



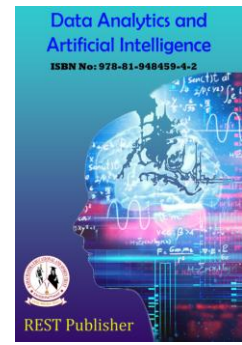
## Data Analytics and Artificial Intelligence

Vol: 3(2), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: <http://restpublisher.com/book-series/daai/>

DOI: <https://doi.org/10.46632/daai/3/2/12>



# Heart Disease Detection -A Machine Learning Approach

\*Josephine Reenamary .S, Rev. Sr. ArockiaValan Rani

ST . Joseph 's College of Arts and Science for Women, Hosur, Tamil Nadu, India.

\*Corresponding Author Email: [Josephinemary775@gmail.com](mailto:Josephinemary775@gmail.com)

**Abstract:** One of the human body's most important organs is the heart. It helps the body's blood to circulate and become cleaner. The global leading cause of death is heart attack. Chest discomfort, a faster heartbeat, and breathing problems were a few indications. The accuracy of this data was regularly checked. This publication presented a broad summary of heart attacks and current treatments. Additionally, a quick overview of the important machine learning methods for heart attack prediction that are available in the literature is provided. The machine learning techniques described include Decision Tree, Logistic Regression, SVM, Naive Bayes, Random Forest, KNN, and XG Boost Classifier. The algorithms are contrasted based on the braced of characteristics.

**Keywords:** Machine Learning, Heart Attack Prediction, KNN, Logistic Regression, SVM, Decision Tree, Naïve Bayes, Random Forest, XG Boost.

## 1. INTRODUCTION

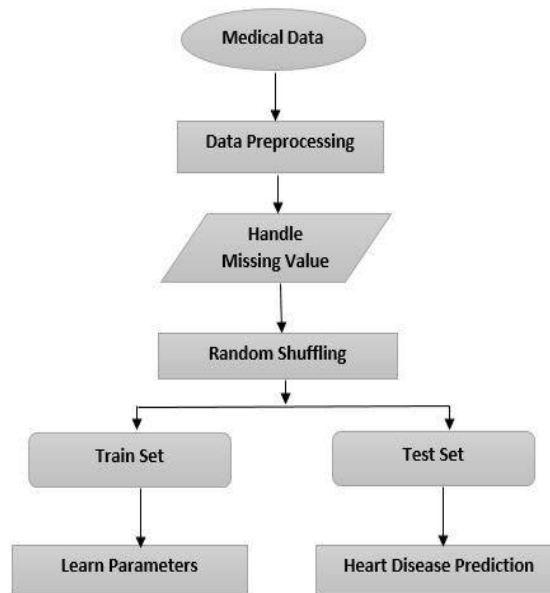
The human body's vast heart serves a vital function. Every person has to take care of their heart in order to retain health. Since many diseases have a connection to the heart, it is necessary to anticipate heart attacks. To do this, a comparative research in this area has been conducted. At the moment, the majority of patients die from heart attacks, which are often diagnosed at the very end of their illness. This is taking place since there aren't enough tools available to employ algorithms to forecast heart attacks with greater precision and efficiency. Initial heart attack prediction, whether the person is anticipated or not, is a challenge for the healthcare businesses in today's world. Researchers worked hard to develop a prototype that is successful in early heart attack prediction. they cannot construct a workable prototype, period. Each component of the structure has benefits and drawbacks of its own. Machine learning algorithms were trained to comprehend how to use and analyse data. The intersection of these two technologies is referred to as "machine intelligence." The definition of machine learning is that it learns from routine events and routine behaviours. All of you compared the accuracy of algorithms like Decision Tree, Logistic Regression, K-Nearest Neighbor, SVM, Naive Bayes, Random Forest, and XG Boost based on natural restrictions like analytical records, such as cholesterol, blood pressure, sex, age, and so on. During this investigation, they In this investigation, they determine which machine learning algorithm is the best based on the results by calculating the accuracy of seven distinct ones[1] [2].During the testing phase, a variable 80% accuracy is reached on the testing set. Putting information from earlier records to use practically takes time. Low accuracy rate. Therefore, they employed the Random Forest technique to create

