Explainable AI-Driven Chatbot System for Heart Disease Prediction Using Machine Learning

Salman Muneer[1], Taher M. Ghazal[2]*, Tahir Alyas[3],
Muhammad Ahsan Raza[4], Sagheer Abbas[5]*, Omar AlZoubi[6], Oualid Ali[7]
School of Computer Science, National College of Business Administration & Economics (NCBA&E), Lahore, Pakistan[1]
Center for Cyber Security-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia (UKM), 43600, Bangi, Selangor, Malaysia[2]
Department of Computer Science, Lahore Garrison University, Pakistan[3]
Department of Information Sciences, University of Education, Lahore, Multan Campus 60000, Pakistan[4]
Department of Computer Science, Prince Mohammad Bin Fahd University, Alkhobar, KSA[5]
College of Arts & Science, Applied Science University, P.O.Box 5055, Manama, Kingdom of Bahrain[6, 7]

***Abstract*—Heart disease (HD) continues to rank as the top cause of morbidity and mortality worldwide, prompting the enormous importance of correct prediction for effective intervention and prevention strategies. The proposed research involves developing a novel explainable AI (XAI)-driven chatbot system for HD prediction, combined with cutting-edge machine learning (ML) algorithms and advanced XAI techniques. This research work highlights different approaches like Random Forest (RF), Decision Tree (DT), and Bagging-Quantum Support Vector Classifier (QSVC). The RF approach achieves the best performance, with 92.00% accuracy, 91.97% sensitivity, 56.81% specificity, 8.00% miss rate, and 99.93% precision compared to other approaches. SHAP and LIME provide XAI methods for which the chatbot's predictions and explanations endow trust and understanding with the user. This novel approach proves the potential of seamless integration of explanations in a wide range of web or mobile applications for healthcare. Future works will extend the work on incorporating other diseases' predictions in the model and improve the explanation of those predictions using more advanced explainable AI approaches.**

Keywords*—*Heart disease prediction; machine learning; chatbot system; XAI

I. Introduction

The rapid development of artificial intelligence (AI) [1] and ML [2] technologies has made various fields, including healthcare [3], solutions to complex medical problems possible. In this regard, HD stands as an important public health issue among the leading causes responsible for mortality across the globe. Therefore, early detection and prevention of HD is important to curb the materialized global impact. AI-enforced solutions are fascinating tools for achieving such goals, with chatbots being a prominent example. Chatbots have emerged as a novel interface for healthcare applications by using conversational AI to communicate with users through real-time conversations and gather valuable health information. These intelligent agents driven by ML models can assess risk factors, reply to questions, and provide health predictions based on user inputs. The promise of Chatbots is most evident in heart disease prediction as they act as virtual assistants to surface early red flags and motivate preventive actions. However, a significant concern surrounding the implementation of these chatbots in life-critical health scenarios is their ability to elucidate their predictions in a way that is transparent and comprehensible to end users. The application of chatbots, empowered by ML techniques, ensures a broader range of capabilities for an interactive conversation system.

*Corresponding Author.

Traditional ML models [4] have proven efficient predictors of clinical scenarios, they remain largely "black boxes" whose inner workings are often not transparent enough to fully account for how certain decisions are arrived at. Such opacity may generate doubts regarding trust and reliability because healthcare demands insight into the reasoning behind a prediction. To avert these challenges, XAI [5] has been fashioned to augment the interpretability of the decision-making processes involved in these models. By providing greater insight into variables predicting heart disease, XAI helps in decision-making by healthcare practitioners and parties involved thus engendering trust in AI solutions.

Explainable AI helps in this regard by allowing chatbots not only to predict HD risk but also to explain the reasons behind their predictions understandably. In contrast with the normal "black box" ML model for which decisions are explained in contextual terms, XAI-driven chatbots such as this can potentially explain why some attributes (blood pressure, cholesterol levels, etc.) lead to a simultaneously unique heart disease risk given all patient profiles: age ranges and themselves's lifestyle choices. Transparency will be an essential aspect of healthcare. It builds trust and clarifies the rationales of AI recommendations to patients as well as providers. XAI-based chatbots [6] for heart disease prediction [7] combine strengths from both conversational agents and sophisticated predictive models. These chatbots aim to interact with users in dialogue, collecting the required health information and providing instant risk estimates. The chatbot incorporates XAI methods to explain how some health measurements, lifestyle habits, or risks in custom history contribute to the overall risk profile, thus providing a transparent vision of the prediction process. These together not only enhance user experience but also make sure that the predictions remain interpretable and actionable. Moreover, the drifting of implementing XAI into HD prediction models coincides with the growing concern in ethical AI, and the necessity of fair systems [8], accountability [9], and transparency [10]. While such qualities are critical in every domain, they gain special significance in healthcare since they may directly affect patient outcomes as well as trust in AI-driven recommendations. XAI can allow clinicians to verify model predictions, see where biases occur, and make adjustments, ensuring results are fair and accurate across patient groups via the chatbot's suggestions. In turn, this works to encourage patients to adopt the recommended lifestyle changes or prompts for further medical evaluation.

The usage of XAI with chatbots for HD prediction [11], the primary advantages would be to improve patient education and engagement. It helps them better understand why and how age, smoking behavior, activity level, or dietary habits could increase the risk of HD. This understanding empowers people to take charge of their health and care and achieve better outcomes overall. It not only helps diagnose [12] but also educates the patient about their condition, acting as a bridge between complicated medical information and a better understanding of patients.

The practice of such systems should be done carefully, with full consideration given to technical and clinical issues. From the technical aspects, ML algorithms used, feature selection, and model training are key to determine what will be predicted by the bot that should have high accuracy with ground truth. For the interpretability of ML models, there are different techniques like SHAP (Shapley Additive Explanations) [13] and LIME (Local Interpretable Model-Agnostic Explanations) [14]. These techniques help you identify the roles different input features play in your final prediction, and so your model decision is more scrutable. From the clinical point-of-view, it is important that this AI chatbot be tested with patient-generated data, and working closely with doctors helps to mature its design. Working alongside cardiologists, GPs and other practitioners should allow the dialog responses of this chatbot as well to help prediction models better adapt in focusing on specific patients at risk of HD. Two trade-offs yet to be considered are the model complexity and interpretation ability when the work starts deploying XAI-based chatbot service for HD prediction. Although more complex models (e.g., deep neural networks) may provide greater accuracy, they are also often less interpretable than simpler ones such as decision trees or logistic regression. Finding the right balance between accuracy and interpretability is crucial to ensure that the chatbot delivers reliable predictions while maintaining transparency. Researchers are dynamically discovering hybrid techniques that integrate the strengths of various models to obtain optimal performance.

Many studies have explored ML and Deep Learning (DL) approaches for HD prediction, but they often lack transparency, limiting their practical adoption in healthcare. The novelty of this study lies in introducing an XAI-driven chatbot system that combines accurate predictions with interpretable outputs. By focusing on transparency and user-friendly design, it bridges the gap between complex AI models and practical healthcare use, addressing the "black-box" limitations of traditional methods. This research work proposes the design of XAI chatbots for HD prediction which marks an important sign of leveraging AI for healthcare. With ML's powerful predictive features and XAI helping in bringing transparency, now healthcare systems can benefit from accurate predictions as well as making these models trustable for both patients and Healthcare Providers. Even more, improvements are planned as the field continues to grow, making these chatbots even more useful when it comes to battling HD and any other health sicknesses that last a lifetime.

The remainder of this paper is organized as follows: Section II discusses the literature review, highlighting the previous work regarding chatbot systems for HD prediction, and Section III provides limitations of previous works. Section IV discusses the proposed methodology, Section V highlights the dataset description, Section VI explains the dataset structure, Section VII explains the system block diagram, Section VIII describes pseudo code, Section IX simulation results, Section X explains the Discussion, Section XI presents conclusion of the manuscript and Section XII describes the limitations and future suggestions. This structure ensures a comprehensive understanding of the study and its contributions.

