Machine Learning in Cardiovascular Disease Detection: Approaches, Challenges, and Open Research Problems

Lawal Olayinka Olusegun[1*]; Alowolodu Olufunso Dayo[2]; Obe Olumide Olayinka[3]; Adetunmbi Adebayo Olusola[3]

[1]Department of Computer Science, Joseph Ayo Babalola University, Ikeji-Arakeji, Osun State, Nigeria
[2]Department of Cybersecurity, Federal University of Technology, Akure, Ondo State, Nigeria
[3]Department of Computer Science, Federal University of Technology, Akure, Ondo State, Nigeria

Corresponding Author: Lawal Olayinka Olusegun[1*]

Publication Date: 2026/05/09

Abstracts: Cardiovascular diseases (CVDs) contribute significantly to global health challenges, demanding prompt early detection and intervention to reduce harmful outcomes. The application of machine learning (ML) techniques offers a viable avenue to revolutionise conventional diagnostic methods in CVD detection. This review examined the ML algorithms for risk prediction and diagnosis of various CVDs, including arrhythmias, heart failure (HF), and coronary artery disease (CAD), emphasising diverse approaches, challenges, and avenues for future research. Machine learning models utilise complex patterns found within extensive Clinical datasets encompassing electronic health records (EHRs) and diagnostic imaging to improve early diagnosis and individualised treatment management strategies for affected individuals at risk for CVD diseases. This study defines open research problems requiring further investigation to enhance the efficacy and clinical applicability of ML models in combating cardiovascular diseases. This work recommends reliable and interpretable ML models, integration of heterogeneous data sources, collaborative efforts to address data scarcity, and advancements in model transparency and explainability. The integration of ML techniques holds great promise for advancing CVD detection and improving patient outcomes. Overcoming the challenges outlined in this review and examining opportunities for future study can unlock the full potential of ML in alleviating the global burden of CVD-related morbidity and mortality.

Keywords: *Machine Learning, Cardiovascular Disease Detection, Risk Prediction, Diagnosis, Challenges, Open Research Problems.*

How to Cite: Lawal Olayinka Olusegun; Alowolodu Olufunso Dayo; Obe Olumide Olayinka; Adetunmbi Adebayo Olusola (2026) Machine Learning in Cardiovascular Disease Detection: Approaches, Challenges, and Open Research Problems. *International Journal of Innovative Science and Research Technology*, 11(4), 3853-3862. https://doi.org/10.38124/ijisrt/26apr1555

I. INTRODUCTION

(CVDs) include a collection of disorders affecting blood vessels and the heart: (CAD), (HF), and arrhythmias and these diseases are one of the leading causes of morbidity and mortality worldwide (WHO, 2021), necessitating ongoing advancements in both understanding and managing these conditions (WHO, 2021). Figure 1 summarises the major types of cardiovascular disease.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1555

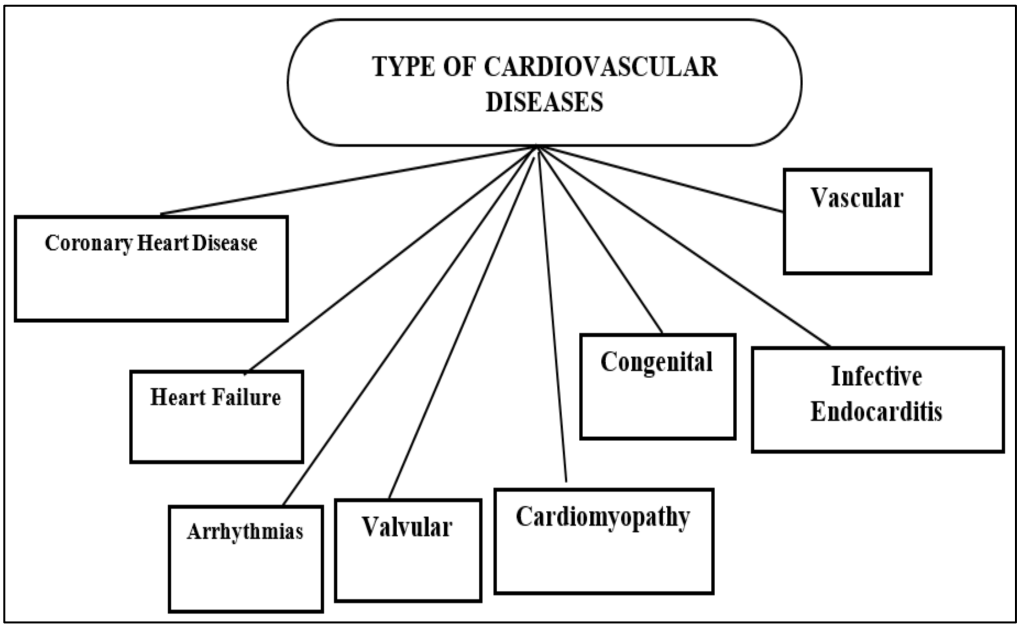

Fig 1 Major Types of Cardiovascular Disease

- *Categories of Cardiovascular Conditions*

- *CAD:*

Occurs when the arteries, responsible for supplying oxygen and nutrients to the heart, become stenosed or occluded due to atherosclerotic plaque buildup. This restricted blood flow can lead to chest pain (angina) or a heart attack (American Heart Association, 2024).

