A Systematic Review of Machine Learning, Deep Learning, and Explainable AI Approaches for Cardiac Disease Prediction

Sunanda Budihal*, Sheetalrani Kawale** & Abhishek Angadi*

DOI: https://doi.org/10.5281/zenodo.21206312

ABSTRACT

The cardiovascular (Cardiac) disease (CVD) is another factor that causes death among the global population most, and this is the reason why there is a high necessity to implement proper, effective, and interpretive diagnostic systems. The usage of machine learning (ML), deep learning (DL) and hybrid algorithms of artificial intelligence to predict heart diseases has enjoyed a widespread use in the recent years with different levels of success. The current paper provides a systematized review of the latest ML-, DL-, and hybrid-based systems to predict heart disease that include ensemble model, deep learning model, explainable artificial intelligence (XAI), and privacy-preserving model. The analyzed studies are evaluated based on datasets, classifiers, validation methods, data balancing and evaluation measures methods. It is illustrated in the analysis that the traditional ML models are most likely to have the prediction accuracy in between 80-90, the ensemble and hybrid models will most likely have the prediction accuracy in between 90-98. Recent explainable and fine-tuned ensemble methods show an accuracy of a level of 99, however, frequently under controlled conditions. Irrespective of the developments made, the issues include low interpretability, excessive benchmark data, imbalance in classes, and computational complexity. According to the research gaps that are mentioned, the given review reflects the necessity of clarifiable hybrid models that include high performing ensemble models like XGBoost and Deep learning-based feature fusion to increase the predictive quality, transparency and clinical usability.

KEYWORDS: *Cardiovascular Disease (CVD), Machine Learning, Deep Learning, Explainable Artificial Intelligence (XAI).*

I. Introduction

Cardiovascular disease (CVD) is the major cause of deaths across the globe where it ranks among the top causes of global deaths annually. The World Health Organization (WHO) shows that CVDs are one of the leading public health concerns, and thus the need to find proper methods to prevent the disease, detect it at its initial stages, and monitor it continuously. Early and accurate diagnosis of heart disease is not only vital in the reduction of mortality, better patient outcomes, but also a reduced healthcare burden in general [1]. Conventional diagnosis of heart diseases depends on clinical skills, laboratory and radiographic methods including electrocardiogram and echocardiography. Although these methods are effective, they are frequently time-consuming, subjective and lack scalability to facilitate the continuous monitoring of larger number of patients. As electronic health records and clinical datasets have been growing at a high rate, data-driven diagnostic methods are progressively becoming more popular.

Reviewed by Budihal and Kawale [2], machine learning (ML) methods promise useful potentials of automated prediction of heart disease through efficient analysis of structured clinical data. For early research, the most common ML algorithms used were Logistic Regression, Decision Trees, Support Vector machines and Random Forests on standard ML benchmark datasets such as the UCI Cleveland heart disease dataset. Research by Miao et al. [3] and Mohan et al. [7] has shown that the prediction systems based on ML can be implemented, but they showed average performance and generalization [3],[4]

* **Sunanda Budihal,** Department of Computer Science, Karnatak State Women's University, Vijayapura.

** **Sheetalrani Kawale,** Department of Computer Science, Government First Grade College, Bagalkot.

*** **Abhishek Angadi,** Assistant Professor, Department of MCA, Secab Institute of Engineering and Technology, Vijayapur.

later studies by Kavitha et al. (2021) showed that the hybrid ML models can enhance the accuracy, but the use of single datasets was still considered a serious issue [5]. The ensemble learning procedures were offered, and notably, Extreme Gradient Boosting (XGBoost) was shown to be more effective than the classical classifiers to face these challenges. Nagavelli et al. [9] and Subramani et al. [10] found that there was better predictive accuracy and improved robustness in the case of ensemble and hybrid ML-DL frameworks. These models are not interpretable all the time; hence, their use in clinical decision-support system is limited despite their effectiveness.

Meanwhile, new deep learning (DL) models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), hybrid ML-DL models have begun to be given more attention. Whereas, it is correct that DL models have been proved successful in the relationship with more complex nonlinearities, complete reviews like the one by Zhou et al. [11] show that there are still problems; one of them being lack of data and complete transparency [8]. High performance models have also been introduced with explainable artificial intelligence (XAI) techniques to increase levels of trust and clinical usability. The recent work of El-Sofany et al. [13] demonstrates that explainability methods such as SHAP with an ensemble classifier have played a crucial role in enhancing the transparency of model without having any effect on the performance of an ensemble classifier [9]. It is based on these findings that the current research trends are defined by the need to devise explainable hybrid schemes that are capable of integrating effective ensemble models like XGBoost with feature fusion realized through a neural network to achieve the right balance between accuracy, explainability, and practicality in the real world.

The remaining part of this paper is arranged in the following manner.

