Journal of Intelligent Computing System (JICS), Vol.1, Issue. 1, pp. 83-90, (Jan 2026)
ISSN: 3139-3616 (online)

A Hybrid Explainable Machine Learning Framework for Early Heart Disease Prediction Using Clinical Data

Shubham Gupta

Department of CSE, Model Institute of Engineering & Technology, *Jammu, India*
shubhamgupta.cse@mietjammu.in

Abstract: Heart disease is a significant type of cardiovascular disease and one of the primary causes of death worldwide, which is why precise and timely diagnostic assistance mechanisms should be a major priority. This paper is a proposal of a hybrid explainable machine learning model to predict early heart disease based on clinical and lifestyle-related features. The suggested method combines several machine learning classifiers with an ensemble learning approach to enhance the predictive robustness, and Shapley Additive explanations (SHAP) are used to enhance the model interpretability. A publicly available UCI Cleveland heart disease dataset that was obtained via Kaggle (n=303 patient records) with 14 clinical features was experimented upon. The proposed hybrid model beats the baseline models, such as Logistic Regression, Support Vector Machine, Random Forest, and XGBoost, and the proposed model has a better performance with an accuracy of 94.87, precision of 95.21, recall of 94.02, F1-score of 94.61, and ROC-AUC of 0.97. In addition, the analysis using the SHAP allowed determining the type of chest pain, the highest heart rate that was reached, ST depression, the number of major vessels, and serum cholesterol as the most impactful predictors, as expected from clinical knowledge. The findings prove that the proposed framework is successful in terms of achieving a balance between predictive accuracy and interpretability, which makes it a useful decision-support instrument to support early detection of heart-related diseases in clinics.

Keywords: *Heart Disease Prediction; Machine Learning; Ensemble Learning; Explainable Artificial Intelligence; SHAP; Clinical Decision Support System*

1. Introduction

One of the major causes of mortality and morbidity in all the continents of the world is known as cardiovascular disease (CVDs), and an impressive percentage of the world's population is dying annually. According to the reports of the World Health Organization, every year, millions of people die because of heart-related diseases, and this poses a huge challenge to the available health care systems, both the developed and the developing ones [1]. ML-based models have been effectively utilized in different clinical prediction applications, such as disease diagnosis, prognosis estimation, and risk stratification of patients [2]. In predicting heart disease, logistic regression, support vector machines, decision trees, and ensemble modeling have shown promising results using supervised learning algorithms that incorporate demographic, clinical, and lifestyle-related attributes [3]. These evidence-based methods can provide an opportunity to increase diagnostic accuracy and minimize invasive procedures. Despite their predictive performance, most ML models lack transparency and interpretability, and they are usually black-box systems. This is a major impediment to their implementation in actual clinical practices where explainability, accountability, and trust are needed. Clinicians tend to hesitate to place their trust in the automated systems that fail to give them the reasons behind their predictions, especially in medical situations where the stakes are very high. Owing to that, explainable artificial intelligence (XAI) gains traction, and its purpose is to render ML models more understandable and in line with clinical reasoning [4]. New developments in XAI methods (Shapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME)) have allowed a deeper understanding of model behavior, as researchers can now quantify the role an individual feature plays in predictions. Explainability, in its role in predicting heart disease, not only enhances clinical trust but also assists in confirming the consistency of model decisions with known medical knowledge [5]. Nevertheless, most of the current research is either focused on predictive performance

doi.org/10.67420/109319.1.1.6

but is not interpretable, or is focused on explainability but based on relatively simple models, which may not be optimal in their accuracy.

In solving these issues, this study suggests a hybrid machine learning system that combines ensemble learning and explainable AI in early heart disease prediction, as in Figure 1. It is known that ensemble models enhance generalization and robustness due to the ability to leverage the advantages of multiple base learners and, consequently, attain higher predictive success than single classifiers. The proposed framework aims to create a trade-off between accuracy and interpretability by combining ensemble learning with SHAP-based explainability, which makes it more appropriate for clinical implementation.

