A Comprehensive Review of Machine Learning and Deep Learning Approaches for Early Prediction and Diagnosis of Cardiovascular Diseases

[1]Satendra Ahirwar, [2]Dr Sunil Phulre, [3]Dr Bhupesh Gour
[1]MTech Scholar, Department of Computer Science and Engineering, Lakshmi Narain College of Technology & Science
[2]Professor, Department of Computer Science and Engineering, Lakshmi Narain College of Technology & Science
[3]HOD, Department of Computer Science and Engineering, Lakshmi Narain College of Technology & Science
[1]satyendraahirwar29@gmail.com , [2]sunilp@lnct.ac.in , [3]bhupeshg@lnct.ac.in

Abstract: Cardiovascular diseases (CVDs) constitute a major cause of death worldwide, posing immense challenges to health systems because of their multifaceted nature and delayed diagnosis. Thus, early detection and effective prognosis are valid parameters defining the loose ends in building strategies, Starpoint image for reducing the mortality rate and improving patient outcomes. This paper makes an appraisal into machine learning (ML) and deep learning (DL) techniques for early prediction and diagnosis of cardiovascular diseases in a comprehensive manner. Traditional diagnostic methods are generally dependent on clinical expertise and may suffer layback due to tardy detection, high cost, and variability in accuracy. To mitigate this, the use of ML and DL models has grown in recent years. This paper considers a broad spectrum of supervised learning algorithms, namely Decision Trees, Random Forests, and Support Vector Machines, alongside modern advances in deep learning structures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and the Transformer architecture. Furthermore, they will be analyzed based on the source of the data in terms of clinic datasets, medical imaging, ECG signals, and electronic health records (EHRs). The performance statistics like accuracy, precision, recall, and AUC are discussed to compare the efficiency between different models. The review brings out the major challenges such as data imbalance, interpretability, and very high computational complexity. The discussion, further, includes new developing main storms like explainable AI and multimodal learning, which could perhaps be considered as future research directions for the construction of robust and proficient CVD prediction systems.

Keywords: Machine Learning, Deep Learning, Cardiovascular Disease, Prediction Models, Healthcare Analytics, Explainable AI

I. INTRODUCTION

Heart disease refers to a group of medical conditions that affects the structure and function of the heart and blood vessels, rather being just a disease that means certain death or near-death. The disease is worldwide and has been taking a toll on public health systems. Some common forms of heart disease include coronary artery disease, heart failure, and arrhythmias. Heart disease results from genetic predisposition or a combination of environmental and lifestyle factors, such as poor nutrition, lack of exercise, smoking, and stress [1]. In order to mitigate mortality, it is important to diagnose and manage the disease before symptoms appear. However, as technology is evolving, machine learning techniques are beginning to be harnessed for more accurate and efficient predictions of heart disease. Figure 1Illustrates the structure of the heart and highlights conditions related to heart disease, including blocked arteries, reduced blood flow, and other cardiovascular abnormalities affecting heart function. This table 1 summarizes the major types of heart diseases along with their causes, symptoms, and associated risk levels for better understanding and classification.

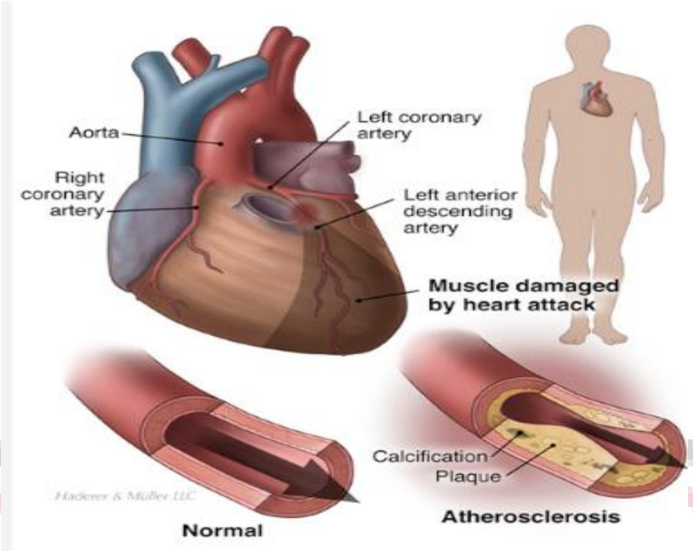

Figure 1: Heart Disease [2]

Table 1: Types of Heart Disease

Type of Heart Disease	Description	Causes	Symptoms	Risk Level
Coronary Artery Disease (CAD)	Narrowing of coronary arteries	Cholesterol buildup, smoking	Chest pain, shortness of breath	High
Heart Failure	Heart unable to pump blood effectively	High BP, diabetes	Fatigue, swelling in legs	Very High
Arrhythmia	Irregular heartbeat	Electrical signal issues	Palpitations, dizziness	Medium
Cardiomyopathy	Disease of heart muscle	Genetic factors, infections	Breathlessness, fatigue	High
Congenital Heart Disease	Present at birth	Genetic defects	Cyanosis, growth issues	Variable
Valvular Heart Disease	Damage to heart valves	Infections, aging	Chest pain, fainting	Medium
Hypertensive Heart Disease	Due to high blood pressure	Chronic hypertension	Headache, chest pain	High

