iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47896

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 10 Issue XII Dec 2022- Available at www.ijraset.com

Heart Disease Prediction Using Machine Learning

Mukund Kulkarni[1], Tanay Patel[2], Kunal Patil[3], Akshita Manwatkar[4], Mustafa Bohra[5], Mustansir Daulat[6]
[1, 2, 3, 4, 5, 6]*Computer Engineering Department, Vishwakarma Institute of Technology, Pune*

Abstract: ***One of the top causes of mortality globally is heart disease. A doctor cannot readily foresee it since it is a complex process that needs experience and superior forecasting knowledge. The health-care institution remains "rich in information" yet "deficient in information." There is a wealth of information available in online health care systems. However, there is a scarcity of appropriate data analysis tools for detecting underlying linkages and patterns. An automated medical diagnostic system can boost medical efficiency while lowering expenses. The objective is to identify hidden patterns in cardiovascular illness using data mining techniques and forecast the existence of heart disease in individuals when presence is assessed on a scale. Predicting cardiac disease necessitates a vast quantity of data that is too complicated and voluminous for normal tools to collect and interpret. The purpose of this research is to develop the most effective and accurate machine learning approach for predicting heart disease.***
Keywords: Prediction, heart disease, machine learning, algorithms, analysis, database, data mining.

I. INTRODUCTION

The leading cause of death in India and the rest of the world is heart disease. According to the World Health Organization, cardiovascular diseases kill 17.7 million people each year, accounting for 31% of all deaths worldwide (WHO). As a result, it is an essential moment to analyse the mortality rate by correctly detecting the disease in its early stages. Health care management can utilise the information gathered to enhance service quality. The amount of data in the healthcare industry is enormous. Patient information, resource management information, and data that has changed are all included. Health-care organisations should have the ability to analyse data. Millions of patients' medical information might be stored, and computer technology and data mining techniques could help answer a few key health-care questions.) Clinical choices are frequently made based on physicians' expertise and knowledge rather than the rich material buried on the internet. This practise results in unintended bias, mistakes, and exorbitant medical expenses, all of which have an impact on the quality of care offered to patients. Combining clinical decision support with computer-based patient records has been recommended to decrease medical mistakes, increase patient safety, eliminate undesirable behaviour variability, and improve patient outcome. This proposal is promising as a model for modelling and analysing

For instance, data mining has the potential to generate a wealth of data that can significantly improve the accuracy of therapeutic judgments.

II. LITERATURE SURVEY

The aim of the paper was finding the most suitable algorithm for predicting the heart disease. The study compares the accuracy prediction of Decision tree, Logistic regression, Random Forest, Naïve Bayes algorithm .Their study stated that Random Forest is the most suitable algorithm for predicting the heart disease with 90.16% of accuracy. [1]

The paper presented technology various ML algorithms to diagnose the heart disease. They created confusion matrix of all the algorithm and compare all the algorithms and get SVM is the finest algorithm to predict the heart disease. [2]

The study is an element of work on heart disease detection and prognosis. It is based on the use of machine learning algorithms, of which we have selected the three most popular ones (Neural Network, SVM, and KNN), on a real data set of Algerian individuals, with excellent results; we reached 93% accuracy with Neural Network. Their ability to test the algorithm's stability on data sets of various sizes and conclude that Neural Networks produce the best outcomes is what makes their study so strong. [3]

This project provides the deep insight into machine learning techniques for classification of heart diseases. The role of classifier is crucial in healthcare industry so that the results can be used for predicting the treatment which can be provided to patients. The existing techniques are studied and compared for finding the efficient and accurate systems. Machine learning algorithms greatly enhance the accuracy of cardiovascular risk prediction, allowing individuals to be detected early in the disease process and get preventative therapy. In some circumstances, their method worked well, while in others it did not. [4]

In this study, two supervised data mining algorithms—Naive Bayes Classifier and Decision Tree Classifier—were applied to the dataset to estimate the likelihood that a patient will develop heart disease. To compare the accuracy of these two methods, the same dataset is used for both of them. The accuracy level of this model's prediction of heart disease patients was 91%, while the accuracy level of the Naive Bayes classifier's prediction of heart disease patients was 87%. [5]

©IJRASET: All Rights are Reserved

III.PROPOSED SYSTEM

Data collection is the first step in a processed technique; for this, obtain data from Kaggle, which has been thoroughly examined by researchers.

A. Data Collection

The first stage in developing a prediction system is data gathering, which is done via the internet [12], followed by data cleaning and selecting the training and testing datasets. In this research, we used 80% of the training dataset and 20% of the testing dataset to train the system. After cleaning, this data collection has 14 columns and 383 rows generated by code (data.info). When the code (data.head) is used, a table with 5 rows is formed, as illustrated in Fig. 2.

