Journal of Informatics and Web Engineering

Vol. 5 No. 1 (February 2026) eISSN: 2821-370X

XAI -Driven Explainability for Cardiovascular Diseases Prediction

Jacqueline Dike[1], Jarutas Andritsch[2*]

[1,2]Department of Science and Engineering, Southampton Solent University, East Park Terrace, Southampton, SO14 0YN,
United Kingdom

corresponding author: (jarutas.andritsch@solent.ac.uk; ORCiD: 0000-0003-2670-7111)

Abstract - The adoption of Artificial Intelligence (AI) in cardiovascular disease prediction has significantly improved risk stratification, offering new avenues for early diagnosis and preventive care. With the growing availability of electronic health records and structured clinical datasets, Machine Learning (ML) and Deep Learning (DL) models have demonstrated strong predictive capabilities. However, despite their performance, its adoption in healthcare is often constrained by the lack of transparency and interpretability in many ML and DL models. This lack of explainability undermines clinical trust and raises ethical concerns. In high-stakes domains such as Cardiovascular Disease (CVD) prediction, clinicians require not only accurate outputs but also clear explanations of how those predictions are derived. This paper presents a comparative evaluation of Explainable AI (XAI) techniques applied to both conventional ML models such as Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest and DL architectures including AutoInt, FT-Transformer, and Category Embedding. Using the Framingham Heart Study dataset, this study integrates Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) to assess model interpretability and feature relevance. Results show that conventional models offer superior explainability with comparable predictive accuracy, while DL models, although slightly less interpretable, demonstrate potential with advanced XAI techniques. The findings advocate hybrid approaches that balance accuracy and interpretability, supporting ethical and practical AI deployment in healthcare.

Keywords—Cardiovascular Disease, Explainable AI, Shapley Additive Explanations, Local Interpretable Model-Agnostic Explanations, Deep Learning, Conventional Model.

Received: 12 July 2025; Accepted: 18 October 2025; Published: 16 February 2026

1. INTRODUCTION

Cardiovascular Disease (CVD) continues to be one of the most significant causes of death globally. It is primarily caused by the narrowing or blockage of blood vessels, which can result in serious and potentially fatal complications such as heart attacks, angina (chest pain), and heart failure. These conditions present major public health concerns due to their high mortality and morbidity rates. For over 15 years, CVD has consistently remained the leading cause of death worldwide. According to the World Health Organization (WHO), CVD was responsible for approximately 15

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.1.10

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

million deaths in 2015. By 2020, this number had risen to 17.9 million, with forecasts predicting the toll could exceed 23.6 million by 2030 [1].

With the advent of Artificial Intelligence (AI) for preemptive detection of CVD, the possibility of misdiagnosis via prediction is significantly reduced. However, one of the most critical oversights that many AI systems suffer is transparency or the lack thereof in model development. Many AI models are "black box" models that fail to disclose how they arrive at predictions, which presents significant barriers to clinical use. Lack of transparency breeds mistrust for the clinician and the patient, and without understanding how a model result was concluded, it cannot be part of the decision-making process [2-3]. With the CVD and its mortality complications that account for an increasingly disproportionate percentage of global death, the predictive models must be developed with intentionality, accuracy, and transparency. Transparent models maintain the trust factor of the clinician, appropriate ethical application, and regulatory observation. In addition, with the onset of big data increasingly available due to Electronic Health Record (EHR) accessibility, genomics, and lifestyle factors, the ability for personal risk prediction is on the rise. Still, it can only be fulfilled through explainable AI models that disclose how they come to their decisions. Explainable AI (XAI) solves this problem by assessing the reason behind AI generated predictions [4]. Thus, the main contribution of this study is to assess the explainability potential of traditional Machine Learning (ML) and Deep Learning (DL) generated models via two applications of XAI: the Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) technique. By doing this, we can explore the trade-offs between interpretability and accuracy and offer insights into the practical deployment of XAI-powered diagnostic tools in clinical decision-making. The paper is organized as follows: Literature works are discussed in Section 2. The research methodology is presented in Section 3 while Section 4 demonstrates and discusses results. Section 5 provides the conclusion and future work.