A synergy between modifiable and non-modifiable risk factors causes heart diseases. Modifiable factors include unhealthy lifestyle choices such as smoking, excessive alcohol consumption, poor diet, physical inactivity, and obesity [3]. They may also include various chronic medical conditions, for example, hypertension, diabetes, and high cholesterol levels. The non-modifiable set is composed of classic factors like age, gender, and genetic predisposition. Stress and the environment also add to these heart-related complications. The interaction of these various risk factors takes a toll on the impairment of heart function over time. Steps for early identification and prevention of such risk factors, complemented by the design of predictive models of heart disease based on machine learning techniques, call for an understanding of these factors [4].
Heart disease remains a significant public health issue globally, accounting for multiple deaths every year. Cardiovascular diseases contribute to nearly one-third of all deaths in the world, as different reports on global health state. The burden in developing countries is increasing rapidly because of urbanization and lifestyle changes [5]. In India, heart diseases qualitatively traumatize most of the urban and rural populations, as their incidence is initiating among younger people attributable to sedentary lifestyles, stress, and harmful eating habits. It is important to adopt precise and proactive measures in order to arrest the steep progress of the disease. The use of machine learning in the health sector can greatly enhance early warning prediction with a considerable reduction of mortality rates and enhanced patient outcomes [6].
According to recent research, the increasing prevalence of heart disease, worldwide, has since become a major public health concern that reiterates the urgent necessity for successful predictive methods. Rapid urbanization and lifestyles that breed unhealthy traits along with attendant vacuous levels of stress and no physical activity are shooting up the cases of cardiovascular disease across all age groups. Diagnostic tests for detecting the disease, which in the conventional course are widely applied by the medical world, often subject the patient to manual analysis, with a resulting delay in the diagnosing process [7]. The more advanced the lethal cardiac cases, the bigger the dam in diagnosis-this is what resulted in the emergence of a clear-cut demand for more reliable and efficient methods of prediction. Hence, machine learning provides the machinery to survey large data sets of health information and uncover hidden patterns so as to be able to predict very accurately with minimal human assistance. It is the duty of classification technology to be such that diagnostic accuracies are heightened, to support clinical decision-making, and to allow for early detection of heart disease, with commensurate decrease in mortality rates and betterment in patient care issues [8]. Table 2 highlights the key drawbacks

of conventional diagnostic approaches, including low accuracy, delayed detection, reliance on manual expertise, and inability to efficiently handle large and complex medical data.

Table 2: Limitations of Traditional Diagnostic Methods

Limitation	Description	Impact on Diagnosis	Example
Time-Consuming Procedures	Requires multiple tests and expert analysis	Delays in diagnosis and treatment	ECG, stress tests over multiple visits
Human Error	Dependence on physician experience and judgment	Possibility of misdiagnosis	Incorrect interpretation of reports
Limited Data Utilization	Cannot efficiently process large-scale patient data	Misses hidden patterns and correlations	Ignoring historical patient records
High Cost	Diagnostic tests and hospital visits are expensive	Reduced accessibility for patients	MRI, angiography costs
Late Detection	Focuses more on symptoms rather than prediction	Disease identified at advanced stages	Detection after chest pain occurrence
Lack of Automation	Manual processes dominate diagnosis	Slower and less efficient system	Paper-based records
Inconsistency	Results may vary between doctors or institutions	Lack of standardization	Different diagnoses for same symptoms
Not Scalable	Difficult to handle large patient volumes	Inefficiency in mass screening	Overburdened hospitals

Cardiovascular diseases (CVDs) constitute one of the most pressing public health challenges confronting the world in the present day as CVDs account for a serious fraction of global morbidity and mortality and put massive pressure on healthcare systems, economies, and societies. CVDs cover a wide array of maladies affecting the heart and blood vessels; they include, among others, coronary artery disease, heart failure, arrhythmias, and stroke-many of which grow very slowly into asymptomatic states until the advanced stages. The surge of CVD prevalence is caused by many factors, such as aging population, suboptimal lifestyles, unhealthy dietary habits, increasing stress levels, use of tobacco, and co-morbid conditions like diabetes, hypertension, and obesity. In developing countries, particularly, quick urbanization and lifestyle transitions make cardiovascular conditions ever more common and early detection and prevention become the most urgent of tasks. The traditional diagnosis of cardiovascular diseases is constructed of clinical evaluations, laboratory tests, imaging techniques, and the know-how provided by physicians [8]. They have some disadvantages, too, in they are often costs, time-consuming methods, and variation in interpretation delay timely intervention. Traditional diagnostic systems remain essentially reactive (and not proactive) in that they are aimed at diagnosing diseases only after the symptoms come on in order to limit treatment efficacy, showing a track record of increased risk for severe complications and death. In this context, a great deal of attention has been to the early prediction and risk assessment of cardiovascular diseases; early diagnosis of high-risk individuals would allow them to be suitably managed with prevention and therapy and improved patient outcomes. Early prediction does not only significantly decrease mortality but also caters to cost containment because patients would have been treated at an early stage rather than at an advanced one with expensive drugs. Rapid developments in computing technology and the availability of large-scale data in health biosciences are shifting the traditional medical paradigm towards data-driven predictive modeling. Machine learning and deep learning are, therefore, emerging as very powerful tools, capable of handling complex, high-dimensional data in medicine to identify patterns that are otherwise invisible using traditional statistical techniques [9]. These systems can assimilate various data sources, like EHRs, clinical measurements, imaging scans, genetic information, and physique signals such as electrocardiograms, to come up with laboratory and reliable outcomes. In addition, their continuous learning and adaptation to new data render them a good ground for repetitive improvement in prediction performance in dynamic healthcare settings. Although all these breakthroughs are taking place, various challenges such as data quality issues, class imbalance, interpretability, and generalizability across different populations and healthcare systems need to be resolved. The ethical discussions, issues on data privacy, and requirements for being transparent in decision-making will play significant roles in the adoption of these technologies in real-world clinical applications. Thus, there is urgent hidden necessity to explore and test modern computational methods able to overcome these limitations for the purposes of providing accurate, efficient, and scalable solutions in predicting cardiovascular diseases. This review aims to present a broad comparative review of actual landscape of machine learning and deep learning approaches that are currently available for early prediction and diagnosis of the cardiovascular diseases with the potential to revolutionize medicine with preventive and personalized care [10].

