information

MDPI

Article

Intelligent Hybrid Modeling for Heart Disease Prediction

Mona Almutairi [1] and Samia Dardouri [1,2,*]

[1] Department of Computer Science, College of Computing and Information Technology, Shaqra University, Shaqra 11911, Saudi Arabia; s446500713@std.su.edu.sa
[2] InnoV'COM Laboratory-Sup'Com, University of Carthage, Ariana 2083, Tunisia
* Correspondence: s.dardouri@su.edu.sa

Abstract

Background: Heart disease continues to be one of the foremost causes of mortality worldwide, emphasizing the urgent need for reliable and early diagnostic tools. Accurate prediction methods can support timely interventions and improve patient outcomes. **Methods:** This study presents the development and comparative evaluation of multiple machine learning models for heart disease prediction using a structured clinical dataset. Algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost, and Deep Neural Networks were implemented. Additionally, a hybrid ensemble model combining XGBoost and SVM was proposed. Models were evaluated using key performance metrics including accuracy, precision, recall, and F1-score. **Results:** Among all models, the proposed hybrid model demonstrated the best performance, achieving an accuracy of 89.3%, a precision of 0.90, recall of 0.91, and an F1-score of 0.905, and outperforming all individual classifiers. These results highlight the benefits of combining complementary algorithms for improved generalization and diagnostic reliability. **Conclusions:** The findings underscore the effectiveness of ensemble and deep learning techniques in addressing key challenges such as data imbalance, feature selection, and model interpretability. The proposed hybrid model shows significant potential as a clinical decision-support tool, contributing to enhanced diagnostic accuracy and supporting medical professionals in real-world settings.

Keywords: heart disease; machine learning; neural networks; predictive modeling early detection

check for updates

Received: 9 September 2025
Revised: 28 September 2025
Accepted: 3 October 2025
Published: 7 October 2025

Citation: Almutairi, M.; Dardouri, S. Intelligent Hybrid Modeling for Heart Disease Prediction. *Information* **2025**, *16*, 869. https://doi.org/10.3390/info16100869

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cardiovascular diseases (CVDs), including heart disease, are the leading cause of death in Saudi Arabia, accounting for approximately 37% to 42% of all mortalities, according to the World Health Organization (WHO) and the Saudi Ministry of Health (MOH). Globally, heart disease remains one of the most critical public health challenges, causing millions of deaths annually and affecting individuals in diverse ways. It encompasses a range of conditions, including coronary artery disease, valvular disorders, and cardiomyopathy all of which can lead to heart failure if not detected and managed promptly [1].

Several risk factors contribute significantly to the development of heart disease, such as high cholesterol, obesity, hypertension, physical inactivity, and genetic predispositions. These risks are particularly pronounced in Saudi Arabia due to rapid urbanization and changing lifestyle patterns. Furthermore, the symptoms of heart disease often overlap with other medical conditions, complicating early diagnosis and intervention.

In recent years, the use of machine learning (ML) and artificial intelligence (AI) in healthcare has gained substantial momentum. These technologies allow for the analysis of large, complex clinical datasets, offering new possibilities for early detection and accurate prediction of heart disease. By leveraging data such as clinical indicators, lifestyle habits, and genomic markers, machine learning models can enhance diagnostic precision and support more informed clinical decisions.

In this study, we implemented and evaluated several machine learning models to predict the risk of heart disease. The algorithms explored include Logistic Regression [2], Random Forest [3], Decision Tree [4], Support Vector Machine (SVM), Gaussian Naive Bayes, and a neural network-based approach. Each model was assessed using accuracy and standard deviation as the primary evaluation metrics.

These findings underscore the potential of machine learning as a powerful tool for the early detection of heart disease. However, they also highlight the need for high-quality, diverse datasets and robust preprocessing techniques to enhance model performance.

As illustrated in Figure 1, AI-aided diagnostic systems are increasingly being used to support clinical decisions in identifying conditions such as atrial fibrillation, valvular heart disease, heart failure, congenital heart disease, cardiomyopathy, and coronary artery disease. These technologies not only enhance the accuracy of diagnosis but also allow for early detection, timely intervention, and optimized treatment pathways. The diversity of CVD types that can be analyzed through AI models highlights the potential for these systems to serve as comprehensive tools in cardiology practice.

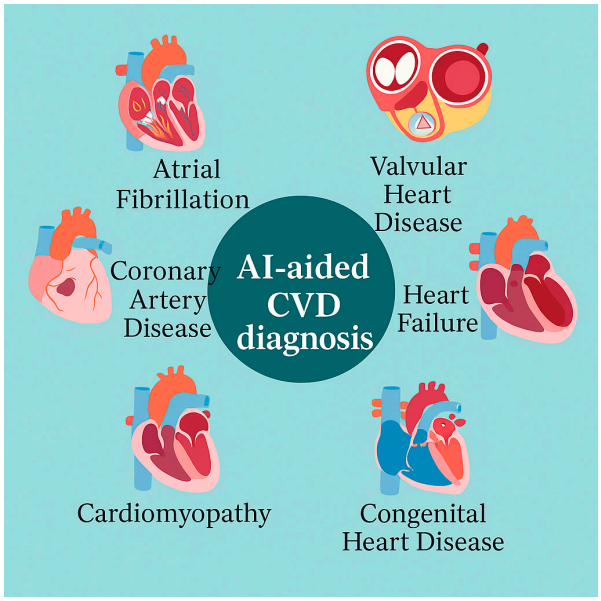


Figure 1. AI-aided diagnosis of various cardiovascular diseases, including atrial fibrillation, coronary artery disease, cardiomyopathy, congenital heart disease, heart failure, and valvular heart disease.

2. Related Work

Heart disease prediction using machine learning has become one of the most studied topics in recent years. Many researchers have tried different techniques to extract the most relevant features and obtain a more accurate diagnosis. Heart disease prediction using machine learning (ML) and artificial intelligence (AI) has gained considerable attention due to the increasing burden of cardiovascular diseases globally [1]. Various models and algorithms have been developed to enhance early diagnosis and improve clinical decision-making.

Traditional models such as Random Forest [2] and Decision Trees [3] have been widely used due to their simplicity and interpretability. Recent studies have leveraged more sophisticated approaches, including neural networks [5], ensemble methods [6,7], and hybrid models [8,9]. For example, Khan et al. proposed a feature selection mechanism using a modified Artificial Bee Colony (M-ABC) and KNN, achieving high prediction accuracy [10].

Martin-Isla et al. presented a comprehensive review of ML in cardiac imaging, highlighting its role in diagnosis and risk stratification [4]. Boukhatem et al. utilized a variety of ML algorithms in their heart disease prediction framework and achieved promising results in terms of accuracy and robustness [5,11].

