



# HEART DISEASE PREDICATION USING MACHINE LEARNING TECHNIQUES

Omkar singh<sup>1</sup>, Amit Kumar Pandey<sup>2</sup>, Rishidev Mishra<sup>3</sup>, Pooja Pandey<sup>4</sup>

<sup>1</sup>HOD(Department of DATA SCIENCE),<sup>2,3,4</sup>PG Students

<sup>1,3,4</sup>Department of DATA SCIENCE, <sup>2</sup>Department of IT ,

Thakur College of Science and Commerce

Thakur Village, Kandivali(East), Mumbai-400101, Maharashtra, India

**Abstract:** The death rate due to heart disease is one of the major global causes of increased mortality. Heart diseases are also called heart and blood vessel diseases. One of the main obstacles in clinical data analysis is to detect heart disease early. This underscores the need for effective early detection and prevention strategies these statistics continue to make cardiovascular disease (CVD) as one of the world's leading killer diseases such that its earlier detection and prevention interventions are warranted. The paper examines the use machine learning models in predicting heart disease by using support vector machines (SVM). In order to develop a model, a comprehensive dataset has been used which captures different attributes of patients including demographic information, lifestyle behaviours, and past medical history. SVMs are particularly suitable for large-scale datasets characterized by high dimensions; this approach allows identifying intricate patterns within sample sets without making any assumptions about how they may relate together or why such relationships exist between items found within your data array. Moreover, this study will critically assess and contrast ML algorithms with regards their sensitivity, specificity as well as overall accuracy. The results indicate that each algorithm exhibits distinct strengths in predicting heart disease, thereby providing valuable information to clinicians and healthcare professionals in applying strategies personalized risk assessment strategy. This research contributes to ongoing efforts to leverage ML for proactive and effective cardiovascular health management.

**Keywords:** Heart Disease, Cardiovascular disease(CVD), Support Vector Machines, Machine learning, Blood Pressure.

## I. INTRODUCTION

Heart disease is one of the leading causes of death in the world today and it greatly affects healthcare systems and society as a whole. It is important that heart diseases are detected early and their likelihood for occurrence accurately determined to enable prompt intervention and prevention of negative outcomes. The recent past has observed positive findings in the use of machine learning technologies in medical diagnosis and prognosis. This paper specifically focuses on how support vector machines (SVM) can be used to predict heart ailments. Our goal will be to build robust models with high accuracy rates that can identify subjects at risk of developing heart diseases through application of these advanced computational methods thus allowing targeted interventions and personalised healthcare strategies. SVM among others are the most commonly used machine learning algorithms due to its ability to handle linear as well as non-linear relationships with data which is appropriate for complex clinical datasets. In SVM, on other hand is good at devising the most suitable hyper plane to separate classes in high dimensions and hence, it can work well for

datasets with non-linear boundaries. In this study, we thus intend to compare their efficiency and performance vis-a-vis prediction capability, computational efficiency, and robustness of diverse machine learning approaches used. Accurate and dependable predictive models that help in detecting heart diseases have the potential of transforming clinical practice by providing proactive interventions and personalized patient care ultimately resulting into reduced global burden of cardiovascular diseases.

## II. LITERATURE REVIEW

These have culminated into an upped attention on early detection and prevention techniques as a result of the worldwide surge in heart disease incidence and its related mortality rates. The traditional clinical approaches for detecting heart diseases are slow and often inaccurate hence calling for more advanced computational methods. In this review, we examine machine learning (ML) models such as logistic regression, support vector machines (SVM), and random forests algorithms that have been used to forecast cardiovascular disease. The research paper of Vijeta Sharma, et.al. which deals with the important issue of early prediction of heart disease using machine learning classification techniques. The aim is that the different methods discussed classify heart diseases more effectively in order to attain greater accuracy and performance appraisal. Heart disease is one of the most common causes of death worldwide, which encompasses factors like lifestyle or genetic belongingness. Researchers have done a lot on various machine learning algorithms for example, SVM, DT, NB and RF on Cleveland Heart Disease Dataset aiming at predicting accurately the likelihood of getting heart diseases. After experimentations have been conducted SVM and Random Forest emerged as best performers recording 98% and 99% accuracy respectively surpassing Naïve Bayes and Decision Tree model. The study reveals the importance of machine learning in healthcare and why right selection of algorithms is paramount in accurate prediction. All in all, this research underscores how machine learning can be used to aid in early detection and management of heart diseases leading to better clinical decision-making and patient outcomes[1]. In Devansh Shah et.al. paper they have used data mining and machine learning techniques are used to predict heart disease. A dataset from the UCI repository is analysed, focusing on 14 vital characteristics related to heart disease. Different algorithms like Naïve Bayes, decision tree, K-nearest neighbor, and random forest are utilized and compared for accuracy. The research finds that K-nearest neighbor algorithm gives the highest accuracy. The research underscores the need for precise and early detection of heart diseases with a view to improving healthcare outcomes. Further study is recommended to discover more about other data mining techniques which can improve prediction accuracy[2]. The aim of research of Rahul Katarya and Sunit Kumar Meena is to explore how heart disease can be predicted by machine learning methods in view of modern lifestyles leading to the growing prevalence of cardiovascular diseases. The performance of different algorithms including Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, Decision Trees, Random Forest, Artificial Neural Network, and Deep Neural Network on relevant datasets was assessed. Findings show that Random Forest is the best performing algorithm with an accuracy of 95.60%, which shows it as a useful tool for predicting heart disease. This research emphasizes the need for early detection and management of heart issues. In future work, deep learning techniques may be implemented to improve prediction accuracies further[3]. The paper of Archana Singh and Rakesh Kumar relates the importance of correct forecasting of heart diseases due to their massive influence on human health. It delves into machine learning techniques like K-Nearest Neighbor, Decision tree, Linear Regression and Support Vector Machine (SVM) using UCI repository dataset. The study looks at the accuracy of these algorithms through Python programming in Anaconda (Jupyter) Notebook. Concentrating on supervisor learning methods including SVM, KNN algorithm is considered as the best approach for prediction of heart disease. Finally, the paper calls for additional studies to improve predictive capability while reducing deaths from cardiovascular ailments by raising awareness[4].

