algorithms

MDPI

Review

A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges

Marwah Abdulrazzaq Naser [1,*], Aso Ahmed Majeed [2], Muntadher Alsabah [3], Taha Raad Al-Shaikhli [4] and Kawa M. Kaky [4]

[1] Department of Computer Engineering, University of Baghdad, Al-Jadriya, Baghdad 10071, Iraq
[2] Department of Basic Science, University of Kirkuk, Kirkuk 36001, Iraq; asoalsalihi@uokirkuk.edu.iq
[3] Medical Technical College, Al-Farahidi University, Baghdad 10071, Iraq; muntadher.m@uoalfarahidi.edu.iq
[4] Department of Computer Engineering Techniques, Al-Nisour University College, Baghdad 10071, Iraq; taha.r.eng@nuc.edu.iq (T.R.A.-S.); kawa.mudher@gmail.com (K.M.K.)
* Correspondence: marwah@dcec.uobaghdad.edu.iq

Abstract: Cardiovascular disease is the leading cause of global mortality and responsible for millions of deaths annually. The mortality rate and overall consequences of cardiac disease can be reduced with early disease detection. However, conventional diagnostic methods encounter various challenges, including delayed treatment and misdiagnoses, which can impede the course of treatment and raise healthcare costs. The application of artificial intelligence (AI) techniques, especially machine learning (ML) algorithms, offers a promising pathway to address these challenges. This paper emphasizes the central role of machine learning in cardiac health and focuses on precise cardiovascular disease prediction. In particular, this paper is driven by the urgent need to fully utilize the potential of machine learning to enhance cardiovascular disease prediction. In light of the continued progress in machine learning and the growing public health implications of cardiovascular disease, this paper aims to offer a comprehensive analysis of the topic. This review paper encompasses a wide range of topics, including the types of cardiovascular disease, the significance of machine learning, feature selection, the evaluation of machine learning models, data collection & preprocessing, evaluation metrics for cardiovascular disease prediction, and the recent trends & suggestion for future works. In addition, this paper offers a holistic view of machine learning's role in cardiovascular disease prediction and public health. We believe that our comprehensive review will contribute significantly to the existing body of knowledge in this essential area.

Keywords: machine Learning; cardiovascular disease; cardiovascular disease types; classification; prediction; cardiac care; feature selection; healthcare; explainable AI (XAI); disease diagnosis; intelligent system; classification; artificial general intelligence (AGI)

check for updates

Citation: Naser, M.A.; Majeed, A.A.; Alsabah, M.; Al-Shaikhli, T.R.; Kaky, K.M. A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. *Algorithms* **2024**, *17*, 78. https://doi.org/10.3390/a17020078

Academic Editor: Jesper Jansson

Received: 26 December 2023
Revised: 6 February 2024
Accepted: 9 February 2024
Published: 13 February 2024

1. Introduction

To date, healthcare systems face significant challenges, including the increasing prevalence of diseases, the simultaneous presence of multiple health conditions, a growing need for healthcare services, disability due to aging, and rising healthcare expenditures [1,2]. However, among other diseases, cardiovascular disease, is, in particular, considered a major public health problem, affecting millions of people across the globe [3–5]. Specifically, cardiovascular disease poses not only a medical challenge on healthcare systems but also an economic and societal one [6,7]. Table 1 summarizes the major cardiovascular disease types and the deception of each of the types. Therefore, with the right treatment and early detection of cardiovascular disease, the symptoms of the disease can be reduced and the function of the heart can be significantly improved [8,9]. It would also help in allowing early intervention, and personalized treatment plans, hence, enhancing healthcare systems [10]. The predicted results of cardiovascular disease can be used to prevent, and thus, reduce the cost of surgical treatment [11,12]. However, conventional methods for cardiovascular disease prediction are either costly or lack efficiency in human cardiovascular disease prediction. Hence, the indispensability of smart and advanced healthcare systems has become apparent, emphasizing the urgent need for their development [13]. Smart healthcare systems enable physicians to conduct remote patient monitoring, facilitating the continuous tracking of disease progression [14,15]. Additionally, these intelligent systems play a crucial role in disease identification, diagnosis, categorization, forecasting, prevention, and treatment [16,17]. To this end, various artificial intelligence (AI) methods, particularly machine learning algorithms, can be applied to healthcare systems [18], and hence, the mortality rate associated with cardiovascular disease can be reduced [19].

Table 1. List of major cardiovascular disease types and a description of each type.

Cardiovascular Disease Type	Description
CAD	CAD occurs when the blood vessels (coronary arteries) that supply the heart with oxygen and nutrients become narrowed or blocked due to the buildup of cholesterol and fatty deposits (atherosclerosis).
Heart failure	Occurs when the heart cannot pump blood effectively, leading to a reduced supply of oxygen and nutrients to the body's tissues and organs. It can result from CAD, hypertension, and heart valve diseases.
Arrhythmias	This condition involves problems with one or more of the heart's valves. It can lead to valve stenosis (narrowing) or regurgitation (leakage). Common valve disorders include aortic stenosis and mitral regurgitation.
Valvular	This condition involves problems with one or more of the heart's valves. It can lead to valve stenosis (narrowing) or regurgitation (leakage). Common valve disorders include aortic stenosis and mitral regurgitation.
Cardiomyopathy	This is related to the muscle and can weaken the heart's ability to pump blood effectively. It can be inherited or acquired, and there are different types, such as dilated cardiomyopathy and hypertrophic cardiomyopathy.
Congenital	This is present at birth and involves structural abnormalities in the heart. It can affect the heart's walls, valves, or blood vessels.
Infective Endocarditis	This is an infection of the inner lining of the heart (endocardium) and the heart valves. It is typically caused by bacteria or other microorganisms that enter the bloodstream and settle in the heart.
Vascular	This refers to conditions or disorders that affect the blood vessels, which include arteries and veins throughout the body, which can disrupt the normal flow of blood.

1.1. Motivation and Paper Contributions

This review paper is motivated by the urgent need to assemble the role of machine learning methods in improving cardiovascular disease prediction, which is to date considered a critical area of public health concern. With the continuous evolution of machine learning techniques and the growing public health impact of cardiovascular disease, this paper seeks to provide an up-to-date and comprehensive evaluation of machine learning techniques for cardiovascular disease prediction. This paper bridges the gap between machine learning and cardiology, emphasizing the importance of interdisciplinary collaboration and domain knowledge. The scope of this paper encompasses a thorough analysis of various machine learning models, feature selection and engineering, evaluation metrics, recent advances, and their public health impact. This paper carries out a discussion encompassing various aspects of machine learning models, including their underlying mechanisms, applications, strengths, and limitations. Moreover, this paper offers an insightful review, emphasizing the innovative applications of machine learning in cardiovascular disease diagnosis. This paper provides an overview of the recent advancements in the application of machine learning for cardiovascular disease, aiming to elucidate the current trends, approaches, and challenges associated with machine learning techniques in cardiovascular disease diagnosis. What sets this review apart is its unique focus on the incorporation of domain knowledge, up-to-date coverage of trends, and the holistic view of the public health impact of machine learning in cardiovascular disease prediction. This paper aims to be an effective resource for researchers and specialists facilitating informed decision-making and fostering advancements in the field. Furthermore, the contents of this review paper are structured to provide a clear and organized presentation of the role of machine learning in cardiovascular disease prediction. Noting that there are various health conditions such as Cardiovascular diseases, Chronic Obstructive Pulmonary Disease (COPD), Influenza (Flu), Tuberculosis (TB), Human Immunodeficiency Virus (HIV), Neurological Diseases, Cancer, Diabetes Mellitus, Osteoarthritis, Gastroesophageal Reflux Disease (GERD), Endocrine and Metabolic Diseases, Depression Inflammatory Diseases, Mental Health Disorders, Gastrointestinal Diseases and Musculoskeletal Disorders. In this paper, we focus on cardiovascular disease due to its widespread prevalence, significant impact on public health, and the imperative need for proactive measures in understanding, preventing, and managing these conditions.

1.2. Papers Selection Strategy

The main objective of this research is to instigate the role of machine learning for cardiovascular disease prediction, and hence identify research papers that align with this specific scope and criteria of the investigation. To this end, we seek to provide a comprehensive overview of the papers that deal with this research topic. To enhance the probability of retrieving high-quality search results, well-known digital databases were selected and queried. These databases include Science Direct, providing access to a broad spectrum of scientific journals in medicine, science, and technology; IEEE Xplore digital library, featuring publications related to engineering and technology; MDPI, PubMed, and Google Scholar, offering diverse articles across various domains. The documents are selected with the English language, which is either a journal or conference format. It primarily focuses on the development of techniques related to machine learning that are used for cardiovascular disease prediction. Documents were categorized as irrelevant if they either did not meet the search criteria or included the specified search terms but did not address cardiovascular disease. The choice of these databases was influenced by their established academic credibility and their representation of various academic disciplines. The study's search terms were (Machine learning OR machine learning model) AND (heart disease OR cardiovascular disease) AND (disease types).