2. LITERATURE REVIEW

2.1 The Rise of AI in Cardiovascular Prediction

AI approaches claim to transform healthcare diagnostics by determining whether patients are at risk for disease. However, many practitioners still favour standard ML techniques like Logistic Regression (LR) and Decision Tree (DT), which are interpretable and uncomplicated [5-7]. For example, in the researcher's analysis of cardiovascular risk, Rudin argues that simpler, easier models to understand, such as DTs, should dominate over more complex techniques. Simpler ones are often faster and more accurate relative to ill-defined black box operations, as demonstrated by successfully learning from the Framingham dataset [8].

Conventional ML techniques often struggle to replicate and predict results for complicated or high-dimensional datasets. When data involves nonlinear relationships or cannot be separated using simple binary classification, these models reach a point of diminishing returns. When trained on robust datasets over time relative to aggregates of health data, whether medical imaging, genomic data, or signals from monitoring sessions, DL techniques outperform routine predictive accuracies [9]. Thus, when assessing thousands of variables related to health risks including but not limited to demographics, clinical history, Electrocardiograms (ECGs), and medical imaging, DL will better determine health risk accurately.

Yet such accuracy does not always succeed when merged within a clinical practice setting. The significant hurdle for implementation is transparency. Many ML and DL models operate as a "black box" where little about how output came to be through hidden layers is known. For ethical, trustworthy, quality decision-making, physicians must understand how and why AI determined what it produced to integrate such findings legally and ethically effectively. Context is critical in an ambiguous medical environment [10]. Therefore, XAI was born to establish that techniques existed to provide transparency. XAI provides the possibility of several techniques making model output explorable. It offers transparency among clinicians, regulators, and patients alike; more informed decisions create safer implementations of AI in healthcare.

2.2. Explainable AI Techniques: SHAP and LIME

The field has implemented SHAP and LIME for explainability. SHAP provides global and local explainability because it aims to explain how predicted values can be attributed to given input features [11]. LIME is more locally driven, as it takes features of interest to create a more simple, explainable model that can approximate the complex black box model in the local region around the prediction of interest [12].

For example, when estimating an individual 10-year risk of developing CVD, You et al. explored many ML models like Light Gradient Boosting Machine (LGBM), eXtreme Gradient Boosting Machine (XGBoost), Random Forest (RF), LR, K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Artificial Neural Networks (ANN) that were applied to the data gathered from their research [13]. While many of these models were black boxes, applying SHAP allowed them to explain their findings by essentially “breaking the black box.” Li et al. predicted the risk of brain metastases from the structured EHR data using Reverse Time AttentIoN (RETAIN) model. The decision results of the model were interpreted using SHAP values based on a feature attribution to identify the factors contributing to the model [14].

Rezk et al. presented a framework that utilized hybrid ensemble learning models which combined LightBoost and XGBoost algorithms together with SHAP and LIME analysis to create an explainable heart disease prediction model. By integrating LIME with ML techniques, the system offers a comprehensive framework to enhance the efficacy and reliability of the heart disease prediction model by showing the features that are important to the prediction. In addition to SHAP, LIME has been employed to generate local, instance-specific explanations, allowing clinicians to understand the reasoning behind individual risk predictions. These case-by-case insights are especially valuable in clinical settings where personalized interpretation is essential.

Overall, incorporating XAI techniques into CVD prediction models enhances their transparency and interpretability. By clearly identifying influential features such as ST slope, oldpeak, chest pain type, max heart rate and cholesterol, the model could potentially help with a clear understanding of its decision-making process [15]. Similarly, Petmezas et.al. presented a prediction model of heart failure using Extremely Randomized Trees (Extra-Trees) and non-linear correlation measures to enhance mortality prediction in HF patients. This model also utilized SHAP to improve the interpretability of the model [16]. XAI facilitates the translation of complex model decisions into actionable, clinically meaningful information. This interpretability is key to building trust among healthcare professionals and promoting the adoption of AI-driven tools in medical practice.

2.3 Balancing Accuracy and Interpretability

Research suggests a trade-off between model performance and explainability. For instance, DTs and LRs offer clarity but may underperform on complex tasks. In contrast, ensemble and DL models achieve high accuracy but require XAI tools to bridge the interpretability gap.

