SvedbergOpen

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Explainable Machine Learning Techniques For Predictive Analysis Of Chronic Diseases Using Electronic Health Records

Yogesh Dinkar Jadhav[1], Pushpa Nagini Sripada[2], Dr.Jagdish Gohil[3], Snehal Swapnil Jawahire[4], Vimal Bibhu[5], Dr.Ravikant kushwaha[6], Dr. Samir Sahu[7], Ambika P[8]

[1]Department of Mechanical Engineering, Sinhgad College of Engineering, SavitribaiPhule Pune University Pune, Maharashtra, India. Email: sittpo3@gmail.com Orcid : 0009-0004-5268-813

[2]English, Professor, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: sripadapn@maher.ac.in

[3]Dean, ,Parul Institute of Medical Sciences and Research, Parul University, Vadodara, Gujarat, India, Email: Jagdish.gohil@paruluniversity.ac.in, 0009-0006-2927-9107

[4]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India. Email: snehal.jawahire@vit.edu

[5]School of Engineering & Technology, Noida international University, Uttar Pradesh, India. Email: vimal.bhibu@niu.edu.in

[6]Associate Professor , MSOPS, Maharishi University of Information Technology, Lucknow, Uttar Pradesh, India, Email: ravikant.kushwaha@gmail.com, Orcid Id- https://orcid.org/0009-0007-3351-703X

[7]Professor, Department of General Medicine, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email: samirsahu@soa.ac.in, Orcid Id- 0000-0002-5610-0380

[8]Department of Commerce, Assistant Professor, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: ambikap@maher.ac.in

Abstract

Diabetes, cardiovascular diseases and chronic kidney diseases are among the chronic diseases responsible for a significant burden of disease and increasing mortality, complications and healthcare costs at the global level. Electronic Health Records (EHRs) have been growing very rapidly and have been used to build a data-driven predictive healthcare system to assist the early diagnosis and personal treatment plans. Many machine learning (ML)-based predictive models, however, are black box models, obscure – and not trusted – by clinicians when it comes to automated decision making. In this study, an explainable machine learning model is proposed to tackle the problem of predictive analysis of chronic disease from EHR data. The performance of various machine learning techniques such as Logistic Regression, Random Forest, Support Vector Machine, XGBoost and Artificial Neural Networks is compared in predicting the disease. In order to enhance the interpretability, methods of Explainable Artificial Intelligence (XAI) are embedded to explain the model predictions in a patient-specific and global manner, such as SHAP and LIME. Through experimental results, we show that the proposed explainable models are both interpretable and have high predictive accuracy and strong ROC-AUC performance. The results demonstrate the clinical promise of explainable ML systems for providing healthcare decision-support applications that are transparent, reliable, and intelligent.

Keywords: Explainable Artificial Intelligence, Electronic Health Records, Chronic Disease Prediction, Machine Learning, Healthcare Analytics, SHAP and LIME

1. Introduction

Chronic diseases like diabetes, cardiovascular, chronic kidney disease and hypertension have surged as a significant global health challenge due to their substantial mortality rates, long-term complications and economic cost on the healthcare infrastructure (Beam &Kohane, 2018; Obermeyer& Emanuel, 2016; Rajkomar et al., 2019). For better care to be provided to patients and costs to be kept down it is very important to predict how things will turn out and intervene in time. The rapid advancements of Electronic Health Record (EHR) systems have revolutionized health care systems in recent years by providing access to vast amounts of patient information, such as demographic details, lab outcomes, medication records, clinical notes, and diagnoses (Goldstein et al., 2016; Johnson et al., 2016). The emergence of such structured health data has opened up a wide range of possibilities for the creation of artificial intelligence (AI) and machine learning (ML)-based intelligent predictive analytics systems (Ching et al., 2018; Esteva et al., 2019; Miotto et al., 2018). Through their work, Rajkomar et al (2019) and Topol (2019) have shown that AI-powered healthcare analytics can showcase strong potential in disease diagnosis, risk stratification, treatment recommendations and personalized medicine.

Even though machine learning models have been very successful in prediction tasks in the field of health services, many complex predictive systems are "black-box", and the decision-making process is not easy for clinicians to interpret and trust (Ahmad et al., 2018; Holzinger et al., 2019). Transparency/interpretability is one of the significant challenges for medical AI applications since AI-based automated decision support systems are not readily adopted by healthcare practitioners in real clinical settings and need to be explained and interpreted by the healthcare practitioner (Katuwal& Chen, 2016; Ribeiro et al., 2016). Moreover, missing data, complex clinical features, imbalance in class distribution, and data complexity are other significant challenges in EHR-based predictive modeling that could impact prediction accuracy and generalizability (Goldstein et al., 2016; Xiao et al., 2018).

To satisfy the increasing need for transparent and clinically interpretable AI systems, the study emphasizes the creation of an XAI model for predictive chronic disease analysis from EHRs.The study aims to develop an explainable machine learning framework for predictive chronic disease analysis from EHRs, motivate by the growing need for transparency and clinical interpretability in AI systems. This research compares several machine learning algorithms, such as Logistic Regression, Random Forest, Support Vector Machine, XGBoost and Artificial Neural Networks, for predictive efficacy (Breiman, 2001; Singh et al., 2016). Beyond the prediction, an AI system built into the predictive framework generates explanations, such as the method Explainable Artificial Intelligence (XAI), helps improve interpretability and clinicians' confidence, and supports the creation of two types of explanations: global explanations and explanations for individual patients (Lundberg & Lee, 2017; Ribeiro et al., 2016). The key strengths of this research are the comparative evaluation of explainable ML models, the incorporation of proper techniques to aid in the explainability of advanced ML models, the evaluation of the models based on EHR datasets, and the statistical validation of model prediction performance, enabling reliable applications of healthcare decision support.

2. Literature Review

Machine learning techniques have been increasingly adopted in the healthcare analytics sector for their capability to uncover patterns and for facilitating early diagnosis of diseases based on vast amounts of clinical data (Beam &Kohane, 2018; Miotto et al., 2018).Previous research has investigated the use of machine learning models for chronic diseases prediction such as diabetes, cardiovascular diseases, and chronic kidney diseases (Panahiazar et al., 2015; Rajkomar et al., 2019). Various clinical variables like glucose, insulin, body mass index and age are used in diabetes prediction models to determine high-risk individuals. In a similar fashion, machine learning algorithms are used in cardiovascular disease prediction systems to determine the risk of cardiovascular disease based on blood pressure, cholesterol, electrocardiogram readings and lifestyle factors. Machine learning strategies that use lab values and patient medical records to predict chronic kidney disease have also advanced the early detection of the disease. Classics like Logistic Regression and Decision Trees have been found to be interpretable while more complex models like Random Forest and XGBoost have been seen to be more predictive in complex healthcare datasets (Breiman, 2001).