### III. METHODOLOGY

This study used methodology to evaluate the performance of various modeling techniques namely logistic regression, random forest and SVM on another data set called “Heart Disease Predication” for predicting Heart disease.

1. **Data Collection and Preprocessing:** Obtain a comprehensive dataset that incorporates demographic information (age, gender) as well as physiological parameters (blood pressure, cholesterol levels) from dependable sources like medical records or research databases. Do necessary steps in data preprocessing such as handling missing values, encoding categorical variables and scaling numerical features so as to make sure that the data across the dataset is uniform and compatible.
2. **Feature Selection and Engineering:** Perform exploratory data analysis (EDA) to learn about the distributional patterns and associations among features. Utilize correlation analysis, recursive feature elimination or importance ranking of attributes (for Random Forest) in selecting pertinent features that are highly significant in heart disease prediction. Employ feature engineering techniques for coming up with new characteristics or transforming existing ones, which may assist in capturing intricate relationships hence improving model accuracy.
3. **Model Training:** Train Logistic Regression models; Random Forest models and SVM models using training data respectively.

**Support Vector Machines (SVM):** Support vector machine is a supervised machine learning algorithm which determines the best hyperplane that separates data points into different classes with the aim of maximizing margin between the classes. SVM can be used for both classification and regression tasks. They are widely used in image classifications, text classifications and bioinformatics. Effective in high dimensions spaces, memory efficient versatile(different kernel functions), robust to overfitting.

#### DATA COLLECTION & DESCRIPTION:

SR NO	Attribute	Description
1	Age	Age of the patient in years
2	Sex	Gender of the patient(0=male, 1=female)
3	Cp	Chest pain type: 0: Typical angina 1: Atypical angina 2: Non-anginal pain 3: Asymptomatic
4	trestbps	Resting blood pressure in mm Hg
5	Chol	Serum cholesterol in mg/dl
6	Fbs	Fasting blood sugar level, categorized as above 120 mg/dl(1 = true, 0 = false)
7	restecg	Resting electrocardiographic results: 0: Normal 1: Having ST-T wave abnormality 2: Showing probable or definite left ventricular hypertrophy
8	thalach	Maximum heart rate achieved during a stress test
9	exang	Exercise- induced angina(1=yes, 0=no)
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	Slop of the peak exercise ST segment: 0: Upsloping 1:Flat 2:Downsloping
12	Ca	Number of major vessels(0-4) colored by fluoroscopy
13	Thal	Thalium stress test result: 0: Normal 1: Fixed defect

		2: Reversible defect 3: Not described
14	target	Heart disease status (0 = no disease, 1 = presence of disease)

#### IV. RESULTS AND DISCUSSION

The results of the experiments show that the highest level of accuracy belongs to **SVM, which achieves an accuracy rate of 86.89%**. Furthermore, through a detailed analysis precision, recall and F1-score reveals more about each algorithm's performances. In addition, confusion matrices show how well the classifiers identified those with heart disease as well as those without.

TABLE I. Accuracy Comparison

SVM			
	Precision	recall	F1-score
0	0.84	0.90	0.87
1	0.90	0.84	0.87
Accuracy- 86.89%			

DESCRIPTION:-

Class	Classification
0	No Disease
1	Disease

#### V. CONCLUSION

#### REFERENCES

- [1] Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. *SN Computer Science*. 2020 Nov;1:1-6.
- Sharma V, Yadav S, Gupta M. Heart disease prediction using machine learning techniques. In *2020 2nd international conference on advances in computing, communication control and networking (ICACCCN) 2020 Dec 18* (pp. 177-181). IEEE.
- M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, London, UK, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.
- 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.
- Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JM, Moni MA. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*. 2021 Sep 1;136:104672.
- Katarya R, Meena SK. Machine learning techniques for heart disease prediction: a comparative study and analysis. *Health and Technology*. 2021 Jan;11:87-97.
- Sangle PS, Goudar RM, Bhute AN. Methodologies and Techniques for Heart Disease Classification and Prediction. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 2020 Jul 1* (pp. 1-6). IEEE.
- Kavitha M, Gnaneswar G, Dinesh R, Sai YR, Suraj RS. Heart disease prediction using hybrid machine learning model. In *2021 6th international conference on inventive computation technologies (ICICT) 2021 Jan 20* (pp. 1329-1333). IEEE.
- Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*. 2018 May;29:685-93.
- Shanmugasundaram G, Selvam VM, Saravanan R, Balaji S. An investigation of heart disease prediction techniques. In *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA) 2018 Jul 6* (pp. 1-6). IEEE.