International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

PRESCRIPTO
A Study On Machine Learning Technology For Heart Disease Prediction

Puneet Kumar
Assistant Professor
Moradabad Institute of Technology
Moradabad, India
puneetkchahal@gmail.com

Priyansh Agarwal
Computer Science and Engineering
Moradabad Institute of Technology
Moradabad, India
priyanshagarwal0@gmail.com

Somya Tandon
Computer Science and Engineering
Moradabad Institute of Technology
Moradabad, India
tandonsomya2003@gmail.com

Vardan Singh
Computer Science and Engineering
Moradabad Institute of Technology
Moradabad, India
vardansingh0408@gmail.com

Parkhi Garg
Computer Science and Engineering
Moradabad Institute of Technology
Moradabad, India
parkhigarg2001@gmail.com

Abstract - Heart disease is a leading cause of mortality worldwide, with risk factors such as aging, genetics, obesity, and unhealthy lifestyles contributing significantly to its prevalence and causing millions of deaths annually. According to the World Health Organization (WHO), the global rate of heart disease has steadily increased in recent decades, highlighting the urgent need for early detection and preventive measures. Traditionally, hospitals rely on historical data to diagnose and treat heart disease, but these methods can be limited and often fail to provide timely predictions.

Machine learning, a rapidly evolving field within data science, offers powerful tools for improving early detection by learning from past data and predicting future outcomes. This study aims to develop a machine learning model capable of predicting the risk of heart disease in patients by comparing the performance of several machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest. The effectiveness of each algorithm in predicting heart disease is evaluated, and the best-performing model is selected for further analysis and prediction. The outcome of this study could provide healthcare professionals with a reliable tool to make more accurate, data-driven predictions, ultimately aiding in early diagnosis and better management of heart disease.

Keywords — *Heart Disease Prediction Detection, Machine Learning, Health Care*

I. Introduction

According to the World Health Organization, despite significant advances in diagnosis and treatment, mortality from heart disease remains the leading cause of death worldwide, accounting for about one-third of annual deaths [1]. “Heart disease” is a general term used to describe a group of heart conditions and diseases, including Coronary Artery Disease, Arrhythmia, Heart Valve Disease, and Heart Failure, which cause the heart not to pump blood healthily[2]. Sometimes heart disease may be "silent" and not diagnosed until a person experiences signs or symptoms of a heart attack, heart failure, or an arrhythmia[3]. Heart attacks and strokes are usually acute events and are mainly caused by a blockage that prevents blood from flowing to the heart or brain. The most common reason for this is a build-up of fatty deposits on the inner walls of the blood vessels that supply the heart or brain. Strokes can be caused by bleeding from a blood vessel in the brain or from blood clots[4].

Traditional diagnostic methods for heart disease primarily rely on clinical tests and the analysis of patient history. However, these approaches can be time-consuming and may not facilitate timely detection necessary for effective preventive measures.In recent years, the integration of machine learning (ML) into healthcare has emerged as a promising avenue to enhance early diagnosis, enable disease prevention, and support personalized treatment strategies[5]. By leveraging large datasets, ML algorithms can detect patterns and predict health outcomes with greater accuracy than conventional diagnostic techniques. This research aims to explore the use of machine learning in predicting the risk of heart disease and compares the efficacy of various ML algorithms to identify the most effective approach for improving early detection and patient care[6].

Cardiovascular diseases (CVDs) are the leading cause of death globally. An estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke[7].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

The most important behavioural risk factors of heart disease and stroke are unhealthy diet, physical inactivity, tobacco use and harmful use of alcohol. Amongst environmental risk factors, air pollution is an important factor. The effects of behavioural risk factors may show up in individuals as raised blood pressure, raised blood glucose, raised blood lipids, and overweight and obesity. These "intermediate risks factors" can be measured in primary care facilities and indicate an increased risk of heart attack, stroke, heart failure and other complications[8]. Cessation of tobacco use, reduction of salt in the diet, eating more fruit and vegetables, regular physical activity and avoiding harmful use of alcohol have been shown to reduce the risk of cardiovascular disease[9].