II. Literature Review

There have been extensive research interests pursued over the years on several autonomous systems, such as virtual assistants, chatbots, and applications for HD prediction employing XAI among other techniques. Their work has opened up ways for making predictive healthcare systems transparent, accurate, and ethical which provides a strong ground on top of where future advancements can be carried out. This section presents some of the most important research ever published in this field. AI-powered health chatbots are a game changer for how mental healthcare can be approached by businesses as well as any other sector including the traditional form of care within Healthcare. Powered by Natural Language Processing (NLP) and AI, these smart chatbots can converse through text as well as voice performing functions similar to those handled by a live human agent. The kinds of jobs they provide are many; everything from helping individuals in crisis to customer service. Smartphone and smart speakers-powered voice-enabled chatbots use high-end speech-to-text capabilities but are sensitive to elements like regional accents or background noise. Text-based chatbots, on the other hand, such as those found in Messenger or Slack that are available within web applications offer more managed interactions and allow for complex questions to be answered making them great tools for people looking for mental health support.

The authors highlighted that in recent years, chatbots have made remarkable progress in addressing various health-related issues, including obesity and weight management, dementia (e.g., Endurance, which engages in meaningful conversations with individuals with Alzheimer’s disease), substance use disorders [15-16], oncology care, and insomnia (e.g., Casper, a chatbot designed to assist those with sleep difficulties). They have also been utilized in prenatal services, HIV and sexual health education, and managing depression and anxiety. Notably, chatbots have demonstrated significant potential in mental health applications, contributing to suicide prevention, virtual cognitive-behavioral therapy, and psychoeducation. This study aims to explore and summarize the role of chatbots in the healthcare industry.

The authors in study [17] presented that AI chatbots have been effective in promoting treatment adherence, smoking cessation, and healthier lifestyles, with features like goal setting, monitoring, and real-time feedback. Studies showed mixed results on feasibility, acceptability, and usability, but these chatbots offer personalized services and scalability via platforms like smartphones and Facebook Messenger. Participants valued the nonjudgmental space chatbots provided, especially for sensitive communications.

The authors in study [18] said the rise of AI-empowered chatbots has also brought new security and privacy risks. Cybercriminals are using social engineering tactics like phishing through cross-platform messaging apps to create a new kind of cyber threat called "Smishing". These fake chatbots look legit but are designed to trick people into giving away sensitive personal data. To combat these threats we need to increase user awareness and be cautious with unfamiliar services. Implementing security measures like using temp tokens during chatbot sessions is crucial when devices are lost or compromised. Also, data anonymization techniques like de-identification, pseudonymization, and generalization should be used to protect sensitive data from unauthorized access. By doing all these organizations can fortify their defenses against privacy breaches and security issues with malicious chatbots.

The research also shows the increasing use of chatbots in healthcare settings to improve service and efficiency. Examples include Healthily [19] and Ada Health [20] which give users health advice and are now health information dissemination tools. In mental health, specialized chatbots like Woebot [21] are being used to deliver Cognitive Behavioral Therapy (CBT) to people with depression, PTSD, and anxiety. They also help patients with autism to learn and practice social skills as part of therapy. Beyond being therapy tools, chatbots also help with administrative tasks in healthcare facilities such as allowing patients to self-book appointments and track doctor availability. This integration in healthcare systems improves patient access and management and thus better healthcare delivery. Chatbots can also collect and process information from patients through structured questions about identity and health conditions. This information is used for patient admission, symptom monitoring, patient-doctor interaction, and documentation. Chatbots can also help with medication reminders for conditions that don't require in-person consultations. The healthcare application of chatbots not only enhances the level of patient experience and outcome but also maximizes organizational efficiency, hence ensuring that such technologies stay increasingly relevant in modern healthcare environments.

Another one of the chatbots is Gyant [22], designed to interact with users having minor non-emergency medical issues by asking questions regarding symptoms as well as general health status. An important feature of Gyant is its use of humor, human-like interaction, emojis, and memes to cope with the user successfully. However, the software is currently not available since it has been phased out. For example, Symptomate [23] is an AI chatbot, that allows patients to report their symptoms and provides them with a list of possible conditions. However, its accuracy is significantly lower than 80% compared to Ada Health. Nonetheless, this proves the fact that digital health solutions, in terms of their development and perfection, will be needed for maximum efficacy and reliability [24].

III. Limitations of Previous Work

Despite the studies in AI-powered chatbots that predict HD, it is clear that some of the previous attempts have significant disabilities. For instance, there is a severe lack of XAI, which usually leaves users and medical professionals unaware of the processes that govern the decisions they make when using a system. Moreover, most of these systems fail with issues of data privacy while dealing with delicate health information as they tend to leave their users vulnerable to breaches or unauthorized access. Another major problem is scalability as most models do not deliver optimally in various healthcare settings or with other patient populations. Moreover, the reliability of the decision-making functionality of these chatbots is incomplete, and conclusions may be as wrong as they are less encompassing. Last but not least, the problems of transparency continue to worsen the mistrust in these systems since the user does not know whether to believe the answers provided by the system. Collectively, these limitations underwrite the need for better methodologies that place more emphasis on user safety, trust, and clarity when using AI technologies in healthcare settings. A few of the limitations are shown in Table I:

TABLE I. Limitations of Previous Works

Ref.	Year	Disease Dataset	ML Algorithm/Decision Support System	XAI Implementation	Chatbot enabled System
[25]	2024	HD	DT	×	✓
[26]	2024	HD	Clinical Decision Support (CDS) system	×	✓
[27]	2023	Prostate Cancer	Hybrid Approach with NLP models (Named Entity Recognition, Intent Recognition, Sentiment Analysis, and Language Detection)	×	✓
[28]	2023	HD	XGBoost	✓	×
[29]	2023	HD	Bagging-QSVC	✓	×
[30]	2023	HD	Random Forest (RF)	×	×
[31]	2022	HD	Cox models plus symbolic regression	×	×
	2024	HD	RF	✓	✓

Table I summarizes the critical analysis of inherent limitations within previous works on the application of ML algorithms and decision support systems, particularly towards targeting HD, the prediction of prostate cancer. Much work never includes XAI techniques in their strategies, but this is a fundamental aspect of improving the transparency and trustworthiness in automated systems. For instance, although some models may lead to successful predictions of health-based outcomes, the inability to interpret may deter healthcare professionals from accepting this system for use in their service provision because they would want to understand the patterns guiding the prediction. Finally, while many works include chatbot functionalities to make interaction easier for users, it usually occurs in isolation from XAI and leads to systems that are user-friendly but less clear about their operational mechanisms. This is a disadvantage when it comes to user confidence, especially in more sensitive fields like healthcare. The deployment of advanced NLP techniques has only been moderately applied in some research works, thereby limiting the capabilities of these systems to better interact with users. Interactions may therefore not holistically deal with the intricacies of patient queries. In summary, these constraints call for further research to integrate both XAI and chatbot technologies with a focus on the construction of more solid, transparent, and user-centric healthcare solutions.

IV. Proposed Methodology

Technological advancement has intended the adoption of autonomous systems in most fields, with the application here being one of the most popular emerging applications of chatbots. Chatbots are becoming very popular, especially in the health sector, as dear guides for patients and health-related questions. However, these chatbots have encountered numerous problems such as issues related to patient medical records, insecure communication, and inability to provide clear and precise answers.

Knowing the problems calls for a rising demand to develop intelligent approaches [32-33] that would yield better results. Thus, this research work aims to provide an intelligent approach to develop a chatbot system that uses XAI. It sets out to mitigate prevailing challenges around chatbots in general with special emphasis on the healthcare industry to ensure that secure communications are implemented while providing clear responses and retrieving accurate information from patient records. The proposed chatbot system is shown in Fig. 1.

Fig. 1 represents the proposed approach, which comprises the training and validation phases. During the training phase, initially, patient data is acquired from the patient through a chatbot for secure and tamper-proof transactions, thereby enhancing data integrity and transparency within healthcare systems. The tamper-proof transactions are then forwarded to the preprocessing layer, involving normalization, handling missing values, and moving averages, and the processed data is then divided into training and testing sets, with respective ratios of 80% and 20%. Subsequently, the approach is trained on 80 percent of the data for predictive analysis using ML algorithm RF. The predictions are directed to XAI for comprehensive output explanations. If the output aligns with the predefined learning criteria, the results are stored in the cloud; otherwise, they are returned to the approach if the learning rate is not achieved.