AI is increasingly being leveraged to enhance the detection and diagnosis of heart-related conditions through advanced machine learning and data-driven approaches. Udoy and Hassan [12] provide a comprehensive review of AI-driven technologies in heart failure diagnosis, emphasizing the role of personalized healthcare solutions, wearable sensors, and predictive algorithms in improving clinical outcomes. Complementing this, Akter et al. [13] explore the use of five distinct datasets to evaluate various machine learning models for heart disease prediction, highlighting the effectiveness of ensemble methods and the importance of diverse data sources for robust model performance.

Comparative analyses, such as those by Patidar et al. [14] and Chakraborty et al. [15], explored the performance of different ML classifiers (e.g., SVM, logistic regression, MLP) on heart disease datasets, confirming the importance of model selection and data preprocessing. Further, Teoh et al. [16] introduced a hybrid evolutionary algorithm to optimize feature selection, enhancing model performance.

Advancements in AI-driven decision support systems have also been significant. Elvas et al. [17] demonstrated a clinical decision support system for early cardiac event detection, while Garza-Frias et al. [18] developed an AI model based on chest radiographs to detect heart failure.

Data mining and machine learning continue to play a pivotal role in enhancing the accuracy of heart disease prediction through the analysis of complex medical datasets. Chourasia and Pal [19] integrate data mining techniques with machine learning algorithms to improve diagnostic precision, demonstrating how meaningful patterns can be extracted from large-scale health data. Similarly, Parthiban and Srivatsa [20] concentrate on diagnosing heart disease in diabetic patients, employing machine learning approaches to manage the heightened risk and diagnostic complexity within this specific population group.

Ethical and data quality considerations are also emerging. Alwakid et al. [21] proposed an ethically aware ML framework for cardiovascular diagnosis, and Singh et al. [22] provided a scoping review on personalized cardiovascular AI risk assessment.

Several studies focused on optimization techniques. Saranya and Pravin [23] used grid search and hyperparameter tuning for feature selection, and Cao et al. [8] applied PSO-XGBoost for cardiovascular disease prediction. Similarly, Bilal et al. [24] proposed a hybrid AI model that significantly improved predictive accuracy.

The integration of these models into clinical settings continues to evolve, as demonstrated by Ogunpola et al. [25], who designed ML-based cardiovascular screening tools, and Al-Mahdi et al. [9], who used ensemble deep learning for improved heart disease detection.

Overall, these studies show that while many methods are promising, challenges like data preprocessing, feature selection and model generalization across diverse populations still remain unsolved, calling for further improvements in predictive modeling of heart disease. Recent advances in semantic modeling and machine learning have significantly contributed to the development of intelligent health prediction systems. Wang [26] emphasized the role of semantic representation in natural language processing for improving

understanding of clinical text, while Kero and Demissie [27] reviewed ontology-driven machine learning approaches that enhance the interpretability of predictive models. Knowledge graph embeddings, as discussed by Ge et al. [28], support relational reasoning between patient attributes, enabling more robust cardiovascular risk assessments. In the context of data integration, deep learning models for schema matching such as SMAT [29] and AdNeV [30] facilitate harmonization of disparate medical datasets. Duan et al. [31] introduced a graph-based method for zero-shot learning, enabling accurate predictions even in underrepresented cardiac conditions. Additionally, Hoseinzade and Wang [32] applied graph neural networks to semantic type detection, automating feature annotation in complex medical datasets. Finally, Xue et al. [33] proposed a reinforcement learning framework for ontology alignment, offering scalable solutions for unifying clinical terminologies across healthcare systems. Collectively, these methods enhance the scalability, interoperability, and generalizability of AI-driven heart disease prediction models. Recent advancements in artificial intelligence have significantly influenced cardiovascular diagnostics and predictive modeling. Xue et al. [33] proposed a deep reinforcement learning-based ontology meta-matching technique, which improves semantic integration across heterogeneous data sources—an essential step for enhancing interoperability in medical informatics and enabling more robust model training. In the domain of predictive modeling, Zhenya and Zhang [34,35] introduced a hybrid cost-sensitive ensemble approach tailored for heart disease prediction, which effectively addresses the common challenge of class imbalance by minimizing misclassification costs.

This study underscores the effectiveness of ensemble and deep learning approaches in medical diagnosis, particularly in predicting heart disease. It addresses key challenges such as data imbalance, feature selection, and model interpretability factors that often hinder the deployment of machine learning in clinical practice. The findings demonstrate that intelligent hybrid models, which combine the strengths of multiple algorithms, can significantly improve predictive accuracy and robustness. Such models not only enhance diagnostic performance but also provide valuable decision support for clinicians, ultimately contributing to more informed and timely interventions.

Although numerous studies have applied traditional machine learning models, ensemble techniques, and deep learning approaches for heart disease prediction, several challenges remain unresolved. Many existing works are limited by small or homogeneous datasets, lack robust handling of class imbalance, and provide limited insight into model interpretability, which is critical for clinical adoption. Furthermore, only a few studies have explored hybrid ensemble approaches that effectively integrate complementary algorithms to balance generalization and diagnostic precision. To address these gaps, this study makes the following key contributions: (1) we propose a novel hybrid ensemble model that integrates XGBoost and SVM to leverage their complementary strengths; (2) we present a comprehensive comparative analysis with traditional and deep learning baselines to demonstrate the superiority of the proposed method; (3) we apply interpretability techniques to provide insights into model decision-making and highlight clinically relevant features; and (4) we design and implement a complete system architecture that supports reproducibility and practical deployment.

3. Methods

3.1. Dataset Description

This study uses a publicly available dataset titled “Heart Failure Prediction” by *Federico Soriano*, hosted on Kaggle. The dataset consists of 918 anonymized patient records and is structured for a binary classification task, aiming to predict the presence or absence of heart disease based on various clinical and demographic features.

The dataset comprises 12 input features and 1 target variable. It includes both demographic information such as *Age* and *Sex* and clinical indicators, such as blood pressure, cholesterol, chest pain type, and electrocardiogram (ECG) readings.

Figure 2 displays the distribution of the target variable (*HeartDisease*) in the dataset. It highlights a noticeable class imbalance, with a substantially higher number of samples labeled as Class 1 (patients with heart disease) compared to Class 0 (patients without heart disease). Such imbalance can bias machine learning models toward the majority class, leading to reduced sensitivity in detecting the minority class. Addressing this imbalance through techniques like resampling, class weighting, or synthetic data generation (e.g., SMOTE) is essential to ensure robust and fair model performance.