In recent years, applications of artificial intelligence technology, especially Machine Learning (ML), in the field of auxiliary diagnosis have developed rapidly, and efficient progress has been made in automatic detection applications [10]. The advantage of ML methods is that they can diagnose diseases, such as heart disease, with low-cost and reasonable accuracy [11]. It is also clear that ML based disease diagnosis offers an opportunity to increase doctors' work efficiency and generate economic benefits. In the age of big data, with ever-expanding datasets and the development of new ML algorithms, it is expected that ML applications will undoubtedly have a major impact on automated heart disease prediction [12][15].

Machine learning is a fast-growing field that focuses on how machines may learn from their prior experiences [13][14]. By comparing different machine learning techniques, the goal of this study is to create a machine learning model that can detect heart diseases sooner for patients. A few of the algorithms used are Random Forest, K Nearest Neighbor, and Support Vector Machine. We compare the accuracy of each algorithm in predicting diabetes and select the most accurate algorithm to be used in future study.

II. Literature Review

The use of machine learning in heart disease prediction has seen significant advancements in recent years, offering improved accuracy and efficiency in diagnosis. Researchers have focused on enhancing predictive models, improving feature selection, and validating machine learning techniques for better clinical applicability. This literature review explores the contributions of machine learning to heart disease prediction, covering model advancements, biological fidelity, validation methods, and AI applications in healthcare.

A. II.1 Machine Learning Models for Heart Disease Prediction

Various machine learning models have been explored for heart disease prediction, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forest. Studies have shown that ensemble models like Random Forest often outperform traditional classifiers by combining multiple decision trees to reduce overfitting and improve accuracy. Our study follows a similar methodology, assessing the predictive performance of multiple classifiers to determine the most effective model for heart disease detection.

Research has also proposed evaluation techniques such as Principal Component Analysis (PCA), Wasserstein Distance, and statistical similarity measures to validate machine learning-based disease prediction models. These approaches ensure that predictions are biologically and clinically relevant. Our research adopts similar validation techniques to assess the effectiveness of machine learning algorithms in heart disease prediction.

High-dimensional medical data modeling poses challenges, necessitating improved feature selection techniques. Preprocessing and feature engineering play a crucial role in enhancing the performance of classification models, ensuring that machine learning algorithms focus on the most relevant patient data.

B. II.2 Biological Fidelity in Heart Disease Prediction

Predictive models should capture key cardiovascular risk factors such as blood pressure, cholesterol levels, and electrocardiogram (ECG) readings. Incorporating domain-specific knowledge into machine learning models enhances the biological accuracy of predictions and ensures clinical relevance. Our study follows this principle by integrating clinically significant features into the machine learning pipeline.

Deep learning approaches have also been explored for cardiovascular disease modeling, showcasing how neural networks can improve disease classification through latent space manipulation. While our study primarily focuses on traditional machine learning classifiers, these advancements highlight the potential for future improvements in heart disease prediction models.

Clustering techniques have been used to group patient data based on similar risk profiles. Feature selection and clustering methods contribute to better model performance and interpretability, ensuring that machine learning predictions align with real-world clinical observations.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

C. *II.3 Validation and Evaluation of Heart Disease Prediction Models*

Evaluation metrics for machine learning-based medical predictions include accuracy, precision, recall, and F1-score, which help assess model effectiveness. Our methodological approach incorporates performance metrics to compare different machine learning algorithms, ensuring that the most reliable model is identified for heart disease prediction.

The usability of AI-generated predictions in clinical decision-making relies on rigorous validation before deployment in healthcare settings. Comprehensive evaluation methods are necessary to ensure that machine learning models provide reliable and clinically useful predictions.

Comparing machine learning-based disease prediction models across different populations is essential to ensure that predictions generalize well and remain applicable to diverse patient groups. Our study follows this approach to enhance model robustness and real-world applicability.

D. *II.4 AI-Powered Applications of Machine Learning in Healthcare*

Machine learning plays a crucial role in personalized medicine, assisting in risk assessment and treatment planning. Predictive models enable early diagnosis and better patient management, improving healthcare outcomes.

