Heart Disease Detection Using Logistic Regression

Aishwarya R[1], Chitra M P[2], Hema L Patel[3], M Ishani Kuttappa[4], Dr. Madhu M Nayak[5]
[1, 2, 3, 4] Dept of CSE
[5]Assistant Professor, Dept of CSE
[1, 2, 3, 4, 5] GSSSIETW, Mysuru, India

Abstract- *Heart disease is one of the leading causes of mortality worldwide, necessitating early and accurate detection for effective treatment and prevention. This project focuses on developing a predictive model for heart disease detection using logistic regression, a robust statistical method widely used for binary classification problems. The primary goal is to analyze patient data and predict the likelihood of heart disease based on various clinical and demographic attributes.*

Keywords- Heart disease, Mortality, Early detection.

I. INTRODUCTION

Cardiovascular diseases (CVDs) are a significant public health challenge, contributing to substantial morbidity and mortality across the globe. According to the World Health Organization (WHO), CVDs are responsible for an estimated 17.9 million deaths each year, which constitutes about 31% of all global deaths. This alarming statistic underscores the critical need for effective strategies to combat heart diseases, which include coronary artery disease, myocardial infarction, and heart failure. These conditions are particularly insidious because they often develop asymptomatically, leading to late-stage diagnosis and limited treatment options.The early detection of heart diseases is pivotal in reducing the overall burden of CVDs. Identifying at-risk individuals before the onset of severe symptoms allows for timely intervention, lifestyle modifications, and targeted therapies that can prevent the progression of the disease. Traditional diagnostic methods, such as electrocardiograms (ECGs), stress tests, and imaging techniques, are valuable but may not always be sufficient in detecting heart conditions at an early stage. These methods often lack the sensitivity and specificity needed for precise and early identification, making the quest for enhanced diagnostic tools an urgent priority

In recent years, the healthcare industry has witnessed a paradigm shift driven by the integration of advanced technologies. Machine Learning (ML), a subset of artificial intelligence (AI), hasgained prominence as a transformative force in medical diagnostics and research. ML algorithms are adept at analyzing vast amounts of data and uncovering intricate patterns that may not be discernible through traditional statistical methods.

II. LITERATURE SURVEY

Literature Survey is most important step in the software development process. Before developing the tool, it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, the next step is to determine which operating system and language can be used in developing the tool.

[1] Performance of different Machine Learning Techniques for the Prediction of Heart Diseasesby M R Sarveshvar, AninditaGogoi, Amrit Kumar Chaubey, SRohit, T R Mahesh.Published in: 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS). Machine Learning (ML), which is the most well-known sub area Artificial Intelligence, is revolutionizing the area of research. ML is primarily an subset of Artificial Intelligence (AI) that has been a crucial component of digitalization solutions that has gotten a lot of buzz in the digital world. In this work, ML is utilized to determine whether or not a person has cardiac disease. ML may be used to determine if a person has a cardiovascular illness based on particular characteristics such as chest discomfort, cholesterol levels, age, and other factors. Cardiovascular disease diagnosis can be simplified using ML classification algorithms that are based on supervised learning. To distinguish those with cardiac illness from those who do not, many ML algorithms are being used including Naïve Bayes Classifier, Logistic Regression Classifier and Random Forest.. The dataset contains certain irrelevant features that are removed during the data cleaning stage, and the data is also standardized for better results. The results and analyses of the publicly available ML Heart Disease dataset are being compared in this paper using different ML methods. The accuracy as well as confusion matrix are also being used to validate a good amount of promising outcomes.

[2] A Hybridized Model for the Prediction of Heart Disease using ML Algorithms by T Penchala Naidu, K Amar Gopal, Sk Rameez Ahmed, R. Revathi, Sk Hasane Ahammad.it is suggested that face recognition-based (mostly on student attendance systems) with OpenCV takes the

number of individuals present within the classroom and takes attendance of each of them using face detection and recognition techniques. It is efficient for students to recognize which of them are present. In this proposed system faces are detected and recognized using the Eigen object detection algorithm. done with the help of OpenCV (CVI) which has face detection present in the OpenCV inbuilt. For this, the system associates HD digital camera is needed to require the input pictures in an exceedingly fastened space wherever the camera is located. The photographs that are taken from the camera are detected with a haar cascade. The frontal faces and eyes are then trained with the Eigen algorithm, the trained faces are stored in a database first, and compared to the trained images after comparing it makes attendance to the recognized persons.

