IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Heart Vigilance: Empowering Heart Health with Predictive Insights

Arshaan Shaikh, Utkarsh Rai, Yogesh Mishra, Etty Tiwari, Prof.Rajesh Gaikwad

Student, Assistant Professor

Department Of Computer Engineering,

Shree LR Tiwari Collage of Engineering, Mumbai, India

Abstract: Heart Vigilance presents a web application comprising three integral modules: a prediction model, an Optical Character Recognition (OCR) model, and a chatbot. This research focuses on recognizing the presence of heart disease in individuals by using these seven key parameters(Age, Sex, Chest pain type, BP, Cholesterol, FBS over 120, EKG results,Max HR, Exercise Angina, ST depression, Slope of ST, Number of Vessels Fluro, Thallium, BMI), validated in consultation with a cardiologist. Various prediction algorithms, including logistic regression, decision tree, and random forest, were employed and assessed for accuracy, with random forest demonstrating the highest accuracy at approximately 87%. The OCR model, leveraging technology from nano nets OCR, aids in extracting parameter information from patients' lab reports, reducing potential human errors. Additionally, a GPT-4-based chatbot is integrated into the application to provide responses to basic inquiries regarding heart disease and patient care. The amalgamation of these modules constitutes the Heart Vigilance application, designed to enhance heart disease prediction and awareness.

Keywords—Heart disease prediction, AIML, web application, prediction model, OCR, chatbot, logistic regression, decision tree, random forest, GPT-4, nanonets OCR.

I. INTRODUCTION

Heart disease poses a notable global risk to health and necessitates effective means for early detection and prevention. The mission of Heart Vigilance is to address this concern and establish a highly functional web application for the project. This introduction presents the key aspects of the project, including its purpose, objectives, methods, and significance. Heart disease is frequently a product of genetic predispositions combined with environmental and personal health conditions. It refers to several conditions influencing the heart’s shape and proper activity.

Heart disease includes coronary artery, heart rhythm, and heart defect illnesses. Early detection Of the risk factors and critical responses effectively counteract the detrimental effects of heart disease. The overwhelming prevalence of sedentary lifestyles, poor, high-fat diets, and stress has escalated heart disease awareness rates. Although medical and technological advances have provided more precise detection of the disease, many individuals remain unaware of the risks.

Therefore, By analyzing large datasets and employing sophisticated modeling techniques, the platform aims to provide accurate risk assessments and tailored recommendations for each user. Furthermore, the web application will serve as a comprehensive resource for heart health education, offering informative content on risk factors, symptoms, and preventive measures. Heart Vigilance seeks to revolutionize heart disease prevention by offering a proactive and personalized approach through innovative technology.

II.FEATURE SELECTION

The appraisal of heart infection hazards requires a strong show prepared on significant information. To start this prepare, we will utilize a comprehensive dataset, such as the Heart disease prediction dataset, which contains relevant data collected from people experiencing heart well-being evaluations

After getting the dataset, the following pivotal step is to preprocess it by evacuating any copy sections, guaranteeing the judgment of the information for demonstrate preparation. The dataset envelops different parameters crucial for precise hazard appraisal, including:

1. Age
2. Sex
3. Chest pain type
4. Blood Pressure (BP)
5. Cholesterol levels
6. Fasting Blood Sugar (FBS) over 120 mg/dl
7. Electrocardiogram (EKG/ECG) results
8. Maximum Heart Rate (Max HR)
9. Exercise-induced angina
10. ST depression induced by exercise relative to rest
11. The slope of the peak exercise ST segment
12. Number of major vessels (0-3) colored

by fluoroscopy

13. Thalassemia (Thal) - a blood disorder
14. Body Mass Index (BMI)

These parameters form the basis for training a predictive model capable of classifying heart disease risk into low and high categories. By analyzing these variables, the model can effectively identify individuals at heightened risk of developing heart disease and provide personalized recommendations for prevention and management.

III. LITERATURE SURVEY

The authors Abhijeet Jagtap, Priya Malewadkar, Omkar Baswat, and Harshali Rambade.Heart disease, a leading cause of death worldwide, poses challenges for prediction due to its complexity. Healthcare data abundance lacks effective analysis tools. A proposed automated medical diagnosis system aims to enhance efficiency and reduce costs. Utilizing data from the Kaggle and Cleveland Foundation, this project employs data mining techniques to extract hidden patterns for heart disease prediction, addressing the need for efficient and accurate machine learning methods.

The authors Apurb Rajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi, and Dr. Poonam Ghuli.Heart disease prediction, crucial in healthcare, faces increasing mortality rates. Utilizing the UCI dataset, this study employs diverse ML techniques like Naive Bayes, Decision Tree, Logistic Regression, and Random Forest. Results show Random Forest achieves 90.16% accuracy, outperforming others. Future improvements involve web application development based on Random Forest and utilizing larger datasets for enhanced predictions, aiding healthcare professionals in effective heart disease prediction.

