WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

CARDIOVASCULAR DISEASE PREDICTION BY MACHINE LEARNING

POOJA R

PG Student, Department of MCA,
Vivekanandha institute of information and management studies

ABSTRACT:

Heart disease is a major cause of death worldwide. Machine learning techniques have shown great promise in aiding the diagnosis and prediction of heart disease. One such technique is the decision tree algorithm, which is a widely used classification algorithm. To develop a machine learning model using the decision tree algorithm to predict heart disease in patients. We used a dataset containing heart disease patients, including age, sex, chest pain, cholesterol levels and etc…. other medical conditions. After preprocessing the dataset and selecting the most relevant features, we trained a decision tree model on the data. We evaluated the model we also used a confusion matrix to visualize the model's performance. Data analytics is useful for prediction from more information and it helps the medical center to predict vision.

INTRODUCTION:

Heart disease is a serious health problem that affects many people and is one of the leading causes of death worldwide.This project uses machine learning to help predict heart disease. By using Python and algorithms like K-Nearest Neighbors (KNN), XGBoost, and Multi-Layer Perceptron (MLP).

OVERVIEW:

Heart disease can be difficult to detect early because symptoms often appear late or are hard to notice the system can quickly analyze the data and give accurate results. This can support doctors in making better and faster decisions, leading to early treatment and better patient visual image of Cardiovascular disease prediction by machine learning.

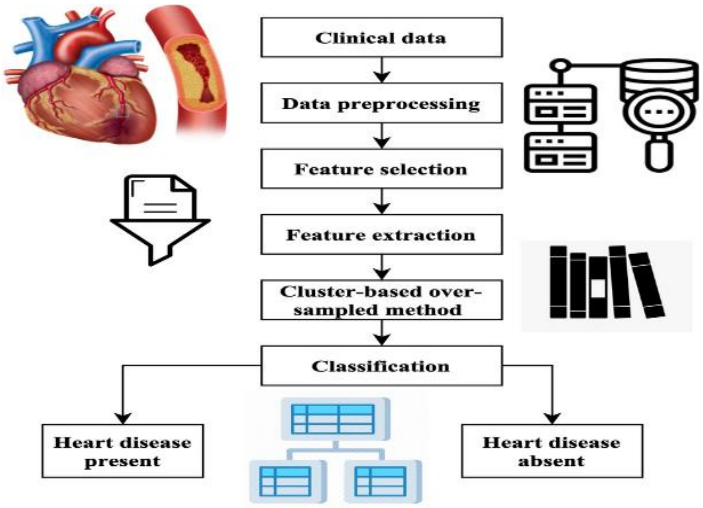


Fig.1. Shows the Comprehensive evaluation and performance analysis of machine learning in heart disease prediction.

AIM:

The aim of this project is to design and develop an intelligent machine learning-based system capable of accurately predicting the presence of heart disease using clinical and patient health data. By utilizing advanced algorithms such as K-Nearest Neighbors (KNN), XGBoost, and Multi-Layer Perceptron (MLP), the system seeks to analyze key medical indicators including age, blood pressure, cholesterol levels, and other relevant features to assess the risk of heart disease. The project is built using Python and leverages libraries like NumPy, Pandas, and Scikit-learn for data manipulation, preprocessing, model training, and evaluation. The ultimate goal is to create a tool that can assist healthcare professionals in making faster and more informed diagnostic decisions, reducing the chances of human error, and supporting early intervention strategies.

This system is intended to be efficient, scalable, and easy to integrate into real-world medical environments. Through the application of machine learning, the project aims to enhance diagnostic accuracy, improve patient outcomes, and contribute to the broader effort of using artificial intelligence in healthcare.

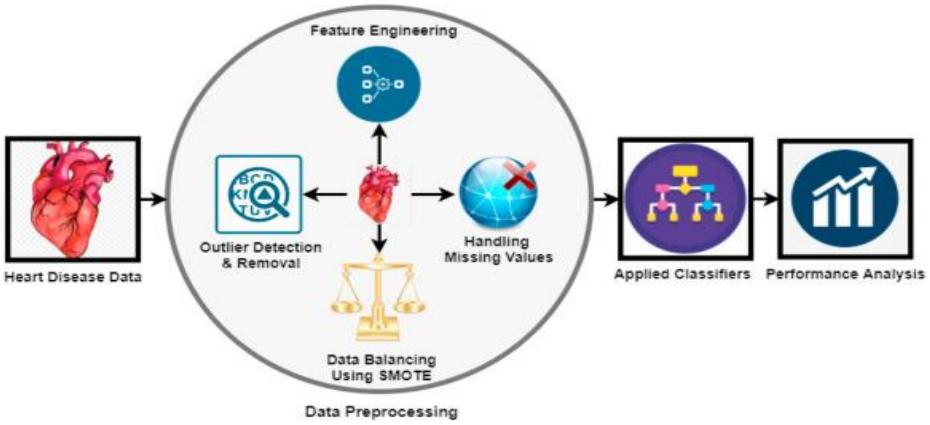


Fig. 2. Predicts using Supervised Learning algorithms.