1.3. Organization of the Paper

The structure of this review paper is as follows. Section 2, provides a review of the fundamentals of machine learning and its significance in cardiovascular disease prediction. Section 3 carries out a discussion about the machine learning models, which can be used for the prediction of cardiovascular disease, accompanied by an in-depth review of the state-of-the-art research. The data collection and preprocessing techniques in healthcare systems are discussed in Section 4. Section 5 describes the commonly used evaluation metrics for predicting cardiovascular disease. In addition, it presents a discussion about model validation and cross-validation and sheds light on the significance of interpretability and explainability for machine learning in the healthcare system. Section 6, provides a list of some open problems and suggestions related to the application of machine learning in cardiovascular disease prediction and highlights some further suggestions that can be considered for future works. Finally, this paper is concluded in Section 7. The following table of contents outlines various sections and subsections, facilitating easy navigation for the readers. This structure ensures that the content flows logically and comprehensively, aiding in a better understanding of the contributions made in this research. Figure 1 demonstrates the contents of this paper.

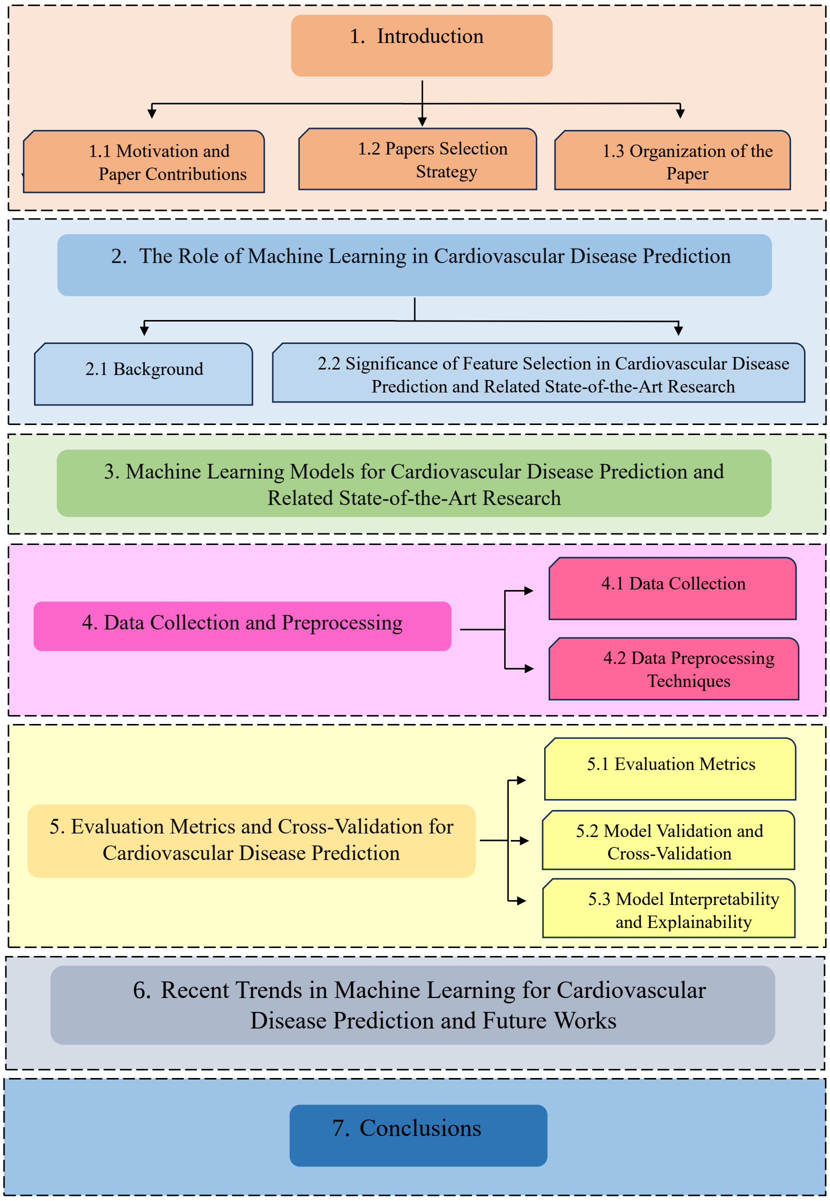

Figure 1. Structure of the review paper demonstrating the contents of each section and subsection.

2. The Role of Machine Learning in Cardiovascular Disease Prediction

This section offers a foundational overview of machine learning, followed by a detailed discussion of its significant role in cardiovascular disease prediction. In particular, this section will discuss the applications and the tools of machine learning that can be used for cardiovascular disease predictions. Furthermore, the discussion is extended to emphasize the importance of feature selection for cardiovascular disease prediction, highlighting its significant role in optimizing model performance.

2.1. Background

Healthcare professionals relied on several conventional methods for cardiovascular disease prediction. An example of conventional methods of cardiovascular disease prediction includes clinical risk factors that are related to age, gender, family history, and personal medical history. In addition, echocardiography can be used for visualization of the heart function [20] where electrocardiogram (ECG) can be used to detect signs of congestive heart failure [21]. In particular, ECG can help in the prognosis and treatment management of patients diagnosed with congestive heart failure [22]. Cardiac catheterization can also be used to diagnose and evaluate coronary artery disease (CAD) and typical issues with the heart and blood vessels [23,24].

However, recently, there has been a growing need for more advanced predictive models, such as those powered by machine learning, to improve the accuracy and efficiency of cardiovascular disease prediction [25]. Machine learning is a subset of artificial intelligence (AI) that uses algorithms to allow computer agents to perceive, acquire knowledge, identify patterns, and make intelligent decisions by analyzing collected data [26–28]. With its ability to evaluate enormous amounts of patient data, machine learning has emerged as a key player for achieving accurate and trustworthy cardiovascular disease prediction [29]. The predictive power of machine learning techniques has emerged as a promising path for revolutionizing the management of cardiovascular disease [17,30,31].

Machine learning can enhance early disease detection, accelerate the development of drugs, provide data-driven insights, enable remote monitoring, acquire crucial information from patient's datasets, allow data-driven decision-making, improve image and speech recognition, and simplify administrative procedures [32–34]. This enables early detection, often before symptoms become severe, allowing for timely intervention and treatment, hence potentially lowering healthcare costs. Analyzing such vast amounts of data in the healthcare field is challenging for humans, if not nearly impossible [35]. Hence, the prevalent use of machine learning proves invaluable in extracting meaningful insights from such extensive datasets. Machine learning algorithms are very beneficial for remote healthcare monitoring [36,37]. Specifically, patients can receive remote monitoring and consultations, reducing the need for frequent hospital visits and improving access to care, particularly in remote areas [38]. Machine learning algorithms can integrate and analyze a vast amount of patient data from various sources, including medical records and notes, genetic information, and diagnostic tests, to identify subtle patterns, to detect early warning signs and risk factors associated with cardiovascular disease [32,39–41]. This would help in providing a comprehensive view of a patient's health and enabling a better understanding of cardiovascular disease risk factors and finding subtle patterns. Examining diverse patient data can flag individuals at risk before symptoms appear, enabling timely intervention and reducing healthcare resource burdens. Machine learning algorithms can analyze medical images and extract the relevant features from medical images, such as X-ray, angiograms and magnetic resonance imaging (MRI), echocardiograms, computed tomography (CT) scans, and clinical records, to identify subtle signs of cardiovascular disease [42–47].

In the literature, there is a predominant focus on supervised machine learning for cardiovascular disease prediction, due to the availability of labeled datasets, leading us to discuss the supervised approach exclusively. Supervised learning involves an algorithm learning from labeled data, allowing it to predict outcomes for new, unlabeled cases, by generalizing knowledge from the provided available data [48]. In particular, the supervised learning algorithm learns to map input data to a specific output, or target variable, based on a labeled dataset. In supervised learning, the algorithm aims to generalize patterns from the training data in order to perform predictions or classifications of new data. As described in [49], supervised machine learning entails having a predefined output attribute and utilizing input attributes. Supervised algorithms initially perform analytical tasks using training data and then build functions to map new instances of the attribute [50].

Classification and regression algorithms are two categories of supervised machine learning algorithms according to [51,52]. Labeled data, or what is known as training set [53], is crucial in supervised learning because it provides the algorithm with the ground truth information that is needed to learn and make predictions. Labeled data consists of input features (independent variables) [54,55] and their corresponding correct target values (dependent variable) or labels [56]. The algorithm uses these labeled examples to identify patterns, associations, and relationships within the data, allowing it to learn how to make predictions on new, unlabeled data. The supervised learning algorithm uses the features to make predictions or classifications about the target variable based on patterns learned from the labeled examples in the dataset [57]. This allows for the creation of predictive models that can help in identifying individuals at risk of cardiovascular disease based on their characteristics. Appendix A provides the attributes of three common datasets for cardiovascular disease prediction, which are the Cardiovascular, Cleveland, and Framingham datasets.