- Section II provides a thorough literature review on the existing works regarding prediction of heart (cardiac) diseases using machine learning, deep learning, and hybrid artificial intelligence methods.
- Section III provides a survey of popular learning models used in the study, their principles, strong sides and weaknesses in heart disease prediction.
- Section IV outlines the datasets applied in the studies being reviewed, benchmark, public, clinical, and imaging datasets, and the challenges encountered.
- Section V sums up performance evaluation measures and validation plans that are presented in the literature.
- Section VI will compare and discuss the strengths, trends, and limitations of ML, DL, hybrid, and explainable AI methods in a comparative manner.
- Section VII presents the research gaps and motivation with the open challenges and research directions.
- Lastly, in Section VIII, the paper concludes by summarizing the main findings and highlighting the way ahead of clinically deployable cardiac disease prediction systems.

II. Literature Review

Here, a literature review of studies about cardiac disease prediction with machine learning, deep learning, and hybrid AI models and explainable AI is performed. The focus is made on datasets, classification models, the strategy of validation, and trends of performance to detect the development of methods and gaps in the research. The initial studies on predicting cardiac diseases mostly depended on conventional machine-learning models that were used on structured clinical data.

The hybrid random forest-linear model (HRFLM) of Mohan et al. [7] was based on the UCI Cleveland data, and its accuracy is 88.7%. Even though the hybrid model performed better as compared to individual classifiers, generalization was limited because it was based on a limited number of features and one benchmark dataset. Nagavelli et al. [9] compared Naive Bayes, SVM, and XGBoost classifiers and gave the results that XGBoost was the most accurate (95.9%). Nevertheless, explainability and sound validation measures were not embraced in the study. On the same note, Biswas et al. [6] showed that prediction of heart disease at an early stage is better when feature selection is based on ML models, however, the computational complexity was a challenge. Recent studies with ML models have shown great performance with optimized ensemble techniques. Teja and Rayalu [17] worked on multiple UCI datasets and 93% accuracy on XGBoost, and Rimal et al. [20] have concentrated on stratified cross-validation and have achieved a 95% F1-score only.

Cardiac disease prediction has been further developed through deep learning since it can formulate non-linear associations that are complex. Miao et al. [3] trained a deep neural network (with a lot of multilayer perceivers) with the UCI Cleveland data set, which had an 83.7 percent accuracy and

AUC of 0.89. It demonstrated the capability of deep learning to work, but as other studies indicated, it has a limited scale, expensive calculation, and the models are not readily interpretable. A systematic review conducted by Zhou et al. [11] established that deep learning models generally outperformed the traditional ML methods in predicting heart diseases. However, the review further revealed that there are also long-term difficulties which include a lack of datasets, absence of standardized benchmarks, and little explainability; with Holste et al. [12] reporting an AUROC of 0.82 on 1 percent labeled data using a self-supervised deep learning model, EchoCLR. The study by Qadeer et al. [16] found that CNN based models were able to predict the instances of heart failure with 99% but the specificity of the disease limited generalization.

Proposals have been made to use hybrid models that can use the advantages of both paradigms, i.e. ML and DL. Subramani et al. [10] tested hybrid ML-DL framework on a hybrid of cardiovascular data and obtained 96% accuracy. Nonetheless, there was a limitation to the reliability of the results reported because of the heterogeneity of the datasets and the reproducibility of the results. Superior ensemble-based hybrid frameworks are said to report extremely high predictive power. Hamid et al. [15] suggested a more refined CatBoost model with 99.02% accuracy but noted that there is a problem of overfitting and external validation. Such a comprehensive experimental study as proposed by Rohan et al. [19], which implemented 21 ML and DL models and used XGBoost as the most successful model with the highest accuracy of 97% and AUC of 0.986, also had a high cost of computation. Navita et al. [22] suggested using stacking ensemble model, combining several classifiers, and SMOTEENN to balance the data, which showed 97.8% accuracy and 0.986 AUC. Although there is high performance the model is complicated and therefore difficult to deploy in real time.

The concept of explainable artificial intelligence (XAI) has become an important condition to enable clinical use of AI-based cardiac disease predictors. El-Sofany et al. [13] combined XGBoost with SHAP and SMOTE to come up with an explainable system with an accuracy of 97.57% and AUC of 0.98 at the feature level. Nonetheless, the framework itself was still computationally intensive and dataset specific. A framework that was ethically and fairness conscious was introduced by Al Gharib et al. [21], which included explainable hybrid framework that focused on transparency and accountability, but the flexibility of the model was impacted by feature limitations. Graph neural networks have been applied to predict heart diseases with accuracy of 91%, but Wajgi et al. [14] have encountered issues regarding scalability and interpretability. Privacy-preserving learning has become a topic of interest because of the sensitive patient information. Tyagi et al. [23] used federated learning to predict heart disease and had a performance that is comparable to centralized models without data privacy loss. Nevertheless, problems related to communication overhead and inability to explain are still ongoing. Budihal et al. [18] explored real-time clinical

Table 1: Summarizes Selected Literature Review of Representative Machine Learning, Deep Learning, and Explainable AI–based Studies for Cardiac Disease Prediction.

