Journal of Ubiquitous Computing and Communication Technologies (ISSN: 2582-337X)
www.irojournals.com/jucct/

CARE-XAI: A Privacy-Preserving, Fair, and Explainable AI Framework for Multimodal Early Cardiovascular Risk Prediction

Revathy S P.[1], Swetha A B.[2], Vaishali Srija P B.[3]

[1]Professor, [2,3]Student, Department of Information Technology, Velammal Engineering College, Chennai, Tamil Nadu, India

E-mail: [1]revathysp91@gmail.com, [2]swethababu2003@gmail.com, [3]vaivesh2112@gmail.com

Abstract

Early prediction of cardiovascular diseases has always been a significant challenge due to data privacy risks, non-interpretable outcomes, and demographic biases associated with current machine learning approaches. In this paper, CARE-XAI (Comprehensive AI Risk Evaluation with Explainable Artificial Intelligence) is presented as a novel framework that combines federated learning, fairness optimization methods, and explainable artificial intelligence to predict cardiovascular diseases based on multiple clinical modalities. CARE-XAI uses the UCI Heart Disease database as the fundamental clinical modality, along with simulated behavioral and wearable datasets to improve predictive accuracy. Federated learning using the FedAvg algorithm ensures privacy-preserving training in a decentralized manner without any leakage of sensitive information about patients. Fairness optimization with the help of AIF360 is used to counteract demographic biases, while model transparency is guaranteed through SHAP explanations. Experimental results show that CARE-XAI provides an accuracy score of 83.1% with 82.5% precision, 84.3% recall, and 0.881 AUC values, while also performing comparably well with centralized frameworks.

Keywords: Explainable AI, Federated Learning, Healthcare, Privacy-Preserving, Heart Disease Prediction · Multimodal Learning, SHAP, Early Risk Detection.

DOI: https://doi.org/10.36548/jucct.2026.2.002
Received:20.02.2026, received in revised form:24.03.2026, accepted:09.04.2026, published: 20.04.2026

1. Introduction

Early detection and prediction have been very important factors in prevention in medicine, because they can facilitate early treatment and better health results. Nowadays, there is great potential in AI in terms of predicting diseases [1], there are several major problems associated with it, such as privacy issues, inability to understand predictions, and bias in the data.Machine learning algorithms need centralized information and therefore have increased chances of being breached. Also, there may be several problems associated with black box models, since they do not allow for understanding the predictions. Lastly, there is always the problem of the data used by AI being biased, leading to unequal health results between certain demographics.

CARE-XAI is a solution that combines the use of privacy preservation, fairness, and explainable AI technologies in order to solve the stated problem. Privacy preservation is accomplished by employing federated learning algorithms that allow training a model in a decentralized fashion and avoid transferring sensitive data [6]. Moreover, fairness-aware learning techniques are used to identify biases in the dataset and eliminate the impact of such biases on the model output. Finally, SHAP-based explainability allows analyzing predictions made by the model.The major difference between CARE-XAI and conventional models lies in their architecture and source of input data. While traditional methods usually employ datasets with a single modality of input data, CARE-XAI incorporates both clinical data from the UCI Heart Disease dataset [11] and synthetic behavioral and wearable device data to provide a better model of patient's condition and improve model's performance. The effectiveness of CARE-XAI was assessed using typical metrics. It showed a high level of performance in producing accurate, unbiased, and explainable predictions.

2. Literature Review

In the field of healthcare, machine learning has made remarkable progress in predicting cardiovascular risks in the last ten years. Early research focused mainly on using statistical techniques like logistic regression and random forests by using structured medical data, which includes blood pressure, cholesterol, and age [1]-[3]. These models produced acceptable predictions, but their effectiveness was hindered by their inability to detect

complicated non-linear connections within medical data. The emergence of deep learning has shown its superiority in this area, leading to better prediction outcomes [4]. Multimodal learning has emerged as a promising technique for predicting medical risks by combining structured medical data with data from wearable sensors and behavioral factors [5].

In spite of these developments, the implementation of artificial intelligence technology in medicine is limited by the problem of data security issues. Traditional algorithms for machine learning are based on using central databases, making them more prone to data leakage. To overcome these obstacles, federated learning techniques have been developed, which allow model training without revealing sensitive data. The algorithm FedAvg has been introduced [10], it allows several organizations to cooperatively develop models through exchanging only model parameters [9]. Further research has shown that federated learning provides similar results to traditional techniques without compromising privacy and is thus preferable in the medical sphere [7], [8].