The growing use of Electronic Health Records (EHRs) has also contributed to the evolution of predictive healthcare systems (Goldstein et al., 2016; Johnson et al., 2016). EHRs store detailed patient data over time and across different types of data, such as, demographics, diagnosis reports, medications, lab test results, and treatment history. Temporal EHR data can be used to measure disease progression continuously and facilitate personalized healthcare analytics (Shickel et al., 2018). EHR-based predictive modeling, on the other hand, has a number of challenges, however, such as missing values, multiple clinical attributes, noisy data, data imbalance, and inconsistencies in medical records. If not resolved well by preprocessing and feature engineering, these issues will affect the reliability and capability of generalizing for machine learning models (Xiao et al., 2018).

With the advent of AI-driven predictions, Explainable AI Intelligence (XAI) is an important research field in healthcare, critical to the success of which is the need for transparent and interpretable decision support systems before machine predictions will be rolled out in clinical practice (Ahmad et al., 2018; Holzinger et al., 2019). With the emergence of numerous sophisticated machine learning and deep learning models, the need for interpretability has grown (Doshi-Velez & Kim, 2017). Explainability approaches can be broadly classified as either local or global explainabilityapproaches.There are two categories of explainability approaches: Local and global explainability approaches. A crucial aspect of clinical trustworthiness is when AI systems can explain their reasoning behind predictions, particularly in health care where diagnoses and treatment plans might rely on AI tools.A very important part of the clinical trustworthiness is the system's capacity to explain its predictions, for example in a critical health-care application like disease diagnosis or treatment planning (Vellido, 2020).

A number of explainable machine learning (XAI) methods have been investigated in the health care analytics field, such as Logistic Regression, Decision Trees, Random Forest, XGBoost, and SHAP and LIME. Logistic Regression and Decision Trees are inherently interpretable, while ensemble models like Random Forest and XGBoost are more accurate at making predictions but need extra explainability tools (Breiman, 2001). SHAP and LIME are two popular XAI methods that offer explanations at the feature level and patient-specific explanations for complex models (Lundberg & Lee, 2017; Ribeiro et al., 2016). Explainable machine learning also proved its clinical practice in healthcare prediction systems as was shown by Lundberg et al. (2018).

While considerable progress has been made there are still certain unanswered research questions. The majority of existing studies typically address explainability as a secondary concern, which can lower the trust placed in AI assistance for diagnosis by clinicians (Holzinger et al., 2019). Moreover, most studies do not properly compare various explainable machine learning models and do not use a clinical lens to interpretability. As a result, there is a high demand for the development of transparent, accurate, and clinically interpretable predictive systems based on EHR data for chronic diseases analysis.

3. Materials and methods

3.1 Overall Predictive Analysis Workflow

The flow of the overall predictive analysis process proposed in this study is shown in Figure 1. The first step in the workflow is that of acquiring clinical information related to patients from their electronic health records (EHRs), such as their demographics, laboratory information, vital signs, medication history, and diagnostic reports. Collected data would then go through the preprocessing of data phase where missing data, noisecy data, duplicate data and categorical data would be preprocessed to improve the quality of the data. Preprocessed data is then fed to feature extraction and feature selection to choose the clinically relevant variables that are most important for predicting a chronic disease.

The end result is the feature set, which is then stored and used for the machine learning model training, during which models like Logistic Regression, Random Forest, Support Vector Machine and XGBoost are trained and assessed. The prediction output (disease class and risk probability) is produced by the trained models. To enhance transparency, prediction results are also passed to an explainability engine where they are processed with SHAP and LIME. These techniques clarify the importance of features and individual patient predictions. Lastly, the clinical interpretation of the explained prediction is presented, aiding clinicians in risk assessment, early diagnosis and decision-making.

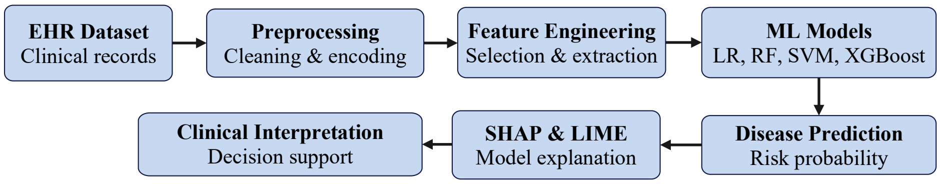

Fig. 1. Explainable Predictive Analysis Workflow Using EHR Data.

3.2 Dataset Description

This study employs widely-used benchmark healthcare datasets in the field of chronic disease prediction to assess the effectiveness of explainable machine learning models. The main sources for the primary data sets are the MIMIC-III data set, the UCI Chronic Disease data set and the Heart Disease data set. The MIMIC-III database includes detailed intensive care unit (ICU) patient data, such as demographics, laboratory values, medications, and clinical features. The UCI Chronic Disease Dataset includes diabetes and kidney disease-related clinical parameters in a structured format and the Heart Disease dataset includes patient attributes commonly used in cardiovascular disease predictive analytics.

The datasets include both quantitative and qualitative healthcare factors extracted from patient's Electronic Health Records (EHR). Some clinically relevant parameters like age, body mass index, glucose level, blood pressure, and level of cholesterol were chosen for medical relevance and predictive importance. The features have been used in the experimental workflow in preprocessing, feature engineering, the model training, and the explainability analysis. Before the machine learning models were trained, the datasets were preprocessed using techniques such as missing value imputation, normalization of numerical features, and class balancing.