- *HF:*

Occurs when the heart is unable to pump blood effectively, reducing the supply of oxygen and nutrients to the body's tissues and organs, resulting in CAD, hypertension, and heart valve diseases (Merck Manual, 2024).

- *Arrhythmias:*

Arrhythmias refer to irregularities in the heart's rhythm, which can affect the pumping efficiency. Common forms include atrial fibrillation and ventricular tachycardia.

- *Valvular Heart Disease:*

Valvular issues are the abnormal functioning of one or more heart valves, causing stenosis (narrowing) or regurgitation (leakage). Examples are aortic stenosis and mitral regurgitation (American Heart Association, 2024).

- *Cardiomyopathy:*

This is a condition that targets the heart muscle making it unable to pump blood properly. These are dilated and hypertrophic cardiomyopathy which may be hereditary or acquired. (Smith et al., 2021).

- *Congenital Heart Disease:*

It means the structural defects present at birth that influence the valves, the walls of the heart, or blood vessels, which affects the normal functioning of the heart (Houyel & Meilhac, 2021).

- *Infective Endocarditis:*

It is a heart infection of the inner lining (endocardium) and the valves, which are usually caused by bacteria getting into the bloodstream and settling in the heart (Liesenborghs & Vanassche, 2020).

- *Vascular Disease:*

These disorders include problems with blood vessels in the body (arteries and veins) that interfere with normal blood flow and may cause complications such as an aneurysm or a clot (Merck Manual, 2024).

One of the most common types of CVD is coronary artery disease, which is defined by the constriction or obstructiveness of the coronary arteries as a result of the accumulation of plaque, which inhibits the blood flow to the heart muscle (American Heart Association, 2024) and may cause angina, myocardial infarction or sudden cardiac death. (Selak et al., 2020). Hypertension, hyperlipidemia, smoking, diabetes, obesity, and sedentary lifestyles are risk factors of CAD and are associated with atherosclerosis, the main pathological process of this disease. (Shetty, 2010).

The common management approaches to CAD include lifestyle changes, pharmacological therapy and in a few instances, surgical therapies like angioplasty and coronary bypass grafting. (Shetty, 2010).

Another significant CVD is heart failure which is caused by the inability of the heart to pump blood effectively, which is usually caused by a heart disease like CAD, high blood pressure, or diabetes. Symptoms are dyspnea, fatigue, and fluid retention, which may have a serious implication on the quality of life (Loades, 2018). It is diagnosed through clinical examination, imaging, including echocardiography, and biomarkers, like the B-type natriuretic peptide (BNP) levels, and treated with medications, including diuretics, ACE inhibitors, and beta-blockers as well as lifestyle changes and device therapy in severe cases (Yancy et al., 2017). Arrhythmias are caused by structural or ischemic changes in the heart, or other systemic diseases, and result in such symptoms as palpitations, dizziness, and syncope (Keller, 2014). Diagnostic procedures include electrocardiography, Holter monitoring, and electrophysiological studies, and treatment is based on the type of arrhythmia, and the simple cases are treated using drug therapy, and complex cases using invasive procedures like catheter ablation (Al-Khatib et al., 2018).

Ageing, changes in lifestyles and the escalating levels of comorbidities such as diabetes and obesity have contributed to the increase in the prevalence of CVDs in the world even with medical technology, diagnostic tools and therapeutic modalities. (Naghavi, et al., 2017). This tendency emphasizes the urgent need to develop more effective detection and management strategies, in particular, the ones that will enable it to detect the problem early and implement the individual treatment options. ML methods may evolve into a valuable tool in CVD detection and management, which is more capable of handling large volumes of data, forecasting the risk of disease, and recognising complicated patterns, which may otherwise not be detected by the conventional approaches (Shah et al., 2020). ML can revolutionise the process of CVDs management by increasing the accuracy of the diagnosis process, providing patients with more individual treatment plans, and finally by dealing with the global burden of CVDs.

The review analyzes different ML algorithms and methodologies applied to the problem of CVD, discussing the issues regarding the quality of data and model interpretability, scalability, and ethics. Moreover, the research paper identifies the limitations and strengths of existing ML methods in the CVD detection and suggests the future research directions to increase the clinical usefulness of such technologies. (Alkalah, 2016). A thorough review of the literature was performed with the help of such databases as PubMed, IEEE Xplore, and Google Scholar, and articles about the use of ML techniques in the detection of CVD published after 2010 were included in the review. The keywords were machine learning, cardiovascular disease, risk prediction, diagnosis and prognosis.

II. ML METHODS IN CVD DETECTION

The (ML) methods are increasingly becoming essential in (CVD) detection that can provide powerful applications in risk prediction, diagnosis and prognosis. Figure 2 represents the various learning forms of (ML) approaches:

