Predicting the Risk of Heart Diseases from Retinal Images Using Machine Learning

Bhamidi Nagamani Chandrika[1], Dr.K. Santhi Sree[2]
[1]*Student, M. Tech, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*
[2]*Professor of CSE and Head of the Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad*

***Abstract*- Heart diseases are one of the major health issues. The death caused by heart diseases, while individual cases can sometimes be sudden or unexpected, the overall pattern of mortality due to heart diseases tends to follow certain trends and risk factors. Blood vessels are the key factor in detect heart related issues as they are spread throughout our body. Cardiovascular diseases such as hypertension and heart attacks have a significant impact on the structure and function of retinal blood vessels. These changes can be assessed using specialized tools like retinal fundoscopy, which provide images indicating the extent of damage caused by hypertension and heart attacks. Machine learning algorithms can detect preclinical signs that may not be obvious to the naked eye. The goal of our project is to investigate how hypertension and heart attacks affect retinal blood vessels by analysing datasets that include patients both with and without these heart-related conditions. The objective of this research is to innovate novel detection or monitoring methodologies that offer reduced invasiveness, heightened accuracy, cost-effectiveness, and widespread accessibility.**

I. INTRODUCTION

Heart is the core oragan of the whole body,it pumps blood through out body using bood vessels. Blockage in these blood vessels are the major concern for heart related health issues. According to a survey conducted by the World Health Organization (WHO), approximately 17.5 million individuals lost their lives to cardiovascular diseases (CVDs) in 2019, accounting for 35% of all global fatalities. Of these deaths, 86% were attributed to heart attacks. By 2023, the number of deaths due to CVDs surpassed half a billion worldwide. Early detection of CVDs is crucial to prompt medical intervention. Alarmingly, over 76% of deaths from cardiovascular diseases occur predominantly in middle-income and low-income countries.

Therefore, early prediction of cardiovascular abnormalities and the development of tools to predict heart diseases such as hypertension and heart attacks can significantly save lives and assist doctors in devising effective treatment plans, ultimately reducing mortality rates from cardiovascular diseases. With the advancement of healthcare systems, there is now a wealth of patient data available. This data can be gathered from Electronic Health Record Systems using Big Data technologies, which can then be leveraged to create predictive models for cardiovascular diseases.

This paper proposes the use of machine learning algorithms, specifically Deep Neural Networks (DNNs), to implement a system for predicting heart diseases based on heart disease prediction datasets. Data mining involves the process of extracting useful information from extensive databases, making it particularly valuable for exploratory analyses where large volumes of data need to be distilled into meaningful information.

II.OJECTIVE AND GOAL

The primary goal of this project is to create a prototype for predicting heart diseases using retinal images. This will be achieved by employing three data mining techniques: Decision Trees, Naïve Bayes, and Deep Neural Networks. These methods aim to uncover hidden insights, such as patterns and relationships related to heart disease, from a comprehensive historical database of heart disease cases. The prototype aims to streamline the diagnostic process for heart disease, offering healthcare providers valuable

support in making informed clinical decisions. This approach contrasts with traditional decision support systems, which can be cumbersome. Moreover, by enabling more effective treatments, the prototype seeks to contribute to reducing healthcare costs associated with managing heart diseases.

A supervised machine learning algorithm is employed for data classification to achieve higher accuracy in predicting outcomes. Specifically, Deep Neural Network (DNN) classification is utilized to train a dataset on heart disease and predict occurrences of heart attacks. The results demonstrate that the developed prediction system and medical prescriptions effectively forecast heart attacks. Additionally, the project utilizes two machine learning techniques, namely k-nearest neighbour (KNN) and existing multi-linear regression, to predict the Stroke Severity Index (SSI) for patients. Decision Tree, KNN, and Naïve Bayes algorithms are evaluated based on metrics such as accuracy, precision, sensitivity, and specificity. The scope of the project is dynamic and adaptable, influenced by advancements in technology, ongoing research trends, and user feedback.