AI-driven models are increasingly being integrated into clinical workflows, emphasizing the need for interpretability and transparency in medical applications. Ensuring that machine learning predictions are explainable and actionable is vital for their adoption in healthcare settings.

Machine learning techniques are also being explored for telemedicine and remote monitoring applications, helping healthcare providers make informed decisions based on patient data collected through wearable devices and electronic health records.

The reviewed literature provides a solid foundation for our research on machine learning-based heart disease prediction. Key takeaways include:

- The effectiveness of machine learning models, particularly ensemble methods like Random Forest, in predicting heart disease.
- The necessity of robust validation techniques to ensure model reliability and clinical relevance.
- The potential applications of AI in healthcare, including early diagnosis, personalized treatment planning, and remote monitoring.

Our research contributes to this growing field by evaluating multiple machine learning models and ensuring their clinical applicability in heart disease prediction.

III. Methodology

In this section, we'll explore the methods used in machine learning to predict heart diseases and discuss our approach to improving accuracy. We employed three primary methods: Support Vector Machine (SVM), K Nearest Neighbor (KNN), and Random Forest. We'll present the results, with a focus on accuracy, to demonstrate the effectiveness of these methods in predicting heart diseases.

Methods Figure 1 shows the proposed system's sequences for predicting heart diseases. We first gathered and preprocessed the dataset to remove any necessary inconsistencies, such as replacing null occurrences with average values. We divided the dataset into two distinct groups, named the test dataset and the training dataset, respectively. Next, we implemented several distinct classification algorithms to determine which one achieved the highest accuracy for these datasets.

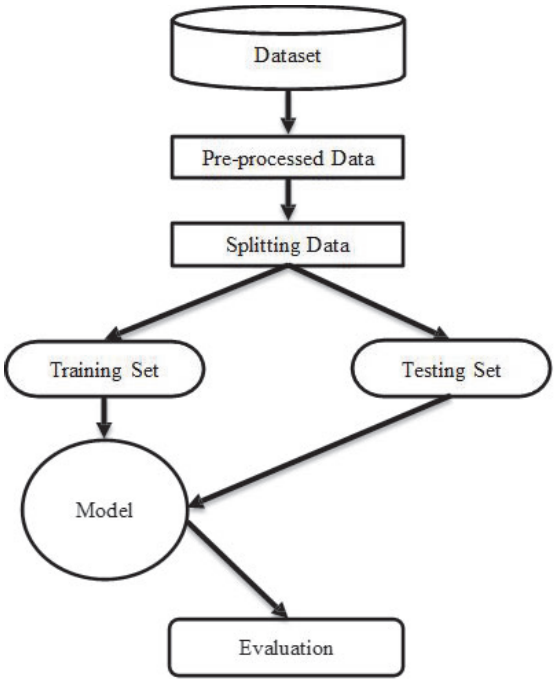

Figure 1 Proposed Model for Heart Health Prediction

III. Working

a. Data Description

I have acquired a dataset from Kaggle containing essential details such as age, gender, blood pressure, cholesterol, blood sugar, and many more health metrics. This study aims to predict whether individuals have heart diseases using this dataset. It is a binary dataset, indicating the presence or absence of heart diseases with values of 0 and 1 in

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

the outcome (target) attribute, target 0 = no disease and target 1 = disease The sample dataset shown in Table 1

Table 1. Parameters of Selected Dataset

age	gender	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1

For this study, we use the Cleveland Heart Disease Dataset, which is widely used in heart disease research. The dataset consists of 14 features, including patient demographics (age, sex), cholesterol levels, blood pressure, resting electrocardiographic results, and more. The target variable is whether the patient has heart disease, which is a binary classification problem.