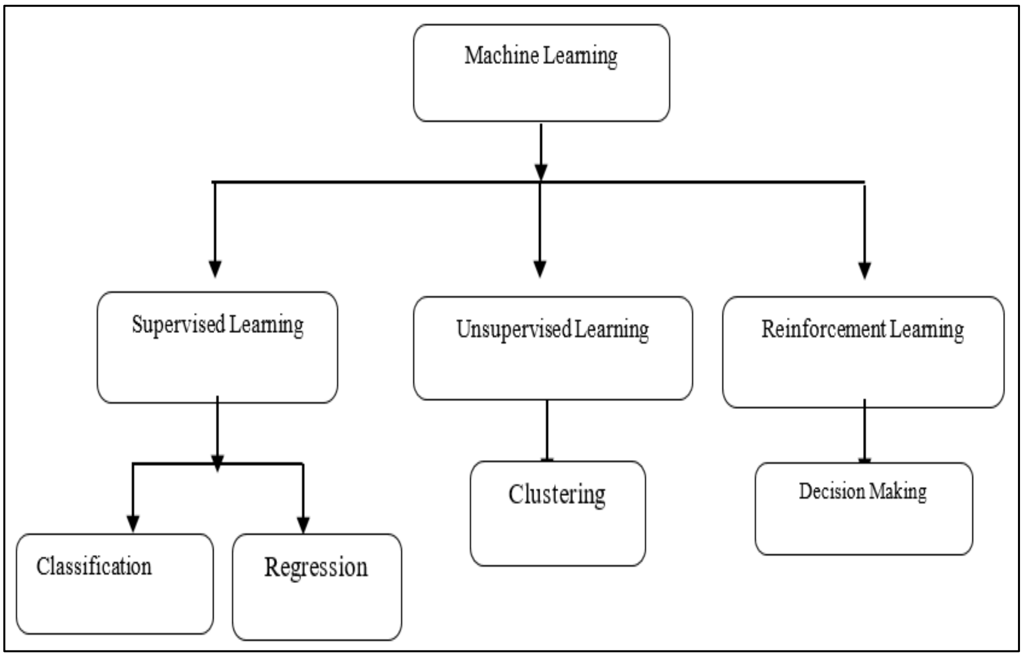

Fig 2 Types of Machine Learning

➢ *Supervised Learning (SL)*

(SL) It is a technique of model training using data that already has labels or correct answers. This includes such algorithms as Support Vector Machine (SVM), Decision Trees (DT), Random Forests (RF), and Neural Networks (NNs). They are applied to forecast the risk of developing CVD using information about patients, including their medical history, lifestyle, and clinical measurements. Algorithms, e.g., logistic regression (LR), SVM, and random forests, have been commonly applied in risk prediction based on the analysis of patient demographics, clinical features, and biomarkers to determine the risk of CVD. (Smith et al.,2021).Figure 3 gives details of supervised learning.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1555

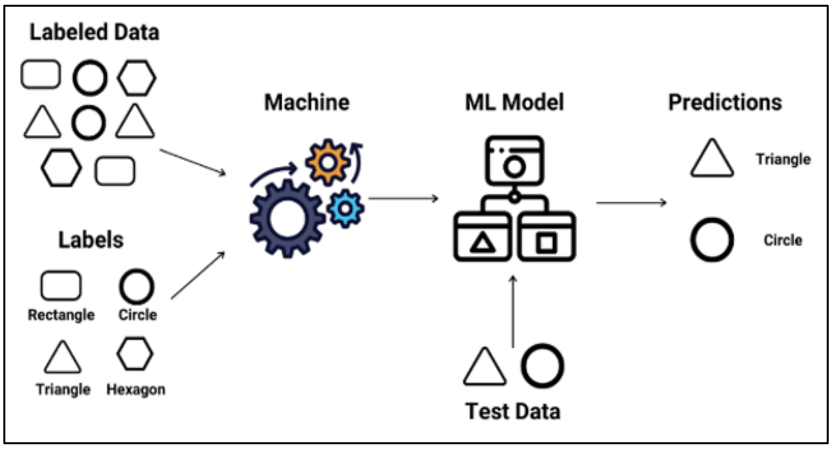

Fig 3 Supervised Learning

➢ *Unsupervised Learning Techniques (USL)*

It is a type of machine learning that works with unlabeled data to identify latent patterns in the data. To give an example, the K-means Clustering algorithm and Principal Component Analysis algorithm are just but a few algorithms that can be used to cluster patients together. The categorisation helps the researchers to understand better the different types of cardiovascular diseases (CVD) like clustering, anomaly detection that is also vital in exploring large cardiovascular datasets. Unsupervised Learning is shown in figure 4. The methods identify the groups of patients and determine outliers, which can be used to determine subtypes of disease and discover abnormal patterns of possible cardiovascular events early (Khan and Zhao, 2021).The clustering algorithms such as K-means and hierarchical clustering group patients with common features, which helps to comprehend the heterogeneity of diseases and personalized treatment (Wang et al., 2023). Meanwhile, anomaly detection algorithms can assist in detecting the rare occurrences and abnormal physiological patterns and contribute to early interventions in high-risk patients (Li et al., 2020).

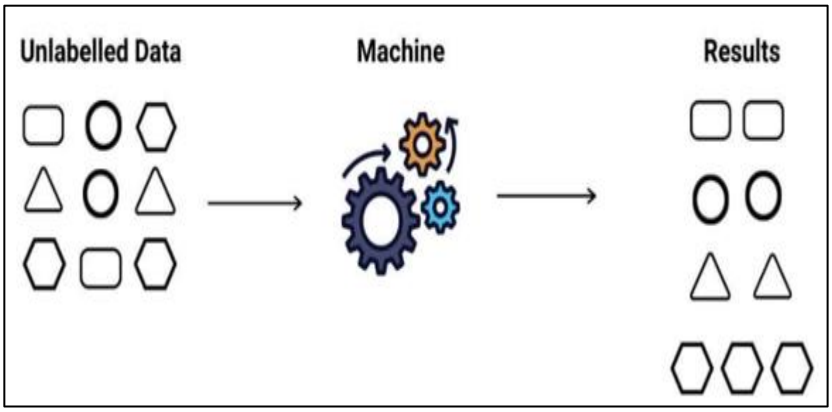

Fig 4 Unsupervised Learning

- *Risk Prediction:*

LR, SVMs, and random forests are supervised algorithms widely used in the CVD risk stratification to use patient data and predict the probability of developing a disease (Anderson & Brown, 2020). Logistic regression is easily interpretable and SVMs and random forests are useful when dealing with high-dimensional data and complicated relationships. (Doe & Nguyen, 2019).