In the validation phase, patient data is directly compared with the imported data stored in the cloud. If the criteria are met, indicating the presence of HD, the system is shown as "found". On the other hand, if the criteria are not met, signifying the absence of HD, the system is discarded.

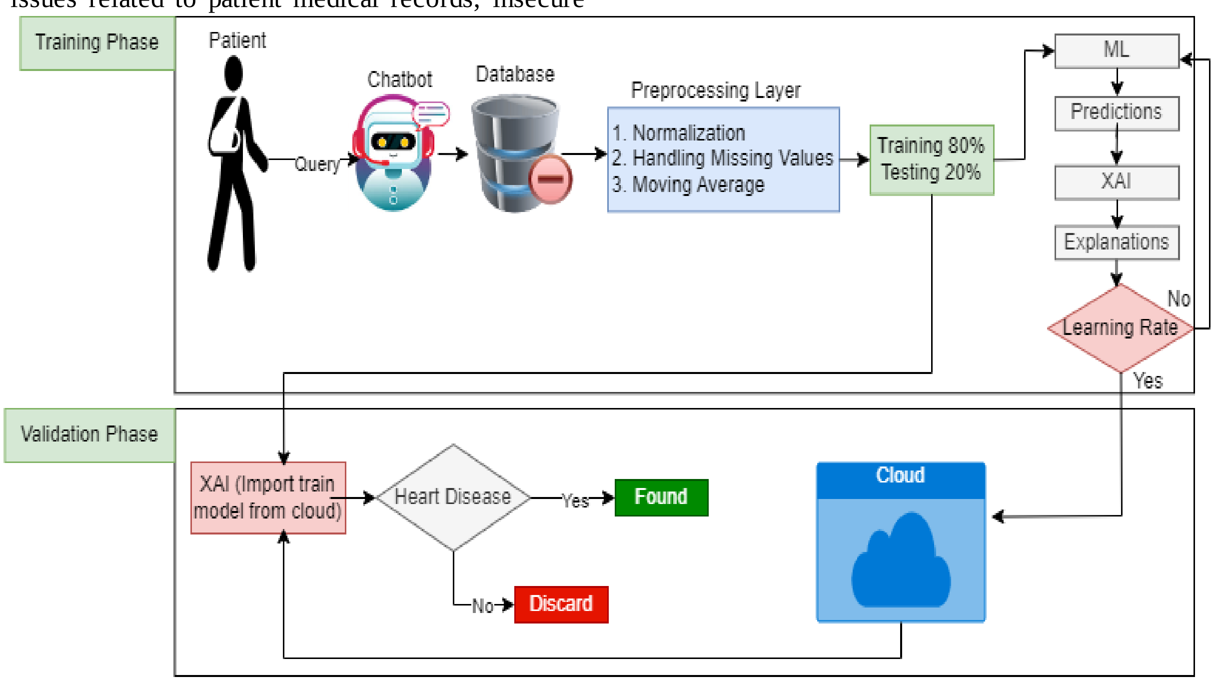

Fig. 1. Proposed chatbot system.

V. Dataset Description

The dataset utilized in this study comprises 308,855 entries & 19 features [34], providing a comprehensive overview of various health-related parameters. It encompasses features such as General Health, which is rated on a scale from 1 to 5 (with 5 indicating excellent health and 1 indicating poor health), recent Checkups (within the previous year), where 0 denotes no checkup, 1 represents a checkup within the last year, 2 for two years ago, 3 for a checkup within the last five years, and 5 for checkups more than five years ago. Exercise is captured through exercise habits, but supplemented by the prevalence of chronic diseases including HD, Skin Cancer, Other Cancers, Depression, Diabetes, and Arthritis by binary indicators where 1 means presence of the condition, and 0 means absence. Demographic information is also captured, in the Sex column where 1 means female, and 0 means male. Other determinants include stature in centimeters and weight in kilograms, which are necessary parameters in the BMI. Characteristics of lifestyle such as smoking and alcohol intake were taken into consideration, and dietary intake was reflected in fruit consumption, green vegetable consumption, and fried potato consumption. This length of variable allows for a very sharp analysis of determinants, which can thereby assess complex interrelations between lifestyle, demographic factors, and health outcomes in the prediction of heart disease with AI-driven chatbots.

VI. Dataset Structure

The Table II below reports the most important characteristics used in the dataset for health-related analysis, categorized by type of data. Such a collection of datasets contains health-related features pertaining to lifestyle habits, medical history, and demographic information. Qualitative attributes-the ones describing, for instance, "General Health," "Exercise," and "Depression"-represent statuses or histories related to a patient's health. Quantitative features refer to features whose variables include such examples as "Height (cm)" and "Weight (kg)". Attributes such as these are represented using numerical data types. They are fundamentally indispensable for developing predictive models for health outcomes, providing a complete view of someone's health profile.

TABLE II. DATASET FEATURES WITH ITS DATATYPE

Feature Name	Data Type
General_Health	Object
Checkup-(Within Previous Year)	Object
Exercise	Object
Heart_Disease	Object
Skin_Cancer	Object
Other_Cancer	Object
Depression	Object
Diabetes	Object
Arthritis	Object
Sex	Object
Age_Category	Object
Height_(cm)	Float64
Weight_(kg)	Float64
BMI	Object
Smoking_History	Object
Alcohol_Consumption	Object
Fruit_Consumption	Object
Green_Vegetables_Consumption	Object
FriedPotato_Consumption	Object

It can be seen in Table II that the predictors of medical history features and the patterns of behavior were the features necessary to form an opinion relating to risks regarding health conditions. For example, "Heart Disease," "Skin Cancer," and "Other Cancer" revealed whether any patient was diagnosed with the said diseases, while "Smoking History" and "Alcohol Consumption" reflected lifestyle decisions likely to lead to various health conditions. The indicators available in this dataset are the intake of nutrition, like "Fruit Consumption" and "Green Vegetables Consumption" to assess the dietary pattern. By integrating all these different features, the dataset offers a complete foundation to analyze health patterns and to predict the possibility of getting ill. It could be a beneficial resource in healthcare research and study.

VII. System Block Diagram

This research work proposed a system that uses historical data for prediction, as shown in Fig. 2. It applies EDA to identify whether there is a need for pre-processing and whether it identifies outliers. Pre-processing includes handling null values, duplicate values, outliers, and class imbalance. Then, the data sets are split into test and training data where 20% is designated for the test set and 80% for the training set. The model trains and tests on those datasets, and then the accuracy, precision, recall, f1-score, and confusion matrix are used to achieve the best model. Then, using the chosen model, accurate prediction about the outcome with an explanation of LIME is applied.

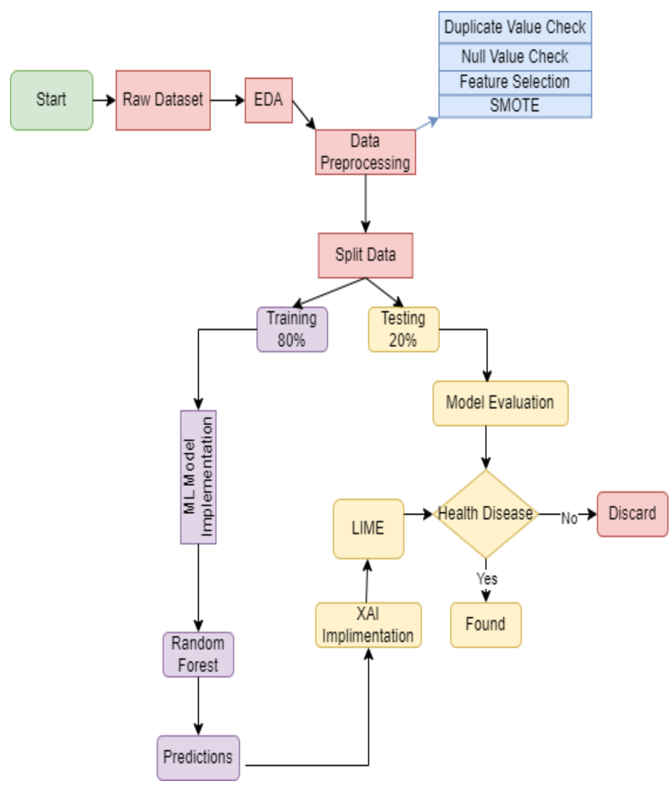

Fig. 2. System block diagram.

