

An Explainable Machine Learning approach for Heart Disease Prediction using Data Analytics and Streamlit Dashboard

T.Nikhitha Aishwarya

Dept. of Computer Science &
Engineering

Bapatla Engineering College
Bapatla, India

aishwaryathathapudi@gmail.com

Sk.Sameera

Dept. of Computer Science &
Engineering

Bapatla Engineering College
Bapatla, India

shaiksameera0830@gmail.com

Sd.Najida Maheveen

Dept. of Computer Science &
Engineering

Bapatla Engineering College
Bapatla, India

syedmaheveen26@gmail.com

T.SambaSivaRao

Dept. of Computer Science &
Engineering

Bapatla Engineering College
Bapatla, India

sambasivarao2731@gmail.com

Abstract - Heart disease is a cause of death in many parts of the world and therefore early disease diagnosis and risk evaluation is the key factor in efficient prevention and treatment of the disease. Conventional diagnostic tools usually involve some clinical knowledge and might not be readily available to the masses. This paper introduces a Heart Disease Prediction System, which uses machine learning methods to deliver precise as well as real-time risk evaluation on the basis of health parameters inputted by users. The system is deployed as a web application that is interactive and is based on Streamlit, allowing users to enter medical and lifestyle information and make immediate predictions in a user-friendly format. Several classification models such as the Logistic Regression, the Decision Tree and the XGBoost are used and the model that performs best is chosen by considering the evaluation measures like accuracy and precision. To increase transparency and user trust, the system incorporates Explainable Artificial Intelligence (XAI) with SHAP that offers information on the contribution of features to the prediction. The system also creates individual health recommendations and facilitates the automatic creation of PDF-based reports to document and analyze. The experimental findings prove that the suggested system can reach high levels of prediction accuracy, and it is also interpretable and usable. Combining machine learning, explainable AI, and web technologies, the system can be used as an

effective tool in early detection and preventive healthcare support.

Keywords - Heart Disease, Machine Learning, XGBoost, SHAP, Explainable AI, Web Application

I. INTRODUCTION

Heart disease is a major killer of human beings in the world and early diagnosis is a very crucial factor in minimizing its effects. Health condition is not known to many people until it becomes serious. The traditional mode of diagnosing the heart disease usually encompasses medical practitioners, physical examination and hospital visits, which are not necessarily easily accessible and affordable. It complicates the process of people monitoring their health regularly and detecting the risks at the early stages. With the advancement of technology machine learning has become a useful practice in the healthcare industry. It assists in the analysis of medical information and prediction of illnesses on the basis of trends and past history. The existing systems however, provide most of the predictions but not the means through which the result is obtained. Such non-transparency makes the users not trust and comprehend the result. Furthermore, the majority of systems are not user-friendly and do not offer customized recommendations on how to become healthier. In order to address these issues, this paper introduces a Heart

Disease Prediction System, a web-based program that will deliver the correct and easy-to-understand predictions. Users can use the system to type in their health data and receive on-the-fly results. It applies machine learning techniques, including Logistic Regression, Decision Tree and XGBoost to handle the data and forecast heart disease risk. The system also has Explainable Artificial Intelligence by SHAP, which aids users to know how various factors affect the prediction. The system also makes personalized health recommendations in addition to prediction and produces a report that the client can refer to in the future. These properties enhance accessibility, transparency, and usability, and the system is beneficial in the early detection and preventive care.

II. RELATED WORK

A number of studies have proposed machine learning solutions to early detection of heart disease.

Detrano et al. [1] developed a popular dataset on heart diseases that have been exploited by numerous researchers to develop predictive models. Various algorithms like Logistic Regression and Support Vector Machine have been used on this data to classify patients using medical features. Even though these models are reasonable in terms of accuracy, they do not offer a user-friendly interface and fail to explain their predictions.

Sharma et al. [2] have created a web-based heart disease prediction system in which the user is able to enter health information and the system provides an online prediction. Although the system is more accessible, it can only predict and lacks model improvement or explainable AI methods to understand the results in a deeper way.