Interpretability is yet another key limitation associated with AI in medicine. Most advanced models tend to be black boxes, hindering their applicability in medical practices. Explanations for AI were developed to tackle this problem and help make models more interpretable. One of them includes SHAP, a technique grounded in Shapley values and game theory that allows one to measure the importance of a certain feature in a prediction, thus making AI interpretable globally and locally [3].

Another important aspect related to AI in medicine is bias and fairness. Researchers have demonstrated that biased models can be trained due to data imbalance or even due to biases in training datasets [1]. This may lead to uneven health care results and even to the ethical problem itself. One way to deal with this problem is the usage of fairness-aware learning algorithms and toolkits, like AIF360. In particular, the toolkit allows measuring the bias in terms of demographic parity and equal opportunity metrics.

Moreover, the advent of multimodal health datasets has opened up novel avenues for achieving better prediction results. Continuous health monitoring via wearables is a supplementary source of information in addition to classical clinical data, which helps in making more complete predictions about patients' future health status. On the other hand, merging multiple different data sources poses certain technical difficulties in regard to varying data structure, size, and time characteristics.

Despite considerable advancements in each area separately, from federated learning to explainability and fairness as well as multimodal learning, existing scientific literature focuses on solving only one particular problem without considering the other factors. Namely, there currently lacks a single approach capable of incorporating privacy protection, bias mitigation, and explainability into a multimodal healthcare prediction framework.

3. Proposed Work

A. Overview of CARE-XAI Framework

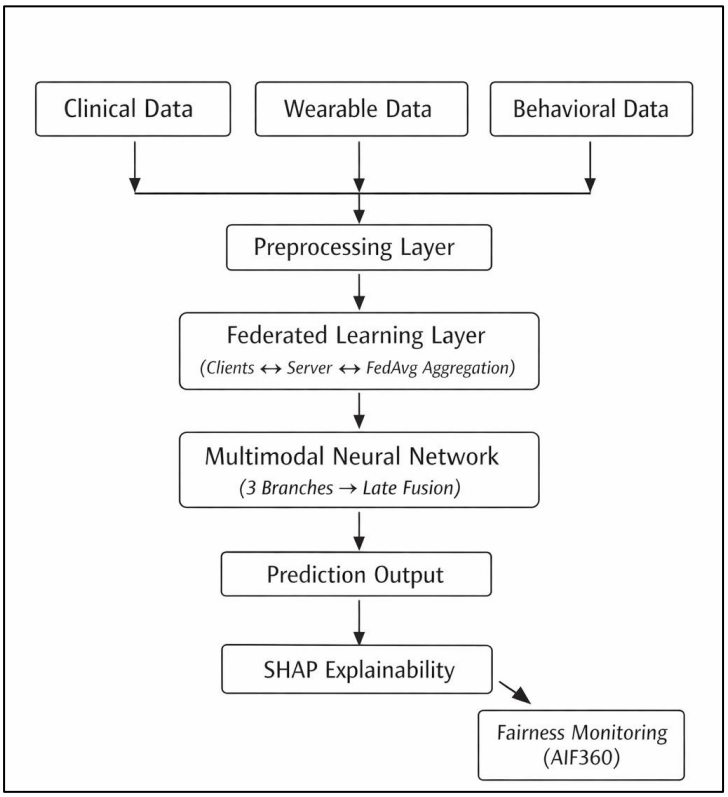


Figure 1. The System Architecture of the Proposed CARE-XAI Framework

The CARE-XAI framework aims at predicting the risk of having a cardiovascular disease early on in a way that preserves privacy, achieves fairness and offers model interpretability. The framework employs a federated learning approach whereby various hospitals collaborate in training a global machine learning model without the exchange of personal data belonging to their patients. The client trains a model locally using the available data for each hospital and sends the learned parameters to an aggregate server through FedAvg algorithm [10]. In such a way, data privacy is maintained and collaborative model learning is facilitated from distributed settings. The framework (Figure 1) incorporates multi-modal learning, fairness-aware learning and explains AI methods in one structure. Multi-modal

learning learns from clinical data, sensors and behavioral features of a patient as means of offering an overall perspective on health status. Fairness mechanisms are incorporated in the learning procedure in order to control biasness among different demographic populations.