[3] Survey of Heart Disease Prediction and Identification using Machine Learning Approaches byRamya G. Franklin, B. Muthukumar. Published in: 2020 3rd International Conference on Intelligent Sustainable Systems Heart disease is highlighted as the major one among the various death factors. Detecting heart disease tends to be a bit complex due to insufficient knowledge and experience of the medical practitioners concerning warning signs of heart failure. There exist innumerable data volumes in the healthcare sector. By adopting the best appropriate data mining techniques, early detection of heart-related diseases can be achieved and also preventing it from occurring. Both the Machine Learning (ML) and Data Mining (DM) techniques prove to be effective and significant in the domain of the medical industry. identification and prediction of heart disease along with evaluating the drawbacks of the existing work. The article utilizes the DM techniques to summarize existing researches concerning heart disease prediction, examining a combination of DM techniques to reveal the best suitable and effective technique. The CNN, LSTM is being proposed for heart disease identification and prediction which yields in improvised output than the rest of the prevailing techniques. Various levels involved in the proposed approach are a collection of dataset, training, and testing, collection of user symptoms, securely forwarding the data by utilizing AES, and eventually, result is being generated in PDF format. Comparative performance in datasets concerning medicine is adopted in predicting heart disease techniques in comparison to the rest of other ML approaches. The CNN, LSTM is being proposed for heart disease identification and prediction which yields in improvised output than the rest of the prevailing techniques. Various levels involved in the proposed approach are a collection of dataset, training, and testing, collection of user symptoms, securely forwarding the data by utilizing AES, and eventually, result is being generated in PDF format.

[4]A Modern Comparison of ML Algorithms for Cardiovascular Disease Predictionby Aviral Chanchal; Ajay Shanker Singh; K Anandhan Published in: 2021 9th International Conference on Reliability, Infocom Technologies Optimization (Trends and Future Directions) (ICRITO) In today's world, with rise in number of people following a sedentary lifestyle, heart disease has become the top reason of death. It is tough for healthcare personnel to forecast such an ailment in advance because it is a complex undertaking that demands competence and a greater knowledge for prediction. We aimed this study towards looking into the topic of predicting cardiovascular disease in advance with use of different ML models, and comparing the accuracies of different older as well newer techniques, on the Framingham heart study dataset from Kaggle. We have investigated the cardiovascular disease prediction using several techniques, namely, Decision Tree, KNN, Naive Bayes, SVM, XGBoost, and Random Forest. The prediction accuracy, ROC curve, and AUC value are all used to evaluate the performance of these machine learning techniques. Based on accuracy scores, KNN, SVM, and RFC perform equally well while scoring an accuracy of 85.33%. However, an analysis of the ROC curve and AUC value provides us with a different picture of how despite slightly lesser accuracy percentages, Naive Bayes, the modern method XGBoost, and Random Forest actually perform better than the techniques anticipated earlier based on accuracy alone The research utilizes a comprehensive dataset sourced from diverse healthcare institutions, encompassing clinical and lifestyle features relevant to heart disease prediction.

5] Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection Tahseen Ullah Ahmadkhan, Syed Irfan Ullah, Khalil Ullah, Muhammad Ishaq, Yazeed Yasin Ghadi And Abdul Mohsenalgarni. The paper titled "Prediction of Heart Disease Using Machine Learning Algorithms" discusses the application of various machine learning (ML) algorithms to predict heart disease, which is a leading cause of death globally. The authors, Islam Aabdalla, Islam Daoud Suliman, and Dr. D.Vasumathi, aim to enhance early detection of cardiac diseases through the use of classification models such as K-Nearest Neighbors (KNN), Decision Trees (DT), Naive Bayes (NB), Logistic Regression (LR), and Random Forest (RF). The study utilizes five cardiac disease datasets, including the Cleveland dataset, and employs techniques like chi-square feature extraction to improve model accuracy. The results indicate that Logistic Regression achieved the highest accuracy of 96.67% on the Cleveland dataset after multiple iterations, outperforming other algorithms. The paper emphasizes the importance of machine learning in healthcare, particularly in identifying individuals at high risk for cardiac events. It also reviews existing literature

on predictive algorithms for heart disease, highlighting the need for improved accuracy in detection methods. The methodology section details the data preprocessing steps, including handling missing values and feature selection, and describes the classification techniques used. Overall, the study contributes to the field by comparing different ML models.