Reddy et al. [3] suggested a prediction system of heart disease based on Decision Tree and K-Nearest Neighbors algorithm. The system assists in the recognition of the patients who are at risk on the basis of the input health parameters. It does not however feature visualization techniques and fails to offer individualized recommendations to users.

Patel et al. [4] applied a heart disease prediction model based on the Naive Bayes and Random Forest algorithms. The system aims at enhancing accuracy in prediction through comparing the various models. Nevertheless, it does not have a decent user interface and interpretation of the prediction results.

Kumar et al. [5] one of the proposed systems involves the use of Artificial Neural Networks (ANN) to predict heart disease. The model can capture complicated trends in data and offers better performance. Nonetheless, the system is a black box and does not provide a description of the way of how the prediction is done that decreases the trust of the users.

Verma et al. [6] developed a system to predict heart diseases based on data mining and classification algorithm. The system offers fundamental prediction capabilities but does not include more sophisticated capabilities, e.g., explainable AI, personalized recommendations, and report generation.

Singh et al. [7] have designed a machine learning-based system that combines various algorithms to achieve an improved prediction performance. The system enhances accuracy as it adds the outcome of the various models. Although this has been improved, it lacks visualization and easy interaction to make it more usable.

Despite the fact that these systems concentrate on various attributes of heart disease prediction, they are all lacking a comprehensive solution that combines high-performance models, such as XGBoost, explainable AI with SHAP, personal recommendations, visualization, and automated report generation in one platform. The suggested system will overcome these shortcomings by providing a comprehensive, transparent, and user-friendly system that is accurate in prediction and preventive healthcare support.

III. EXISTING SYSTEM

Currently, heart disease prediction mainly depends on the traditional medical diagnosis and the simplest machine learning models are the primary sources of heart disease prediction.

The general routine is of patients going to hospitals, taking clinical tests, and doctor evaluation. This takes time, medical experience and facilities which do not necessarily exist to all people, particularly in remote locations. In most instances, prediction systems are trained on machine learning algorithms that are trained on medical data.

The model predicts the outcome of a prediction which requires the users to key in health parameters in these systems. What is more, the majority of such systems are not created with user-friendly interfaces, so they are hard to comprehend and use by the common users. Also, the data entry is frequently manual and lacks due validation and this can influence the accuracy of findings. The other significant weakness is that in most cases, the current systems only give the final prediction and fail to explain the process through which the result is achieved. There is no possibility of users to know which factors were taken into account to make the prediction and without this, they will not trust the system. Also, the visualization and interpretation of results do not exist properly, and it is hard to make meaningful conclusions based on this prediction. Moreover, personalized health recommendations and reports are not available in most systems. The user is also provided with a barebones

output of whether the user is at risk or not without any advice how to better their health situation.

This limits the feasibility of the system in preventative healthcare in the real world. This implies that the existing strategies are characterized by the following flaws inaccessibility, inexplicability, non-interaction with the user, and the inability to make a decision, which results in the necessity to find a more developed and convenient to use alternative.

IV. PROPOSED SYSTEM

The Heart Disease Prediction System is created to address the challenges presented in the current system. It is a fully online app in which the user may easily measure his or her risk of heart disease through the application of machine learning technology. The system automates the prediction system, which used to rely on manual diagnosis and medical consultation. It incorporates data input, prediction, explainability, visualization, and report generation on one platform. This system is composed of a number of modules, and they are as follows:

A. User Input Module

In this module, use a basic and straightforward web interface created with Stream lit, which allows users to interact with the system. The interface is user friendly in that people who do not have technical or medical skills can easily operate the system. The users must fill in different health related parameters like age, gender, type of chest pains, blood pressure on rest, cholesterol level, fasting blood sugar, and maximum heart rate among other pertinent parameters. These are the required inputs in the likelihood of predicting heart disease.