Ref. No.	Author (Year)	Title	Dataset Used	Model / Classifier	Validation Parameters	Results	Limitations	Future Work
[3]	Miao et al. (2018)	Coronary heart disease diagnosis using deep neural networks	UCI Cleveland	DNN (MLP)	Accuracy, AUC-ROC	83.7% acc, AUC 0.89	Small dataset; high computational cost; overfitting risk; low interpretability	Larger datasets; optimized architectures; regularization; XAI integration
[7]	Mohan et al. (2019)	Effective heart disease prediction using hybrid ML techniques	UCI Cleveland, Staglog	Hybrid RF–Linear (HRFLM)	Accuracy	88.7% accuracy	Restricted features; single dataset; no external validation; limited explainability	Multi-dataset testing; feature expansion; XAI; clinical validation
[9]	Nagavelli et al. (2022)	ML technology-based heart disease detection models	UCI Cleveland	NB, SVM, XGBoost	Accuracy	95.9% (XGBoost)	Small dataset; no XAI; limited validation; imbalance	Larger datasets; XAI adoption; robust validation; balanced learning

[10]	Subramani et al. (2023)	Cardiovascular diseases prediction by machine learning incorporation with deep learning	Combined datasets	ML – DL hybrid	Accuracy	96% accuracy	Dataset heterogeneity; high complexity; reproducibility issues	Standardization; normalization; lightweight hybrids; external validation
[11]	Zhou et al. (2024)	Review of DL-based HD prediction	Multiple studies	CNN, RNN, Hybrid DL	Accuracy, AUC	DL outperformed ML	Dataset scarcity; lack of benchmarks; limited explainability; training cost	Benchmark creation; explainable DL; multimodal fusion; clinical trials
[12]	Holste et al. (2024)	Efficient deep learning-based automated diagnosis from echocardiography with contrastive self-supervised learning	Echo videos	EchoCLR (SSL-DL)	AUROC	AUROC 0.82 (1% labels)	Imaging dependency; infra cost; modality limitation; annotation bias	Multimodal fusion; scalable SSL; diverse datasets; clinical validation
[13]	El-Sofany et al. (2024)	A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method	Private datasets	NB, SVM, XGBoost, RF, LR, others	Accuracy, AUC	97.57% acc, 0.98 AUC (XGboost)	High complexity; feature sensitivity; dataset bias; scalability	Lightweight XAI; real-time use; diverse datasets; fairness analysis
[14]	Wajgi et al. (2024)	Heart disease prediction using GNN	Kaggle	GNN (RMSprop)	Accuracy	91% accuracy	Graph construction cost; scalability; preprocessing overhead; low interpretability	Scalable GNNs; auto-graph learning; GNN-XAI; real-world data
[18]	Budihal et al. (2025)	Heart Disease Prediction Using Machine Learning On A Real-Time Clinical Dataset With Multi-Diagnostic Features	Clinical (1,100 records)	LR, SVM, DT, KNN, RF, XGBoost	Accuracy, ROC-AUC	93% acc, 0.98 AUC (RF and XGboost)	Class imbalance; single-center data; limited XAI; scalability	Balanced learning; multi-center data; XAI; deployment studies
[15]	Hamid et al. (2025)	Fine tuned CatBoost machine learning approach for early detection of cardiovascular disease through predictive modeling	Public dataset	CatBoost	Accuracy	99.02% accuracy	Overfitting risk; benchmark dependency; no external validation; complexity	External datasets; regularization; explainable CatBoost; clinical trials
[16]	Qadeer et al. (2025)	Heart Failure Prediction Through a Comparative Study of Machine Learning and Deep Learning Models	Clinical dataset	NB, DT, CNN, RF, KNN	Accuracy	99% (CNN)	Disease-specific; limited generalization; DL cost; interpretability	General CVD models; XAI; lightweight DL; broader datasets
[17]	Teja & Rayalu (2025)	Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance	Multiple UCI datasets	RF, KNN, LR, NB, GB, AdaBoost, XGBoost	Accuracy, AUC	XGboost and NB: 93%	Dataset bias; demographic limits; fairness gaps; no external validation	Bias mitigation; demographic diversity; fairness metrics; real deployment
[19]	Rohan et al. (2025)	An extensive experimental analysis for heart disease prediction using artificial intelligence techniques	Public datasets	21 ML & DL models	Accuracy	97% (XGBoost)	High computation; complexity; limited XAI; scalability	Efficient models; pruning; explainable AI; benchmarking

[20]	Rimal et al. (2025)	Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy	Multiple datasets	LR, SVM, KNN, RF	F1-score	RF F1 = 95%	No hyperparameter tuning; limited metrics; dataset bias; no XAI	Optimization; richer metrics; explainability; external datasets
[22]	Navita et al. (2025)	Advanced Hybrid Machine Learning Model for Accurate Detection of Cardiovascular Disease	Multiple datasets	RF, KNN, AdaBoost, LR	Accuracy, AUC	97.8% acc, 0.986 AUC	High complexity; computation cost; interpretability issues; deployment difficulty	Model simplification; lightweight stacking; XAI; real-time systems
[23]	Tyagi et al. (2025)	Heart Disease Detection Using Federated Learning	UCI datasets	Federated NN	Accuracy	Comparable to centralized	Communication overhead; heterogeneity; latency; no XAI	Efficient FL; FL+XAI; scalability; clinical validation

[Note: The table above reflects the data provided in source 78.]

deployment with 93% accuracy on the clinical datasets with RF and XGBoost but the problem of scalability and class imbalance remained.