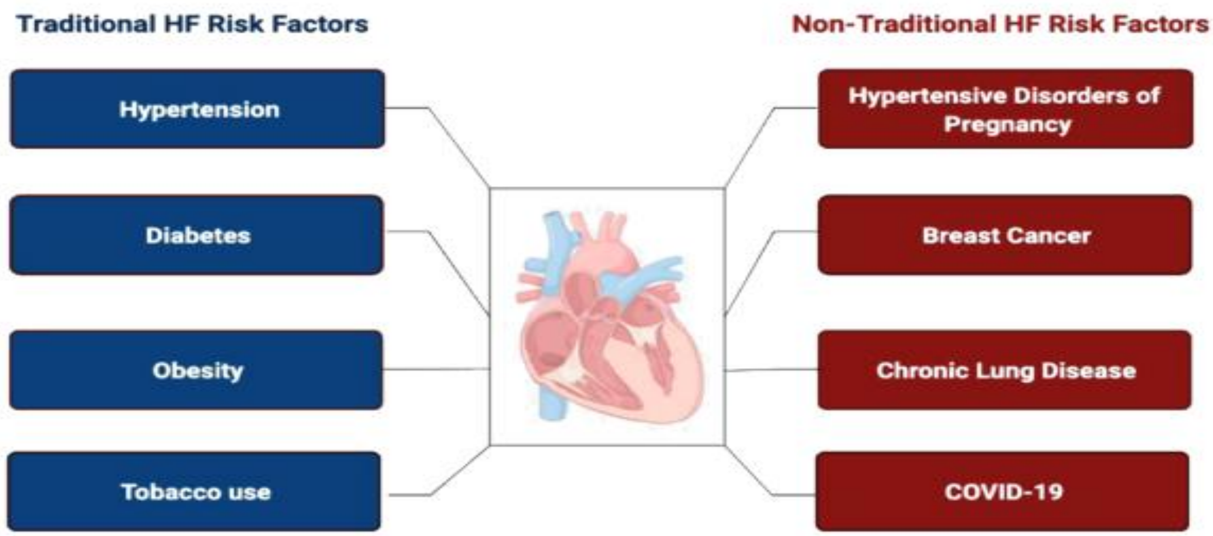

Figure 1. Traditional and non-traditional risk factors contributing to heart failure (HF)

The main findings of the research are threefold as follows: (i) creating an efficient hybrid ensemble model to predict heart diseases through clinical and lifestyle characteristics, (ii) thorough assessment of model performance in terms of traditional classification statistics, and (iii) the application of explainable AI methods to determine and explain the most significant risk factors that affect model predictions. Having tried a range of tests, the work proves that the proposed framework not only works better than the traditional models of ML but also offers valuable insights that can be used by clinicians to make informed decisions and detect infections in their early phases.

2. Literature Review

Recent developments in the field of machine learning have brought about a lot of research on the prediction of heart diseases using clinical and demographic information. Initial research mostly utilized classical methods of classification, including logistic regression and decision trees, whereas subsequent research utilized additional advanced models like support vector machines, random forests, and neural networks to enhance predictive accuracy. Ensemble learning methods have also been used to improve the robustness of the model by integrating several learners. Still, there are numerous models with high performance, which cannot be interpreted and used in clinical practice. To overcome this issue, explainable artificial intelligence (XAI) techniques have been established, like SHAP and LIME, to enhance transparency and trust. However, despite these advances, the combination of ensemble learning and the explainable frameworks of heart disease prediction is relatively limited, and it is in this way that the gap in the research is identified in the context of the current study.