## 2. LITERATURE REVIEW

The application of different Machine Learning algorithms in heart attack prediction systems has been the subject of several studies in the medical literature. JohnSanthana Krishnan [3] Making use of a machine learning system to forecast heart attacks In order to predict heart attacks in patients, this study applied categorization techniques. This article effectively describes heart attacks, including their veracity, common forms, and risk factors. This system uses a decision tree and Naive Bayes to forecast heart attacks. The most often employed prediction methods are decision trees, such as ID3 Algorithms and Naive Bayes Techniques. Naive Bayes classifier outperforms the others in terms of accuracy. Heart Attack Prediction using Effective Machine Learning Techniques was proposed by Avinash Golande [4], where specialists employ a few data mining rules to assist decision-makers or clinicians in differentiating between different types of heart attacks.

### 3. PROPOSED SYSTEM

Based on a few criteria, the system is suggested with the ability to anticipate frequent heart attacks. The suggested includes various blocks, as seen in Fig. 1 above.



**Data Acquisition:** The process of computing actual physical conditions and converting them into numerical numbers that a computer can manage is known as data acquisition.

**Data Pre-Processing:** Data pre-processing is the process of combining raw data for usage in a machine learning model. It is crucial to the creation of machine learning models, but while working on this project, they are unable to access the data to be cleaned and prepared. access to clean, ready data is not always available. when you don't always have access to cleaned-up, prepared data To enable everyone to access pre-processing services, data must first be cleansed and unnecessary data must be eliminated.

**Model Stacking:** All the regression classification models that may be employed in two layer estimators are gathered through the process of model stacking. The base line models are used to anticipate the results to create the first layer on the test data set. The input is used as the baseline model prediction by the layer two repressors or meta-classifiers, which then produces new output. Deterministic tree, Logistic Regression, Navie Bayes, KNN, Random forest, XG Boost, and SVM are the algorithms employed in this model.

**Logistic Regression:** Since the output of a probability between the dependent variable jumps between 1 and 0, logistic regression is a type of statistical model that is used for classification and predictive analysis. It calculates the likelihood that an event will occur based on an independent variable on the available data set. The odds, which represent the likelihood of success vs the likelihood of failure, are applied in this regression via the log it transformation. It's referred to as log odds. where  $p(\mu)=1/22$  is the midpoint of the curve, and  $p(s)$  is a location parameter. Hazra, A., Mandal, S., Gupta, A. and Mukherjee [7][8] "A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology

**K-Nearest Neighbor Classifier:** The K-Nearest Neighbor supervised learning approach assigns the new case to the group that is common to the available groups and uses a simple machine learning algorithm to forecast using the available examples and the new data. As a result, current data arises that can be easily categorised from a well-applied category by utilising the k closest classifier. This is because the KNN algorithm copies all of the current data and groups new data based on the comparable data. The distance between the data points is calculated using the Euclidean distance formula.

$$A \text{ and } B = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

**Decision Tree:** In addition to utility, event outcomes, and resource costs, a decision tree is a supporting tool that may be utilised as a tree comparable to decision making models and their workable outputs. A decision tree is one of the ways to illustrate a decision tree algorithm that can have conditional control statements.

Naïve Bayes: The Navie Bayes algorithm, a machine learning tool used to identify classification issues, was derived from the bias theorem. It is one of the simplest possible machine learning algorithms that has applications in many different sectors.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