B. Element Selection

Dataset Elements are property of dataset that are used for analysis and prediction. There are many elements such as sex, age, slope and many more that are shown in TABLE.1 for system analysis.. .Block Diagram of Prediction System is .displayed in Fig.2. Table

C. Element Selection

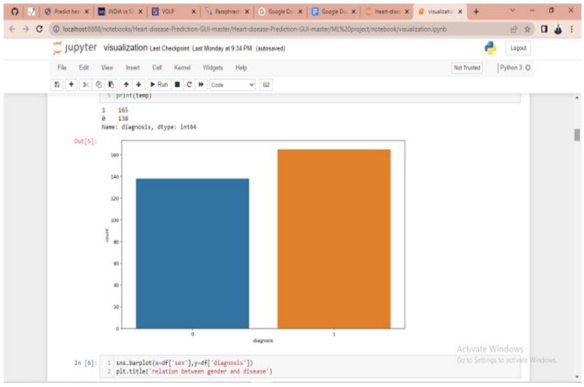
Dataset Elements are dataset properties used for analysis and prediction. Many factors, such as gender, age, slope, and others, are presented in TABLE.1 for system analysis. Figure 2 shows a block diagram of a prediction system.

S.NO	EXPLANATION	ELEMENTS
1	PATIENT's AGE	age
2	MALE, FEMALE	sex
3	CHEST PAIN	Cp
4	REST BLOOD PRESSURE	trtbps
5	COLESTROL	cho
6	FASTING BLOOD SUGAR	fbs
7	REST-ELECTROCARDIOGRAPH	Restecg
8	MAX HEART RATE	thalachh
9	SLOPE	Slp
10	THALASSEMIA	Thall
11	OUTPUT(Heart disease Patient	Output

D. Pre-processing of Data

Pre-processing needed for achieving prestigious result from the machine learning algorithms.

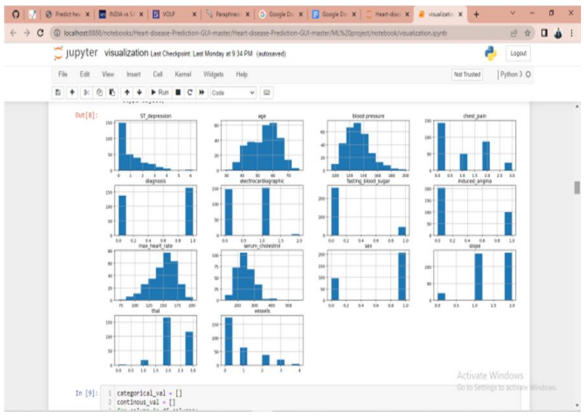
Some ML algorithm does not support missing values for this we have to manage null values from original raw data. Some attribute of data set have been detected that is not useful for prediction such as education city etc. Fig 3 shows green color bar represents non disease patient and red color bar presents heart disease patient. This diagram has created by the code.



Graph showing Boolean value of patients with Heart Attack

E. Histogram of Elements

Elements displays the variety of dataset properties. Using the code [dataset.hist()], In a frame, several histograms are formed.



Graph of each cardiality

IV. METHODOLOGY

A. Machine Learning Algorithms

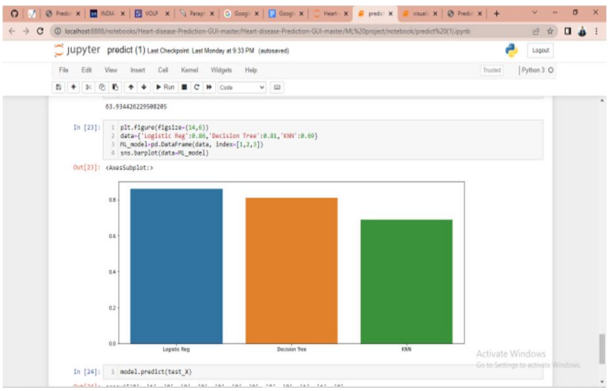
1) *Logistic Regression [LR]:* The supervised ML learning approach is known as LR.. The relationship between them is illustrated by an equation of line, which is linear in nature, which is why this technique is known as linear regression. As shown in Fig.5, it provides a connection equation to forecast a dependent variable value "b" based on an independent variable value "a," implying that the linear regression approach provides a linear relationship between a(input) and b. (output).
2) *Decision Tree [DT]:* Decision Tree is an algorithm that categorises parameters despite the presence of arithmetic data. DT generates a tree-like structure. Because of its simplicity, DT has evaluated several huge data sets linked to medicine. For analysis, it uses tree nodes. Node of the Leaf: Indicate the outcome of each Test. Internal Node: Other nodes rely on the main node [Root Node] to manage a range of components. Data will be divided into two or more parallel sets using this method. The entropy of each parameter is then computed. After that, split the data using a predictor with a high information gain and a low entropy.
3) *K-nearest Neighbour [KNN]:* KNN is a classification method from 'The Supervise' learning family. It classifies entities that rely on their nearest neighbours. KNN is a widely utilised approach that may be used as a classifier and regression in a variety of fields such as image processing, data processing, pattern recognition, and other applications. The algorithmic program's output is determined by the K-nearest neighbour class, which is enforced by locating K-number of coaching points closest to the requested character and considering the votes among the K objects. The algorithmic software is quite simple. is, on the other hand, capable of learning very complicated non-linear call boundaries and regression functions . KNN's intuition that related cases should have similar category labels (in classification) or target values (regression).

