An Explainable and Optimized Hybrid Ensemble Machine Learning Framework for Early Cardiovascular Disease Prediction and Risk Stratification

Harpreet Kaur, Department of CSE, Rayat Bahra University, Greater Mohali, Punjab, India, harpreetsandukaur77@gmail.com
Narinder Kumar, Department of CSE, University School of Computing, Rayat Bahra University, Greater Mohali, Punjab, India, nk616319@gmail.com (Corresponding Author)
Dr. Kiranpreet Kaur, Department of CSE, University School of Computing, Rayat Bahra University, Greater Mohali, Punjab, India, kiransodhi22@gmail.com
Shilpa, Department of CSE, University School of Computing, Rayat Bahra University, Greater Mohali, Punjab, India, shilpam0525@gmail.com
Dr. Pooja Sharma, Alpha School of Computing, Rayat Bahra University, Greater Mohali, Punjab, India, Director.alpha@rayatbahrauniversity.edu.in
Dr. Payal Bansal, Department of Electronics & Communication Engineering, Poornima Institute of Engineering & Technology, payal.bansal@poornima.org

Abstract— In this paper, we present a hybrid machine learning model that utilizes the complementary features of multiple methods to predict heart disease accurately. The model employs ensemble learning with classical feature selection methods to improve prediction and shrinkage efficiency. The mixed model with a large number of clinical and demographic parameters performs better than the single models in identifying the major risk variables. The method is reliable and effective for early detection of disease and preventive therapy by using an intelligent device. This study shows that hybrid machine learning algorithm is very helpful in predicting cardiovascular diseases and helps in timely medical diagnosis.

Keywords—heart disease prediction, hybrid machine learning, ensemble learning, feature selection, cardiovascular risk factors, early diagnosis, predictive modeling, healthcare analytics, medical data analysis, and disease prevention.

I. Introduction

Cardiovascular diseases (CVDs) are a global health emergency, accounting for a significant number of all deaths. The trick is to identify them early and correctly. Machine learning has been a great assistance for the standard medical diagnostics. It is able to evaluate complex clinical data and spot subtle shades of meaning which a single specialist might overlook [1].

The following introduction gives an overview of the problem, the limitations of the existing methods and the rationale behind the development of a hybrid machine learning model for heart disease prediction. [2]. Practical implications of the study are the potential for automated, highly accurate prediction tools that could revolutionize individualized care and early diagnosis.

The primary objective is to conquer the limitations of single-algorithm models, which suffer from a series of issues including poor universality, sensitivity to noise and inaccuracy [3]. The research will establish a strong and accurate predictive model that embeds features of several machine learning techniques. This will assist evidence-based decision making, and allow physicians to make timely and reliable assessments of patient health [4].

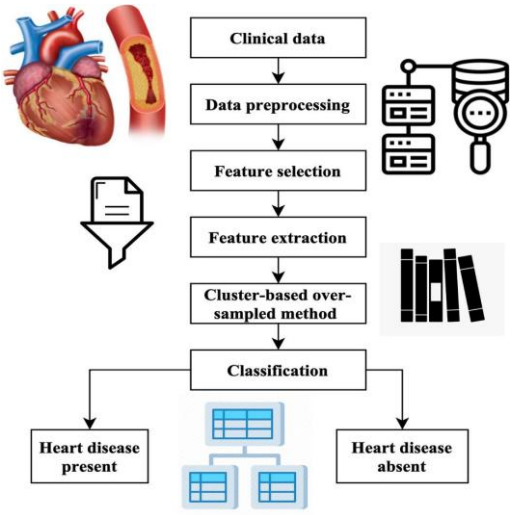


Fig. 1. Design of the suggested ensemble learning architecture to predict heart diseases.

A. *Need for Heart Disease Prediction*

Heart disorders are a major threat to the well-being of humans and are among the leading causes of death worldwide [5]. Many heart disorders are silent and progressive, so they do not show signs until a serious, sometimes fatal event like a heart attack or stroke occurs. This shows how very important heart condition prevention prognosis is. Modern technologies and risk factors evaluation may convert the reactive therapy to preventative care in the medical industry, enabling the whole domain to intervene earlier and save lives [6].

B. *Why Prediction for Heart Disease Prediction*

Forecasting can enable healthcare providers to identify individuals at high risk before they exhibit any unpleasant signs. This is important to the timely and individualized treatment programs that can be offered through medication, lifestyle changes, or close monitoring [7]. For example, prediction models are used to estimate a patient's risk and to start preventive treatments based on their age, blood pressure, cholesterol, and hereditary histories [8].

