iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60528

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Prediction of Heart Disease by using Machine Learning Algortihms

Dr.Srinivas Ambala[1], Yash Salunkhe[2], Divya Sawant[3], Tejas Waydande[4], Gayatri Shingate[5]
Department of Computer Engineering Pimpri Chinchwad College of Engineering Nigdi Pune, India

Abstract: ***Heart disease is considered as one of the common health problem, and machine learning can be a powerful tool for reducing the burden of disease. Using algorithms to learn from data and make predictions about the future is the process of creating a machine learning-based heart disease prediction model. the information that was utilized to train machine learning algorithms from various sources. By identifying risk factors and offering tailored treatment and lifestyle advice, the machine learning algorithm's predictions can help avoid heart disease.The increasing incidence of cardiac disease highlights the critical requirement for early detection instruments. With the use of machine learning techniques, this study creates a heart disease prediction system (K-NearestNeighbors, SVM, DT, and RF). By analyzing relevant medical attributes, the system aims to identify individuals at higher risk of heart disease. Performance evaluation demonstrates improved accuracy compared to previous models, offering potential to reduce misdiagnosis and lower healthcare costs. Implemented in a .pynb format, this project contributes to the advancement of predictive healthcare for heart disease.***
Keywords: Heart diseases, Early detection, SVM Classifier, RF Classifier, KNN Classifier.

I. INTRODUCTION

Machine learning's vastness and growing applications offer exciting possibilities. Its supervised, unsupervised, and ensemble learning classifiers facilitate data analysis and prediction. This project aims to leverage these techniques in a Heart Disease Prediction System (HDPS), demonstrating the potential of machine learning for impactful healthcare solutions. One of the most deadly illnesses in the world, heart disease has a devastating impact on a person's quality of life. Over the world, heart disease has become one of the main causes of death. The World Health Organization estimates that heart-related illnesses claim 17.7 million lives annually, or 31% of all fatalities worldwide. These days, heart disease is rather prevalent. Treatment and prevention of heart failure depend on an early and accurate diagnosis of heart disease. The machine learning methods are applied directly to the dataset in order to produce a meaningful prediction analysis. Cardiac failure can be predicted using machine learning algorithms based on a variety of characteristics, including age, sex, CP, Trestbps, and others. Here, we're concentrating on heart disease prediction using patient datasets and patient data for which we need to estimate the likelihood of occurrence. There are number of Machine learning algorithms which can be used for prediction purpose. Few of them are K-Nearest Neighbors, Support Vector Machine (SVM), Random Forest(RF),DecisionTree..

In developed countries, cardiovascular diseases are one of the leading reasons for death worldwide. A high risk of CVD is also because of factors like high blood pressure, obesity, stress, diabetes, alcohol, cholesterol, and smoking that can prevent also treat with salutary practices changes. However, other risk factors can be ungovernable such as age,gender, and history of the family. Initial phase detection of cardiovascular diseases can prevent the mortality rate; people are not conscious of the causes of the cardiovascular disease earlier due to the absence of awareness. Health care organizations are trying to diagnose the disease at the initial phase. Most of the time, the disease is noticed at the last phase or after death. We aim to diagnose the disease at an initial phase [1-2]. Using machine learning algorithms, we can recognize the disease at an initial stage and help to cure disease with a conventional diagnosis. To develop the model, we have used different machine learning algorithms to resolve the problem; we also tried to find the significant factor while predicting cardiovascular disease.

II. LITERATURE REVIEW

This paper developed a model using Random Forest, XGBoost utilizing**.**The dataset is available in UCI machine learning Repository having an accuracy of 92.85%. Benefits of this research include the ability to predict early cardiac disease detection using the Random Forest algorithm. Disadvantages are after analyzing the quantitative data generated from the computer simulations, moreover their performance is closely competitive showing slight difference.

©IJRASET: All Rights are Reserved

More features or risk factors, more accurate methodologies, and a larger number of records from the genuine dataset—that is, patients with heart disease from Indonesia's largest national cardiovascular center—will be used in future study.

This proposed the technique based on Random Forest Algorithm. Using the Heart condition dataset available in UCI machine learning repository. And got an accuracy of 90.16%. Advantages are Model Stacking Performance Evaluation: Heart Attack Prediction. Disadvantages are Accuracy of Bayes predication is not excepted as compare to random forest algorithm. Future scope incuded developing an internet application supported by the random forest algorithm

The proposed the Naiye Bayes, Decision Tree, LR and Random Forest. Dataset used was available in UCI machine learning Repository. Accuracy obtained was 90.16%. This research have developed that Random Forest Algorithm is the most accurate algorithm for the prediction heart disease and more advance parameter s are used in order to obtain the test. The future goal to develop an internet application for employing a large dataset.