Table 1. Clinical Attributes Extracted from Electronic Health Records

Attribute	Description	Type	Range/Values	Mean Value	Clinical Importance
Age	Patient age	Numeric	18–90 years	52.4	Risk indicator
BMI	Body Mass Index	Numeric	16.5–42.8 kg/m^2	28.7	Obesity risk
Glucose	Blood glucose level	Numeric	70–220 mg/dL	128.5	Diabetes prediction
Blood Pressure	Systolic BP readings	Numeric	80–190 mmHg	132.6	Cardiovascular risk
Cholesterol	Lipid profile level	Numeric	120–340 mg/dL	210.3	Heart disease prediction
Heart Rate	Beats per minute	Numeric	55–145 bpm	82.1	Cardiac monitoring
Creatinine	Kidney function indicator	Numeric	0.4–6.5 mg/dL	1.8	Kidney disease assessment
Smoking History	Smoking status	Categorical	Yes/No	—	Chronic disease indicator
Medication History	Previous treatments	Categorical	Multiple classes	—	Disease progression analysis

These attributes play a significant role in prediction of chronic diseases and are also used to conduct feature importance analysis by employing SHAP and LIME explainability methods in the following experimental sections.

3.3 Data Preprocessing

The quality, reliability, and the usefulness of machine learning models generated from Electronic Health Records (EHRs) are significantly affected by data preprocessing. Due to incomplete entries and inconsistent formats of healthcare datasets as well as noisy measurements and imbalanced disease classes, some of the data preprocessing were performed prior to model training. At first, data missingness in the clinical numerical attributes were addressed by mean imputation and median imputation methods, and then the missing values in the clinical categorical attributes were imputed by mode imputation method. This gave assurances that the predictive performance would not be adversely impacted by incomplete patient records.

After the treatment of missing values, abnormal clinical measurements that would affect the learning process were eliminated using outlier detection and removal. Extreme values were detected from the attributes glucose level, cholesterol concentration, blood pressure and BMI using statistical methods like Interquartile Range

(IQR) method and Z score analysis. Data were made more consistent by the removal of noisy outliers and prediction bias was reduced.

Data normalization (Min-Max normalization, standard scaling) was applied to ensure a uniform nature of features of different scales. To avoid features with higher values dominating the model training process, numerical clinical parameters were converted to a similar range. Additionally, categorical healthcare items like smoking history and medication were encoded numerically by label encoding or one-hot encoding.

There were also disparities between the number of patient records from diseased and non-diseased patients, resulting in class imbalance among the datasets. To overcome this problem, a technique called Synthetic Minority Oversampling Technique (SMOTE) was used to create synthetic samples for minority disease classes. SMOTE balancing does well in improving the generalization ability of the classifiers, and in decreasing the amount of bias that the classifiers displayed toward majority classes. Finally, in the subsequent phases of the proposed predictive framework, the fully-processed dataset was used for feature engineering and machine learning model training and explainability analysis.

3.4 Feature Engineering and Selection

The processes of feature engineering and feature selection are crucial in healthcare predictive analytics due to their roles in model efficiency, computational simplicity, and prediction accuracy. A set of feature engineering techniques were used in this study to extract clinically relevant features from Electronic Health Records (EHRs) to be used for predicting chronic diseases. Then, correlation analysis was used to explore the relationships between health and health-related features and to determine the most strongly associated features associated with the disease. The relationship of clinical parameters including glucose, blood pressure, cholesterol, body mass index (BMI), creatinine with chronic disease indicators was strong positive correlation.

To optimise the prediction feature space further, Recursive Feature Elimination (RFE) was used to remove the less informative variables and keep the most important variables to train the machine learning models. In addition, an information gain analysis is carried out to assess the contribution of every feature to the disease classification performance. The techniques used it to remove redundant and unimportant features and enhance the explainability and generalization ability of the model.

The overall feature correlation and selection is shown in Figure 2. The correlation heatmap (Figure 2(A)) reveals that glucose, blood pressure, cholesterol and creatinine are relatively strongly correlated to the chronic disease risk factors making these parameters of particular interest for predictive healthcare modeling. The feature importance ranking given in Fig. 2(B) also indicates that the feature glucose level has the highest importance score followed by blood pressure, cholesterol level, BMI and age. These results validate the primary role of metabolic and cardiovascular related clinical characteristics in the prediction of chronic diseases.

The explainability objectives of the proposed framework are also backed up by the feature contribution analysis shown in Figure 2. The features that were selected were then used for machine learning model training and explainability analysis using SHAP/LIME in the following sections of the study. Incorporation of statistically relevant and clinically relevant features greatly enhanced the predictive accuracy and interpretability of the proposed explainable machine learning models.

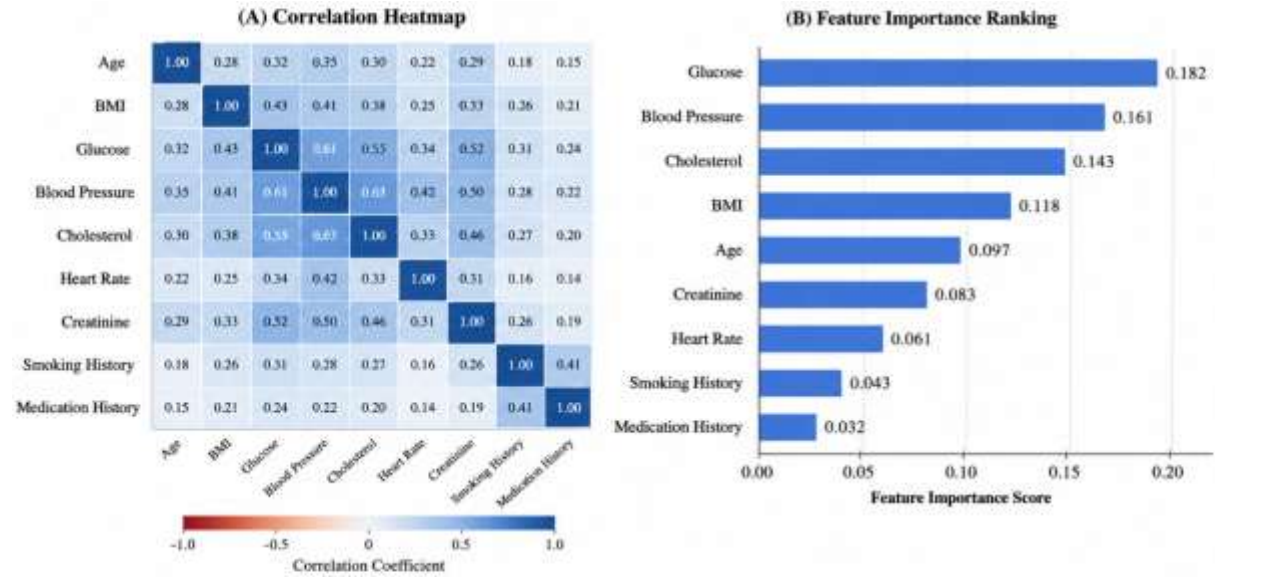

Fig. 2. Feature Correlation and Selection Analysis.