Machine learning classification algorithms have the potential to identify patients at risk of cardiovascular disease based on their medical data records, allowing for early medication and treatment. Classification algorithms can categorize patients into risk groups, enabling healthcare providers to prioritize patients at higher risk. Classification algorithms can assist in triaging patients, ensuring that those with the most urgent cardiac issues receive immediate treatments. Therefore, machine learning classification algorithms are essential for cardiovascular disease management, helping efficiently in early disease detection, risk assessment, personalized treatment, and more accurate diagnostics.

In summary, machine learning is invaluable in healthcare, especially in the prediction and management of cardiovascular disease. It empowers healthcare providers with the tools to make more informed decisions, enhances patient outcomes, and advances the efficiency of the healthcare system.

2.2. Significance of Feature Selection in Cardiovascular Disease Prediction and Related State-of-the-Art Research

In healthcare systems, features are considered as input variables that describe the characteristics of patients [58]. Each individual (patient) in the dataset is represented by a set of feature values [29,59]. In particular, healthcare datasets may contain irrelevant features that may introduce noise into the model, hence leading to decreased prediction accuracy [60]. Hence, feature selection approaches aim to reduce the input variables by removing redundant or irrelevant features and selecting the most informative and relevant features [61]. To this end, feature selection can be used to improve prediction accuracy and efficiency in healthcare systems [57]. It is crucial in the development of accurate and interpretable predictive models for cardiovascular disease [17]. Additionally, feature selection enhances the classification accuracy and minimizes the model execution time [62]. However, feature selection requires a precise selection of relevant variables from a large set of possible features [63,64].

Furthermore, feature selection techniques can reduce the dimensionality of the datasets which can be achieved by ignoring the noisy features, and hence, the predictive models can be more accurate. Principal component analysis (PCA) is one of the dimensionality reduction methods that can be used to minimize the number of features while retaining most of the variance [65,66]. For example, dimensionality reduction using the PCA technique has been applied in [67–69] for cardiovascular disease detection. Other techniques such as firefly algorithm [70] and minimum redundancy maximum relevance [71] have also been applied for dimensionality reduction in cardiovascular disease prediction. Applying such efficient dimensionality reduction techniques can improve model efficiency and interpretability. Therefore feature engineering with dimensionality reduction has the ability to improve the data selection, hence improving prediction accuracy [72]. Furthermore, a model with too many features relative to the number of instances in the dataset is at risk of overfitting, where it fits the noise rather than the underlying patterns [73]. Feature selection can mitigate this overfitting risk by simplifying the model and reducing its complexity. Feature selection methods are categorized into filter and wrapper methods [74]. The wrapper methods choose an optimal subset of features by incorporating the classifier, whereas the filter methods select features independently of the classifier.

The most common feature selection methods are univariate, weighted least square, rough sets, fast correlation-based filter (FCBF), and sequential forward selection [5,75–79]. The work in [62] has proposed three feature selection algorithms, which are Relief, minimal-redundancy-maximal-relevance (mRMR), and least absolute shrinkage and selection operator (LASSO) to identify the most crucial and strongly correlated features that significantly impact the prediction of cardiovascular disease. Univariate and Relief feature selection methods were proposed in [5], where the univariate method utilizes a statistical approach to select a subset of features that has the strongest relationship with a class label. In contrast, the Relief technique gives each feature in the dataset a weight, and these weights are modified over time. In [80], a new approach known as a hybrid random forest with a linear model (HRFLM) is introduced. This method focuses on identifying important features using machine learning techniques, hence, leading to improved accuracy in cardiovascular disease prediction. The prediction model incorporates various feature combinations and employs several established classification methods. As a result, the proposed approach enhances the performance of cardiovascular disease prediction.

Recently, evolutionary methods have emerged as a significant class of techniques that can be utilized efficiently for feature selection and prediction of cardiovascular diseases. For example, the work in [81] focused on involving the identification and selection of crucial features, along with the exploration of machine learning techniques, to augment the predictive capacity of classification models for accurately predicting cardiovascular disease. To this end, a hybrid ensemble model using genetic algorithm (GA) and linear discriminant analysis (LDA) was proposed to improve the prediction accuracy. The work in [82] proposed a combination of convolutional neural network (CNN) and jellyfish search optimizer (jSO) approach for the prediction of cardiovascular diseases. In particular, the jSO optimization algorithm is exploited to tune the CNN hyperparameters and improve the accuracy. The work in [83] introduced a new model named hyOPTXg, which was designed for predicting cardiovascular disease through an optimized XGBoost classifier. Consequently, fine-tuned hyperparameters of XGBoost and conducted model training using the optimized parameters were proposed to achieve a superior performance enhancement in cardiovascular disease prediction. The work in [84] provided a comparative investigation that integrates machine learning algorithms with meta-heuristic algorithms for feature selection, aiming to enhance the classification capabilities of machine learning algorithms by identifying features that significantly influence accuracy. The findings affirm that the amalgamation of machine learning and meta-heuristic algorithms leads to superior classification accuracy with a reduced number of features. Hybrid methodologies that integrate hyper-parameter optimization algorithms with two highly effective classification techniques namely: Support Vector Machines (SVMs) and Long Short-Term Memory (LSTM) neural networks have been proposed in [85] to further improve the accuracy of cardiovascular disease diagnosing. The results were achieved based on the Cleveland dataset and its extension Statlog. The work in [86] proposed a multiobjective approach with a fuzzy system for the classification of cardiovascular risk. The proposed approach involved addressing computational elements such as configuring the fuzzy system, optimizing the process, selecting an appropriate solution from the optimal Pareto front, and ensuring the interpretability of the fuzzy logic system post-optimization.

Leveraging multiobjective optimization and Pareto dominance allows the acquisition of a set of optimal solutions that embody the most effective equilibrium between two optimization objectives. The work in [87] focused on creating and automating a disease prediction model to facilitate early detection of cardiovascular disease and its associated risk factors. To this end, feature selection was executed using non-linear Particle Swarm Optimization (NL-PSO). Subsequently, classification was carried out using the Improved Deep Evolutionary model with Feed Forward Neural Networks (IDEBDFN). The algorithm's learning nature was leveraged to assess the characteristics of the hidden layers, ensuring optimal results. The findings illustrate that the proposed model exhibits superior predictive accuracy. The work in [88] introduced an alternative training technique for a multilayer perceptron (MLP) that incorporates a particle swarm optimization (PSO) algorithm for cardiovascular disease detection. The results demonstrated that the proposed hybrid MLP-PSO classifier empowers practitioners to diagnose cardiovascular disease earlier, with enhanced accuracy and efficacy. An approach involving a radial basis function neural network (RBFNN) was proposed in [89], which was coupled with a robust hybrid particle swarm optimization (HPSO). The HPSO incorporated a spiral-shaped mechanism (HPSO-SSM) to enhance the PSO algorithm performance by addressing constraints such as slow convergence and the local minimum challenge. The work in [90] proposed evolutionary algorithms based on Genetic Algorithm (GA) and PSO for the feature selection to further improve the accuracy of machine learning algorithms. The results demonstrated that the feature selection based on GA achieved the best prediction accuracy. Several research papers have also found that genetic algorithm (GA) is a highly effective method for feature selection, see e.g., [91–95].

Overall, as the development of machine learning continues to shape the future of healthcare, the feature selection approach remains an essential component to improve cardiovascular disease prediction.

3. Machine Learning Models for Cardiovascular Disease Prediction and Related State-of-the-Art Research

This section provides an exploration of machine learning models designed for the prediction of cardiovascular diseases, accompanied by an in-depth discussion of the state-of-the-art research. In particular, this section comprehensively addresses the advancements, methodologies, and key findings related to cardiovascular disease prediction using machine learning models.

Advanced machine learning models are used to analyze heterogeneous data that come from various sources [96]. The choice of machine learning model can significantly impact the quality and trustworthiness of cardiovascular disease prediction [31]. With the abundance of medical data and the ever-increasing computational capabilities, machine learning models have become indispensable [97]. Predictive analytics models are used to stratify patients into different risk categories, enabling targeted interventions. These models can help in identifying individuals at high risk of cardiovascular disease and enable personalized treatment plans. Machine learning models are being integrated into clinical workflows to provide real-time decision support to healthcare professionals [31]. In this subsection, we will explore a range of machine-learning models that are commonly employed for cardiovascular disease prediction.

The common machine learning models are Logistic Regression [98], Decision Trees [99–101], Random Forests [102], support vector machines (SVM) [103,104], K-Nearest Neighbors (KNN), Adaptive Boosting, commonly referred to as AdaBoost [105,106], Naïve Bayes [107,108], and Conventional Neural Network (CNN) or what know as deep learning [13]. Table 2 provides a summary of the machine learning models in the prediction of cardiovascular diseases along with the strengths and weaknesses points.

To this end, several research works have been done so far to investigate the application of machine learning models in cardiovascular disease prediction. For example, the work in [109] constructed a cardiovascular disease classification system using a Logistic Regression classifier machine learning techniques, which achieves an accuracy of 77%. Cleveland dataset is employed, along with global evolutionary and feature selection methods. The work in [110] designed a diagnostic system for cardiovascular disease classification, utilizing multi-layer perception and SVM algorithms, achieving an accuracy rate of 80.41%.