II. BACKGROUND AND PRIOR RESEARCH STUDIES

Cardiovascular diseases (CVD) are coupled with research over the last few decades, moving from traditional statistical analysis to more complex micro-structured data, machine learning (ML), and deep learning (DL). In the early stages of cardiovascular risk prediction, researchers relied primarily on the formidable statistics tools like logistic regression and

Cox proportional hazards regression to capture a limited number of clinical risk factors such as age, gender, blood pressure, total cholesterol, smoking, and lastly, diabetes. But these prognostic rules lacked the ability to capture unconventional, complex nonlinear associations between risk factors and that would have ensured accuracy in the estimation of risk. In this environment of very rapid data growth and increasing computational possibilities, ML algorithms emerged for these extended purposes. Techniques such as Decision Trees, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Random Forests, among others, are used in a very wide range on cardiovascular datasets for improving the classification success and in stratification of risk. For instance, several studies have shown that ensemble learning methods have significantly outperformed traditional statistical approaches, doing so by reducing overfitting and effectively handling high-dimensions of the data. Furthermore, several feature selection and dimensionality reduction techniques, which include Principal Component Analysis (PCA), mutual information methods, have been experimented with extensively to advance model efficiency and interpretability. Integration of structured clinical materials with heterogenous sources such as electronic health records (EHR), medical imaging [11], ECG signals, and wearable sensor data has broadened the palette of cardiovascular investigations. Recent studies have shown that ML models trained on EHR data could effectively predict outcomes like heart failure, sudden cardiac death, and long-term mortality with the highest levels of accuracy, allowing for better-informed decisions for smart patient care. Over time, deep learning made its way into this realm, where it introduced possibilities to carry out automatic feature extraction and representation learning from raw data. Convolutional Neural Networks (CNNs) found ample use in medical image analysis, such as MRI, CT, and echocardiographic heart images, with astonishing accuracy and precision across all disease classifications, segmentation, and anomaly detection. In the similar context, recurrent neural networks (RNN) and longitudinal short-term memory (LSTM) networks have been applied on sequential data for ECG signals or time-series patient records to pick up about temporal patterns and a health dynamic. Transformer-based architectures [12], leveraging attention mechanisms, offer another alternative to their capability to use far-field dependencies to produce higher accuracy in complex data set predictions. Hybrid models fall within another category, combining classical ML techniques with deep learning architecture models. One promising example in this direction is the combination of manual features with learned representations that improve diagnostic performance in ECG-based heart disease detection. Simultaneously, the number of Explainable AI (XAI) techniques, among which SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) feature predominantly, have helped address one of the major drawbacks hampering the usability of ML models that is lack of interpretability. These techniques enable a clinician to understand the contribution of individual features towards model predictions to generally win trust and promote clinical translation. An turther array of studies put their sight on the fairness and mitigation of bias in ML models. Despite major advancements in research of cardiovascular disease primary prediction, many limitations and challenges still exist, among which include data imbalance and missing data [13], heterogeneous and variable datasets, and limitations in generalizing to different populations and healthcare systems. To note, many of the studies have been small of scope, only generalizing to limited testing sets that would not truly represent any patient populations, while the operational complexity in deep learning models might cut short their real-world application in real-time in severely resource-constrained settings. Moreover, the matter of privacy and security surrounding sensitive patient data poses a great barrier to sharing of data and model construction. Without doubt, there are emerging trends of research that are meant to tackle the existing challenges: federated learning, which proposes training models across settings while keeping the raw data contained privately in other institutions, and multimodal learning, which seeks to combine numerous data types, giving a more complete grasp on the patient's health. With technological development of edge computing and Internet of Things (IoT), real-time monitoring and prediction of cardiovascular conditions vis-à-vis wearable devices and remote healthcare systems are now possible [14]. Basically, the advancement in cardiovascular disease prediction research indicates a shift toward advanced, accurate, and patient-centered approaches, thanks to the amalgamation of cutting-edge computational approaches and multifarious data sources. This new research underpins the scope of ML and DL model contributions to transform cardiovascular care by enabling early-detection, personalized treatment, and improved outcomes, while insisting on the necessity of continued research to address existing challenges and ensure operationalization of these technologies in real-world settings.