- *Diagnosis:*

CNNs and RNNs have proven useful in the diagnosis of medical imaging and analysis of physiological data, respectively. CNNs interpret different imaging types to identify structural abnormalities and RNNs process sequential data, such as ECGs, to find the time-dependent patterns linked to CVD. (Wu et al., 2019).

- *Prognosis:*

ML models are based on longitudinal data to forecast disease progression and treatment outcomes to aid clinicians to determine high-risk patients and optimize care plans (García-Ordás et al., 2023).

Such techniques as clustering and anomaly detection provide clues on disease subtypes, and enable the early detection of unusual features in large datasets, enabling more personalised medicine strategies. (Zarzà et al., 2023).

Combining these ML techniques, healthcare specialists can increase the accuracy of the CVD detection and management, and improve patient outcomes. (Yu et al., 2021).Categories and examples of algorithms for SL, USL, and reinforcement learning are illustrated in Tables 2.1, 2.2, and 2.3, respectively, according to Domdouzis et al. (2021).

Table 2 Categories of Supervised Learning (Domdouzis et al, 2021).

Category	Algorithm Examples
Classification	Decision Trees, SVM, k-nearest Neighbours (k-NN), Naive Bayes, Logistic Regression, (NNs) (e.g., Multi-Layer Perceptron)
Regression	Linear Regression, Support Vector Regression (SVR), DT, NNs (e.g., Feedforward Neural Networks), (k-NN)
Ensemble Methods	RF, Gradient Boosting Machines (GBM), AdaBoost, Bagging, Stacking

Category	Algorithm Examples
Linear model	Multi-layer perceptron (MLP), Deep Learning, Logistic Regression, Linear Regression, Rigid Regression, Support Vector Regression (SVR), (SVM)
Non – Parametric Model	(k-NN),Kernel Regression (KR), Logical Regression (LR), Kanel Density Estimation(KDE)
Non – Metric model	Classification and Regression Tree (CART), (DT)
Parametric Model	Naive Bayes, Gaussian Discriminant Analysis (GDA), Hidden Markov Model(HMM), Probabilistic graphical Model
Ensemble Methods	Random Forest, Gradient Boosting Machines (GBM), AdaBoost, Bagging, Stacking

Table 3 Categories of Unsupervised Learning (Domdouzis et al, 2021).

Category	Algorithm Examples
Clustering	k-Means, Hierarchical Clustering, DBSCAN, Agglomerative Clustering, Spectral clustering
Density Estimation	Gaussian Mixture Models (GMM), Graphical Models
Dimensionality Reduction	Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (t-SNE), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Singular Value Decomposition (SVD)

Table 4 Categories of Reinforcement Learning (Domdouzis et al, 2021).

Category	Algorithm Examples
Value-Based Methods	Q-Learning, Deep Q-Networks (DQN), Double DQN
Policy-Based Methods	REINFORCE, Trust Region Policy Optimisation (TRPO), Proximal Policy Optimisation (PPO)
Actor-Critic Methods	Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C), Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC)
Model-Based Methods	Model Predictive Control (MPC), Dyna-Q
Multi-Agent Reinforcement Learning	Multi-Agent Deep Deterministic Policy Gradient (MADDPG), Cooperative Multi-Agent Reinforcement Learning (CMARL)

➢ *Deep Learning (DL) Models,*

DL is a part of (ML) that utilises NNs with many layers to process data. (CNNs) They are especially good at analysing images, such as echocardiograms (heart ultrasound images). However, RNNs are designed for handling time-series data, like electrocardiograms (ECGs), which track heart activity over time. Particularly, CNNs and RNNs are highly effective for automated feature extraction and classification in medical imaging and physiological signal analysis, respectively (Saxena & Paul, 2020). CNNs excel in detecting CVD-related abnormalities from imaging modalities like X-rays and MRI, while RNNs analyse time-series data such as electrocardiograms (ECG) for anomaly detection (Yu et al., 2021).

➢ *Ensemble Methods*

Ensemble methods are techniques that combine several models to enhance overall performance. There are two popular methods: Bagging and Boosting. These techniques are useful in enhancing accuracy of cardiovascular disease (CVD) detection models by reducing errors, namely, variance (unreliability of predictions) and bias (bias).

III. DATA COLLECTION, PREPROCESSING AND APPLICATIONS OF ML IN CVD DETECTION

➢ *Data Collection*

Machine learning (ML) models applied to cardiovascular disease (CVD) detection are based on data collected by various sources, such as (EHRs), wearable devices (such as a fitness tracker), and medical imaging. In order to make the data effective in the analysis, some important preprocessing procedures have to be followed including the normalisation of data values, repairing or eliminating missing data and the choice of the relevant features. It is important to use high-quality and diverse datasets to train powerful and robust ML models.