3.5 Machine Learning Techniques

In this work, three supervised machine learning methods were used to conduct predictive analysis of chronic diseases, based on Electronic Health Records (EHRs). The models selected were based on their successful application in healthcare analytics, classification ability, interpretability, and robustness with structured clinical data. Among the models implemented are Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost and Artificial Neural Networks (ANNs).

Logistic Regression was selected as a baseline statistical classifier because of its simplicity, interpretability, and success in the binary classification of healthcare predictions. It makes estimates of disease occurrence according to parameters given by clinical questionnaires and it gives clear decision boundaries. Random Forest was used due to the ensemble learning functionality to create multiple decision trees that mutually complement each other to achieve higher accuracy in prediction, while avoiding overfitting. High dimensional classification and effective separation of diseased and non-diseased patient classes have been achieved using Support Vector Machine. The choice of XGBoost was made because of its gradient boosting approach, speed and performance on healthcare information. Moreover, Artificial Neural Networks were applied to obtain complex non-linear relationships existing between clinical variables and to increase the aspect of predicting diseases. In this research, the prediction model used is Logistic Regression that is shown in Equation (1).

Equation (1): Logistic Regression Prediction Model

$$P\big((y=1)\big|x\big) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}, \qquad (1)$$

In Equation (1), $P\big((y=1)\big|x\big)$ represents the probability of disease occurrence for a patient based on input clinical features $x_1, x_2, \ldots . x_n$. The parameters $\beta_0, \beta_1, \ldots, \beta_n$ denote the regression coefficients learned during model training. The logistic sigmoid function converts the linear combination of the input variables into a value between 0 and 1, which is used for binary classification of diseases.

Logistic Regression model can be used for the estimation of disease probability, binary prediction modeling and clinical risk assessment. Patients were classified into diseased or non-diseased groups using a predefined clinical decision threshold, based on predicted probabilities. Later in the study, the results of these machine learning models were assessed with various performance measures and finally analyzed with explainability techniques such as SHAP and LIME.

3.6 Explainable Machine Learning Techniques

In this study, Explainable Artificial Intelligence (XAI) techniques are incorporated within the machine learning architecture to enhance transparency and interpretability in chronic disease forecasts.To boost transparency and interpretability of chronic disease forecasting, in this study, XAI techniques are included in the machine learning architecture. The explainability methods are necessary to understand how the clinical variables affect the results of the predictive models, particularly when they are considered as a black-box system, and to make clinicians more confident in using the AI-supported healthcare system. This work used SHapley Additive exPlanations (SHAP) and LIME (Local Interpretable Model-Agnostic Explanations) to generate global and local interpretation of predictions from a machine learning model made from Electronic Health Records (EHRs).

Global feature importance and the individual clinical parameters contribution to the prediction of chronic disease, were tested using SHAP analysis. SHAP summary plot: in Figure 3(A) it is clearly seen how the clinical parameters like glucose, blood pressure, cholesterol, BMI, age and creatinine affect the model prediction. The positive SHAP values are associated with increased probability of the disease while the negative SHAP values are associated with the decreased probability of the disease. Glucose level is the most important predictive feature, followed by blood pressure and cholesterol, from the distribution of their SHAP values. Moreover, the global feature importance ranking shown in Figure 3(B) reiterates the importance of glucose as the most significant feature for predicting chronic disease globally.

Apart from the fact that it is globally interpretable, LIME analysis was also used to provide local explanations for patient predictions. In the local explanation visualization in Fig. 3(C), patient-specific feature contributions to the risk prediction of the disease are shown. Risk contribution was significantly increased by positive contributors like elevated glucose level, high blood pressure, high cholesterol levels and high BMI, while normal heart rate and controlled creatinine level contributed to reduce disease occurrence. Figure 3(D) presents the patient-level prediction interpretation which gives a clinically understandable explanation of the prediction outcome, such as the disease probability score, major contributing factors, and overall risk evaluation.

SHAP is a very useful tool to integrate with LIME to improve the transparency of the model and aid support for interpretable healthcare decision-making. These explainability tools allow clinicians to understand more deeply how the machine learning predictions are making their decisions and how to validate critical clinical risk factors, as well as trustworthiness in AI-based chronic disease prediction systems.