The work [111] introduced a model to determine the most effective machine learning algorithm for early-stage prediction of cardiovascular disease, ensuring high accuracy. The results showed that the best accuracy for cardiovascular disease classification has been achieved using a random forest algorithm with a rate of 95.4%. The work in [112] developed a cardiovascular disease classification system by integrating a neural network with Fuzzy logic, resulting in an accuracy of 87.4%. The work in [113] introduced an ANN ensemble-based diagnostic system for cardiovascular disease, coupled with the statistical measurement system Enterprise Miner. Their system yielded an accuracy of 89.01%, a sensitivity of 80.09%, and a specificity of 95.91%. The work in [114] developed a machine learning-based cardiovascular disease diagnosis system, incorporating the ANN algorithm and a feature selection algorithm, achieving commendable performance. The work in [115] proposed an expert medical diagnosis system for cardiovascular disease identification, using predictive machine learning models like Naïve Bayes, Decision Tree, and ANN. These models attained an accuracy of 86.12%, 88.12%, and 80.4%, respectively.

Table 2. A summary of machine learning models in prediction of cardiovascular disease.

Machine Learning Model	Strengths	Weaknesses
Logistic Regression (LR): A statistical model that assesses the probability of an individual having a cardiovascular disease based on various input features, such as age, cholesterol levels, and blood pressure.	1. Simplicity and interpretability in disease diagnostics. 2. Computationally efficient for large datasets. 3. Less prone to overfitting, making it robust when working with moderate-sized datasets.	1. Assumes linear input–outcome relationship. 2. Less flexible in complex medical scenarios. 3. Potential lower predictive accuracy with intricate data patterns. 4. Struggles with non-linear relationships without feature engineering or transformation.
Decision Trees (DT): Applicable for both classification and regression problems. It generates a tree-like structure of decision rules by recursively splitting the data according to input attributes. Helps to determine the most important elements or symptoms of cardiovascular disease.	1. Easy to interpret and hence can explain the decision-making process easily. 2. Handles non-linear relationships in medical data with both categorical and numerical features and identifies key factors for cardiovascular disease. 3. Low computational costs during prediction.	1. Prone to overfitting, impacting generalization. 2. May not capture complex relationships as well as other models. 3. Less stable, with small data changes leading to different tree structures.
Random Forest (RF): An ensemble learning method that combines multiple decision trees to make predictions. It is less prone to overfitting compared to individual decision trees, resulting in the improved generalization of new data. It provides higher predictive accuracy due to the combination of multiple trees and the reduction of bias and variance.	1. Mitigates decision tree overfitting by averaging predictions. 2. It offers robustness, high accuracy, and handling of non-linearity. 3. Provides accurate and robust results, even in the presence of noisy or complex data. 4. Robust, with effective handling of high-dimensional data. Less sensitive to noisy data.	1. Computationally intensive, may need more resources. 2. Larger model sizes can be limited in resource-constrained environments. 3. Requires proper hyperparameter tuning for optimization.
Support Vector Machines (SVMs): A powerful algorithm that can be used for both classification and regression tasks. It obtains the optimal hyperplane, which seeks the best separation of the data points into different classes. Proficiency in managing high-dimensional data; hence, it is suitable for datasets with many features.	1. Handles non-linear relationships with kernels. 2. Versatile and effective in complex pattern capture. 3. Find complex decision boundaries in high dimensions. 4. Works well with clear class separation margins.	1. Computationally expensive, complex in high dimensions. 2. Sensitive to kernel choice. 3. Time-consuming training on large datasets. 4. Essential hyperparameter tuning for optimal performance. 5. May lack interpretability compared to logistic regression or decision trees.
K-Nearest Neighbors (KNN): A simple algorithm that finds the *k*-nearest data points in the training dataset, such as the Euclidean distance. It predicts the class of the new data point by taking a majority vote from the KNN. Effective when similar patients with similar feature profiles are likely to have similar cardiovascular disease outcomes.	1. Non-parametric algorithm, adapts well to diverse and non-linear patterns of data distributions. 2. Simple to understand and implement. 3. Works well with small datasets, making it applicable in different scenarios. 4. Less sensitive to outliers and noisy data points.	1. Computationally expensive for large datasets. 2. Sensitive to the choice of *k*. 3. May struggle with imbalanced datasets. 4. Critical to properly select *k* for best predictive performance.

Table 2. *Cont.*

Machine Learning Models	Strengths	Weaknesses
AdaBoost: Combines several weak classifiers into a unified and robust classifier. Its mechanism involves assigning higher weights to samples that pose greater classification challenges while assigning lower weights to well-categorized samples. It finds application in both categorization and regression analysis.	1. High accuracy and performance in classification. 2. Versatile with different data types and base classifiers, making it versatile in different machine learning scenarios. 3. Effectively handles noisy data and outliers. 4. Assign higher weights to misclassified instances, allowing it to focus on correcting mistakes.	1. Handles noisy data but sensitive to outliers or mislabeled instances. 2. Computationally intensive with a large number of weak learners, which may affect training time and resource requirements. 3. Struggles with complex relationships or dependencies in datasets, as it relies on relatively simple weak learners.
Naïve Bayes: A probabilistic model that works based on the Bayes theorem. Given the class title, it assumes that features are conditionally independent. It calculates the probability that a patient has cardiovascular disease given their feature values, such as age, cholesterol, and blood pressure.	1. Computationally efficient with high-dimensional data. 2. Performs well with small to moderately-sized datasets. 3. Its probabilistic nature allows for the simple interpretation of results. 4. Handles a large number of features efficiently.	1. Assumes feature independence, limiting accuracy in capturing complex dependencies among features. 2. May not capture intricate relationships between features, which can affect predictive accuracy.
Deep Learning: It employs various evaluation criteria, including accuracy and specificity, to guide feature extraction inversely. Involves the use of artificial neural networks with multiple hidden layers to learn complex representations from data. Can identify intricate patterns that may be challenging for human interpretation, leading to more accurate diagnosis of diseases.	1. Automatically discovers relevant features for data, reducing the requirement for manual feature engineering. 2. Effectively handles missing data, providing accurate predictions even when some data are unavailable. 3. Achieves high predictive performance when trained on large and diverse datasets. 4. Handles diverse data types including images, text, and numerical data.	1. Complex and computationally intensive. 2. Requires substantial computational resources for training and inference. 3. Relies on large amounts of labeled data for effective training. 4. Challenging to interpret, limiting utility in medical diagnostics, where interpretability is essential.

The work in [116] developed a three-phase technique based on ANN for cardiovascular disease prediction in angina, achieving an accuracy of 88.89%. The work in [117] designed an integrated medical decision support system for cardiovascular disease diagnosis, incorporating ANN and Fuzzy AHP. The proposed method achieved an accuracy of 91.10%. The work in [118] presented a cardiovascular disease classification system employing relief and rough set techniques, which achieved a classification accuracy of 92.32%. In another study [103], a cardiovascular disease identification method using feature selection and classification algorithms was proposed. To this end, the Sequential Backward Selection Algorithm (SBS FS) was utilized for feature selection and tested the KNN classifier on both full and selected feature sets, obtaining high accuracy. The work in [80] designed a cardiovascular disease prediction method using hybrid machine learning techniques and introduced a novel feature selection method for effective machine learning classifier training and testing, achieving a classification accuracy of 88.07%. The work in [119] developed cardiovascular disease identification techniques employing an improved SVM-based duality optimization technique. While the aforementioned techniques have employed various methods to detect cardiovascular disease at early stages, they exhibit limitations in terms of prediction accuracy and computational time.

In [120], the researchers developed a classifier utilizing a blend of diverse support vector machines (SVMs) to classify ECG signals, focusing on the extraction of features from intervals between consecutive beats. Additionally, they tackled the challenge of highly imbalanced data by employing both over and under-sampling techniques on the Arrhythmia dataset. The work in [121] proposed a classifier that is capable of identifying 17 distinct types of Arrhythmia, which employs a CNN model specifically designed for long-duration ECG signals in the Arrhythmia dataset. The work in [122] utilized an adaptive implementation of CNNs to classify the Arrhythmia dataset, resulting in a rapid and precise patient-specific ECG classification and monitoring system. The work in [123] proposed a random forest and CNN algorithms for cardiovascular disease prediction. To this end, several imbalance techniques have been discussed. In [124], prediction models for coronary cardiovascular disease (CHD), which is known as cardiovascular disease have been proposed. To this end, various supervised machine learning algorithms, including Gaussian Naïve Bayes, Bernoulli Naïve Bayes, and Random Forest, are employed in cardiovascular disease prediction. The results demonstrated that the Gaussian Naïve Bayes, Bernoulli Naïve Bayes, and Random Forest algorithms achieved accuracy rates of 85%, 85%, and 75%, respectively.