III. Machine Learning Models

The review will analyze some of the supervised machine learning (ML) methods that have been heavily used in the premature detection and forecasting of heart disease. In monitored ML, the training of predictive models is done using labeled datasets where the patterns and relationships between the input clinical features and the associated disease outcomes are learned. After training, such models are tested on unseen or unlabelled test-data to determine their capability to categorically correctly label patients as diseased and non-diseased. Structured clinical datasets can be easily used to train supervised machine learning methods, and therefore they have been widely used in the prediction of heart diseases. In this part, a brief description of the popular ML-based classification methods presented in the literature on predicting heart disease and a description of their underlying principles and applicability to clinical decision support are presented.

3.1 Machine Learning Classification Techniques

Classification is one of the most basic supervised ML tasks whereby a model is trained to classify input data based on defined classes. In cardiac diseases prediction, the classification algorithms are normally employed to differentiate between patients with and without cardiac diseases according to their clinical factors like age, blood pressure, cholesterol, and electrocardiographic outcomes. Most common used ML classification methods in the study of heart diseases are summarized in the following subsections.

4.1. Decision Tree

A Decision Tree (DT) is a technique of supervised learning and is usually utilized in the classification of problems, but can also be implemented in regression. A decision tree is hierarchical in that it contains root node, internal decision nodes, branches, and leaf nodes. Every internal node is one of the decisions made depending on a certain feature and every leaf node corresponds to a particular class label. The ease and the interpretability of decision trees make them popular in the medical field because the process of decision-making can be easily visualized and explained to clinicians. DTs are applied to predict decision rules by clinical parameters in predicting heart diseases. Nevertheless, decision trees are subject to overfitting, especially when they are trained on small or noisy data, are likely to decrease their generalization.

4.2 Random Forest

Random Forest (RF) is an ensemble learning algorithm that is used to ensemble several decision trees in order to enhance the accuracy and strength of prediction. Every tree, in the forest, is trained on a random sample of the data and attributes and the eventual prediction is one that is produced by majority voting. The prevalence of the use of the random forest models in prediction of heart diseases is explained by the fact that they can handle high-dimensional data, decrease overfitting, and provide better accuracy than individual decision trees. Although RF models are highly performing, unlike individual decision trees, RF models are not interpretable because of the ensemble structure, which hides the direct paths of decision.

4.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised ML data structure that builds an optimal hyperplane in order to classify data points to different classes. SVMs are also effective in feature spaces of high dimensions and they can be used in nonlinear relations by use of kernel functions. SVMs have shown high classification in heart disease prediction with proper scaling of the dataset and features selected. Nevertheless, with large datasets, SVMs are computationally expensive, and one must be careful when tuning the kernel parameters in order to obtain a good performance.

4.4 Logistic Regression

Logistic Regression (LR) is a classification method based on statistics and machine learning that is popular in medical diagnosis. It calculates the probability of a binary event happening in the form of a linear combination of the input variables and applies logistic regression to ensure the output is between 0 and 1.

The logistic regression has been appreciated because of its simplicity, efficiency and interpretability, which is why it is useful as a baseline model in the heart disease prediction studies. It can however perform poorly in the case of complex nonlinear relationships that are there in clinical data.

4.5 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a strong ensemble learning algorithm that is based on the principles of gradient boosting. It constructs decision trees sequentially by reducing prediction errors of earlier models and also uses regularization methods to eliminate overfitting. XGBoost has been found to be among the most handy classification classifiers in predicting heart diseases with high level of accuracy in all data sets. It is particularly suitable when there is a structured data in healthcare because of its ability to handle missing data, data interactions and class imbalance. However, XGBoosts are generally computationally expensive and lack interpretability unless explainable AI techniques are employed.

On the whole, the Machine learning classification models are instrumental in prediction of heart diseases by facilitating automatic analysis of clinical data. Although other traditional methods, including Logistic Regression and Decision Trees, are interpretable, more predictive methods, such as Random Forest and XGBoost are ensemble techniques. The accuracy versus interpretability trade-off is the major factor that raises its demand, which explains the trend in integrating explainable AI methods into current studies.

3.2 Deep Learning Models

The use of deep learning models in prediction of heart disease has received enormous interest because of their capacity to automatically extract and learn nonlinear and intricate relationships with clinical and biomedical datasets at scale. As opposed to the classical machine learning techniques, which assume manual feature selection, the DL models use several hidden layers to derive hierarchical feature descriptions. Deep neural networks have been used to analyze structured clinical data as well as unstructured medical data in the past few years, including medical images and signals. This segment is a summary of deep learning structures that are widely used in the literature to predict heart diseases with their pros and cons.

5.1 Artificial Neural Network

The Artificial Neural Network (ANN) is one of the basic deep learning architectures and it is based on human brain structure. It comprises of interlinked layers of neurons, an input layer, a hidden layer or layers, and an output layer. All neurons approximate non-linearities between the characteristics of inputs and target results using weighted links and activation functions. ANNs have been extensively applied in the prediction of heart diseases as they are flexible and can represent complicated interactions among clinical variables. Numerous papers indicate better predictive accuracy with the ANN-based models than with conventional ML classifiers. However, ANNs are very sensitive to hyperparameter experiments and large datasets to avoid overfitting and their black-box nature limits their explanatory ability.

