International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST52310382

Heart Disease Prediction using Machine Learning

Prof. Sachin Sambhaji Patil[1], Vaibhavi Dhumal[2], Srushti Gavale[2], Himanshu Kulkarni[2], Shreyash Wadmalwar[2]

[1]AssistantProfessor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

[2]BE Students, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 05 May 2023
Published: 24 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
398-404

ABSTRACT

Heart disease is a significant global health concern, accounting for a substantial number of deaths worldwide. Early prediction and detection of heart disease play a pivotal role in improving patient outcomes and reducing mortality rates. Machine learning techniques have demonstrated considerable potential in accurately predicting heart disease based on patient data. In this research paper, we propose a novel approach to heart disease prediction using machine learning algorithms, with a particular focus on creating a user-friendly graphical user interface (GUI) for enhanced accessibility and ease of use. The proposed approach leverages a diverse dataset encompassing demographic information, medical history, laboratory results, and diagnostic tests, providing a comprehensive view of a patient's health status. Multiple state-of-the-art machine learning algorithms, including logistic regression, support vector machines, random forests, and artificial neural networks, are employed to build robust prediction models. These models are trained, validated, and evaluated using appropriate performance metrics to ensure accuracy and reliability. To facilitate practical implementation, a user-friendly GUI is designed to provide an intuitive interface for healthcare professionals and individuals without extensive programming or machine learning expertise.

Keywords: Heart disease, Graphical User Interface, Machine Learning, Logistic Regression, KNN, accuracy

I. INTRODUCTION

Heart disease remains one of the leading causes of mortality and morbidity globally, imposing a considerable burden on healthcare systems and societies at large. Timely prediction and identification of heart disease risk factors are critical for effective management, prevention, and early intervention. Traditional risk assessment models heavily rely on statistical techniques and clinical expertise. However, the increasing availability of vast amounts of patient data and the advancements in machine learning offer

new opportunities for accurate prediction and risk assessment.

Machine learning algorithms have demonstrated great potential in the healthcare domain, enabling predictive models to extract valuable insights from complex and diverse datasets. These algorithms can analyze large volumes of patient data, including demographic information, medical history, laboratory results, and diagnostic tests, to identify patterns and relationships that may contribute to heart disease. By leveraging these patterns, machine learning models can effectively predict the likelihood of heart disease and help healthcare professionals make informed decisions.

While machine learning models have shown promising results in heart disease prediction, their adoption and practical implementation can be challenging for healthcare professionals and individuals without specialized knowledge in programming or data analysis. Therefore, the development of user-friendly interfaces, such as graphical user interfaces (GUIs), can bridge this gap and facilitate the utilization of machine learning algorithms in real-world scenarios.

In this research paper, we propose a GUI-based approach for heart disease prediction using machine learning techniques. Our objective is to create an intuitive and accessible interface that allows healthcare professionals and individuals to input their health data easily and obtain reliable predictions regarding their heart disease risk. The GUI will serve as a user-friendly tool that leverages the power of machine learning algorithms to provide accurate risk assessments and support proactive management strategies.

To achieve this, we will employ a diverse dataset encompassing various clinical features, including age, gender, blood pressure, cholesterol levels, and other relevant parameters. Several state-of-the-art machine learning algorithms, such as logistic regression, support vector machines, random forests, and artificial neural networks, will be employed to build prediction models. These models will be trained, validated, and evaluated using appropriate performance metrics to ensure accuracy and reliability.

The developed GUI will provide an interactive platform where users can easily input their health information, initiate the prediction process, and obtain personalized risk assessments for heart disease. The system will utilize the trained machine learning models to analyze the input data and present the results in a clear and understandable format, allowing users to make informed decisions and take necessary preventive measures.

II. LITERATURE SURVEY

In recent years, numerous experiments and research have been conducted in the field of medical science and machine learning, leading to the publication of significant papers.

