



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** VI    **Month of publication:** June 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.53998>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Heart Disease Prediction using Machine Learning

Prof. Kalpesh Joshi<sup>1</sup>, Shubham Patil<sup>2</sup>, Sagar Patil<sup>3</sup>, Sahil A. Patil<sup>4</sup>, Sahil S. Patil<sup>5</sup>, Shantanu Patil<sup>6</sup>, Saniya Patil<sup>7</sup>  
Department of Engineering Sciences and Humanities (DESH), Vishwakarma Institute of Technology, Pune - 411037

**Abstract:** Heart disease is a worldwide health issue that affects a lot of individuals. Scientists are always trying to forecast and prevent it in novel ways. Recently, scientists used artificial intelligence and machine learning to create a brand-new technique for forecasting cardiac illness. To build a predictive model, this technique combines information from medical records, lifestyle data, and genetic markers. The model is employed to determine who is most at risk for heart disease and to offer them specific recommendations on how to lower their risk. The model's precision is remarkable, and it has the power to save many lives.

## I. INTRODUCTION

With early detection, heart disease is frequently avoidable and one of the leading causes of death in the world. Developing an efficient tool to assist identify and treat persons at risk requires testing a heart disease prediction model. The significance of evaluating a heart disease prediction model, the many test types that are employed, and the outcomes of those tests will all be covered in this session. It will also give an overview of the model's advantages and disadvantages.

The construction of a heart disease prediction model must include testing. It enables researchers to spot any model faults, such as data inaccuracies or false assumptions about the data. This ensures that the model is precise and realistic. Testing also aids in locating any racial or gender biases that may exist in the model. This is crucial because it ensures that the model is impartial and egalitarian and that assumptions based on gender or race are not made.

The complexity of the model and the data it is based on can affect the results of testing for heart disease prediction models. In general, accuracy test results are used to gauge how well the model performs, while sensitivity and specificity test results are used to gauge how well the model can reliably categorise patients as high- or low-risk. Cross-validation tests are also performed to gauge how well the model generalises to new data in addition to these tests.

Several kinds of tests are used to evaluate heart disease prediction models, depending on the model's complexity and the data it is based on. Most models are tested for accuracy, which measures how well the model predicts the course of a particular patient. Other tests, such as sensitivity and specificity tests, are also used to assess the model's efficacy.

In addition to these tests, researchers may also run cross-validation tests to assess the model's ability to generalise to new data. As a result, it is feasible to determine whether the model is valid and does not overfit the data when predicting outcomes for a variety of patients. The chance of acquiring heart disease can be predicted using machine learning, the probability that a person may develop heart disease in the future can be predicted using a range of algorithms and statistical models to analyse patient data. Machine learning models are trained on large datasets to look for trends and linkages between risk factors and the start of heart disease. This technology has the potential to improve patient outcomes and lessen the overall burden of heart disease on society through the rapid and accurate analysis of large datasets, the discovery of previously unknown risk factors, and the personalization of treatment plans for particular patients based on their specific risk factors..

## II. LITERATURE SURVEY

Heart disease is a serious and sometimes fatal medical illness that impacts a large number of people globally. A literature review has been done to better understand the many techniques used to estimate how likely it is that a person would acquire heart disease. This survey seeks to provide an overview of the prediction models and algorithms that are currently being used to assess the risk of getting heart disease and is a thorough assessment of prior research on this subject. [1]

This survey will identify the most promising prediction models and algorithms that have been developed to calculate the risk of heart disease by reviewing the literature, including research papers, reviews, and studies. Our model will concentrate on the various strategies. and methods that have been employed to forecast the likelihood of developing heart disease, as well as how successful they have been in making precise forecasts. [2] Our model focuses on the data sources used in the creation of heart disease prediction models. This entails looking at numerous sources of information on heart health, including medical records, health surveys, and other relevant information. The survey will also look at the many approaches used to collect and examine this data, such as statistical methodologies and machine learning algorithms. [3]