III. MACHINE LEARNING TECHNIQUES FOR CVD PREDICTION

Machine learning (ML) has been shown to have a profound impact on cardiovascular disease (CVD) prediction and early detection methods by way of contributing to the efficacy over traditional statistical or rule-based methods. Predicting CVD requires complex descriptions of how the high-dimensional medical data concerning patient demographics, clinical history, lifestyle characteristics, and a myriad of diagnostic test results indicates those individuals who are at risk. ML methods are indeed highly satisfactory in that they can derive hidden patterns from historical training data, therefore improving their own predictive accuracy along the way in an explicit programming-free environment. In recent years, a host of ML algorithms are applied to CVD prediction, with supervised learning, unsupervised learning, and ensemble methods contributing in diverse, unique ways to improving the diagnostic tool and decision system in healthcare.

Supervised learning techniques are most frequently employed for cardiovascular risk prediction, with a great emphasis on labeled medical datasets. These methods involve training a model is trained on input elements, like age, blood pressure, cholesterol level, heart rate, and glucose level along with known output labels indicating whether a patient had CVD or not. Logistic Regression is quite simple but can provide more interpretability. It makes use of a sigmoid function to estimate the probability of a person to have CVD. Owing to this, Logistic Regression has been significantly praised in medical applications where the model must be able to convey some meaning. Despite its simplicity, Logistic Regression is more sophisticated in a situation where linear forms of relationship exist, establishing the Logistic Regression model as a sturdy baseline for CVD prediction frameworks [15].

Another widely used method for CVD prediction is the Decision-Tree classifier. Models like this one divide the dataset into smaller sets of instances based on the values of feature as decided by a node from the tree structure. In general, Decision Trees provide an intuitive model that lends itself to good support for clinical decisions-specifically owing its simplicity and its being so easy to interpret. However, much attention needs to be paid to overfitting when the model is trained with small datasets. For this reason, more sophisticated ensemble techniques such as Random Forest are adopted. Random Forest creates a number of Decision Tree classifiers coupled with a majority voting scheme for the aggregation of all outputs, which would generally improve accuracy and alleviate issues associated with overfitting. It is often effective at modeling complex nonlinear relationships and missing data; the very characteristics typical of healthcare-related datasets. Support Vector Machines (SVM) have also found good applications in CVD prediction because these machines are capable of defining high-dimensional feature space. SVM does the job by finding the hyperplane that completely separates classes with the maximum margin. The machine learning model particularly comes in handy when the dataset hosts complex boundaries between healthy and case groups. SVM, being well associated with Kernel functions, can effectively learn these nonlinear relationships, such as the Radial Basis Function (RBF). The SVM with its kernel is thus able to predict risks at a higher percentage level. SVM has been positively seen as optimal in most cases, but is loaded with computational expenses for large data, thus not recommended for real-time applications if not optimized [16].

Various procedures like k-Nearest Neighbors (KNN) are basic representations extracting knowledge in data-giving sections for an array of medical prediction circumstances. KNN is used for new patient classification by the majority class in its nearest neighbors in pattern space. The seat would not require a training function, which explain its hassle-free implements. Their implementation relies heavily on the selection of a distance function and the value of the numeric variable, thus not simultaneously sacrificing performance. The KNN classifier becomes computationally expensive when used during predictions pertinent to gigantic datasets. Naïve Bayes is a probabilistic algorithm used for CVD prediction. This method is proposed by applying Bayes theorem under the assumption that all features are independent of each other. Despite this strong assumption, Naïve Bayes has demonstrated great light in the classification of medical data owing to simplicity and efficiency; great performance may be obtained over huge datasets as well as on a couple of features set as filenames and a couple of cases. Though, its accuracy may sharply decrease in the case when the bulk of the individual features are not independent within the development data. Besides these traditional supervised learning methods, ensemble learning has recently attracted much attention. They contribute to enhancing the overall predictive performance by combining various models. Moreover, techniques, such as Gradient Boosting Machines (GBM), XGBoost, and AdaBoost, are widely used in CVD prediction systems, and XGBoost in particular is robust in the presence of missing values, regularized to prevent overfitting, and computationally efficient [17]. This methods encapsulate interactions among clinical variables very well. This can make an immense difference to classification accuracy.

Unsupervised learning algorithms have an important part in the analysis of cardiovascular conditions, especially when there are no label data. For example, clustering of individuals by their health profiles can be based on the application of the K-Means and Hierarchical Clustering algorithms. Clustering analyses of health profile groups help pinpoint some less prominent risk patterns concerning early CVD diagnosis. For example, similarity in high cholesterol levels, blood pressure, lifestyles can resolve into a high-risk cluster, maybe even before a medical diagnosis of CVD could have been detected. Principal component analysis (PCA) stands another method for unsupervised analysis to reduce dimensionality in allowing complex sets to have equally complex features preserved.