Machine learning (ML) in cardiovascular disease (CVD) detection is quite a versatile field, with solutions to various problems being presented, which eventually improves the care of patients. These applications are used in risk prediction, diagnosis and prognosis and combine powerful algorithms to analyse various datasets and derive actionable information. (Stark et al., 2024; Zhang et al., 2023).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1555

In predicting risks, ML algorithms are used to predict the probability of a patient to develop CVDs in a given time frame based on patient-specific data. The inclusion of variables like demographic features, clinical history, lifestyle behaviours and biological indicators, with algorithms to define risk and assist healthcare providers to identify people who are at greater risk of CVD. (Anderson & Brown, 2020). The method allows the use of early intervention measures and increased surveillance of high-risk people and eventually results in the successful prevention and management of cardiovascular diseases. (Colizzi et al., 2020).

ML models are also crucial in the diagnosis of the medical imaging data in identifying the abnormality and lesions of CVD. These models aid clinicians in making reliable diagnoses and treatment choices by automatically examining imaging data such as plaque development in the coronary arteries, cardiac performance, as well as tissue makeup through X-rays, MRI scans, and CT scans (Chen et al., 2021). Moreover, with the help of ML methods, it is possible to unite various imaging modalities and clinical data sources and conduct a holistic evaluation of CVDs and enhance the quality of diagnosis (Krittanawong et al., 2020).

To predict disease progression, treatment outcomes and future cardiovascular events, ML algorithms use longitudinal patient data to predict prognosis (Smith et al.,2021). Through the analysis of time-dependent trends in physiological parameters, disease markers, and response to treatments, such algorithms help clinicians to identify the patients at increased risk of unfavourable outcomes and design the individualised treatment plans (García-Ordás et al., 2023). This predictive feature enables medical professionals to manage patients in a more optimal way, better resource allocation, and positive patient outcomes due to timely intervention and prevention (Ski et al., 2023).

Generally, the uses of ML in CVD detection are boundless potentials to revolutionize the cardiovascular healthcare. With the ability to predict possible risks, diagnose accurately, and create individualised prognosis through the power of sophisticated algorithms to analyse the intricate datasets, ML facilitates better patient care and patient outcomes in the cardiovascular disease battle. (Smith et al.,2021: Zhang et al., 2023).

➤ *Evaluating (ML) Models:*

Involves using specific metrics to measure their performance. Main system of measurement include:

- Accuracy: The overall percentage of correct predictions.
- Precision: The percentage of true positive predictions among all positive predictions made.
- Recall: The percentage of true positive predictions among all actual positive cases.
- F1-Score: The harmonic mean of precision and recall, providing a balance between the two.
- ROC-AUC: An indicator of the model discrimination between classes, according to the receiver operating characteristic curve.

The correct metrics should be selected, particularly when working with uneven datasets because certain metrics may provide false outcomes and may not be reflective of the model performance.

➤ *Challenges in ML-based CVD Detection*

Difficulties with machine learning (ML)-based cardiovascular disease (CVD) detection represent a great setback despite the encouraging results of such methods. Among these issues are data availability/quality, interpretability of models, scalability, generalizability to different healthcare environments, ethical aspects of data privacy, and bias reduction, among others, which further complicate the development and implementation of ML-based CVD detection systems (Choi et al., 2017; Zhang et al., 2023).Challenges related to data in the context of cardiovascular disease (CVD). include:

➤ *Data-Related Challenges*

- *Data Privacy and Security:*

It is important to safeguard confidential health data. The quality and availability of data are inherent issues in the detection of CVD using ML. To solve these problems and provide credible model predictions, efficient data preprocessing and curation methods are needed (Rangineni, 2023).

- *Imbalanced Datasets:*

Positive cases (e.g., patients with CVD) are often significantly fewer than negative cases (e.g., healthy individuals), and it is challenging to train effective models. Medical data is usually characterized by missing data, data imbalance, and inconsistencies, which undermine the performance and overall generalizability of the ML models (Raji et al., 2019).

- *Variability in Data Sources:*

Variations in the data gathering of different sources may affect the efficiency of the models to be applied in the analysis.

➤ *Model-Related Challenges*

- *Overfitting:*

This happens when a model fits the training data overfitting it, with noise, and anomalies, making it unaffected by new, unseen data.

- *Underfitting:*

It occurs when a model is overly simplistic and does not represent the underlying patterns or relationships in the data and thus performs poorly on both training and unseen data.

- *Interpretability and Explainability:*

To apply (ML) models in clinical practice, health care practitioners need to know how these models arrive at their predictions. The results of the model should be trusted. Another challenge that is critical is the interpretability of the ML models especially the deep learning architectures. DL models can be considered black-box systems, which can be hard to comprehend the decision-making process, and therefore may require interpretable ML models to gain clinician and patient trust and acceptance (Alzubaidi et al., 2023)

- *Generalization to Diverse Populations:*

Another important issue is scalability in a variety of healthcare settings. The training of ML models based on data of one institution or population might not be highly accurate in other institutions because of clinical practices, patient demographics, and data collection procedures (Aggarwal et al., 2021). The process of scalability demands the creation of transferable models and data sharing and standardization to coordinate the performance in various healthcare settings (Raji et al., 2019).

It is important the models perform well in various groups of people so that they can give accurate predictions, irrespective of the differences in demographics. This poses a major problem in the creation of powerful ML solutions.

