# Heart Disease Prediction Using Artificial Intelligence And Machine Learning

**Kiran Pustode[1], Rahul Navkhare[2]**

[1]*Student of M-Tech Artificial Intelligence & Data Science Engineering Department in WCEM, Nagpur*
[2]*Professor of M-Tech Artificial Intelligence & Data Science Engineering Department in WCEM, Nagpur*

---***---

**Abstract -** Heart disease is one of the most significant causes of global mortality since its intricacy and the rate of misdiagnosis have brought a great challenge to medical workers. As machine learning has shown robust efficacy in decision-making and predictions, it is essential to construct a machine learning model to assist with heart disease diagnosis. The cost of healthcare is outpacing both corporate and governmental budgets as a result of silent illnesses like cardiovascular disorders. Early diagnosis and treatment of these illnesses are therefore desperately needed. One of the newest and most popular technologies is machine learning, which is being used to anticipate diseases in a variety of industries worldwide, including healthcare. The purpose of this study is to use logistic regression to forecast the overall risks and find the most important heart disease predictors. In order to find the predictors, the binary logistic model, one of the classification techniques in machine learning, is employed in this work.

***Key Words*:** Machine learning, Logistic regression, classification algorithms, heart diseases

## 1. INTRODUCTION

The term heart disease refers to a broad range of disorders that impair arteries, blood vessels, and other organs and cause abnormal heart function. The SARS-CoV-2 virus significantly impairs heart health and may even cause heart destruction because it interferes with the human respiratory system and tries to reduce oxygen levels in the lungs. Heart disease is indicated by the formation of atheromatous plaques, abnormal lipid metabolism, and the accumulation of lipids and other substances in the coronary artery circulation. It may lead to luminal narrowing or occlusion, which may result in oxygen deprivation, myocardial ischemia, or necrosis, presenting as symptoms such as tightness in the chest, discomfort in the chest, and myocardial infarction. In the modern world, heart disease is the primary cause of death. The World Health Organization estimates that cardiac conditions will kill 12 million people globally year. 8.9 million people died as a result of coronary heart disease in 2017, according to 10.6 million new cases reported worldwide. 126.5 million persons worldwide suffered from coronary heart disease in 2017. In the US, medical costs related to heart disease are projected to increase by 41% between 2010 and 2040, from $126.2 billion to $177.5 billion.

The healthcare sector is undergoing a dramatic transformation because to machine learning (ML) applications. A subset of artificial intelligence (AI) technology called machine learning (ML) aims to improve the accuracy and productivity of medical professional's work. For nations dealing with a physician shortage and an overburdened healthcare system, machine learning presents a great potential. Because it can identify patterns in large data sets and make it easier to identify diagnostic markers linked to risk or disease, it is crucial to medicine. Machine learning techniques can support clinical management and experts investigate superior performance in numerous medical applications, including medical image analysis, language processing, and tumor or cancer cell detection. Algorithms for machine learning categorization are now being used in clinical therapy. Techniques for classification can be used to extract knowledge. Practitioners can enhance patient outcomes and make better judgements by using accurate cardiac disease prediction. Machine learning-based techniques, such as support vector machines (SVM), K-nearest neighbors (K-NN), artificial neural networks (ANN), decision trees, logistic regression (LR), AdaBoost (AB), Naive Bayes (NB), fuzzy logic (FL), etc., have been widely used in the early diagnosis of cardiac disease. The ratio of mortality from heart disease has decreased because to these machine learning-based expert medical decision systems.

We need to enhance the performance of the previous work design by employing a preprocessing technique that comprises a normalization step that uses the mean value of each feature to fill in any missing data. Data preparation, an essential stage in the machine learning process, enhances the quality of the input data and extracts valuable information. The primary goals of that work are to assess several machine learning algorithms for the detection of heart illness and the prognosis of coronary artery disease. A machine can learn without needing to be explicitly trained to do so thanks to a process called machine learning. A subfield called artificial intelligence makes use of sophisticated software to enable gadgets to carry out jobs competently. We use logistic regression to get around classification problems. It accomplishes this by forecasting categorical outcomes as opposed to utilising linear regression, which indicates continuous improvement. The prediction of heart disease, which has been progressively increasing in prevalence globally, is an example of a binomial, which has two possible outcomes in the most basic case. Using logistic regression improves accuracy in comparison to the existing method. In the modern healthcare industry, machine learning is commonly utilised to diagnose and forecast the existence of diseases using data models.One such widely used machine learning approach for investigations involving complex illness risk assessment is logistic regression. Thus, the goal of the study is to use logistic regression to estimate the overall risk and identify the most important predictors of cardiovascular illnesses.