Fig. 2 shows that the research work illustrates an entire workflow in a system block diagram for a healthcare prediction model that involves the integration of ML and XAI methodologies. Starting from the bottom, the raw healthcare data is gathered, followed by EDA to explore the dataset and draw meaningful conclusions based on it. The data would then be subjected to thorough preprocessing with checking for duplicate and null values, selection of best features, and balancing with SMOTE. With preprocessing, the dataset was divided into the training and the test sets. For this example, the training set was used to construct the Random Forest model, while the testing set was reserved to analyze its performance. The output from the model is then analyzed to determine whether any disease is present. If a disease is identified, it is further passed over the XAI method LIME [35] to give a clear explanation of which features contributed to the prediction. This approach enhances the interpretability and transparency of the model for healthcare professionals, such that they can trust the outcomes of the model more easily. The system ends by flagging some cases where a health condition is detected or discarding cases where no disease is found and then sending it out as a responsible healthcare delivery process.

VIII. Pseudo Code

The step-by-step pseudo code for the prediction of HD is shown below in Table III.

TABLE III. PSEUDO CODE

```
START
   Step 1: Data Loading
         Load dataset from source
   Step 2: Data Preprocessing
         Perform exploratory data analysis (EDA)
         Remove duplicates and check for null values
         Handle missing values if any
         Feature selection based on relevance
         Apply SMOTE for handling class imbalance (if needed)
        Normalize/Standardize numerical features like "Height" and
"Weight"
   Step 3: Splitting Data
         Split data into Training set (80%) and Testing set (20%)
   Step 4: Model Implementation
         Select a machine learning model (e.g., Random Forest)
         Train the model on the Training set
         Evaluate the model using the Testing set
   Step 5: Model Evaluation
        Calculate performance metrics (e.g., Accuracy, Precision,
Recall)
         Check if the model detects 'Health Disease' or 'No Disease'
         IF model performance is satisfactory THEN
         Proceed to explainable AI (XAI) implementation
   ELSE
         Modify model or preprocessing steps and retrain
   Step 6: XAI Implementation (e.g., LIME)
        Generate explanations for model predictions using XAI
techniques
         Display or store the explanations
   Step 7: Make Predictions
         Input new data into the trained model
        Predict whether the patient has 'Health Disease' or 'No Disease'
         Display predictions and explanations
     Step 8: Decision and Action
           IF 'Health Disease' is predicted THEN
         Recommend further medical consultation
     ELSE
         No further action required.
 END
```

As presented in Table III, pseudo-code illustrates a systematic approach to analyzing healthcare data and making predictions regarding health conditions. It begins with loading the dataset and running preprocessing steps as follows: exploratory data analysis, and missing value handling, and it applies SMOTE to the dataset for class balance. After this step is performed, the data set is split into a training and a testing set to permit the training of an ML algorithm chosen by the researcher. After evaluating the model's performance, the code incorporates explainable AI methods to enhance the interpretability of predictions, ultimately guiding healthcare decisions based on the results.

IX. Simulation Results

This research work proposed a chatbot system for HD prediction using the Random Forest (RF) approach and implemented it on the dataset containing 308855 samples. The data were distributed into 80 % training (247084 samples) and 20% validation (61771 samples). As stated in the equations, this approach finds the result using multiple statistical measures.

$$Accuracy\ (Acc) = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Sensitivity\ (TPR) = \frac{TP}{TP+FN} \tag{2}$$

$$Specificity\ (TNR) = \frac{TN}{TN+FP} \tag{3}$$

$$Miss\ Rate\ (FNR) = 1 - Acc \tag{4}$$

$$Fall\ out\ (FPR) = \frac{FP}{FP+TN} \tag{5}$$

$$LR + ive\ (LR+) = \frac{TPR}{FPR} \tag{6}$$

$$LR - ive\ (LR-) = \frac{FNR}{TNR} \tag{7}$$

$$Precision\ or\ Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP} \tag{8}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN+FN} \tag{9}$$

TABLE IV. CHATBOT SYSTEM TRAINING PHASE FOR HD PREDICTION USING RF

	Total number of samples (49000)	Result (Output)	
Input	Expected output	Predicted Positive	Predicted Negative
		True Positive (TP)	False Positive (FP)
	229978 Positive	229750	228
		False Negative (FN)	True Negative (TN)
	17106 Negative	16806	300

It is shown in Table IV that the proposed chatbot system predicts the HD during the training period using RF. During training, 247084 samples are divided into 229978, 17106 positive, and negative samples. 229750 true positives are successfully forecasted, and no HD is recognized, but 228 records are mistakenly predicted as negatives, indicating the No Disease is recognized. Likewise, 17106 samples are obtained, with negative showing HD is identified and positive indicating No Disease. With 300 samples correctly identified as negative, showing the HD is recognized, and 16806 samples inaccurately foreseen as positive, representing No Disease is identified despite the presence of the HD.

It is shown in Table V that the proposed chatbot system predicts the HD during the validation period using RF. During validation, 61771 samples are divided into 56774, 4997 positive, and negative samples. 56736 true positives are successfully forecasted, and No Disease is recognized, but 38 records are mistakenly predicted as negatives, indicating the HD is recognized. Likewise, 4997 samples are obtained, with negative showing HD is identified and positive indicating No Disease. With 50 samples correctly identified as negative, showing the HD is recognized, and 4947 samples inaccurately foreseen as positive, representing No Disease is identified despite the presence of the HD.

TABLE V. CHATBOT SYSTEM VALIDATION PHASE FOR HD PREDICTION USING RF

	Total number of samples (49000)	**Result (Output)**	
Input	Expected output	Predicted Positive	Predicted Negative
		True Positive (TP)	False Positive (FP)
	56774 Positive	56736	38
		False Negative (FN)	True Negative (TN)
	4997 Negative	4947	50

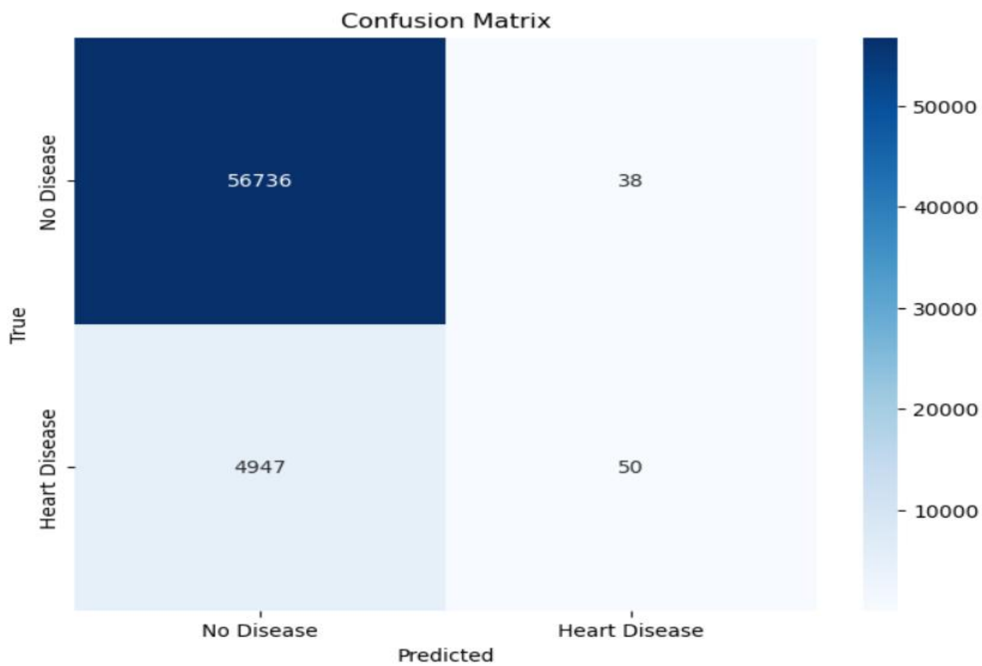

Fig. 3. Confusion matrix.

It is shown in fig. 3 that confusion matrix visually represents the performance of a system in predicting HD. The matrix shows that the system accurately identified 56,736 instances of HD (true positives) and misclassified 3,947 cases of HD as no disease (false negatives). It also identified 50 cases correctly as no disease (true negatives) but incorrectly flagged 38 as having HD when they did not (false positives). This matrix helps assess the balance between correct and incorrect predictions, indicating a higher rate of HD detection but with a considerable number of missed cases.

