IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Heart Disease Prediction using Machine Learning

Yash Nevase[1], Fardeen Patel[2], Suffiyan Shaikh[3], Ayush Raskar[4], Prof. Vijaya Avati[5]
Students, Department of Mechanical Engineering[1,2,3,4]
Asst. Professor,Department of Mechanical Engineering[5]
JSPM'S Jayawantrao Sawant College of Engineering, Hadapsar, Pune, Maharashtra, India

Abstract: *Machine Learning is applied in a variety of fields all* handle. *over the world. The healthcare sector is no exception. Machine Learning can help forecast the existence or absence of locomotor problems, heart disease, and other diseases. In this research, we compare classifiers such as decision trees, KNN, Logistic Regression, and Random Forest and propose an ensemble Regression, and Random Forest and propose an ensemble weak classifiers.*

Keywords: Machine Learning, KNN, Logistic Regression,Decision tree

I. INTRODUCTION

The World Health Organization estimates that heart disease causes 12 million deaths worldwide each year. One of the leading causes of morbidity and mortality among the global population is heart disease. One of the most crucial topics in the data analysis area is predicted cardiovascular disease. Since a few years ago, the prevalence of cardiovascular disease has been rising quickly throughout the world. Numerous studies have been carried out in an effort to identify the most important risk factors for heart disease and to precisely estimate the overall risk.

Heart disease diagnosis is a difficult undertaking that can provide automated predictions about the patient's heart state to improve the effectiveness of subsequent treatment.Heart disease is typically diagnosed based on the patient's indications and symptoms. Several techniques, including K-nearest neighbour (KNN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Logistic regression (LR), are used to categorise the severity of the condition.

II. EXISTING MODELS

In previously employed machine learning models like Decision Trees, we start from the root of the tree to predict a class label for a record. We contrast the root attribute's values with that of the attribute on the record. We follow the branch that corresponds to that value and go on to the next node based on the comparison. The value of the attributes in the dataset is also explained by the decision tree. They also made use of the Cleveland data collection. Using some techniques, the data set is divided into 70% training and 30% testing. 91% of the time, this method is accurate.While KNN works by calculating the distances between a query and each example in the data, choosing the K examples that are closest to the query, and then voting for the label that is used the most frequently. Thus KNN gave us better results compared to others by more than 92% accuracy.

1.1 Objectives

The objectives of the project are as follows:

- To identify complex patterns in user data and then predict outcomes based on those patterns to respond to business inquiries and address business concerns.
- To analyze the data and find trends with the use of machine learning.
- To create a method for predicting cardiac disease.
- To develop a code for heart disease prediction.
- To analyze data for effectiveness and precision of present procedure.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

III. LITERATURE SURVEY

Heart Disease Prediction using ML (Mr. Santhana Krishnan. J, Dr. Geetha S)

In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of 91% occurring heart disease of the patients in terms of percentage. This is performed through data mining classification techniques.

Heart Disease Prediction using ML (Hassan, C.A.u.; Iqbal, J.; Irfan, R.; Hussain, S.; Algarni, A.D.; Bukhari, S.S.H.; Alturki, N.; Ullah, S.S.)

Effectively Predicting the Presence of Coronary HeartDisease Using Machine Learning Classifiers.

Received: 7 April 2022
Accepted: 27 July 2022 Published:23,September2022

IV. METHODOLOGY

1 Collection of dataset:

We first gather data for our technology that forecasts cardiac illness. We divided the dataset into training and testing data after it was collected.

The learning of the prediction model takes place on the training dataset, and the evaluation of the prediction model occurs on the testing dataset. 30% of the data are utilised for testing in this project, while 70% are used for training.

Heart Disease UCI served as the project's data source. The dataset has 76 properties, of which the system uses 14 for its operation..

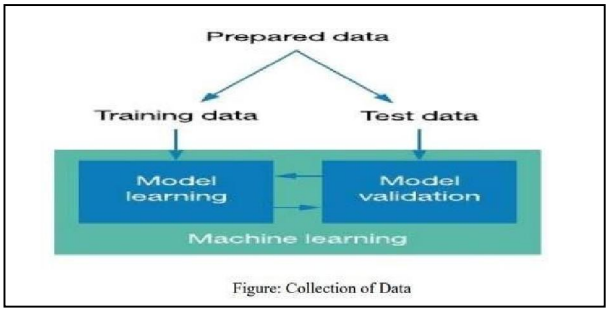


Figure: Collection of Data

2 Selection of attributes:

The choice of acceptable attributes for the prediction system is included in attribute orfeature selection.

This is done to make the system more effective. For the prediction, a number of patient characteristics are used, including gender, the nature of the patient's chest discomfort, fastingblood pressure, serum cholesterol, and exang.

For this model's attribute selection, thecorrelation matrix is used.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

3. Pre-processing of Data :

The pre-processing of data is a critical stage in thedevelopment of a machine learning model.
Data that isn't initially clean or in the model's required format can lead to inaccurate results. Pre-processing involves transforming data intothe format we need. It is used to handle the dataset's noise, duplication, and missing values.
Activities like importing datasets, partitioning datasets, attribute scaling, etc. are all part of datapre-processing.
Preprocessing the data is necessary to increasethe model's accuracy.