These new deep learning techniques in the offing certainly assisted the CVD prediction as they delivered automatic feature extraction from big, imbalanced data. The Artificial Neural Networks (ANN) are adopted for approximating nonlinear relationships between input features and disease outcomes. The feedforward neural network of Multilayer Perceptron (MLP) is often applied in structured medical datasets. Multilayer perceptrons have many hidden layers that learn hierarchical feature representations allowing them to achieve more accurate predictions. However, ANN-needs large datasets and substantial computational resources. Mainly CNNs are used for the image-based diagnosis of cardiovascular issues, including analysis of ECG signals or medical imaging data like echocardiograms. The idea is to automatically extract spatial features from raw input data, therefore reducing the need for manual feature engineering. RNNs and LSMT networks are used for time-series data analysis too, such as continuous heart rate monitoring. They can learn from the historical data regarding time-series, thereby making them suitable for predicting future cardiac events given the patient's historical data. Hybrid models are a part of the quest for developing optimization-based ML models that are being more widely used in CVD prediction research. In these models, a machine-learning method is combined with an optimization Philosophy to impel the best performance from both worlds. For this example, we can take some models such as feature selection algorithm into account to improve classification model performance-wise and ignore so much additional features [18]. Optimization using PSO, GA, and FA is yielding its currency here by looking at fine-tuning hyper-parameters and optimizing feature selection.

Feature engineering and selection are another key factor for an ML-based CVD prediction program. In most cases, medical datasets would almost certainly have redundant or irrelevant features which could affect a model's performance. Feature selection techniques such as Recursive Feature Elimination (RFE), correlation analysis, and mutual information help identify the crucial predictors for CVD. To put it mildly, proper feature selection will improve accuracy while minimizing computational complexity and, once that is done, make the model interpretable as well. Model evaluation critically ensures the effectiveness of ML-based CVD prediction systems. Common evaluation metrics include accuracy, precision, recall, F1-score, and Area Under the Curve (AUC); recall is particularly crucial in medical diagnostics because it quantifies the model's ability to effectively detect CVD in sufferers, thus reducing false negatives. Techniques such as cross-validation are critical in order to secure that the model generalizes well in an unknown world. While cardiovascular risk prediction systems using machine learning models have shown remarkable success, they face quite a few important challenges, such as data imbalances, poor quality data sets, lack of interpretability for sophisticated models, and privacy that tends to be a particularly significant issue regarding patient data. Data imbalances tend to be an even greater pain point, with abundant patients who are healthy far outnumbering those diagnosed with CVDs. This leads to the bias in predictive results. Counter-method measures are the SMOTE method (Synthetic Minority Over-sampling Technique), for overcoming imbalanced datasets [19].

To conclude, machine learning has completely revamped the prediction of cardiovascular diseases by ensuring their early diagnosis, increasing accuracy, and facilitating clinical decision-making. From straightforward models like Logistic Regression to sophisticated ones like deep learning and ensembles, each model has its own distinctive contribution towards enhancing the predictive performance. By combining the use of machine learning within healthcare systems, machine learning has the potential to decrease mortality by enabling timely intervention and personalized treatment strategies. The future is likely to focus on explainable intelligence with real-time predictive systems and integrate with external wearable health-monitoring devices, too; thereby augmenting the role of machine learning in cardiovascular healthcare.

IV. DEEP LEARNING APPROACHES IN CARDIOVASCULAR DIAGNOSIS

Deep learning (DL) has brought major transformation in cardiovascular disease (CVD) diagnosis through automated extraction of features, dealing with high-dimensional data, and improved predictability as compared to conventional machine learning algorithms. It stands in contradistinction to classical methods that rely largely on hands-on engineering of features; instead, deep learning models obtain learned hierarchical representations directly from the raw medical data, such as electrocardiograms (ECG), echocardiograms, cardiac MRI images, and the electronic health records (EHR) of patients. This is where DL helps strongly in intricate cardiovascular diagnostics cases where patterns are often nonlinear and are beyond grasp for conventional statistical methods. In recent years, deep learning has become a backbone for intelligent healthcare systems for the earliest stages of prediction, risk assessment, or the classification of cardiovascular conditions. Artificial Neural Network is the most commonly used deep learning model for cardiovascular diagnosis. It consists of interconnected neurons arranged into layers, i.e., the input layer, hidden layers, and output layer, and it processes medical data through weighted interconnections. Such networks can learn the complexity of the interactions between clinical features, be it related to blood pressure, cholesterol level, heart rate, glucose level, etc.; and Multi-Layer Perceptron (MLP), a variant of feedforward neural network, is quite popular for structured, tabular medical data. By grasping the nonlinear interactions between features, MLP significantly improves accuracy of prediction and is thus suited for tasks such as binary classification, for example distinguishing between cardiovascular disease's presence and absence. One of the most hyped neural networks for cardiovascular imaging is the convolutional neural network (CNN), acclaimed for its ability to study spatial patterns in medical images like angiograms, echocardiograms, and MRI scans. CNN models work by extracting meaningful features from input images. It is through these features that characteristically distinct groups of cells come into existence. Said differently, in this method of learning where the network must extract features from the images, the process converts unusable pixel information into meaningful feature-extracted and pooled representations. Furthermore, changes in pixel intensity could lead to subtle patterns and changes in the digitized images [20].