V. RESULT ANALYSIS

Jupyter notebook is the tool for simulation of programming and open source network application. By using this, we can work with python programming.

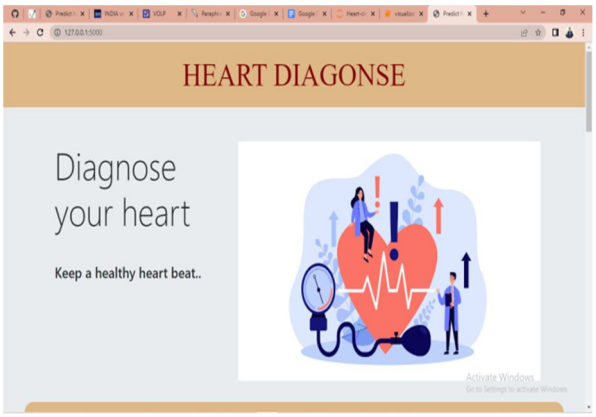
It is very comfortable tool .It contains coding, scripting language elements that have links, equations, figures, plots and many more. By importing various libraries of python programming, work on large dataset and analysis and visualize with different graphs of data in real time. With Jupyter notebook, data cleaning, numerical imitation, statistical modelling and many work have to be done.

S.NO	MODEL	ACCURACY
1	LR	86
2	DL	81
3	KNN	63

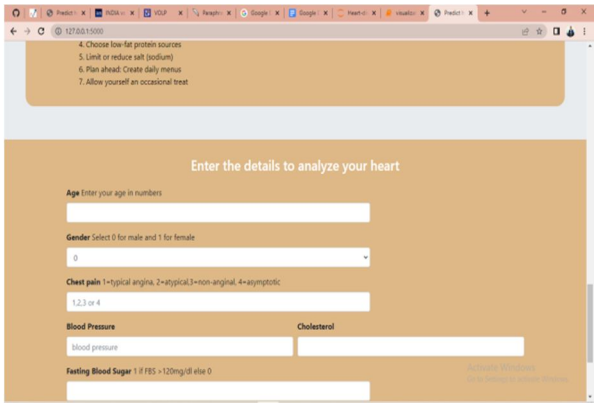


Accuracy of Predictions

A. Web Application



Our website's Home page



Input window

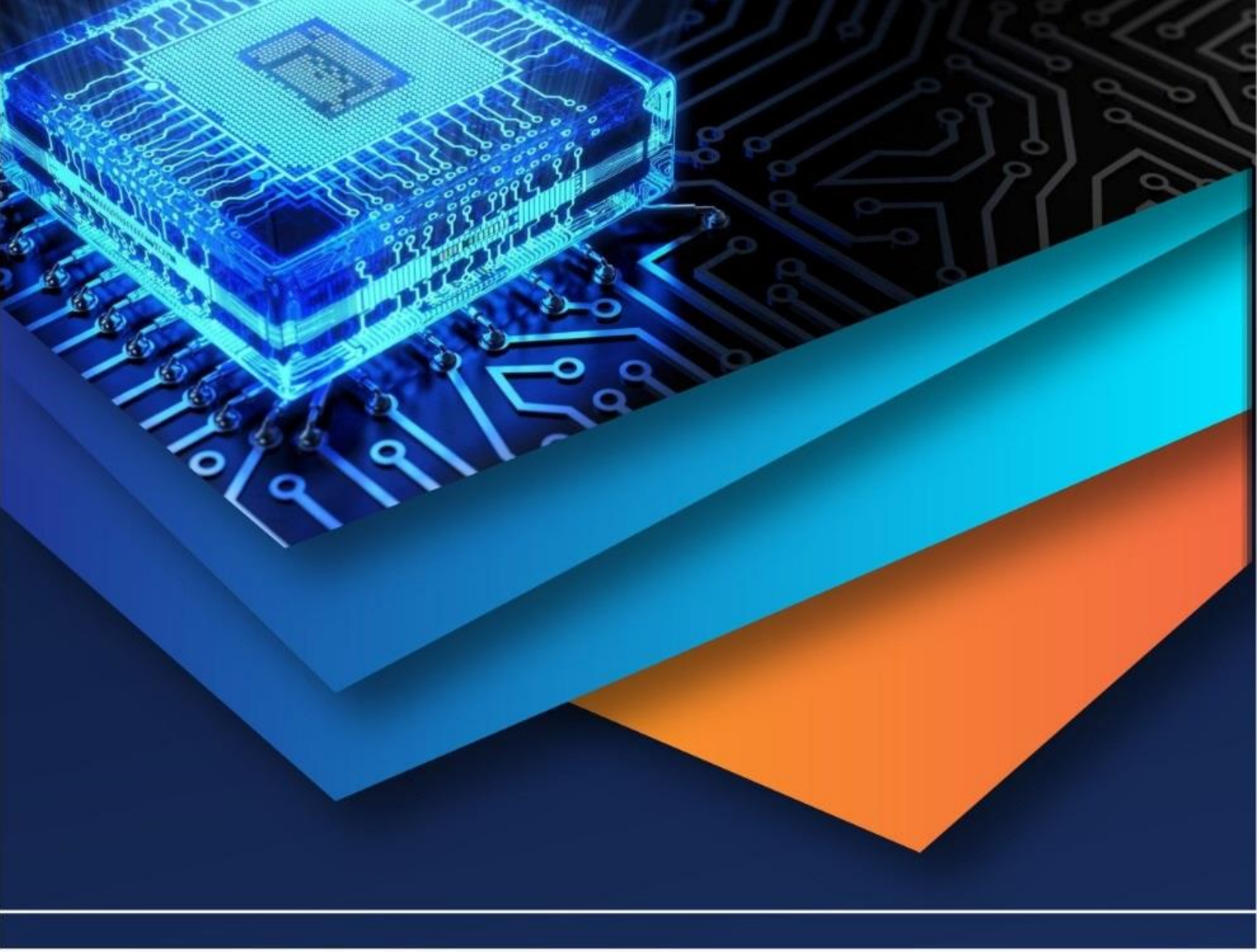
VI. CONCLUSION

This assignment will provide you a thorough grasp of machine learning approaches for heart disease detection. Role of classifier is important in the healthcare industry so that the results are used to predict possible treatment is provided to patients. Existing strategies are researched and compared to find the most effective and efficient programs.

Machine learning algorithms dramatically enhance the accuracy of predicting cardiovascular risk in people who can be identified early in the disease and may benefit from preventative therapy. It may be concluded that machine learning methods for predicting heart illness or heart-related ailments have a wide range of applications.

REFERENCES