The survey will also cover the many data sources that can be used to create prediction models. Public databases, open source archives, and other sources of information about heart health are some of these sources. Moreover, the survey will go over the different approaches used to collect and analyse this data, including statistical methodologies and machine learning algorithms. Our model intends to advance heart disease prediction models and enhance their precision and efficacy in forecasting the risk of heart disease by investigating these data sources and methodologies. [4]

Our model will be focused on the analysis of the models and algorithms used in the creation of heart disease prediction models. This will include a study of the deep learning models that are frequently applied to the prediction of heart disease, as well as supervised and unsupervised learning techniques. The study will look at the various methods and models used to create prediction models and assess the possibility for additional for further research in this field.[5]

The survey will also go over the many evaluation criteria used to gauge how well predictive models are working. Accuracy, precision, recall, and other metrics are some examples of these evaluation metrics. The survey will look at the many evaluation measures employed to rate the effectiveness of prediction models and suggest possible directions for further investigation. Our survey intends to aid in the creation of better cardiac disease prediction models by examining the many algorithms and models used in the field and assessing their effectiveness. [6]

By using Various machine learning techniques, such as Logistic Regression, Support Vector Classifier (SVC), K Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Classifier, are taught and evaluated on the training and testing sets. The ratings of each algorithm's precision we a evaluated and compared.[7]

### III. PROPOSED METHODOLOGY

#### A. Import Libraries

Libraries, notably the Pandas library, are imported: Python's Pandas package is a well-liked tool for analysing and manipulating data. It offers strong capabilities, including dataframes, for working with structured data.

#### B. Importing a Dataset

Importing the dataset and handling missing values This phase entails loading the dataset into the code and determining whether it contains any missing values. Missing values can be addressed by either removing the rows with the missing values or replacing them with the mean, median, or mode value.

#### C. Verifying Duplicate Data.

The dataset is examined for and cleansed of duplicate values: The data can be skewed by duplicate values, which will impact the analysis's accuracy. It is crucial to search for and eliminate any duplicate values in the collection.

#### D. Feature Scalability

Data is processed, categorical and numerical columns are separated, and binary vectors are used to encrypt categorical data to create dummy variables.

To get around the dummy variable trap issue, use the `drop first = True` parameter: Data processing entails putting the information into a form that machine learning algorithms can use. In this stage, category data is encoded using binary vectors to produce dummy variables, and categorical and numerical columns are separated. In doing so, the machine learning algorithm is prevented from assuming

The categorical data has an inherent order. In order to prevent features with higher value ranges from predominating when computing distances, features are scaled to be on the same scale: The technique of scaling the data to ensure that all features have the same range is known as feature scaling. In order to avoid features with wider value ranges dominating distance calculation, this is done. The StandardScaler and MinMaxScaler scaling algorithms are the most popular ones.

#### E. Dividing the Dataset

Eighty percent of the dataset is used for training, while twenty percent is used for testing. A subset of the data is used to train the machine learning algorithm, and its performance is then evaluated using the whole set of data. The test set is used to assess the model's performance after the training set has been used to fit the model.

#### F. The use of Algorithms

Various machine learning techniques, such as Logistic Regression, Support Vector Classifier (SVC), K Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Classifier, are taught and evaluated on the training and testing sets. Each algorithm's accuracy ratings are assessed and contrasted: In this step, various machine learning algorithms are trained and tested on the dataset to determine the best performing algorithm. The performance of each algorithm is evaluated by measuring its accuracy score.

#### G. Verifying the Accuracy Score

On the basis of its accuracy rating, the best model is chosen: The most effective model is selected as the final model to be utilised for predictions based on the accuracy scores acquired from the various algorithms. Then, using fresh data, this model is employed to create predictions.