III. PROPOSED WORK

According to Figure 1, the system is implemented using five major steps in prediction cardiovascular disease and to measure the accuracy of each of the models.

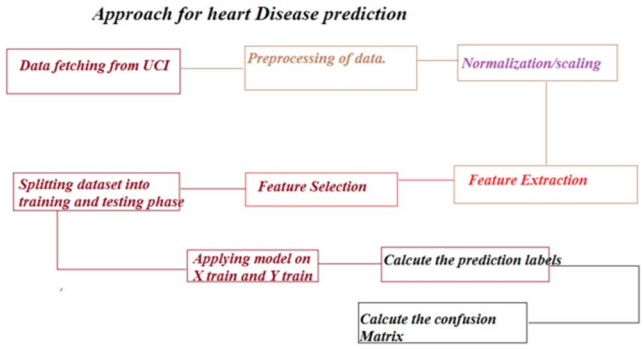


Fig .1 Workflow of the GUI

IV. METHOD EXPLANATION

A. Data Description

There are three important features of input classified as Objective, Examination, and Subjective:

1) Objective features for factual information from patient;
2) Examination features for results collected after medical examination;
3) Subjective features for information collected from the patient.

Features used in the implementation are: Age, Glucose, Height, Cardiovascular disease (Presence),Weight, Physical activity, Gender, Alcohol intake, Systolic blood pressure, Smoking, Diastolic blood pressure, Cholesterol are the features of dataset used in model for description.

B. Data Pre-processing

The dataset is consisting of 12 rows and 1400 columns (patient records). After removing similar records, theremaining 1399 patient records used. For the features of the provided dataset, the binary classification and the multiclass parameter is proposed. To examine the presence or absence of the heart disease, the multiclass parameter is used. The value 1 indicates that the patient has heart disease. State 0 specifies that the heart disease is absent in the patient. The medical records are transformed into the detection values during the pre-processing. The data pre-processing of the patient records designated that the 940 cases designate the value of 1 indicating that the heart disease is present. The value of 0 shown by the rest 460 records signifies the absence of the heart disease.

C. Feature selection

Out of the 13 properties in the collection, the patient record is uniquely recognized by two features: age and sex, which assign unique IDs. Information about medical conditions makes up the remaining features. Key characteristics that help predict heart disease are found in the medical data. In order to identify the features with a high and positive correlation feature, correlation was conducted on all 13 characteristics using the goal value.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Features	Correlation
Exang	0.436757
Cp	0.433798
Oldpeak	0.430696
Thalach	0.421741
Ca	0.391724
Slope	0.344029

Fig 2. Feature Selection using correlation

D. Classification

Classification means classifying the data in different groups based on the similarities present in different data points. Here classification is used in the prediction of heart disease. Various machine learning models are available, but in the proposed method, any one of the following algorithms or models can be used [12] The given problem is a classification and regression problem. We are using various algorithms to find (or predict) the relation between the target variable (i.e. survived or not) with other variables (Sugar Gender, Age, and more). The following algorithms are used:

1) *k-NN Neighbours:* K-Nearest Neighbors (KNN) is probably one of the simplest algorithms utilized in machine learning for both regression and classification techniques. KNN calculations use the data and characterize new data points dependent on resemblance measures (e.g., distance function). In basic, we can say The KNN calculation accepts that comparative things are close to one another. In KNN, Classification occurs by considering the majority vote to its neighbors. The data point goes to the class that has the most intimate neighbors. As we increment the number of nearest neighbors, the estimation of k and accuracy may increment.The predictor variables anticipate the target variable or thedependent variable. For our situation, our target variable isthe chance of coronary illness, and by utilizing the KNN algorithm, our model will predict the result
2) *Support Vector Machine(SVM):* Support Vector Machine (SVM) is a regulated ML calculation that we utilize for both regression and classification tasks. Notwithstanding, we generally use it in classification issues. In our model to implement SVM, we plot every single data item as a point in n-dimensional space(where n is the number of features you have) with the estimation of each component being the estimation of a particular coordinate. At that point, we perform classification by finding the hyper-plane that separates the two classes well overall.
3) *Decision Tree:* We executed the decision tree classification algorithm since it is truly outstanding and most utilized regulated learning strategies. Building a decision tree that is steady into a given data set is simple. The difficulty lies in building proper decision trees, which implies concise decision trees. We used ID3 ((Iterative Dichotomiser 3), which is the most widely used type among all the decision tree algorithms' types.
4) *Random Forest:* Several decision trees are combined in Random Forest, an ensemble learning technique, to produce forecasts that are more accurate. It is commonly used for classification and regression tasks and is known for its ability to handle high-dimensional data and reduce overfitting.
5) *Logistic Regression* Logistic regression is a supervised machine learning algorithm used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The logistic regression model transforms the linear regression function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1.