2.3.1 AutoPrognosis vs. Clinical Scores

Alaa et al. compared an automated ML ensemble (AutoPrognosis) to the Framingham Risk Score in predicting cardiovascular events. AutoPrognosis, an optimized ensemble of many models, was significantly more accurate, achieving an AUC of 0.77 and correctly predicting 368 more cases of CVD over 5 years than Framingham. AutoPrognosis combined hundreds of features and models (a clear example of black box to clinicians). This exemplifies the accuracy gain vs loss of interpretability. Interestingly, the study noted that adding a wealth of novel risk factors (e.g. walking pace, health rating) boosted accuracy more than the choice of complex model itself, suggesting richer data plus moderately complex models may represent a balanced approach [17].

2.3.2 Hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM)

Hossain et al. proposed a hybrid DL model combining a CNN and LSTM to identify CVD. Also, the model combines feature engineering and explainable AI to enhance the accuracy and interpretability of the prediction. The proposed model achieved an accuracy of 73.52%. These studies showed that creatively designed ensembles prediction model with explainable components can balance performance and interpretability [18].

2.3.3 Atrial Fibrillation Monitoring With XAI

She et al. introduced an explainable AI tool named "AF'fective" to assist cardiologists in monitoring patients with atrial fibrillation following catheter ablation. In a real-world pilot study, the system processed ECG data and other patient inputs to predict the risk of AF recurrence. Its output was both a prediction with an associated explanation i.e., a portion of an ECG with deviation and the output number of episodes/year and an explanation without input. These outputs served as a suggested, but not definitive, diagnosis to the cardiologist, such as "number of previous ablation attempts." The need for explanations was validated through workshops and focus group sessions with cardiologists who needed to align their medical experience with the AI-generated results [19].

Nevertheless, even with the benefits of using XAI in cardiovascular risk assessment, some challenges persist. First, despite attempts at external validation of clinically created models, few achieve success using larger, reputable databases with known applicability to the real world like the Framingham Heart Study. This creates generalizability concerns for many results, especially those from multicenter, heterogeneous patient populations [20]. Second, many models fail to address seamless clinical implementation or intended end-user use. A significant percentage of the articles published focus on breakthroughs related to the algorithm as opposed to yielding strong implementation considerations through clinically practicable and usable interpretability [21]. This is crucial for everyday use, as clinicians require results that are accurate, sensitive, and specific; they also require that AIs can demonstrate the output that can be practically applicable to assessing patient risk to make a difference. Accomplishing this requires a paradigm shift in AI so that quality of transparency, and quality of relevant output exist only then does reliable implementation of AI into modern medicine become feasible [22].

3. RESEARCH METHODOLOGY

This study adopts a four-stage Define-Collect-Select-Apply (DCSA) methodology where conventional ML and DL are employed through XAI for CVD diagnosis. The subsequent procedures are sequential and expounded upon in Figure 1.

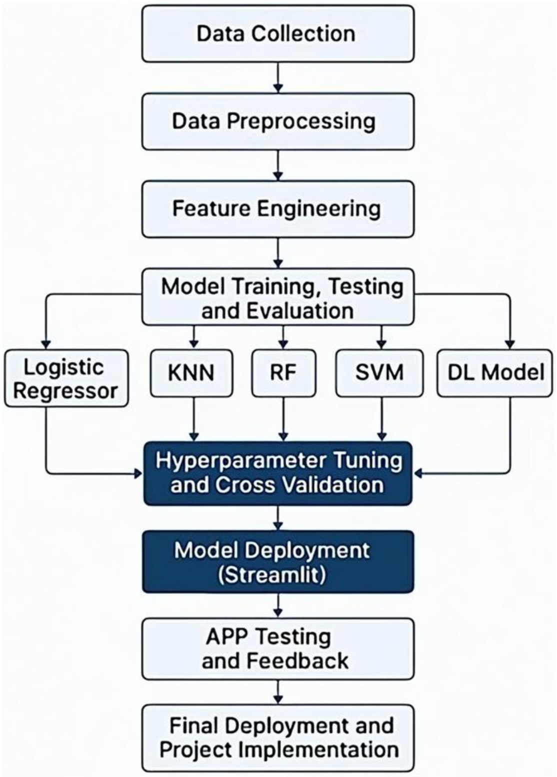
Data Collection
Data Preprocessing
Feature Engineering
Model Training, Testing and Evaluation
Logistic Regressor
KNN
RF
SVM
DL Model
Hyperparameter Tuning and Cross Validation
Model Deployment (Streamlit)
APP Testing and Feedback
Final Deployment and Project Implementation