B. System Architecture

CARE-XAI architecture can be viewed as an arrangement of different functional layers that work together to create an environment for safe and interpretable predictions. Table 1 below provides an overview of the CARE-XAI architecture layers and their functions. It also explains the flow of data within the CARE-XAI system starting from the input processing layer to the output explanation layer.

Table 1. Functional Layers of the CARE-XAI System Architecture

Layer	Function Description
Data Layer	Handles clinical, wearable, and behavioral data inputs
Preprocessing Layer	Performs cleaning, normalization, and encoding
Federated Learning Layer	Enables distributed training using Flower and FedAvg
Model Layer	Implements multimodal neural network with fusion
Explainability Layer	Provides SHAP-based interpretation

C. Dataset and Multimodal Integration

The main dataset used in this research is UCI Heart Disease Dataset [11], which includes data on 303 patients. Despite the fact that the initial set of attributes includes 76 variables, all machine learning studies, as well as the current paper, consider 14 relevant clinical parameters. These features include important characteristics such as age, gender, chest pain type, trestbps (resting blood pressure), chol (serum cholesterol), fbs (fasting blood sugar), restecg (resting electrocardiographic results), thalach (maximum heart rate achieved), exang (exercise-induced angina), oldpeak (ST depression), slope (the slope of the ST segment), ca (number of major vessels), and thal (thalassemia). The target variable is the presence or absence of heart disease.

Table 2. Feature Categories and Attributes in the UCI Heart Disease Dataset

Feature Category	Attributes
Demographic	Age, Sex
Clinical Measurements	Resting blood pressure, Cholesterol, Fasting blood sugar
Cardiac Test Results	Chest pain type, Rest ECG, Max heart rate, Exercise angina
Diagnostic Indicators	Oldpeak, Slope, Number of vessels (ca), Thalassemia
Target Variable	Presence of heart disease (0/1)

The table 2 showcases the selected 14 clinical attributes from the UCI Heart Disease dataset, categorized in an organized manner. This table highlights how the chosen attributes have been classified under different categories such as demographics, clinical measures, cardiac test results, and diagnostic criteria. The UCI Heart Disease dataset acts as the backbone for CARE-XAI's clinical modality. As mentioned previously, it should be highlighted that this dataset lacks any wearable or behavioral attributes. Thus, further modalities need to be added independently to enhance the predictive ability of the system. For multimodal learning, the artificial generation of the wearable and behavioral modalities is done by integrating them with the clinical modality.

The behavioral modality contains parameters like whether one smokes or not, stress level, and behavior pattern. These supplementary parameters are created under statistical assumptions and mapped with the clinical data set for the same feature scaling and distribution. Late fusion is performed for integrating the output of each modality, which helps to capture different perspectives of representation for the clinical as well as simulated data. By doing so, the dataset's originality is retained without compromising its use for multimodal learning analysis using the suggested framework. In addition, the dataset is used for evaluating fairness among different demographics like gender and age.

D. Implementation

As shown below, table 3 presents software libraries, frameworks, and hardware requirements adopted for developing the CARE-XAI model. This table summarizes the

resources that assist with deep learning, federated learning, fairness assessment, XAI, and deployment of the CARE-XAI framework.

Table 3. Development Environment Used in CARE-XAI Implementation

Component	Specification / Tool	Purpose
Programming Language	Python 3.8	Core implementation of the framework
Deep Learning Framework	TensorFlow 2.x	Model development and training
Federated Learning Framework	Flower	Distributed training and aggregation (FedAvg)
Fairness Toolkit	IBM AIF360	Bias detection and mitigation
Explainability Tool	SHAP (v0.41)	Model interpretability and feature contribution analysis
Data Processing Libraries	NumPy, Pandas	Data manipulation and preprocessing
Machine Learning Utilities	Scikit-learn	Data preprocessing and evaluation support
Backend Framework	Flask	Deployment of web-based interface
Frontend	HTML, CSS	User interface for visualization
Hardware	Intel Core i7 Processor	Computational processing
Memory	16 GB RAM	Handling model training and data processing

The preprocessing step guarantees high-quality and consistent data from all modalities. For the missing values in the UCI Heart Disease dataset, the mean is used as an imputation technique for numeric features and the mode for categorical variables. Normalization techniques for continuous features including blood pressure and cholesterol are done via Z-score standardization whereas encoding is adopted for categorical variables. In order to overcome the imbalance of data, SMOTE is used to balance the minority classes.