5.2 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of deep learning framework mainly used in the processing of spatial and temporal data. CNNs are designed to perform classification by first automatically computing local patterns and features of input data using convolutional layers and then pooling and full connected layers. CNNs have been used in prediction of heart disease not only on medical imaging data (including echocardiograms) but also on transformed clinical data. Models based on CNN have been shown to be highly accurate because they model complicated feature interactions. However, CNNs are computationally time-consuming and need high annotated datasets, which are not necessarily found in clinical practice.

5.3 Recurrent Neural Network and Long Short-Term Memory

The recurrent neural networks (RNNs) and their advanced form, Long Short-Term Memory (LSTM) networks are designed to learn the serial and temporal dependencies of the data. LSTMs eliminate the vanishing gradient problem that is present in the standard RNNs. These models are especially applicable when a time-series of health data is to be analyzed, including patient monitoring logs and longitudinal clinical measurements. RNNs and LSTMs can be used to predict the progression of the disease in heart diseases. Nevertheless, they are too complex and demand much computation which makes them difficult to implement in real-time.

Although deep learning models are powerful predictors, they have a number of weaknesses when used in clinical settings. These are high computational cost, large labeled dataset dependence and limited interpretability. The insufficiency of transparency in the process of making decisions based on the use of the DL causes the appearance of issues with clinical trust, accountability, and ethical adoption, which explains the introduction of explainable AI methods into current research.

The next-generation machine learning and deep learning models were required to overcome the drawbacks of standalone machine learning models and deep learning models with the recent studies paying more attention to the hybrid and explainable artificial intelligence (XAI) models. Hybrid models combine the strengths of a learning paradigm(s) more than one to enhance predictive performance, robustness and generalization. To add transparency and clinical trust, explainable AI approaches are integrated to reveal the insights on the model decisions. In this section, hybrid learning strategies and explainable AI techniques that are applied in predicting heart diseases are reviewed.

3.3 Deep Learning and Hybrid Machine Learning Models

Hybrid models are a combination of the traditional ML algorithms and deep learning architecture to evolve both the feature learning and classification functions. As an example of feature extraction or decision fusion, neural networks are combined with ensemble classifiers (Random Forest or XGBoost). Hybrid ML-DL systems have also shown higher performance than single models especially in the application of complex clinical data. These methods enhance precision and stability, but more complicated models and higher costs of computation are still important issues.

3.4 Hybrid Models, Which Are Based on Ensemble Learning

The hybrid models are an ensemble-based model, which incorporates the predictions of various classifiers to achieve a better performance. Some of these techniques include boosting, bagging and stacking, which are commonly used when studying the prediction of heart diseases. It has been demonstrated that boosting-based algorithms, especially XGBoost and CatBoost, have been effective in achieving high predictive accuracy because of their capability to deal with feature interactions and imbalanced classes. There are also stacking ensembles which learn the best combinations of base models. Even though they are effective, ensemble models may need mechanisms used to explain the model in order to assist in clinical interpretation.

3.5 Explainable Artificial Intelligence Strategies

Explainable AI (XAI) techniques attempt to reveal the model predictions and explain it to human users. One widely used technique to incorporate high-performing classifiers with SHapley Additive exPlanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME) is to give the expected importance of features and the local explanation. XAI can also improve the performance of clinicians in heart disease prediction because, as opposed to other methods, it is a transparent decision support system that allows clinicians to understand how particular clinical variables influence prediction. Experiments that combine XGBoost with SHAP indicate that it is possible to have both explainability and predictive accuracy. Nonetheless, the addition of XAI raises the complexity of the system and computation cost.

3.6 AI Models That Are Privacy-Preserving and Ethical

The recent studies have also covered privacy-controlling and ethically informed AI models to deal with issues of patient data confidentiality and fairness. With federated learning, the raw patient data are not shared when training the model, which makes it appropriate to multi-institutional healthcare settings. Although privacy-preserving models demonstrate encouraging findings, there are still research problems like overhead of communication, heterogeneity of the system, and lack of integration with explainable frameworks. Balancing predictive accuracy, interpretability and clinical applicability, hybrid and explainable AI models are a promising future in the quest of predicting heart diseases. Explainable AI and privacy-preserving solutions, ensemble learning, deep neural networks are becoming increasingly seen as the keys to the development of reliable and implementable healthcare decision support systems.

IV. Datasets Used

Machine learning (ML), deep learning (DL) and hybrid models are comprised of datasets that are used to create and test heart disease prediction models. The datasets used to train models have a strong impact on model performance and generalization capability in the literature that has been reviewed. This part of the paper outlines the key datasets that were reported in the chosen articles and points out their features and drawbacks.

UCI Heart Disease dataset is the most broadly utilized benchmark in research on heart disease prediction. The most widely used of those is the Cleveland dataset, which is structured and has the objective clinical variables of age, sex, type of chest pain, blood pressure at rest, cholesterol, fasting blood sugar, electrocardiographic findings and highest heart rate. This dataset is being used in many of the early and recent works to evaluate ML, DL, and hybrid classifiers. Nonetheless, it has a rather limited sample size and lacks demographic diversity which limits the transferability of reported findings.