SYSTEM:

Many deep learning models require substantial computational power, making them impractical for real-time deployment in resource-constrained health care settings.

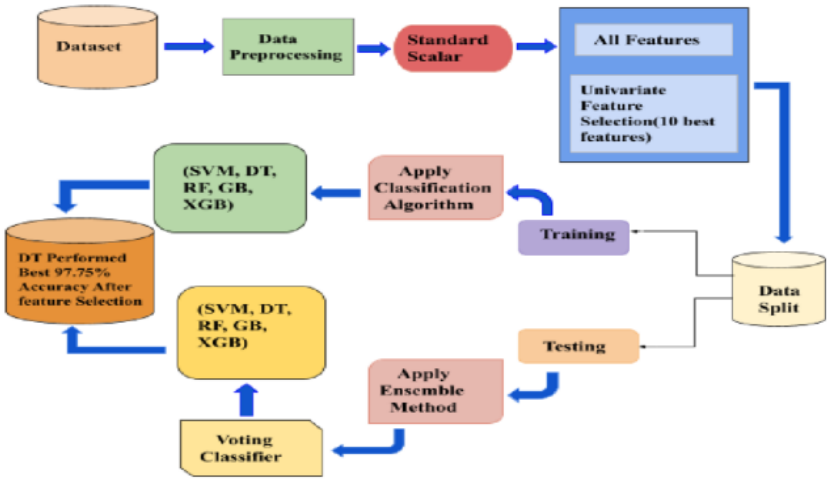


Fig.3. Shows the Proposed system and existing system by predicting the machine learning.

EXISTING SYSTEM:

Existing systems for heart disease prediction utilize a range of machine learning and deep learning techniques to enhance diagnostic accuracy. Traditional models such as Decision Trees, Logistic Regression, and K-Nearest Neighbors have been widely used but often struggle with complex, non-linear relationships in medical data. Advanced systems incorporate ensemble methods like Random Forest and XGBoost, which improve performance by combining multiple classifiers. More recent approaches use deep learning architectures, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, to automatically extract and learn high-level features from clinical data.

Hybrid models combining machine learning and deep learning, such as CNN-LSTM or MLP-based systems, have demonstrated higher accuracy and robustness.

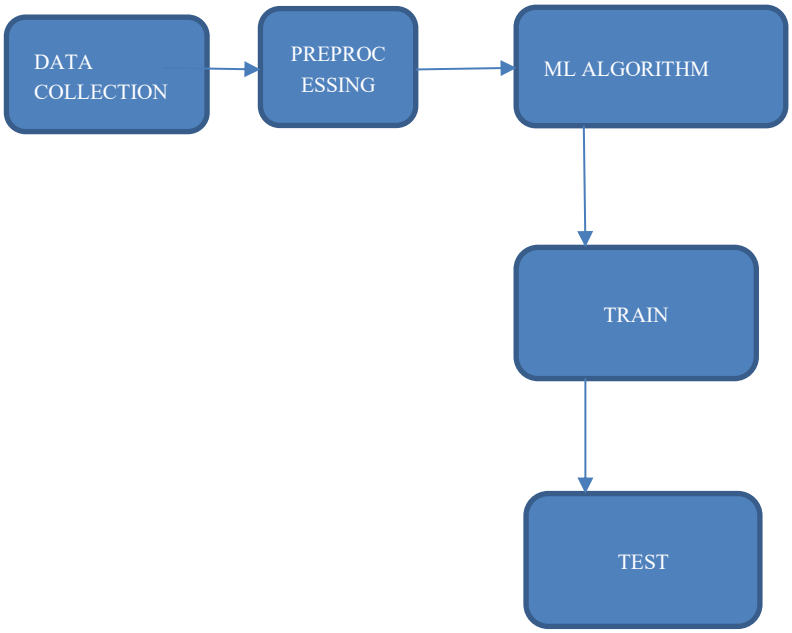
3.2 PROPOSED SYSTEM:

The proposed system aims to develop an intelligent, machine learning-based solution for predicting heart disease based on clinical data. The system will use patient information such as age, gender, blood pressure, cholesterol levels, and other key health indicators to assess the risk of heart disease. The process begins by collecting and preprocessing the data to remove inconsistencies and prepare it for analysis. Machine learning

4.MODULE DESCRIPTON:

Module involves gathering relevant clinical data required for heart disease prediction. This module is responsible for cleaning and preparing the dataset for analysis.