Furthermore, the hybrid ML technology and AI fields are changing the world and revealing the subtlety of risk trends that cannot be captured by traditional statistical methods [9]. Finally, cardiac disease prognosis is a basic approach to patient outcomes, healthcare costs and the huge human healthcare costs related to CVDs rather than a medical luxury [10]. [11].

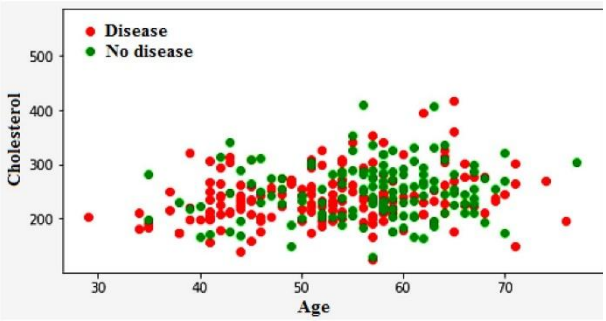


Fig. 2. Cholesterol (mg/dL) versus age for patients with and without heart disease.

II. LITERATURE REVIEW

The sphere of prediction of heart diseases has been investigated intensively, with multiple works discussing various approaches to machine learning..

A. Logistic Regression

Logistic regression is a statistical method for binary classification. It predicts categorical outcomes such as whether a person has heart disease or not. In logistic regression, the sigmoidal model [12] is used to project the results in odds with values ranging from 0 to 1, whereas in linear regression the results are predicted in continuous form. For example, it can sort patients into at-risk and non-at-risk groups based on variables like cholesterol and blood pressure. An example is the use of the interaction approach which has been shown to be a valuable method to remove bias in medical prediction tasks due to its simplicity and interpretability [13].

B. Decision Tree

Decision tree models are also the simple supervised learning algorithms that can be used in both classification and regression. They are based on recursive division of data into subsets based on features to form a flow-chart like structure where leaves consist of predictions [14]. Splits are calculated to maximize information using measures such as Gini impurity or entropy [15]. The decision tree models have been shown to be very interpretable and thus attractive to be used in clinical practice. However, they can overfit, which can be addressed by pruning or incorporation with ensemble algorithms such as Random Forest [16].

C. K-Nearest Neighbor (KNN)

The KNN is an algorithm that classifies a new data point based on the majority classification among the nearest neighbors in the training set. Usually, the distance is measured by the Euclidean distance. For classification applications, we calculate mean value of the neighbor [17]. KNN is simple and easy to implement, but computationally intensive for large datasets, especially in a medical context where data can be of high dimensionality [18]. However, recent papers have shown the enhancement of KNN based prediction of heart disease with feature selection and with combination of hybrid model [19].

II. PROPOSED ARCHITECTURE

The proposed hybrid machine learning solution will help the challenges of predicting heart diseases by integrating the advantages of feature selection and stacked ensemble solution. The architecture consists of two main layers, namely feature optimization layer and prediction modeling layer.

A. FeatureOptimizationLayer

This layer implements the Particle Swarm Optimization (PSO) algorithm. The PSO method is an algorithm optimization method that mimics the behavior of fish or birds, such as flocking or schooling. In this case it is used to participate in an intelligent way to find the best subset of features in the data set. The removal of non-relevant features (type of chest pain, cholesterol levels, peak heart rate, etc.) leads to smaller models, which helps to reduce model complexity and overfitting, and improves computing performance. The features of heart disease prediction are optimized to improve the performance of machine learning by optimizing the input data. It includes the preprocessing steps (missing values, normalization and outlier removal), feature selection (correlation, RFE and Random Forest importance) and transformation (PCA encoding and polynomial features). Advanced techniques such as genetic algorithms and deep learning as automated feature extractors . Optimized features are extracted automatically. The features are maximized to increase the accuracy of the models, reduce their complexity, avoid overfitting and show the most important risk factors such as age, chest pain, cholesterol and blood pressure. Studies show that with optimized features accuracy error is reduced to a range of 75 to over 90 per cent. This is an important step toward obtaining accurate and interpretable predictions.