The authors Shrutikirti Singh and Seba Susan, study examines healthcare QA systems, highlighting text-based QA models and neural networks, particularly emphasizing attention mechanisms in LSTM- and transformer-based models. Results suggest BiLSTM and transformer models with attention mechanisms provide faster, more human-like responses. Multi-lingual chatbots are deemed essential due to global connectivity. While general healthcare chatbots lack personalization, future research may focus on personalized systems like kBot for specific user needs. Specialized QA agents for disease diagnosis, including for COVID-19, are envisioned. Initiatives like RxWhyQA for large-scale clinical QA datasets are recommended for future development

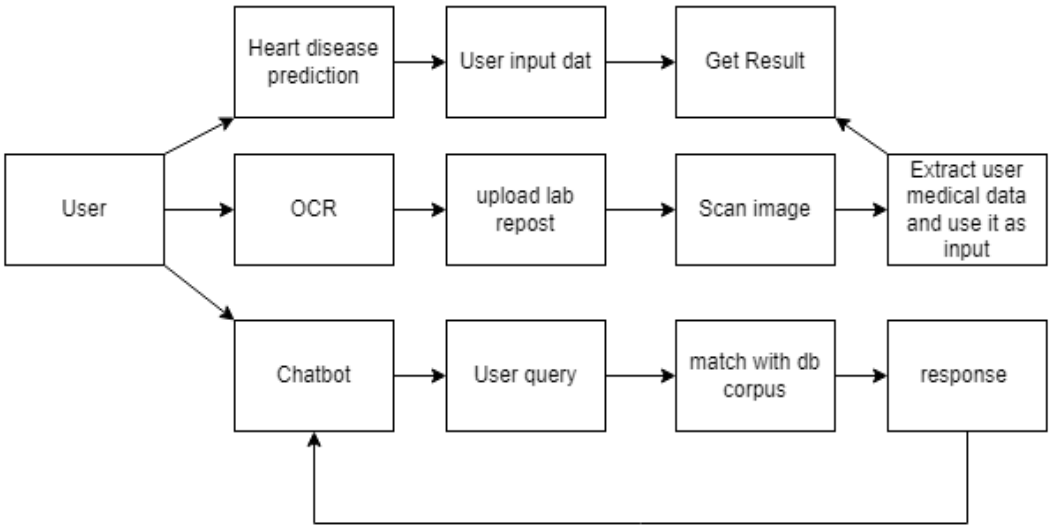
The World Health Organization identifies cardiovascular disease as a leading cause of global mortality. Researchers advocate for automated systems utilizing data mining techniques like Naive Bayes, SVM, k-NN, DT, NN, LR, RF, and Gradient Boosting to predict coronary heart disease risk. Utilizing feature selection

methods and machine learning algorithms such as Random Forest with PCA achieved 92.85% accuracy, showcasing potential for early heart disease diagnosis

IV. PROPOSED SYSTEM

The proposed system is an innovative web application designed to predict Heart Disease risk and provide preventive measures to users. This project aims to address healthcare challenges by leveraging the power of machine learning algorithms, natural language processing, chatbot technology, and Optical Character Recognition (OCR). Heart Vigilance is intended to empower individuals with the knowledge and tools necessary to proactively manage their health, reduce the risk of Heart disease, and make informed lifestyle choices.

We implement chatbot functionality in systems that can handle user queries regarding Heart diseases and their preventive measures. This chatbot can allow users to understand their disease and make lifestyle changes to improve their health. The chatbot is built using the langchain framework. The web application also contains a prediction tool that will make use of various machine



learning algorithms. This model will predict if a user is at risk of heart disease. The web application also contains an ocr module that will scan lab reports uploaded by users to extract information that can be used to make predictions if the user is at risk of a heart disease or not.

The prediction tool is trained on the following machine learning models.

Logistic Regression: This algorithm is used for binary classification to assess the probability of an individual developing Heart disease.

KNN: k-nearest neighbors (KNN) algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Random Forest: Random Forest is an ensemble method that enhances prediction accuracy by combining multiple decision trees.

Support Vector Machine: Support vector machines are supervised max-margin models with associated learning algorithms that analyze data for classification and regression analysis.

V.METHODOLOGY

Our project is multifaceted, and each aspect plays a significant role in addressing the problems of heart disease detection, prevention, and healthcare awareness.

The key methodology and contribution to the project are discussed below:

1. Machine Learning Algorithms Our approach uses machine learning algorithms; Logistic Regression, Random Forest, and Decision Trees provide the predictive analysis of complicated datasets, including the risk

of heart disease. Random Forest was the most accurate algorithm, and it was used as the main risk assessment model.

2. OCR Technology Our project uses OCR technology to facilitate user-entered data and avoid mistakes in medical diagnosis and results We used Nanonets-API to extract medical parameters.