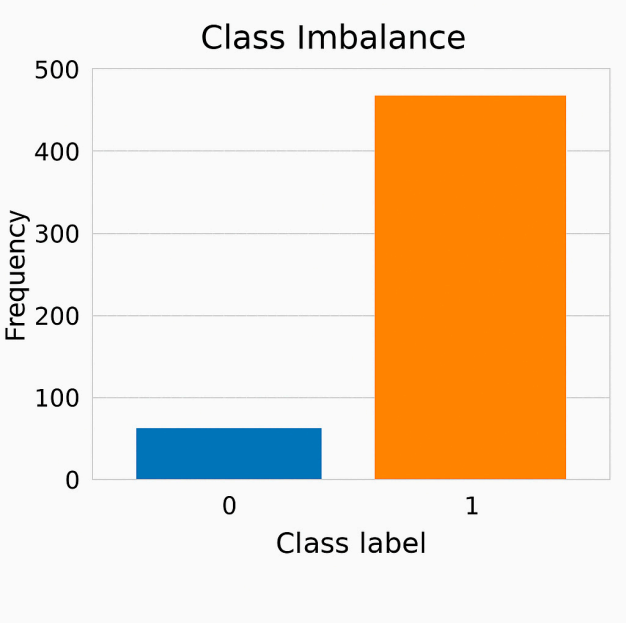


Figure 2. Class imbalance in the dataset.

3.2. Model Development

The model development process began with extensive data preprocessing, including handling missing values, encoding categorical features, and applying feature scaling using StandardScaler. The cleaned dataset was then divided into training and testing sets with a stratified 80:20 ratio to preserve class balance. Several machine learning models were developed and trained to evaluate their predictive performance. The baseline models included Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost. Each model was trained using default hyperparameters initially, followed by tuning to enhance performance. For deep learning, a fully connected neural network (Multilayer Perceptron) was built with dropout layers and batch normalization to reduce overfitting and accelerate convergence.

In addition, a proposed hybrid ensemble model was developed by combining XGBoost and SVM using a soft voting strategy. This ensemble leveraged the strength of gradient boosting (XGBoost) in handling structured data and the capability of SVM in finding optimal decision boundaries, resulting in improved generalization. The performance of all models was evaluated based on accuracy, precision, recall, F1-score, and confusion matrix analysis. Throughout the development, early stopping and learning rate scheduling techniques were employed to optimize deep learning models, while repeated experimentation was conducted to compare models and finalize the proposed one. The proposed hybrid model achieved the highest accuracy of 89.3%, confirming the effectiveness of ensemble learning in medical prediction tasks.

3.3. Mathematical Formulation

To provide a deeper understanding of the proposed model architecture, the mathematical representation of the key components is described as follows:

3.3.1. Support Vector Machine (SVM)

SVM aims to find the optimal hyperplane that maximizes the margin between the classes. The decision boundary is defined as:

$$f(x) = \text{sign}(w^t\, x + b) \tag{1}$$

where

- w is the weight vector
- x is the input feature vector
- b is the bias term

The optimization objective becomes:

$$\text{minimize } (1/2)\ ||w||^2 \text{ subject to } y_i(w^t\, x_i + b) \geq 1 \tag{2}$$

3.3.2. XGBoost Classifier

XGBoost constructs an ensemble of weak learners (typically decision trees). At each iteration t, the model is updated as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{3}$$

The objective function is minimized as:

$$L^{(t)} = \sum_n l(y_i, \hat{y}_i^{(t)}) + \sum_k \Omega(f_k) \tag{4}$$

where

- l is the loss function (e.g., logistic loss)
- Ω is the regularization term for controlling model complexity

3.3.3. Soft Voting Ensemble

The final prediction $\hat{y}$ of the hybrid model is computed by averaging the predicted probabilities from both classifiers:

$$\hat{y} = \text{argmax_c } [\alpha \cdot \text{P_XGBoost(c)} + (1 - \alpha) \cdot \text{P_SVM(c)}] \tag{5}$$

where

- P_XGBoost(c) and P_SVM(c) are the predicted probabilities for class c
- $\alpha \in [0, 1]$ is the weighting parameter (e.g., 0.5 for equal weight)

3.4. Proposed Hybrid Model

To further enhance prediction performance, a hybrid model was proposed by combining the strengths of two powerful classifiers: XGBoost and Support Vector Machine (SVM). The motivation behind this integration was to leverage the ensemble learning capabilities of XGBoost particularly its efficiency with structured data and the robust classification boundaries provided by SVM.

In the proposed architecture, illustrated in Figure 3, both classifiers were trained independently using the same input features. The final prediction was obtained using a soft

voting strategy, which calculates the average predicted probabilities from each model and selects the class with the highest combined probability. This ensemble mechanism helps to reduce bias and variance simultaneously, producing more stable and accurate predictions.

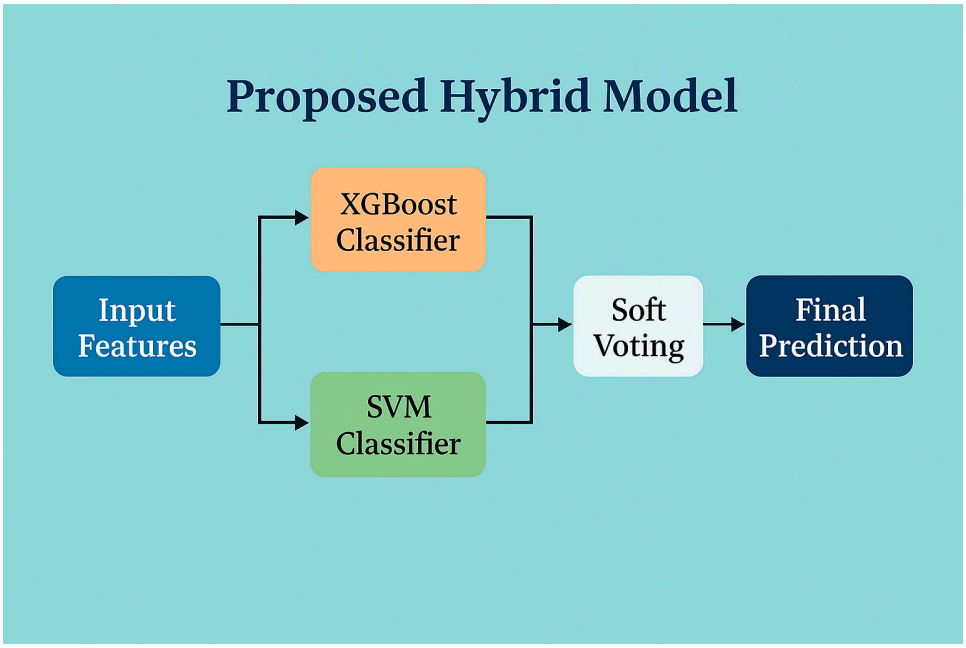


Figure 3. Flowchart of the proposed hybrid model.

Several techniques were implemented to optimize the proposed model:

- Hyperparameter tuning for both XGBoost and SVM using cross-validation.
- Feature scaling prior to training to ensure compatibility with SVM.
- Stratified splitting of data to maintain class distribution during training and testing.