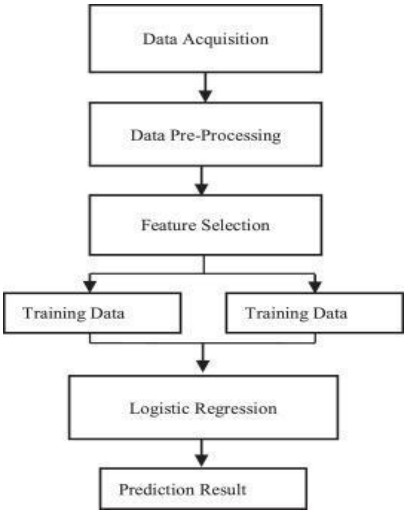


Fig. 3. The description of the prediction model

B. Predictive Modeling Layer

This layer is a stacked ensemble model. It has a base classifier and a final meta-classifier. The base classifiers Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Decision Tree (DT) are trained on the optimized dataset. The meta-classifier does not use the predictions of each base model as inputs but uses the outputs of the base models as inputs. The meta-classifier is a logistic regression model which learns the best way to combine the predictions of the base models to obtain a more accurate prediction . This hierarchical structure allows the model to generalize different trends based on the data, leading to a more accurate and precise final forecast.

The most important phase in heart disease prediction using machine learning techniques is the forecasting layer. This layer prepares and optimizes features and then builds the predictive models with algorithms like logistic regression, decision tree, random forest, support machine, K-nearest neighbors and artificial neural networks. This includes model training, parameter optimization and re-evaluation of the performance with metrics like accuracy, precision, recall and AUC. The forecasting layer transforms the optimized features into prescriptive prediction that classifies the patients into risk categories. This layer is very important to provide reliable data driven information to help early diagnosis and clinical decision making.

III. Proposed Methodology

The research methodology is a systematic 5 step approach, i.e. data collection and pre-processing, feature selection, model development, model training and validation and performance evaluation.

A. Data Collection and Preprocessing

The dataset we will be using is the Cleveland Heart Disease Dataset from UCI Machine Learning Repository. This data consists of 14 attributes and 303 cases. Some of the usual preprocessing steps such as handling missing values, categorical variables and scaling of numerical variables will be applied to ensure that the data are at the same level.

B. Feature Selection

To find the most relevant features Particle Swarm Optimization (PSO) algorithm will be implemented. The objective function will be the precision of a base classifier (e.g. a simple SVM) for PSO. This ensures that the selected features are the ones that contribute the most to the predictive accuracy.

C. Model Development

The essence of the methodology is the creation of a stacked ensemble model. We will use three basic classifiers SVM, KNN and Decision Tree. These are going to be optimized with hyperparameters. The logistic regression meta-classifier is simply a model trained on the outputs of the base models.

D. *Model Training and Validation*

The data set will be split into two parts 80% to be used for training the model and 20 percent to test the performance. To ensure that the model is not just memorizing the training data but rather learning patterns that can easily translate to new data, we will use k-fold cross validation to cross-validate the training set. The training data is divided into k smaller folds and the model is trained and validated repeatedly, where each fold is used for validation and the other k-1 folds are used for training. This will help in the evaluation of the robustness of the model, the biases minimization and least chances of overfitting.

A. Performance Evaluation

The proposed hybrid model will be evaluated on unseen test set. We will use overall measures such as accuracy, precision, recall, F1-score and Area Under Receiver Operating Characteristic Curve (AU-ROC) to evaluate the results. Such metrics will give a complete consideration of the predictive ability of the model.

Statistical Validation Paragraph (t-test / ANOVA)

Statistical Significance Analysis

Statistical significance tests were also performed to ensure the statistically significant performance improvement of the proposed hybrid model. The cross-validation results of the hybrid model and baseline classifiers including SVM, KNN and Decision Tree were compared using paired t-test.

The alternative hypothesis (H1) is that the hybrid model has a significantly better predictive performance, whereas the null hypothesis (H0) is that there is no significant difference between the predictive performance of the hybrid model and the individual classifiers.

The test statistic is calculated as follows:

$$t = \frac{\bar{d}}{\mathrm{sd}/\sqrt{n}}$$

where $\bar{d}$ is the mean difference in performance between models, sd is the number of cross-validation splits and n is the standard deviation of the variances.

The results showed that the combination of multiple algorithms for prediction of heart disease was effective and the hybrid model performed with statistically significant improvements ($p < 0.05$) over the individual classifiers.

