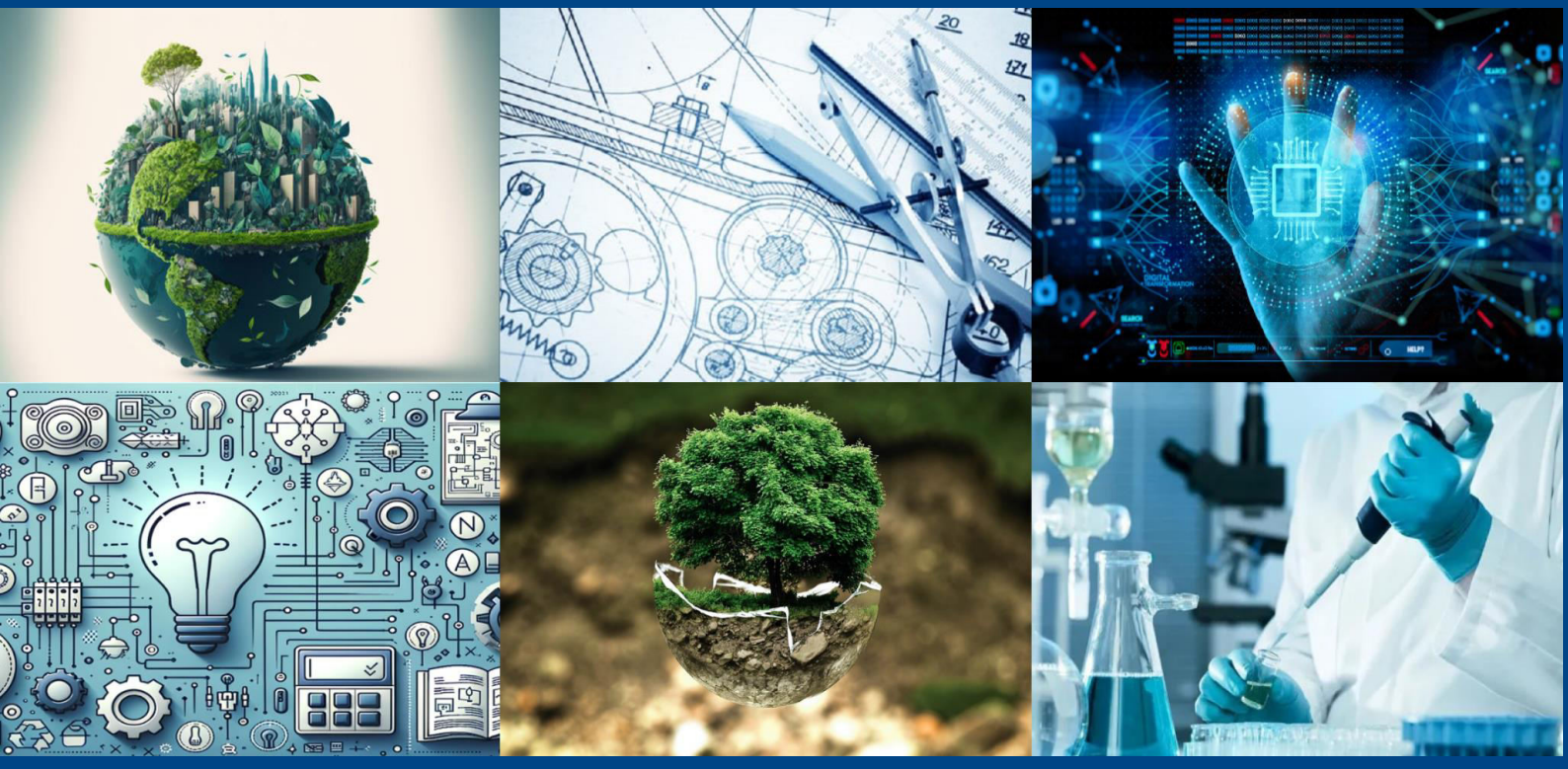
ISSN: 2582-7219

# International Journal of Multidisciplinary Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.206** **Volume 8, Issue 5, May 2025**

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805266

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Cardio Predict: Machine Learning- Powered Heart Disease Detection

**Akash A. Borge, Prof. Sonali Mutha**

Master of Computer Applications, Anantrao PAwar College of Engineering and Research, Pune, India

**ABSTRACT:** This study aims to create an applicative machine learning model in order to predict heart disease using a patient's medical report. We pre-processed the data, splitting it to train and test, tuning the hyper parameters fine and could reach an accuracy of about 86%. The findings show that machine learning has great promise in medical diagnosis, according to which the cutting-edge technology is a useful tool for health care professionals as in screening and diagnosing heart disease.

**KEYWORDS**: Heart disease prediction, machine learning, random forest classifier, medical diagnostics, hyper parameter tuning.

## I. INTRODUCTION

Cardiac disease is still one of the most common causes of death worldwide. Early detection and prevention are both crucial in influencing patient outcomes and lessening the burden of these diseases. Conventional diagnostic techniques are frequently inadequate for handling the complexity of medical data, and are liable to miss nuanced risk factors. This is where machine learning (ML) comes in. Machin learning offers new techniques to investigate large, complex datasets in search for underlying patterns which otherwise would be indiscernible. In spite of the advances in medical technologies, it is still difficult to precisely predict heart disease on account of the multi-aspect of the clinical data. In this article, we create a machine learning model which can predict the tendency of getting heart disease with a wide range of patient information. We empirically measure how well ML algorithms work when applied to the most.