[1]A paper entitled "Prediction of Heart Disease using Machine Learning Algorithms" was proposed by Aadar Pandita and Sarita Yadav. They used Several Machine Learning algorithms such as Naïve Bayes, K-Nearest Neighbor, Decision Tree and Random Forest are correlated to find the most precise model. In this project the algorithm is given with input variables and actual output obtained then algorithm compares between the actual and predicted output to identify errors and modifies the model precisely. The heart disease database is from the UCI repository. The accuracy of their model is 0.9907. And they conclude it with the statement that with the help of KNN algorithm we can know the patients who are suffering from heart diseases.

[2] M.Snehith Raja, M. Anurag, Ch. Prachetan Reddy proposed a paper “Machine learning based heart disease prediction system”. The study relies on the analysis of data collected in SAQ. They used Several Machine Learning algorithms such as KNN, Random Forest, K-mean clustering, Decision Trees. The algorithm constructs N of Decision trees and outputs

the class that is the average of all decision trees output.So, accuracy of prediction at early stages was achieved effectively.

[3] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar proposed a paper titled "Prediction of Heart Disease Using Machine Learning" with the aim of developing an application capable of predicting heart disease vulnerability based on basic symptoms. They used ANN, multi-layer perceptron (MLP) to train and test the dataset. The system was developed using python code using PyCharm IDE. With the help of python library sci-kit learn, the system was implemented successfully.

[4] Proposed by Baban U. Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagare, and Manisha Darade, the paper titled "Heart Disease Prediction Using Machine Learning" was put forward.They used some machine learning algorithms such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM). This project aims to know whether the patient has heart disease or not. The records in the dataset are divided into the training set and test sets. After pre-processing the data. The data classification technique namely supports vector machine, artificial neural network, random forest was applied. In the project, proper data processing was employed for the analysis of the heart disease patient dataset. The three models were trained and tested, achieving the following maximum scores: Support Vector Classifier with 84.0%, Neural Network with 83.5%, and Random Forest Classifier with 80.0%.

III.METHODOLOGY

In this model, four algorithms are used to predict the output to check whether a person has a heart disease or not. These algorithms are:

Logistic Regression- Logistic regression is a statistical modeling technique used for binary classification problems.It models the relationship between the predictor variables and the binary outcome using the logistic function.The logistic function maps the linear combination of predictors to a probability value between 0 and 1.

kNN- K-nearest neighbors is a non-parametric classification algorithm that determines the class of a new instance based on its nearest neighbors in the training dataset.The algorithm assigns a class label to a test instance by majority voting among its K nearest neighbor.The value of K determines the number of neighbors considered for the classification decision.

Decision tree- Decision trees are hierarchical structures that make decisions based on a set of conditions or rules. The tree is constructed by recursively partitioning the data based on the most informative features. At each node, a feature and a threshold value are selected to split the data, creating branches.

Random forest- Random Forest is an ensemble learning method that combines multiple decision trees. It creates a collection of decision trees by randomly sampling the training dataset with replacement (bootstrapping) and selecting random subsets of features. For prediction, the final outcome is determined by aggregating the predictions of all individual decision trees (e.g., voting or averaging).