III.BACKGROUND

The primary goal of this project is to enhance healthcare services by reducing costs and minimizing delays in treatment. Clinical decisions, which can sometimes be time-consuming, may result in delays in patient care. Most hospitals utilize applications like hospital information systems to manage vast amounts of healthcare and patient data, which include numbers, text, charts, and images. However, these data often remain underutilized in clinical decision-making processes.

The project aims to address the challenge of transforming this data into actionable information that enables healthcare practitioners to make more effective clinical decisions. Data mining plays a pivotal role in this process by systematically exploring large datasets to uncover patterns and extract valuable insights. By employing various statistical analyses and machine learning techniques, data mining transforms extensive data volumes into meaningful information.

Medical data mining holds immense potential for discovering hidden patterns within clinical datasets, which can significantly enhance healthcare diagnostics. Ultimately, the project seeks to leverage data mining to provide a user-friendly approach for uncovering novel insights and patterns within medical data, thereby facilitating better healthcare decision-making and improving patient outcomes.

IV.PROBLEM STATEMENT

Many hospital information systems are primarily designed to handle tasks such as patient billing, inventory management, and basic patient data storage including age, height, weight, and hereditary health issues. However, these systems have limitations in answering complex queries such as identifying preoperative risks affecting hospital stay length or predicting the likelihood of a patient developing heart disease. Clinical decisions are often reliant on doctors' experience and intuition rather than tapping into the wealth of data hidden within databases. This approach can introduce biases, errors, and unnecessary medical costs, impacting the quality of patient care. Integrating clinical data with computer-based patient records holds promise in reducing medical errors, enhancing patient safety, standardizing practices, and improving clinical outcomes.

This integration aims to streamline diagnosis times and ensure timely treatment for patients. Data modelling and analysis tools, particularly data mining, offer significant potential in creating a knowledge-rich environment that can greatly enhance the quality and effectiveness of clinical decision-making. This approach seeks to leverage comprehensive data insights to improve healthcare delivery and patient outcomes.

V.TECHNOLOGY AND TOOLS USED

1.*Python*: Python serves as the primary programming language for developing pose estimation algorithms and applications. Its rich ecosystem of libraries, including TensorFlow, PyTorch, and OpenCV, facilitates rapid prototyping, experimentation, and deployment.

2.*Deep Learning Frameworks*: Deep learning frameworks such as TensorFlow, PyTorch, and Keras provide essential building blocks for implementing and training neural network architectures for pose estimation. These frameworks offer high-level APIs, pre-trained models, and optimization tools that streamline the development process.

3. ***OpenCV***: OpenCV (Open-Source Computer Vision Library) is a versatile library that provides a wide range of functionalities for image and video processing, including image loading, transformation, feature extraction, and visualization. It serves as a foundational tool for preprocessing input data and post-processing pose estimation results.

4.Web Development Technologies:

○ HTML, CSS: HTML defines the structure of web pages, CSS styles the layout and appearance.

○ JavaScript and jQuery: JavaScript enhances the interactivity and dynamic behaviour of web applications. jQuery simplifies DOM manipulation and event handling, making it easier to write concise and efficient JavaScript code.

○ MySQL: MySQL provides robust SQL-based query capabilities, transaction support, and scalability, making it suitable for managing user profiles and admin login information.

5.API Development and Testing:

Flask simplifies the development of API endpoints, management of HTTP requests, and conversion of data between JSON format and Python objects. Its flexibility makes it a preferred choice for developers looking to create efficient and scalable web applications with ease.

VI.SYSTEM ARCHITECTURE

The System Architecture in here is classified into four modules. They are Data Pre-processing, Feature, Classification and prediction. Here, Classification and Prediction come under Algorithms

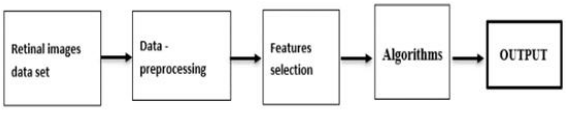

Fig.1.System Architecture

Data Pre-Processing: This file contains all the pre-processing functions needed to process all input documents and texts. First, we need to the train the datasets, test its effectiveness and validate data files then performed some preprocessing like tokenizing, stemming etc. Exploratory data analysis (EDA) involves examining the distribution of the response variable and assessing data quality by checking for missing or null values. This initial step helps in understanding the characteristics of the dataset without making any assumptions.