### Machine Learning Model Development for Heart Disease Prediction

In this section, we'll walk you through how we worked with our data and built our machine learning model step by step. First, we explored the dataset to understand the most important features using basic statistics and visualizations. Then, we cleaned and prepared the data to make sure it was ready for analysis. After that, we trained several machine learning models to see which one could best predict the risk of heart disease. We checked how well the models performed using measures like accuracy and precision. In the end, we fine-tuned our model and got it ready for real-world use, especially in healthcare, where this kind of tool can help doctors make better decisions.

## II. LITERATURE REVIEW

Experience-based researches have been carried out by researchers who exploit machine learning to predict cardiovascular diseases over the years. Numerous research works report that ML models are superior to traditional diagnostic approaches, unearthing potential patterns in medical data.

In a well known study, Detrano et al. applied logistic regression model to predict heart disease, obtaining reasonable accuracy but was not suitable for learning non-linearly correlated data. More recently, advanced models, such as Random Forest, Support Vector Machines (SVM) and Neural Networks, have been applied, with higher levels of accuracy and generalization performance.

Some UCI researchers tried Decision Trees and K-Nearest Neighbor (KNN) (Holden et al., 2020) on the same dataset and found that Random Forest was consistently better, not only in terms of accuracy, but also in stability (Holden et al., 2020).

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805266

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**Objectives**

- Your task is to reproduce the results - Examine the heart disease dataset and do some investigation of what the significant features are and what patterns might exist.
- Prepare the data - Organize the data, correcting for missing or inaccurate data and ensuring it's in the proper format for processing.
- Train the model - Apply different machine learning algorithms and have them learn the patterns on whether a person has a heart disease or not.
- Evaluate its performance - Test the models to determine how well they perform, how accurate and how predictive they are, using things like accuracy and precision.
- Only keep the best one - Evaluate the models and keep the one that performs the best, and then do further fine-tuning.

## III. METHODOLOGY

We took the UCI heart disease data set, including age, cholesterol, type of chest pain and looked at significant health data. The data was prepared, which fills missing and codes for the variables. The numerical variable was expanded to be comparable.

After dividing the data into training (80) and (20) tests, we created a searching analysis to determine the convenience relationship. Statistical and model -based methods were used for important convenience choices.

Various machine learning algorithms (logistics regression, SVM, Random Forest, KNN, Grade -Boosting) were tried. The performance was expanded by setting hyperparameters with networks and random discovery.

**Implementation**

We tested several models, including Logistic Regression, Random Forest, and XG Boost. Each model was trained, fine-tuned using Grid Search, and evaluated using accuracy, precision, recall, and other metrics.
Finally, we created a simple interface (with Stream lit) where users can enter health details and get instant predictions about heart disease risk.

**Machine Learning Model Development for Heart Disease Prediction**

This is the section where we guide you through exactly how we manipulated our data and created our machine learning model. First, we conducted a data look-see to see which features looked most significant based on some elementary stats and vis. Next: We cleaned up and formatted the data to ensure that it was prepared for analysis. We then trained various machine learning models to try to figure out which could best predict the risk of heart disease. We tested how good the models were with things like accuracy and precision. In the end, we refined our model and prepared it for real-world deployment, particularly in health care, where such a tool can assist doctors in making better decisions.

**Suggestions to the challenges**

- Missing or messy data -

Once in a while, the dataset included an empty or wrong value something that could throw the model off.

 What we did: We filled in missing data with averages or smart guesses from similar records. If anything was too incomplete, we excluded it.

- Not a sufficient number of positive cases (individuals with heart disease) -

The dataset contained far more healthy cases than diseased ones, creating a potential bias in the model.

 What we did: We balanced the categories by creating synthetic examples of the underrepresented group with a technique called SMOTE.

- Difficult to identify the best model -

With so many algorithms released into the wild, picking the right one is not a simple task.

 What we did: We examined several models and compared them based on different performance measures not just

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805266

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## IV. RESULTS AND DISCUSSION

Our findings indicated that the performance of the machine learning models was adequate for the prediction of heart disease. Random Forest and XG Boost performed best with same high accuracy and closely matched in all other performance metrics for all models we tried.

These models didn't just accurately identify patients who were destined to have heart disease but also did a good job of preventing false alarms. They achieved on average about 85% to 90% accuracy, which is pretty good for a medical prediction tool.

The F1-score — equal parts precision and recall — was also high, indicating the models made good predictions in general. Visual aids such as confusion matrices and ROC curves helped us to gain a clearer understanding of the performance of each model.

## V. CONCLUSION

This shows that data papers can first help doctors catch heart problems, which can cause a major change in treatment. Although the device cannot replace a physician, it can serve as a useful assistant to help identify people who may quickly be at risk.

On the whole, cardio is just an example of how technology can make the health care system more smarter and more on your fingers.

**Future Work**

More and better diverse data sets so that the model can provide the right prediction for individuals in all people. We also want to use refined techniques such as deep learning, which can enable the system to search for more complex patterns in data.

Another purpose is to connect the model to health data in real time, such as data from smart watches or training tracks, to make predictions more dynamic and useful in everyday life.

## REFERENCES


- https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset"
- https://en.wikipedia.org/wiki/Correlation
- https://www.kaggle.com/
- Cleveland Clinic Foundation. (n.d.). UCI Machine Learning Repository: Heart Disease Data Set. Retrieved from
- https://archive.ics.uci.edu/ml/datasets/Heart+Disease
  The dataset used for training and testing in this project.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

## IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com