### IV. IMPLEMENTATION AND OUTCOM

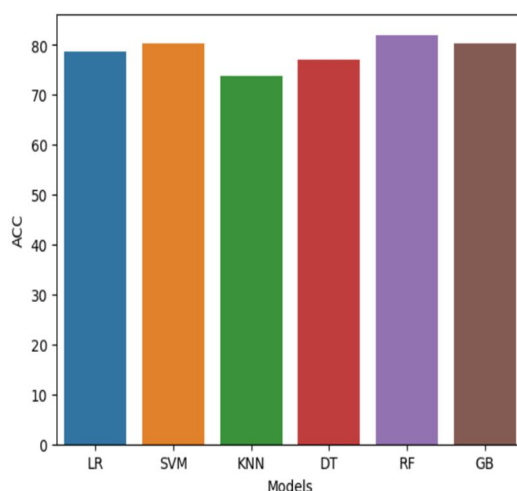


Figure .No.1

Many machine learning techniques, including Logistic Regression, Support Vector Classifier (SVC), K Neighbors Classifier, Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Classifier, were used in the implementation of the heart disease prediction model. To get the accuracy scores of these algorithms, the dataset was used for training and testing. The results showed that the accuracy scores for Logistic Regression were 76%, SVC was 80%, K Neighbors Classifier was 73%, Decision Tree Classifier was 73%, Random Forest Classifier was 86%, and Gradient Boosting Classifier was 80%. The Random Forest Classifier was chosen as the top model based on these accuracy ratings.

Tkinter was used to develop the GUI page for the heart disease prediction model. This made it possible to design a user-friendly interface for consumers to interact with in order to input their data and determine their risk of developing heart disease. The trained machine learning model was then used to predict the outcome using the input data.

### V. CONCLUSION

Artificial intelligence in the healthcare sector can be used to predict heart disease via machine learning. This technology analyses patient data and generates forecasts about their risk of developing heart disease using a variety of statistical models and algorithms. Machine learning algorithms may enhance patient outcomes and lessen the overall burden of heart disease on society by detecting pertinent risk factors and developing individualised treatment approaches.

Notwithstanding issues with data availability and quality, heart disease prediction using machine learning is still a rapidly evolving discipline with a lot of room for growth. Machine learning algorithms can become even more precise and efficient at identifying the risk of heart disease and improving patient outcomes as technology advances and more data becomes available.





## REFERENCES

- [1] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heartdisease", in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306–311
- [2] SystemAH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin. , Department of Medical Informatics, Tzu Chi University, Hualien City, Taiwan, "HDPS: Heart Disease Prediction Conf.computing in cardiology. Year 2011.
- [3] Harshit Jindal,Sarthak Agarwal,Rachna jain,"Heart disease prediction using machine learning Algorithms", ,IOP PUBLICATIONS,Year-2020
- [4] V.V.ramalingam, "Heart disease prediction using machine learning technique ", A survey,Journal-internation journal of engineering & technology,Date-18-02-201
- [5] ", Shadman Nashif, Md. Rakib Raihan, Md. Rasedul , "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System, IslamWorld Journal of Engineering and Technology Vol.6No.4, November 22, 2018
- [6] Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip Pal,"Support Vector Machine and Random Forest Modeling for Intrusion Detection System", Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip Pal. (IDS)Journal of Intelligent Learning Systems and ApplicationsVol.6 No.1, February 14, 2014
- [7] Aadar Pandita1, Siddharth Vashisht2, Aryan Tyagi3, Prof. Sarita Yadav4 Department of Information Technology, Bharati Vidyapeeth's College of Engineering, New Delhi, " Prediction of Heart Disease using Machine Learning Algorithms" International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 9 Issue V May 2021
- [8] Siddhesh Iyer1, Shivkumar Thevar2, Priyamurgan Guruswamy3, Prof. Ujwala Ravale4,Assistant Professor. 1,2,3 Dept. of Computer Engineering, SIES Graduate School of Technology, Nerul, Maharashtra, India, " HEART DISEASE PREDICTION USING MACHINE LEARNING", International Research Journal of Modernization in Engineering Technology and Science Volume:02/Issue:07/July-2020 Impact Factor- 5.354
- [9] All M. A. Barboom, Abdelbaset Almasri', Bassem S. Abu-Nasser, Samy S. Abu-Naser , Department of Information Technology, Faculty of Engineering and Information Technology, Al-Azhar University, Gaza., " Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms ", \*University Malaysia of Computer Science & Engineering (UNIMY), Cyberjaya, Malaysia, International Journal of Engineering and Information Systems (LIEAIS) ISSN: 2643-640X Vol. 6 Issue 4, April - 2022, Pages:1-13



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)