CARE-XAI uses a multi-modal neural network architecture (Figure 2) which is made up of three branches representing clinical, wearables, and behavior data, respectively. All three branches have fully-connected layers with 64 and 32 neurons, followed by ReLU activation and dropout layers. The outputs of all three branches are then merged using late fusion approach and passed through fully connected layers.

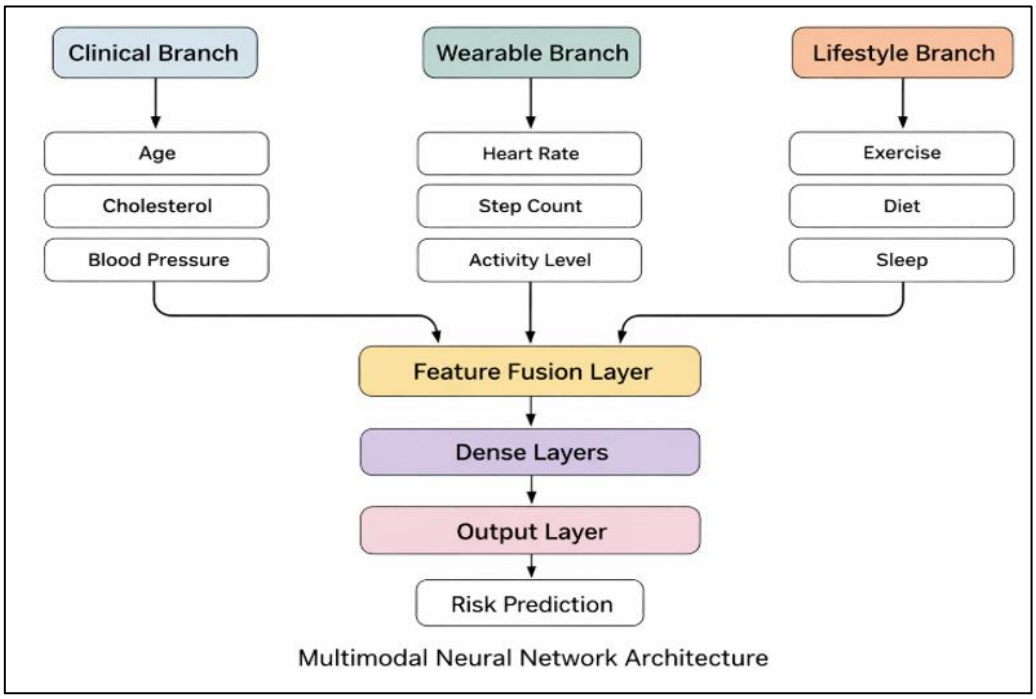


Figure 2. The Multimodal Neural Network Architecture

Table 4. Configuration of the Multimodal Neural Network Architecture

Component	Configuration
Hidden Layers	64 → 32 neurons
Activation Function	ReLU
Optimizer	Adam (0.001 learning rate)
Loss Function	Binary Cross-Entropy
Fusion Method	Late Fusion

Table 4 above illustrates the major configuration parameters for the multimodal neural network model, which include layer configurations, activation function types, optimization process, and integration technique employed. This offers a peek at the design considerations made when modeling clinical, wearable, and behavioral data.

Federated learning is achieved via the Flower framework. A global model is sent to the participating clients and trained locally. After that, updated parameters are sent back to the server and aggregated via the FedAvg algorithm [10].

The fairness of the model is measured using two metrics, namely demographic parity and equal opportunity. When discrepancies go above the set threshold value, reweighing and retraining of the model are used to ensure balanced predictions on the basis of different demographics. The explainability of the model is obtained via SHAP and provides insight into the effect of the individual features on the prediction of the model.

4. Results and Discussion

CARE-XAI is tested on the Heart Disease data set from the UCI collection, containing 303 patient records with 14 medical variables. Here, the data set serves as the dominant modality for the model, while some wearables and behavioral data are introduced as other modalities for multimodal learning. Data splitting into train and test sets is followed by evaluation of model accuracy through classification metrics such as Accuracy, Precision, Recall, F1-Score, and AUC, and evaluation of fairness and privacy.