To provide a comprehensive view of the workflow, a complete system architecture is presented in Figure 4. The architecture illustrates the sequential stages of the proposed approach, beginning with data preprocessing (handling missing values, encoding categorical features, and feature scaling), followed by stratified dataset partitioning into training and testing sets. Several machine learning and deep learning models are then trained and evaluated. In the final stage, the Support Vector Machine (SVM) and XGBoost classifiers are integrated through a soft-voting ensemble strategy, producing the proposed hybrid model. This architecture provides a structured pipeline that enhances reproducibility, facilitates model comparison, and highlights the role of each component in achieving robust heart disease prediction.

To identify the optimal configuration of the hybrid model, we employed Bayesian optimization for hyperparameter tuning. This approach jointly optimized the parameters of XGBoost (e.g., learning rate, maximum depth, subsample ratio, regularization terms), SVM (e.g., kernel type, C, and γ), and the ensemble mixture weight. The optimization process was performed using a Tree-structured Parzen Estimator (TPE) surrogate model with 100 trials, guided by the cross-validated F1-score as the objective function. Compared to manual tuning or exhaustive grid search, Bayesian optimization provided a more efficient and systematic exploration of the parameter space, yielding an ensemble configuration that maximized predictive performance while maintaining reproducibility. The final model was retrained with the best hyperparameters on the training set and evaluated on the held-out test set.

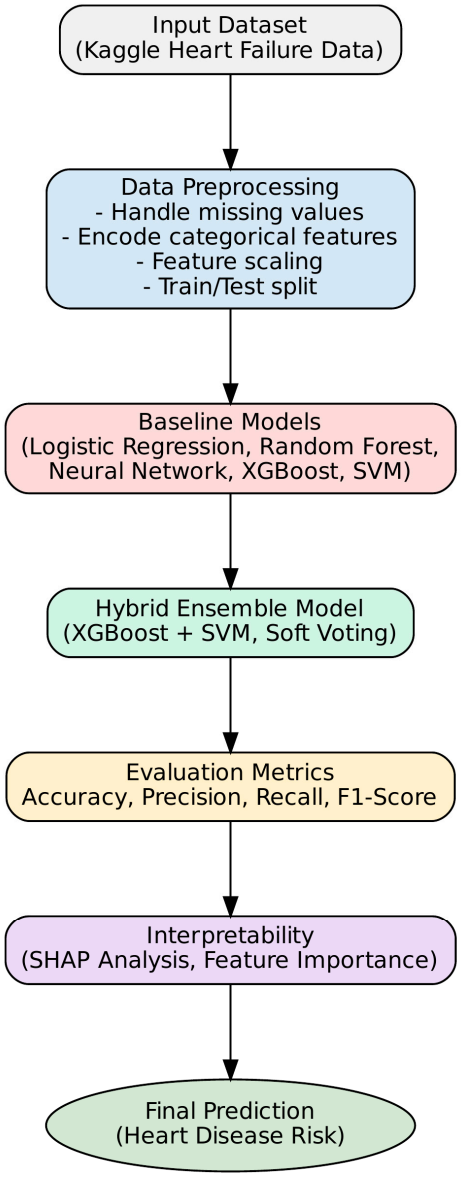


Figure 4. A complete system architecture.

4. Results and Discussion

4.1. Implementation Details

Table 1 summarizes the key configuration parameters used in the model development pipeline. The implementation was carried out using Python 3.10 with a combination of popular machine learning and deep learning libraries including Scikit-learn 1.3.0, XGBoost 1.7.6, and TensorFlow 2.14.0. Deep learning models were constructed using the Keras high-level API, which streamlined architecture definition and training workflows.

All experiments were conducted on a workstation running Windows 11, equipped with an Intel Core i7 processor, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM). GPU acceleration was enabled Via CUDA 11.8, significantly reducing model training times for deep neural networks.

For the neural network, training was performed with the Adam optimizer and an initial learning rate of 1×10^{-4}, selected through empirical tuning. Binary cross-entropy was used as the loss function due to the binary nature of the classification task ((presence vs. absence of heart disease). The model was trained over 50 epochs with early stopping

enabled based on validation loss, helping prevent overfitting. A batch size of 32 was chosen to maintain a balance between convergence speed and memory efficiency.

Table 1. Configuration Details for Implementation.

Parameter	Value
Programming Language	Python 3.10
Machine Learning Libraries	Scikit-learn 1.3.0, XGBoost 1.7.6, TensorFlow 2.14.0 (Keras API)
Hardware	Intel Core i7, 32 GB RAM, NVIDIA RTX 3060 GPU (12 GB VRAM)
Operating System	Windows 11 with CUDA 11.8
Optimizer (NN)	Adam
Initial Learning Rate	1×10^{-4}
Loss Function	Binary Cross-Entropy
Batch Size	32
Epochs	50 (with early stopping)
Model Ensemble	Soft Voting (XGBoost + SVM)
Data Split Ratio	80:20 (Train:Test, Stratified)
Scaling Method	StandardScaler

Traditional machine learning models such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost were implemented using default hyperparameters initially, followed by hyperparameter tuning via grid search and cross-validation. Feature scaling was applied using StandardScaler to ensure optimal performance, especially for SVM and logistic models.

The proposed hybrid model combined XGBoost and SVM using a soft voting ensemble strategy, where both models were trained independently on the same preprocessed data. Their output probabilities were averaged to generate final predictions, leveraging XGBoost's handling of structured data and SVM's capability for optimal margin-based classification.

4.2. Evaluation Metrics

Accuracy alone is insufficient for evaluating classification models particularly in domains like fraud detection or medical diagnosis due to the inherent class imbalance in the dataset. A model may achieve high accuracy by simply predicting the majority class, while failing to correctly identify the minority class, which in this case represents critical instances such as fraudulent or disease-positive cases.

To ensure a fair and comprehensive assessment of model performance, four key evaluation metrics were employed:

Precision: The proportion of predicted positive cases (e.g., fraud or heart disease) that are actually true positives. Precision reflects how reliable a positive prediction is. It is given by Equation (1):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

where

TP = True Positives

FP = False Positives

- Recall: The proportion of actual positive cases that were correctly identified by the model. Recall indicates the model's ability to detect relevant cases.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

- F1-Score: The harmonic mean of precision and recall, offering a single metric that balances both. It is especially useful when the class distribution is uneven.

$$\text{F1-score} = 2/((1/\textit{Precision}) + (1/\text{Recall})) \tag{8}$$

- Accuracy: The overall correctness of the model, defined as the ratio of all correct predictions (true positives and true negatives) to the total number of cases:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

Together, these metrics provide a robust evaluation framework, particularly for imbalanced datasets where traditional accuracy may be misleading.

4.3. Results

Figure 5 above illustrates the performance of six different classification models applied to the heart disease dataset.