The system has the proper validation of input data to prevent wrong or unrealistic values. As an illustration, numerical inputs are limited in a range of acceptable values, and categorical inputs are chosen in dropdown lists. After the user has entered all the necessary information, the system will preprocess data including scaling and formatting and send the data to the prediction module. The module makes data collection easier and allows the removal of manual medical calculations required to complete the data collection process and continue with the prediction.

B. Prediction Module

The predictive module is the central part of the system in which machine learning algorithms will be used in order to process the input of the user and identify the risk of heart disease. In the development stage, several algorithms are trained and tested on a heart disease dataset and they include Logistic Regression, Decision Tree, and XGBoost. The performance measure of all the models is in the form of accuracy, precision and recall.

Among these models, XGBoost is chosen as the ultimate model because it has better performance and can manage complicated relationships among the data. Once the user enters his or her input, the system then processes the information and inputs it into the trained model. The model then forecasts the chances of getting heart disease and classifies the finding into various levels of risk as low, moderate, or high. The module ensures fast and reliable forecasts to ensure that users can get the results in time. An optimized model is used to improve the effectiveness and the overall accuracy of the system.

C. Explainability Module

The system has a SHAP (Shapley Additive) (SHAP) explainability module that makes itself more transparent and creates trust with users. This module is an explanation of the contribution of each of the input features towards the decision, unlike in other traditional models that just give a prediction outcome only. It determines the most effective factors that raise or lower the risk of heart disease.

The visual plots and graphs are used to display the SHAP values, and thus the users can easily interpret the results. As an illustration, users can easily determine the influence of their risk level by high cholesterol or blood pressure factors. Such a degree of explanation makes users realize the rationale of the prediction, and boosts confidence in the system. The module converts the system into a black-box model into a transparent and understandable solution, which is particularly critical when the system is used in healthcare.

D. Recommendation Module

The recommendation module gives personal health recommendations depending on the prediction outputs and input parameters. The system does not just show the user whether he/she is at risk, but also provides actionable advice. The suggestions are created dynamically based on the health state of the user and the risk level that is predicted.

As an example, users that are at a higher risk level can get the advice concerning a balanced diet, exercise, lowering the level of cholesterol, stress management, and a healthcare professional. Less risky users can be provided with preventive tips in order to preserve their existing health state. This module renders the system more realistic and applicable in the real life by directing the users to make healthier lifestyle decisions. It is important in preventive health care as it causes the users to act early.

E. Report Generation Module

The system has report generation module whereby the user can save his or her prediction results in the form of a

PDF document. This characteristic can be helpful to keep records and distribute the outcomes to medical practitioners to further discuss them. Report generated will contain rich information, i.e. user input values, predicted level of risk, probability score, summary overview of the explanations and personalized recommendations.

Automatic generation of the report takes place once the prediction process is complete and it can be downloaded in one-click. The system will produce its reports automatically, so there will be no need of documentation. The users receive a concise and well structured report on their health. The fact that the system is easy to use as it allows downloading the report makes it appropriate when sharing it with healthcare professionals as well as personal use. Besides this, the report allows users to maintain a record of their health status over time that may prove to be useful in future reference and comparison. It also minimizes the possibility of errors which normally take place during manual record keeping.

The organized nature of the report makes sure that all the significant information, including input values, prediction outcomes, and recommendations, are well explained. This simplifies the process of comprehending the results by the users and doctors.

In general, this characteristic enhances the functionality of the system and its practicality in the application to real-life healthcare.

F. Visualization Module

The visualization module assists the user to comprehend the prediction in simple and straightforward manner. The system does not just present the results in text format, but in a graphical and chart mode. These images simplify the process of users to comprehend the influence of their health parameters on the forecast.

SHAP visualizations are employed to indicate the input of each feature to the ultimate prediction in this system. As an example, the user can visualize how cholesterol, blood pressure, or age level up or down contribute to the risk of heart disease. The most significant factors that affect the outcome are evident in these plots.