## 2. BACKGROUND

The dataset from a continuing cardiovascular study in Framingham, Massachusetts is available on the Kaggle website (https://www.kaggle.com) and was used for the logistic regression analysis. This study's classification objective is to determine the patient's 10-year risk of developing heart disease in the future. There are 15 attributes and 4238 patient entries in the Framingham dataset. Python programmers can perform data analysis using more potent and versatile data science platforms. Logistic regression is a binary classification algorithm commonly used for predicting the probability of an instance belonging to a particular class. In the context of heart disease prediction, logistic regression can be employed to predict the likelihood of an individual having heart disease.

## 3. LITERATURE REVIEW

Ramat Enki Sateesh et al. have proposed an ensemble classifier using Support Vector Machine, K Nearest Neighbor, Multiple Kernel KNN Classification and Regression Trees and achieved an accuracy of 93.41%. Rutuja Gujare, D.Viji, Simran Bhatt have proposed a model to predict heart disease using ensemble classifier . Bagging, Boosting and voting were used in this model and analyzed also. In another ensemble model to predict heart disease has been proposed with Linear Discriminant Analysis , Classification and Regression Trees , Support Vector Machines, K-Nearest Neighbors and Naïve Bayes classifier are used as base learners . I. Yekkala, S. Dixit and M. A. Jabbar have proposed a model to predict heart disease using ensemble and Practical swarm optimization technique. In the proposed model to predict cardiovascular disease has achieved an accuracy 87.04%. C. Beulah Christalin LathaS. Carolin Jeeva have proposed a model to predict cardiovascular disease based on bagging method and achieved 7% increased accuracy than base classifiers. Xiao-Yan Gao, et al. have analyzed heart disease prediction mechanism using ensemble learning. KNN, DT, RF, and NB have used in the proposed model authors have proposed a model to predict heart disease using Bagging mechanism and achieved accuracy 81.41%. Author has proposed a model to predict heart disease using bagging ensemble method. Base classifiers used are Decision Tree, Logistic Regression, Artificial Neural Network, Naive Bayesian and K Nearest Neighbor.

## 4. LOGISTIC REGRESSION MODEL

Logistic regression is a statistical method that is used for building machine learning models where the dependent variable is dichotomous: i.e. binary. Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables. The independent variables can be nominal, ordinal, or of interval type. The name "logistic regression" is derived from the concept of the logistic function that it uses. The logistic function is also known as the sigmoid function. The value of this logistic function lies between zero and one. The following is an example of a logistic function we can use to find the probability of a vehicle breaking down, depending on how many years it has been since it was serviced last.

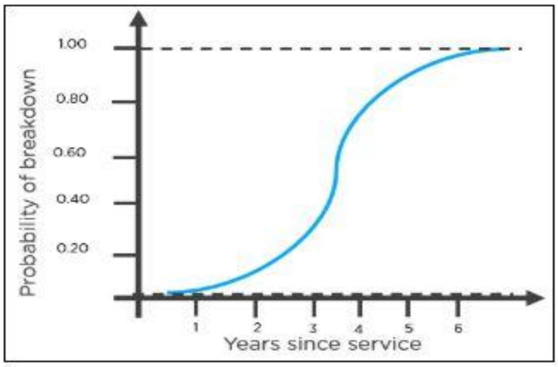


**Fig -1: Logistic Regression**

Despite its name, logistic regression is not a regression model rather, it is a classification model. For binary and linear classification issues, logistic regression is a straightforward and more effective approach. It's a classification model that performs exceptionally well with linearly separable classes and is quite simple to implement. In the industry, this method is often used for classification. Similar to the Adaline and perceptron, the logistic regression model is a statistical technique for binary classification that can be extended to multiclass classification.

Logistic regression is a supervised machine learning algorithm that predicts the probability of an outcome, event, or observation. The model delivers a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false. Logistic regression works by computing a sum of the input features and calculating the logistic of the result. It analyzes the relationship between one or more existing independent variables to predict a dependent data variable.

## 5. METHODOLOGY

This method was developed with the main objective of estimating the risk of heart disease 10 years from now. We employed a range of feature selection techniques, such as logistic regression and backward elimination, as a machine learning method to train our system. These algorithms' specifics are discussed below.