Other studies use openly available cardiovascular datasets that have been acquired through open repositories like Kaggle, in addition to benchmark datasets. Such datasets usually have more samples and more clinical variables, thus they are appropriate to train ensemble and hybrid models. However, such problems as the absence of values, irregular defining of features, and insufficient clinical validation are frequently mentioned.

Few studies use realistic patient distributions based on real-world clinical or hospital-based data, which are more realistic and enhance clinical relevance. These datasets allow assessing the prediction models in a realistic healthcare setting. Limitations to access and use of such datasets are however, frequently set by privacy, ethical and regulatory issues which limits access and size of the dataset.

The most recent studies experiment with imaging-based and multimodal data, especially the echocardiography data, to assist deep learning and the self-supervised learning strategies. Through these datasets, extraction of both spatial and temporal features is possible which could not be obtained in the traditional tabular clinical data. Imaging datasets can only be used at very large scale with specialized equipment and with a lot of annotation, and this restricts their use. In general, it can be noted that the studied papers have a high reliance on benchmark and public datasets, and they are not extensively cross-dataset and externally validated. Such issues as class imbalance, small sample sizes, and standardized benchmarks are commonly used as challenges inherent to datasets. The failures highlight the need to have larger, more extensive and clinically valid data that will support feasible and implementable heart disease prediction systems.

V. Performance Evaluation

Performance evaluation plays one of the most important roles in determining the efficacy of machine learning (ML), deep learning (DL) and hybrid architecture in heart disease prediction. Performance evaluation takes the form of an analysis and comparison of reported results of existing studies in a review context as opposed to their new experimentation. This section outlines generally adopted measures of evaluation, validation and the trends in performance that have been observed in the literature reviewed.

The majority of studies assess heart disease prediction models based on conventional classification metrics derived based on the confusion matrix, where accuracy is the most frequently reported one, but which is also deceptive in the event of class imbalance, which is common in cardiovascular datasets. To overcome this shortcoming, other measures in precision, recall (sensitivity), specificity, F1-score and AUCROC are commonly used, where recall is especially important in the clinical sense because it represents the success of heart disease Gnanavelu et al., [24]. Train-test splits, k-fold cross-validation, stratified cross-validation are validation strategies that are typically used to test model generalization, with the latter being many preferable on smaller datasets where external validation may be done with independent datasets.

Patterns in the literature reviewed reveal that the traditional machine learning models typically obtain a range of 80-90, whereas the ensemble models, like the Random Forest, the XGBoost, and the CatBoost, always report higher accuracy rates of between 90-98 and deep learning or hybrid ML-DL methods sometimes report up to 99 accuracy rates under controlled experimental conditions. Data balancing strategies like SMOTE and SMOTE-ENN yield an efficient improvement in recall and F1-score, resulting in more clinically meaningful assessments, and the substitute transparency with the combination of explainable AI technologies does not affect the predictive performance significantly. In general, the findings point to the existence of a trade-off between accuracy, interpretability, and generalizability, which implies an essentiality of broad-based evaluation frameworks, including robust validation, balanced data processing, and explainability to facilitate faithfully clinically deployable heart disease prediction systems.

VI. Comparison of Various ML Techniques Used in Predicting Heart Disease

This section gives a tabular comparison between all the research papers presented above. This comparison will be made on the basis of accuracy and this can be observed in Figure 1. These elements in the table are few as demonstrated below.

The figure 1 makes a comparison between the accuracy reported on different ML, DL, and hybrid models on heart disease prediction. Early and traditional models have an accuracy of 88-92 percent whereas ensemble and boosting-based methods have better accuracy. XGBoost and ML-DL hybrid models attain over 96 percent accuracy which is better than that of standalone classifiers. Fine-tuned CatBoost (99.02%) and stacking ensembles (97.8%) are the most accurate, which shows the efficiency of advanced ensemble models. In sum, ensemble and hybrid models can always be seen to be demonstrating better predictive performance compared with other models because they have limitations, including inability to provide interpretability in their performance, as well as reliance on individual datasets and clinical trust in the models.

To address the shortcomings of existing studies, which include the lack of interpretability in the high-performing models, we present an explainable artificial intelligence (XAI)-based framework that combines ensemble models and fusion deep learning. XGBoost can be used to have powerful feature learning and high-quality prediction, whereas the fusion ANN is used to learn intricate nonlinear connections among the chosen features. To guarantee the transparency and clinical interpretability, explainable AI methods are integrated to measure the importance of features and give an insight into the decision-making of the model. The proposed hybrid and explainable framework is designed to meet the high predictive validity, better generalization, and increased trustworthiness to fill the critical research gaps and ensure the feasible implementation in the real-world cardiovascular disease diagnostics system.

VII. Research Motivation and Gaps

Although machine learning (ML), deep learning (DL), and hybrid artificial intelligence methods have advanced a lot in predicting heart diseases, the comparative study of the existing literature indicates that there are still multiple gaps in the research, which restrict their clinical implementation in the real world. The literature review has identified major gaps in research that still restrict the clinical implementation of an article in machine learning-based cardiac disease prediction systems.