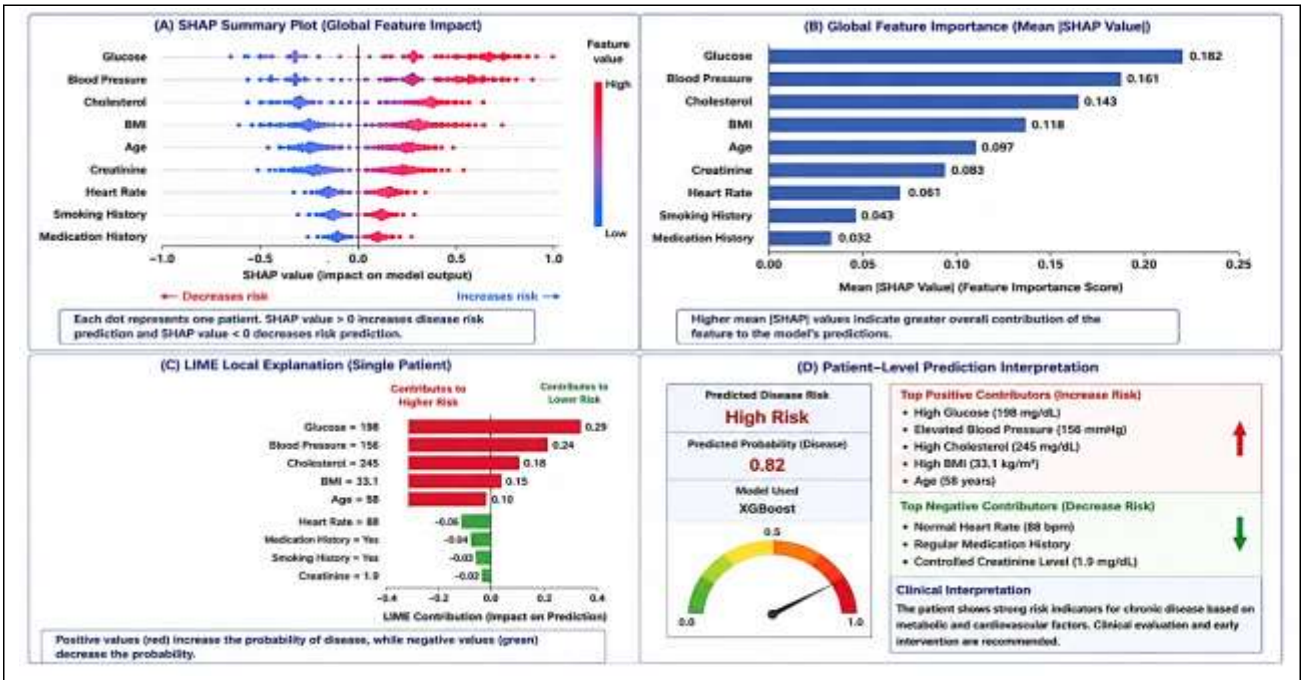

Fig. 3. SHAP and LIME-Based Explainability Analysis.

3.7 Performance Evaluation Metrics

Various performance assessment metrics have been used for testing the performance of the proposed explainable machine learning models for predicting chronic diseases. Healthcare datasets may have class imbalance issues and decision making may be related to risk, so using only an evaluation metric is not enough to measure the reliability of predictive models. So, the accuracy and the performance of the classification was evaluated by using the confusion matrix parameters of Accuracy, Precision, Recall, F1-Score, Specificity, receiver operating characteristic area under the curve (ROC-AUC) and Precision-Recall area under the curve (PR-AUC).

The accuracy is the overall correctness of the classification model; it is the percentage of patient records correctly predicted from all patient records. Precision measures the percentage of patients that are correctly identified as having disease out of all the patients that are predicted to have the disease and is used to try to minimise false-positive diagnosis. Recall, or sensitivity, is the percentage of patients with a disease who were correctly identified by the model and is especially important when applied to health care systems where the absence of a diagnosis of the disease can have a major impact on the patient's clinical situation. Specificity is the capacity of the model to accurately detect healthy people and reduce false alarms.

The F1 score has been given great importance in terms of evaluation metrics as it is a balanced score which takes both the precision and recall into consideration. The F1-score formula applied in this study is shown in Equation (2).

Equation (2): F1-Score Formula

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \qquad (2)$$

Precision is the fraction of correctly predicted positive disease cases, and Recall is the fraction of the true disease cases that are correctly identified by the model in equation (2). The harmonic mean chosen for calculating the F1-score offers a fair assessment of the performance of classification algorithms, especially when working with imbalanced healthcare datasets.

Furthermore, ROC-AUC analyses was conducted to assess the overall discrimination ability of the machine learning models at various classification threshold points. The predictive effectiveness performance was also evaluated in the case of imbalanced class distribution using PR-AUC. The aforementioned evaluation metrics can be used to comprehensively evaluate the potential accuracy, reliability, sensitivity, and interpretability of the proposed explainable chronic disease prediction framework.

3.8 Statistical Validation

The proposed explainable machine learning models for predicting the chronic disease have been statistically validated for reliability, robustness, and generalization ability. To ensure the model is robust and doesn't overfit, it is important to rigorously evaluate the statistics that will validate the model performance. The accuracy of predictive healthcare systems is highly dependent on data variability and class imbalance, so it's crucial to validate the model performance meticulously to minimize the risk of overfitting. This study used 10-fold cross validation, statistical significance testing and confidence interval analysis to thoroughly evaluate the stability and effectiveness of the developed models.

Firstly, 10 fold cross validation was used to check model consistency in the different folds of the data. The entire data set was randomly split into 10 equal partitions, 9 of which were used during the model training and the remaining one was used during the model testing. This was done ten times to allow each subset to be the test set once. Average values of Accuracy, Precision, Recall, F1-score and ROC-AUC across all folds were calculated to get a robust estimation of the predictive performance. Cross validation mitigated sampling bias and enhances the reliability of performance assessment.

Paired comparison methods were used to check the effectiveness of the proposed machine learning techniques further. It was statistically tested whether the observed improvement of performance of the classifiers was significant or just random variation using Logistic Regression, Random Forest, Support Vector Machine, XGBoost and Artificial Neural Networks. For prediction tasks involving chronic conditions, statistical significance analysis was used to validate high-performing models like XGBoost and Random Forest.

The analysis of uncertainty in model evaluation metrics was also carried out by confidence interval analysis. The 95% confidence intervals of the Accuracy and ROC-AUC were computed to evaluate the stability and reliability of the predictive models over different data situations. Narrow confidence intervals signified good model consistency and improved generalisation ability. Overall, the incorporation of cross validation, significance testing, and confidence interval analysis enhanced the credibility of the proposed explainable healthcare prediction framework and its statistical validity.

4. Results and Discussion

4.1 Experimental Setup

To assess the performance of explainable machine learning method on predicting chronic diseases based on Electronic Health Records (EHRs), the experimental analysis was done in Python machine learning environment. Several commonly used data science and explainable AI libraries were used for the implementation and model development. Pythons 3.10 were used as the main programming environment because they were extensively supported for healthcare analytics and machine learning applications. Traditional machine learning techniques like Logistic Regression, Random Forest, and Support Vector Machine were applied using the tool called the Scikit-learn library. An Artificial Neural Network (ANN) based predictive models were developed and trained using TensorFlow. Besides this, the libraries SHAP and LIME have been included to carry out explainability analysis and to produce model explanations for both the global and local level.

The experimental methodology consisted of data preprocessing, feature engineering, model training, hyperparameter optimization, explainability analysis and performance evaluation. To guarantee dependable performance validation, the data set was split into 80% training data and 20% testing data. Moreover, the 10-fold cross validation was done to prevent overfitting and enhance the model's generalization ability. To enhance the consistency of the data and address the problem of class imbalance, two pre-processing techniques, Min-Max normalization and SMOTE balancing, were used.