The work in [125] proposed an ensemble model adopted to enhance predictive accuracy by combining the strengths of multiple classifiers. To this end, ensemble learning is employed, amalgamating five classifier models—SVM, ANN, Naïve Bayes, regression analysis, and random forest—to predict and diagnose cardiovascular disease. The work in [126] aimed to explore the utilization of machine learning in predicting cardiovascular attacks based on historical health records of patients. The study concentrated on the application of the random forest and CNN for prediction purposes. The results suggested that random forest exhibited superior performance compared to other classifiers, particularly in terms of classification accuracy. In [127], an SVM based on a decision tree system has been proposed for cardiovascular disease prediction. The results demonstrated that the proposed approach acquired knowledge of decision boundaries with various orientations, thereby demonstrating greater flexibility in learning diverse datasets.

In [128] a back-propagation neural network and Logistic Regression algorithms have been proposed for cardiovascular disease prediction. The results showed that the back-propagation neural network and Logistic Regression algorithms achieved an accuracy of 85.074% and 92.58%, respectively. The work in [129] encompasses two domains: signal processing and statistical learning. Leveraging signal processing techniques, authors have successfully segmented and represented each heartbeat through a vector of distinctive characteristics. Then, a cardiac electrocardiogram signal (ECG) is used, which is collected by sensors and proposed an SVM and Cuckoo search-optimized neural network algorithms for automatic cardiovascular disease detection. The work in [130] developed a smart scoring system and investigated the use of SVM for cardiac arrest prediction. The results showed that the intelligent scoring system has proven its capability to produce risk scores that are comprehensible to humans and has exhibited efficacy as a robust predictor of cardiac arrest occurring within a 72-hour timeframe. The work in [131] proposed a CNN with a depth of 9 layers to autonomously recognize five distinct categories of heartbeats, which are ventricular ectopic, supraventricular ectopic, unknown beats, non-ectopic, and fusion, within ECG signals. The results demonstrated that after training the CNN with augmented data, it achieved an accuracy of 94.03% and 93.47% in classifying heartbeats diagnostically in both original and noise-free ECGs, respectively.

The work in [132] introduced an innovative method for heartbeat recognition, employing a principal component analysis network for feature extraction from noisy ECG signals. To this end, an SVM algorithm is proposed. The work in [133] proposed KNN and decision tree algorithms for cardiovascular disease diagnosis. The work in [134] proposed a Logistic Regression technique for predicting cardiovascular disease within a cardiovascular dataset by employing dimensionality reduction, aiming to enhance the accuracy of the prediction. To this end, a feature scaling approach is to make the essential features fit with Logistic Regression to analyze the performance. In [65], the ANN algorithm and PCA-based technique are explored to improve the accuracy of cardiovascular disease prediction. The results showed that the accuracy has been improved achieving 97.7%. In [135], the prediction of cardiovascular disease through data mining methods is examined. A thorough evaluation of various techniques has been conducted, including the KNN algorithm, decision tree algorithm, neural network classifications, and Bayesian classification algorithms. Additionally, the utility of GA for feature selection is investigated to identify crucial cardiovascular disease-related features. The experiments revealed that the decision tree model achieved a high level of accuracy. The work in [136] utilized an unsupervised clustering method to assess cardiac involvement in systemic sclerosis, revealing unknown connections between samples for cardiovascular disease prediction. The work recommended employing big data methods such as HDFS and SVM for optimal attribute detection. The study also explored various data mining algorithms for cardiovascular disease detection, suggesting storing extensive data across nodes with HDFS and implementing the prediction algorithm using SVM across multiple nodes simultaneously.

In [137], the application of a backpropagation neural network for predicting cardiovascular disease is explored. The study involved the use of a deep-learning model in disease prediction. The work employed a neural network for both learning and prediction, utilizing the Cleveland dataset. To enable a real-time diagnosis approach, the work in [40] proposed an integration of a deep learning algorithm into the 6G-enabled Internet of Medical Things (IoMT) to facilitate online sharing of medical records and monitoring results. The work in [138] investigated the prediction of cardiovascular disease using various machine-learning techniques. The study involved the application of classification and regression models, specifically the Decision tree, KNN algorithm, SVM, and linear regression. Experimental results demonstrated that the KNN algorithm achieved the highest level of accuracy in cardiovascular disease prediction. In [139], five different machine learning algorithms for cardiovascular disease prediction have been employed, which are Logistic Regression, KNN, SVM, Decision tree, and Random Forest. To this end, a new hyperparameter tuning model has been proposed and compared with conventional approaches. The experimental results demonstrated that the proposed prediction approach achieved high accuracy in making predictions related to cardiovascular disease for the considered scenario. The work in [140] investigated cardiovascular disease prediction using four different datasets: Cleveland, StatLog, Hungarian, and Z-Alizadeh Sani. To this end, a two-tier ensemble PSObased feature selection was proposed.

In [141] two supervised machine learning algorithms, which are KNN and Random Forest, were exploited to predict cardiovascular disease prediction. Besides, KNN, decision trees, and random forests were investigated in [142]. A new approach was proposed in [143] to predict cardiovascular disease, utilizing a hybrid technique that combines a decision tree and an artificial neural network. To this end, a data mining-based approach in [144] was employed to explore, analyze, and extract data from various fields, unveiling meaningful information. In other words, raw data is transformed into valuable and useful information. An approach based on Naïve Bayes and decision tree algorithms have been implemented in [145] to predict cardiovascular disease, aiming for a concise diagnosis and analysis using a minimal set of attributes. Researchers frequently employ diverse feature selection techniques in conjunction with machine learning models for cardiovascular disease diagnosis. For example, in [95], the combination of SVM and an AI method was utilized to identify crucial features in classifying cardiovascular disease. The genetic algorithm serves as the feature selection technique, complementing the SVM, which functions as the classification algorithm. A cardiovascular disease diagnosis system was proposed in [146] using a fusion of rough sets-based attribute reduction and fuzzy logic. The work in [147] introduced an approach based on the ensemble Quine McCluskey Binary Classifier (QMBC) to identify patients with cardiovascular disease.

A new hybrid machine learning model technique based on the combination of Random Forest and Decision Tree techniques has been proposed in [148]. The results revealed an accurate prediction rate for cardiovascular disease using the hybrid model. In addition, a technique named (CardioHelp) has been introduced in [149], which employs a deep learning CNN approach to estimate the likelihood of cardiovascular disease in a patient. The experimental results demonstrated that the proposed method achieved an accuracy of 97%, which outperforms the existing approaches. Various machine learning algorithms are employed to build predictive models for cardiovascular disease prediction. Some popular algorithms used for this purpose are discussed above. Each one of these algorithms has its unique characteristics, strengths, and weaknesses, making them suitable for different scenarios. Each one of these models leverages distinct algorithms and methodologies, offering unique advantages and insights. From the simplicity of Logistic Regression to the intricacy of deep neural networks, these models have been pivotal in advancing our ability to assess and mitigate the risk of cardiovascular disease [150].

Noting that the choice of algorithm depends on various factors, such as the size of the dataset, the quality of the dataset, and the need for interpretability. Machine learning models, including neural networks, hold the promise of accurately diagnosing coronary cardiovascular disease [151]. Some algorithms, like Decision Trees and Random Forests, are favored when interpretability is crucial. SVMs and KNN can be effective when dealing with complex, high-dimensional data. Logistic Regression and Naïve Bayes are simple and suitable for binary classification tasks. It is also worth noting that ensemble methods, like Random Forests, are often preferred in cardiovascular disease prediction to combine the strengths of multiple algorithms and improve overall model performance. The choice of algorithm should be made based on a thorough understanding of the data and the specific goals of the cardiovascular disease prediction task. Recent results developed by [12], indicated that deep learning algorithm surpasses the capabilities of traditional machine learning methods, especially when it comes to handling large datasets, enabling the simultaneous training of substantial amounts of data for high prediction accuracy. The deep learning method consistently delivers accuracy levels of 85%. Moreover, combining various techniques within a single model is expected to yield promising forecasting results. For instance, machine learning models employing supplementary techniques like feature selection can achieve a maximum accuracy of 97% as stated in [12].

Table A4 provides comparisons between different algorithms with different datasets that are used for cardiovascular disease prediction. The comparison is carried out in terms of accuracy, precision, recall, and F1-score.

4. Data Collection and Preprocessing

This section initiates with a discussion about data collection in healthcare systems. Then, an explanation of data preprocessing techniques is provided.

4.1. Data Collection

Data collection in healthcare systems refers to the process of gathering and recording information about patients, medical conditions, treatments, and various healthcare-related factors. Typically, providing up-to-date information regarding the patient's condition can be very helpful to medical professionals. The availability of such information allows a reliable cardiovascular disease prediction to be achieved. The primary purpose of data collection in healthcare systems is to collect and maintain patient information to monitor health status, provide treatment, and make informed clinical decisions. In particular, this data is essential for managing patient care, enabling timely decisions using the patterns that exist in the data, healthcare administration, developing strategies for health promotion and disease prevention, and decision-making within the healthcare system [37].