The other major class of deep learning models is Recurrent Neural Networks (RNN) that are employed for cardiovascular uses, especially when it comes to multiple-level data and time series data; these are equipped to hold the memory of the earlier inputs, thereby being the case with the analysis of long continuous heart monitoring data. But the standard RNNS can achieve vanishing gradients when handling very long sequences; hence, Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) are more often used as improvements and are in a good position for long-range dependency in physiological signals other than just the heart rate variability and blood pressure trends. The LSTM-based models offer their best capabilities in predicting future cardiac events based on the historical patient data.

To enhance inter-model integration during cardiovascular diagnosis, the use of hybrid deep learning representations has lately been on the rise. These architectures fuse features from various architectures, as in CNN-LSTM, in order to make the most of the space-time characteristics. The CNN-LSTM model, for example, is particularly effective in ECG-based diagnosis since CNN primarily identifies the spatial patterns in the signal segments and the LSTM network-handles temporal dependencies. Such hybrid schemes confirm that the model remains robust while dealing with actual clinical data on a much better level. Another major DNN procedure greatly acclaimed is Autoencoders. These are used in unsupervised learning to observe increased fragments and detect aberrations in the cardiovascular dataset, i.e. the dataset is compressed into a lower-dimension representation followed by reconstruction by the model, patiently learning to teach model valuable circuitry. Here they are used for recognizing abnormal heart patterns, with one direct attribute being the identification of error in reconstruction of subjects. The VAE further extend this potential by the input of probabilistic modeling, making them valuable in recognizing latent or uncommon cardiac diseases [21].

Deep learning in medical image segmentation and classification is very important. U-Net is a special architectural layout to handle CNNs and sees application in segmenting heart structures like ventricles and arteries from MRI and CT scans. Accurate segmentation is of paramount importance for further inferences of cardiac function parameters like ejection fraction and stroke volume. Similarly, the ResNet (Residual Networks) helps in the training of very deep models by controlling the fixation of the vanishing gradient problem with skip connections, allowing very deep networks performing effectively in the cardiovascular image classification tasks. Another breakthrough comes with the use of the attention mechanism and transformers-based models in cardiovascular diagnosis. Attention mechanisms enable models to center on the bits from the input data they need to understand, which raises interpretability and performance. Vision Transformer (ViT) models are being looked at in medical image analysis, especially for heart disease detection, due to their capability to capture global dependencies while operating on image data. These models are especially versatile in catching subtle indicators that might be missed by traditional CNNs [22].

Prediction of cardiovascular disease is carried out in the area of predictive analytics through the adaptation of deep learning systems. Given the patient's age, sex, stress level, nutrition, fitness, lifestyle, and genetics, deep neural networks are used to come up with the likelihood of some cardiovascular cataclysms in the near future. Aligning personal treatment with the high-risk sequela determined using these models is supposed to benefit patients in every way. Wearable devices like smartwatches and fitness trackers will make the continuous monitoring of heart activity possible and cardiovascular health assessment easy. [23]Deep learning that has proved its mettle for cardiac diagnoses faces a host of challenges owing to the same arguments. From the bulk of obstacles, a significant one will be its enormous hunger for large, well-labeled datasets hard to get, more so in the medical field with all its privacy and constraints on data sharing. The interpretability of the model is another serious challenge here because deep learning models have that impenetrable, "black box" feel that makes it challenging for clinicians or just about anybody to know how these predictions are made. Explainable AI (XAI) techniques like SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations) are being harnessed in healthcare to induce transparency. In other words, computational efficiency is a major concern, and training the deep learning models would require a lot of processing power and memory resources. This keeps the deployments limited to healthcare environments where resources are in shortage; nevertheless, cloud computing and edge AI provides ways to mitigate some of these limitations by running these algorithms further and faster.

COMPARASION OF MACHINE LEARNING AND DEEP LEARNING TECNIQUES FOR CVD

Machine learning techniques for CVD prediction rely on manual feature engineering and work well with structured data, offering high interpretability and lower computational cost. Deep learning methods automatically extract complex features from large datasets like ECG and images, achieving higher accuracy but requiring more data, computing power, and reduced interpretability.

Table 2 Comparative Analysis of Machine Learning and Deep Learning-Based Heart Disease Prediction Models

Ref.	Methodology	Results	Limitations
[1]	ML models (DT, SVM, RF) with preprocessing & feature selection	RF achieved **96.2% accuracy**	Limited dataset size, lacks real-time validation
[2]	Clinical data + XGBoost, Logistic Regression	**AUC = 0.89**	Focused only on elderly patients, limited generalization
[3]	DL heart-sound classification using SMOTE & temporal models	Accuracy **~94%**	Requires high-quality audio data, computational cost
[4]	Retinal image analysis with ML regression	**R^2 > 0.80**	Limited to diabetic patients, imaging dependency
[5]	CNN-based cardiac MRI segmentation	Dice score **0.92**	Requires large annotated datasets
[6]	ML using inflammatory index (RF)	**AUC ≈ 0.85**	Population-specific cohorts
[7]	ML risk prediction with feature engineering	Accuracy **88–90%**	Moderate accuracy, feature sensitivity
[8]	EHR-based ML ensemble models	**AUC = 0.87**	Data imbalance and missing values
[9]	ML on routine health data	**AUC > 0.86**	Retrospective design limits real-time use
[10]	Hybrid ECG (classical + DL features)	Accuracy **95%**	Complexity in feature fusion
[11]	Gender-aware ML with bias evaluation	Accuracy **~91%**, bias ↓15%	Trade-off between fairness and accuracy