Fig. 4 shows that the ROC curve in the image evaluates the performance of a classification model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The curve shows a balanced trade-off between sensitivity (ability to detect positives) and the false alarm rate. The area under the ROC curve (AUC) is 0.83, indicating that the model performs well, but there is room for improvement. A perfect classifier would have an AUC of 1, meaning it distinguishes perfectly between positive and negative classes.

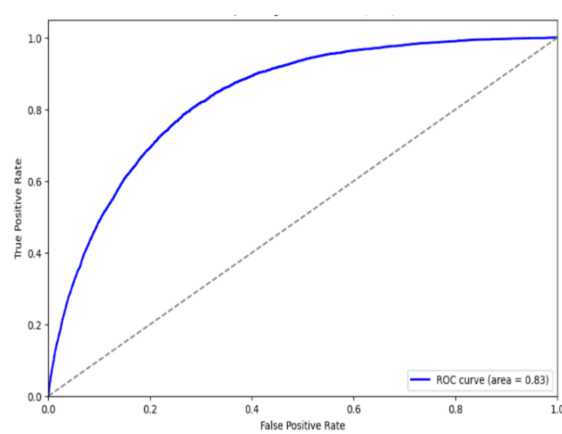

Fig. 4. Receiver operating characteristic (ROC) curve.

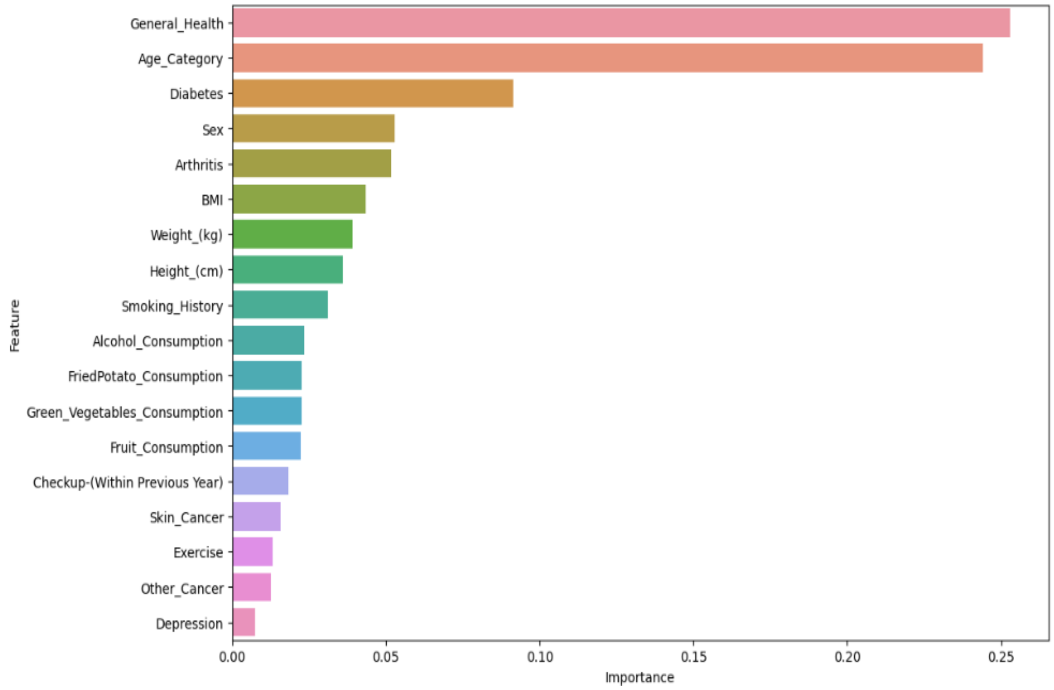

Fig. 5. Feature importance.

It is shown in Fig. 5 that the bar chart illustrates the feature importance ranking in a system, where each feature's contribution to the prediction outcome is displayed. General Health and Age Category are the most influential factors, contributing significantly to the system’s predictions, followed by Diabetes and Sex. Features like Depression, Other Cancer, and Exercise have the least importance. The chart helps to identify which variables have the greatest impact, allowing for better interpretation and possible improvements in system design or understanding of the data.

TABLE VI. CHATBOT SYSTEM VALIDATION PHASE FOR HD PREDICTION USING RF

RF	Acc (%)	TPR (%)	TNR (%)	FNR (%)	FPR(%)	LR+	LR-	PPV (%)	NPV (%)
Training	93.10	93.18	56.82	6.90	0.43	2.16	0.1210	99.90	0.018
Validation	92.00	91.97	56.81	8.00	0.43	2.13	0.0014	99.93	0.010

Table VI shows that the Proposed chatbot system performance in terms of accuracy, sensitivity, specificity, miss rate, and precision during training using RF provides 93.10, 93.18, 56.82, 6.90, and 99.90, respectively. The suggested approach yields 92.00, 91.97, 56.81, 8.00, and 99.93 during the validation phase's accuracy, sensitivity, specificity, miss rate, and precision. Furthermore, the proposed chatbot system for HD prediction using RF approach yields 0.43, 2.16, 0.1210, and 0.018 in terms of fall-out likelihood positive ratio, likelihood negative ratio, and negative predictive value during training and 0.43, 2.13, 0.0014, 0.010 in terms of validation.

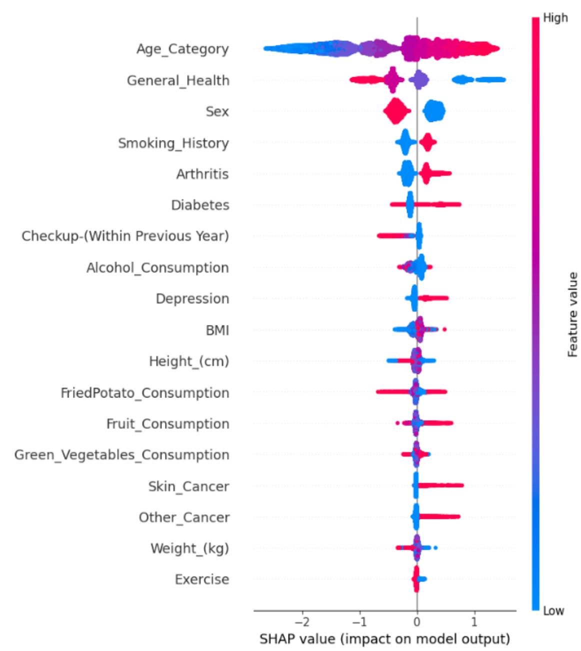

Fig. 6. SHAP value.

It is shown in fig. 6 that the SHAP (SHapley Additive exPlanations) plot highlights the impact of various features ona system's output. Each dot represents a single instance, with the position indicating the SHAP value, showing whether the feature drives the prediction towards a positive or negative outcome. Features like Age Category and General Health have the most influence, with higher feature values (shown in red) often leading to positive model outcomes. Conversely, low feature values (blue) push the predictions in the opposite direction. The plot helps explain how individual features contribute to the system's decisions.

Fig. 7 SHAP waterfall plot shows how individual features contribute to a single prediction. The base value E[f(x)]=−2.517 is the average prediction and features either increase (shown in red) or decrease (shown in blue) the predicted value. General Health has the most significant negative impact (-0.48), reducing the prediction, while Sex increases the prediction by +0.35. Other features such as Smoking History and Arthritis also contribute to the prediction, either positively or negatively. The final prediction value f(x)=−2.964 is the cumulative effect of all features.

Fig. 8 shows the prediction probabilities for a system evaluating HD risk. The system predicts a 94% probability for 'No Disease' and a 6% probability for HD. On the right, the feature contributions are split between the two outcomes: features like General Health (value = 4) and Age Category (value = 8) push the prediction towards HD, while features like Diabetes and Smoking History push towards no disease. The final prediction reflects the cumulative influence of all these factors on the system's output.