III. METHODOLOGY

1 Data Collection and Preprocessing Data Source: The dataset for this project is sourced from the UCI Machine Learning Repository, which provides a well-documented heart disease dataset. This dataset includes a variety of features such as age, gender, cholesterol levels, blood pressure, and other relevant medical parameters. Data Preprocessing: Preprocessing steps involve handling missing values through imputation or removal, encoding categorical variables using techniques like one-hot encoding, normalizing numerical features to standardize the scale, and splitting the dataset into training and testing subsets. This ensures that the model is trained on a diverse and representative sample of data, which is crucial for its generalizability.

2 Model Development Model Selection: Logistic regression is selected due to its interpretability and effectiveness in binary classification tasks. The logistic regression algorithm estimates the probability that a given input belongs to a specific class, making it well-suited for predicting the likelihood of heart disease. Training the Model: The model is trained using the training subset of the dataset. The training process involves optimizing the logistic regression parameters to minimize the loss function,Model Validation: Cross-validation techniques, such as k-fold cross-validation, are used to evaluate the model's performance on different subsets of the data. This helps in assessing the model's robustness and identifying any potential issues related to overfitting or underfitting.

3 Implementation on Google Colab: Google Colab is chosen as the development environment due to its accessibility, ease of use, and powerful computational resources. The logistic regression model is implemented and tested on Colab, leveraging its integration with Python libraries such as Scikit-learn, Pandas, and NumPy. Colab also facilitates collaboration and sharing of the project with other researchers and practitioners.

4 User Interface with Gradio: Gradio is used to create an interactive user interface that allows users to input their medical and demographic data and receive real-time predictions from the logistic regression model. The interface is designed to be intuitive and user-friendly, ensuring that users with no technical background can easily navigate and utilize the tool.

Interface Features:

1. Input fields for relevant medical and demographic parameters.
2. Real-time display of the prediction results.
3. Explanations and guidance based on the prediction, helping users understand their risk profile and potential next steps.

5 Model Evaluation: The final step involves evaluating the performance of the logistic regression model and the Gradio interface. Key performance metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's diagnostic capabilities. User feedback is also gathered to refine the interface and ensure it meets the needs of non-technical users.

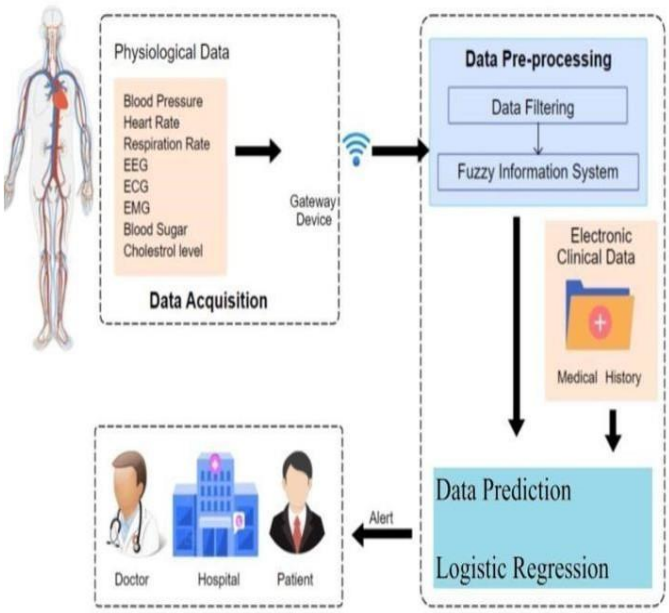


Figure 1:Design Diagram of Heart Disease Detection.