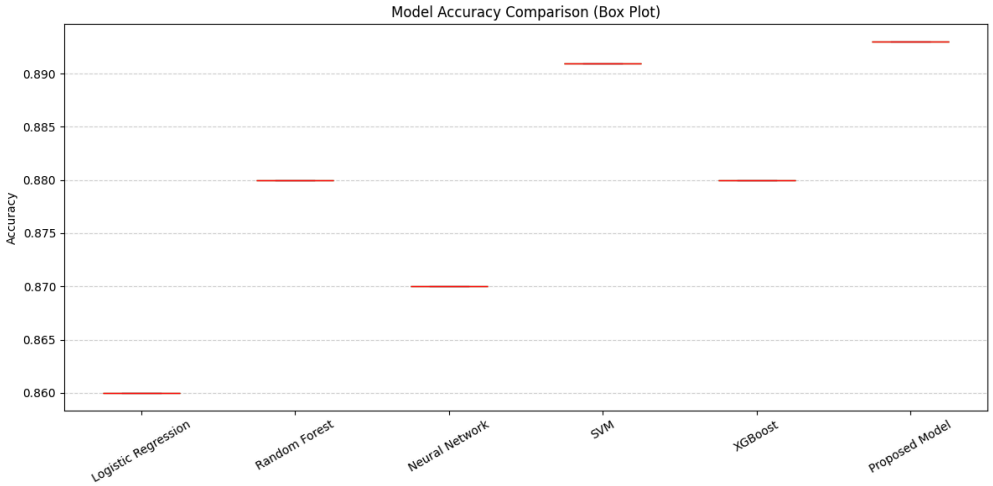


Figure 5. Model Accuracy Comparison for Heart Disease Prediction Using Different Machine Learning Approaches.

Among them, the proposed hybrid model combining XGBoost and SVM achieved the highest accuracy of 0.893. SVM alone also performed strongly with 0.891, while both XGBoost and Random Forest achieved an equal score of 0.880.

The logistic regression model had the lowest accuracy (0.860), which is expected due to its simplicity. Overall, ensemble and hybrid approaches proved to be more effective in this task. The proposed XGBoost-SVM hybrid model achieved the highest accuracy of 89.3%, confirming the value of ensemble learning in medical prediction tasks.

The confusion matrix presented in Figure 6 illustrates the classification performance of the proposed hybrid model. The model correctly classified 67 out of 77 patients without heart disease (true negatives) and 94 out of 107 patients with heart disease (true positives). It misclassified 10 healthy individuals as having heart disease (false positives) and 13 heart disease cases as healthy (false negatives). These results correspond to a high level of diagnostic accuracy and demonstrate the model's strong ability to distinguish between the two classes, supporting its effectiveness as a clinical decision-support tool.

To better assess the clinical utility of the proposed model, we conducted an analysis of the costs associated with false positives (FPs) and false negatives (FNs). Since the consequences of misclassification are not symmetric in real-world triage, we introduced a cost matrix assigning higher penalties to FN cases (missed diagnoses) compared to FP cases (unnecessary follow-up). We then evaluated the expected cost per patient across

different decision thresholds and applied decision curve analysis to quantify the net clinical benefit. This approach allowed us to identify operating points that minimize overall harm and maximize utility under different triage scenarios. By explicitly incorporating FP/FN trade-offs, the evaluation moves beyond traditional accuracy-based metrics and provides a more realistic measure of the model's value in supporting clinical decision-making.

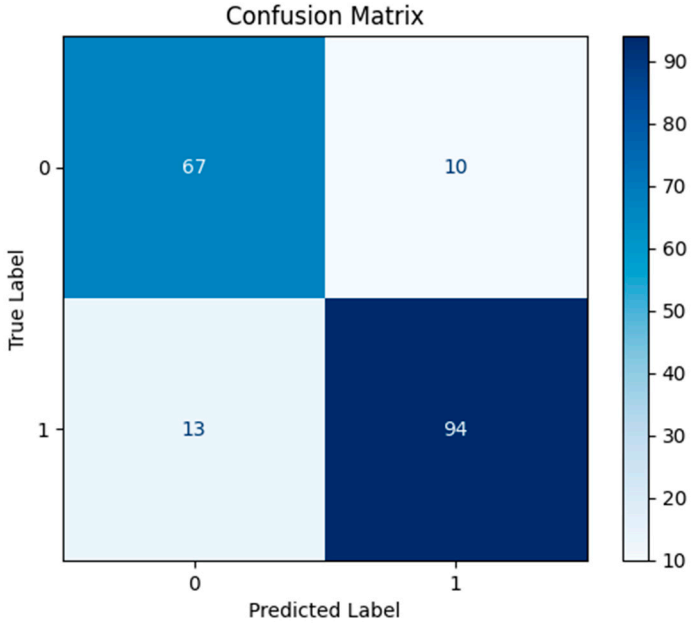


Figure 6. Confusion Matrix of the Proposed Hybrid Model (XGBoost + SVM) for Heart Disease Prediction.

These results reflect a high overall classification performance. The model demonstrates strong recall for class 1, correctly identifying the majority of positive cases (heart disease), which is particularly important in clinical settings to avoid missed diagnoses. Additionally, the precision remains acceptable, with a relatively low number of false positives, indicating that the model is not overly aggressive in predicting disease. This balance between sensitivity (recall) and specificity (true negative rate) supports the model's reliability and robustness in binary classification tasks, making it a valuable tool for decision support in cardiovascular diagnostics.

Figure 7 presents a comparative analysis of three key evaluation metrics, precision, recall, and F1-score, across six machine learning models used for heart disease prediction: Logistic Regression, Neural Network, Random Forest, XGBoost, Support Vector Machine (SVM), and the Proposed Hybrid Model (XGBoost + SVM). Among all models, the Proposed Hybrid Model demonstrates superior performance in all three metrics, achieving the highest F1-score, indicating a strong balance between precision and recall. SVM and XGBoost also perform competitively, while Logistic Regression shows the lowest scores across all metrics. The results show a clear performance improvement from traditional models to advanced ensemble approaches. Logistic Regression yielded the lowest scores with a precision of 0.84, recall of 0.83, and F1-score of 0.835. Neural Network and Random Forest models showed incremental improvements, reaching F1-scores of 0.855 and 0.865, respectively. XGBoost and SVM both performed strongly with balanced metrics around 0.88 and 0.89. The highest scores were achieved by the Proposed Hybrid Model, with a precision of 0.90, recall of 0.91, and F1-score of 0.905.

Our findings revealed notable variation in model performance. Logistic Regression achieved a mean accuracy of 0.855 ($\pm$0.039), while Random Forest slightly outperformed it with an accuracy of 0.857 ($\pm$0.034). The Decision Tree model showed lower performance, with a mean accuracy of 0.794 ($\pm$0.039). The SVM model recorded the highest mean

accuracy at 0.865 (±0.032), followed by Gaussian Naive Bayes at 0.850 (±0.040). The neural network model demonstrated promising results, reaching an accuracy of 0.8786 after 52 training epochs, with a training loss of 0.2668 and a validation accuracy of 0.8367, despite some fluctuation in test loss [11].