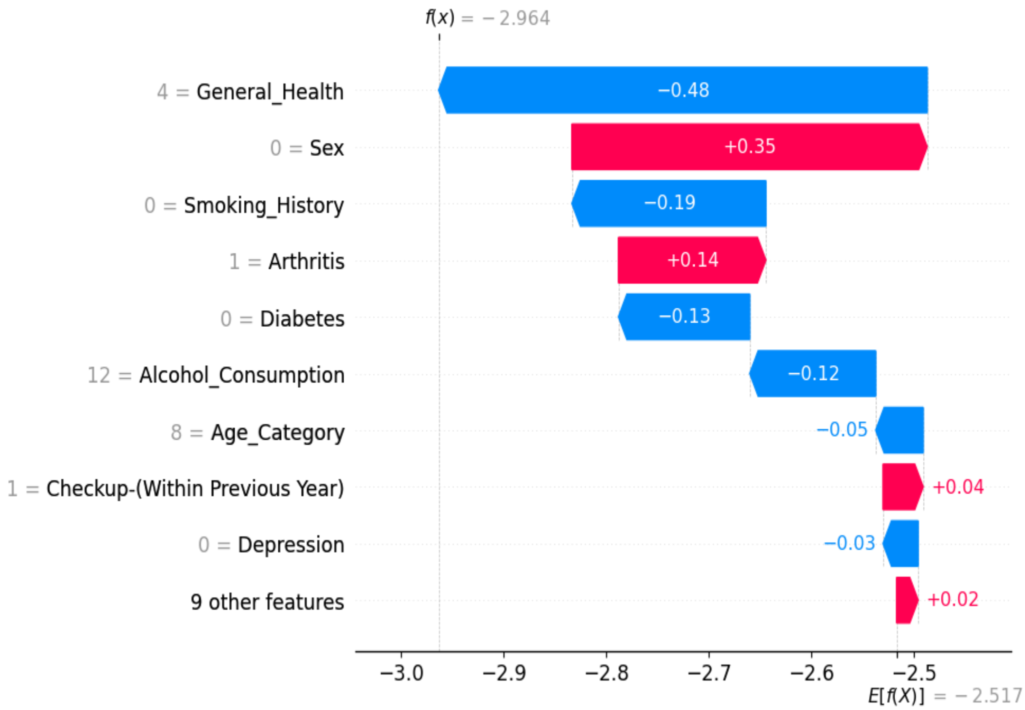

Fig. 7. SHAP waterfall impact.

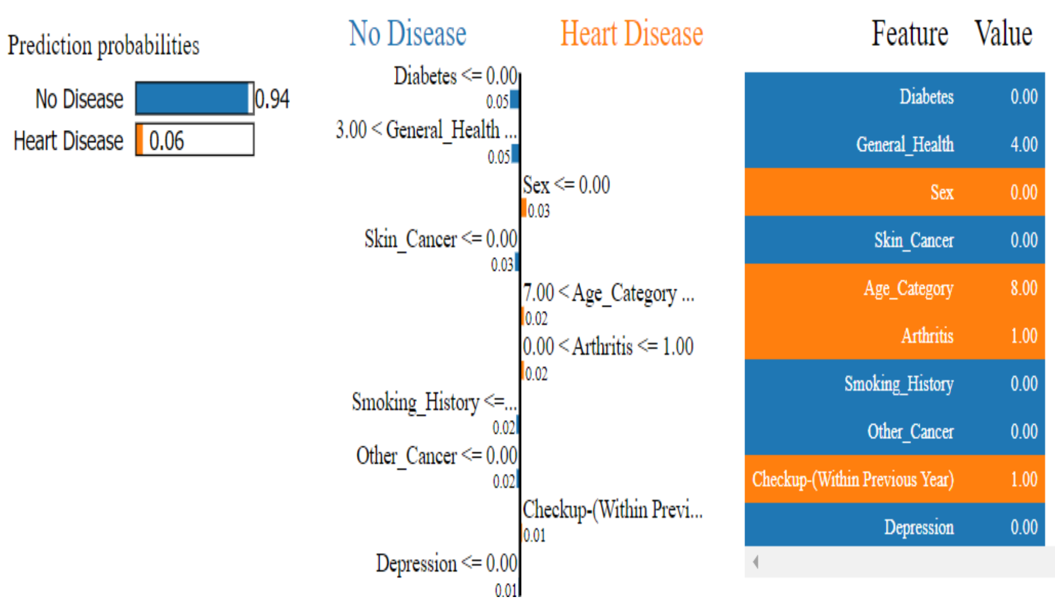

Fig. 8. LIME prediction as 'No Disease" found.

It is shown in Fig. 9 that LIME (Local Interpretable Model-agnostic Explanations) works by generating simplified, interpretable systems that approximate a complex system's decision-making process. In the figure 9, LIME highlights which features positively or negatively influence the outcome, shown by green (positive impact) and red (negative impact) bars. This visualization helps understand the contribution of each feature, such as diabetes, general health, or sex, to the system's prediction. LIME makes AI more interpretable by clarifying the factors driving specific decisions.

It is shown in Fig. 10 that the LIME prediction shows a 51% probability of HD and a 49% probability of no disease. Factors such as age, general health, and diabetes have strong contributions toward HD prediction, as highlighted in orange. Each feature listed, like the age category being greater than 9 and diabetes being present, increases the likelihood of HD, while features like "Other Cancer" lower it slightly. LIME explains how individual factors influence the system's decision, making the prediction more interpretable and understandable.

It is shown in Fig. 11 that in the LIME explanation, the features contributing to the prediction of HD are visualized with green (positive impact) and red (negative impact) bars. Significant positive contributors include age category, general health, and diabetes, indicating these factors increase the likelihood of HD. On the other hand, "Other Cancer" and alcohol consumption slightly reduce the probability, as shown in red. LIME effectively shows how each feature impacts the system's decision, highlighting HD is predicted.

Table VII highlights some previous chatbot systems with multiple approaches. The applied approaches also show the accuracies and miss rate.

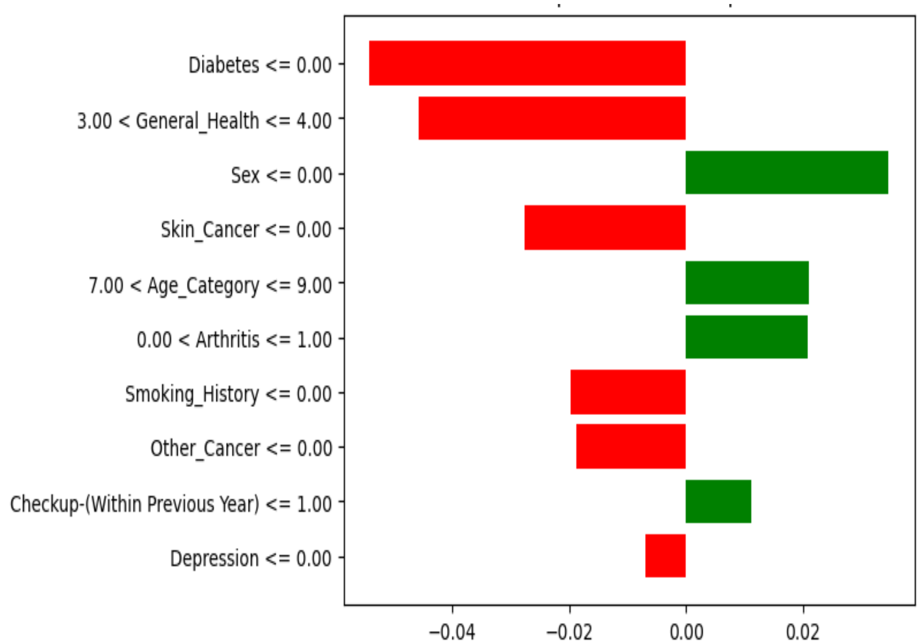

Fig. 9. LIME explanation as 'No Disease" found.