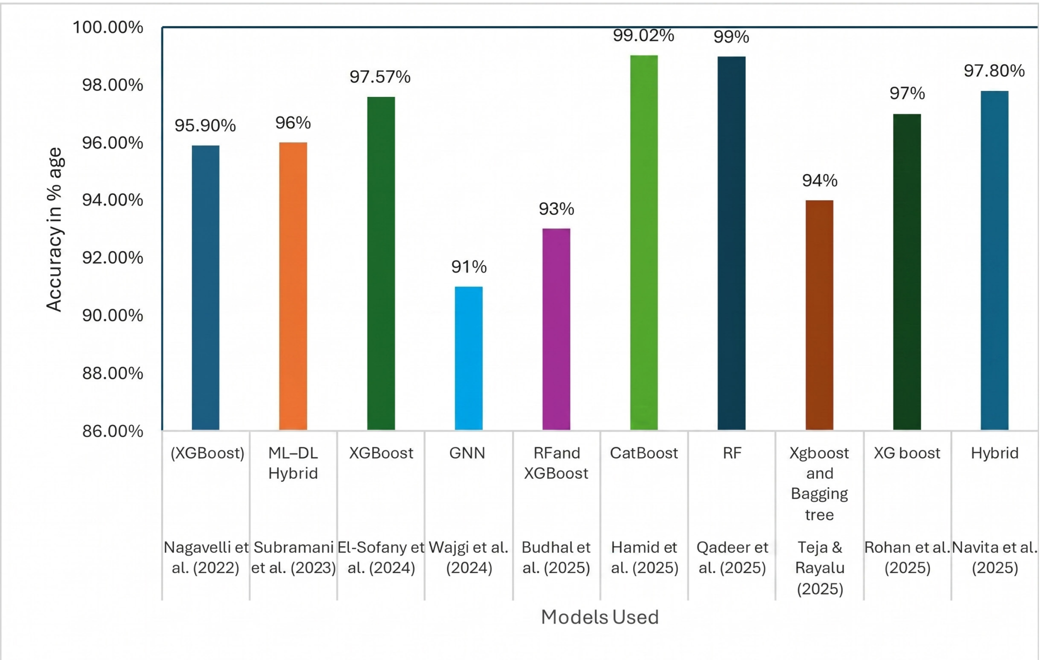

Figure 1. Comparison of machine learning, deep learning and hybrid models of predicting heart disease in the literature with regard to accuracy.

The excessive dependence on benchmark datasets, especially the UCI Cleveland one, limits the applicability of the suggested models to the real world and the real population of patients. Even though ensemble learning and deep learning methods often have a high predictive accuracy, their interpretability is often low, which compromises clinician trust and prevents their practical implementation in healthcare institutions. Moreover, the successful combination of explainable artificial intelligence methods, including SHAP, and high-performing hybrid or ensemble models is not widespread in the available literature. The other gap is poor management of the class imbalance and reporting that might result in the biased prediction and unreliable performance estimates. Moreover, several state-of-the-art deep learning and ensemble models are associated with high computation expenses, which makes them difficult to scale and implement in a clinical setting in real-time. Lastly, there is a lack of attention to issues associated with data privacy, ethical considerations, and fairness, and rather limited literature suggests privacy-conserving or fairness-aware models that are also highly performing and explainable.

Research Motivation It is based on the above gaps that there is need to develop future heart disease prediction frameworks which are less dependent on accuracy in the future. According to the literature, the ensemble models that have shown high performance like the XGBoost have always performed better than traditional classifiers, and the neural networks can adequately express nonlinear relationships between features. But their potentials have not been adequately researched in a framework which can be explained and relied upon in clinical practice. Thus, this review suggests the rationale behind the creation of explainable hybrid models incorporating a strong ensemble learning (e.g., XGBoost) and a deep learning implementation of feature fusion, with the help of the relevant data balancing policies. These frameworks have the potential to balance predictive accuracy, interpretability, fairness and clinical applicability, which are the major limitations of the current methods.

VIII. Conclusion

This review is a critical examination of current studies on machine learning, deep learning and hybrid artificial intelligence-based prediction methods on heart disease. Articles that appeared after 2018 and were studied systematically in terms of datasets, classifiers, learning paradigms, data balancing strategies, evaluation metrics, and reported performance showed that there was an apparent trend in the development of traditional machine learning models to ensemble, deep learning, explainable, and privacy-preserving models. Although the early machine learning solutions usually reached an accuracy of between 80-90%, more modern ensemble and hybrid models, which include Random Forest, XGBoost, CatBoost, and ML-DL bridges, have continually shown higher performance between 90-98%, and some fine-tuning and interpretable approaches have even reached 99% accuracy when run in a carefully controlled experiment.

Although these have been achieved, there are still issues of low interpretability, heavy reliance on benchmark datasets, imbalance in classes, computational complexity, and lack of external validation that limit its use in real-world clinical settings. Therefore, the advancement of explainable and privacy aware hybrid systems able to combine high performance ensemble model with deep learning feature fusion to help in improving predictive accuracy, transparency, and practical clinical implementation is recommended in future studies.