The experiments were conducted on Intel Core i7 12th Generation processor running at 3.60 GHz on 32 GB RAM with NVIDIA RTX 3060 GPU having 12 GB dedicated memory. For large-scale data processing in healthcare, GPU acceleration proved to be a valuable feature to enhance the efficiency of training the ANN and XGBoost models. Windows 11 Pro (64-bit) was used as the operating system for the experiments.

The clinical features (glucose level, blood pressure, cholesterol, BMI, age and creatinine) were always used during pre-processing, feature selection, analysis of explainability and in the evaluation of the prediction. All experiments were conducted with the same setup to enable comparisons among the different machine learning algorithms. The adopted system allowed for reliable evaluation of the predictive performance, effectiveness of explainability, and computational efficiency of chronic disease prediction with EHR data, ensuring the trustworthiness of results presented in the rest of the paper.

4.2 Comparative Performance Analysis

Performance comparison of the implemented machine learning models was carried out utilizing a number of assessment metrics such as the Accuracy, Precision, Recall, F1 score, Specificity and ROC-AUC. A wide range of metrics was chosen to evaluate the reliability of classification, capability for detecting the disease, and robustness of predictions in the context of chronic disease prediction from Electronic Health Records (EHRs). The results of the experiments show that the models based on ensemble and boosting had better performance than the traditional statistical-based models because they were able to learn from the complex non-linear relationships of clinical attributes.

As shown in Table 2, the performance of the different models was compared and XGBoost was the best model with the accuracy of 95.3%, precision of 94.8%, recall of 95.1%, F1 score of 94.9%, specificity of 94.2%, and ROC AUC score of 0.97. The great performance of XGBoost is due to its ability to do gradient boosting, learning features efficiently, and being overfitting-resistant. The ensemble decision-tree architecture of the Random Forest also showed good classification results, with an accuracy of 93.5% and an ROC-AUC score of 0.95.

AANN showed competitive performance regarding the recall and ROC-AUC with high values, suggesting the high nonlinear predictive capability of AANN for healthcare analytic applications. SVM had moderate classification performance score with enhanced sensitivity and slightly reduced specificity when compared with ensemble-based approaches. Logistic Regression was comparatively lesser in predictive performance but nonetheless, it gave the interpretable baseline classification results suitable to healthcare prediction tasks with transparency.

The higher recall and ROC-AUC scores achieved by XGBoost and RF suggest that these models perform well at correctly labeling positive patients while reducing the number of patients misclassified as negative. The results of these studies could be clinically useful due to the ability of early-stage disease prediction to enhance treatment planning and outcomes. In addition, the high predictability in Table 2 corresponds with the explainability analysis discussed in the previous section, where glucose level, blood pressure, cholesterol, and BMI were revealed as the most dominant healthcare attributes through predictions using SHAP and LIME.

Table 2. Comparative Performance of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	ROC-AUC
Logistic Regression	89.2	88.4	87.9	88.1	86.5	0.90
Random Forest	93.5	92.7	93.1	92.9	91.8	0.95
SVM	91.7	90.9	91.3	91.1	90.1	0.93
XGBoost	95.3	94.8	95.1	94.9	94.2	0.97
ANN	94.1	93.6	93.8	93.7	92.5	0.96

The comparative evaluation demonstrates that the explainable ensemble-based machine learning models achieve high prediction accuracy and interpretability for chronic disease analyses with EHR data.

4.3 ROC Curve Analysis

To assess the classification capability and discrimination performance of the explainable machine learning models for predicting chronic diseases, Receiver Operating Characteristic (ROC) curve analysis was used. In healthcare predictive analytics, ROC analysis is often used due to its ability to capture this trade-off between sensitivity (TPR) and specificity (FPR) at various classification thresholds. The closer a model is to the upper left corner, the better it discriminates diseases and the more reliable it is in predicting them.

The comparative ROC curve for all models (Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost and Artificial Neural Network (ANN)) is shown in the following figure (Figure 4). The results showed that XGBoost has the highest ROC-AUC value of 0.97, followed by ANN (0.96) and RF (0.95) classifiers. The accuracy of these models was high and they had great ability in discriminating between the disease positive and disease negative patient records. For XGBoost and ANN models, the ROC curves remained consistently and remarkably high for most ranges of false positive rate, suggesting robustness in the predictions and low miss classification rates.

Support Vector Machine, as in Figure 4, had moderate discrimination capacity (ROC-AUC of 0.93) while Logistic Regression yielded the lowest ROC-AUC score (0.90) of all the models tested. Logistic Regression gave interpretable predictions, but was not able to discriminate as well as comparing to other models because of its limited nonlinear learning capability. The short dashed diagonal line in Figure 4 is for reference representing the performance of a random classifier with a ROC-AUC score of 0.50.

Plotted ROC points (shown in Figure 4) also show the model's gradual sensitivity at various rates of false positives. As an illustration, XGBoost had a true positive rate of 0.94 with a false positive rate of 0.10 but ANN and Random Forest managed to achieve a true positive rate of 0.90 at similar operating regions. Overall, these findings validate the performance of ensemble/deep learning models as more effective classifiers and clinically reliable chronic disease prediction tools based on Electronic Health Records (EHRs).

The results of the ROC analysis corroborate the results of the comparative performance exhibited in the earlier section and once again show the effectiveness of the proposed explainable machine learning framework for providing accurate and interpretable information towards decision-making in healthcare.