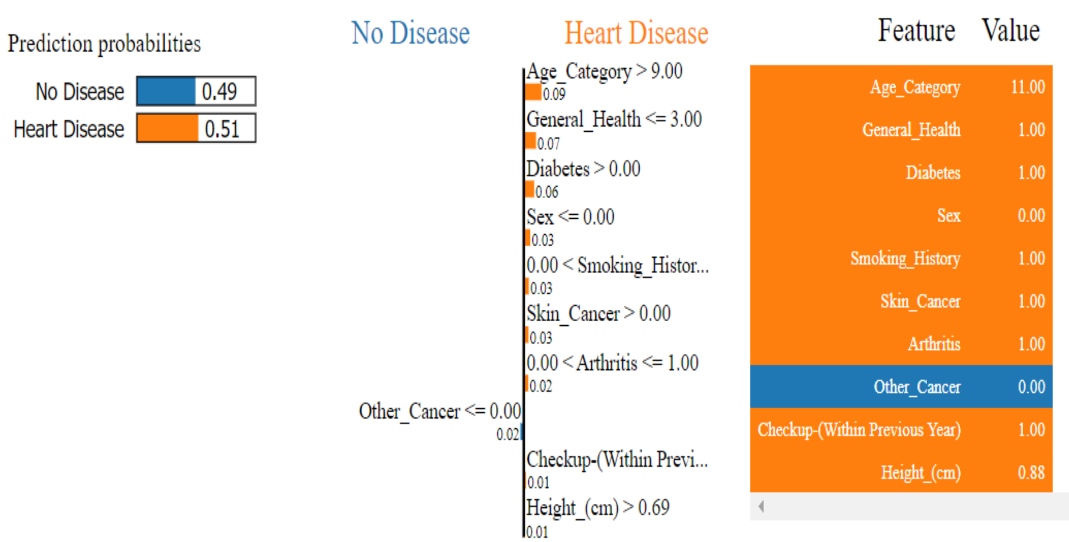

Fig. 10. LIME prediction as 'HD" found.

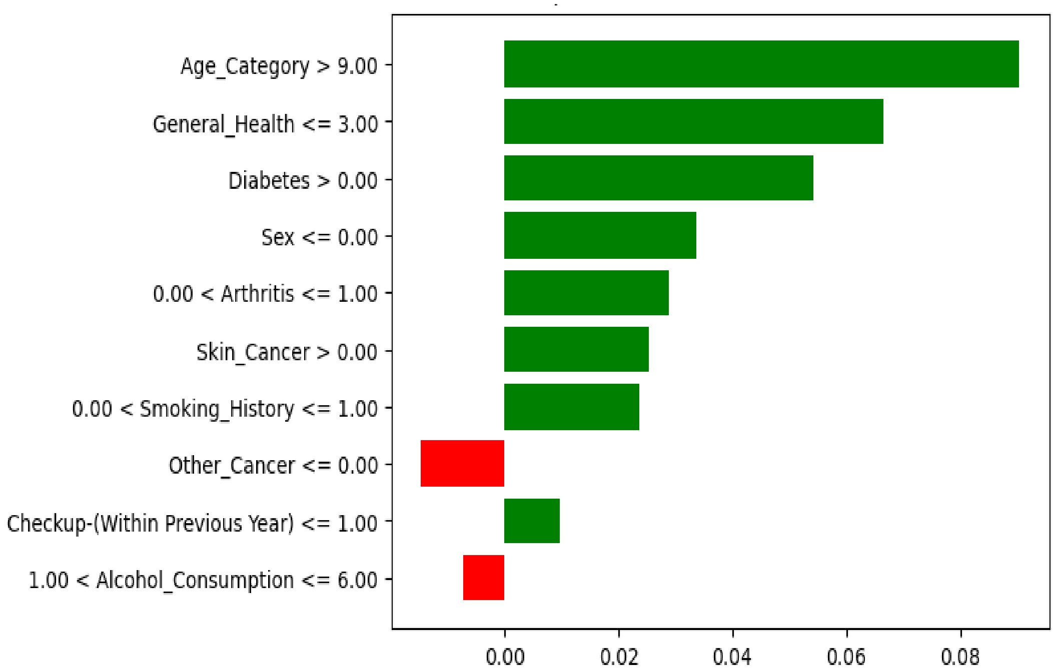

Fig. 11. LIME explanation as "HD" found.

TABLE VII. COMPARING THE PROPOSED CHATBOT SYSTEM PERFORMANCES WITH PREVIOUS APPROACHES

Ref	Year	Approach	Acc (%)	Miss-rate	Chatbot based Approach	XAI Implementation
[36]	2022	SVM	85.40	14.60	No	No
[37]	2021	Logistic Regression & KNN	87.5	12.5	No	No
[38[	2023	Multilayer perceptron with cross-validation	87.28	12.72	No	No
[39]	2021	KNN	84.86	15.14	No	No
[40]	2021	Extra Tree	87.0	13.0	No	No
[41]	2022	SVM	82.5	17.5	No	SHAP
[42]	2023	RF	74.0	26.0	No	SHAP
[43]	2024	DT	91.9	8.1	Yes	No
[44]	2023	Bagging-Quantum Support Vector Classifier (QSVC)	90.16	9.84	No	SHAP
		Proposed system (RF)	**92.0**	**8.0**	**Yes**	**SHAP & LIME**

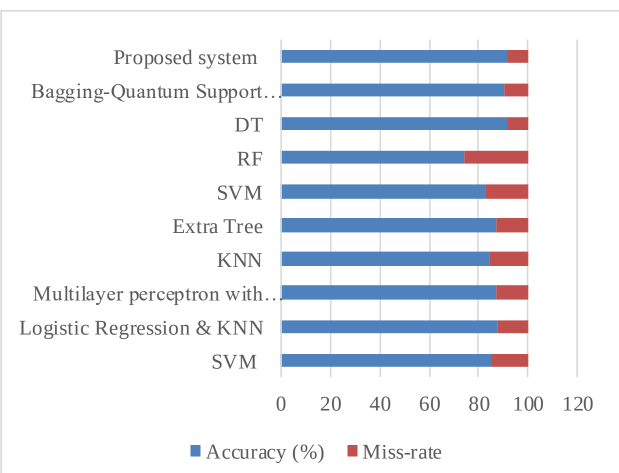

Fig. 12. Graphical representation of the previous approaches with the proposed system.

Table VII and fig. 12, compares the performance of the proposed chatbot system with previous ML approaches to predict HD. It is clearly shown that the proposed approach is better than the previous results in terms of accuracy and miss rate.

X. DISCUSSION

The proposed XAI-driven chatbot demonstrates strong potential in HD prediction by combining accurate ML models with user-friendly, interpretable outputs. This approach addresses the "black-box" limitations of traditional models, enhancing transparency and trust in healthcare applications. While the system shows promising results, its performance may vary across diverse populations, and further validation is needed to ensure its scalability and effectiveness in real-world settings. Future work will focus on addressing these limitations and integrating additional features to improve usability and adaptability.

XI. Conclusion

Heart disease is a significant global health challenge and a leading cause of mortality, with early detection being essential for preventing the disease's progression. Consequently, the objective of this study was to develop an explainable AI-driven chatbot system for predicting heart disease risk, effectively integrating advanced ML techniques with XAI methods. The novelty of the proposed system lies in its remarkable accuracy rate of 92% and an 8% miss rate, demonstrating its superiority over traditional approaches. By utilizing SHAP and LIME, the system boosts transparency in the decision-making process, permitting users to understand the rationale behind predictions. This innovative chatbot interface provides an accessible platform for interaction, ultimately fostering trust and promoting informed decision-making among healthcare providers and patients. The incorporation of XAI with ML in this context sets a new standard for responsible AI applications in healthcare, focusing on the requirement for continuous advancements in predictive modeling to address the ongoing challenges of HD.

XII. Limitations and Future Suggestions

Although XAI-driven chatbot demonstrates strong predictive performance, offering real-time, personalized HD risk assessments. Its ability to provide clear, interpretable results enhances trust and usability, particularly in healthcare settings where transparency is crucial. However, there are limitations: bias during the training of the system with potential biases in the training data, and reliance on specific input features that may not capture the overall picture of health indicators. Future research should continue to enhance this dataset using a larger, more diversified population, systems' robustness, and lifestyle and genetic factors, among others. Additionally, adding health monitoring data in real time will improve predictive accuracy. The long-term future path for perfecting the performance and experience of the chatbot will include continuous user feedback and iterative improvements on the developments themselves, thus ensuring that they continue to be valuable tools in preventive healthcare.

References

[1] Bagheri M, Bagheritaba M, Alizadeh S, Parizi MS, Matoufinia P, Luo Y. AI-Driven Decision-Making in Healthcare Information Systems: A Comprehensive Review.