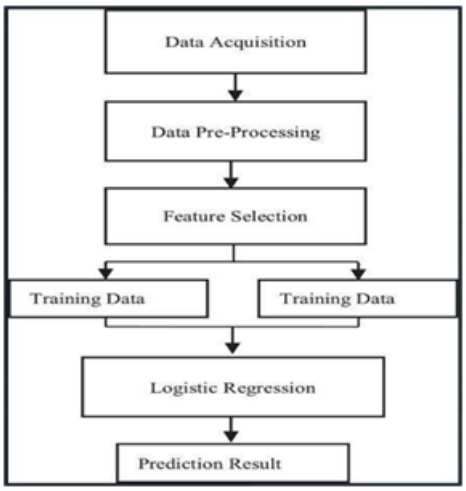
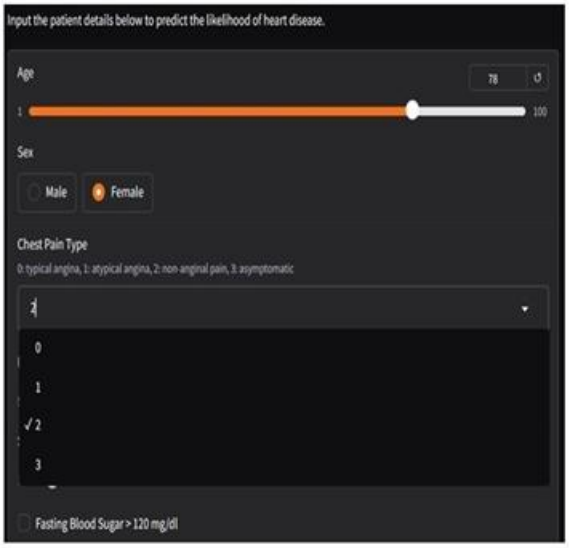
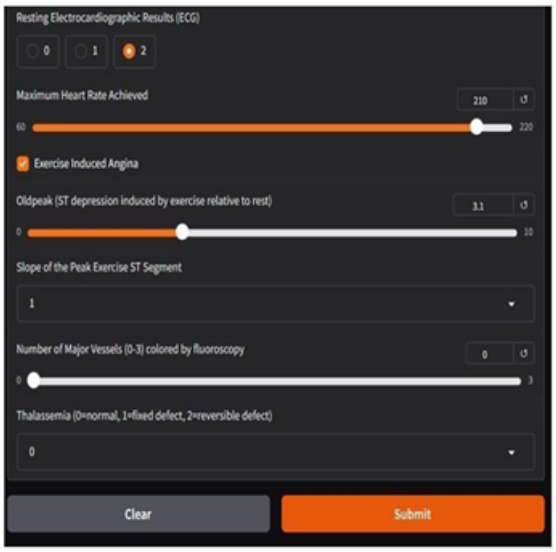


Figure 2:Data Flow Diagram of Heart Disease Prediction.