IV.SYSTEM ARCHITECTURE

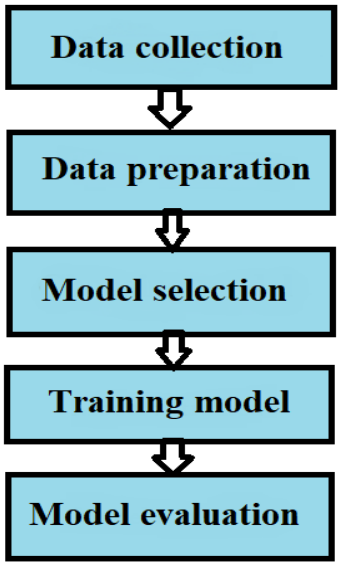

Fig 1. Workflow of the model

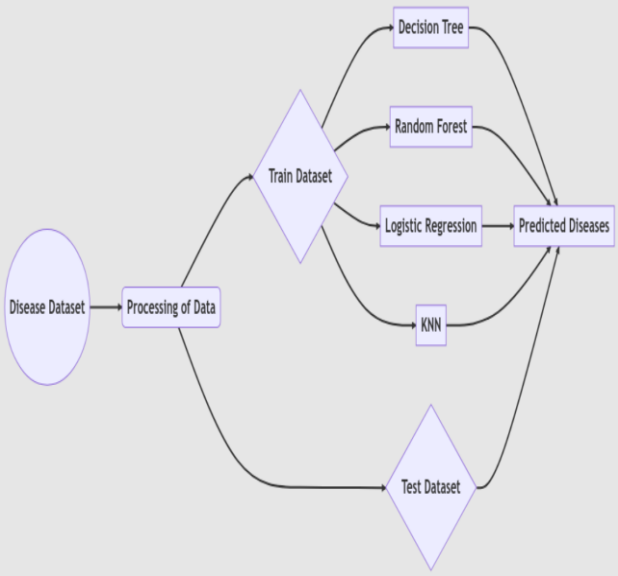

Fig 2. Systemarchitecture of the model

1. Disease Dataset Collection

The first step in designing a system for heart disease diagnosis is to collect a dataset of patients with the disease of interest, in this case, heart disease. The dataset contains a variety of patient attributes, such as age, gender, blood pressure, cholesterol levels, medical history, and other relevant factors. The dataset can be collected from medical records, hospitals, or generated through clinical trials or surveys. The larger the dataset, the better the accuracy of the algorithms, as the algorithms can learn more from a larger dataset.

The dataset used in this system was downloaded from Kaggle:
https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset

2. Data Preprocessing

After the dataset has been collected, the next step is to pre-process the data. Pre-processing involves cleaning the data, dealing with missing values, and normalizing or standardizing the data to make it easier for the algorithms to process.

3. Train Dataset and Test Dataset Split

The next step is to split the dataset into a train dataset and a test dataset. The train dataset is used to train the machine learning algorithms, while the test dataset is used to evaluate their performance. The split ratio between train and test datasets can be 70:30, 80:20, or 90:10 depending on the size of the dataset. We used the ratio of 80:20 for splitting the data. The split ensures that the algorithms learn from a subset of the data and are evaluated on a separate subset to avoid overfitting and underfitting.

4. Implementing Algorithms

The next step is to implement machine learning algorithms to predict whether a patient has heart disease or not. In this case, decision tree, random forest, logistic regression, and K-nearest neighbor (KNN) are the algorithms of interest. Each algorithm is trained using the train dataset and evaluated using the test dataset. The accuracy of each algorithm is compared, and the algorithm with the highest accuracy is selected for further evaluation.

5. Predicted Output

The final step is to predict whether a patient has heart disease or not based on their attributes using the selected algorithm. The algorithm takes the patient attributes as input and outputs a probability value or a binary value indicating the presence or absence of heart disease. The predicted output is further evaluated using metrics such as accuracy, precision, recall, and F1 score to determine the performance of the algorithm.

A graphical user interface (GUI) is designed to detect heart disease by incorporating various machine learning models that analyze patient data. In this system, GUI was designed using a python module “tkinter”. After entering the required medical risk factors, it provides the output by showing “Possibility of heart disease” or “No Heart Disease”.

V. RESULT

The accuracies obtained by the algorithms after implementing the model are as follows:

i. Logistic Regression – 90.16 %
ii. KNN – 86.88 %
iii. Decision Tree – 80.32 %
iv. Random Forest – 86.88 %

Out of all the algorithms used, Logistic regression predicted the output with the highest accuracy i.e., 90.16 %