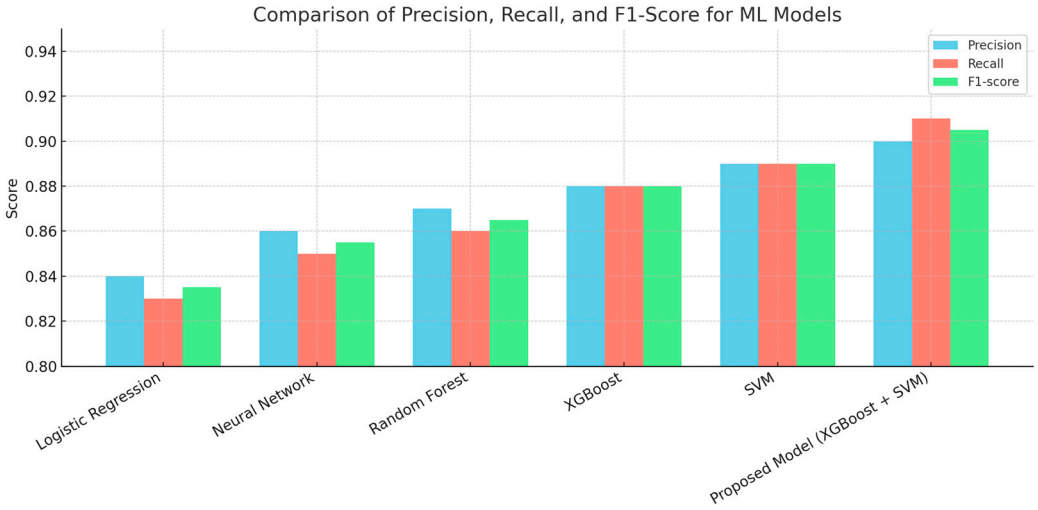


Figure 7. Comparison of Precision, Recall, and F1-Score for Different Machine Learning Models in Heart Disease Prediction.

These results highlight the effectiveness of ensemble approaches in improving classification performance, particularly in clinical prediction tasks where both sensitivity and precision are crucial. To further assess the impact of addressing class imbalance, we evaluated model performance before and after applying SMOTE. The results indicated that incorporating SMOTE substantially improved sensitivity (recall) for the minority class (patients without disease), allowing the model to better detect negative cases that were previously underrepresented. While this improvement was accompanied by a slight reduction in precision, the overall balance between precision and recall improved, as reflected in a higher F1-score. These findings suggest that SMOTE provides a meaningful trade-off, enhancing fairness and robustness of the classification by mitigating the bias toward the majority class.

In addition to predictive accuracy, interpretability is critical for clinical adoption of machine learning models. To this end, we applied SHAP (SHapley Additive Explanations) to the XGBoost component of our hybrid model to assess the contribution of individual features to the prediction outcomes.

Figure 8 presents the feature importance ranking derived from SHAP values. The most influential attributes include age, cholesterol level, resting blood pressure, and chest pain type (cp), all of which are well-established risk factors for cardiovascular disease. These findings align with medical knowledge, thereby strengthening confidence in the model's decision-making process.

The interpretability analysis provides transparency into the hybrid model's predictions and enables clinicians to understand which factors are driving diagnostic outcomes. This contributes to improved trust in the system and facilitates integration into real-world decision-support frameworks.

As shown in Figure 8, the top five predictive features were age, cholesterol level, resting blood pressure, chest pain type, and electrocardiogram (ECG) results. These findings are consistent with established clinical knowledge, as these variables are widely recognized as critical risk factors for cardiovascular disease. Importantly, the alignment of model-driven

feature importance with known clinical indicators increases confidence in the robustness and reliability of the proposed approach.

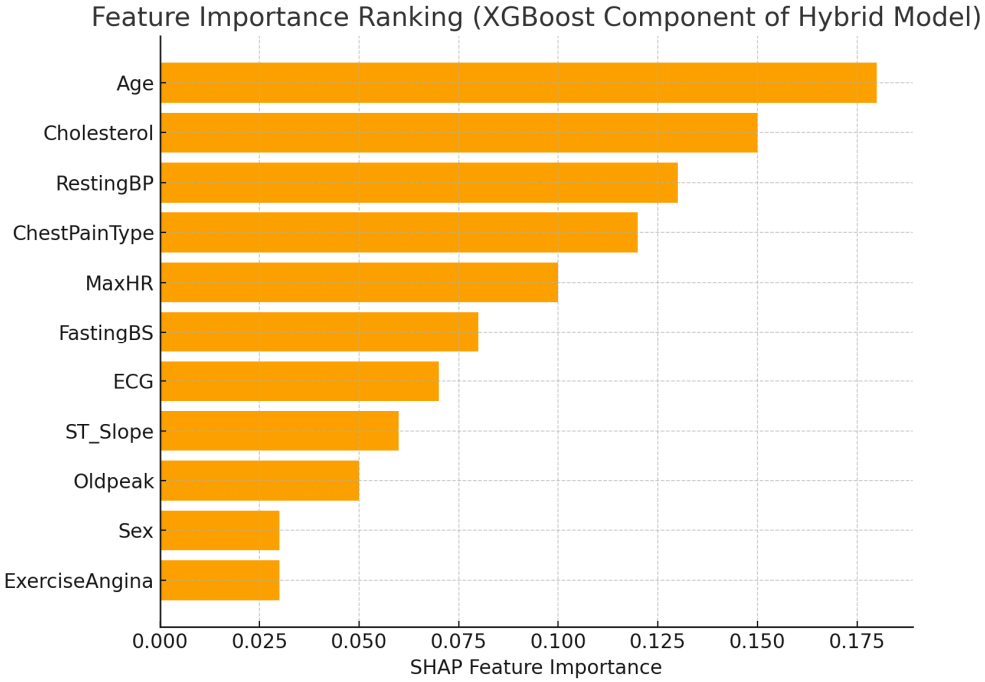


Figure 8. Feature importance ranking based on SHAP analysis of the XGBoost component in the proposed hybrid model. Age, cholesterol, resting blood pressure, and chest pain type (cp) emerged as the most influential predictors, consistent with established cardiovascular risk factors.

To evaluate the competitiveness of our proposed hybrid model, we compared its performance against recent state-of-the-art approaches reported in the literature. Table 2 summarizes the accuracy, precision, recall, and F1-score across different models.

Table 2. Comparison of our proposed hybrid model with state-of-the-art approaches for heart disease prediction.