[12]	Ensemble ML (RF, GB) with optimization	Accuracy **95.6%**	High computational complexity
[13]	Clinical + behavioral ML models	**AUC = 0.88**	Behavioral data variability
[14]	Multimodal AI (clinical + imaging)	Accuracy **93%**	Integration complexity, data heterogeneity
[15]	Proteomic data + ML (RF)	**AUC = 0.90**	Expensive biomarker acquisition
[16]	DL-based MRI T1 mapping	**AUC = 0.91**	Requires advanced imaging equipment
[17]	ML classifiers (DT, NB)	Accuracy **92%**	Lower performance than DL models
[18]	DL for battery damage assessment	Accuracy **94%**	Not directly healthcare-related
[19]	ML comparison (screening methods)	Accuracy **93%**	Dataset dependency
[20]	Transformer-based DL model	Accuracy **96%**	High computational cost, requires large data

CHALLENGES AND RESEARCH GAPS

Several drawbacks make the real-world deployment of CVD prediction models using ML and DL highly difficult. The key issue here is the lack of highly useful and balanced datasets, as the medical data often contains values that are missing and a highly imbalanced setting between the healthy and the diseased cases. The particular rest sets models with highly reduced generalization capabilities. Another key issue is data privacy and security, because patient health records are extremely sensitive and cannot be efficiently shared among institutions for a large-scale cooperative partnership. Pretty much is the explain ability of black-box models in the case of complex models, especially in cases of deep learning techniques such as CNNs or LSTMs. What the clinicians want is complete prime facie opinion based sorting; this is something that most DL models do not do when they make predictions. Furthermore, data overfitting may lead to poor generalization when the model is trained on narrow or non-diverse datasets. The quite high computational require data handling power and memory, deterring usage of these models in real-time or resource-constrained healthcare environments. Evaluation frameworks the way forward are unconventional in AI, namely, there are no standard references of comparison between different datasets and evaluation of classifiers or predictive metrics.

Construction and not specific are probationary interest areas in explainable AI (XAI), methods on the management of imbalanced medical datasets in a better way, and design of real-time low-weight prediction models. The absence of integration of multi-modal data sources such as ECG, imaging, and genetic data readily compiles the prediction more accurately. From then on, privacy-preserving learning techniques like federated learning and deployable systems should be further investigated.

V. CONCLUSION AND FUTURE DIRECTIONS

Cardiovascular disease (CVD) is one of the major causes of death globally, and the significance of early prediction in diminishing this mortality rate and better patient outcomes is immense. This review discusses the ability of machine learning (ML) and deep learning (DL) technology to greatly improve the efficiency and accuracy in diagnosis and prediction of CVD. Logistic Regression, Support Vector Machines, Decision Trees, Random Forest, and ensemble methods are the popular classification methods that are well-liked for usage with structured clinical data. With good interpretability and computational efficiency, such models are appropriate for clinical support systems. However, despite this sophistication, quite a few challenges essentially still seem to be manifest. There are a number of roadblocks, including data imbalances, a lack of quality and sufficiently large medical datasets for AI model making, regulatory privacy laws, too much computational demand for real-world deployment, and a likely lack of trust on the part of an MHD. The non-interpretable nature of deep learning models may still be the determining deals breaker to gain trust. Future research needs to be focused upon interpretability such that explainable AI (XAI) models should be developed to make decisions more transparent.

REFERENCES

[1] H. F. El-Sofany, “Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques,” *IEEE Access*, vol. 12, pp. 106146–106160, 2024, doi: 10.1109/ACCESS.2024.3437181.
[2] Fu, Shihui, et al. "Machine learning-based prediction of three-year mortality in elderly inpatients with coronary artery disease combined with heart failure." International Journal of Medical Informatics (2026): 106307.
[3] Ameen, Asmaa, et al. "Advancing cardiovascular screening: deep learning-based heart-sound classification using SMOTE and temporal modeling." Scientific Reports (2026).
[4] Nguyen, Tuyet Thao, et al. "Factors Associated with Machine Learning-Based Predictions of Retinal Aging using Teleretinal Screening Images from Patients with Diabetes." Ophthalmology Science (2026): 101112.
[5] Chai, Wen-Yen, et al. "A Deep Learning-Based Fully Automated Cardiac MRI Segmentation Approach for Tetralogy of Fallot Patients." Journal of Magnetic Resonance Imaging 63.1 (2026): 264-276.