The system can also show simple charts in addition to SHAP plots to show the risk level or comparisons. These graphics enable users to grasp the essence of the overall output fast without necessarily having expert skills. Visualization can be used to make the system more interactive, informative and user-friendly, which enhances the experience.

G. System Integration and Workflow

The system integrates all the modules into one web based platform to facilitate easy running of the system and real time processing. Workflow starts when the user

enters input data and then validates and preprocesses it. The processed data is forwarded to the prediction module, in which the machine learning model produces the output. This is then passed to the explainability module which gives detailed information then the recommendation module which produces personalized suggestions.

Lastly, the system enables the people to see the results and download a comprehensive report. It is all done in a few seconds and gives users immediate response. Machine learning, explainable AI, visualization and reporting will be integrated to make the system efficient, reliable, and easy to use. This holistic solution renders the suggested system an effective instrument of preventive healthcare and early detection.

V.SYSTEM ARCHITECTURE AND IMPLEMENTATION

The Heart Disease Prediction System adheres to a modular design employing the machine learning, explainable artificial intelligence, and web-based technologies. The system is categorized into various layers to guarantee effective processing, scalability and maintainability. Each of the layers has a particular role, such as interaction with the user, data processing, prediction, and visualization of the results.

A. Presentation Layer (Frontend)

The presentation layer will be built with the help of the Streamlit framework that offers an interactive and convenient web interface. It allows the users to enter health-related data, including age, blood pressure, cholesterol level, and lifestyle through user-friendly elements, including sliders, checkboxes, and input fields. The frontend will take user input and provide prediction results as a risk percentage and category, and visualizations (bar charts and SHAP plots). It also gives individualized suggestions and enables the user to save a detailed PDF report. Streamlit facilitates real-time communication and updates without reloading the page thus increasing the responsiveness and usability of the system.

B. Application Layer (Backend)

Application layer performs the fundamental functionality of the system, and it is written in Python, which has a lot of support of machine learning and data processing. The input of the user is first preprocessed by converting it to a structured format that is understood by the machine learning model in this layer. To have consistency, StandardScaler is used to perform feature scaling. The processed data is then sent to the trained machine learning models which are Logistic Regression, Decision tree and XGBoost among them where XGBoost is chosen as the final model since it has better



FIGURE. 5. SHAP EXPLANATION



FIGURE. 6. BAR CHART

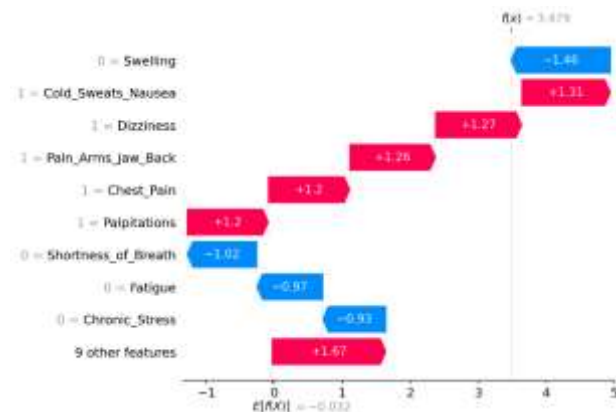


FIGURE. 7. WATERFALL PLOT



FIGURE. 8. RECOMMENDATION SECTION



FIGURE. 9. DOWNLOAD REPORT SECTION

VII. FUTURE WORK

Although the Heart Disease Prediction System is a successful performance, it is possible to suggest some

improvement that can be offered to make the performance more applicable to the real world.

A. Live Data Integration

It can be incorporated with the hospital databases and wearable health-related devices that can gather real-time patient data. It would allow monitoring continuously and predict better with the help of dynamic health parameters.

B. Mobile Application Development

Examples of the systems that can be utilized to develop a mobile application are Flutter or React Native. This would allow the users to conveniently use the system to obtain instant notification and monitor their health status on the mobile devices.