Figure 1. XAI-Driven Explainability for CVDs Prediction Research Process

The aim of this research is to find the best prediction models of CVD within an XAI paradigm that makes results interpretable and clinically applicable to the following independent variables linked to the estimated prior expected chance of having CVD as a dependent variable: clinical variables (cholesterol, blood pressure), demographic variables (age, sex), lifestyle variables (smoking, exercise), socioeconomic variables (income, education). Therefore, the research question was set to explore and achieve the research intention:

What is the performance of different AI prediction models in assessing cardiovascular risk when combined with XAI strategies?

3.1 Data Collection

This study employs data obtained from the Teaching Request Department of the National Heart, Lung, and Blood Institute (NHLBI). The dataset, derived from the Framingham Heart Study (FHS), was provided upon formal request and with the institutional approval. It consists of 4,434 participants, with a total of 11,627 recorded observations. It includes 39 features, encompassing a broad spectrum of demographic, behavioural, medical, and physical health indicators, such as age, gender, smoking habits, blood pressure, cholesterol levels, BMI, glucose levels, and a medical history of diabetes or hypertension.

3.2 Data Pre-processing and Feature Engineering

The data preprocessing phase involved several key steps to ensure data integrity and readiness for analysis. The category variables like age groups and gender (female vs. male) were used as independent variables. The reason for this inclusion is that by using these as categorical independent variables, the study can analyze risk by specific subgroup, thereby also creating equity in the XAI results across a diverse population in its findings. Column names were standardized for consistency, and redundant fields were removed to enable us to focus on the columns needed mainly for prediction. Missing values were addressed using the pandas null methods to identify the null values in each attribute of the dataset. It was discovered that the High-Density Lipoprotein Cholesterol (HDLC) and Low-Density Lipoprotein Cholesterol (LDLC) columns had 73% null values in the dataset, and these columns could not be imputed due to the large percentage of null values. Therefore, these columns were removed. Other columns with less than 20% missing data were treated using appropriate imputation techniques. Duplicate entries were identified by using the pd.duplicates() method with all the columns in the dataset as subsets and it was discovered that the dataset has no duplicate records, data types were verified to understand the type of analysis that would be done on each column, and outliers were identified using boxplots and z-score analysis across continuous variables. Variables such as systolic blood pressure and Body Mass Index (BMI) exhibited mild to moderate skewness, which was noted for consideration during model selection and evaluation. While no outliers were removed at this stage, their presence informed the choice of robust algorithms that are less sensitive to distributional variance.

Categorical variables, such as gender were encoded using one-hot encoding, this technique creates separate binary variables for each category within a feature, ensuring that the model treats them as distinct entities. One-hot encoding is especially useful for nominal categorical variables, where categories lack a natural ranking or hierarchical structure. To mitigate the influence of varying feature scales on ML algorithms, continuous variables were normalized using Standard Scaling transforming them into a mean of zero and a standard deviation of one. This ensures balanced feature contribution during model training, especially for algorithms sensitive to feature magnitude.

We also checked the CVD variable, which was the target variable for the prediction model. The record of CVD variables showed a class imbalance between normal (which is 0, with 8,728 records) and Coronary Artery Disease (CAD) diagnosed (which is 1, with 2,899 records). Therefore, the Synthetic Minority Oversampling Technique (SMOTE) was implemented to handle class imbalance in the training dataset.

3.3 Model Selection

This study employed LR, SVM, DT, and RF as the primary models for assessing cardiovascular risk. These models were selected due to their balance between predictive accuracy and interpretability, which is essential for implementing XAI techniques. Each model contributes unique advantages, enabling both precise predictions and transparent insights into the factors influencing cardiovascular health. Below is a breakdown of each model and its relevance to XAI.