BLOCK DIAGRAM



5. SOFTWARE DESCRIPTION:

Each software leverages machine learning algorithms to detect the disease in patients based on medical and lifestyle data. It is designed to assist healthcare professionals in early diagnosis and preventative treatment by analyzing patterns in patient records.

5.1.PYTHON:

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. Python code tends to be shorter than comparable codes. Although Python offers fast development times, it lags slightly in terms of execution time. Compared to fully compiling languages like C and C++, Python programs execute slower. Of course, with the processing speeds of computers these days, the speed differences are usually only observed in benchmarking tests, not in real-world operations.

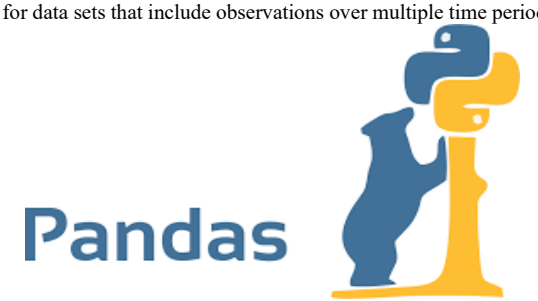
5.2.NUMPY:

NumPy stands for numeric python which is a python package for the computation and processing of the multidimensional and single dimensional array elements.

NumPy in combination with SciPy and Mat-plotlib is used as the replacement to MATLAB as Python is more complete and easier programming language than MATLAB. Prerequisite Before learning Python Numpy, you must have the basic knowledge of Python concepts.

5.3.PANDAS:

Pandas are a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.[2] The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

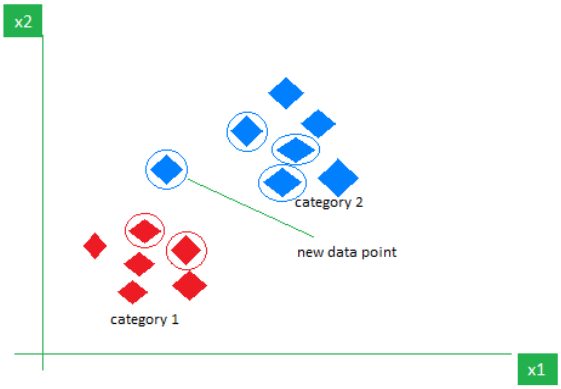


5.4.SCKIT-LEARN:

Scikit-learn is probably the most useful library for machine learning in Python. It is on NumPy, SciPy and matplotlib, this library contains a lot of effiecient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction. Scikit-learn is an open source Python library that has powerful tools for data analysis and data mining.

5.5.K-NEAREST NEIGHBOUR:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.It assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.



K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data. It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

6.CONCLUSION:

This demonstrates the potential of machine learning in the healthcare domain, specifically for the prediction of heart disease using clinical parameters. By utilizing Python and popular libraries like NumPy, Pandas, and Scikit-learn, and implementing algorithms such as K-Nearest Neighbors (KNN), XGBoost, and Multi-Layer Perceptron (MLP), the system successfully predicts the presence of heart disease with high accuracy. Among all models, XGBoost delivered the best results due to its advanced gradient boosting mechanism and ability to reduce overfitting.

REFERENCES:

1.A. S. Abdullah and R. R. Rajalaxmi, "**A data mining model for predicting the coronary heart disease using random forest classifier"**, Proc. Int. Conf. Recent Trends Comput. Methods Commun. Controls, pp. 22-25, Apr. 2012.

2. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori and M. Abdar, **"Using PSO algorithm for producing best rules in diagnosis of heart disease"**, Proc. Int. Conf. Comput. Appl. (ICCA), pp. 306-311, Sep. 2017.

3.N. Al-milli, **"Backpropogation neural network for prediction of heart disease"**, J. Theor. Appl.Inf. Technol., vol. 56, pp. 131-135, 2013.

4.C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya and R. Deepika, "**Analysis of neural networks based heart disease prediction system**", Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), pp. 233-239, Jul. 2018.

5. P. K. Anooj, **"Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules",** J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27-40, Jan.2012.

6. C.-A. Cheng and H.-W. Chiu, "**An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database"**, Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), pp. 2566-2569, Jul. 2017.