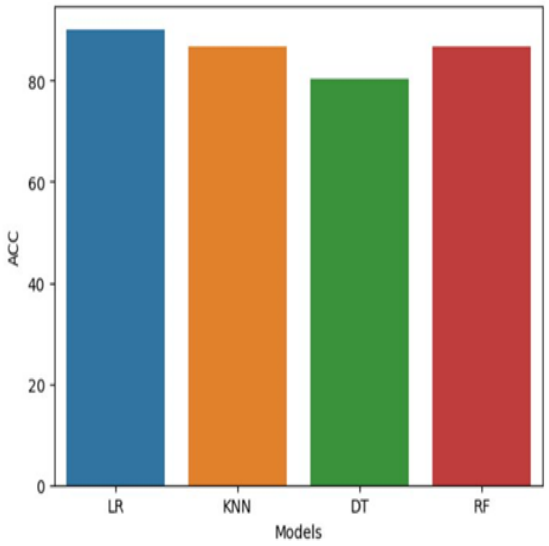

Fig 2.Accuracies obtained bythealgorithms

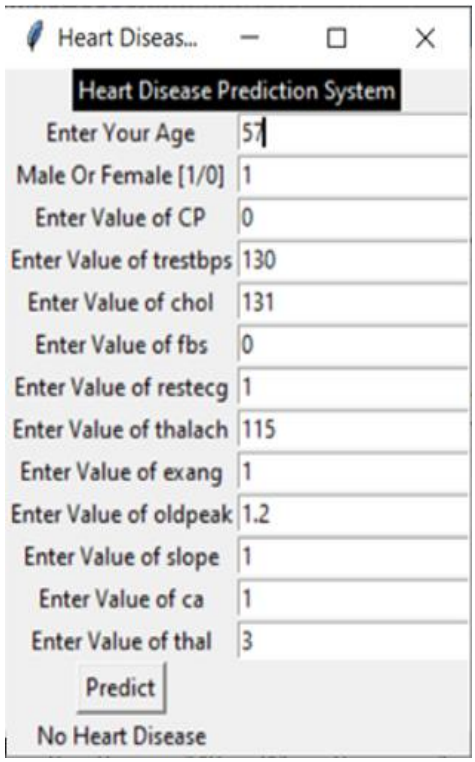

Fig 3. GUI result when no heart disease

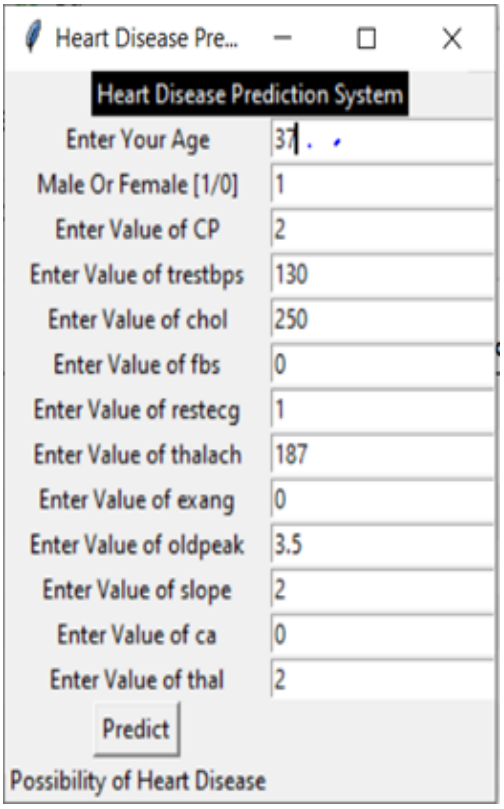

Fig 4. GUI result when possibility ofheartdisease

VI.DISCUSSION

The use of machine learning techniques for heart disease prediction, coupled with a user-friendly graphical user interface (GUI), has significant implications for improving healthcare outcomes and empowering individuals in managing their heart health. This study proposed a comprehensive approach to heart disease prediction, involving data collection, data preparation, model selection, model training, and model evaluation.

Overall, this study demonstrated the potential of machine learning techniques in accurate heart disease prediction. The selection of appropriate models, such as the Random Forest, Logistic Regression, Decision Tree, and KNN algorithms, allowed for robust and reliable predictions. The incorporation of a user-friendly GUI enabled seamless interaction and accessibility, promoting individual awareness and proactive heart health management.These findings hold significant implications for healthcare professionals, individuals, and healthcare systems at

large. Early detection, proactive intervention, and personalized care can improve patient outcomes and reduce the burden on healthcare resources. By bridging the gap between predictive algorithms and practical implementation, this research contributes to the advancement of cardiovascular healthcare, potentially saving lives and enhancing overall well-being.