Data collection in healthcare can involve various sources, including electronic health records (EHRs), ECG, IoT devices that can be kept in a body, time-series data, clinical assessments, medical records, wearable devices, patient-reported data, and medical imaging. EHRs in particular contain a wealth of patient information, including demographics, medical history, vital signs, and diagnostic tests [152]. EHRs are a primary source of healthcare data for cardiovascular disease prediction. Wearable devices like smartwatches and fitness trackers can collect real-time data on heart rate, activity levels, and sleep patterns. These devices are used to collect data and monitor and predict cardiovascular disease risk. Medical imaging, including echocardiograms, MRIs, and CT scans, provides detailed information about cardiac structure and function. Note that integrating data from such various sources allows for more comprehensive patient profiling and accurate predictions [153].

The data produced by the sensors exhibit the traits of significant volume, speed, and diversity, typical of big data [154]. Collecting patient-reported data, such as symptoms and lifestyle factors, can engage patients in their care and help healthcare professionals better understand individual health needs. The collected data can be structured (e.g., numerical measurements and categorical information) or unstructured (e.g., clinical notes and medical images) [32]. This data can be analyzed to identify trends, assess outcomes, and develop new medical insights.

Data collection is often necessary to comply with healthcare regulations and quality reporting requirements. Hence, accurate and secure data handling is critical for regulatory compliance. Healthcare data can also have missing values, errors, and inconsistencies. Therefore, ensuring data quality is critical for accurate modeling. Current works in healthcare systems [155–157] predict data uniformly, so neglecting urgency. This indeed would cause delays in treating severe patient conditions. Besides, managing, storing, and processing this data in real time can be a significant challenge. To this end, in [158] medical decision assistance is defined as furnishing clinicians with intelligently filtered computer-generated clinical knowledge and patient-related information to improve patient care. Various clinical databases are commonly used for cardiovascular disease prediction. These databases include Cleveland cardiovascular disease dataset obtained from the University of California Irvine (UCI) [159], the Framingham cardiovascular disease prediction dataset [160], Cardiovascular Disease dataset [161], Physikalisch Technische Bundesanstalt (PTB) diagnostic ECG dataset [162], and the stroke prediction dataset [163]. Researchers often use these databases for cardiovascular disease research.

Overall, data collection in healthcare systems can vary widely. There are some technical challenges related to healthcare data collection such as data privacy, reliability, and security. In particular, healthcare data is highly sensitive, and patient privacy is a paramount concern. Complying with regulations like the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) is crucial to protecting patient information [164–167]. Therefore, data collected should be accurate, secure, and privacy-compliant, as it plays an essential role in patient care and healthcare management.

4.2. Data Preprocessing Techniques

In healthcare systems, managing extensive databases becomes challenging, and hence, data preprocessing techniques become necessary. Data preprocessing may involve data creating, transforming, data cleaning, and data reading to improve model performance. Data preprocessing may also involve image normalization, noise reduction, data splitting, and standardizing image sizes to ensure consistency. Data preprocessing is essential for the best representation of data in machine learning [168]. To ensure effective training models, techniques such as handling missing values, standard scaling (Standscale (SS)), MaxAbs, quantile transformer, normalization known as (zero-mean normalization), robust scaler, and min-max (MinMax) scaling can be employed on the dataset [30,62,169]. Other techniques such as replacing missing values with estimates, cleaning data, removing rows or columns with too many missing values, and predictive modeling can also be used for data preprocessing [170]. The work in [171] has excluded independent variables (symptoms), which may have minimal or no impact on the target variable (disease), to simplify the analysis. In general, the numerical features of the dataset are normalized. This prevents certain features from dominating the modeling process. Missing values are addressed by simply removing the corresponding rows from the dataset.

In addition, data augmentation can involve techniques like rotation, scaling, and flipping to increase the training data and reduce the risk of overfitting [172]. Oversampling techniques like Synthetic Minority Over-sampling Technique (SMOTE), random oversampling (ROS), and adaptive synthetic sampling (ADASYN) can be used to address the imbalanced data for efficient cardiovascular disease prediction [173–175]. Data augmentation can also include creating composite features, data normalization, one-hot encoding categorical variables, and extracting relevant information from unstructured data [30]. Split the dataset into training, validation, and test sets to evaluate model performance. Further discussion about data preprocessing for cardiovascular disease can be found in [168]. Figure 2 shows the structure of the data preprocessing and machine learning model applications in cardiovascular disease prediction.

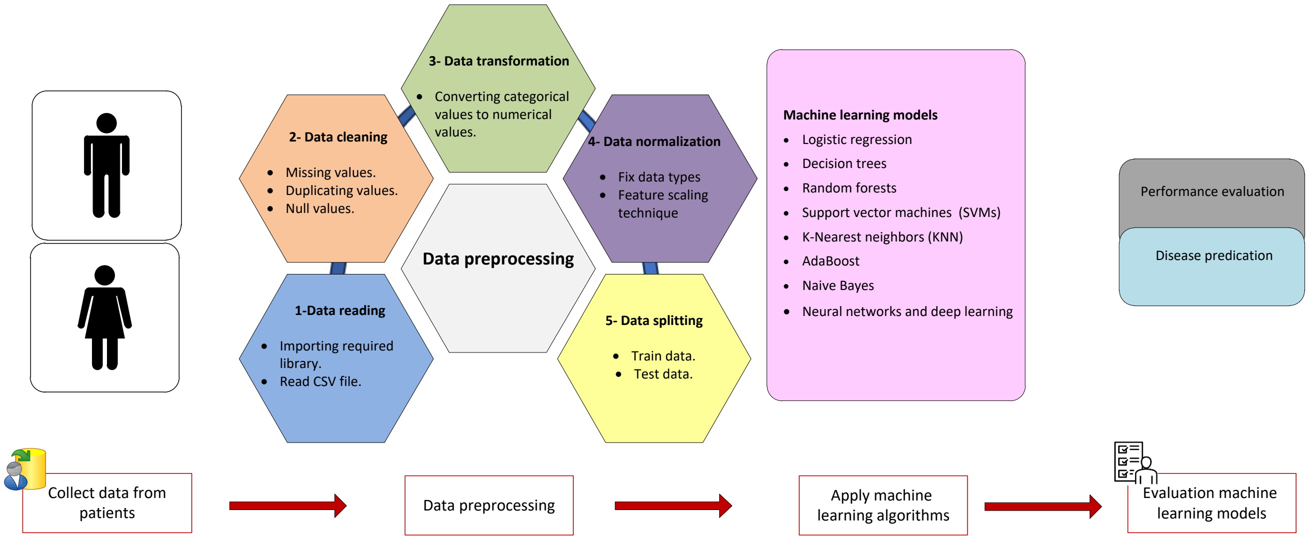

Figure 2. Structure of the data preprocessing and machine learning model applications.

5. Evaluation Metrics and Cross-Validation for Cardiovascular Disease Prediction

This section outlines the commonly used evaluation metrics for predicting cardiovascular diseases and provides the mathematical formulations for each of these metrics. Furthermore, this section provides a brief discussion about model validation and cross-validation and sheds light on the significance of interpretability and explainability for machine learning in the healthcare system.

5.1. Evaluation Metrics

When assessing the performance of cardiovascular disease prediction models, it is essential to use a range of evaluation metrics that provide a comprehensive view of their effectiveness. Specifically, assessing the performance of machine learning models in cardiovascular disease prediction is crucial for determining their effectiveness in clinical applications. Various evaluation metrics have been used to measure the model's performance and to evaluate the effectiveness of classifiers [176]. These metrics are computed using the confusion matrix. The common evaluation metrics used to assess the performance of machine learning models in cardiovascular disease prediction, including accuracy, recall (sensitivity), specificity, precision, F1-score, Matthews correlation coefficient (MCC), the area under the curve (AUC) and receiver operating characteristic (ROC) curve [29,62,176–180]. In cardiovascular disease prediction, evaluation criteria are crucial [17]. Table 3 describes the key of each of the evaluation metrics as well as provides the mathematical formulation of each of the evaluation metrics, which can be used to evaluate the effectiveness of machine learning algorithms for cardiovascular disease.

The abbreviations that are commonly used in cardiovascular disease prediction are given as follows. TP (True Positive): When the predicted output is identified as true positive (TP), it indicates that the subject with cardiovascular disease is correctly classified, confirming the presence of cardiovascular disease. TN (True Negative): In the case of a predicted output classified as true negative (TN), it signifies the accurate classification of a healthy subject, correctly identifying them as not having cardiovascular disease. FP (False Positive): If the predicted output is false positive (FP), it implies the misclassification of a healthy subject, incorrectly indicating that they have cardiovascular disease. FN (False Negative): When the predicted output is false negative (FN), it indicates the misclassification of a subject with cardiovascular disease as healthy, incorrectly suggesting the absence of cardiovascular disease.

Table 3. Evaluation metrics that are utilized to investigate the effectiveness of machine learning algorithms in cardiovascular disease prediction.