[6] Zhang, Pengyu, et al. "Association between systemic inflammatory response index and cardiovascular disease risk in rheumatoid arthritis patients: a machine learning based on US and Chinese cohorts." Clinical Rheumatology (2026): 1-12.
[7] Deng, Yifan, et al. "Development and validation of a machine learning-based predictive model for coronary heart disease risk in middle-aged and young adults." Heart & Lung 76 (2026): 113-123.
[8] Youssfi, Younès, et al. "Machine learning-based prediction of sudden cardiac death in the general population using electronic health record data." European Journal of Preventive Cardiology (2026): zwag087.
[9] Chuang, Shan-Shan, Fu-Tien Chiang, and Shih-Wei Lin. "Machine learning–based prediction of CAC-defined cardiovascular risk using routine health examination data: a retrospective cross-sectional study in a Taiwanese population." Journal of the Formosan Medical Association (2026).
[10] Karampour, Ramin Aghili, and Alireza Fallahi. "A Hybrid Approach for Early Diagnosis of Cardiac Ischemia from Electrocardiogram Signal: Combining Classical and Deep Learning-Based Features." Biomedical Signal Processing and Control 113 (2026): 108967.
[11] Coban, Onder, and Ayse Kartal. "Evaluating the Performance Disparity and the Role of Gender-Aware Approaches in Machine Learning Based Disease Detection." Computers and Electronics in Medicine 3.1 (2026): 1-10.
[12] Sovia, Rini, et al. "Robust Predictive Model for Heart Disease Diagnosis Using Advanced Machine Learning Techniques." Journal of Applied Data Sciences 7.1 (2026): 541-553.
[13] Çakmak, Abdulkadir, et al. "Machine Learning-Based Prediction of Coronary Artery Disease Using Clinical and Behavioral Data: A Comparative Study." Diagnostics 16.2 (2026): 318.
[14] Abdullayev, I. N., O. E. Jiyanbayev, and R. Kh Nasimov. "Early Detection and Prognostic Assessment of Ischemic Heart Diseases Based on Multimodal Artificial Intelligence Algorithms." Technical Science Integrated Research 2.1 (2026): 33-42.
[15] Xiang, Shizhen, et al. "Proteomic signatures and machine learning based-prediction models for cardiovascular risk in survivors of myocardial infarction." BMC Cardiovascular Disorders (2026).
[16] Birns, Anna Y., et al. "Development and Validation of Deep-Learning Based T1 Mapping Quantification and Risk Prediction in Cardiac Amyloidosis." Journal of Cardiovascular Magnetic Resonance 28 (2026).
[17] Singh, Shashank Kumar, et al. "Heart disease prediction system using machine learning." Artificial Intelligence and Sustainable Innovation. CRC Press, 2026. 749-755.
[18] Deng, Yunfei, Jiangtao Li, and Xianglin Huang. "Deep learning-based real-time damage assessment of lithium-ion batteries under dynamic impact." Journal of Power Sources 662 (2026): 238737.
[19] Hanmanop, Sasipatcha, et al. "A Data-Driven Framework for Diabetes Prediction: Machine Learning-Based Comparison of Invasive and Non-Invasive Screening." Journal of Current Science and Technology 16.1 (2026): 159-159.
[20] Arif, Saba, et al. "A Diagnosis Tool for Early Detection and Classification of Heart Disease in Individuals Using Transformer Mechanisms." Computer Methods and Programs in Biomedicine (2026): 109248.
[21] Hosseinabadi, Fatemeh, and Seyedhassan Sharifi. "Heart Disease Detection with Machine Learning Algorithms." (2026).
[22] Guzmán-Muñoz, Eduardo, et al. "Supervised Machine Learning-Based Prediction of In-Hospital Mortality Following Hip Fracture in Older Adults." Diagnostics 16.4 (2026): 612.
[23] Moirangthem, Miranda, et al. "Developing a Machine Learning-Based Cardiometabolic Disease Model for Predicting Liver Disease." AIJR Abstracts 8.1 (2026): 71-71.
[24] Zeng, Liangming, et al. "Machine learning-based estimation of trunk fat percentage and its association with cardiometabolic risk leveraging two large national cohorts." *Frontiers in Nutrition* 13 (2026): 1715570.
[25] Castrejón-Peralta, César, et al. "Artificial Intelligence and Deep Learning-Based Methods and Devices for Measuring Vital Signs: A Systematic Review." *Applied Sciences* 16.2 (2026): 1126.
[26] Jeong, Sihyeon, et al. "Deep learning-based multi-view echocardiographic framework for comprehensive diagnosis of pericardial disease." *European Heart Journal-Digital Health* 7.3 (2026): ztag028.
[27] Tabassum, Saadia, et al. "A machine learning-based framework for heart disease diagnosis using a comprehensive patient cohort." *Computers, Materials & Continua* 84.1 (2025): 1253-1278.
[28] Asim, Muhammad, et al. "A DEEP LEARNING-BASED APPROACH FOR HEART DISEASE DETECTION AND EARLY WARNING." *Contemporary Journal of Social Science Review* 3.2 (2025): 2856-2872.
[29] Patey, Olga, et al. "Prenatal detection of congenital heart defects using the deep learning-based image and video analysis: protocol for Clinical Artificial Intelligence in Fetal Echocardiography (CAIFE), an international multicentre multidisciplinary study." *BMJ open* 15.6 (2025): e101263.
[30] Raj, J. Relin Francis, et al. "A Deep Learning-Based Approach for Classifying Heart Disease Using CNN on Medical Imaging." *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*. IEEE, 2025.