%.

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806014

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Random Forest (RF): The Random Forest algorithm, widely employed in machine learning, amalgamates outputs from numerous decision trees to produce a singular result. True to its name, this technique contemplates multiple decision trees before arriving at a conclusive output, forming an ensemble of decision trees. Founded on the premise that a greater number of trees yield more accurate decisions, Random Forest utilizes a voting system for classification and computes the mean of decision tree outputs for regression. Particularly effective in handling large datasets with high dimensionality, this algorithm is adept at robustly addressing complex scenarios by leveraging the collective wisdom of diverse decision trees.
Here when the analysis was done the accuracy level was up to 85.25%.

K-Nearest Neighbors (KNN): KNN stands as a fundamental and crucial classification algorithm within the realm of Machine Learning, falling under supervised learning. Its applications span pattern recognition, data mining, and intrusion detection. Its widespread utility in real-world scenarios is attributed to its non-parametric nature, signifying an absence of assumptions about data distribution. The focus here is on training data to classify coordinates, assigning them to groups based on a specific attribute.
Here when the analysis was done the accuracy level was up to 63.93%.

Naive Bayes: Naive Bayes classifiers comprise a range of classification algorithms grounded in Bayes' Theorem. This family of algorithms operates on the fundamental principle of assuming independence between every pair of features being classified. It's not a singular algorithm but a versatile collection. The Naïve Bayes classifier stands out as a straightforward yet highly effective classification method, facilitating the swift creation of machine learning models known for their rapid and efficient prediction capabilities.
Here when the analysis was done the accuracy level was up to 85.25%.

V. RESULT

S.no:	Methodologies:	Accuracy:
01	Logistic Regression	85.25%
02	Random Forest	85.25%
03	Support Vector	68.85%
04	KNN	63.93%
05	Naive Bayes	85.25%

Here Logistic Regression, Random Forest and Naïve Bayes has more accuracy with 85.25%.

VI. CONCLUSION

A heart disease detection model has been developed using five Machine Learning classification modelling techniques. We developed a heart disease detection model employing five distinct Machine Learning classification techniques, aiming to identify the optimal classifier among them. Our comparative analysis reveals commendable performance across all classification methods when applied to the dataset. The machine learning approaches employed for heart disease prediction include Logistic Regression, Random Forest, Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). Notably, Logistic Regression, Random Forest, and Naive Bayes consistently outperformed others, achieving accuracies of 85.25%. KNN exhibited an accuracy rate of 63.93%, while Support

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Vector Machines achieved 68.85%. These results successfully fulfill the objective of enhancing prediction accuracy in heart disease detection. Future endeavors will focus on a more in-depth exploration of evolutionary computation techniques, aiming to assess their efficacy in addressing the complexities of the problem at hand. This ongoing research aims to further refine and optimize the heart disease prediction model for improved accuracy and reliability.

REFERENCES

1. Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath. "Heart disease prediction using machine learning algorithms". Published under licence by IOP Publishing Ltd. Citation Harshit Jindal 2021 IOP Conf. Ser.: Mater.Sci. Eng. 1022 012072 DOI 10.1088/1757-899X/1022/1/012072
2. C. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab, 2022,pp.16,doi:10.1109/ASET53988.2022.973 4880. Publisher: IEEE.
3. T. Chithambaram. "Heart Disease Detection Using Machine Learning" October 2020. DOI:10.21203/rs.3.rs-97004/v1.
4. K.Karthick, S.K.Aruna, Ravi Samikannu, Ramya Kuppusamy, Yuvaraja Teekaraman and Amruth Ramesh Thelkar. "Implementation of a Heart Disease Risk Prediction Model Using Machine Learning". Volume 2022 | Article ID 6517716.
5. Sonam Nikhar , A.M. Karandikar. "Prediction of Heart Disease Using Machine Learning Algorithms". [Vol-2, Issue-6, June- 2016]. ISSN : 2454-1311 .
6. K. Battula, R. Durgadinesh, K. Suryapratap and G. Vinaykumar, "Use of Machine Learning Techniques in the Prediction of Heart Disease," 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering(ICECCME),Mauritius,2021,pp.1- 5,doi:10.1109/ICECCME52200.2021.9591026
7. P. Sujatha and K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-7, doi: 10.1109/INOCON50539.2020.9298354.
8. A. G. Pemmaraju, A. Asish and S. Das, "Heart Disease Prediction Using Feature Selection and Machine Learning Techniques," 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India, 2022, pp. 28-33, doi: 10.1109/MLCSS57186.2022.00014.
9. R. Williams, T. Shongwe, A. N. Hasan and V. Rameshar, "Heart Disease Prediction using Machine Learning Techniques," 2021 International Conference on Data Analytics for Business and Industry (ICDABI), Sakheer, Bahrain, 2021, pp. 118-123, doi: 10.1109/ICDABI53623.2021.9655783.
10. Suraj Raut, Rishabh Magar, Rohan Memane, Prof. V. S. Rupnar, "HEART DISEASE PREDICTION USING MACHINE LEARNING",International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:23495162, Vol.7, Issue 6, page no.2081-2085, June-2020
11. Malavika G , Rajathi N , Vanitha V and Parameswari P,"Heart Disease Prediction Using Machine Learning Algorithms", e Vol 13 No 11 (2020).
12. Md. Julker Nayeem Nayeem, Sohel Rana, Md. Rabiul Islam, "Prediction of Heart Disease Using Machine Learning Algorithms", Published Nov 30, 2022.
13. A. Singh and R. Kumar, "Heart Disease Prediction using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.
14. Nouh Sabri Elmitwally Faculty of Computing, Engineering and the Built Environment at Birmingham City University, Birmingham, "Heart Disease Prediction Using Machine Learning Method"
15. Mr.Valle Harsha Vardhan, Mr.Uppala Rajesh Kumar, Ms.Vanumu Vardhini, Ms. Sabbi Leela Varalakshmi, Mr.A.Suraj Kumar, "HEART DISEASE PREDICTION USING MACHINE LEARNING "Vol 14 Issue 04,2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com