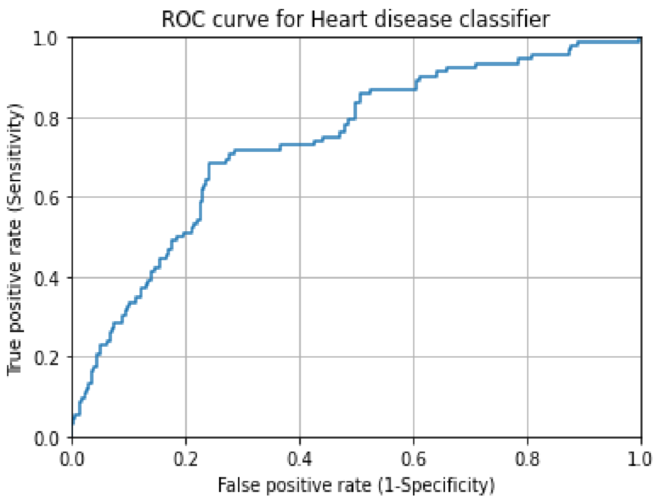

**Fig -2: ROC Curve for Heart Disease Classifier**

The 4240 observations in the Kaggle dataset are used to predict the risk of heart disease in a given location for a given patient. In this work, we used the patient data to predict heart disease using the SK Learn programmed. This patient data will be useful in predicting the course of heart disease over the next ten years. Preprocessing and loading were done with the data that was gathered. The acquired data was used for pre-processing and data loading. The preparation process involves clearing the primary error and any unnecessary information from the database. Finding missing data in a database is another application for this method. After that, information relevant to the prognosis of heart conditions is extracted via feature selection.

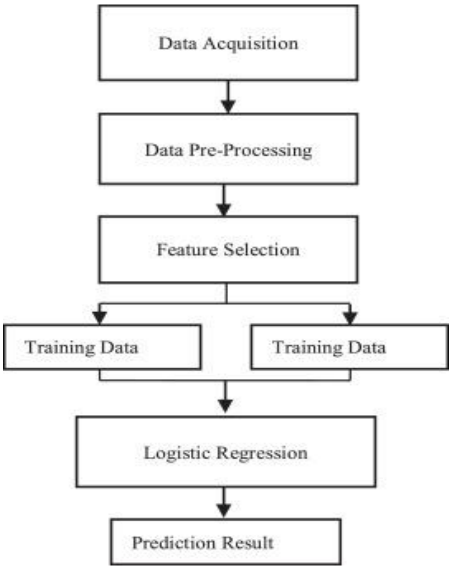


**Fig -3: Workflow of Logistic Regression Model**

### 5.1 Data Pre-Processing:

In order to build up more accurate ML model, data preprocessing is required. Data pre-process is the process of cleaning the data. This includes identification of missing data, noisy data and inconsistent data.

Exploratory Data Analysis is used to analyze the processed data. This stage ascertains whether a predictive model is a workable analytical instrument to accurately complete a given assignment. The user can then obtain an accurate result by refining the data to make sure they are categorised and relevant. The information needed to make the forecast has been divided. Logistic regression is used to train the separated data before it is tested with test data. Logistic regression is used to get the exact findings.

## 6. FUTURE DIRECTIONS

To enhance the efficacy of Logistic Regression models, future research should focus on incorporating advanced feature engineering techniques, optimizing hyperparameters, and exploring ensemble models. Additionally, investigations into real-time implementation and the integration of diverse data sources are vital for practical application. Effective feature selection and engineering are critical for model performance. Literature emphasizes various approaches, including statistical tests and domain knowledge integration, to identify relevant features for Logistic Regression models.

## 7. CONCLUSION

The reviewed literature underscores the significance of Logistic Regression in heart disease prediction using AI and ML. While acknowledging its limitations, Logistic Regression remains a valuable tool, particularly in scenarios where model interpretability is paramount. As the field continues to evolve, a holistic approach considering various ML algorithms and data sources will contribute to more robust and accurate predictions.

## REFERENCES

1. Mozaffarian, D., Benjamin, E., Go, A., Arnett, D., Blaha, M.Cushman, M. et al. (2015). Heart Disease and Stroke Statistics—2021 Update. Circulation, 131(4).
2. Divyansh Khanna, Rohan Sahu, Veeky Baths, and Bharat Deshpande Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease, International Journal of Machine Learning and Computing, Vol. 5, No. 5, October 2020.
3. Dr. T Lalitha, future prediction of heart disease through exploratory analysis of data, smart green connected societies, vol. 1 no. 01, 2021
4. Dinesh Kumar G, Prediction of cardiovascular disease using machine learning algorithms, preceding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, 2019.
5. A, S Thanuja Nishad papers, Predicting Heart Diseases in Logistic Regression of Machine Learning Algorithms by Python Jupyter lab, International Journal of Advanced Research and Publications,Vol3, Number 8, 2021.
6. Miguel-Hurtado, O., Guest, R., Stevenage, S., Neil, G., & Black, S. (2019). Comparing Machine Learning Classifiers and Linear/Logistic Regression to Explore the Relationship between Hand Dimensions and Demographic Characteristics.
7. Abhijna Bhat, Pragathi, Pranamya M S, Smitha, Prediction of Heart Disease Using Logistic Regression, International Research Journal of Engineering and Technology, Vol 07, Issue: 06June, 2022.
8. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee, Asmita Mukherjee, Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review, Advances in Computational Sciences and Technology Vol 10, Number 7, 2020.