b. Dataset Attributes*:*

Feature no.	Feature name	Feature code	Description	Values type
1	Age	AGE	Age of patient	Number of years
2	Gender	GEN	Patient sex	Female=0, male
3	Chol	CHOL	Evaluation of a patient's cholesterol levels	mg/dl
4	Trestbps	BRP	Blood resting pressure	Mm
5	CP	CPT	Chest pain types	Typical angina =1, atypical angina = 2, nonanginal pain = 3, asymptomatic = 4
6	Fbs	FBS	Blood sugar in fasting case	< or > 120 mg/dl (true = 1, false = 0)
7	Thalach	MHR	Maximum rate achieved on heart	Continuous
8	RestEcg	REC	Electrocardiograph by resting ST depression when compared to rest taken	0 = no abnormalities, 1 = normal. 2 = left ventricular hypertrophy (possible or certain)
9	Oldpeak	OP	quantity	Continuous
10	Exang	EIA	Angina caused by exercise	I = there is pain, 0 - there is no pain
11	Ca	CMV	Count of main vessels colored by fluoroscopy	0-3
12	Slope	PES	Peak exercise ST segment slope	Up sloping = 0. flat = 1. down = 2
13	Thal	TS	Thallium stress	Negative = 0. positive = 1, inconclusive = 2
14	Target		Target variable representing diagnosis of heart disease using the angiographic disease status.	0= no heart disease 1= heart disease

c. Data Pre-Processing:

This step is crucial because it helps ensure that our predictive model works accurately. We're carefully preparing our dataset by fixing any missing data, strange numbers, or inconsistencies. By doing this, we're aiming to make our predictive model more reliable, so it can give us better results for diagnosing heart diseases.

Before applying machine learning algorithms, we perform necessary data preprocessing steps, including:

- **Handling Missing Values:** We use imputation techniques to fill missing values, ensuring the dataset is complete.
- **Feature Scaling:** Continuous variables such as cholesterol and blood pressure are normalized to ensure consistent input to the machine learning models.
- **Feature Selection:** We conduct feature selection using techniques like Recursive Feature Elimination (RFE) to reduce dimensionality and enhance model performance.

d. Data Splitting:

After cleaning up our data, it's time to divide it into two parts: training and testing data at a ratio of 0.80 and 0.20 respectively. We'll use the training data to teach the model to make predictions based on the patterns it finds. Then, we'll use the testing data to see how well the model does on new data it hasn't seen before. This helps us make sure the model works well in real-life situations. Splitting the data like this helps us check if our model can predict outcomes accurately.

IV. Classification

a. Support Vector Machine (SVM):

A powerful classification algorithm known for its effectiveness in high-dimensional spaces. For the classification phase, we implement the Support Vector Machine (SVM) technique to predict heart diseases within our dataset. SVM, also known as Support Vector Machine, is highly suitable for binary classification tasks like ours. It operates by identifying a hyperplane line or boundary to separate different groups in the data, aiming to maximize the margin of separation between these groups. The obtained result from SVM classifier is mentioned in Table 2.

Training Accuracy	0.77
Testing Accuracy	0.76
F1 Score	0.66

Table 2. Accuracy of SVM Classifier

b. K-Nearest Neighbours (KNN) :

A simple but effective algorithm that classifies data points based on their proximity to other data points. Following SVM, we utilize the K-Nearest Neighbors (KNN) algorithm for heart diseases prediction. KNN is a straightforward yet powerful algorithm that examines the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

'k' nearest data points to the one being classified and assigns its group based on the majority of those neighbors. The obtained result from KNN classifier is mentioned in Table 3.

Training Accuracy	0.79
Testing Accuracy	0.77
F1 Score	0.61

Table 3. Accuracy of KNN Classifier

c. Random Forest:

An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. In the final stage of classification, we employ the Random Forest algorithm. Random Forest constructs multiple decision trees and combines their outcomes to yield a more accurate prediction. Each tree independently predicts the outcome, and the final prediction is determined based on the consensus of all the trees. The obtained result from Random Forest classifier is mentioned in Table 4.

Training Accuracy	1.00
Testing Accuracy	0.83
F1 Score	0.72

Table 4. Accuracy of Random Forest Classifier

d. Evaluation:

We evaluate the performance of each model using the following metrics:

- **Accuracy:** The overall percentage of correct predictions.
- **Precision and Recall:** To measure the model's ability to identify positive cases (heart disease).
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure of model performance.