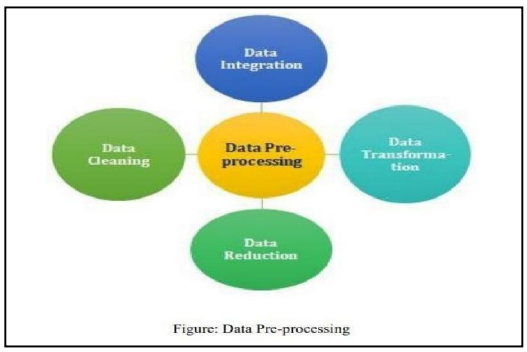
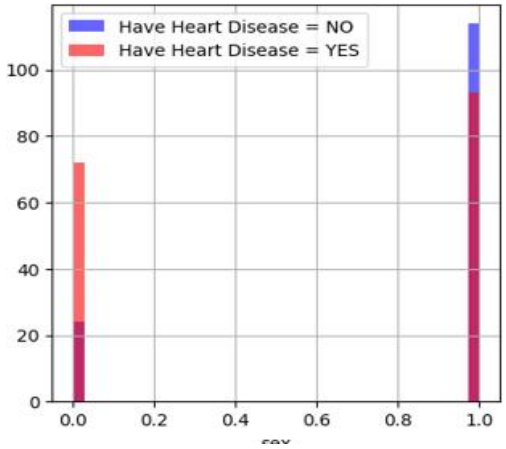


Figure: Data Pre-processing

4. Balancing of Data

There are two approaches to balance unbalanced datasets. They are both under- and over-sampling.
Under Sampling: In Under Sampling, the size of thelarge class is reduced in order to balance the dataset.
When there is enough data, this process is taken intoaccount.
Over Sampling: In Over Sampling, the size of the scarce samples is increased in order to balance thedataset.
When there is not enough data, this process is takeninto account



5. Prediction of Disease

For classification, a variety of machine learningtechniques are utilized, including SVM, Naive Bayes, Decision Trees, Random Trees, LogisticRegression, Ada-boost, and Xg-boost.
Algorithms are compared, and the algorithmthat predicts heart disease with the best degree of accuracy is chosen.

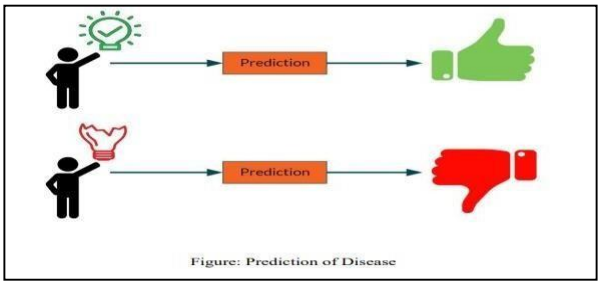


Figure: Prediction of Disease

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

IV. RESULT AND CONCLUSIONS

The project is created with the help of python programming, machine learning.
The model is trained with 300 patient's data and 14 different parameters .The model has accuracy of 95% using KNN.
The project helps in detection of heart disease of thepatient using his medical data provided by the doctor

ACKNOWLEDGMENT

- It is our proud privilege and duty to acknowledge the kind of help and guidance received from several people in preparation of this report. It would not have been possible to prepare this report in this form without their valuable help, co-operation andguidance.
- We express our sincere gratitude to our guide, **Prof. Vijaya Avati**, Department of Mechanical Engineering, JSCOE, Hadapsar, Pune for guiding us in investigations for this Projectand in carrying out experimental work. We hold him in esteemfor guidance, encouragement and inspiration received from him.
- Our sincere thanks to **Prof. (Dr.) Pradeep A. Patil**, Head of the Department of Mechanical Engineering, JSCOE, for hisvaluable suggestions and guidance throughout the period of thisreport.
- First and foremost, we wish to record our sincere gratitude to Management of this college and to our beloved Principal, **Prof. (Dr.) R. D. Kanphade**, Principal, JSPM's Jayawantrao Sawant College of Engineering, Hadapsar, Pune for his constant support and encouragement in preparation of this report and for making available library and laboratory facilities needed to prepare this report.
- The Project on **"Heart disease prediction using machine learning"** was very helpful to us in giving the necessary background information and inspiration in choosing this topic for the Project. Our sincere thanks to **Prof. (Dr.) Mahesh Gaikwad,** Project Coordinator for having supported the work related to this Project. His contributions and technical supportin preparing this report are greatly acknowledged.

REFERENCES

[1]. Heart Disease Prediction using ML(Mr. Santhana Krishnan.J, Dr.Geetha S) In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of 91% occurring heart disease of the patients in terms of percentage. This is performedthrough data mining classification techniques.
[2]. Heart Disease Prediction using ML (Hassan, C.A.u.; Iqbal, J.; Irfan, R.; Hussain, S.; Algarni, A.D.; Bukhari, S.S.H.; Alturki, N.; Ullah, S.S.) Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers.Received: 7 April 2022 Accepted: 27 July 2022 Published:23,September2022