V. RESULTS

To evaluate the efficiency of the classification, the predicted type was compared with the original type. In the propounded methodology, for classification objective, the dataset was examined by seven different classifiers mentioned above (1 -7). For this dataset, the decision tree and random forest classifiers have produced the maximum accuracy.

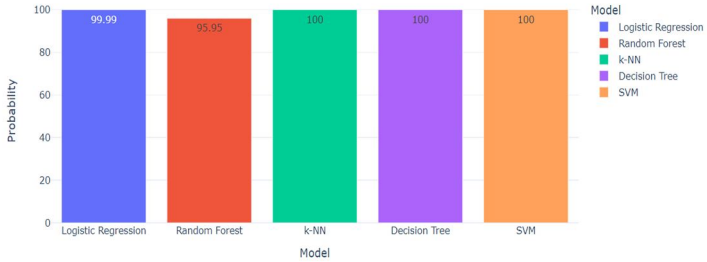


Fig 3. Prediction percentage of the models

According to the graph k-NN, Decision tree and SVM have the most accurate prediction of 100%.
The accuracy score of the algorithms is as follows:

Algorithm	Accuracy
Logistic Regression	97.5
Random Forest	97.5
k-NN	96.0
Decision Tree	96.5
Support Vector Machine	97.5

The prediction of the heart disease will be done using the Logistic Regression algorithm since this algorithm has the highest accuracy.
These results are based on the crucial factors of the patients data including his cholesterol level,age,sex and maximum heart rate.
These factors are important as these will help to detect whether the patient would have a heart disease and what would be the treatment for the disease would be given by the doctor and be sent throught him via email or the patient is at the low risk of having the heart disease.Even if he is at the lower risk what precautions should be taken would be adviced to the patient by the doctor.