[1] Baban.U. Rindhe1, Nikita Ahire2, Rupali Patil3, Shweta Gagare4, Manisha Darade5 ," Heart Disease Prediction Using Machine Learning" , International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) ,Volume 5, Issue 1, May 2021
[2] Dhai Eddine Salhi, Abdelkamel Tari2, and M-Tahar Kechadi , "Using Machine Learning for Heart Disease Prediction" , LIMOSE Laboratory, University of Mhamed Bougara, Boumerdes, Algeria ,February 2021
[3] Apurb Rajdhan, Avi Agarwal,Dr. Poonam Ghuli, Dundigalla Ravi, Milan Sai ,"Heart Disease Prediction using Machine Learning" International Journal of Engineering Research & Technology (IJERT) , Published by : www.ijert.org ,Vol. 9 Issue 04, April-2020
[4] Harshit Jindal1, Sarthak Agrawal1, Rishabh Khera1, Rachna Jain2 and Preeti Nagrath ,"Heart disease prediction using machine learning algorithms" ,ICCRDA 2020
[5] Rati geol ," Heart Disease Prediction Using Various Algorithms of Machine Learning" ,March 2021
[6] Vijeta Sharma , Shrinkhala Yadav, Manjari Gupta , "Heart Disease Prediction using Machine Learning Techniques", ICACCCN 2020
[7] Asmit Srivastava , Ashish kumar Singh , "Heart Disease Prediction using Machine Learning " , ICACITE 2022
[8] Kuldeep Vayadande , Rohan Golawar , Sarwesh Khairnar , Arnav Dhiwar ,Sarthak Wakchoure , Sumit Bhoite ,Darpan Khadke , " Heart Disease Prediction Using Machine Learning and Deep Learning Algorithms " , CISES 2022
[9] S.Usha , S.Kanchana , " Effective Heart Disease Prediction using Machine Learning Techniques " , ICEARS 2022
[10] M.Snehith Raja , M.Anurag , Ch.Prachetan Reddy , Nageswara Rao Sirisala , " Machine Learning Based Heart Disease Prediction System " , ICCCI 2021



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)