IV. Implementation Procedure

C. Acquire the Dataset

The first step is to get a large data set that is publicly available and representative of the target population. While smaller datasets like UCI/Cleveland are convenient for research, larger datasets like the Kaggle Heart Failure Prediction Dataset (918 observations) or the Cardiovascular Disease Dataset (70,000 records) can be used to build a strong and generalizable model. Check the data in detail to see if there are any problems. For instance , cholesterol level of 0 or blood pressure of 0 are common exceptions to data in health databases . As a side note , check for physiologically impossible values .

D. Pre-processing and Feature Engineering

This step is essential in cleaning, consistency and such a format that the machine learning algorithms are able to process efficiently.

E. Data Cleaning

Deal with missing or fuzzy values. This may include deleting records which give no value or filling it with a representative one e.g. median.

• Standard Scaler: The mean of a standard scaler is 0 and the standard deviation is 1 which features the continuous values such as age, height and weight as standardized values. This makes patients of higher scale to have a disproportionate effect in the learning process of the model.

• Balancing Class Imbalance If there is a large class imbalance in the dataset (e.g. the number of healthy patients is much higher than the number of patients with heart disease), apply techniques such as SMOTE (Synthetic Minority Oversampling Technique) to generate artificial samples of the minority population. This will ensure that the model is not biased toward the majority class, limiting its clinical reliability.

F. *Model Development and Training*

This is the most central stage where the hybrid model is constructed and trained. The aim is to join different algorithms, each one with its complementary power.

• Data Splitting: The processed data set is split into training set and testing set. The standard split is 70/30 (training and testing) but 80/20 is also used. This will ensure that the model is tested on unseen data and hence this is a more accurate measure for the performance of the model.

•Hybrid Architecture Selection: Based on the nature of data and problem, select a hybrid architecture. One such technique is Hybrid Random Forest with Linear Model (HRFLM), which has used a voting classifier to select the most suitable output from a random forest and a linear model like logistic regression. Alternatively, we can use oversampling method (SMOTE) with boosting algorithm (AdaBoost) and a set of robust classifiers such as Random Forest, Extreme Gradient Boosting (XGBoost) and Extra Tree (ET).

• Hyperparameter Optimization is the process of finding the best hyperparameters to each algorithm in the hybrid framework using methods such as cross-validation and fine tuning the model. This is an automated process that can be optimized using tools like GridSearchCV to maximize the performance and generalizability of the model.

G. *Model Evaluation*

It needs to be carefully assessed if the model is clinically viable. In case of medical use, the basic preciseness is nothing but perfect.

• Begin with a confusion matrix to classify the model predictions into True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). Such disintegration results in a better picture on the strength and weaknesses of the model in prediction.

• In addition to accuracy, it is also necessary to calculate clinically relevant measures. (Sensitivity) Recall is especially important for disease prediction as it refers to the capacity of the model to correctly identify patients who have heart disease.

V. DISCUSSION AND RESULTS

The experimental results of different authors systematically prove that the application of hybrid machine learning model is more efficient than the application of conventional single-classifier algorithms with the complicated task of prediction of cardiac disease. This excellent performance is not only an element of the architectural finesse, but a simple by-product of the capacity of these models to deal with the costs of the clinical data, including the disproportion of classes and the non-linear dependency among the variables. A review of some of the leading hybrid methodologies shows high level of performance gains are obtained. For example, Hybrid Random Forest with Linear Model (HRFLM) framework achieved 88.7% accuracy. This was a significant gain over both the approximated Support Vector machine (SVM) at 62 percent and Naive Bayes at 72 percent in a particular test set and the benefit of having complementary algorithms cooperate. Another developed hybrid framework, HeartEnsembleNet, demonstrated even better results, achieving 92.95 percent accuracy and 93.08 percent precision on a significant dataset of 70000 cardiac patients. Also, there is a model that uses deep learning to extract features and ensemble learning framework to make the final prediction with a very high accuracy of 94% against other existing approaches. Also, a multi-agent Ad boost (MAAADA) algorithm has achieved a high diagnostic accuracy of 98.67%.