Features: Extraction In this file is performed for extraction and selecting methods from sci-kit learn python libraries. And for another features methods like simple bag-of words and n-grams are used where n-grams is a class extends the Python 'set' class with efficient fuzzy search for members by means of an N-gram similarity measure and it also has static methods to compare a pair of strings.

Classification: All the classifiers for the retinal images detection which are used for heart diseases prediction. The extracted features are fed into different classifiers. Naive-bayes, Logistic Regression, Linear SVM, and DNN Classifications classifiers from sklearn. Once fitting the model, compared the f1 score and then checked the confusion matrix. After fitting all the classifiers, 2 best performing models were selected as candidate models for heart diseases classification. Finally selected model was used for heart disease detection with the accuracy.

Prediction: The classifier algorithm that showed the best performance was selected as our final model and saved to disk under the name "final_model.sav". When this repository is closed, the model will be transferred to the user's machine and utilized by the "prediction.py" file for heart disease classification. The user will input an image, and the model will provide the final classification output along with the probability of accuracy.

Sequence Diagram: A Sequence diagrams are instrumental in depicting how different parts of a system interact over time, providing a dynamic view that helps in understanding system behaviour and communication patterns.

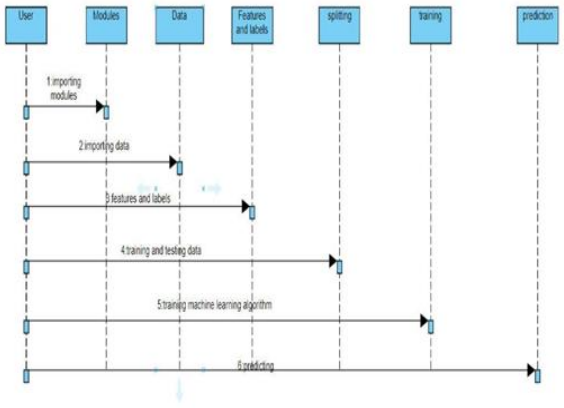

Fig.2.Sequence Diagram

Activity Diagram: An activity diagram provides a graphical representation of the sequence of actions within a system, outlining how control flows from one activity to the next. Activity diagrams are tailored to depict the dynamic behaviour of a system, offering clarity on how different activities interact and progress towards achieving specific objectives or outcomes.

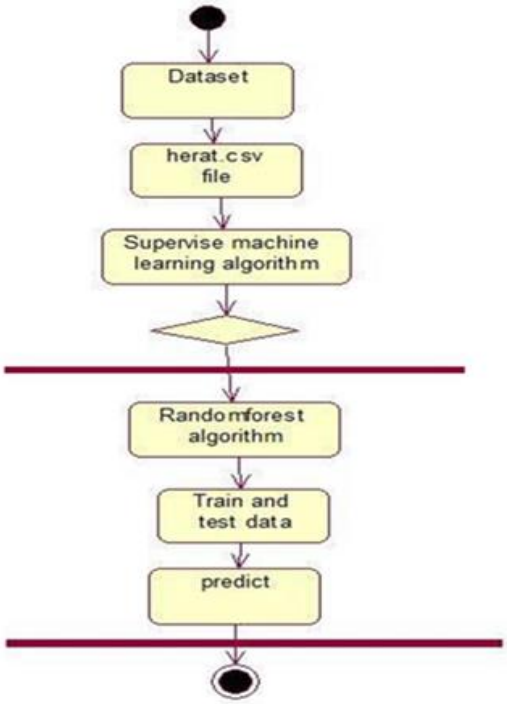

Fig.3.Activity diagram

VII.USAGE

This section provides a comprehensive guide to use predicting the heart diseases from retinal images application effectively. Whether you're a beginner getting started with the application or an experienced user looking to maximize its capabilities, this guide covers all the essential steps and features.

1. Launching the Application: Start by launching the application either through the command line or by navigating to the web interface using your preferred web browser. Ensure that all required components, Python environment, and API server, are up and running smoothly.