References:

1. World Health Organization, Cardiovascular diseases (CVDs), Geneva, Switzerland, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Budihal and Kawale, "A systematic review of machine learning models for cardiac disease prediction," International Journal of Computer Applications, vol. 186, no. 63, pp. 27–32, Jan. 2025, ISSN: 0975-8887.
3. K. H. Miao, J. H. Miao, and G. J. Miao, "Diagnosing coronary heart disease using deep neural networks," IEEE Access, vol. 6, pp. 66095–66102, 2018.
4. A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart disease dataset," UCI Machine Learning Repository, Univ. of California, Irvine, 1988. [Online]. Available: https://archive.ics.uci.edu/dataset/45/heart+disease.
5. The Devastator, "Predicting heart disease risk using clinical variables," Kaggle Dataset, 2025. [Online]. Available: https://www.kaggle.com/datasets/thedevastator/predicting-heart-disease-risk-using-clinical-var. Accessed: 2025.
6. N. Biswas et al., "Machine learning-based model to predict heart disease in early stage employing different feature selection techniques," BioMed Res. Int., vol. 2023, Art. no. 6864343, pp. 1–14, 2023, doi: 10.1155/2023/6864343.
7. S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.
8. P. Shah, M. Shukla, N. H. Dholakia, and H. Gupta, "Predicting cardiovascular risk with hybrid ensemble learning and explainable AI," Scientific Reports, vol. 15, Art. no. 17927, 2025, doi: 10.1038/s41598-025-01650-7.
9. U. Nagavelli, S. Samanta, and P. Chakraborty, "Machine learning technology-based heart disease detection models," J. Healthcare Eng., vol. 2022, 2022.
10. S. Subramani, R. Balakrishnan, and V. Ramalingam, "Cardiovascular disease prediction using machine learning and deep learning techniques," Frontiers in Medicine, vol. 10, Article ID 1150933, 2023.
11. C. Zhou, P. Dai, A. Hou, Z. Zhang, and F. Wang, "A comprehensive review of deep learning-based models for heart disease prediction," Artif. Intell. Rev., vol. 57, Article ID 263, 2024.
12. G. Holste, E. K. Oikonomou, B. J. Mortazavi, Z. Wang, and R. Khera, "Efficient deep learning–based automated diagnosis from echocardiography using contrastive self-supervised learning," Communications Medicine, vol. 4, Article ID 133, 2024.
13. H. El-Sofany, B. Bouallegue, and Y. M. Abd El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," Scientific Reports, vol. 14, Article ID 23277, 2024.
14. R. Wajgi, S. Patil, and P. Kulkarni, "Heart disease prediction using graph neural networks," IEEE Access, vol. 12, pp. 55421–55433, 2024.
15. M. Hamid, A. Rahman, and M. Islam, "Fine-tuned CatBoost machine learning approach for early detection of cardiovascular disease," IEEE Access, vol. 13, pp. 99876–99888, 2025.
16. M. Qadeer, H. Ali, and S. Khan, "Heart failure prediction through comparative study of machine learning and deep learning models," Sensors, vol. 25, no. 3, Article ID 456, 2025.
17. D. Teja and G. M. Rayalu, "Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance," BMC Cardiovascular Disorders, vol. 25, Article ID 212, 2025.
18. Budihal et. al.: "Heart Disease Prediction Using Machine Learning On A Real-Time Clinical Dataset With Multi-Diagnostic Features". International Journal of Applied Mathematics. Vol. 38 No. 12s (2025) https://doi.org/10.12732/ijam.v38i12s.1584
19. D. Rohan, G. P. Reddy, Y. V. P. Kumar, K. P. Prakash, and C. P. Reddy, "An extensive experimental analysis for heart disease prediction using artificial intelligence techniques," Scientific Reports, vol. 15, Article ID 6132, 2025.
20. Y. Rimal, N. Sharma, S. Paudel, A. Alsadoon, and S. Gill, "Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy," Scientific Reports, vol. 15, Article ID 13444, 2025.
21. S. Al Gharib, J. Charafeddine, F. Dornaika, and S. Haddad, "Hybrid learning framework for explainable cardiovascular disease detection," IEEE Access, vol. 13, pp. 130168–130182, 2025, doi: 10.1109/ACCESS.2025.3591241.
22. Navita, P. Mittal, Y. K. Sharma, U. K. Lilhore, S. Simaiya, and K. Saleem, "Advanced hybrid machine learning model for accurate detection of cardiovascular disease," Int. J. Comput. Intell. Syst., vol. 18, Article ID 51, 2025.
23. M. Tyagi, N. Jain, and S. Verma, "Heart disease detection using federated learning," IEEE J. Biomed. Health Inform., vol. 29, no. 4, pp. 2345–2356, 2025.
24. C. Gnanavelu et al., "Cardiovascular disease prediction using machine learning performance metrics," J. Young Pharm., vol. 17, no. 1, pp. 226–233, 2025, doi: 10.5530/jyp.20251231.
25. S. Mall, "Heart attack prediction using machine learning techniques," in Proc. 4th Int. Conf. Advance Computing and Innovative Technologies in Engineering (ICACITE), Ghaziabad, India, 2024, pp. 1778–1783, doi: 10.1109/ICACITE60783.2024.10617300.