VI. CONCLUSION

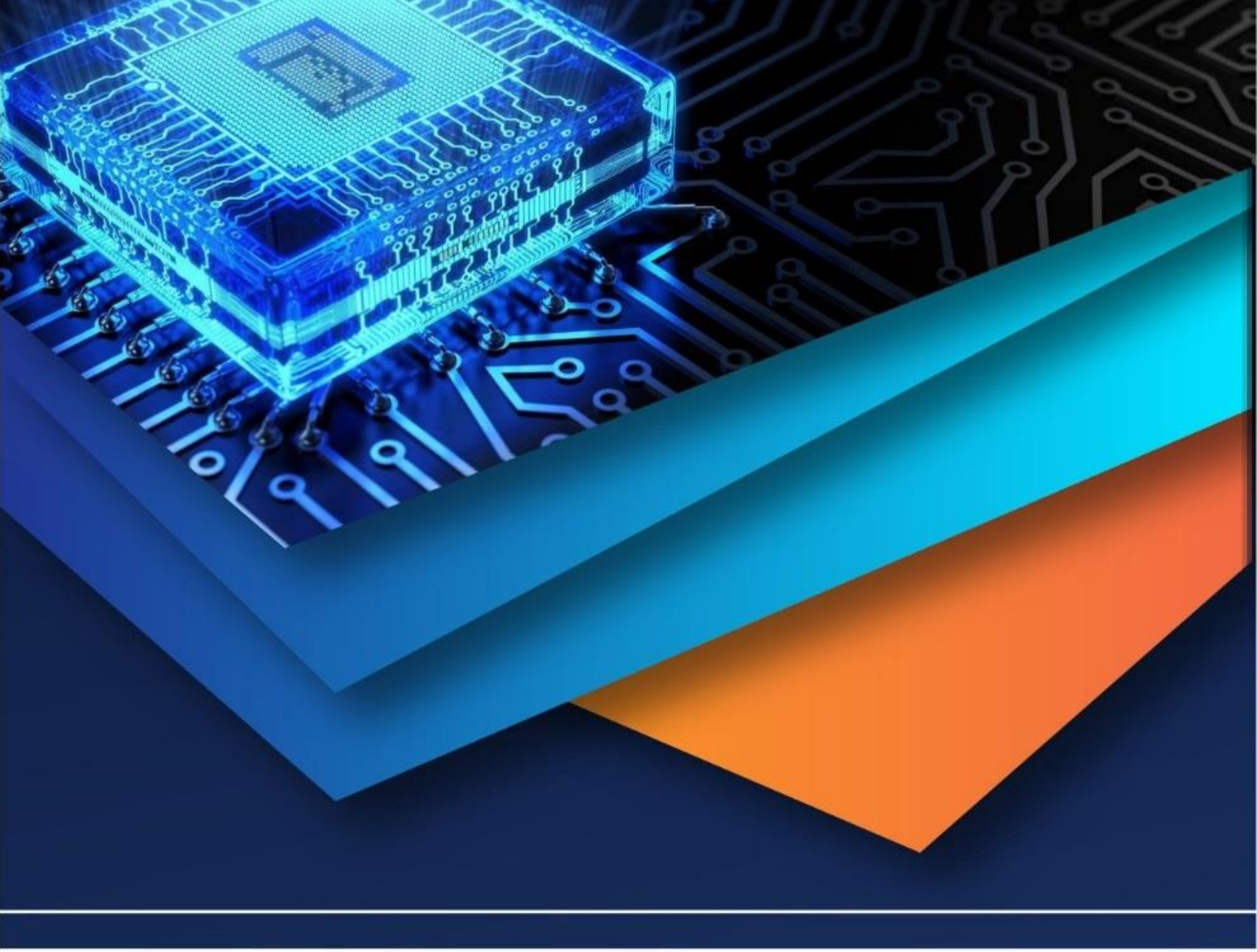
To sum up, there is a lot of promise for machine learning in the diagnosis and prognosis of heart disease. Thanks to advancements in data collection, computation, and machine learning algorithms, it is now possible to design models that accurately predict an individual's risk of developing heart disease based on a variety of demographic and health-related factors. While machine learning has the potential to alter the way we detect and prevent heart disease, there are still difficulties involved in deploying these models, such as ensuring that they are reliable, impartial, and fair. The construction of accurate models also depends critically on the quality and representativeness of the training data.
Heart disease prediction is essential as well as challenging work in the medical Field. Nevertheless, the mortality rate can be reduced if the disease is recognized at the initial stages, and precautions and proper treatment are possible. This paper illustrates various automated computerized Cardiovascular Disease Prediction methodologies, which can be performed by Supervised Learning plus Classification and Regression methods. The algorithms are tested using various features.
Future research should focus on resolving these problems and improving machine learning models' predictive power for heart disease. Nevertheless, the field of research on utilizing machine learning to predict cardiac disease is exciting and promising, and it has the potential to significantly impact both public health and medicine.

REFERENCES

[1] Lakshmi C.N, Bindhushree., Jaya Poojary, Manish, Shylaja., “Heart disease Prediction using machine learning algorithm.”(2022) International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 10 Issue VII July 2022
[2] Aditi Gavhane, Isha Pandya, Gouthami Kokkula, Prof. Kailas Devadkar (PhD), “Predication of heart disease using machine learning algorithm” Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1 (2021).
[3] Rahul chaurasia, Saksham Gupta, shipra singh siddhu, “ Prediction of heart disease using machine learning” (2018). IJCRT Volume 6, Issue 2 April 2018.
[4] Pooja Anbuselvan, “Heart Disease Prediction using machine learning techniques.”(2020). IJIRT .

[5] Penmetsa Reshma Devi, Dr. N.K. Kameswara Rao ” Heart disease prediction using machine learning”(2022).
[6] Rishab Magar, Roham Memane, Suraj Raut, Heart disease prediciton using Machine Learning (2020) International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2020)
[7] Apurb Rajdhan, Avi Agarwal, Dr.Poonam Ghuli, Milan Sai, Dundigalla Ravi” Heart disease prediction using Machine Learning” (2020)
[8] Rachit Misra, Pulkit Gupta, Prashuk Jain” Prediction of heart disease using Machine Learnin Algorithms”(2020) IJIRT
[9] Abhijeet Jagtap, Priya Malewadkar, Omkar Baswat, Harshali Rambade” Heart Disease prediction using Machine Learning”(2019) International Journal of Research in Engineering, Science and Management Volume-2, Issue-2, February-2019
[10] Rashmi G.Saboji, Prem Kumar Ramesh “A Scalable Solution For Heart Disease Predication Using Classification Mining Technique.” (2017) International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)