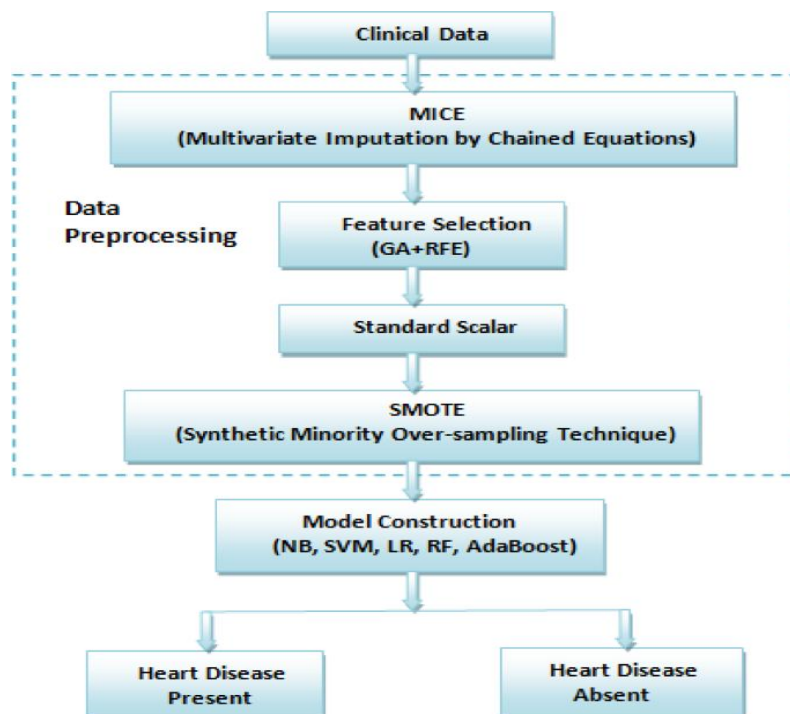
SVM (Support Vector Machine): One of the most effective exceptions to the supervised learning approach is the support vector machine (SVM). The fundamental goal of the SVM model is still to identify the best line or resolution boundary for segmenting hyper planes in N-dimensional space into distinct modules so that future additions or modifications of the statistics data may be easily inserted into the same type.

Random Forest: A well-defined model called random forest is utilised as a supervised machine learning approach. Everyone in this model uses this method for both regression and categorization difficulty The random forest model makes use of the ensemble learning concept. One of the classifiers, it determines the anticipated accuracy of the provided data set based on the supplied label and accommodates a variety of subsets of the given data set in the decision tree.

XG Boost: The boosted tree techniques' processing power limit is pushed by the flexible and highly accurate gradient boosting executable known as XG Boost. It is mostly utilised for rapid computing and also energises performance.

Performance Evaluation: One of the most crucial aspects of the machine learning process is performance evaluation, which requires meticulous execution. Data resampling, performance assessment, and outcomes data with statistical significance are the three primary subtasks that are being evaluated.

Heart Attack Prediction: After completing the aforementioned steps, users receive the forecast for their own input, and as a result, the project's anticipated conclusion will be the prediction of an accuracy score for a certain dataset and whether or not the patient should be diagnosed with a heart attack. Kailash Devadkar[11]Prediction of Heart Disease using Machine Learning



Symptoms: Typical signs of a heart attack include: Soreness in the chest that may be pressure, tightness, pain, hurting, or stiffness, spreading to the shoulder, arm, back, neck, jaw, teeth, or even the upper belly.

- Chest pain & chest burning
- Tingling & Numbness
- Nausea & Vomiting
- Breathless
- Palpitation
- Sweat & chillness

**Information on current medical condition:**

- Total cholesterol level (Tot Chol)
- Systolic blood pressure, or Sys BP
- Diastolic blood pressure, or Dia BP
- Body Mass Index (BMI)
- Heart Rate: Although heart rate is really discrete, it is still referred to be continuous in medical research due to the wide range of potential values.
- Glucose: The amount of blood sugar

**4. RESULT ANALYSIS**

The main objective of this study is to control the likelihood that someone may experience a heart attack. And offer suggestions about how to move forward. With the Random Forest method, great accuracy rates are achievable. They utilised the following data set, as an example:

- The data in Table 1 is sufficient to assess whether or not a person is at risk of having a heart attack.[9][10] Each statistic in the data set results from cardiac functions.
- For instance, the four tenets of the kind of chest pain (cp). (1. Symptomatic angina 2. An unusual case of gina
- Non-angular pain 3. 4. Symptomless Table 1 lists the features that the statistics set reflects.
- Trestbps- Plasma pressure level in the resting mode.
- Cholesterol is measured in mg/dl in serum.
- fbs - Fasting blood sugar levels (if greater than 120 mg/dl, displayed as 1, otherwise as 0).
- Exang – Exercise-induced angina (0–No, 1–Yes); Old Peak – Exercise-induced ST depression as relation to resting state.