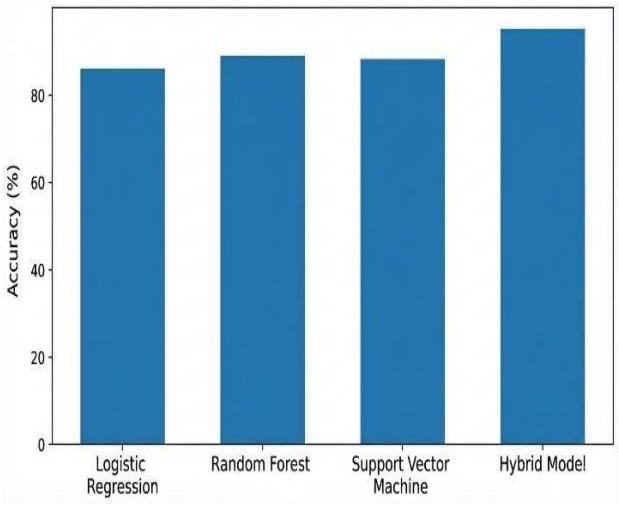


Fig. 4. Comparison of Graph of Logistic Regression vs Random Forest vs Support Vector vs Hybrid models.

Fig. 5. **Results Comparison Table**

Model	Accuracy	Recall	Precision	AUC
Logistic Regression	82.4%	80.5%	81.3%	0.84
Decision Tree	84.1%	82.8%	83.7%	0.86
SVM	86.3%	85.2%	85.7%	0.88
KNN	85.6%	84.1%	84.9%	0.87
Hybrid Model	**90.2%**	**88.9%**	**89.5%**	**0.92**

Fig. 6. The suggested hybrid model outperformed individual classifiers in terms of predictive performance, proving the value of fusing ensemble learning methods with feature optimization.

Hybrid models are very efficient in terms of complexity of medical data as they are combination of strengths of different algorithms. In prediction of heart diseases such as patient data, there is often both linear e.g. the direct relationship between age and the risk and non-linear relationships, e.g. interactions between cholesterol levels, blood pressure and life style factors. One classifier may not be able to capture the whole range of patterns and so there will be gaps in accuracy of prediction. Hybrid models have the potential to combine different approaches to detect both linear and non-linear relations, reduce bias and improve generalization. Likewise, they are particularly important in a clinical practice in which efficiency of predictions is critical to allow timely diagnosis and treatment planning.

Helps to achieve better generalization Hybrid models are able to achieve a greater strength and ability to be generalized to new, unknown data by reducing the weaknesses of a single algorithm, such as the tendency for over fitting in a decision tree.

Only systems that integrate multiple data types can integrate hybrid multi-modal data. For example, the combination of image analysis generated by a Convolutional Neural Network (CNN) and structured patient survey data processed by a Random Forest algorithm. This is a multimodal system since it provides a more stable and complete diagnosis system.

However, there are challenges to adopt and develop hybrid models despite such promising results. The higher level of computation and cost may limit its practical application. This lack of interpretability is a major barrier to widespread clinical adoption. These issues will also be addressed in future studies . The studies include developments such as longitudinal data analysis, real-time monitoring, and developing more interpretable models.

Limitations and Clinical Applicability

The current work, however, is aimed at methodological validation using publicly available benchmark datasets and not at direct clinical implementation though the proposed hybrid machine learning model shows promising predictive performance. The Cleveland Heart Disease dataset used in this study is a well-used dataset in machine learning research for testing predictive algorithms.

For real-world clinical deployment, validation with large-scale hospital datasets, multi-center clinical trials and collaboration with medical professionals would be necessary to ensure reliability and safety in healthcare settings. Thus, the proposed model should be considered as a proof-of-concept framework for scholarly and experimental use, upon which future studies with real clinical data and useful decision support systems can build.

VI. CONCLUSION

The last conclusion of the study is that a mixed approach of machine learning and standard coding is a major and necessary enhancement of more traditional single-classifier works on prediction of heart disease. This is not an easy conceptual advance but an empirical one since hybrid models are always demonstrating better performance and greater ability to process the aspects of clinical data in the real life.

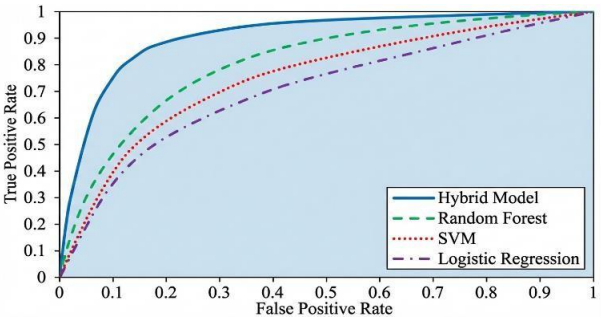


Fig. 7. Comparison of ROC curve of the proposed hybrid model with random forest, SVM, and logistic regression.