Evaluation Metric	Mathematical Equation	Description
Accuracy	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$	Accuracy is a basic evaluation metric and represents the proportion of correct predictions out of the total predictions. While accuracy is a useful metric, it may not be sufficient for imbalanced datasets, where one class (e.g., healthy patients) dominates the other (e.g., patients with cardiovascular disease).
Recall (Sensitivity)	$Sensitivity = \frac{TP}{TP+FN}$	Sensitivity measures the model's ability to correctly identify positive cases (individuals with cardiovascular disease). High sensitivity is crucial in healthcare to minimize false negatives, ensuring that individuals with cardiovascular disease are not missed.
Specificity	$Specificity = \frac{TN}{TN+FP}$	Specificity measures the model's ability to correctly identify negative cases (individuals without cardiovascular disease). High specificity is vital to minimize false positives, reducing unnecessary interventions for individuals without cardiovascular disease.
Precision	$Precision = \frac{TP}{TP+FP}$	Precision is the proportion of true positive predictions out of all positive predictions. It measures the model's accuracy when it predicts positive cases. Higher precision indicates a lower rate of false positives. Precision is important when minimizing false positives is crucial, as in medical diagnosis, where a false positive can lead to unnecessary treatments or anxiety.
F1-score	$F1\text{-}score = \frac{2TP}{2TP+FP+FN}$	The F1-score is the harmonic mean of the precision and recall (sensitivity). It provides a balanced measure that considers both false positives and false negatives. It is beneficial when there is an imbalance between the positive and negative classes.
Matthews correlation coefficient (MCC)	$MCC = \frac{(TPTN-FPFN)}{\sqrt{((TP+FP)(TP+FN)(TN+FP)(TN+FN))}}$	Denotes the predictive capacity of a classifier, expressed through values ranging from -1 and $+1$. For example, an MCC classifier value of $+1$ signifies ideal predictions, while -1 indicates entirely inaccurate predictions. A value close to 0 suggests that the classifier is making predictions randomly.
Area under the curve (AUC)	$AUC = \frac{1}{2}(\frac{FP}{(FP+TN)} + \frac{TN}{TN+FP})$	The AUC represents the region beneath the receiver operating characteristic (ROC) curve, a graphical depiction of the true positive rate against the false positive rate. A higher AUC value corresponds to superior model performance, with an ideal model approaching an AUC close to 1.

Different medical conditions and scenarios require different trade-offs between sensitivity and specificity. For example, in a cardiac emergency setting, high sensitivity may be more important to detect as many cases as possible. In contrast, in routine screenings, a balance between sensitivity and specificity may be more appropriate. Imbalanced datasets could also pose a technical challenge to the prediction model. Imbalanced datasets are common in healthcare applications, including cardiovascular disease prediction. Hence, it is also important to address the challenges of imbalanced datasets to ensure that cardiovascular disease prediction models are both accurate and clinically relevant. Specifically, addressing the imbalance datasets is essential to prevent models from becoming overly biased toward the majority class [123]. In such datasets, one class (e.g., patients with cardiovascular disease) is significantly smaller than the other (e.g., healthy patients). This can lead to challenges, such as biased models, misleading accuracy, focus on specific metrics, and resampling techniques. For example, models trained on imbalanced data may exhibit a bias toward the majority class, leading to poor performance in detecting the minority class [123]. Accuracy can be misleading in imbalanced datasets, as a model that predicts all instances as the majority class can still achieve high accuracy. In imbalanced datasets, metrics like precision, recall, and F1-score become more important as they provide insights into the model's performance on the minority class. Techniques like oversampling (increasing the size of the minority class) or undersampling (reducing the size of the majority class) can be used to address the imbalance issue. Therefore, assessing machine learning models for cardiovascular disease prediction requires a combination of general and specific evaluation metrics, considering the trade-off between sensitivity and specificity. Balancing the trade-off between minimizing false positives and false negatives is particularly significant in healthcare applications. By adjusting the classification threshold, one can balance sensitivity and specificity according to the specific requirements of the application. In addition, ensemble techniques like bagging and boosting can help to improve the performance of models on imbalanced data by combining multiple models to make predictions.

5.2. Model Validation and Cross-Validation

Model validation and cross-validation are crucial for ensuring the robustness of cardiovascular disease prediction models [17,181–184]. It could help in assessing how well a model generalizes to new. Cross-validation techniques, such as *k*-fold cross-validation, split the dataset into multiple subsets, training the model on different portions and testing it on others [185,186]. This would help in identifying the potential overfitting and provide a more reliable estimate of a model's performance, ensuring it can make accurate predictions for diverse patient diseases.

5.3. Model Interpretability and Explainability

There is an essential need for making machine learning models more interpretable and transparent, especially in the healthcare system. In particular, one of the main challenges in machine learning methods is dealing with complex models that are often considered black boxes [172,187]. While machine learning methods have demonstrated exceptional predictive power, understanding their decision-making processes can be a very challenging issue [188,189]. This is a major concern in healthcare systems, where decisions need to be justified and trusted. In other words, it is important to understand why a machine-learning model makes certain predictions. Therefore, model interpretability and explainability become particularly essential in healthcare, especially for cardiovascular disease prediction. Specifically, In healthcare systems, the significance lies not only in the quantitative algorithmic performance but also in the essential features that the algorithm employs for disease detection [187]. Hence, incorporating interpretability and explainability for machine learning models enhances the practical application of such models in real-world scenarios [190]. Understanding which features had the most influence on a prediction is a fundamental form of interpretability.

Techniques like feature importance scores can help in identifying the most significant predictors in cardiovascular disease prediction [191]. Considering simpler and more interpretable models like Decision Trees or Logistic Regression, especially when clinical decision-making requires transparency. Employ local interpretable models like Local Interpretable Model-Agnostic Explanations (LIME) or SHapley Additive exPlanations (SHAP), which is a game theoretic approach, to explain individual predictions [192–198]. These models provide explanations and visualizations for specific instances, which can be valuable in healthcare decision-making since they also help healthcare professionals to understand model outputs [199,200]. Applying such models can involve creating rule-based systems that align with medical expertise and regulations, ensuring that model predictions are consistent with established healthcare standards. If a model produces predictions that contradict established clinical knowledge, it should raise a red flag and prompt further investigation. Models may drift or degrade over time, and ongoing vigilance, especially in real-time systems, is essential to ensure they remain trustworthy [201].

In a nutshell, interpretability, and explainability are crucial for ensuring trust and accountability in clinical decision-support systems. Ensuring that models can be understood and trusted is vital for making responsible and effective clinical decisions [202].

6. Recent Trends in Machine Learning for Cardiovascular Disease Prediction and Future Works

This section provides a list of some open problems and suggestions regarding the application of machine learning in cardiovascular disease prediction and highlights some further suggestions that can be considered for future work.

- Delays in diagnosing cardiac disease continue to have a major impact on the treatment of patients [123]. More efficient classification and early prediction of cardiovascular disease methods are required.
- Deep learning techniques have gained prominence in medical image analysis and ECG data interpretation due to their significant role in providing more accurate diagnosis and prediction of diseases [31]. However, the adoption of such techniques in clinical practice comes with some challenges. These challenges include data privacy, the need for large and diverse datasets, regulatory compliance, model robustness, and generalization issues [203]. Therefore, future research should address these challenges. Furthermore, the interpretability of deep learning models remains an ongoing research topic, as understanding the inner workings of complex neural networks is crucial for clinical acceptance.
- Explainable AI (XAI) is essential in cardiovascular disease prediction, attracting significant research attention [204]. XAI techniques empower clinicians by elucidating the importance of each feature in predictions, enabling informed decision-making and building trust [187]. Collaboration between healthcare providers and AI-driven systems can be facilitated through XAI, translating complex model outputs into actionable insights and enhancing clinicians' diagnostic and treatment capabilities. Further research studies in this essential direction are needed.
- The transformative impact of IoT technology on healthcare is evident, enabling remote patient monitoring and facilitating telemedicine through wireless sensors [205–207]. The integration of machine learning models with wearable devices, IoT sensors, and mobile health applications for continuous cardiovascular disease monitoring represents a compelling research avenue [32,208]. Exploring the efficiency of remote medical applications by employing machine learning algorithms is crucial for enhancing data analysis, increasing reliability, and achieving fast and accurate decision-making. Further research in this domain is essential for advancing healthcare systems.
- Machine learning can enhance risk assessment and treatment recommendations by analyzing the historical data of patients [209,210]. As such, machine learning can identify early warning signs of cardiovascular disease, enabling timely interventions [211,212]. Therefore, future works should focus on developing more accurate models to enhance risk assessment and treatment recommendations.
- The Generative Adversarial Network (GAN) stands out as a widely embraced technique in the machine learning domain [73,213]. Leveraging this method enables the creation of synthetic data closely resembling real data, making GAN a promising solution for addressing challenges associated with data scarcity in cardiovascular discussions. Hence, understanding the intricate connections between GAN and data privacy could be explored in future studies. Besides, utilizing GAN to classify unbalanced signals in fetal cardiovascular rate data can be investigated in future [214].