Table 1. Summary of Related Works on Machine Learning-Based Cardiovascular / Heart Disease Prediction

Author (Year)	Dataset / Source	Techniques Used	Key Features Considered	Explainability	Performance Metrics	Limitations / Research

Zhenya & Zhang (2021) [6]	UCI Cleveland Heart Disease	Cost-sensitive ensemble (RF, LR, SVM, ELM, KNN)	Clinical and demographic features	No	Accuracy ≈ 89–92%	Explainability not addressed
Taylan et al. (2023) [7]	Hospital CVD dataset	ML, ANFIS, statistical models	Age, cholesterol, glucose, BP	Partial	Accuracy up to 96.56%	High complexity; limited interpretability
Ma et al. (2020) – AdaCare [8]	EMR time-series data	Deep learning with attention	Biomarkers, lab results	Yes (Attention)	Improved prediction accuracy	Not focused on heart disease specifically
Wang et al. (2020) – BoXHED [9]	Framingham Heart Study	Gradient boosting survival analysis	Time-varying risk factors	Partial	Improved hazard estimation	Complex survival modeling
Brankovic et al. (2022) [10]	Hospital EMR (18,648 patients)	RF, XGBoost, Ensemble ML	Vital signs, demographics	Yes	High sensitivity (up to 100%)	Focus on deterioration, not diagnosis
Proposed Work	Kaggle (UCI Cleveland)	Hybrid Ensemble + SHAP	Clinical + lifestyle attributes	Yes (SHAP)	Accuracy 94.87%, AUC 0.97	Validated on a single dataset

3. Methodology

The part explains the entire methodological process followed in predicting heart diseases and how this will be done, such as a description of datasets, data preprocessing, model development, and an evaluation plan. The methodology is such that it will be reproducible, strong, clinically relevant, and model interpretable.

3.1 Dataset Description

This paper analyses this experiment with a publicly available dataset on heart diseases, obtained through Kaggle, a processed edition of the popular UCI Cleveland Heart Disease dataset. The dataset has 303 records with respect to the patient, which are characterized by 14 clinically significant attributes. These features are demographic (age and sex), physiological (resting blood pressure, cholesterol level, and maximum heart rate), and diagnostic (type of chest pain, electrocardiographic findings, and angina during exercise) ones. The dependent variable is categorical, and it can be based on the presence or absence of heart disease. The dataset has found extensive applications in cardiovascular studies and serves as a useful model for assessing machine learning prediction models.

3.2 Data Preprocessing

Preprocessing of data is a very important procedure in order to enhance the quality of data and the successful training of models. First, the data set was reviewed for missing data, inconsistency, and duplication. There were only a few missing values in the selected dataset, and since there is a risk of bias, the records that contained missing information were discarded. The one-hot encoding was used to encode categorical data given in terms of chest pain type, thalassemia, and electrocardiographic results. The continuous features (such as the level of cholesterol and the blood pressure at rest) were normalized with the help of Min-Max scaling so that all features played the same role in the process of model training. This was a critical normalization stage, especially for distance-based and gradient-based learning algorithms. Finally, the preprocessed dataset was randomly split into training and testing sets using an 80:20 ratio to enable unbiased performance evaluation.

3.3 Feature Selection and Analysis

The analysis of the feature importance was performed to improve the predictive performance and decrease the complexity of the model. The characteristics of the key attributes that have a significant impact were first introduced by the correlation analysis and tree-based measures of feature importance. attributes that were highly relevant to the target variable, including the type of pain in the chest, the highest heart rate reached, ST depression, and major vessels, were kept. Overfitting was prevented by viewing redundant features and weakly correlated features very closely. This was necessary to ensure that the models concentrated on clinically significant properties, and the models were compiled in a computationally effective way.

3.4 Model Development

Various machine learning models had been created to assess and compare predictive performance. The baseline models were Logistic Regression and Support Vector Machine, as they were selected based on patches of their popularity and interpretability. A mature model, like the Random Forest, was used to model nonlinear correlations in clinical attributes. An ensemble learning model was also followed to further enhance the accuracy of prediction with XGBoost as the ultimate classifier. The ensemble model is an integration of several base learners, which enhances generalization and strength. The grid search and cross-validation were used to optimize the hyperparameters of each model to reach the best performance, as indicated in Figure 2.