3.4 Model Development

For model development, conventional ML algorithms, LR, SVM, RF and DT were employed to predict the occurrence of CVD. For the DL models, algorithms such as FTT Model, AutoInt, and Category Embedding were utilized. The dataset was split into training and testing subsets in the ratio of 70:30, with models trained on the former and evaluated on the latter. With a solid foundation established through preprocessing and model training, Explainable AI techniques were integrated to enhance model transparency. Tools such as LIME and SHAP were applied to both ML and DL models to interpret prediction outputs. This allowed for the identification of the most influential features contributing to each model's decisions and facilitated a comparative analysis of interpretability across algorithms.

3.5 Model Evaluation

In this work, four evaluation metrics were utilized: Accuracy, Precision, Recall, and F1-score. These metrics are commonly used in classification problems and provide valuable insights into the performance of the algorithms. To evaluate and compare the effectiveness of different classifiers, a comparative analysis was conducted based on these indicators. For comparing the performance of the different prediction models with XAI techniques, the dataset for testing the models was prepared during the data preprocessing step. This same dataset was used to train and test the selected models, as presented in the Model Selection section. The performance results from each model were recorded to enable comparison later.

4. RESULTS AND DISCUSSIONS

This section presents the comparative performance evaluation of the models used in this study, highlighting their effectiveness based on key metrics, as shown in Table 1.

Table 1. Comparative Performance Evaluation of Models

Metrics	Model						
	Conventional ML				DL		
	LR	DT	SVM	RF	AutoInt	FTT Model	Category Embedding
Accuracy	90.10%	**93.50%**	91.40%	**93.40%**	77.14%	88.14%	86.96%
Avg.Precision	86.00%	**93.00%**	92.00%	**93.00%**	57.27%	89.39%	86.83%
Avg.Recall	88.00%	**89.00%**	85.00%	**89.00%**	75.68%	88.14%	75.84%
Avg. F1-score	87.00%	**91.00%**	88.00%	**91.00%**	65.20%	88.51%	85.86%
Explainability	High	**Very High**	Moderate (Post hoc with SHAP/LIME)	**Moderate**	Moderate (Post hoc with SHAP)	Low (Post hoc with Deep SHAP)	Moderate (Post hoc with SHAP)

In response to the research question identified earlier, we analysed the performance of many conventional and DL techniques that utilized XAI techniques in the model. These models were assessed based on accuracy, F1-score and explainability. As reflected by Table 1, conventional models have good prediction abilities and are relatively transparent with less computational burden. For instance, LR provides 90.10% accuracy and 87.00% F1-score, establishing a confident, anticipated baseline. In addition, it is completely interpretable as its model coefficients reflect positive or negative contributions of each feature relative to the output; thus, it is clinically aligned with risk scoring.

Decision trees outperform LR with 93.50% accuracy and 91.00% F1-score yet also possess high interpretability and low computational burden. Their structure of if-then rules closely mimics the if-then considerations clinical professionals make. Although decision trees tend to overfit, hyperparameter tuning showed commendable performance statistics with low complexity. SVMs also showed competitive performance (91.40% accuracy, 88.00% F1), but their limited transparency, especially with non-linear kernels, reduces their applicability in clinical environments. RF, an ensemble of decision trees, matched the accuracy and F1-score of individual decision trees (93.40%, 91.00%) while improving generalization. However, this came at the cost of reduced interpretability, as individual decision paths are less traceable.

Post-hoc XAI tools such as SHAP and feature importance plots were employed to mitigate this limitation and support model transparency. DL models, including transformer-based architectures, were also evaluated. AutoInt, despite its potential to capture complex feature interactions, underperformed with 77.14% accuracy and a 65.20% F1-score likely due to dataset limitations and overfitting. The FT-Transformer (FTT) improved upon this, achieving 88.14% accuracy and 88.51% F1-score by leveraging attention mechanisms to model feature dependencies. However, its default opacity necessitated the use of Deep SHAP to interpret predictions. The most important predictors out of all three trained models were Time since CVD diagnosis (TIMECVD), Time since Angina Pectoris (TIMEAP), and Time to Death (TIMEDTH) regardless of RF, ANN, or decision tree classification. In addition, SHAP was the XAI application that increased explainability via the traditional and DL approach as it assessed visually the contribution of each specific input feature over the others. The most effective models in accuracy achievement were decision tree and LR (90-93% accuracy), which were also the most explainable without any XAI application due to their inherent interpretability features. This indicates that explainability does not come at the cost of accuracy.