The evaluation metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{\text{Precision . Recall}}{\text{Precision} + \text{Recall}}$$

Predictive performance comparison between CARE-XAI and centralized baseline models is made. Accuracy for the centralized model is 85.3%, while for the federated CARE-XAI model it is 83.1%. Even though there was a slight decrease in the accuracy score, the federated model still provides high levels of performance in all evaluation criteria.

Table 5. Performance Comparison between Centralized and Federated Models

Metric	Centralized Model	CARE-XAI (Federated)
Accuracy	85.30%	83.10%
Precision	84.70%	82.50%
Recall	86.20%	84.30%
F1-Score	85.40%	83.40%
AUC	0.892	0.881

The table 5 presents comparative analysis of the prediction accuracy between the centralized baseline model and CARE-XAI federated model based on evaluation criteria. It is obvious that federated approach can provide competitive results while maintaining privacy.

The results from the analysis on fairness indicate that the baseline algorithm demonstrates disparities between different demographic groups with demographic parity difference and equal opportunity difference of 0.18 and 0.22 respectively, but after applying fairness-aware learning, they become 0.07 and 0.11 respectively.

Table 6. Fairness Metrics Before and After Bias Mitigation

Fairness Metric	Before Mitigation	After Mitigation
Demographic Parity (Gender)	0.18	0.07
Equal Opportunity (Gender)	0.22	0.11
Demographic Parity (Age)	0.15	0.09

The explainability of the model is determined by SHAP, which determines the influence of each attribute on the results produced by the model. It is revealed from the analysis that cholesterol level, age, maximum heart rate, and chest pain category are some of the most critical features in the medical data (Figure 3). In addition, simulated behavioral characteristics such as smoking status and exercise frequency play a role in the outcomes produced.

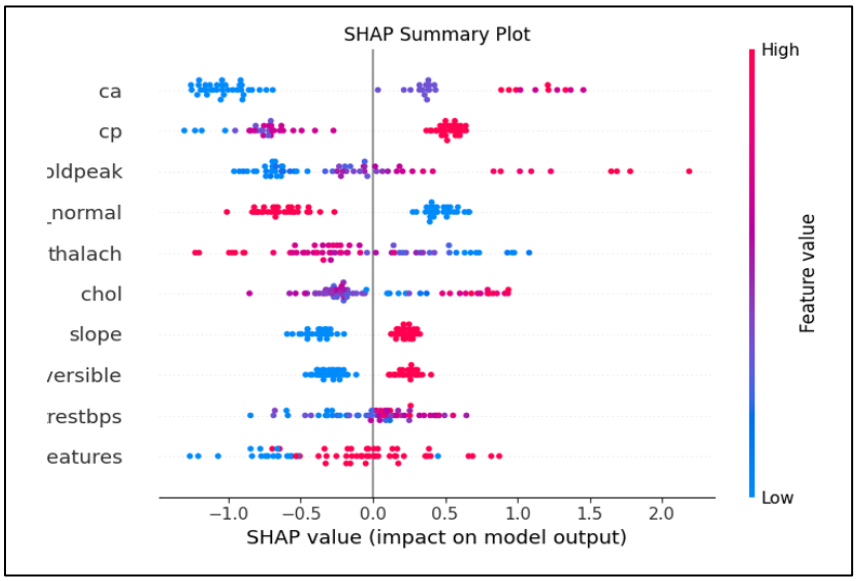
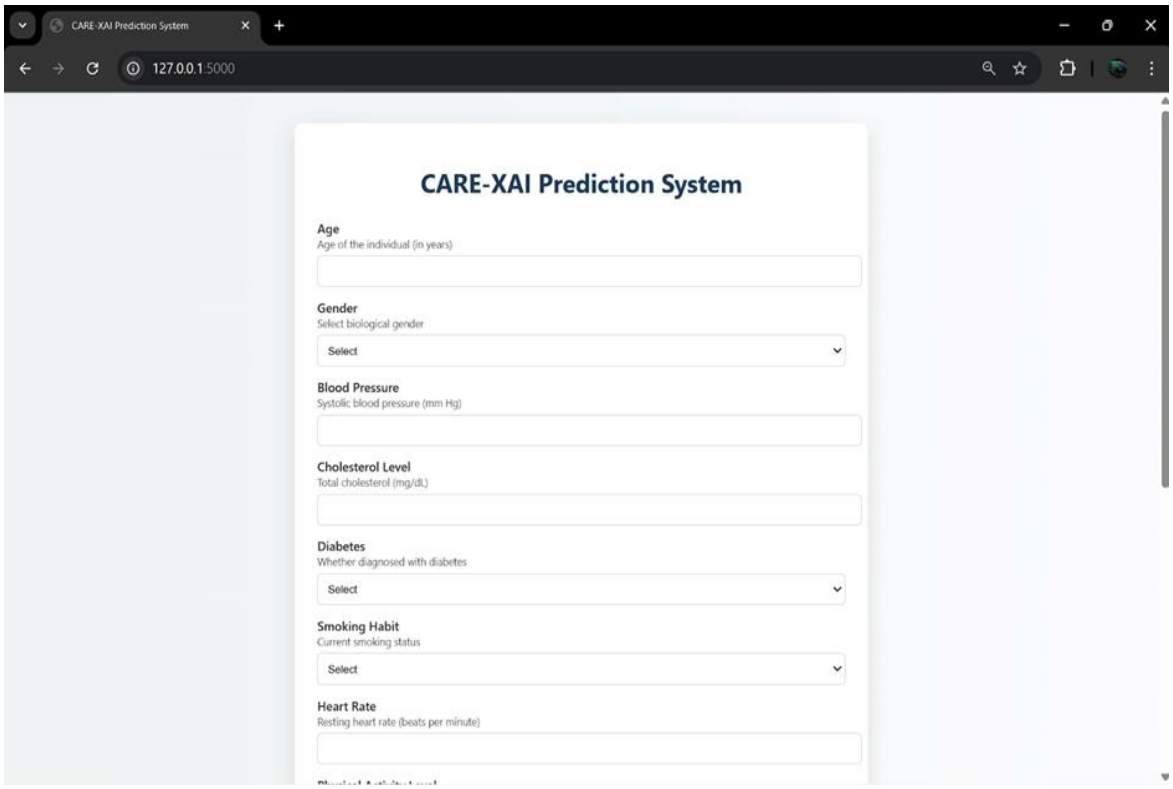