The ethical aspects of the creation and implementation of ML-based CVD detection systems are critical. Ethical patient data-handling practices are essential to guarantee patient privacy and confidentiality without affecting the usefulness and accessibility of the data (Piano, 2020). Moreover, it is important to address bias in training data and model results to establish fair healthcare delivery and decrease disparities in the diagnosis and treatment of CVD with the help of ML-posed detection systems (Obermeyer et al., 2019) The management of these ethical issues is crucial to achieve trust and acceptance of the ML-based CVD detection systems by the healthcare professionals and patients. (Siala & Wang, 2022).

Despite the potential of ML techniques to improve the detection and management of CVDs, the data quality, interpretability, scalability, and ethics problem needs to be addressed to realize the potential of this technique in clinical practice. Having conquered these challenges, ML-based approaches will be able to revolutionize the detection of CVD, and contribute to more efficient and equitable health services (Topol, 2019).

➢ *Implementation Challenges*

The implementation of machine learning (ML) models in the clinical workflows has a number of challenges. Care should be taken to ensure that such models are simple to operate and that they can be incorporated into the current healthcare systems. Critical regulatory and ethical concerns also have to be tackled, such as the necessity to provide fairness and exclude biases within the algorithms. Moreover, the implementation of ML solutions in healthcare settings might be hindered by the lack of financial and resource resources.

IV. OPEN RESEARCH PROBLEM

Even though there has been advancement in research, there are a number of challenges. First, we should have a more varied and representative data to ensure that models are applicable to a broad population. Second, one should develop simple to understand and explain models because it will assist in building trust among clinicians and motivate their use. Third, using various data, such as electronic health records (EHR) and imaging data, can provide a more complete picture of the health of a patient. We also require measures that will enable real-time and continuous patient monitoring on how to be proactive in managing cardiovascular disease (CVD). To promote fairness and equity in healthcare, it is important to eliminate biases in machine learning models.

There are a number of open research questions in machine learning (ML)-based cardiovascular disease (CVD) detection that require additional research to enhance the performance and clinical relevance of ML models in this area. These research issues include different facets, such as the development of the model, data management, transparency, and privacy considerations (Smith et al.,2021).

To start with, robust and interpretable ML models to detect CVD need to be developed. Although ML algorithms have demonstrated their potential to analyse complex medical data, the reliability and interpretability of model predictions is critical in their implementation in the clinical environment. Interpretable models give information on the forces behind the predictions, which improves trust and knowledge among healthcare providers and patients. (Frasca et al., 2024).

The other important issue in ML-based CVD detection is the incorporation of diverse data sources. Healthcare data are usually shared in various systems and forms, which may include electronic health records, medical imaging, genomics, and wearable devices. By creating methods to combine and align these various sources of data, it is possible to enhance the comprehensiveness and validity of ML models to detect CVD (Johnson et al., 2023).

The issue of data scarcity can be addressed by collaborative efforts to improve the development of ML-based CVD detection. Availability of high-volume and heterogeneous data is essential to train strong and generalizable ML models. Data sharing and pooling through collaborative efforts of various healthcare institutions and research organisations can be used to overcome the constraints of single data silos. (Chen et al., 2021)

Another research challenge is to enhance the transparency of models. The more complex the ML models are, the harder it is to understand how they reason their decisions. Clinicians can interpret and trust the output of the ML algorithms when they are presented in transparent models that explain their predictions, leading to a higher uptake of the algorithms in clinical decision-making. (Frasca et al., 2024).

One promising area of research on ML-based CVD detection is the investigation of federated learning as a privacy-preserving training of distributed healthcare data. Federated learning is a method to train ML models using data that is distributed to more than one device or institution,

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1555

without centralised data aggregation, and overcomes privacy issues related to data sharing. By investigating the possibilities of federated learning and its usefulness in the environment related to CVD detection, one can contribute to the creation of privacy-sensitive ML solutions when applied to healthcare. (Stark et al., 2024; Zhang et al., 2023).

It is important to address open research questions, including creating interpretable models, heterogeneous data sources, data scarcity, model transparency, and research into privacy-preserving methods, like federated learning, to advance ML-based CVD detection. Overcoming these obstacles, researchers can help to create more efficient and reliable ML models to enhance the detection and management of cardiovascular diseases. (Smith et al.,2021).

V. CONCLUSION

Machine learning has the potential to significantly enhance the process of cardiovascular diseases detection by providing more precise and effective diagnostics tools. Nevertheless, it continues to be associated with challenges and research questions to be addressed to realize this potential to the fullest. Continuous research and development and cooperation are essential in advancing in this field and improving the patient outcomes. The application of machine learning (ML) methods in the detection of cardiovascular disease (CVD) has a major potential in enhancing patient outcomes and healthcare delivery. ML algorithms have the potential to process complicated medical data to support the early detection, risk forecast, diagnosis, prognosis, and customized treatment plans of CVDs. Nevertheless, various challenges and gaps in research need to be mitigated in order to maximize the advantages of ML in the fight against CVDs and help to lower morbidity and mortality rates globally. The main concerns are an increase in the quality of data, better interpretability of ML models, their scalability to a variety of healthcare settings, and the issue of ethical concerns. Data quality could be enhanced by it with improved preprocessing, and understandable, transparent ML models could establish trust between healthcare providers and patients. Data sharing and model standardisation should result in scalability, but ethical considerations, including data privacy, security, and bias mitigation, should be considered in order to implement ML responsibly.