The convergence of findings from the coefficient-based analysis and the SHAP analysis strengthens our confidence in the results. Both methods identified TIMECVD, TIMEAP, and TIMEDTH as the primary drivers of the model's predictions, indicating that these time-to-event measures are critical factors in the outcome under study. Using the LR coefficients, we gained a clear, quantitative sense of each feature's global effect. The SHAP analysis provided insights into the local behavior of the model; it tells us how much each feature is contributing to each individual's risk prediction. This is particularly advantageous for complex or non-linear models. However, even in our LR (which is linear by design), SHAP helped visualize the consistency of effects across individuals and detect any potential outliers or interaction effects. One strength of SHAP is that it can handle feature interactions and non-linear relationships by attributing contributions in a game-theoretic manner. In our case, the largely linear pattern in the SHAP plots suggested that there were no strong interaction effects driving the predictions. The model behaved as a mostly additive, linear combination of features, which is expected to give the LR framework. A slight limitation of SHAP is that it is computationally more intensive and can be trickier to explain to stakeholders not familiar with the concept. In contrast, odds ratios from LR are a long-standing common language in clinical research.

There are benefits and shortcomings with independent research as subjects to projects. For instance, there are gaps in the analysed domain knowledge, where ideally multi-disciplinary professions should fill these collaborative efforts [23]. For example, information input from clinicians- or at least feedback with a peer review-would have added a comprehensive layer of understanding, confirming whether the research found an expected conclusion or beyond expectation. Additionally, if effective feedback was given, revisions could have added a deeper nuanced understanding to the research intent. Another challenge lies in the computational demands of advanced XAI techniques. Methods such as SHAP and LIME on DL models require substantial processing power, which can hinder their application in resource-constrained environments. During model testing, performance significantly slowed. This highlighted the potential impracticality of deploying such techniques in real-time clinical settings without high-performance computing resources [24]. Finally, not all measures of interpretability are feasible. There are no agreed-upon metrics or frameworks adopted by industry that allow determination of how understandable AI systems are. Thus, because no clinician was involved in the discussions, it was increasingly hard to determine if the generated predictive outputs made sense for diagnostics or if they aligned with clinically understood options that would work for professional practitioners [25].

5. CONCLUSION AND FUTURE WORK

This comparative evaluation underscores the critical trade-offs between accuracy, interpretability, and computational cost necessary to derive explainable AI for healthcare. Ultimately, the more simplistic approaches-LR and decision trees-provided not only appropriate levels of accuracy (~90–93%) but also the transparency needed for healthcare adoption in a clinical setting. Notably, the decision tree, among the most interpretable models, was one of the top performers, affirming that high explainability does not require sacrificing performance. LR also remains attractive for its simplicity and direct risk attribute.

In contrast, while DL models (transformers and embedding-based networks) offered additional modelling power, they introduced interpretability challenges and higher computational demands. In clinical settings, where understanding and trust are paramount, marginal improvements in accuracy may not justify the use of opaque models. XAI tools such as SHAP are vital for making complex models more transparent. By applying these techniques, we aimed to "open the black box" and translate model decisions into actionable insights for clinicians. Evidence suggests that clear and interpretable explanations increase trust in AI-driven predictions and are an essential factor in healthcare applications.

To build upon the current research, future work should prioritize integrating multidisciplinary collaboration, particularly with clinicians and biomedical informaticians. Incorporating expert-annotated datasets and domain-guided feature engineering could enhance model interpretability and improve the clinical validity of feature attribution methods such as SHAP and LIME. Addressing the computational overhead associated with advanced XAI techniques is also critical. Future implementations should explore distributed computing environments, GPU-accelerated frameworks, or cloud-based platforms (e.g., AWS SageMaker, Google Cloud AI Platform) to support the real-time generation of explanations on a scale.