Shows that the application created using the Random Forest technique has a greater level of precision than additional approaches.

**5. CONCLUSION AND FUTURE SCOPE**

A potent collaborative learning system for regression and classification operations is the Random Forest algorithm. The technique builds N decision trees, and it then returns the session that represents the average of all the outputs from the decision trees. As a consequence, accurate early prediction is efficiently achieved. The management of medical data, particularly those pertaining to the heart, can help in the early detection of heart attacks or abnormal cardiac problems, ultimately saving lives.

Age	49	64	43	69
Cp	1	3	2	0
Trestbps	120	150	172	135
Chol	239	219	283	233
Fbs	0	1	0	1
Thalach	178	163	174	114
Exang	0	1	0	1
Old peak	1.4	0.6	174	0.8
Thal	1	2	0	2
Target	0	1	1.8	1

It 0 is quite difficult to anticipate heart attacks in the modern society. By entering the report standards, the patient or user can use this programmed to predict a heart attack if they are unable to reach a surgeon. And can choose whether to seek medical help or not. Future Scope: Future updates to this programmed might include new features, such the ability to notify the patient's family members if a heart attack is expected. The nearby hospital must also receive the information. Online doctor discussions with the other doctor would be another option offered. It's important to note that ML applications built on a variety of effective algorithms are applied not only to the study and prediction of heart attacks, but also to radiology, bioinformatics, and medical imaging analysis.

## REFERENCES

- [1]. Azizkhan F Pathan, Chetana Prakash, Attention-based positionaware framework for aspect-based opinion mining using bidirectional long short-term memory, *Journal of King Saud University - Computer and Information Sciences*, 2021, ISSN 1319-1578.
- [2]. Jayaprakash, S., Nagarajan, M.D., Prado, R.P.D., Subramanian, S. and Divakarachari, P.B., 2021. A systematic review of energy management strategies for resource allocation in the cloud: Clustering, optimization and machine learning. *Energies*, 14(17), p.5322.
- [3]. Mr.SanthanaKrishnan.J, Dr.Geetha.S,” Prediction of Heart Disease Using Machine Learning Algorithms”,2019 1st International Conference on Innovations in InformationandCommunicationTechnology(ICICT),doi:10.1109/ICICT1.2019.8741465.
- [4]. AvinashGolande, Pavan Kumar T,” Heart Disease Prediction Using Effective Machine Learning Techniques”, *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878,Volume-8, Issue-1S4, June 2019.
- [5]. V.V. Ramalingam, AyantanDandapath, M Karthik Raja,” Heart disease prediction using machine learning techniques: a survey”, *International Journal of Engineering & Technology*, 7 (2.8) (2018)
- [6]. Hazra, A., Mandal, S., Gupta, A. and Mukherjee, “ A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review” *Advances in Computational Sciences and Technology* , 2017.
- [7]. M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, “Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools” *International Journal of Scientific Research in Computer Science,Engineering and Information Technology* ,IJSRCSEIT 2019.
- [8]. Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Diseases Prediction”, 4th International Conference on Computing Communication And Automation(ICCCA), 2018.
- [9]. Amandeep Kaur and Jyoti Arora,“Heart Diseases Prediction using Data Mining Techniques: A survey” *International Journal of Advanced Research in Computer Science* , IJARCS 2015-2019.
- [10].Senthil kumarmohan, chandrasegarthirumalai and Gautam Srivastva, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” *IEEE Access* 2019.
- [11].Aditi Gavhane, GouthamiKokkula, Isha Panday, Prof. Kailash Devadkar, “Prediction of Heart Disease using Machine Learning”, *Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA)*, 2018.