[2] Chakraborty C, Bhattacharya M, Pal S, Lee SS. From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology. 2023 Nov 22:100164.

[3] Sahithya B, Prasad G, Sahithi B, Devarlla AC, Yashavanth TR. Empowering Healthcare with AI: Advancements in Medical Image Analysis, Electronic Health Records Analysis, and AI-Driven Chatbots. In2024 3rd International Conference for Innovation in Technology (INOCON) 2024 Mar 1 (pp. 1-7). IEEE.

[4] Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA network open. 2020 Jan 3;3(1):e1918962-.

[5] Muneer S, Rasool MA. AA systematic review: Explainable Artificial Intelligence (XAI) based disease prediction. International Journal of Advanced Sciences and Computing. 2022 Jun 30;1(1):1-6.

[6] Reis L, Maier C, Mattke J, Weitzel T. Chatbots in healthcare: Status quo, application scenarios for physicians and patients and future directions.

[7] Khan MA, Abbas S, Atta A, Ditta A, Alquhayz H, Khan MF, Naqvi RA. Intelligent Cloud Based Heart Disease Prediction System Empowered with Supervised Machine Learning. Computers, Materials & Continua. 2020 Oct 1;65(1).

[8] Lodi A, Olivier P, Pesant G, Sankaranarayanan S. Fairness over time in dynamic resource allocation with an application in healthcare. Mathematical Programming. 2024 Jan;203(1):285-318.

[9] He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M, Cambria E. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. arXiv preprint arXiv:2310.05694. 2023 Oct 9.

[10] Mensah GB. Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in AI Systems. Preprint, November. 2023;10.

[11] Das S, Sultana M, Bhattacharya S, Sengupta D, De D. XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI. The Journal of Supercomputing. 2023 Nov;79(16):18167-97.

[12] Ahmad G, Khan MA, Abbas S, Athar A, Khan BS, Aslam MS. Automated diagnosis of hepatitis b using multilayer mamdani fuzzy inference system. Journal of healthcare engineering. 2019;2019(1):6361318.

[13] Mienye ID, Jere N. Optimized ensemble learning approach with explainable AI for improved heart disease prediction. Information. 2024 Jul 8;15(7):394.

[14] Paudel P, Karna SK, Saud R, Regmi L, Thapa TB, Bhandari M. Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. InProceedings of the 10th International Conference on Networking, Systems and Security 2023 Dec 21 (pp. 69-78).

[15] Ogilvie L, Prescott J, Carson J. The use of chatbots as supportive agents for people seeking help with substance use disorder: A systematic review. Eur Addict Res. 2022; 28(6): 405-18. PMID: 36041418 DOI: 10.1159/000525959 [PubMed].

[16] Ahmed A, Hassan A, Aziz S, Abd-Alrazaq AA, Ali N, Alzubaidi M, et al. Chatbot features for anxiety and depression: A scoping review. Health Informatics J.2023; 29(1): 14604582221146719. PMID: 36693014 DOI: 10.1177/14604582221146719 [PubMed].

[17] Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. J Med Internet Res. 2023; 25: e40789. PMID: 36826990 DOI: 10.2196/40789 [PubMed].

[18] Sebastian, G. (2023). Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4454761.

[19] Healthily Chatbot. Available online: https://www.livehealthily.com/app (accessed on 20 March 2023).

[20] Ada Health Chatbot. Available online: https://ada.com/pt/ (accessed on 20 March 2023).

[21] Qi Y. Pilot Quasi-Experimental Research on the Effectiveness of the Woebot AI Chatbot for Reducing Mild Depression Symptoms among Athletes. International Journal of Human–Computer Interaction. 2024 Jan 16:1-8.

[22] Dechert, M. Implementation and Evaluation of a Chatbot to Crowdsource Geotagged Images to Detect Mosquito Breeding Sites. Master's Thesis, University of Bremen, Bremen, Germany, 2019.

[23] Symptomate Chatbot. Available online: https://symptomate.com/about/ (accessed on 20 March 2023).

[24] Swick, R.K. The Accuracy of Artificial Intelligence (AI) Chatbots in Telemedicine. J. South Carol. Acad. Sci. 2021, 19, 17.

[25] Wahyudi D, Ayuningsih E. The Application of Machine Learning in Predicting The Risk of Heart Disease With Decision Tree Algorithm. Instal: Jurnal Komputer. 2024 Jun 30;16(02):120-30.

[26] Dalakoti M, Wong S, Lee W, Lee J, Yang H, Loong S, Loh PH, Tyebally S, Djohan A, Ong J, Yip J. Incorporating AI into cardiovascular diseases prevention–insights from Singapore. The Lancet Regional Health–Western Pacific. 2024 Jul 1;48.

[27] Görtz M, Baumgärtner K, Schmid T, Muschko M, Woessner P, Gerlach A, Byczkowski M, Sültmann H, Duensing S, Hohenfellner M. An

artificial intelligence-based chatbot for prostate cancer education: design and patient evaluation study. Digital Health. 2023 May;9:20552076231173304.

[28] Zambrano Chaves JM, Wentland AL, Desai AD, Banerjee I, Kaur G, Correa R, Boutin RD, Maron DJ, Rodriguez F, Sandhu AT, Rubin D. Opportunistic assessment of ischemic heart disease risk using abdominopelvic computed tomography and medical record data: a multimodal explainable artificial intelligence approach. Scientific reports. 2023 Nov 29;13(1):21034.

[29] Abdulsalam G, Meshoul S, Shaiba H. Explainable heart disease prediction using ensemble-quantum machine learning approach. Intell. Autom. Soft Comput. 2023 Jan 1;36(1):761-79.

[30] Venkat V, Abdelhalim H, DeGroat W, Zeeshan S, Ahmed Z. Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine. Genomics. 2023 Mar 1;115(2):110584.

[31] Wilstrup C, Cave C. Combining symbolic regression with the Cox proportional hazards model improves prediction of heart failure deaths. BMC Medical Informatics and Decision Making. 2022 Jul 25;22(1):196.

[32] Narendran M, Sathya A, Annoosh R, Hari S. HealthBot Analytics: Optimizing Healthcare Efficiency Through Intelligent Integration. In2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) 2024 Apr 18 (pp. 1-7). IEEE.

[33] Rajkumar K, Ragupathi T, Karthikeyan S. Intelligent Chatbot for Hospital Recommendation System. In2024 2nd International Conference on Disruptive Technologies (ICDT) 2024 Mar 15 (pp. 664-668). IEEE.

[34] https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset?resource=download.

[35] Paudel P, Karna SK, Saud R, Regmi L, Thapa TB, Bhandari M. Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. InProceedings of the 10th International Conference on Networking, Systems and Security 2023 Dec 21 (pp. 69-78).

[36] Ahamed J. et al., 2022. CDPS-IoT: cardiovascular disease prediction system based on IoT using machine learning.

[37] Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. InIOP conference series: materials science and engineering 2021 (Vol. 1022, No. 1, p. 012072). IOP Publishing.

[38] Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms. 2023 Feb 6;16(2):88.

[39] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. 2021. Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience 2021 (2021).

[40] Sakinat Oluwabukonla Folorunso, Joseph Bamidele Awotunde, Em-manuel Abidemi Adeniyi, Kazeem Moses Abiodun, and Femi Emmanuel Ayo. 2021. Heart disease classification using machine learning models. In International Conference on Informatics and Intelligent Applications. Springer, 35–49.

[41] Pratiyush Guleria, Parvathaneni Naga Srinivasu, Shakeel Ahmed, Naif Almusal- lam, and Fawaz Khaled Alarfaj. 2022. XAI framework for cardiovascular disease prediction using classification techniques. Electronics 11, 24 (2022), 4086.

[42] Moreno-Sanchez PA. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Frontiers in Cardiovascular Medicine. 2023 Aug 1;10:1219586.

[43] Wahyudi D, Ayuningsih E. The Application of Machine Learning in Predicting The Risk of Heart Disease With Decision Tree Algorithm. Instal: Jurnal Komputer. 2024 Jun 30;16(02):120-30.

[44] Abdulsalam G, Meshoul S, Shaiba H. Explainable heart disease prediction using ensemble-quantum machine learning approach. Intell. Autom. Soft Comput. 2023 Jan 1;36(1):761-79.