RECOMMENDATIONS FOR FUTURE RESEARCH

The concept of federated learning and transfer learning is a recent development in the field of machine learning (ML) and can be applied to improve cardiovascular diseases (CVD) detection. Researchers, health care practitioners and policymakers must work together to ensure that the benefits of ML in healthcare are maximized. To improve on this area, there is a need to conduct continuous research and development in order to improve patient outcomes. Future research should take into account how to create interpretable ML models, the simple means of explaining their predictions, integrating different data sources to enhance model performance, and exploring privacy-conserving strategies, such as federated learning, to utilize distributed healthcare data to train ML models without compromising patient privacy. To address these challenges, it is important to collaborate and work on multidisciplinary teams to make the clinical applicability of ML-based CVD detection systems more effective.

Moreover, data infrastructure, education, and training can be invested in to empower healthcare professionals to utilize the ML technologies effectively in the detection and management of CVDs. As these issues are resolved and the research directions are followed, ML will be capable of helping to achieve a more substantial improvement in CVD detection, management, and prevention, especially in terms of the improved patient outcomes and the reduced healthcare burden worldwide.

REFERENCES

[1]. Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S. W., Karthikesalingam, A., King, D., Ashrafian, H., & Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. Npj Digital Medicine, 4(1). https://doi.org/10.1038/s41746-021-00438-z

[2]. Al-Khatib, S. M., Stevenson, W. G., Ackerman, M. J., Bryant, W. J., Callans, D. J., Curtis, A. B., Deal, B. J., Dickfeld, T., Field, M. E., Fonarow, G. C., Gillis, A. M., Granger, C. B., Hammill, S. C., Hlatky, M. A., Joglar, J. A., Kay, G. N., Matlock, D. D., Myerburg, R. J., & Page, R. L. (2018). 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Hea. Journal of the American College of Cardiology, 72(14), e91–e220. https://doi.org/10.1016/j.jacc.2017.10.054

[3]. Alkalah, C. (2016). 済無No Title No Title No Title. High-Performance Medicine: The Convergence of Human and Artificial Intelligence. Nature Medicine, 19(5), 1–23.

[4]. Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data, 10(1). https://doi.org/10.1186/s40537-023-00727-2

[5]. Anderson, J., & Brown, R. (2020). Supervised learning methods for cardiovascular risk prediction. Journal of Cardiovascular Research, 32(4), 123–134.

[6]. Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Healthcare. Annual Review of Biomedical Data Science, 4, 123–144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

[7]. Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the

American Medical Informatics Association, 24(2), 361–370. https://doi.org/10.1093/jamia/ocw112
[8]. Colizzi, M., Lasalvia, A., & Ruggeri, M. (2020). Prevention and early intervention in youth mental health: Is it time for a multidisciplinary and trans-diagnostic model for care? International Journal of Mental Health Systems, 14(1), 1–14. https://doi.org/10.1186/s13033-020-00356-9
[9]. De Zarzà, I., de Curtò, J., Hernández-Orallo, E., & Calafate, C. T. (2023). Cascading and Ensemble Techniques in Deep Learning. Electronics (Switzerland), 12(15), 1–18. https://doi.org/10.3390/electronics12153354
[10]. Doe, A., Smith, L., & Nguyen, T. (2019). Advanced machine learning techniques in heart disease detection. IEEE Transactions on Biomedical Engineering, 62(2), 678–690.
[11]. Domdouzis, K., Lake, P., Crowther, P. (2021). Graph Databases. In the Concise Guide to Databases. SpringerBriefs in Applied Sciences and Technology, 223–235.
[12]. Frasca, M., La Torre, D., Pravettoni, G., & Cutica, I. (2024). Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. Discover Artificial Intelligence, 4(1). https://doi.org/10.1007/s44163-024-00114-7
[13]. García-Ordás, M. T., Bayón-Gutiérrez, M., Benavides, C., Aveleira-Mata, J., & Benítez-Andrades, J. A. (2023). Heart disease risk prediction using deep learning techniques with feature augmentation. Multimedia Tools and Applications, 82(20), 31759–31773. https://doi.org/10.1007/s11042-023-14817-z
[14]. Houyel, L., & Meilhac, S. M. (2021). Heart Development and Congenital Structural Heart Defects. 257–284.
[15]. Keller, K. (2014). National Heart, Lung, and Blood Institute. Encyclopedia of Obesity. https://doi.org/10.4135/9781412963862.n312
[16]. Khan, M., & Zhao, Y. (2021). Clustering algorithms for analysing cardiovascular disease datasets. International Journal of Data Science, 5(1), 32–45.
[17]. Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H. J., Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J. L., & Tang, W. H. W. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific Reports, 10(1), 1–11. https://doi.org/10.1038/s41598-020-72685-1
[18]. Li, H., Wang, X., Liu, C., Zeng, Q., Zheng, Y., Chu, X., Yao, L., Wang, J., Jiao, Y., & Karmakar, C. (2020). A fusion framework based on multi-domain features and deep learning features of the phonocardiogram for coronary artery disease detection. Computers in Biology and Medicine, 120(March), 103733. https://doi.org/10.1016/j.compbiomed.2020.103733
[19]. Liesenborghs, L., & Vanassche, T. (2020). Coagulation : At the heart of infective endocarditis. Journal of Thrombosis and Haemostasis, 18(5), 995–1008. https://doi.org/10.1111/jth.14736
[20]. Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. Humanities and Social Sciences Communications, 7(1), 1–7. https://doi.org/10.1057/s41599-020-0501-9
[21]. Loades, J. (2018). Understanding heart failure. Practice Nurse, 48(6), 25–30. https://doi.org/10.4172/2324-8602.1000296
[22]. Mahmud, M., Kaiser, M. S., McGinnity, T. M., & Hussain, A. (2021). Deep Learning in Mining Biological Data. In Cognitive Computation (Vol. 13, Issue 1). Springer US. https://doi.org/10.1007/s12559-020-09773-x
[23]. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342
[24]. Raji, R. A., Rashid, S., & Ishak, S. (2019). The mediating effect of brand image on the relationships between social media advertising content, sales promotion content and behavioural intention. Journal of Research in Interactive Marketing, 13(3), 302–330. https://doi.org/10.1108/JRIM-01-2018-0004
[25]. Rangineni, S. (2023). An Analysis of Data Quality Requirements for Machine Learning Development Pipelines Frameworks. International Journal of Computer Trends and Technology, 71(8), 16–27. https://doi.org/10.14445/22312803/ijctt-v71i8p103
[26]. Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y., & Patra, J. (2009). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the global burden of disease study 2016. The Lancet, 373(9682), 2223.
[27]. Saxena, S., & Paul, S. (2020). Deep learning applications in medical imaging. Deep Learning Applications in Medical Imaging, 27(3), 1–274. https://doi.org/10.4018/978-1-7998-5071-7
[28]. Selak, V., Poppe, K., Grey, C., Mehta, S., Winter-Smith, J., Jackson, R., Wells, S., Exeter, D., Kerr, A., Riddell, T., & Harwood, M. (2020). Ethnic differences in cardiovascular risk profiles among 475,241 adults in primary care in Aotearoa, New Zealand. New Zealand Medical Journal, 133(1521), 14–27.
[29]. Shah, R. U., Steyerberg, E. W., Windecker, S., & Holzmann, M. J. (2020). Machine learning in cardiovascular disease prediction: algorithms and practices. Nature Reviews Cardiology. 17(6), 330–342.
[30]. Shetty, P. (2010). Nutrition transition and its health outcomes. Indian Journal of Paediatrics, 77(3), 252–255.
[31]. Siala, H., & Wang, Y. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Social Science and Medicine, 296(June 2021), 114782. https://doi.org/10.1016/j.socscimed.2022.114782
[32]. Ski, C. F., Cartledge, S., Foldager, D., Thompson, D. R., Fredericks, S., Ekman, I., & Hendriks, J. M. (2023). Integrated care in cardiovascular disease: a statement of the Association of Cardiovascular Nursing and Allied Professions of the European

Society of Cardiology. European Journal of Cardiovascular Nursing, 22(5), E39–E46. https://doi.org/10.1093/eurjcn/zvad009
[33]. Smith, J., Brown, L., & Green, T. (2021). Advanced machine learning models in healthcare: Current challenges and future directions. Health Information Science and Systems, 9(1), 47–55. https://doi.org/10.1007/s13755-021-00327-0
[34]. Smith, P., Johnson, Q., & Lee, M. (2021). Machine learning approaches in predicting cardiovascular events. Circulation: Cardiovascular Quality and Outcomes, Https://Doi.Org/10.1161/CIRCOUTCOMES.120.007895, 14(6).
[35]. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
[36]. Wang, Y., Liu, L., & Wang, C. (2023). Trends in using deep learning algorithms in biomedical prediction systems. Frontiers in Neuroscience, 17(November), 1–32. https://doi.org/10.3389/fnins.2023.1256351
[37]. WHO. (2021). Cardiovascular diseases (CVDs) key facts. World Health Organisation, June 1–5. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[38]. Wu, C. C., Yeh, W. C., Hsu, W. D., Islam, M. M., Nguyen, P. A. (Alex), Poly, T. N., Wang, Y. C., Yang, H. C., & (Jack) Li, Y. C. (2019). Prediction of fatty liver disease using machine learning algorithms. Computer Methods and Programs in Biomedicine, 170, 23–29. https://doi.org/10.1016/j.cmpb.2018.12.032
[39]. Yancy, C. W., Jessup, M., Bozkurt, B., Butler, J., Casey, D. E., Colvin, M. M., Drazner, M. H., Filippatos, G. S., Fonarow, G. C., Givertz, M. M., Hollenberg, S. M., Lindenfeld, J. A., Masoudi, F. A., McBride, P. E., Peterson, P. N., Stevenson, L. W., & Westlake, C. (2017). 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Journal of the American College of Cardiology, 70(6), 776–803. https://doi.org/10.1016/j.jacc.2017.04.025
[40]. Yu, H., Yang, L. T., Zhang, Q., Armstrong, D., & Deen, M. J. (2021). Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives. Neurocomputing, 444, 92–110. https://doi.org/10.1016/j.neucom.2020.04.157
[41]. Zhang, A. P., Wang, G. X., Zhang, W., & Zhang, J. Y. (2023). Cardiovascular disease classification based on a multi-classification integrated model. Networks and Heterogeneous Media, 18(4), 1630–1656. https://doi.org/10.3934/nhm.2023071