C. History Tracking And User Account

The user authentication and data storage can be introduced in such a way that users would be able to track their previous health history to trace the alterations in the level of risks over time.

D. Clinical Decision Support Integration

The system may also be expanded to support the healthcare professionals by including clinical guidelines and making a decision towards the diagnosis and treatment planning.

VIII. CONCLUSION

The Heart Disease Prediction System aims at predicting the chances of heart diseases among patients based on data analysis on their health data in advance. Rather than just focusing on giving advice on leading a healthy life, the system leverages medical parameters like age, cholesterol level, blood pressure, heartbeat, and other features to make precise predictions.

Some patients might not know about their risks because they are not familiar with the data associated with medical records. This system helps overcome this challenge by offering an assessment of heart diseases based on data. Users enter their data in the system and get instant predictions and the contributing factors to their heart disease. Additionally, this project uses an explainable AI framework called SHapley Additive explanations (SHAP) to explain how the model makes predictions based on the provided input values.

Moreover, the use of the system facilitates early detection and preventive medicine through allowing users to check their health status regularly. Based on the prediction outcomes, people will be able to act quickly and seek a doctor's advice, change their lifestyle, or go for additional tests.

Using big data analysis and machine learning techniques, the suggested system provides more opportunities for effective decision making for both patients and physicians.

IX. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Computer Science and Engineering for providing the necessary resources and support to successfully complete this project. The authors are deeply thankful to their project guide for valuable guidance, continuous encouragement, and constructive suggestions throughout the development of this work. The authors also thank the placement cell staff and students who participated in the testing and evaluation of the system.

X. REFERENCES

- [1] World Health Organization, "Cardiovascular Diseases (CVDs)," 2022. Available: <https://www.who.int/health-topics/cardiovascular-diseases>
- [2] Z. F. M. Apandi, N. H. Ngah, and N. J. Mahat, "Data Analytics Approach for Heart Disease Prediction," International Journal of Synergy in Engineering and Technology, vol. 4, no. 2, pp. 64–72, 2023.
- [3] Kumar, S., Verma, A., and Gupta, R., "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study," Journal of Healthcare Engineering, 2022.
- [4] Sharma, P., and Mehta, D., "Hybrid Deep Learning Approach for Cardiovascular Disease Prediction," IEEE Access, vol. 11, pp. 45012–45025, 2023.
- [5] Patel, K., Shah, M., and Trivedi, N., "Challenges in AI-based Healthcare Systems: Interpretability and Usability," International Journal of Medical Informatics, 2024. <https://doi.org/10.1016/j.ijmedinf.2024.105678>
- [6] UCI Machine Learning Repository, "Heart Disease Dataset," 2023. Available: <https://archive.ics.uci.edu/dataset/45/heart+disease>
- [7] M. A. Tusher, "Heart Disease Risk Prediction Dataset," Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/mahatiratusher/heart-disease-risk-prediction-dataset/>
- [8] S. R. Islam, M. M. Rahman, and M. A. Hossain, "Heart Disease Prediction Using Machine Learning Techniques," IEEE Access, vol. 8, pp. 168982–168996, 2020.
- [9] A. K. Gupta and R. Sharma, "Prediction of Heart Disease Using Machine Learning Algorithms,"

International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 6, pp. 1–7, 2021.

- [10] S. K. Yadav and P. K. Jha, "Prediction of Heart Disease Using Machine Learning Techniques," Journal of Healthcare Engineering, 2022.

- [11] National Heart, Lung, and Blood Institute (NHLBI), "Heart Disease," 2023. Available: <https://www.nhlbi.nih.gov/health/heart-disease>

- [12] Mayo Clinic, "Heart Disease – Symptoms and Causes," 2024. Available: <https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes>

- [13] Centers for Disease Control and Prevention (CDC), "Heart Disease Facts," 2023. Available: <https://www.cdc.gov/heartdisease/facts.html>