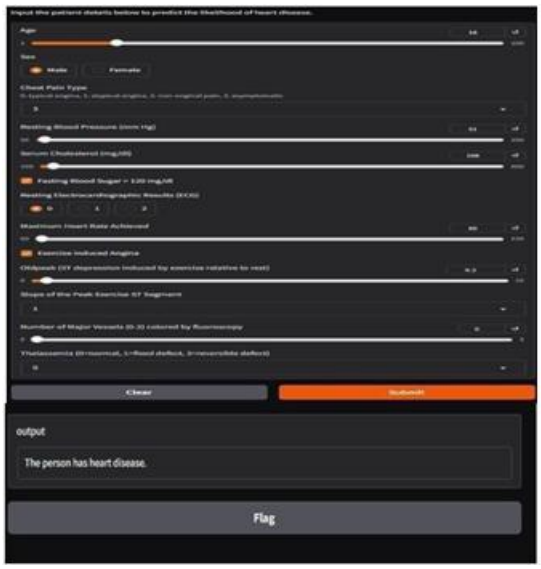
IV. SNAPSHOTS



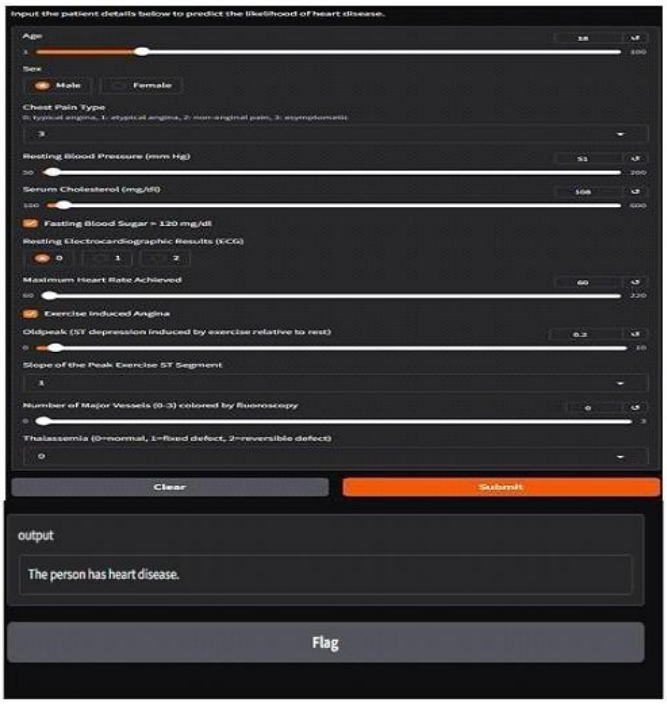
Snapshot 1: Input Page 1



Snapshot 2: Input page 2



Snapshot 3: Predicted result as patient heartdisease



Snapshot 4: Predicted result as patient does not have heart disease

V. CONCLUSION

The Heart Disease Detection project stands at the forefront of leveraging machine learning to address the complexities of cardiovascular health diagnostics. The project's emphasis on the logistic regression algorithm reflects a commitment to harnessing cutting-edge technologies to enhance early detection rates and improve diagnostic efficiency. As the project progresses, it holds promise for contributing not only to advancements in healthcare technology but also to the broader narrative of proactive health

management in the 21st century. By providing predictive insights and facilitating proactive healthcare management, the project aligns with the broader goals of reducing the burden of cardiovascular diseases on individuals and society.

The integration of ethical considerations underscores the project's commitment to responsible data handling and equitable healthcare practices. As the project advances, it is expected to offer valuable contributions to the intersection of healthcare, technology, and data science, ultimately shaping the future landscape of cardiovascular health diagnostics. Finally, ethical and collaborative efforts will be essential for the successful implementation of ML in healthcare.

Addressing biases in training data, ensuring fairness, and complying with data privacy regulations will foster trust among stakeholders. Collaboration between technology developers, healthcare providers, and policymakers will standardize practices, ensure interoperability, and create robust infrastructure for deploying these technologies. Together, these advancements will make ML an integral part of heart disease prevention and management, significantly reducing global cardiovascular mortality rates.

REFERENCES

[1] Performance of different Machine Learning Techniques for the Prediction of Heart Diseases. M R Sarveshvar; Anindita Gogoi; Amrit KumarChaubey; S Rohit; T R Mahesh Published in: 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS).
[2] A R. Alanazi, "Identification and prediction of chronic diseases using machine learning approach,"Journal of Healthcare Engineering, vol. 2022.
[3] Aatmaj Amol Salunke "Comparative Analysis of Heart Disease Prediction Models: Unveiling the Most Accurate and Reliable Machine Learning Algorithm".
[4] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, ‘‘Heart disease identification method using machine learning classification in E-healthcare,’’ IEEE Access, vol. 8, pp. 107562–107582, 2020.
[5] Zhang Y, Fogoros R, Thompson J, Kenknight B H, Pederson M J, Patangay A & Mazar S T (2011). U.S. Patent No. 8,014,863. Washington, DC: U.S. Patent and Trademark Office.
[6] A Hybridized Model for the Prediction of Heart Disease using ML Algorithms T Penchala Naidu, K Amar Gopal, Sk Rameez Ahmed, R. Revathi, Sk Hasane Ahammad.
[7] Survey of Heart Disease Prediction and Identification using Machine Learning Approaches Ramya G. Franklin; B. Muthukumar Published in: 2020 3rd International Conference on Intelligent Sustainable Systems.