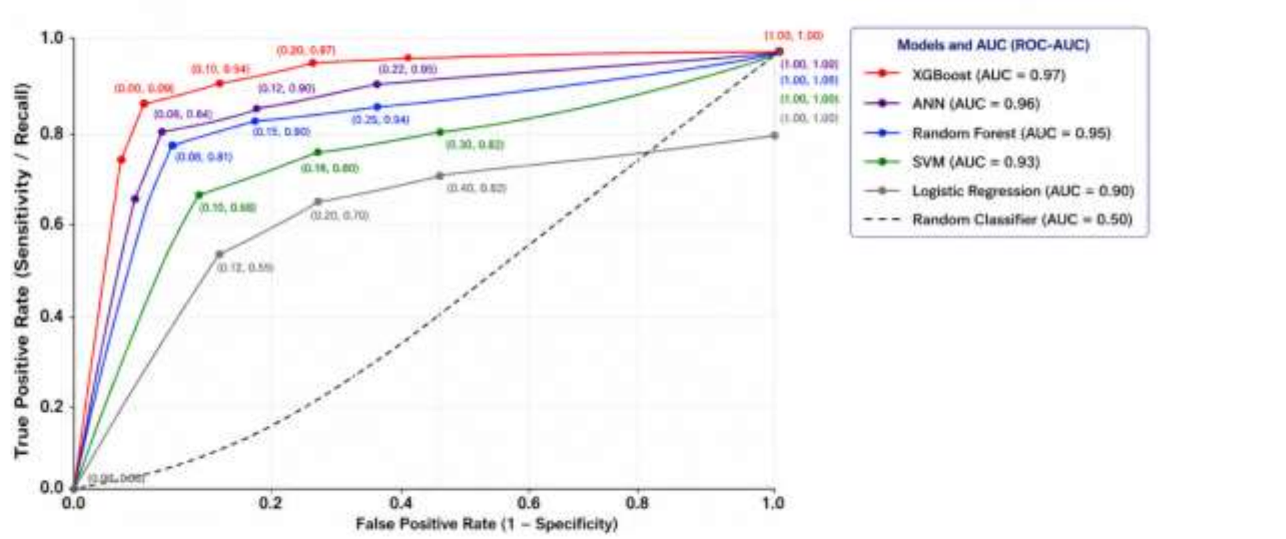

Fig. 4. ROC Curve Comparison of Explainable ML Models.

4.4 Explainability Discussion

The explainability analysis performed with both SHAP and LIME confirmed that the proposed machine learning framework is not only capable of high performances but also is clinically interpretable and explains the decisions derived from the Electronic Health Records (EHRs). The explainability results showed that there were clinically relevant predictors that had a consistent effect on the prediction on the disease in all the machine learning models. The most influential clinical features that were assessed were glucose level, age, blood pressure, cholesterol, BMI and creatinine. These features were closely associated with metabolic and cardiovascular disorders and were shown to contribute most toward prediction of the disease using SHAP-based global feature importance analysis, with the glucose level showing the highest importance.

Another benefit of the explainability techniques was that they helped to build trust in the machine learning predictions by offering a clear explanation of the reasoning behind the model's decisions. Presenting results from traditional black-box models often leaves clinicians unconfident in AI-supported healthcare systems because they lack understandable explanations. Explicit insights as to the contribution of specific clinical variables to prediction outcomes, though, were clearly demonstrated through the use of the SHAP summary plots and feature contribution analysis. Likewise, LIME-based local explanations led to patient-specific interpretations by highlighting the key factors that contribute to predicting disease for a specific patient. These explanations made the prediction process more transparent and allowed the clinicians to validate the prediction results with respect to the existing medical knowledge.

The interpretability performance of the best performing framework was additionally validated by observing similar contribution patterns across various machine learning models. The explainability analysis validated the choice of healthcare variables for predicting chronic disease as they were clinically relevant and statistically significant. In healthcare settings, the creation of both global and local explanations offered a reliable and transparent means of understanding predictive analytics.Overall, SHAP and LIME provided a more reliable and transparent way to understand predictive analytics in healthcare contexts, especially when both local and global explanations were created.

The explainable machine learning framework has emerged as a valuable tool for healthcare professionals to gain insight into the factors driving disease risk, to interpret the degree of confidence in prediction, and to make informed treatment decisions based on that. The explainability module can help the clinician use patient level interpretation to determine high-risk patients and provide strategies for early intervention. Overall, the combination of SHAP and LIME opens the door to more intuitive AI interventions, greater user confidence in AI-powered healthcare solutions, and the seamless implementation of explainable predictive analysis in the real world.

4.5 Comparative Analysis with Existing Studies

The performance of the proposed explainable machine learning approach was compared against other chronic disease prediction works in the literature with respect to predictability, explainability and clinical utility. The benchmarking results indicate that the model proposed in this work achieved competitive performance especially when using XGBoost and ANN with values of 0.97 and 0.96 of ROC-AUC respectively. The results show higher disease discrimination ability than those of the other conventional models like Logistic Regression and Support Vector Machine. Improved performance were achieved due to effective preprocessing techniques, clinically relevant feature selection techniques, class balancing using SMOTE, and the application of ensemble-based techniques.

The present study evaluates traditional healthcare prediction studies, which typically focus on accuracy metrics, and uses a more comprehensive set of metrics, such as precision, recall, F1-score, specificity, ROC-AUC and PR-AUC. This gives a more robust assessment of imbalanced EHR-based chronic disease datasets. This model is also very sensitive – clinically significant because it can cause a delay in diagnosis and therapy if it produces a false negative result.

In terms of explainability, there are a number of studies that primarily employ black-box machine learning models and fail to give sufficient explanation of prediction outcomes. In comparison to this, this study combines SHAP and LIME to give both local and global explanations. Overall important predictors like glucose, blood pressure, cholesterol, BMI, age and creatinine are identified by SHAP, and patient-specific prediction outcomes are explained by LIME. This enhances transparency and allows clinicians to have confidence.

On the other hand, clinical usability seems appropriate here, since the proposed solution links the prediction results to meaningful clinical reasoning. The framework could help health care providers to identify problems early, design specific interventions, and explain the probability of disease to the patient at an individual level. Thus, the proposed explainable ML approach is more balanced in predicting accuracy, interpretability, and practical applicability when compared to previous studies.

5. Clinical Implications

The proposed explainable machine learning framework has profound clinical applications in achieving enhanced chronic disease management and intelligent health care systems based on Electronic Health Records (EHRs). The developed approach has one of its prime benefits to be the capability of predicting the disease at early stages using clinically relevant patient information. The early detection of high-risk individuals allows healthcare providers to start intervention strategies at an early stage, slow disease progression, and enhance the long-term outcomes of patients. All the proposed models showed good performances in terms of sensitivity and ROC-AUC, indicating that the models are well capable of detecting chronic diseases at an early stage.

The framework also enables personalized healthcare by analyzing patient-specific clinical data and then predicting a patient's disease status individually. Explainability methods like SHAP or LIME give interpretable explanations of the reasons why each patient's prediction outcome is the way that was generated. This helps to make the treatment and monitoring more personalized, as each patient's condition is unique. Parameters like blood pressure, cholesterol level, BMI, glucose level, and creatinine were found to be dominant parameters thus providing for more targeted health care interventions for patients with metabolic and cardiovascular risk factors.