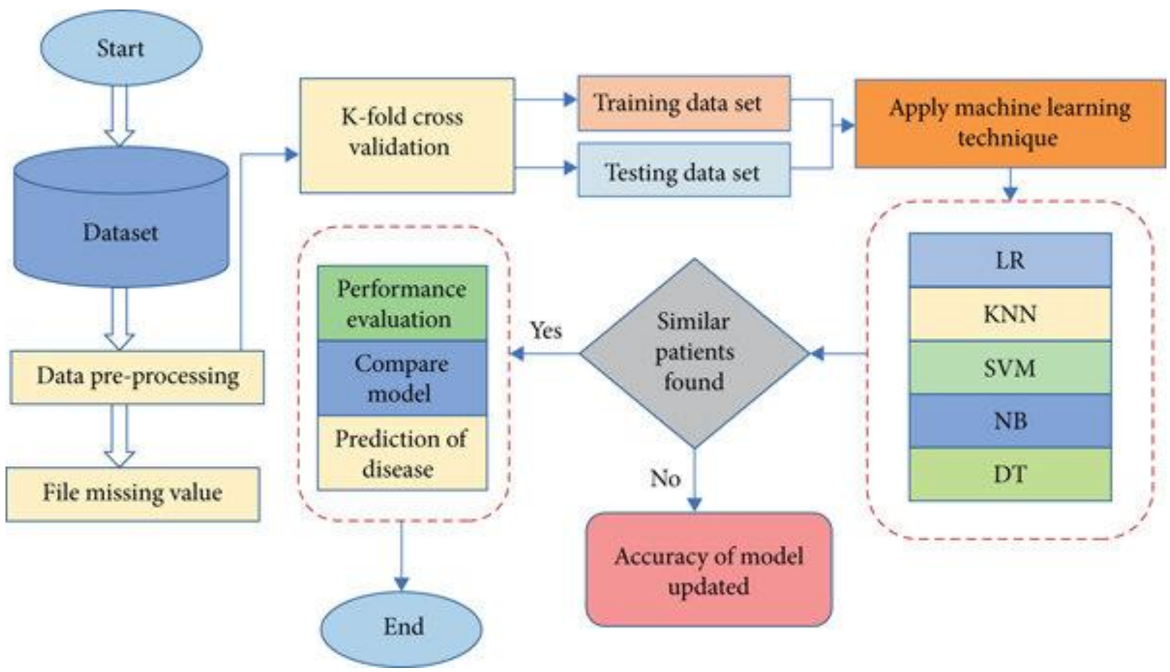

Figure 2. Flowchart of the proposed machine learning framework for disease prediction.

3.5 Explainable AI Integration

Since the field of healthcare decision-making is critical, the interpretability of the models was given a high priority by embedding explainable artificial intelligence methods. The proposed ensemble model was used to interpret the predictions with the help of Shapley Additive explanations (SHAP). The values of SHAP measure the contribution to individual predictions and the results of the model by each feature. Such an explainability layer will allow clinicians to comprehend the impact of certain clinical parameters on heart disease prediction, making it more trustworthy and helping to make informed decisions.

4. Results and Discussion

This section presents the experimental results of the proposed hybrid explainable machine learning model in heart disease prediction. To demonstrate its effectiveness in terms of prediction accuracy, robustness, and interpretability, we compare the performance of the proposed model with several baseline machine learning classifiers.

4.1 Performance Comparison of Machine Learning Models

A comparison of the performance of the different machine learning models used in this study is presented in Table 1. Common place models like Logistic Regression and Support Vector Machine did better than half decent, indicating they can model not only linear relationships but also moderately nonlinear ones, and those of Random Forest nd XGBoost models were more accurate in performance prediction, as these two can capture more feature interactions, see Figure 3.