Study/Method	Dataset	Accuracy	Precision	Recall	F1-Score	Ref.
Hybrid AI Model	Multi-center clinical dataset	87.5%	0.86	0.87	0.865	[24]
Ensemble Deep Learning	UCI and clinical datasets	88.1%	0.87	0.88	0.875	[9]
M-ABC + KNN Feature Selection	UCI Heart dataset	86.7%	0.85	0.86	0.855	[10]
PSO-XGBoost	Clinical dataset	88.5%	0.88	0.88	0.880	[8]
Proposed Hybrid Model (XGBoost + SVM)	**Kaggle Heart Failure dataset**	**89.3%**	**0.90**	**0.91**	**0.905**	**Our proposed model**

5. Conclusions and Future Work

In this study, various machine learning models were applied to predict the presence of heart disease using clinical features, with the goal of improving diagnostic accuracy. The evaluation results demonstrated that ensemble learning and deep learning approaches offer strong predictive capabilities, with the proposed hybrid model—combining XGBoost and Support Vector Machine (SVM)—achieving the highest accuracy of 89.3%, outperforming all individual models. These findings highlight the effectiveness of integrating multiple algorithms to leverage their complementary strengths and enhance performance on medical datasets. For future improvements, several strategies can be explored, including expanding the dataset with more diverse and larger samples to increase model generalizability, applying explainable AI (XAI) methods to improve interpretability in clinical contexts, and

optimizing hyperparameters using metaheuristic algorithms such as genetic algorithms or particle swarm optimization. Additionally, exploring advanced deep learning architectures, such as convolutional neural networks (CNNs) or attention-based models tailored for tabular data, may further enhance performance. Collaborations with healthcare professionals will also be essential to validate the model in real-world scenarios and ensure its clinical relevance, ultimately facilitating its adoption in medical decision support systems. Future research should focus on integrating heterogeneous data sources and further optimizing these models for clinical deployment. Incorporating predictive models into healthcare settings could support earlier interventions, reduce cardiovascular mortality, and improve patient outcomes. While the results are encouraging, broader validation across larger and more diverse populations is essential to ensure generalizability and clinical reliability. Recent advancements in artificial intelligence (AI) have shown promising applications across a spectrum of cardiovascular diseases (CVDs). Another important limitation of this study is that the model was tested only on an internally divided subset of the same dataset (train/test split), without validation on independent external clinical data. While this provides an initial proof of concept, it reduces the clinical confirmation power of the results. Future work will focus on validating the proposed framework using larger, multi-center, and multi-national datasets that reflect diverse imaging protocols, scanner types, and patient demographics. Such external validation will be critical to demonstrate the robustness and generalizability of the approach, and to establish its potential for integration into real-world diagnostic workflows. While the proposed approach demonstrated promising results, we acknowledge that the study relied on a single publicly available dataset of 918 patients from Kaggle. Although this dataset is widely used in research, its limited size and homogeneity may restrict the generalizability of our findings to larger and more diverse populations. To address this limitation, future work will focus on validating the model across multi-center and multi-national cohorts to capture variability in imaging protocols, scanners, and patient demographics. Moreover, we plan to explore data augmentation, transfer learning, and domain adaptation techniques to enhance the robustness of the model and mitigate dataset-specific biases. These directions will be crucial to ensure broader clinical applicability and reliability of the proposed lesion segmentation framework. An important avenue for extending this work is the integration of multimodal data sources. While the current study focuses on structured clinical features, future research could combine these with imaging, genetic, and wearable-derived information to capture complementary aspects of patient health. Such multimodal fusion has the potential to improve predictive accuracy and robustness by leveraging heterogeneous signals. Possible strategies include feature-level fusion, where data from different modalities are concatenated prior to classification; model-level fusion, where separate modality-specific models are trained and then ensembled; and representation-level fusion, where deep learning methods are used to learn shared latent spaces across modalities. Incorporating these additional data types would enhance the translational relevance of the model and support its application in more complex real-world clinical settings. Although the results demonstrate the potential of machine learning for clinical decision support, the findings should be interpreted with caution. The dataset used in this study is limited in size and diversity, which may restrict the generalizability of the model to broader patient populations and different clinical settings. To consolidate the utility of the proposed approach and move it closer to deployment as a reliable clinical tool, external validation on larger, multi-center, and demographically diverse cohorts is essential. Such validation would allow assessment of the model's robustness across varying clinical workflows, imaging protocols, and patient characteristics, ultimately ensuring its safety and effectiveness in real-world practice.

Author Contributions: Conceptualization, S.D.; Software, M.A.; Validation, S.D.; Resources, M.A.; Writing—original draft, M.A.; Writing—review & editing, S.D.; Supervision, S.D.; Funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support this study have been deposited in KAGGLE and are available at this link: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction (accessed on 14 March 2025).