ACKNOWLEDGEMENT

We would like to express my sincere gratitude to the FHS team at Boston University and the NHLBI for granting access to the dataset utilized in this research.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Jacqueline Dike: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation; Jarutas Andritsch: Supervision, Writing – Review & Editing.

CONFLICT OF INTERESTS

No conflict of interests was disclosed.

ETHICS STATEMENTS

The dataset is sourced from the NHLBI in which we received consent for. All records in the dataset include anonymized clinical test records from the Framingham Heart Study, which provide critical epidemiological insights into CVD. No data was collected from social media platforms. However, in the event any such data is integrated in future extensions of this research, informed consent will be obtained where applicable, all data will be fully anonymized, and the redistribution policies of the respective platforms will be strictly followed.

DATA AVAILABILITY

The data that support the findings of this study are available from the FHS (https://www.framinghamheartstudy.org/fhs-for-researchers/data-available-overview/). Restrictions apply to the availability of these data, which were used under license for this study.

REFERENCES

[1] WHO, “World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals." World Health Organization. Accessed: Jun. 23, 2025. [Online]. Available: https://digitallibrary.un.org/record/3935247?ln=en&v=pdf.

[2] A. I. F. Poon, and J. J. Y. Sung, “Opening the black box of AI-medicine”, *Journal of Gastroenterology and Hepatology*, vol. 36, no. 3, pp. 581-584, 2021, doi: 10.1111/jgh.15384.

[3] E. Marcus, and J. Teuwen, “Artificial Intelligence and explanation: How, why, and when to explain black boxes”, *European Journal of Radiology*, vol. 173, pp. 111393, 2024, doi: 10.1016/j.ejrad.2024.111393.

[4] A. Sethi, S. Dharmavaram, and S. K. Somasundaram, “Explainable Artificial Intelligence (XAI) approach to heart disease prediction," *2024 3rd International Conference on Artificial Intelligence for Internet of Things (AIIoT)*, Vellore, India, pp. 1-6, 2024, doi: 10.1109/AIIoT58432.2024.10574635.

[5] A. Kilic, “Artificial Intelligence and machine learning in cardiovascular health care”, *The Annals of Thoracic Surgery*, vol. 109, no. 5, pp. 1323-1329, 2020, doi: 10.1016/j.athoracsur.2019.09.042.

[6] A. S. M. Faizal, T. M. Thevarajah, S. M. Khor, and S. Chang, “A review of risk prediction models in cardiovascular disease: Conventional approach vs. Artificial Intelligent approach”, *Computer Methods and Programs in Biomedicine*, vol. 207, pp. 106190, 2021, doi: 10.1016/j.cmpb.2021.106190.

[7] N. Joshi, and T. Dave, “Improved accuracy for heart disease diagnosis using machine learning techniques,” *Journal of Informatics and Web Engineering*, vol. 4, no. 1, 2025, doi: 10.33093/jiwe.2025.4.1.4.

[8] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead", *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206-215, 2019, doi: 10.1038/s42256-019-0048-x.

[9] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang *et al.*, “Interpreting black-box models: A review on Explainable Artificial Intelligence”, *Cognitive Computation*, vol. 16, no. 1, pp. 45-74, 2023, doi: 10.1007/s12559-023-10179-8.

[10] A. Adadi, and M. Berrada, “Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2870052.

[11] C. M. Chituru, S. B. Ho, and I. Chai, “Diabetes risk prediction using shapley additive explanations for feature engineering,” *Journal of Informatics and Web Engineering*, vol. 4, no.2, 2025, doi: 10.33093/jiwe.2025.4.2.2.

[12] Z. Sadeghi, R. Alizadehsani, M. A. CIFCI, S. Kausar, R. Rehman, P. Mahanta *et al.*, “A review of Explainable Artificial Intelligence in healthcare,” *Computers and Electrical Engineering*, vol. 118, 2024, doi: 10.1016/j.compeleceng.2024.109370.

[13] J. You, Y. Guo, J.-J. Kang, H.-F. Wang, M. Yang, J.-F. Feng, J.-T. Yu *et al.*, “Development of machine learning-based models to predict 10-year risk of cardiovascular disease: A prospective cohort study,” *Stroke and Vascular Neurology*, vol. 8, no. 6, 2023, doi: 10.1136/svn-2023-002332.