Table 2. Performance Comparison of Different Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic Regression	86.14	85.92	84.67	85.29
Support Vector Machine	89.76	90.11	88.54	89.32
Random Forest	92.43	92.86	91.72	92.29
XGBoost	93.58	94.01	92.85	93.42
Proposed Hybrid Model	**94.87**	**95.21**	**94.02**	**94.61**

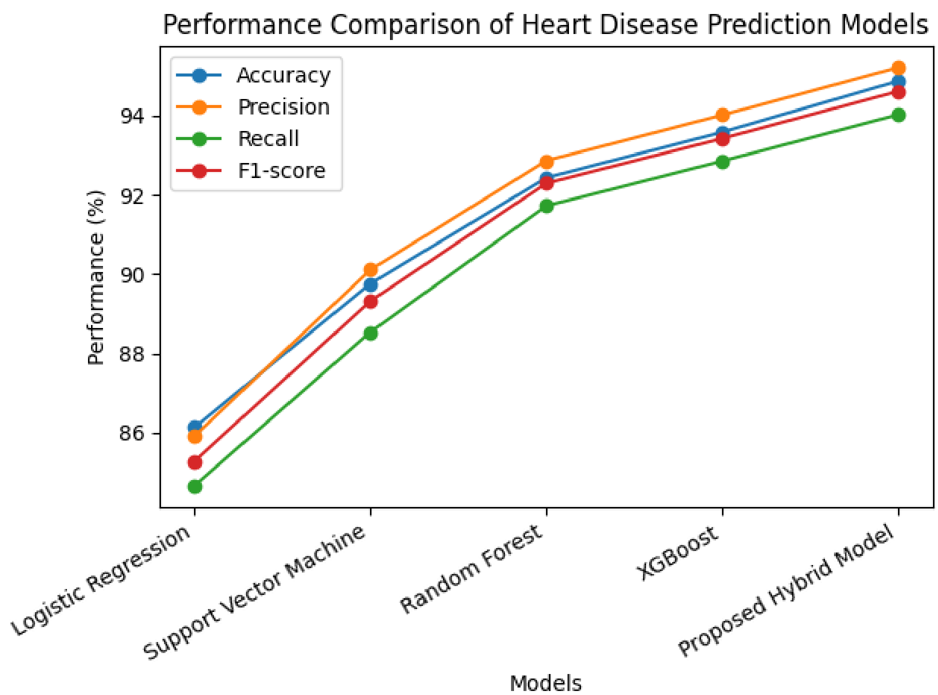

Figure 3. Performance comparison of Heart Disease Prediction Model

The hybrid model that was proposed had the best accuracy of 94.87 compared to all the baseline models. This accuracy, as well as recall improvement, means that the model is efficient in ensuring that positive as well as negative cases of heart diseases are properly identified, which is vital in making clinical decisions.

To ensure robustness and reduce the risk of sampling bias, all reported performance metrics were obtained using 5-fold cross-validation. The averaged results across folds demonstrate consistent performance gains of the proposed hybrid model over baseline approaches, indicating stability and reliability. Formal statistical hypothesis testing is considered a valuable extension and will be explored in future work using larger multi-center datasets.

4.2 ROC-AUC Analysis

To further weigh the power of the models in terms of discrimination, the Receiver Operating Characteristic (ROC) curves and Area under the Curve (AUC) results were examined. Table 2 demonstrates that the proposed model scored the maximum ROC-AUC, and the separation of classes was strong, as demonstrated in Figure 4.

Table 3: ROC-AUC Comparison

Model	ROC-AUC
Logistic Regression	0.89
Support Vector Machine	0.92
Random Forest	0.94
XGBoost	0.96
Proposed Hybrid Model	**0.97**