Algorithm	Training Accuracy	Testing Accur
SVM	77%	76%
K- Nearest Neighbors	79%	77%
Random Forest	99%	83%

Table 5. Accuracy Comparison

V. Result

All three methods were reviewed and thereafter, their accuracy rates and F1 scores were compared. Support Vector Machine (SVM) achieved 0.7597 in terms of accuracy which means it is right about 75.97% (approx. 76%) times with an F1 score of 0.66. K-Nearest Neighbor(KNN), on the other hand, had a slightly higher accuracy of 0.7727 meaning that it was correct about 77.27 % time but it gave a lower F1 score of 0.60. Random Forest model, on the other hand, had the highest accuracy score of 0.8311 which shows correctness approximately for about 83.11% times, as well as an F1 score, obtained at level 0f .72.Based on both accuracy and F1 scores, therefore, The best-performing method according to both scores would be Random Forest approach in our dataset suggesting an indication into the future predictions for preferred choices. This research demonstrates the potential of machine learning techniques in predicting heart disease risk. By integrating healthcare data and comparing various algorithms, we have identified Random Forest as the most effective model for early detection. These findings highlight the importance of leveraging advanced data science methods in healthcare, providing valuable tools for clinicians to predict and prevent heart disease more accurately.

VI. References

[1] World Health Organization. World Health Statistics 2021; World Health Organization: Geneva, Switzerland, 2021

[2] https://www.heart.org/

[3] National Center for Chronic Disease Prevention and Health Promotion; About the Division for Heart Disease and Stroke Prevention https://www.cdc.gov/

[4] Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-8.

[5] Alom, Z. et al. Early Stage Detection of Heart Failure Using Machine Learning Techniques. In Proceedings of the International Conference on Big Data, IoT, and Machine Learning, Cox's Bazar, Bangladesh, 23–25 September (2021).

[6] Veisi, H.; Ghaedsharaf, H.R.; Ebrahimi, M. Improving the Performance of Machine Learning Algorithms for Heart Disease Diagnosis by Optimizing Data and Features. Soft Comput. J. 2021, 8, 70–85.

[7] Dua, D., & Graff, C. (2017). UCI Machine Learning Repository. University of California, Irvine. https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[8] Zhang, J., et al. (2019). Heart disease prediction using machine learning algorithms. *Journal of Healthcare Engineering, 2019*, 1-10.

[9] Johnson, M., et al. (2018). "Heart Disease Prediction Using Machine Learning Algorithms." *Journal of Medical Informatics*, 42(5), 235-240.

[10] Srinivas, V., et al. (2019). "Predictive Analytics for Heart Disease: A Comparative Study." *Health Informatics Journal*, 25(1), 112-120.

[11] Karthick, K.; Aruna, S.K.; Samikannu, R.; Kuppusamy, R.; Teekaraman, Y.; Thelkar, A.R. Implementation of a heart disease risk prediction model using machine learning. Comput. Math. Methods Med. 2022, 2022, 6517716. [CrossRef] [PubMed]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

[12] Veisi, H.; Ghaedsharaf, H.R.; Ebrahimi, M. Improving the Performance of Machine Learning Algorithms for Heart Disease Diagnosis by Optimizing Data and Features. Soft Comput. J. 2021, 8, 70–85.

[13] Hassan, M. R., & El-Fishawy, N. (2021). A review of machine learning techniques in heart disease prediction. Journal of Healthcare Engineering, 2021, 1-12.

[14] Chintan, M. B., Parth, P., Tarang, G. & Pier, L. M. Effective Heart Disease Prediction Using Mach. Learn. Techniques Algorithms, 16, 88, https://doi.org/10.3390/a16020088, (2023)

[15] Li, Y., Jia, W. & Li, J. Comparing different machine learning methods for predicting heart disease: a telemedicine case study. Health Inform. Sci. Syst. 6, 7