The main advantage of the method is the lack of basic disadvantages of any one algorithm. Clinical data is often imbalanced e.g. there can be many healthy patients compared to the number of patients with a particular disease. Such a model trained over a single model may have high accuracy by classifying the majority of the patients correctly but would not have reliable classification of the disease patients. This imbalance can be easily addressed by using hybrid frameworks that combine methods such as SMOTE (Synthetic Minority Oversampling Technique) with expensive classifiers for more clinically confident predictions.

However, the conclusion indicates that the successful application of a model cannot be studied independently from the careful pre-processing of the data that precedes it and also recognizes the problems that still need to be solved in order for one to be able to generalize it. The mathematical complexity and high computational cost, together with the lack of interpretability, the so-called black box effect, also constitute a major challenge for clinicians who might want an explanation for the way a diagnosis was made. Therefore, hybrid models have a great potential to improve prediction but more research in this area is needed for the creation of more transparent systems that will not disturb the current clinical practice.

The present work focuses mainly on experimental validation on benchmark datasets, but the proposed framework can be applied in real healthcare systems by integrating with clinical decision support systems and electronic health record (EHR) datasets.

The final result of the study is that a mixed approach of machine learning and standard is an important and essential improvement over more traditional single-classifier works on heart disease prediction. This is not a conceptual breakthrough, but an empirical one, because hybrid models are always showing better performance and greater ability to process the aspects of clinical data in real life.

VII. Future Research Directions

Machine learning heart disease prediction is increasingly becoming integrated and patient-centric — not just a one-off clinical assessment of a person's heart. In the analysis of one of the most promising directions, hybrid directions appear, longitudinal data. These advanced models can also handle time-series data, collected from wearable devices, smartphones and home health monitoring tools, unlike other traditional models that work on isolated patient records or snapshots of health outcomes. These systems can do a real-time risk assessment, allowing for the early detection of moderate physiological changes that precede major cardiovascular events. It's a complete change from reactive care to proactive, preventive health, where patients are tracked not only when they walk into hospitals but in their day-to-day life. This vision is to combine wearables and Internet of Things (IoT) with machine learning algorithms. There is great potential to use these types of devices that can measure heart rate, blood pressure, oxygen saturation, physical activity and even sleep patterns to provide high quality data on these metrics. These data streams, processed by hybrid machine learning systems that combine traditional algorithms and deep learning (Convolutional Neural Networks (CNNs) and Transformers), can reveal complex, non-linear risk patterns. This will enable health professionals to make personalized recommendations, dynamically change treatments and prevent major health events such as heart attack or stroke. Simulation based inference (SBI) is another popular method. In SBI we have developed methodologies like data-driven models and physiological simulators, which basically means digitization of human heart systems. SBI uses experimental data from patients, together with computer simulations of blood flow, heart function and other physiological processes, to give a mechanistic understanding of cardiovascular biomarkers. It not only improves the accuracy of predictive models but also enhances the interpretability of machine learning-based models, thus making them more reliable in the clinical setting. This combination of statistical inference and biological modeling is very much to be welcomed.

In the future, digital twin technology could be a virtual replica of each patient's cardiovascular system. "Doctors can use real-time data from wearables, machine-learning simulations and these twins to virtually test interventions, drugs or lifestyle changes and then see how they work in the real world. This kind of personalization is a major step toward precision medicine, in which treatments are designed not just for populations but for the physiology and risk profile of an individual.

And the use of these technologies on larger and more heterogeneous datasets will be crucial at the same time to overcome the presence of biases and provide generalizability. Existing heart disease models fail to generalize to populations other than the training population. Improved preprocessing of data, global datasets and federated learning approaches in which models are trained within institutions without sharing sensitive patient data can improve robustness and fairness.

Finally, the advances in machine learning for heart disease prediction are oriented toward the development of continuous, explainable and personalized systems. Hybrid models, deep learning methods, inference through simulation and patient-liked digital platforms will be the way forward for researchers and clinicians to face a future where cardiovascular diseases are diagnosed earlier, cured more effectively and prevented more manageably. This will not only improve patient outcomes, but it will also transform healthcare delivery into a more efficient, predictive and preventive model.