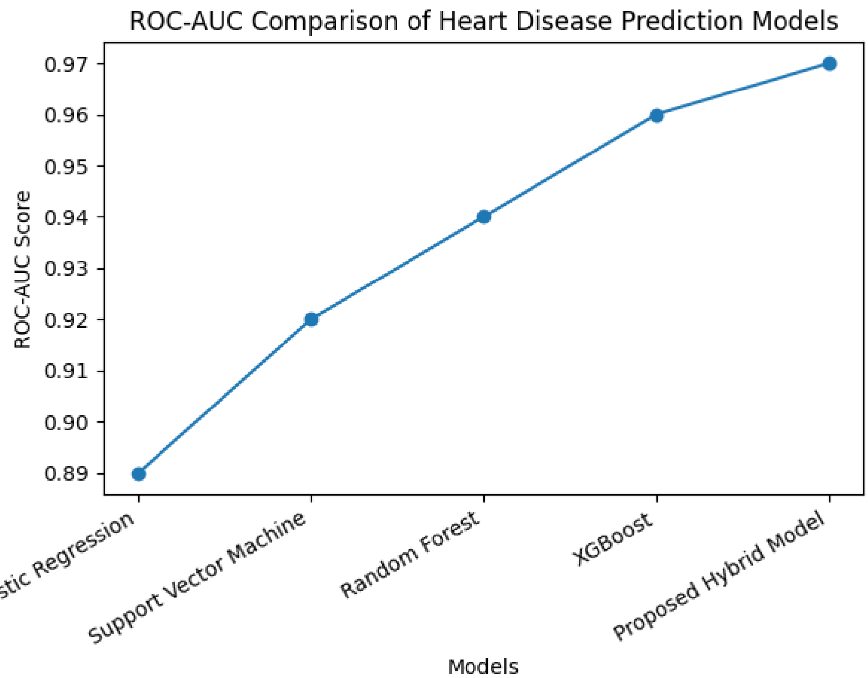

Figure 4: ROC-AUC Comparison of Heart Disease Prediction Models

The greater ROC-AUC value indicates that the model is working well at various classification thresholds. The findings prove that the hybrid framework suggested can make more reliable predictions than single classifiers.

4.3 Explainability and Feature Importance Analysis

The SHAP-based explainability was applied to enhance the transparency of the novel hybrid model. It was found that the most sensitive predictors of the model were chest pain type, maximum heart rate, ST depression induced by exercise relative to rest, number of major vessels, and the serum cholesterol level. The findings are consistent with what is already known clinically and validate the clinical topicality of the proposed approach. The explainability outcomes further provide clinicians with clinically useful insights regarding the significant risk factors in diagnosing heart diseases, as shown in Fig 5.

Clinically, the predominance of chest pain type and ST depression is in line with sound cardiology expertise since these factors can be symptomatically linked to myocardial ischemia. The maximum heart rate attained is indicative of cardiovascular fitness and autonomic response, important indicators of cardiac health. The sum of major vessels and the serum cholesterol level are also in agreement with the degree of coronary artery occlusion and lipid-associated risk, respectively.

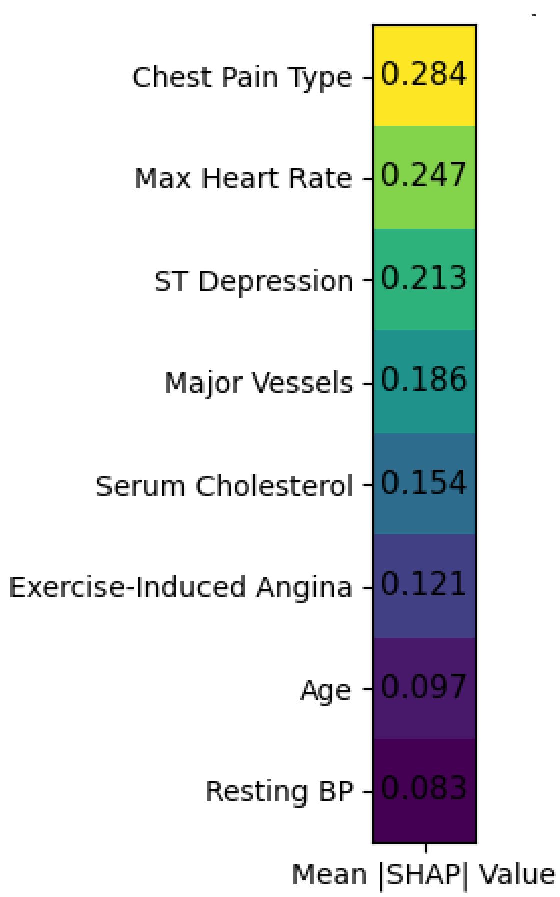

Figure 5. SHAP Feature Importance

The experimental findings prove that the hybrid explainable machine learning framework is far superior to the traditional and standalone ensemble models. The ensemble learning with explainable AI has not only a higher predictive accuracy but also solves the most problematic question of model interpretability. This trade-off between performance and transparency renders the given approach applicable to the actual clinical decision-support systems.