VII. CONCLUSION

In conclusion, this research explored the application of machine learning techniques for heart disease prediction, complemented by a user-friendly graphical user interface (GUI). Through a comprehensive methodology involving data collection, data preparation, model selection, model training, and model evaluation, the study demonstrated the potential of machine learning algorithms in accurately predicting heart disease.The evaluation of various algorithms, including decision tree, random forest, logistic regression, and K-nearest neighbor (KNN), revealed that logistic regression achieved the highest accuracy among the tested modelsi.e., 90.16%. Furthermore, the integration of a user-friendly GUI facilitated seamless interaction and accessibility for healthcare professionals and individuals.By enabling individuals to actively participate in assessing their heart health and encouraging proactive measures, the GUI facilitated the input of health data and generated personalized risk assessments.Overall, this study highlights the potential of machine learning, particularly logistic regression, in accurately predicting heart disease. By incorporating a user-friendly GUI, it fosters engagement, awareness, and proactive management of cardiovascular health. The advancements made in this research contribute to the field of cardiovascular healthcare, providing valuable tools for healthcare professionals and individuals to make informed decisions and improve patient care.

VIII. REFERENCES

[1]. Amit Jain, Suresh Babu Dongala and Aruna Kama "heart disease prediction using machine learning techniques", Publication history: Received on 07 June 2022; revised on 18 July 2022; accepted on 20 July 2022.

[2]. M.Snehith Raja, M.Anurag, Ch.Prachetan Reddy, NageswaraRao Sirisala "Machine learning based heart disease prediction system" Published on 27 Jan 2021.

[3]. Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD) "Prediction of Heart Disease Using Machine Learning", Published on 2020.

[4]. Baban.U. Rindhe, Nikita Ahire, Rupali Patil, Shweta Gagare, and Manisha Darade"Heart Disease Prediction Using Machine Learning". Published on 1 may 2021.

[5]. Pooja Anbuselvan "Heart Disease Prediction using Machine Learning Techniques". On 25th March 2017, an online review titled "Machine learning-based decision support systems (DSS) for heart disease diagnosis" was published.

[6]. C. Boukhatem, H. Y. Youssef and A. B. Nassif, "Heart Disease Prediction Using Machine Learning," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-6, doi: 10.1109/ASET53988.2022.9734880.

[7]. S. Ouyang, "Research of Heart Disease Prediction Based on Machine Learning," 2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Wuhan, China, 2022, pp. 315-319, doi: 10.1109/AEMCSE55572.2022.00071.

[8]. G. Shanmugasundaram, V. M. Selvam, R. Saravanan and S. Balaji, "An Investigation of Heart Disease Prediction Techniques," 2018 IEEE International Conference on System,

Computation, Automation and Networking (ICSCA), Pondicherry, India, 2018, pp. 1-6, doi: 10.1109/ICSCAN.2018.8541165.

[9]. P. S. Sangle, R. M. Goudar and A. N. Bhute, "Methodologies and Techniques for Heart Disease Classification and Prediction," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225673.

[10].M. A. Jabbar and S. Samreen, "heart disease prediction system based on hidden naïve bayes classifier," 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), Bangalore, India, 2016, pp. 1-5, doi: 10.1109/CIMCA.2016.8053261.

[11]."GUI based Heart Disease Prediction system using Machine Learning", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.9, Issue 7, page no.f289-f300, July-2022

Cite this Article

Prof. Sachin Sambhaji Patil, Vaibhavi Dhumal, Srushti Gavale, Himanshu Kulkarni, Shreyash Wadmalwar, "Heart Disease Prediction using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 398-404, May-June 2023. Available at doi : https://doi.org/10.32628/IJSRST52310382
Journal URL : https://ijsrst.com/IJSRST52310382