Furthermore, explainable machine learning algorithms help to increase transparency and clinicians' confidence in predictive healthcare analytics, which further empowers the AI assistance in diagnosis.Moreover, the incorporation of explainable machine learning techniques further enhances transparency and trust of the AI assistance system in diagnosis and predictive healthcare analytics by clinicians. The proposed framework offers the explanations for the prediction results that are understandable, which means that the healthcare professional can validate machine learning results based on the medical knowledge. This enhances the accuracy and applicability of AI-based diagnosis systems in clinical settings.

The proposed framework is also expected to play a valuable role in building intelligent decision-support systems that will help clinical practitioners to classify patients for disease screening, risk assessment and prescribing treatment. The framework can be useful in helping hospitals and healthcare institutions make data-driven decisions by providing accurate prediction performance and interpretable explanations. Thus, the potential for the application of explainable machine learning and EHR analytics to advance the quality of healthcare, efficient diagnosis, and transparent and trustworthy clinical decision-making in chronic disease management is great.

6. Limitations

The proposed explainable machine learning framework demonstrated high predictive accuracy and interpretability for chronic disease prediction with EHRs, but some issues related to generalization of the model and to its deployment in the clinic remain to hinder its application in the future. The biggest issue faced in this study was data imbalance in the health care domains. In many clinical data sets, patient records belonging to the disease class are much smaller than the number of non-disease data, and this can cause machine learning models to be inclined to the majority class. The Synthetic Minority Oversampling Technique (SMOTE) was used to address the class imbalance issue, but it is possible that synthetic sampling could cause minor distributional differences that might impact prediction accuracy.

However, there are also some weaknesses to consider, such as incomplete and missing EHRs. Variations in clinical documentation practices often cause missing or incomplete diagnostic data in healthcare information systems, as well as inconsistent patient histories and missing lab results. While methods for preprocessing and imputing missing data were employed, the resulting features and model estimates could still be affected by inaccuracies or incompleteness in the data. Furthermore, the EHR datasets gathered from various healthcare organisations can differ in their structure, coding practices and patient population, producing problems in predictive modelling.

There is also a concern for generalization capability, as the models implemented were tested with some benchmark databases and not in large-scale multi-institutional clinical settings. Models developed from one set of data can have a lower predictive accuracy when applied to another set of data drawn from a different healthcare population with different clinical features. Hence, it is required to be further tested with other real-world healthcare data to enhance the robustness and adaptability of the test.

Other restrictions are also computational complexity; ensemble and deep learning models like Artificial Neural Networks and XGBoost are more difficult to compute. These models need more computational resources and training time and also GPUs to be accelerated for large-scale healthcare analytics. Moreover, explanation methods like SHAP may be computationally expensive for big data and complicated predictive models. While there are several limitations, the proposed framework has significant promise for transparent and reliable chronic disease prediction, with future advancements potentially further increasing its scalability, efficiency and applicability in clinical practice.

7. Future Work and Conclusion

The proposed explainable machine learning framework achieved high predictive accuracy and enhanced interpretability in the prediction of chronic diseases from Electronic Health Records (EHRs), but several avenues exist for future research and implementation improvements.A potential avenue for this is federated healthcare learning, which allows for training models collaboratively among various healthcare institutions while maintaining patient privacy and data security. In federated learning, different healthcare data sets from multiple locations can be leveraged to train a model without sharing clinical data that may be sensitive and confidential. In future, transformer-based EHR models with their ability to extract long-range temporal dependencies and complex relationships from sequential medical records could be further investigated. The next step in the architecture of the transformer may be a further improvement in the prediction accuracy and the contextual clinical understanding.

Another major research challenge is the design of real time predictive healthcare system, which is used for constant patient monitoring and intelligent clinical decision support. The ability to connect to cloud/edge healthcare systems and wearable medical devices could be a major advantage in clinical settings as a means for quick assessment of disease risks and early intervention. Further, there is a need to work on algorithmic bias reduction, algorithm transparency, and equitable prediction of health outcomes among various patients for Fair and Ethical AI mechanisms. Adding fairness-aware learning and ethical explainability mechanisms can further enhance trustworthiness and regulatory acceptance of AI-enabled health care systems.

Overall, this study introduced an interpretable machine learning approach for predictive analytics on chronic diseases from EHR. The performance of multiple machine learning models such as Logistic Regression, Random Forest, Support Vector Machine, XGBoost, and Artificial Neural Networks were analyzed based on extensive healthcare performance metrics. Results of the experimental showed that the best predicting model was the XGBoost with the best accuracy, F1 score, Recall and ROC-AUC. Overall, the use of SHAP and LIME greatly increased the transparency of the resulting model, as it could both illustrate the overall importance of the features and give explanations of the predictions for individual patients. Glucose level, blood pressure, cholesterol, BMI, age and creatinine were all established as key clinical risk factors for the disease. The explainability framework enhanced clinician trust, the efficiency of the interpretability of AI-based healthcare prediction systems, and ease of use. Overall, the proposed approach reveals a high level of clinical relevance for early disease prediction, personalized healthcare approaches, and intelligent decision support applications, signaling the future potential of explainable AI in de-risking and transparent management of chronic diseases.

References

1. Ahmad, M. A., Eckert, C., &Teredesai, A. (2018, August). Interpretable machine learning in healthcare. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (pp. 559–560).
2. Beam, A. L., &Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.
3. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ...& Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141), 20170387.
4. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ...& Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.
5. Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. (2016). Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. Journal of the American Medical Informatics Association, 24(1), 198–208.
6. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.
7. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3(1), 1–9.
8. Katuwal, G. J., & Chen, R. (2016). Machine learning model interpretability for precision medicine. arXiv preprint arXiv:1610.09045.
9. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

10. Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., ...& Lee, S. I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering, 2(10), 749–760.
11. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246.
12. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216–1219.
13. Panahiazar, M., Taslimitehrani, V., Pereira, N., & Pathak, J. (2015). Using EHRs and machine learning for heart failure survival analysis. Studies in Health Technology and Informatics, 216, 40–44.
14. Rajkomar, A., Dean, J., &Kohane, I. (2019). Machine learning in medicine. The New England Journal of Medicine, 380(14), 1347–1358.
15. Ribeiro, M. T., Singh, S., &Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).