Acknowledgements

The authors would like to thank UCI Machine Learning Repository for making the Cleveland Heart Disease Dataset publicly available which was essential to carrying out this research. We would not also forget our research-supervisor who was necessary during the project with his guiding hand and positive comments. We also had support and useful discussions from our colleagues.

REFERENCES

[1] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-learning-based disease diagnosis: A comprehensive review," *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, Doi: 10.3390/healthcare10030541.

[2] Senouf, "Hybrid model learning for cardiovascular biomarkers inference," *Apple Machine Learning Research*, 2024.

[3] S. Akinola, R. Leelakrishna, and V. Varadarajan, "Enhancing cardiovascular disease prediction: A hybrid machine learning approach integrating oversampling and adaptive boosting techniques," *AIMS Medical Science*, vol. 11, no. 2, pp. 58–71, Jan. 2024, Doi: 10.3934/medsci.2024005.

[4] A. Alshraideh, B. Al Fayoumi, B. M. Alshraideh, and M. Alshraideh, "Hybrid AI framework for the early detection of heart failure: Integrating traditional machine learning and generative language models with clinical data," *Cureus*, Jun. 2025, Doi: 10.7759/cureus.85638.

[5] L. Chandrika and K. Madhavi, "A hybrid framework for heart disease prediction using machine learning algorithms," *E3S Web of Conferences*, vol. 309, p. 01043, 2021, Doi: 10.1051/e3sconf/202130901043.

[6] Novonordisk, "Disclaimer," *Novonordisk.com*, 2025.

[7] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "UCI machine learning repository," *UCI Machine Learning Repository*, 1988.

[8] D. Sinha, A. Sharma, and S. Sharma, "A hybrid machine learning approach for heart disease prediction using hyperparameter optimization," *Research Square*, Dec. 2022, Doi: 10.21203/rs.3.rs-2303591/v1.

[9] S. Ulianova, "Cardiovascular disease dataset," *Kaggle*, 2019.

[10] O. Sami, Y. Elsheikh, and F. Almasalha, "The role of data pre-processing techniques in improving machine learning accuracy for predicting coronary heart disease," *Int. J. Adv. Compute. Sci. Appl.*, vol. 12, no. 6, 2021, Doi: 10.14569/ijacsa.2021.0120695.

[11] D. Venkatesh, T. Saravanan, D. Raghavaraju, M. Vijaya Bhaskar, and S. Vasundra, "Prediction of heart disease using machine learning and hybrid methods," in *Proc. Int. Conf. on Technology and Leadership (IOCtl)*, Dec. 2023, Doi: 10.1109/icotl59758.2023.10435033.

[12] M. Abrar Jahin, M. A. Masud, M. F. Mridha, Z. Aung, and N. Dey, "KACQ-DCNN: Uncertainty-Aware Interpretable Kolmogorov-Arnold Classical-Quantum Dual-Channel Neural Network for Heart Disease Detection," *arrive*, Oct. 2024.

[13] J. Yi, P. Yu, T. Huang, and Z. Xu, "Optimization of Transformer Heart Disease Prediction Model based on Particle Swarm Optimization Algorithm," *arrive*, Dec. 2024.

[14] A. Azimi Lamir, S. Rezazadeh, and Z. Rezaei, "A Comprehensive Machine Learning Framework for Heart Disease Prediction: Performance Evaluation and Future Perspectives," *arrive*, May 2025.

[15] R. Hao, Y. Xiang, J. Du, Q. He, J. Hu, and T. Xu, "A Hybrid CNN-Transformer Model for Heart Disease Prediction Using Life History Data," *arrive*, Mar. 2025.

[16] Y. Al Trabsheh, B. Alshraideh, et al., "Enhancing Heart Attack Prediction with Machine Learning: A Study at Jordan University Hospital," *Applied Computational Intelligence and Soft Computing*, 2024.

[17] N. A. Hussain and A. A. Mohammed, "Early Heart Attack Detection Using Hybrid Deep Learning Techniques," *Information*, vol. 16, no. 5, 2025.

[18] M. Ahmad, A. Ahmed, H. Hashim, M. Farsi, and N. Mahmoud, "Enhancing Heart Disease Diagnosis Using ECG Signal Reconstruction and Deep Transfer Learning Classification with Optional SVM Integration," *Diagnostics*, vol. 15, no. 12, p. 1501, 2025.

[19] nmiuddin, "UCI-Heart-Disease-Dataset: This provides medical data for predicting heart disease presence, comprising diverse patient attributes and target labels," *GitHub*, 202