2. Accessing the Web Interface: Open your web browser and enter the URL provided by the pose estimation application to access the web interface. If the application is hosted locally, you can typically access it using the address http://localhost:port, where port refers to the port number specified during the application setup.

3.login: If you are the admin who is going to run the application then you can use the admin login this type of login is used by an organisation and if you are the personal use, you can login with personal login details.

4. Uploading Images: Once you've accessed the web interface, you'll typically find an option to upload images to train the application before actually using for better accuracy.

5. Initiating the application: After uploading the image, initiate the process by clicking the appropriate button or triggering the corresponding API request. The system will then analyse the input data using the pre-trained models and generate the results based on it.

6. Viewing the Results: Once the estimation is complete, the system will display the results on the web interface or provide them in the response payload of the API request. Depending on the configuration, the results may include blood pressure, probability of getting heart attack and other relevant information.

VIII.PERFORMANCE EVALUATION

$$\frac{1}{1+e^{-x}} = \frac{e^{x}}{1+e^{x}}$$

Logistic Regression: Logistic Regression is a widely used statistical method for predicting binary outcomes where the dependent variable y can take values 0 or 1. It's specifically designed to estimate the probability of y=1 based on given input variables x. The predictions of Logistic Regression are probabilities that indicate the likelihood of the event y=1 occurring. Logistic Regression it demonstrates the data points using sigmoid curve and it is given by the equation:

Logistic regression relies on several key assumptions and considerations:

1. Binary Dependent Variable: Logistic regression necessitates that the dependent variable is binary, meaning it has two possible outcomes.

2. Dependent Variable Representation: In binary logistic regression, the factor level 1 of the dependent variable typically represents the desired outcome or the event of interest.

3. Relevant Variables: It is essential to include only meaningful and relevant independent variables in the logistic regression model.

4. Independence of Variables: The independent variables should be independent of each other to avoid multicollinearity issues, which can distort the model's accuracy.

5. Sample Size Requirements: Logistic regression generally requires larger sample sizes compared to linear regression due to the nature of binary outcomes and the estimation process.

6. Multinomial Logistic Regression: While logistic regression is commonly used for binary outcomes, it can also extend to handle categorical dependent variables with more than two classes. In such cases, it is referred to as multinomial logistic regression.

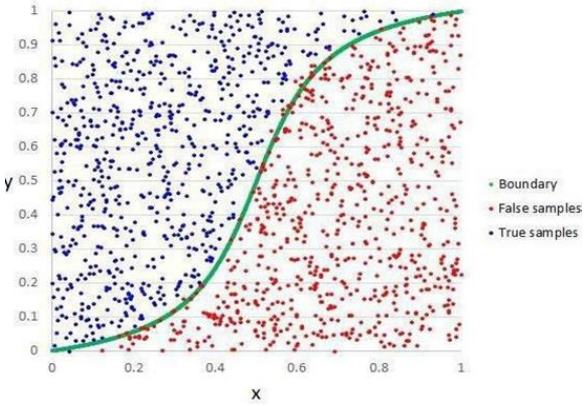

Fig.4.logidtic regression

DNN Classifications In DNN Classification, decision trees are created based on subsets of data samples. Each tree independently predicts outcomes, and a final decision is made through a voting mechanism where the most common prediction among all trees is chosen. This ensemble approach enhances performance compared to a single decision tree by reducing the risk of overfitting through averaging results.

Overall, DNN Classification is valued for its capability to handle complex data and deliver accurate predictions across diverse classification scenarios. Working of DNN Classifications with the help of following steps:

•First, begin with the selecting some random samples from a given dataset.

•Next, this algorithm will create a decision tree for every sample. Then the algorithm will prediction the result from every decision tree.

•Next step, to predict which result is the best voting will be conducted for the prediction results.

•At last, select the most voted prediction results and the winner will be the final prediction result.