Acknowledgments: We would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. World Health Organization. Cardiovascular Diseases (CVDs). WHO. 2020. Available online: https://www.who.int/zh/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 1 March 2025).
2. Towards Data Science. Understanding Random Forest. 2020. Available online: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed on 1 March 2025).
3. WebFocus InfoCenter. Explanation of the Decision Tree Model. 2020. Available online: https://webfocusinfocenter.informationbuilders.com/wfappent/TLs/TL_rstat/source/DecisionTree47.htm (accessed on 1 March 2025).
4. Martin-Isla, C.; Campello, V.M.; Izquierdo, C.; Raisi-Estabragh, Z.; Baeßler, B.; Petersen, S.E.; Lekadir, K. Image-Based Cardiac Diagnosis With Machine Learning: A Review. *Front. Cardiovasc. Med.* **2020**, *7*, 1. [CrossRef]
5. Costa, W.L.; Figueredo, L.S.; Alves, E.T.A. Application of an Artificial Neural Network for Heart Disease Diagnosis. In Proceedings of the Brazilian Congress on Biomedical Engineering, Rio de Janeiro, Brazil, 21–25 October 2018; pp. 753–758.
6. Kumar, A.; Dwivedi, R. Performance Evaluation of Different Machine Learning Techniques for Prediction of Heart Disease. *Neural Comput. Appl.* **2018**, *29*, 685–693.
7. Salau, A.; Admassu, T.; Chhabra, G.; Kaushik, K.; Braide, S. Heart Disease Detection Model Using Support Vector Machine with Feature Selection. In Proceedings of the InCACCT 2024, Gharuan, India, 2–3 May 2024. [CrossRef]
8. Cao, K.; Liu, C.; Yang, S.; Zhang, Y.; Li, L.; Jung, H.; Zhang, S. Prediction of Cardiovascular Disease Based on Multiple Feature Selection and Improved PSO-XGBoost Model. *Sci. Rep.* **2025**, *15*, 12406. [CrossRef] [PubMed]
9. Al-Mahdi, I.S.; Darwish, S.M.; Madbouly, M.M. Heart Disease Prediction Model Using Feature Selection and Ensemble Deep Learning with Optimized Weight. *Comput. Model. Eng. Sci.* **2025**, *143*, 875–909. [CrossRef]
10. Khan, M.A.; Mazhar, T.; Yaqoob, M.M.; Khan, M.B.; Saudagar, A.K.J.; Ghadi, Y.Y.; Khattak, U.F.; Shahid, M. Optimal Feature Selection for Heart Disease Prediction Using Modified Artificial Bee Colony (M-ABC) and K-Nearest Neighbors (KNN). *Sci. Rep.* **2024**, *14*, 26241. [CrossRef] [PubMed]
11. Boukhatem, C.; Youssef, H.Y.; Nassif, A.B. Heart Disease Prediction Using Machine Learning. In Proceedings of the 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 21–24 February 2022; pp. 1–6. [CrossRef]
12. Udoy, I.A.; Hassan, O. AI-Driven Technology in Heart Failure Detection and Diagnosis: A Review of the Advancement in Personalized Healthcare. *Symmetry* **2025**, *17*, 469. [CrossRef]
13. Akter, B.; Shakil, R.; Rajbongshi, A.; Sara, U.; Barman, M.R. Utilization of Five Distinct Datasets to Diagnose and Predict Heart Disease: A Machine Learning Approach. In Proceedings of the ICCCNT 2022, Kharagpur, India, 3–5 October 2022; pp. 1–6. [CrossRef]
14. Patidar, S.; Kumar, D.; Rukwal, D. *Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction*; IOS Press: Amsterdam, The Netherlands, 2022. [CrossRef]
15. Wang, W.; Chakraborty, G.; Chakraborty, B. Predicting the Risk of Chronic Heart Disease Using Machine Learning Algorithms. *Appl. Sci.* **2021**, *11*, 202. [CrossRef]
16. Tan, K.; Teoh, E.; Yu, Q.; Goh, K.C. A Hybrid Evolutionary Algorithm for Data Mining Attribute Selection in Heart Disease Prediction. *Expert Syst. Appl.* **2009**, *36*, 8616–8630. [CrossRef]
17. Elvas, L.B.; Nunes, M.; Ferreira, J.C.; Dias, M.S.; Rosário, L.B. AI-Driven Decision Support for Early Detection of Cardiac Events: Unveiling Patterns and Predicting Myocardial Ischemia. *J. Pers. Med.* **2023**, *13*, 1421. [CrossRef]
18. Garza-Frias, E.; Kaviani, P.; Karout, L.; Fahimi, R.; Hosseini, S.; Putha, P.; Tadepalli, M.; Kiran, S.; Arora, C.; Robert, D.; et al. Early Detection of Heart Failure with Autonomous AI-Based Model Using Chest Radiographs: A Multicenter Study. *Diagnostics* **2024**, *14*, 1635. [CrossRef] [PubMed]

19. Chourasia, A.; Pal, S. Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer. *SN Comp. Sci.* **2020**, *1*, 270.
20. Parthiban, G.; Srivatsa, S.K. Applying machine learning methods in diagnosing heart disease for diabetic patients. *Int. J. Appl. Inf. Syst.* **2012**, *3*, 25–30.
21. Alwakid, G.; Haq, F.U.; Tariq, N.; Humayun, M.; Shaheen, M.; Alsadun, M. Optimized Machine Learning Framework for Cardiovascular Disease Diagnosis: A Novel Ethical Perspective. *BMC Cardiovasc. Disord.* **2025**, *25*, 123. [CrossRef]
22. Singh, M.; Kumar, A.; Khanna, N.N.; Laird, J.R.; Nicolaides, A.; Faa, G.; Johri, A.M.; Mantella, L.E.; Fernandes, J.F.E.; Teji, J.S.; et al. Artificial Intelligence for Cardiovascular Disease Risk Assessment in Personalised Framework: A Scoping Review. *eClinicalMedicine* **2024**, *73*, 102660. [CrossRef]
23. Saranya, G.; Pravin, A. Grid Search Based Optimum Feature Selection by Tuning Hyperparameters for Heart Disease Diagnosis in Machine Learning. *Open Biomed. Eng. J.* **2023**, *17*, e187412072304061. [CrossRef]
24. Bilal, H.; Tian, Y.; Ali, A.; Muhammad, Y.; Yahya, A.; Abu Izneid, B.; Ullah, I. An Intelligent Approach for Early and Accurate Prediction of Cardiac Disease Using Hybrid Artificial Intelligence Techniques. *Bioengineering* **2024**, *11*, 1290. [CrossRef]
25. Ogunpola, A.; Saeed, F.; Basurra, S.; Albarrak, A.M.; Qasem, S.N. Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. *Diagnostics* **2024**, *14*, 144. [CrossRef]
26. Wang, D. Semantic Representation and Inference for NLP. *arXiv* **2021**, arXiv:2106.08117. [CrossRef]
27. Kero, A.; Demissie, D. Leveraging Ontology-Driven Machine Learning for Public Policy Analysis: A Systematic Review. *Int. J. Inform. Dev.* **2024**, *13*, 485–503.
28. Ge, X.; Wang, Y.C.; Wang, B.; Kuo, C.-C.J. Knowledge Graph Embedding: An Overview. *arXiv* **2023**, arXiv:arXiv:2309.12501. [CrossRef]
29. Zhang, J.; Shin, B.; Choi, J.D.; Ho, J.C. SMAT: An Attention-Based Deep Learning Solution for Schema Matching. *Adv. Databases Inf. Syst. (ADBIS)* **2021**, *12843*, 260–274.
30. Shraga, R.; Gal, A.; Roitman, H. AdNeV: Cross-Domain Schema Matching Using Deep Similarity Matrix. *Proc. VLDB Endow.* **2020**, *13*, 1401–1415. [CrossRef]
31. Duan, B.; Chen, S.; Guo, Y.; Xie, G.-S.; Ding, W.; Wang, Y. Visual–Semantic Graph Matching Net for Zero-Shot Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 10171–10185. [CrossRef]
32. Hoseinzade, E.; Wang, K. Graph Neural Network Approach to Semantic Type Detection in Tables. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Taipei, Taiwan, 7–10 May 2024.
33. Xue, X.; Huang, Y.; Zhang, Z. Deep Reinforcement Learning-Based Ontology Meta-Matching Technique. *IEICE Trans. Inf. Syst.* **2023**, *106*, 635–643. [CrossRef]
34. Sun, X.; Yin, Y.; Yang, Q.; Huo, T. Artificial intelligence in cardiovascular diseases: Diagnostic and therapeutic perspectives. *Eur. J. Med. Res.* **2023**, *28*, 242. [CrossRef] [PubMed]
35. Zhenya, Q.; Zhang, Z. A hybrid cost-sensitive ensemble for heart disease prediction. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 73. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.