Figure 3. SHAP-based Feature Importance Explaining Model Decisions



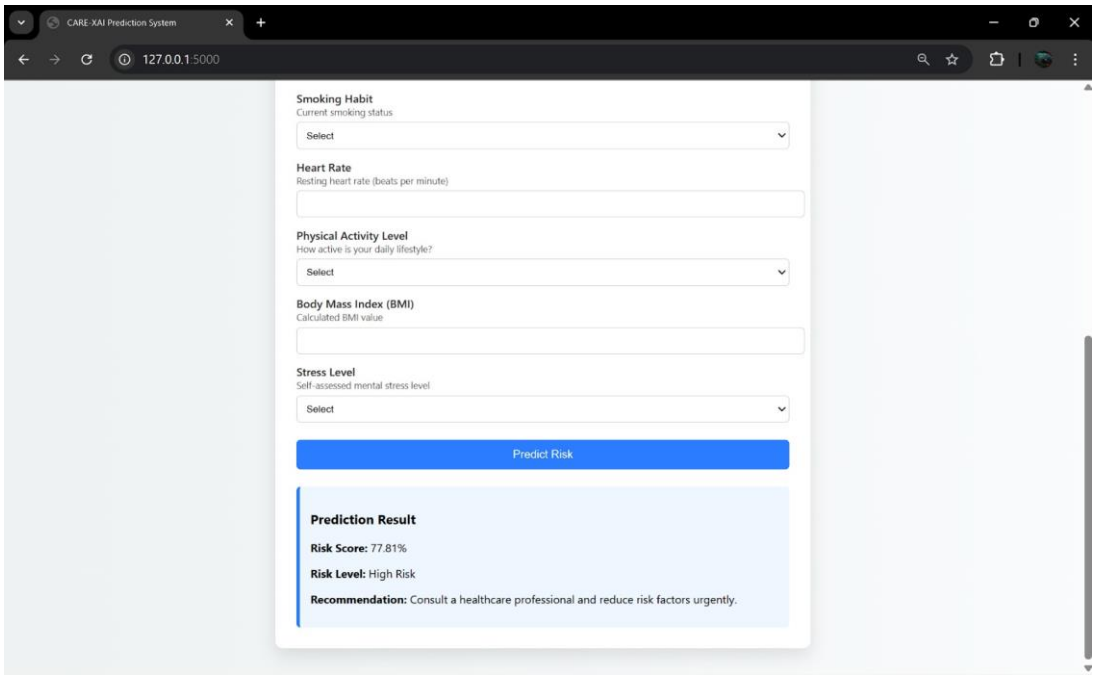


Figure 4. Sample Prediction Result with Risk Score and Explanation

Figure 4 shows how the CARE-XAI framework outputs a sample risk score and the explanations associated with it. The risk score is calculated in quantitative terms for the specific patient, along with SHAP values that show the importance of each feature in the score calculation. For example, positive values refer to features that cause an increase in the score, such as high levels of cholesterol and smoking. On the contrary, negative values refer to protective factors, which include physical exercise and low heart rates, among others. Thus, through such an output, clinicians can easily determine what features affect the final score.

5. Conclusion

In this study, CARE-XAI was introduced for predicting cardiovascular risks at an early stage by leveraging federated learning, fairness-aware optimization, and XAI in a multimodal setting. The suggested framework can be viewed as a solution to multiple challenges related to health-care AI applications, such as patient privacy, explainability of models, and discrimination against specific populations. In particular, the proposed framework uses the UCI Heart Disease dataset together with artificially generated wearable and behavioral data to improve the prediction accuracy. Experiments have shown that CARE-XAI has 83.1% accuracy and 82.5% precision and 84.3% recall, as well as AUC of 0.881, which is comparable to a centralized version of the model. Using fairness-aware learning leads to considerable improvements in eliminating the demographic bias, with demographic

parity increasing from 0.18 to 0.07 and equal opportunity improving from 0.22 to 0.11. Finally, the proposed use of SHAP helps to achieve better explainability of models, which contributes to their practical implementation in health care. Therefore, CARE-XAI can be considered a promising approach in terms of achieving the balance between accuracy, privacy, fairness, and explainability.

References

[1] Krittanawong, Chayakrit, HongJu Zhang, Zhen Wang, Mehmet Aydar, and Takeshi Kitai. "Artificial intelligence in precision cardiovascular medicine." Journal of the American College of Cardiology 69, no. 21 (2017): 2657-2664.

[2] Dey, Damini, Piotr J. Slomka, Paul Leeson, Dorin Comaniciu, Sirish Shrestha, Partho P. Sengupta, and Thomas H. Marwick. "Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review." Journal of the American College of Cardiology 73, no. 11 (2019): 1317-1335.

[3] Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. "Machine learning in medicine." New England Journal of Medicine 380, no. 14 (2019): 1347-1358.

[4] Esteva, Andre, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. "Deep learning-enabled medical computer vision." NPJ digital medicine 4, no. 1 (2021), p. 5.

[5] Johnson, Kipp W., Jessica Torres Soto, Benjamin S. Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley, and Joel T. Dudley. "Artificial intelligence in cardiology." Journal of the American College of Cardiology 71, no. 23 (2018): 2668-2679.

[6] Alaa, Ahmed M., Thomas Bolton, Emanuele Di Angelantonio, James HF Rudd, and Mihaela Van der Schaar. "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants." PloS one 14, no. 5 (2019): e0213653.

[7] Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. "Federated learning for healthcare informatics." Journal of healthcare informatics research 5, no. 1 (2021): 1-19.

[8] Warnat-Herresthal, Stefanie, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara "Swarm learning for decentralized and confidential clinical machine learning." Nature 594, no. 7862 (2021): 265-270.

[9] Li, Tian, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. "Federated optimization in heterogeneous networks." Proceedings of Machine learning and systems 2 (2020): 429-450.

[10] Kairouz, Peter, and H. Brendan McMahan. "Advances and open problems in federated learning." Foundations and trends in machine learning 14, no. 1-2 (2021): 1-210.

[11] Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. Heart Disease [Dataset]. UCI Machine Learning Repository - https://archive.ics.uci.edu/dataset/45/heart+disease