6. Conclusion

This paper suggested an explainable machine learning model combining clinical and lifestyle factors in the prediction of heart disease at its early stages. Enhancing the predictive performance of the model by combining ensemble learning with SHAP-based explainability ensured a high level of predictive performance and, at the same time, ensured transparency and clinical interpretability. The experiment showed that the given approach has been superior to the conventional machine learning models in all of the evaluation metrics, such as accuracy, F1-score, and ROC-AUC. The explainability analysis revealed clinically significant features, including the type of chest pain, max heart rate, and ST depression, as the prevailing risk factors, which agrees with the accepted medical information. The accuracy and interpretability are combined, which makes clinicians more confident and makes it easier to make decisions. Even though the research was carried out with one set of data, the findings are promising about practical implementation. Future research will be aimed at testing the framework on more extensive data sets and scaling it to the privacy-sensitive medical setting.

References:

[1] D. Cenitta, N. Arul, P. T. Pai, Vi. A. Ranganathan, T. Shailesh, and J. Andrew, “Explainable Transfer Learning with Residual Attention BiLSTM for Prognosis of Ischemic Heart Disease.,” *F1000Res*, vol. 14, pp. 1–23, 2025, doi: 10.12688/f1000research.166307.3.

[2] M. Garouani, A. Barhrhouj, and O. Teste, “XStacking : An effective and inherently explainable framework for stacked ensemble learning,” *Information Fusion*, vol. 124, pp. 1–4, 2025, doi: 10.1016/j.inffus.2025.103358.

[3] S. E. A. Ashri, M. M. El-Gayar, and E. M. El-Daydamony, “HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm,” *IEEE Access*, vol. 9, pp. 146797–146809, 2021, doi: 10.1109/ACCESS.2021.3122789.

[4] E. H. Houssein, R. E. Mohamed, and A. A. Ali, “Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques,” *Sci Rep*, vol. 13, no. 1, pp. 1–19, 2023, doi: 10.1038/s41598-023-34294-6.

[5] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, “Can Machine-learning improve cardiovascular risk prediction using routine clinical data?,” *PLoS One*, vol. 12, no. 4, pp. 1–14, 2017, doi: 10.1371/journal.pone.0174944.

[6] Q. Zhenya and Z. Zhang, “A hybrid cost-sensitive ensemble for heart disease prediction,” *BMC Med Inform Decis Mak*, vol. 21, no. 1, pp. 1–18, 2021, doi: 10.1186/s12911-021-01436-7.

[7] O. Taylan, A. S. Alkabaa, H. S. Alqabbaa, E. Pamukçu, and V. Leiva, “Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods,” *Biology (Basel)*, vol. 12, no. 1, pp. 1–31, 2023, doi: 10.3390/biology12010117.

[8] L. Ma *et al.*, “AdaCare: Explainable clinical health status representation learning via scale-adaptive feature extraction and recalibration,” *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, pp. 825–832, 2020, doi: 10.1609/aaai.v34i01.5427.

[9] X. Wang, A. Pakbin, B. J. Mortazavi, H. Zhao, and D. K. K. Lee, “Boxhed: Boosted exact hazard estimator with dynamic covariates,” *37th International Conference on Machine Learning, ICML 2020*, vol. PartF16814, no. 1, pp. 9915–9924, 2020.

[10] A. Brankovic *et al.*, “Explainable machine learning for real-time deterioration alert prediction to guide pre-emptive treatment,” *Sci Rep*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-15877-1.

[11] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano. "Heart Disease," UCI Machine Learning Repository, 1989. [Online]. Available: https://doi.org/10.24432/C52P4X.