The following diagram will illustrate its working-

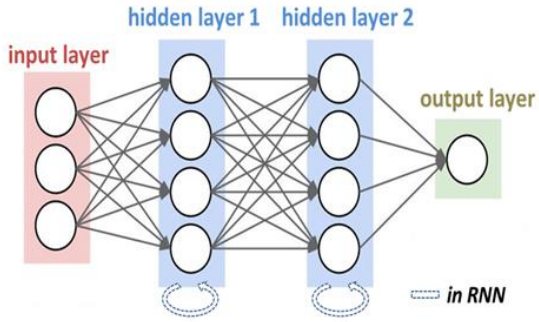

Fig.5.DNN Classification

Performance evaluation is predicted through many layers of testing before giving the actual accuracy. The test cases performed are unit testing, integration testing, Acceptance Testing.

Unit testing: unit testing is to detect and fix bugs early in the development process, ensuring that each component works correctly and contributes to the overall reliability and quality of the software.

Integration Testing: Integration testing verifies the accuracy of interactions between different components within a system. This phase is crucial for identifying and resolving any inconsistencies or bugs that emerge during the integration of components, ensuring the overall reliability and functionality of the system.

Acceptance Testing: It is a crucial project phase that heavily involves end users. Its primary purpose is to verify that the system aligns with functional requirements.

IX.CONCLUSION

This project shows how heart diseases don't always need to be predicted using the regular known ones but the least known retinal images come in handy, as they are less time taking low of cost, which helps patients during emergency. Just this proves how advancement in technology has a great impact in everyday life. Saving ones live being the key in the today's society. With improvement in sciences and technology previous diseases which were hard to detect and hard to solve are being slowly solved and, in this process,

new and different types of problems are arising related to health. By comparing DNN and Logistic Regression, we show that DNN offers superior accuracy. The system provides quick diagnostic results, which is crucial for timely medical intervention in emergency situations. This tool not only facilitates faster diagnosis but also supports efficient healthcare delivery for heart disease patients.

REFERENCE

[1] P .K. Anooj, ―Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules‖; Journal of King Saud University – Computer and Information Sciences (2012) 24, 27–40. Computer Science & Information Technology (CS & IT) 59

[2] Nidhi Bhatla, Kiran Jyoti"An Analysis of Heart Disease Prediction using Different Data Mining Techniques".International Journal of Engineering Research & Technology

[3] Jyoti Soni Ujma Ansari Dipesh Sharma, Sunita Soni. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction".

[4] Chaitrali S. Dangare Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques" International Journal of Computer Applications (0975 – 888)

[5] Dane Bertram, Amy Voida, Saul Greenberg, Robert Walker, "Communication,Collaboration, and Bugs: The Social Nature of Issue Tracking in Small, Collocated Teams".

[6] Schmarje L., Santarossa M., Schröder S.M., Koch R. A survey on semi-, self-and unsupervised learning for image classification. IEEE Access. 2021; 9:82146–82168. doi: 10.1109/ACCESS.2021.3084358.

[7] Ankita Dewan, Meghna Sharma," Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification", 2nd International Conference on Computing for Sustainable Global Development IEEE 2015 pp 704-706. [2].

[8] Goutam B., Hashmi M.F., Geem Z.W., Bokde N.D. A Comprehensive Review of Deep Learning Strategies in Retinal Disease Diagnosis Using Fundus Images. IEEEAccess.2022doi:10.1109/ACCESS.2022.317837

[9] Wagner S.K., Fu D.J., Faes L., Liu X., Huemer J., Khalid H., Ferraz D., Korot E., Kelly C., Balaskas K., et al. Insights into systemic disease through retinal imaging-based oculomics. Transl.Vis. Sci. Technol. 2020;9:6. doi: 10.1167/tvst.9.2.6.

[10] C. Kalaiselvi, PhD, "Diagnosis of Heart Disease Using K-Nearest Neighbour Algorithm of Data Mining", IEEE, 2016

[11] Keerthana T. K., "Heart Disease Prediction System using Data Mining Method", International Journal of Engineering Trends and Technology", May 2017.

AUTHORS

First Author -Bhamidi Nagamani Chandrika, Mtech, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad

Second Author- Dr.K. Santhi Sree, Professor of CSE and Head of the Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad.