Impact of Interpretability on Bias in ECG-Based Models for Cardiac Disease Detection

Saroj Kumari, Raghav Mehra

Department of Computer Engineering & Applications, Mangalayatan University, Aligarh, UP, India

Corresponding author: Saroj Kumari, Email: 20230220_saroj@mangalayatan.edu.in

Cardiac disease is a major health issue worldwide and a leading cause of mortality. Recent advancements in machine learning (ML) show potential for early heart disease detection using patient data and electrocardiograms (ECGs). Early identification can significantly reduce death rates and mitigate the impact of heart disease. Delayed and misdiagnosis therapeutic are two major issues with the traditional diagnostic approaches that can worsen diseases and shoot up the healthcare expenses. To cut these medical costs and avoid any faulty diagnostic ML techniques provides a promising solution in health care sector. The most non-invasive and affordable technique adopted for heart diagnostic by healthcare professionals is ECG. In medical ECG is widely used for the diagnosis, detection, and prevention of many cardiac problems. Disregard of many advantages, there are still issues, such as the lack of qualified cardiologists, comorbidities, and the resemblance of heart disease symptoms in ECG readings. Additionally, patient data and ECGs are often unbalanced, complicating the impartial performance of classical ML models. Traditional ECG diagnostics has improved by applying ML techniques and, helping doctors in interpreting complex cardiac disease processes and boosting computer-assisted treatments. Despite their potential, ML models face skepticism from medical professionals due to their "black box" nature and poor explainability. Many ML models for ECG-based heart diseases detection suffer from bias and lack transparency, raising ethical, legal, and social concerns. To address this, interpretable machine learning (IML) models can boost doctor confidence by providing evidence-based, understandable diagnoses. The detection of cardiac diseases using ECG data, investigating bias and fairness in IML models, and suggesting strategies to guarantee equitable model performance across heterogeneous patient populations are the main focus of this systematic literature review.

Keywords: Heart diseases, CVD, ECG, Machine learning, Bias, Interpretability, Explainability.

2025. In Sandeep Kumar & Kavita Sharma (eds.), *Computational Intelligence and Machine Learning*, 131–148. Computing & Intelligent Systems, SCRS, India. DOI: https://doi.org/10.56155/978-81-975670-5-6-11

1 Introduction

The heart is the most vital organ in our body, responsible for supplying blood to all other organs and tissues. But it's also prone to diseases and trauma, which can result in ailments like cardiovascular disease (CVD). A number of conditions included under CVD affects the heart and blood arteries. Obesity, inactiveness, smoking, high blood pressure, high cholesterol, poor diet, and inadequate nutrition are the major risk factors included for causing CVD by raise symptoms including weakness, exhaustion, and breathing difficulties. In 2019, the World Health Organization (WHO) [1] estimates that, cardiovascular disease takes 17.9 million lives worldwide. Among the CVD, heart attacks and strokes alone accounted for more than 85% of all deaths, of these deaths take place in bulk in low and middle-income nation.

As the population expands, providing cheap diagnoses becomes continuously more difficult and the number of heart disease patients rises, particularly in less developed nations like Bangladesh, India, and numerous African countries. Due to budgetary limitation and restricted availability of sufficient medical resources and infrastructure, the appropriate cardiac disease screening protocols in these areas is frequently absent. Besides these difficulties and challenges, it's important to remember that, with the early detection and right treatment, many cardiovascular diseases can be avoidable. For CVD timely and correct diagnosis is essential to save peoples live and by addressing associated risk factors and guaranteeing access to high quality medical diagnosis and treatment, we can drastically lower the extensiveness of CVD worldwide. To diagnose different cardiac conditions such as heart failure, arrhythmia, and myocardial infarction, Electrocardiograms (ECGs) are often used in clinical practice. Figure1 depicts the different methods physicians used to detect heart disease. The conventional techniques that were used earlier require physical examination for diagnosing heart disease, along with a medical history review, and a clinical evaluation. However, especially in the early stages of the CVD, these methods may not always be sufficient to accurately diagnose cardiac diseases. With the technical revolution, different methods namely invasive and noninvasive, are now available for early detection of CVD. The Non-invasive methods like echocardiograms, electrocardiograms (ECG), coronary computed tomography angiograms (CCTA), cardiac magnetic resonance imaging (MRI), and invasive techniques like blood testing are among the other clinical and more accurate approaches. One of the previously stated diagnostic techniques for early detection of cardiac disorders is electrocardiogram (ECG), which is affordable, non-invasive, adaptable, faster and easy to use. Consequently, an ECG-based diagnosis can be used to identify and early diagnose a number of cardiac diseases, namely- arrhythmia, pericardia, myocardia, electrolyte imbalances, and pulmonary diseases [2].

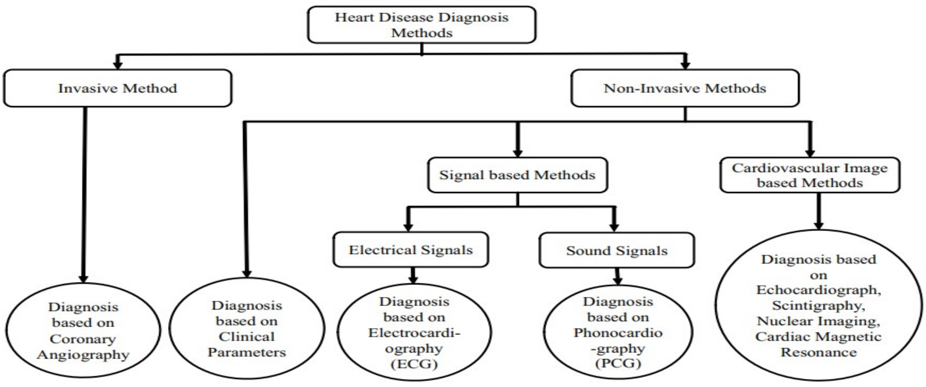

Figure 1. Methods for Diagnosis of Heart Diseases

Over more than 50 years, computer-assisted interpretation of ECGs has been playing a crucial role of clinical processes, improving the interpretation of physician. Conventional techniques use computer assistance to identify and quantify established ECG features, such as waves, segments, and intervals, which are then categorized as normal or abnormal according to predetermined guidelines. However, because of their reliance on inaccurate tracings and their use of antiquated classification techniques, these conventional models frequently have poor accuracy. With the application of data driven methods, especially Machine Learning(ML), the accuracy of automated cardiac disease identification has recently been improved by using a variety of physiological data, such as impedance cardiography (ICG) signals, magneto cardiography (MCG), heart sounds (HS), and electrocardiogram (ECG)[3].Physicians continue to employ ECG-based techniques to identify cardiac problems among them. Nonetheless, because cardiac disease presentations might resemble any other diseases on ECG signals, it could be very challenging to distinguish between them accurately. In addition to these challenges, the age, race, and overall physical state of the patients may all have an great impact on the differences in the ECG signal recording for the same medical condition. These difficulties will be overcome by the deployment of machine learning techniques, which have made it possible to analyze huge amounts of data and helping doctors to diagnose and predict cardiac disease with greater accuracy. The need for new breakthroughs and precision in Machine Learning models and algorithms has expanded dramatically due to the field's fast development and integration with the health care sector. The time and cost required for medical interpretation can be greatly reduced by development of diagnostic technologies, and enabling cardiologists to diagnose ECG recordings more rapidly and precisely. A number of ML based diagnostic tools have created over the past few decades to distinguish between distinct cardiac diseases [4]. Now days, computer have been employed to analyze ECG recordings to decrease the obstacles and assist clinicians in diagnosing cardiac conditions. Despite of computerized interpretation of ECG, many researches has shown that this strategy has significant obstacle as well as the limits of automated ECG interpretation [5]. Therefore, a doctor must still assess the final ECG interpretation to assure the diagnosis, even with efforts to increase the accuracy of automated ECG interpretation systems. However, progress has been hampered by the lack of readily available medical data. This challenge highlights the need for reliable models that can effectively utilize the data at hand to enable more accurate and efficient identification and diagnosis of cardiac diseases.

2 Review of Related Literature

Machine Learning has shown great potential in the medical diagnostics field, particularly in the interpretation of electrocardiogram data for the diagnosis of cardiac diseases. The dependability and equity of machine learning algorithms may be impacted by the serious questions about bias and fairness raised by the application of these technologies in healthcare sector. In this section, review of the literature looks at the study on fairness and bias in interpretable ECG-based cardiac diseases detection models, emphasizing important findings, research approaches, and knowledge gaps.

To categorized the many biases in cardiac diseases the following categories may be used in the detection models based on ECG that have been discovered during the development of machine learning algorithms designed for cardiac disease predictions:

Selection Bias: When the training data is not a true representation of the target population, selection bias arises. Numerous ECG datasets are skewed towards particular demographic groups, according to studies, which results in biased model performance. For example, models trained primarily on data from middle-aged Caucasian males frequently perform worse when applied to females or older persons, according to Ribeiro et al. [6].

Measurement Bias: When the data used to train, validate, or test models is not reflective of the real environment or population to which the model will be applied, systematic mistakes known as measurement bias develop in machine learning. A number of factors may contribute to this bias, which might produce models that are unjust, erroneous, or non-generalizable. According to Zhao and Li [7] the consistency and reliability of the data that is used to train machine learning algorithms might be affected by variability introduced in ECG recordings by various clinical settings and equipment.

Label Bias: In machine learning, label bias for electrocardiogram data refers to errors or inconsistent labelling practices, and it can significantly degrade the performance and reliability of machine learning models. There can have many factors and causes for this type of bias to appear and degrade performance of the model's capacity to reliably categorize or predict the diseases from ECG data. Li et al. [8] in their study, discussed how subjective interpretations by different cardiologists might lead to inconsistent annotations, which would add bias into the model. Since human opinions varies, typical for jobs requiring human interpretation. Bias may result from the representation and alteration of attributes. If some attributes are poorly scaled or more prone to measurement errors, they may have a disproportionate impact on the performance of model.

In this study, welook at the causes and effects of biases in the medical field in an effort to refute their validity and generalisability. Also this study will make an effort to give researchers and medical professional the necessary practical skills to recognize, comprehend, and decrease these biases while also promoting the creation of successful and equitable Al solutions for every patient by adhering to technological best practices. There are wide ranges of computational technologies that are integrated with the rapidly emerging field of medical to provide advanced clinical decision-support systems. The ML algorithms help medical practitioners to diagnose patients and design personalized treatment plans by interpreting complex medical data and producing predictions or interpretations [9]. Nevertheless, biases in these algorithms could lead to systematic errors that give preference to some groups over others, which is particularly concerning in clinical contexts [10]. The development process can take multiple phases where biases can make its appearance. Among these phases, data collecting, algorithm construction and refinement, testing and accuracy evaluation are included and lastly, applying the algorithms to real-world clinical scenarios.

Regarding the problems with the datasets used to construct the ML models, many data gathering flaws are reported in the literature. For example, the models' inability to address regional and ethnic inequalities is impacted by the absence of data for several geographic locations, which might result in unfair and biased predictions or even discrimination based on the individuals' particular features. Other types of problems with data collection could result from, for example, using devices made by different manufacturers with different preprocessing techniques and data handling strategies, or from limiting the sample to a particular segment of the overall CVD population (home care, in-hospital care, etc.), which increases the likelihood of over fitting and inadequate generalization. Even while top-tier databases with meticulously collected ECGs and well characterized individuals are commonly used to construct models, these models may perform poorly when applied to ECGs from regular clinical settings in the real world. The quality of the dataset must come first in order to get around these problems since poor quality datasets cannot be made up for by even the most sophisticated model tweaks. Consequently, external validation using multi-vendor ECG devices in geographically heterogeneous multicenter populations Thus, external validation in geographically heterogeneous multicenter populations using multi-vendor ECG systems would be essential for both increasing the model's acceptance and strengthening its generalizability performance. Therefore, to prevent bias in the predicted performance metrics, the individuals with a "atypical presentation" must be included in the representativeness of the group that the prediction model is targeting.

Keeping with dataset concerns, a sizable portion of studies (30%) have very small sample sizes (less than 100 patients), which restricts the performance of the model's and calls into question their dependability and generalizability. These datasets were often gathered in private settings with the intention of supporting the study's unique research concerns. Their modest size, however, could not account for the necessary variation among research participants to enable the model's acceptance and deployment in a larger population. ML approaches fail to generalize to unknown test data and overfit their performance to the training set when working with tiny datasets [11]. Ensuring that model decisions are interpretable, attaining generalization across heterogeneous populations, and incorporating models into clinical practice are among the issues associated with biased models. Keeping up with technical changes, resolving ethical problems, and conducting rigorous validation all add to the complexity of this field of study. Large accuracy frequently requires tiny sample quantities, according to systems. It turns out that ML algorithms don't work well with big sample sizes. On the other hand, the same classifier performs exceptionally well with unique features in small and effective ways, which has a big influence on predicted accuracy [12]. The reliability is questionable when data quality and diversity is small and when working with smaller or unbalanced datasets, it might be difficult to create reliable prediction models due to the lack of access to high-quality, diverse medical data, such as imaging, genetic information, and clinical records. Interpretability is still difficult to ensure that predictive models are interpretable, especially when using deep learning techniques. This is because complicated models can struggle to provide the clarity needed to explain the logic behind certain predictions, which is essential for clinical adoption [13]. Predictive models must be reliable and relevant in a variety of clinical and demographic scenarios. This presents a substantial difficulty in developing models that generalize successfully across varied patient groups and healthcare settings.

AI has the potential to worsen racial bias and healthcare and health inequities; therefore it's important to be conscious of this. The various uses of the ECG for risk classification may be quite sensitive to race. To guarantee the responsible application of AI in medicine, recommendations included external validation, the upkeep of varied data sets, regular subgroup reporting, and vigilant observation. For the detection of low LVEF(Left Ventricular Ejection Fraction) using the ECG across a wide range of racial/ethnic subgroups. This is not to say that the ECG is race invariant (in fact, a separate CNN was able to discern race using ECG features); rather, it suggests that the ECG features associated with ventricular dysfunction are race invariant[14]. ECG datasets frequently under represent specific demographics, such as women or ethnic minorities, Smith et al.[15] point out, which causes models to perform badly on these populations. Eltrass [16] suggested method for ECG multi-class classification has shown promise, but more research should be done on it by analyzing bigger ECG datasets involving a greater number of patients in order to distinguish between various heart disease classes, such as myocardial infarction and rhythm classes. Even if the suggested method works quite well, there are a few issues that need to be looked into further. The fact that there is frequently no adequate justification for the behaviors of the models during the final predictions is one of the possible disadvantages of machine learning based approach [17]. For example, a Deep Learning-based model has several hidden layers, yet most of the time it is hard to understand how each layer affects the final prediction. In addition, another possible issue is the ML algorithms' biased performance in favour of the majority class. A majority class, often referred to as an unbalanced dataset, arises when a data set has a maximum value in one class relative to other classes. Thus, concerns about the interpretability, biasness, and fairness of the ML-based models' performance persist. Saraswat et al. [18] stated that ecosystems has turned to digital wellness, with analytics-driven decision models that provide real-time forecasts and informatics assistance. To support this move, white box analytics is favoured over traditional black box models. As concerns about AI models' validity and interpretability have grown, AI models have changed to incorporate EXAI decision modules. In addition to enabling interpretability and model debugging, which improve performance by reducing bias, EXAI increases confidence in clinical procedures.

The absence of several publicly accessible, large-scale, and longitudinal databases containing real-life ECG values, in addition to a dearth of machine learning techniques that are both somewhat diversified and extremely accurate, have presented challenges [19-22]. A prevalent obstacle in identifying the frequency of acute diseases in biological datasets is the unequal distribution of classes, which frequently necessitates the use of innovative detection techniques [23]. Ensuring that the data is divided at random to prevent segments from the same subject from appearing in both the training and testing sets is vital. The latter would add a bias that would negatively impact the model's performance with unseen data. Furthermore, it is crucial to test the model using stringent internal and external validation methods in order to prevent falling into an optimistic bias trap regarding the predictive performance of the model. Moreover, the lack of a common framework for comparing model performance across institutions makes it difficult to develop an open platform that promotes the exchange of concepts, datasets, and pre-trained model weights. Mahajan et al. [24] state that data variability may result from differences in the settings and quality of ECG equipment throughout hospitals, which might affect the model's accuracy (measurement bias), caused by differences in the methods employed to collect the data. Nearly every machine learning methodology, including Deep Learning approaches, is currently being researched, as is the way by which the model arrived at its final predictions. The ML-based model is still unable to provide improved explanations for the model's interpretability and dependability, even in light of the recent offers of a number of explainable or interpretable techniques [25].

The capacity to understand, decipher, and draw conclusions from a predictive model's predictions is known as explainability. Establishing trust between patients and healthcare practitioners is essential in the healthcare industry. Explainability assist patients and healthcare professionals to providers better grasp the underlying assumptions of the cardiac disease prediction models and the factors they should consider. Patients are more minded to accept and heed the model's interpretation as a result of its transparency. As a result of explainability, clinical decision support systems are developed which aids medical practitioners in making well-informed clinical decisions. Understanding the theory and reasoning behind the model, clinicians can evaluate the validity, reliability, and dependability of its predictions. They can diagnose patient's overall health, and treat patients more accurately with the use of this information taking into account other clinical factors and interpreting the results in the context of the patient's overall health.

There needs to be open communication and a trust bond between doctors and other medical professionals for the applications of ML to be adopted and implemented in actual real world practice. Interpretable machine learning (ML) offers methods for comprehending and verifying the operation of the ML model, enabling stakeholders to comprehend the fundamentals of the choice and have faith in the former [26]. This gives ML algorithms a white-box quality and permits analytical transparency. In order to guarantee that prognostic models advance health equity, sensitive characteristics like socioeconomic variables at the person level must be specifically taken into account during the model's creation and validation process. In the absence of these factors, medical decision-makers would be forced to fill in health gaps by using their intuition and medical domain expertise, which might introduce more biases into the healthcare system. Furthermore, researchers should disclose subgroup-specific measures of prediction validity and create and verify algorithms utilizing data from varied populations. Prior to developing a model, particular consideration should be paid to the discovery, investigation, and appropriate discussion of societal and data biases. Algorithms that are biased may result from failing to take societal and data biases into account. Additional factors that should be properly investigated include differences in the frequencies of missing data and misclassification, both of which might skew results in a similar way [27]. For these reasons, it is essential that any algorithms be comprehensible and interpreted by all parties involved in order to detect and rectify any potential biases. Establishing crucial benchmarks and consistently observing clinically applied algorithms can also help identify which algorithms are operating equitably and, if not, what adjustments should be done to enhance their performance.

To make ML with ECG applications more useful and applicable in real-world medical settings, more researchis needed. Enhancing AI models' interpretability is crucial if we want medical professionals to be able to understand and trust the models' diagnosis recommendations. It is essential to create a uniform assessment procedure and dataset in order to provide reliable and consistent results throughout several studies [28]. By concentrating on these research areas, we may be able to remove the obstacles preventing ML from being widely used in ECG analysis and broaden its usage in a range of medical contexts. Furthermore, it is crucial to test the model using stringent internal and external validation methods in order to prevent falling into an optimistic bias trap regarding the predictive performance of the model. Moreover, the lack of a common framework for comparing model performance across institutions makes it difficult to develop an open platform that promotes the exchange of concepts, datasets, and pre-trained model weights. On the other hand, it can facilitate cooperation and remove what seem to be obstacles to institutional growth [29].

For heart disease, a leading cause of mortality globally, efficient detection methods are essential to better treatment and prevention. The conventional method of ECG interpretation, which necessitates specific expertise, is one of the main challenges. Machine learning (ML) is gaining traction and showing promise as a technology for developing and deploying intelligent systems in the healthcare industry. Patients can save money and time by using ML-based initiatives, which also provide early clinical assistance. AI applications may perform procedures, diagnostics, investigations, prognoses, and patient interpretations more precisely when they are integrated with machine learning models. This aids in the final conclusion-making process for physicians and radiologists.

3 Performance Caparison of ML Algorithms & Nature of Biasness

The identification and diagnosis of heart problems using ECG data has been the subject of much research in the literature on machine learning (ML) methods. These algorithms (as shown in Figure 2) identify a range of cardiac conditions, from myocardial infarction to arrhythmias, by utilizing the abundance of information included in ECG signals. Below is a detailed overview of some of the most often used machine learning techniques in this field:

Decision Tree and Random Forest: In order to diagnose different cardiac conditions healthcare sector uses two flexible machine learning methods namely Decision Tree and Random Forests for tasks like ECG analysis to identify cardiac problems. A non-parametric type of supervised machine learning known as a decision tree builds a hierarchical structure where each node representing a feature, the branches signifying decision rules, and the leaves providing predictions. The tree is constructed by recursively partitioning data depending on characteristics.

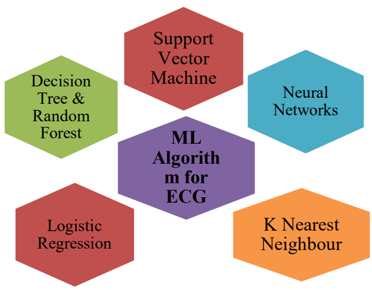

Figure 2. ML Algorithm commonly used for diagnosis method for Heart Diseases

Gini impurity and Information Gain used as metrics in order to optimize class purity in the resultant subsets. Overfitting in the trees is reduced by performed tree Pruning, the technique in which low impact nodes are eliminated. Decision trees can handle both numerical and categorical data and are easily interpretable without the need for preprocessing in order to capture non-linear connections. Small data differences can result in severe overfitting, though, if tree depth or pruning are not well controlled. Nevertheless, if tree depth or pruning are not well controlled, they are prone to overfitting, and even slight differences in the data can have a major impact on the structure of the tree. Multiple decision trees are built during training to increase robustness and generalizability in Random Forests, an ensemble learning approach. Every tree uses random feature sampling and is trained on a random subset of data called bagging. In terms of overall performance Random Forests outperform individual decision trees because they take average of predictions from multiple trees formed, which reduces overfitting and variation. They are effective at handling big, multidimensional datasets. However, when there are a lot of trees and data, training might be computationally demanding. Even if every decision tree may be understood separately, Random Forests' ensemble approach makes comprehension more difficult as a wbole. Because of their interpretability, both Decision Trees and Random Forests perform exceptionally well in ECG analysis for diagnosing cardiac conditions. in finding important characteristics (e.g., waveform, intervals) and categorizing ECG signals (e.g., normal vs. arrhythmia).

Support Vector Machine (SVM): Powerful supervised learning algorithm for outlier detection, regression analysis, and classification are support vector machines (SVMs). They are extensively used in many different domains, such as healthcare, where they perform well on jobs like deciphering heart conditions from ECG data. To find the optimal hyperplane in a high-dimensional space to partition data points into multiple classes, support vector machines maximize the separation between the closest points of different classes. Both linearly and non-linearly separable data can be handled by employing kernel functions to transform the input space into a higher-dimensional feature space. Support vector machines are strong against noise and perform well in predicting the border of a decision based on fresh data. In order to avoid overfitting, they employ a regularization parameter and perform well even in high-dimensional situations. SVM training, however, may be computationally expensive, particularly when dealing with larger datasets. Non-linear kernels might be difficult to interpret, even with their precision, in terms of the decision boundary and the relative relevance of various variables. Statistically significant and comprehensible predictions are essential in clinical settings, and support vector machines (SVMs) are useful in ECG analysis because they efficiently extract pertinent information to differentiate between different heart diseases.

Neural Networks: Deep learning models in particular have gained popularity as neural networks for ECG data analysis and heart disease prediction. Because these networks automatically train and extract characteristics from raw data, they perform very well with high-dimensional, complicated datasets such as ECG signals. Layers of linked neurons make up neural networks; input layers store raw data, hidden layers analyze it, and output layers provide predictions. Convolutional Neural Networks (CNNs) are used to recognize spatial patterns like P waves and QRS complexes; recurrent neural networks (RNNs) are used to capture temporal relationships like heart rate variability; and feed forward neural networks (FNNs) are used for fundamental tasks in ECG research. CNNs and RNNs are used in hybrid models to provide thorough feature extraction.

ECG signals are gathered from several sources, preprocessed to remove noise and normalize data, and then structured correctly in order to employ neural networks for ECG analysis. To optimize the network's weights, labeled ECG data is used in the design and training phases of the neural network architecture. By automatically identifying and extracting clinically meaningful characteristics, neural networks are able to anticipate disorders like myocardial infarction and arrhythmia. Metrics including accuracy, sensitivity, specificity, and AUC-ROC are used to assess the performance of the model. Neural networks may be tailored to different ECG datasets and cardiac disease types, manage high-dimensional data, and provide automated feature extraction. The requirement for sizable labeled datasets, computing requirements, and interpretability present difficulties in the Neural networks. Notwithstanding these obstacles, improvements in explainability and neural network methodology are making neural networks more useful in clinical settings for the diagnosis and monitoring of heart illness, including the integration of wearable technologies for the prediction of cardiac events in real time.

K- Nearest Neighbour: A straightforward, non-parametric, lazy learning technique for regression and classification is called K-Nearest Neighbours (KNN). Due to its simplicity of use and adaptability to tiny datasets, it is useful in electrocardiogram (ECG) analysis for the prediction of cardiac disorders.

KNN uses distance metrics such as Manhattan, Minkowski, and Euclidean to classify a data point in the feature space according to the dominant class of its "K" closest neighbours. The model's sensitivity to noise and decision limits is dependent on the value of 'k'. The obtained ECG signals undergo preprocessing (normalization, segmentation, noise reduction), after which features are retrieved either automatically or manually using techniques like PCA and wavelet transformations. KNN learns from training data, classifies fresh ECG samples by averaging the votes of 'K'neighbours, and assesses performance using k-fold ROC-AUC, accuracy, and sensitivity measures with k-fold cross-validation to prevent overfitting. Although KNN is simple enough to enable rapid prototyping and handling of complex data, it can be memory- and computationally-demanding; choosing the best 'K' and distance metric is important and frequently necessitates domain-specific knowledge; KNN is helpful in detecting abnormal ECG patterns, predicting heart diseases, and detecting arrhythmias. Despite these drawbacks, KNN is still an important tool in medical diagnostics, helping to improve patient outcomes and identify heart diseases.

Logistic Regression: In order to forecast cardiac illnesses and other associated disorders using ECG (electrocardiogram) studies, one useful statistical model for binary classification tasks is logistic regression. Logistic regression is a technique used to estimate the probability of a binary outcome, such as the presence or absence of a disease, based on one or more predictor variables, or characteristics.

It makes the assumption that there is a linear relationship between the outcome's log-odds and the predictor factors. It uses the Sigmoid Function, which the logistic regression model uses to convert predictions into probabilities and is also sometimes referred to as the logistic function:

$$\mathrm{P}(y=1|x)=\frac{1}{1+e^{-\beta 0+\beta 1x1+\beta 2x2+\cdots\ldots\ldots\beta nxn}} \qquad (1)$$

where : P(y=1|x) represents the probability of the positive class given predictor variables x.
$\beta 0, \beta 1, \beta 2, \ldots \ldots \beta n$are coefficients learned duringtraining.

The model predicts the positive class (e.g., presence of a cardiac condition) if the probability P(y=1|x) is greater than a threshold (typically 0.5) and the negative class otherwise.

Wearable technology and medical records are two common sources of data for ECG analysis. The data is first preprocessed by splitting it up into individual heartbeats, normalizing it, and removing noise. While methods like PCA or wavelet transformations automate feature extraction to decrease complexity, features like wave shapes and intervals are extracted manually. In order to predict binary outcomes, such as cardiac diseases, logistic regression models are built. The best coefficients are found by techniques such as maximum likelihood estimation. Metrics for assessing the model's generalizability include the F1 score, ROC-AUC, accuracy, sensitivity, specificity, and cross-validation. For complicated ECG data, the linear assumption of logistic regression may not always hold true, necessitating careful feature selection and management of unbalanced datasets. It facilitates longitudinal monitoring, diagnostic help for cardiac disease, and risk prediction Logistic regression is still a useful technique in medical diagnostics for ECG analysis, supporting clinical judgments and enhancing patient outcomes in cardiology, despite its presumptions and limits.

Table 1. Performance Evaluation & Comparison in History

Author	Algorithm Applied	Performance Evaluation Metrics	Result	Database
Nikhar, and Karandikar [30]	Decision Tree & Naïve Bayes classifiers	Information gain calculations	Decision Tree performs more accurately than the Naïve Bayes classifier	Cleveland Heart Disease dataset
Nashif, Raihan,	Random Forest,	Precision, Recall,	SVM outperform by	Cleveland Heart

Islam, and Imam [31]	Naïve Bayes classifiers, Neural Networks, Logistic Regression, SVM	F-1 score, Accuracy, Sensitivity, Specificity	others with accuracy level of 97.53%, along with corresponding sensitivity and specificity 97.50% and 94.94%.	Disease dataset &Statlog Heart Disease datase
Bhardwaj, Kundra, Gandhi, Kumar, Rehalia, and Gupta [32]	Random Forest Classifier	Precision, Recall, F-1 score	Accuracy of 89.4% with the default hyper parameter	UCI Repository
Jeyaganesan, Sathiya, Keerthana, &Aiyer [33]	Decision Tree, Logistic Regression, Random Forest, and Naive Bayes	Stability, Accuracy and Precision	Random Forest is most effective with an accuracy score of more than 90%	UCI Repository
Malavika et al. [34]	Decision Tree, Logistic Regression, Random Forest, Naive Bayes, &SVM	Confusion Matrix, per class accuracy and classification accuracy	Random Forest algorithm achieves the highest accuracy of 91.8% a	UCI Repository
Nayan et al. [35]	Decision Trees, KNN, Artificial Neural Networks ,Linear Discriminant Analysis, and Linear and quadratic SVM.	Specificity, Sensitivity, and Accuracy	ANN outperformed the other six algorithms in terms of prediction accuracy with 90% in all	Field Data
Ibrahim et al. [36]	Convolutional Neural Network and Recurrent Neural Network, XGBoost	Accuracy, F1 score, AUROC. Sensitivity, Specificity	XGBoost model has the highest accuracy of 97.5%	MIT-BIH Arrhythmia database
Jindal et al. [37]	KNN, Random Forest Classifier, and Logistic Regression	Accuracy and Performance using various performance metrics.	KNN is highest between the three algorithms with accuracy of 88.52%	UCI Repository
Sahoo, and Jeripothula [38]	SVM, Naïve Bayes, Logistic Regression, Decision Trees	Precision, Recall, F-1 score, Accuracy	SVM proved to be the most accurate, yielding results as high as 85.2%	UCI Repository
Taqdees, Akhtar, and Dawood [39]	KNN, Neural Networks, Decision Tree, and Naive Bayes and Random Forest	Confusion Matrix	Naive Bayes gives the highest accuracy which is 88%	UCI Repository
Younas [40]	Decision Tree, SVM, K-NN, Naive Bayes, Logistic Regression, and Random Forest	Confusion Matrix	Logistic regression achieved an accuracy of level of 86.89%.	UCI Repository
Ahmed [41]	Naïve Bayes, Decision Trees, SVM, Bagging & Boosting, and Random Forest	Precision, Recall, F1-Score, Support	Random Forest Classifier, predict cardiac disease with an accuracy of 89.4%.	UCI Repository

Bora, Gutta, and Hadaegh [42]	Logistic Regression, Naïve Bayes, SVM, KNN, Random Forest, Extreme Gradient Boost	Precision, Recall, F1-Score	Random Forest yielding the best accuracy of 93.31%.	UCI Repository& Kaggle Data set
Dixit, Mohan and Terni [43]	KNN, Random Forest, and Logistic Regression	Precision, Recall, F1-Score, Support	Logistic Regression algorithm, which has an accuracy score of 89%	UCI Repository
Guleria, Naga Srinivasu, Ahmed, Almusallam, Alarfaj [44]	SVM, AdaBoost, KNN, Bagging, Logistic Regression and Naive Bayes	AUC, ROC, Sensitivity, Specificity, and the F1-Score	SVM exhibited a better performance with an accuracy of 82.5%,	UCI Repository
Nayeem, Rana, and Islam [45]	KNN, Naive Bayes, and Random Forest	Accuracy, Precision, Recall, F1-score, and ROC	Random Forest has the highest classification accuracy of 95.63%	Kaggle Data set
Patil, and Annadate [46]	SVM, Random Forest, Naïve Bayes, Neural Network, and Decision Tree	Precision, Recall, F-1 score, Accuracy	Decision Tree and SVM algorithms have the best accuracy, 98.05%	Cleveland Heart Disease Data set , Hungary, Switzerland, and Long Beach V. B
Rath et al. [47]	SVM, Logistic Regression and AdaBoost	Accuracy, F1-score, and AUC	AdaBoost outperforms in both dataset	PTB-ECG, and MIT-BIH datasets
Bhatt, Patel, Ghetia, and Mazzeo [48]	XGBoost, Multilayer Perceptron, Random Forest , and Decision Tree	Precision, Recall, Accuracy, F1 score, and Area Under the ROC curve	Multilayer Perceptron hadhighest accuracy of 87.23%	Kaggle Data set
Biswas, Ali , Rahaman, Islam, Mia, Azam, Ahmed, Bui, Al-Zahrani, and Moni [49]	Logistic Regression, SVN, KNN, Rrandom Forest, Naive Bayes, and Decision Tree	Accuracy, Sensitivity, Specificity, Area Under ROC Curve (AURC), and log loss	Random Forest outperforms	UCI Repository
Hossain et al. [50]	Logistic Regression, Naïve Bayes, K-NN, SVM, Decision Tree, Random Forest, and Multilayer Perceptron	Accuracy, Precision, Recall, Specificity, F1-score, and AUC-ROC	Random Forest has the best accuracy rate of 90%.	Field Data
Nandal, Goel, and Tanwar [51]	SVM, Logistic Regression, Naïve Bayes, and XG Boost.	Accuracy, F1 score, Recall, Precision, Area Under the Curve	XG Boost offered the most accurate forecast of 0.94	UCI Repository
Rao et al. [52]	Decision Tree, Gradient boosting,	Accuracy	Logistic Regression outperform others	Cleveland Heart Disease

	Logistic Regression, SVM, and Random Forest		with accuracy of 90.16	Data set
Selvakani, Vasumathi, and Aadhiseshan [53]	Logistic Regression, SVM, and Random Forest	Accuracy, F1 score, Recall, Precision	SVM produces the greatest accuracy	Cleveland Heart Disease Data set, Hungary, Switzerland, Long Beach V, and UCI Repository and Kaggle Dataset
Srinivasan et al. [54]	Neural Network, Naïve Bayes &Radial Basis Functions	Sensitivity, Accuracy, Specifcity, Recall, Precision, and F-Score	Naïve Bayes have accuracy of 94.78%	UCI Repository
Vardhan, Kumar, Vardhini, Varalakshmi, and Kumar [55]	Decision Tree, Random Forest, Naive Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting	F1-score, Accuracy, Precision, Recall, and Confusion matrix	The extreme gradient boosting classifier has the highest accuracy of 81%	UCI Repository& Kaggle Dataset
Sreeja and Supriya [56]	Deep Convolutional Neural Network along with LIME and Grad-CAM	Accuracy, Precision, Recall, F1-score, and ROC AUC	Model Accuracy = 0.985, Precision = 0.982, Recall = 0.982 and F1-score = 0.981	MIT-BIH Arrhythmia
Abbas, Ojo, Hejaili, Sampedr, Almadho, Zaidi, and Kryvinska [57]	Random Forest, KNN, Decision Tree, Extreme Gradient Boosting, Multilayer Perceptron, and Deep Neural Network, & 1D-Convolutional Neural Network	Accuracy Precision, Recall, and F1-score along with P-value.	Multilayer Perceptron model produced the best results, with an accuracy of 95.65% and a P-value of 1.0 decimals	Pascal Challenge database

Table 1summarizes research papers focusing on cardiovascular disease diagnosis, each with a distinct study emphasis and employing techniques, and it was evident that mostly research adopted similar algorithm, performance evaluation method and most importantly the dataset as shown in Figure 5. Therefore, westrongly suggest that more progress be made by focusing on useful datasets rather than theoretical frameworks and classification. It suggests investigating improved forecasting strategies and developing novel feature selection techniques to increase comprehension and accuracy in the prediction of heart disease. Automation of ECG operations and improved comprehension of heart disease subgroups can be achieved by Machine Learning with CVD data. Standardized benchmarks are required since publically accessible datasets are scarce, making them essential for fostering research involvement and applying advanced representation learning methods to echocardiographic data.Numerous opportunities arise when machine learning models are applied to the cardiovascular domains, allowing for individualized therapy. Physicians need to be prepared for the ongoing evolution of cardiology, especially in the area of cardiac imaging.

4 Result

The study aimed to identify literature evidence related to heart disease and address biases stemming from imbalanced datasets. The information available in Table 1 provides an overview of how well different machine learning algorithms performed when predicting heart disease using diverse datasets. The most widely utilised algorithms include Neural Networks, Random Forest, Decision Trees, Support Vector Machines (SVM), Logistic Regression, and Naive Bayes as shown in Figure 3. The high-scoring models were those that reviewed studies using diverse heart disease datasets, not limited to UCI. Although the UCI heart disease dataset utilized in over 75% of systems as shown in Figure 5, similar results observed with various datasets and machine learning techniques. In several tests, Random Forest and SVM came out on top; they frequently outperformed other algorithms in terms of accuracy and overall performance as shown in Figure 4. For example, Random Forest frequently attained accuracy levels greater than 90% in a number of instances, and SVM also performed admirably, especially in research utilizing datasets such as the UCI Repository and Cleveland Heart Disease. Furthermore, additional models like as Multilayer Perceptron and Extreme Gradient Boosting (XGBoost) also demonstrated great accuracy; in certain trials, Multilayer Perceptron achieved 95.65% accuracy and XGBoost up to 97.5% accuracy.

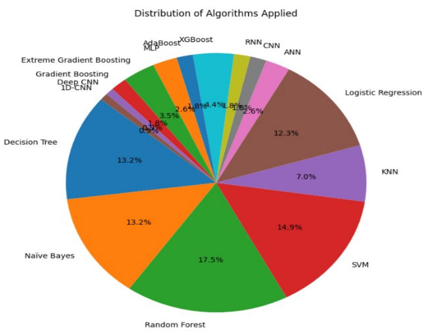

Figure 3. Distribution of Algorithms

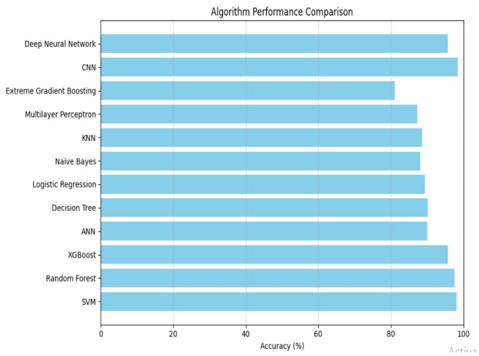

Figure 4. Performance evaluation of Algorithms

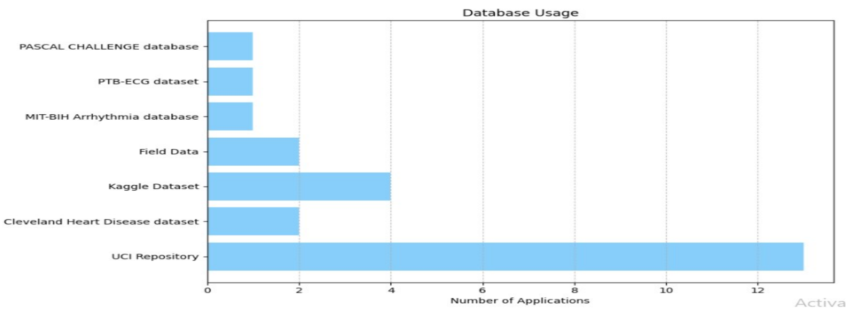

Figure 5. Database Applied on Algorithms

The analysis found that 48 studies displayed biases such as selection bias, measurement bias, and label bias, particularly in gender, race, sample size, and ECG model settings. The information in Table1 indicates a number of potential biases that can affect the studies' findings, which is given as below:

Algorithmic bias: A lot of research strongly depends on particular algorithms, such as Decision Trees, Random Forest, and SVM. Confirmation bias can result from this, when some algorithms seem to perform better just by virtue of being used more frequently or having better tuning. Researches that concentrate on a small number of algorithms may fail to consider the advantages of alternative approaches, which could distort comparisons.

Dataset bias: A large number of studies employ comparable datasets, especially the Kaggle, UCI, and Cleveland Heart Disease datasets. Repeating this process could lead to dataset bias, in which case the findings are unduly restricted to certain datasets and may not transfer well to other data sources. Different results may be obtained from studies that use less common datasets or field data, but these are not as regularly reported.

Performance Metrics Bias: The interpretation of results may be skewed by the use of different performance evaluation metrics (such as accuracy, precision, recall, and F1-score) in different research. It is possible for algorithms to be tuned for particular measures, so that the optimal method is determined by those particular metrics rather than by performance as a whole. For example, an algorithm may perform exceptionally well in accuracy but poorly in recall or precision, two other critical metrics.

Hyperparameter Tuning Bias: While some studies may require significant tuning, others may report with the default hyperparameters. This discrepancy has the potential to introduce bias because the stated performance can really be a reflection of the level of tuning rather than the algorithm's intrinsic superiority.

Model Complexity Bias: More complex models, which frequently exhibit higher accuracy, are preferred, such as XGBoost and neural networks. But sometimes the extra complexity isn't worth it, especially if simpler models function just well. Complex models can lead one to exaggerate accuracy at the cost of generalisability and interpretability.

Publication Bias: If research with improved accuracy or fresh results is more likely to be published and cited, the perception of the "best" model may be skewed. This bias may cause exaggeration of the performance of some algorithms or approaches.

According to the study, developing efficient cardiac disease detection systems requires improving the interpretability of predictions and carrying out more intense machine learning experiments using real-time patient data. The data in the study indicates that biases in the model's complexity, publishing patterns, hyperparameter modifications, datasets, algorithms, and performance indicators may have an impact on the outcomes. Considering these biases when evaluating the findings is important since they may restrict the extent and quality of the findings' implementation. The comprehensive analysis of the 48 studies, that were selected for review shows that more investigation is needed to attain interpretable and consistent performance in medical sector.

5 Conclusions

The world's largest cause of death is the cardiovascular disease, taking millions of lives annually. Early diagnosis and detection of cardiac disease can lead to a reduction in both the overall impact and death rate. Yet, there are a number of drawbacks to using conventional diagnostic techniques, such as incorrect diagnosis and treatment postponements that reduce access to quality care and raise medical expenses. A solution to these problems is provided by artificial intelligence, namely via machine learning. In particular, the precise prediction of cardiovascular disorders highlights the critical role that machine learning plays in cardiac health. The study intends to fully utilize machine learning's potential to improve diseases prediction, given the continual breakthroughs in this field and the serious public health problems associated with cardiovascular disease.'

Each study paper in the review focuses on a distinct aspect and methodology, covering a broad range of issues. Efficient machine learning categorization, data preparation, and assessing the effectiveness of congenital heart disease diagnostics are among the main topics of focus. Features, algorithms, and sample sets are highlighted by a variety of methods, including content analysis and machine learning approaches. The data analysis identifies a number of biases that could affect how well machine learning algorithms are thought to predict heart diseases. The widespread usage of particular datasets, like the UCI and Cleveland Heart Disease repositories, may skew comparisons, while some algorithms, such Random Forest and SVM, are overused. The results could also be further skewed by differences in hyperparameter settings and a propensity for sophisticated models. These elements emphasize the necessity of carefully evaluating the published results since they might have an effect on the conclusions' validity and generalizability.

6 Proposed Future Work

In the future, we plan to use machine learning algorithms on ECG data to predict cardiac disease. The process of predicting cardiac disease starts with gathering raw data from different patients, including their characteristics and whether or not they have cardiac disease. The data is pre-processed by addressing missing values by imputation or removal and standardising characteristics. To extract pertinent features from the ECG data, DWT is used in feature extraction. After that, RFE is used to identify the most crucial feature. After that, the dataset is divided into subgroups for testing and training. Using the training data, a variety of model architectures are defined and trained, such as CNN, MLP, LSTM, and ViT. Several indicators are used to assess the models on the testing set. The performance of the model will be optimised through hyperparameter adjustment. Lastly, an output showing the presence or absence of cardiac disease is produced by using the top-performing model to forecast the condition in fresh cases.

References

[1] Fact Sheet: Cardiovascular Diseases (2022). Available online: https://www.who.int/newsroom/factsheets/detail/cardiovascular diseases-(cvds) (accessed on 23 May 2022).

[2] Ayano,Y. M., Schwenker,F., Dufera, B.D. andDebelee.T.G.(2023).Interpretable Machine Learning Techniques in ECG-Based Heart Disease Classification: A Systematic Review. Diagnostics, 13: pp. 111. https://doi.org/10.3390/ diagnostics13010111

[3] Lichman, M. (2013).UCI Machine Learning Repository 2013. Available online: https://archive.ics.uci.edu/ (accessed on 11 August 2023)

[4] Faezipour, M. Saeed, A.Bulusu, S.C., Nourani, M. andMinn, L. (2010).A patient adaptive profiling scheme for ECG beat classification, IEEE Trans InfTechnol, 14(5): pp. 1153–1165.

[5] Martínez-Losas,Higueras,P.J., Gómez-Polo, J.C., Brabyn, Ferrer, J.M.F., Cañadas, V. and Villacastín, J.P. (2016). The influence of computerized interpretation of an electrocardiogram reading. Am. J. Emerg. Med., 34,: 2031–2032. [CrossRef]

[6] Ribeiro, A.H. et al. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network.Nature Communications.

[7] Zhao, J. and Li, H. (2020).ECG Signal Quality and Measurement Bias in Wearable Devices. Sensors, 20(8): 2351.

[8] Li, Q. et al.(2021). Labeling inconsistencies in ECG data and their impact on machine learning models.IEEE Transactions on Biomedical Engineering.

[9] Gluud, L.L. (2006). Bias in Clinical Intervention Research American Journal of Epidemiology.
[10] Ferrara,E. (2023). Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impact And Mitigation Strategies (Version 2). arXiv.
[11] Alizadehsani, R. Abdar, M., Roshanzamir, M.,Khosravi, A., Kebria, P., Khozeimeh, F.,Nahavandi, S., Sarrafzadegan, N. and Acharya, U. (2019).Machine learningbased coronary artery disease diagnosis: a comprehensive review.Comput.Biol. Med, 111.
[12] Moreno-Sanchez, P. A. et al. (2024). ECG-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review. Computers in Biology and Medicine,172.
[13] Aabdalla, I. D. S. and Vasumathi, D.(2024). A Comprehensive Systematic Review For Cardiovascular Disease Using Machine Learning Techniques. International Journal of Artificial Intelligence and Applications (WAIA), 15(1).
[14] Noseworthy,P.A. et al. (2020).Assessing and Mitigating Bias in Medical Artificial Intelligence: The Effects of Race and Ethnicity on a Deep Learning Model for ECG Analysis. Assessing and Mitigating Bias in Medical AI, 13:208-214.
[15] Smith,S.W. et al. (2018). A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. Journal of Electrocardiology, 52:pp. 88-95.
[16] Eltrass, A. S., Tayel, M.B. and Amma, A. I.(2022).Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures.Neural Computing and Applications, 34: 8755–8775.
[17] Ahsan,M.M.andSiddique, Z. (2022). Machine learning-based heart disease diagnosis: A Systematic literature review.Artificial Intelligence in Medicine,ScienceDirect, 128.
[18] Saraswat, D. et al. (2022). Explainable AI for Healthcare 5.0: Opportunities and Challenges. IEEE Access, 10, 84486 -84517.
[19] Schläpfer, J. and Wellens, H.(20170. Computer-interpreted electrocardiograms: Benefits and limitations.J. Amer. College Cardiol.,70(9):1183–1192.
[20] Ribeiro, A.H., HortaRibeiro, M., Paixáo, G. M. M. , Oliveira, D. M. , Gomes, P. R. , Canazart, J. A. , Ferreira, M. P. S. , Andersson, C. R., Macfarlane, P. W., Meira Jr., W., Schön, T.B. and Luiz P. Ribeiro, A.(2019). Automatic diagnosis of the 12-lead ECG using a deep neural network.arXiv:1904.01949. [Online]. Available: http://arxiv.org/abs/1904.01949
[21] Ledezma, C. A., Zhou, X., Rodríguez, B., Tan, P. J. and Díaz-Zuccarini, V.(2019).A modeling and machine learning approach to ECG feature engineering for the detection of ischemia using pseudo-ECG. PLoS ONE, 14(8), Art. no. e0220294.
[22] Xiao, R., Xu,Y., Pelter, M.M, Mortara, D.W. and Hu, X. (2018).'A deep learning approach to examine ischemic st changes in ambulatory ecg recordings. AMIA Summits Transl. Sci. Proc., 2018: 256.
[23] Elrahman,S. M. A. and Abraham, A.(2013).A review of class imbalance problem. J. Netw. Innov.Comput., 1(8):332–340.
[24] Mahajan, R.et al, (2019). Variability in ECG devices and implications for heart disease diagnosis.Journal of Electrocardiology.
[25] Fletcher, R.R., Nakeshimana, A. and Olubeko, O. (2021).Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health.Frontiers in Artificial Intelligence, 3 (116).
[26] Ahmad, M. A., Teredesai, A. and Eckert, C. (2018). Interpretable machine learning in healthcare.in Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), New York, NY, USA, 447.
[27] Obermeyer, Z., Powers, B., Vogeli, C. et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations, Science. 366:447–53.
[28] Khan Mamun, M.M.R, Elfouly,T.(2023). AI-Enabled Electrocardiogram Analysis for Disease Diagnosis.Appl. Syst. Innov., 6(95). https://doi.org/ 10.3390/asi6050095
[29] Moreno-Sanchez,P.A.et.al.(2024). ECG-based data-driven solutions for diagnosis and prognosis of cardiovascular diseases: A systematic review. Computers in Biology and Medicine, 172.
[30] Nikhar, S. and Karandikar, A.M. (2016).Prediction of Heart Disease Using Machine Learning Algorithms.International Journal of Advanced Engineering, Management and Science (IJAEMS), Infogain Publication,2(6):617- 621, ISSN : 2454-1311. (Infogainpublication.com)

[31] Nashif,S.,Raihan, M. R., Islam, M. R. and Imam, M. H. (2018).Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System.World Journal of Engineering and Technology, 6:854-873. https://doi.org/10.4236/wjet.2018.64057

[32] Bhardwaj, A.Kundra, A., Gandhi, B., Kumar, S., Rehalia, A., Gupta, M. (20219).Prediction of Heart Attack Using Machine Learning.IITM Journal of Management and IT, 10(1):20-24, 2019.

[33] Jeyaganesan, J., Sathiya, A., Keerthana, S. and Aiyer,A.(2020). Diagnosis And Prediction Of Heart Disease Using Machine Learning Techniques.Elementary Education Online, 19 (2) 1817-1827.

[34] Malavika, G., Rajathi, N., Vanitha, V. and Parameswari, P.(2020).Heart Disease Prediction Using Machine Learning Algorithms.Biosc.Biotech. Res. Comm., 13(11):24-27.

[35] Nayan,N. A.,Hamid, H. A.,Suboh, M. Z.,Jaafa,R., Abdullah, N., Yusof, N.A.M.,Hamid, M. A., Zubiri, N. F.,Arifin,A. S. K.,Daud, S. M. A.,Kamaruddin, M. A. and Jamal, A. R. A. (2020). Cardiovascular Disease Prediction from Electrocardiogram by Using Machine Learning.iJOE , 16:7. http://www.i-joe.org

[36] Ibrahim,L., Mesinovic, M., Yang, K. and Eid,M. A.(2020). Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values. IEEE Access,8:210410-210417.

[37] Jindal,H., Agrawal, S., Khera, R., Jain, R. and Nagrath, P.(2020). Heart disease prediction using machine learning algorithms. IOP Conference Series: Materials Science and Engineering, ICCRDA 2020, 1022 (2021) 012072:1-10. doi:10.1088/1757- 899X/1022/1/012072

[38] Sahoo, P. K. and Jeripothula, P. (2021).Heart Failure Prediction Using Machine Learning Techniques.

[39] Taqdees, S., Akhtar, N. and Dawood, K.(2021). Heart Disease Prediction.Conference: Heart Disease Prediction At: Rawalpindi, Pakistan.

[40] Younas.M. Z. (2021).Effective Heart Disease Prediction using Machine Learning and Data Mining Techniques.International Research Journal Of Engineering And Technology (IRJET), E-ISSN: 2395-0056 ,8(4):3539-3546. www.irjet.net

[41] Ahmed, I.(2022). A Study Of Heart Disease Diagnosis Using Machine Learning And Data Mining. Master of Science.California State University, San Bernardino.

[42] Bora, N., Gutta, S. and Hadaegh, A. (2022).Using machine learning to Predict Heart Disease.Wseas Transactions On Biology And Biomedicine, E-ISSN: 2224-2902:1-9. DOI: 10.37394/23208.2022.19.1

[43] Dixit,S. Mohan,P.V. and Terni, S. R.(2022).Prediction of Heart Disease Using Machine Learning Algorithms.International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN: 2321-9653, 10(3):1275-1282. Available at www.ijraset.com

[44] Guleria, P., Naga Srinivasu,P., Ahmed, S., Almusallam, N. and Alarfaj,F.K. (2022). XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques.Electronics 2022, 11(4086): 1-30.

[45] Nayeem, M. J., Rana, S. andIslam, M. R.(2022). Prediction of Heart Disease Using Machine Learning Algorithms.European Journal of Artificial Intelligence and Machine Learning, ISSN: 2796-0072, 1(3):22-26. DOI: 10.24018/ejai.2022.1.3.1

[46] Patil, S. and Annadate, M.(2022).Implementation of Machine Learning Model to Predict Heart Problem.International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878 (Online), 10(5): 117-122.

[47] Rath, A., Mishra, D. and Panda,G.(2022).Imbalanced ECG signal-based heart disease classification using ensemble machine learning technique. Front. Big Data, 5(1021518).

[48] doi: 10.3389/fdata.2022.1021518

[49] Bhatt,M., Patel,P.,Ghetia, T. and Mazzeo,P. L. (2023).Effective Heart Disease Prediction Using Machine Learning Techniques.Algorithms, MDPI, 16(88): .1-14.

[50] https:// doi.org/10.3390/a16020088

[51] Biswas, N., Ali ,M. M., Rahaman,M. A.,Islam,M., Mia, M.R.,S. Azam, Ahmed, K., Bui,F. M. Al-Zahrani,F. A.andMoni, F.A.(2023). Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques. BioMed Research International, Hindawi , Article ID 6864343, 1-15.

[52] Hossain, M. I.,Maruf, M. H.,Khan, M. A. R.,Prity, F. S., Fatema,S. ,Ejaz,M. S. and Khan,M. A. S.(2023). Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison.Iran Journal of Computer Science ,Spinger, 6:397–417.

[53] Nandal,N., Goel, L. and Tanwar, R.(2023). Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. F1000Research 2022, 11 (1126):1-19.

[54] Rao,CH. D.,Priya, B. A. R. P. Sowmya, B. V. S., Chandrasekhar, T., Geethanjali,V. and Kumar, G.S.(2023). Machine Learning Algorithms For Prediction Of Heart Disease. Machine Learning Algorithms For Prediction Of Heart Disease, e-ISSN: 2582-5208, 5(4): 984-990.www.irjmets.com

[55] Selvakani, S., Vasumathi, K. and Aadhiseshan, V.(2023).Application of Machine Learning in Predicting Heart Disease. Asian Basic and Applied Research Journal, Article no.ABAARJ.1255, 5(1): 61-68.

[56] Srinivasan, S., Gunasekaran, S., Mathivanan, S.K., Malar,M. B. B. A., Jayagopal, P. and Dalu,G. T. (2023). An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database. Scientifc Reports , 13(13588):1-19.

[57] Vardhan, V. H.,Kumar, U. R.,Vardhini,V.,Varalakshmi, S. L. and Kumar,A.S.(2023). Heart Disease Prediction Using Machine Learning.Journal of Engineering Sciences, ISSN:0377-9254, 14 (4): 440-450

[58] Sreeja M. U., and Supriya M. H. (2023). A Deep Convolutional Model for Heart Disease Prediction based on ECG Data with Explainable AI. Wseas Transactions On Information Science And Applications, E-ISSN: 2224-3402, 20:254-264. Doi: 10.37394/23209.2023.20.29

[59] Abbas, S.,Ojo, S.Hejaili,A.A.,Sampedro,G.A., Almadhor, A., Zaidi, M.M. and Kryvinska, N. (2024). Artifcial intelligence framework for heart disease classifcation from audio signals.Scientifc Reports, 14 (3123).