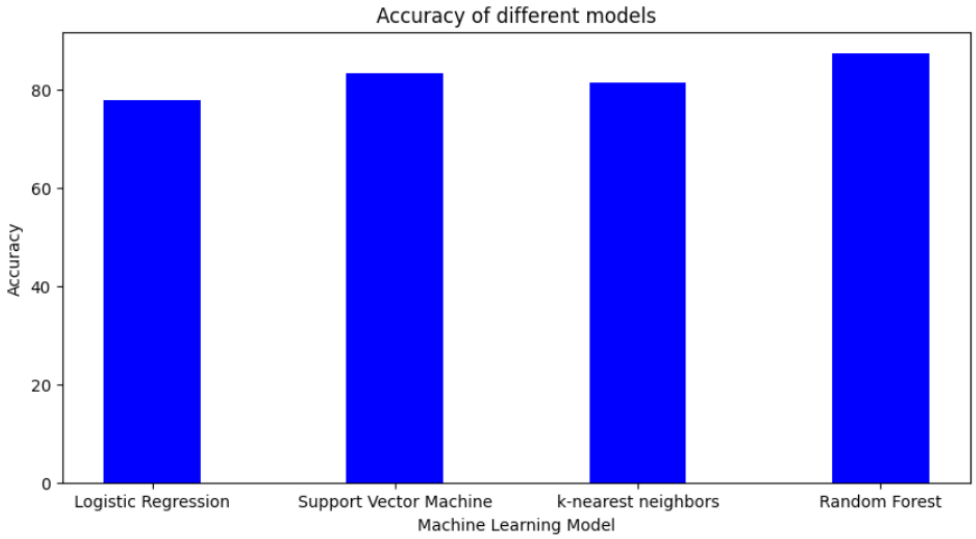
3. Development of Chatbot Interface: To further improve user engagement and interaction, our project consists of a chatbot interface developed using OpenAI's language model. The chatbot is used as a virtual assistant who responds to the user's query and gives them personalized insights and instructions on preventing and managing heart diseases. The chatbot uses the natural language processing technique to understand and interpret the query given by the user and exhibit evidence-based responses according to the medical guidelines and recommendations provided by the experts. Furthermore, this chatbot

promotes interaction and engagement with the users, facilitating getting insights, knowledge corrections, and real-time feedback on heart health to reflect the discussion-consulting with Healthcare Professionals.

4. User-Friendly Interface Implementation:

To create an effortless user experience, our project integrates an easy-to-navigate user-friendly interface created using the Flask frame. The interface allows users to input their medical details, evaluate their risk, and obtain their tailored recommendations in a few steps. It also includes data visualization and informative content to help patients make informed decisions about their health. Moreover, the interface supports many devices, resulting in a high level of usability and usage for our platform.

VI. RESULTS



Our project is multifaceted, and each aspect plays a significant role in addressing the problems of heart disease detection, prevention, and healthcare awareness.

The key methodology and contribution to the project are discussed below

1. Machine Learning Algorithms Our approach uses machine learning algorithms; Logistic Regression, Random Forest, and Decision Trees provide the predictive analysis of complicated datasets, including the risk of heart disease. Random Forest was the most accurate algorithm, and it was used as the main risk assessment model.

2. OCR Technology Our project uses OCR technology to facilitate user-entered data and avoid mistakes in medical diagnosis and results. We used Nanonets-API to extract medical parameters.

VII. CONCLUSIONS

Our project represents a major advancement in heart disease proactivity and caution management. Utilization of cutting-edge technology such as machine learning, OCR, and chatbot integration has allowed us to develop an encompassing web-based platform that can determine the risk of heart disease. The most accurate prediction method was brought to you by our study, namely the Random Forest algorithm.

Our project is multifaceted, and each aspect plays a significant role in addressing the problems of heart disease detection, prevention, and healthcare awareness.

VIII. REFERENCES

[1] Jagtap, Abhijeet, et al. "Heart disease prediction using machine learning." International Journal of Research in Engineering, Science and Management 2.2 (2019): 352-355..

[2] Rajdhan, Apurb, et al. "Heart disease prediction using machine learning." INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) 9.O4 (2020).

[3] Singh, Shruti Kirti, and Seba Susan. "Healthcare Question–Answering System: Trends and Perspectives." Proceedings of the International Health Informatics Conference: IHIC 2022. Singapore: Springer Nature Singapore, 2023.

[4] HEART DISEASE PREDICTION BY USING MACHINE LEARNING ALGORITHMS P. Shashwith*1, B. Sai Prasad*2, CH. Shashi Kumar Reddy*3,N. Abhinav Krishna*4, K. Ramesh*5-06 june 2022.

[5] Implementation of Disease Prediction Chatbot and Report Analyzer using the Concepts of NLP, Machine Learning and OCR Hiba Hussain1, Komal Aswani2, Mahima Gupta3, Dr. G.T.Thampi(4-april 2020)