[14] Z. Li, R. Li, Y. Zhou, L. Rasmy, D. Zhi, P. Zhu *et al.*, “Prediction of brain metastases development in patients with lung cancer by Explainable Artificial Intelligence from electronic health records,” *JCO Clinical Cancer Informatics*, vol. 7, 2023, doi: 10.1200/CCI.22.00141.

[15] N. G. Rezk, S. Alshathri, A. Sayed, E. E.-D. Hemdan, and H. El-Behery, “XAI-Augmented voting ensemble models for heart disease prediction: A SHAP and LIME-based approach,” *Bioengineering*, vol. 11, no. 10, 2024, doi: 10.3390/bioengineering11101016.

[16] G. Petmezas, V. E. Papageorgiou, V. Vassilikos, E. Pagourelias, D. Tachmatzidis, G. Tsaklidis *et al.*, “Enhanced heart failure mortality prediction through model-independent hybrid feature selection and explainable machine learning,” *Journal of Biomedical Informatics*, vol. 163, 2025, doi: 10.1016/j.jbi.2025.104800.

[17] A. M. Alaa, T. Bolton, E. D. Angelantonio, J. H. F. Rudd, and M. v. d. Schaar, “Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK biobank participants,” *PloS one*, vol. 14, no. 5, 2019, doi: 10.1371/journal.pone.0213653.

[18] M. M. Hossain, M. S. Ali, M. M. Ahmed, M. R. H. Rakib, M. A. Kona, S. Afrin *et al.*, "Cardiovascular disease identification using a hybrid CNN-LSTM model with Explainable AI," *Informatics in Medicine Unlocked*, vol. 42, 2023, doi: 10.1016/j.imu.2023.101370.

[19] W. J. She, P. Siriaraya, H. Iwakoshi, N. Kuwahara, and K. Senoo, "An Explainable AI Application (AF'fective) to support monitoring of patients with atrial fibrillation after catheter ablation: Qualitative focus group, design session, and interview study," *JMIR Human Factors*, vol. 12, 2025, doi: 10.2196/65923.

[20] N. Kahouadji, "On the generalizability of machine learning classification algorithms and their application to the Framingham heart study," *Information*, vol. 15, no. 5, 2024, doi: 10.3390/info15050252.

[21] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma *et al.*, "Explainable AI for healthcare 5.0: opportunities and challenges," *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3197671.

[22] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to Explainable Artificial Intelligence in health care," *The Lancet Digital Health*, vol. 3, no. 11, 2021, doi: 10.1016/S2589-7500(21)00208-9.

[23] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Muller, "Causability and Explainability of Artificial Intelligence in medicine," *Wiley interdisciplinary reviews: data mining and knowledge discovery*, vol. 9, no. 4, 2019, doi: 10.1002/widm.1312.

[24] A. Abusitta, M. Q. Li, and B. C. M. Fung, "Survey on Explainable AI: Techniques, challenges and open issues," *Expert Systems with Applications*, vol. 255, 2024, doi: 10.1016/j.eswa.2024.124710.

[25] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and challenges in healthcare," *AI*, vol. 4, no. 3, 2023, doi: 10.3390/ai4030034.

BIOGRAPHIES OF AUTHORS

	Jacqueline Dike is a Master's student in Data Science and Applied AI at Southampton Solent University. She holds a Bachelor's degree in Information Science from Abia State University, Nigeria. Her research interests centre on Deep Learning Models and their practical applications. Jacqueline is passionate about advancing AI solutions through innovative research and data-driven approaches. For academic or professional inquiries, she can be contacted at 0dikej91@solent.ac.uk.
	Jarutas Andritsch, PhD is a lecturer in Computing at the Department of Science and Engineering, Southampton Solent University, UK. She received a doctoral degree (PhD) in computer science from University of Southampton, UK. Her research focuses on data analytics, machine learning, and AI, with a strong interest in using technology to support healthcare systems and improve quality of life. She has worked on several research projects using health and educational data to predict patterns and emerging situations. She is also passionate about enhancing teaching and learning through the use of technology. She can be contacted at jarutas.andritsch@solent.ac.uk.