- Efforts are being made to improve interoperability and data sharing among healthcare systems and institutions [215]. The ability to access and share medical data is crucial for training efficient machine learning models. Research is ongoing to improve the robustness and generalization of machine learning models in healthcare systems. Future research could involve further investigation into interoperability and data sharing among healthcare systems.
- Aggregating larger data at a central server confronts issues. To address these issues, the idea of federated learning is introduced, where the focus is on sharing model knowledge instead of sharing raw data [216]. Federated learning establishes a more secure system by enabling model training on decentralized devices, preserving privacy, and minimizing the risk of centralized data breaches [217]. This approach results in a robust system with enhanced security and data access controls that safeguard privacy [218]. A suggested future work would involve the development and implementation of a standardized, interoperable, and secure federated learning framework for data sharing. This framework would allow multiple healthcare institutions and researchers to collaborate while preserving patient privacy [219].
- The management of vast historical data and the constant influx of streaming data in healthcare services pose a formidable challenge for conventional database storage and machine learning approaches. Addressing this challenge in real-time data processing has led researchers to explore big data approaches [220–222]. Further research studies in this direction are needed.
- Machine learning involves predictive models that analyze patient data for early detection. Future research should refine predictive models for early anomaly detection and develop advanced algorithms for more accurate predictions. Efforts should be focused on creating personalized care plans and improving machine learning integration into telemedicine platforms for informed decisions during remote consultations.
- Future work should address the ethical concerns issue in healthcare systems. In particular, Strategies to mitigate bias and enhance fairness in healthcare systems should be developed. This includes the development of novel algorithms and methodologies that prioritize equity and fairness.
- Achieving the best possible level of security and privacy protection is important, which need to be considered in the future [223]. Data sharing and collaboration among healthcare institutions are crucial for building large, diverse datasets. This collaboration allows the development of more accurate and robust machine-learning models for medical applications. However, ensuring data privacy, handling data heterogeneity, and addressing issues like missing values and imbalanced datasets is crucial, and should be investigated in the future. In addition, collecting and preprocessing healthcare data for cardiovascular disease prediction is still a challenging issue that needs to be addressed.
- As contemporary technology continues to advance, the acquisition of high-resolution and multidimensional data becomes increasingly feasible. In dealing with such high-quality data, the conventional machine-learning approaches may exhibit some limitations. Exploring the potential of employing a blend of multiple machine learning models could prove to be a promising avenue for addressing the challenges posed by high-dimensional data in future research.
- To advance the application of machine learning in healthcare systems, future work should focus on establishing strong collaborative networks between data scientists, clinicians, and domain experts [224]. This collaboration will ensure that machine learning models are designed and tuned to enhance patient care and decision-making.
- Future research should emphasize the potential of machine learning to address cardiovascular disease on a global scale, with a specific focus on regions with limited access to healthcare resources.

- Future research should also aim to bridge healthcare disparities by providing accessible and effective cardiovascular disease prediction and management tools to underserved populations.
- Artificial General Intelligence (AGI), or strong AI, mimics human intelligence by learning, reasoning, generalizing, and exhibiting self-awareness [225,226]. AGI may have a significant impact on cardiovascular disease, due to its adaptability in accomplishing diverse tasks, integrating medical knowledge, and personalizing treatment planning [227]. AGI can achieve continuous learning and adaptability to ensure up-to-date information for managing cardiovascular disease. Research efforts should aim to enhance AGI cognitive abilities and reasoning.

7. Conclusions

Cardiovascular disease among other diseases stands as the primary contributor to worldwide mortality. According to recent statistics, approximately 17.9 million individuals lost their lives due to cardiovascular diseases, accounting for 32% of all fatalities on a global scale. Strategies for prevention and early detection, coupled with advancements in medical technology, including the utilization of advanced artificial intelligence techniques, play a vital role in minimizing the influence of cardiovascular disease on public health systems. Early identification and efficient management of cardiovascular disease can markedly alleviate the strain on healthcare systems globally. To this end, machine learning techniques can play an essential role in advancing cardiovascular disease prediction and patient care, hence contributing significantly to the healthcare systems. Machine learning technology offers several key advantages that improve the accuracy, reliability, and efficiency of cardiovascular disease detection and management. This paper provided a current perspective by covering the latest trends and advancements in the role of machine learning for cardiovascular disease prediction. In particular, this paper provided a comprehensive perspective on the role of machine learning in predicting cardiovascular disease and its implications for public health. This review paper covered a wide range of topics, spanning the assessment of machine learning models, the importance of machine learning, the prevalence of cardiovascular disease and its various types, feature selection, data collection, and preprocessing. Additionally, this paper explained the evaluation metrics used for predicting cardiovascular disease and explored recent trends in this field. Based on the findings of this paper, we emphasize that the multidimensional impact of machine learning, from early detection to personalized treatment, predictive analytics, and real-time monitoring, has the potential to reduce the burden of cardiovascular disease.

Author Contributions: Conceptualization, M.A.; methodology, M.A. and M.A.N.; validation, M.A.N. and M.A.; investigation, M.A.N. and A.A.M.; writing—original draft preparation, M.A. and M.A.N.; writing—review and editing, M.A.N., M.A., T.R.A.-S., A.A.M. and K.M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the University of Baghdad and Al-Farahidi University for their general support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This Appendix provides Tables showing the attributes of three common datasets for cardiovascular disease prediction.

Table A1. The attributes of the cardiovascular datasets used for cardiovascular disease prediction.

ID	Attribute	Type of Attribute	Values
1	Id	Discrete	Unique identifier
2	Age	Discrete	Age of patient in days
3	Gender	Discrete	Female = 1, male = 2
4	Height	Discrete	In cm
5	Weight	Continuous	In kg
6	Ap hi	Discrete	Systolic blood pressure
7	Ap low	Discrete	Diastolic blood pressure
8	Cholesterol	Discrete	1 = normal, 2 = above normal, 3 = well above normal
9	Gluc	Discrete	1 = normal, 2 = above normal, 3 = well above normal
10	Smoke	Binary	Whether patient smokes or not (yes = 1, no = 0)
11	Alcohol	Binary	Whether patient drinks or not (yes = 1, no = 0)
12	Active	Binary	Physical activity (yes = 1, no = 0)
13	Cardio	Binary	Presence or absence of cardiovascular disease (yes = 1, no = 0)

Table A2. The attributes of the Cleveland dataset for cardiovascular disease prediction.

ID	Attribute	Type of Attribute	Values
1	Sex/gender	Discrete	Male = 1 or female = 0
2	Age	Continuous	Age of patient in years
3	Cp (chest pain)	Discrete	1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic
4	RestBP (resting blood pressure)	Continuous	90–200
5	Chol (cholesterol level)	Continuous	126–564
6	Fbs (fasting blood sugar)	Discrete	Fasting blood sugar > 120 mg/dL 1 = true, 0 = false
7	Restecg (resting Electrocardiography results)	Discrete	0 = normal, 1 = ST-T wave abnormality, 2 = showing probable or defined left ventricular hypertrophy by Estes criteria
8	Thalach (maximum heart rate achieved)	Continuous	71–202
9	Exang (exercise-induced angina)	Discrete	Yes = 1 or no = 0
10	Old peak ST (depression level)	Continuous	0 to 6.2
11	Slope (slope of the peak exercise segment)	Discrete	1 = upward sloping, 2 = flat, 3 = downward sloping
12	Ca (fluoroscopy value)	Discrete	0 to 3
13	Thal (severity of chest pain or trouble breathing)	Discrete	3 = normal, 6 = fixed defect, 7 = reversible defect
14	Target	Discrete	Yes = 1 or no = 0

Table A3. The attributes of the Framingham dataset for cardiovascular disease prediction.

ID	Attribute	Type of Attribute	Values
1	Sex	Nominal	Male = l or female = 0
2	Age	Continuous	Age of patient in the whole number
3	Education	Continuous	Values = 1–4. Some High School = 1, High School or GED = 2, Some College or Vocational School = 3, College = 4
4	Current Smoker	Nominal	Yes = 1 or no = 0
5	Cigarettes per day	Continuous	Number of cigarettes smoked per day
6	BP Meds	Nominal	Yes = 1 or no = 0 was BP patient or not
7	Prevalent Stroke	Nominal	Yes = 1 or no = 0 was stroke patient or not
8	Prevalent Hyp	Nominal	Yes = 1 or no = 0, whether the patient was hypertensive
9	Diabetes	Nominal	Yes = 1 or No = 0 was diabetes patient or not
10	Tot Chol	Continuous	Total cholesterol level
11	Sys BP	Continuous	Systolic blood pressure
12	Dia BP	Continuous	Diastolic blood Pressure
13	BMI	Continuous	Body mass index
14	Heart Rate	Continuous	Heart rate or pulse rate
15	Glucose	Continuous	Glucose level
16	Ten-Year CHD (Target)	Nominal	Yes = 1 or no = 2, the 10-year risk of coronary heart disease (CHD)

Appendix B

This Appendix provides a comparison between different algorithms with different datasets that are used for cardiovascular disease prediction. Noting that the approbations in the table are given as follows: Classification and regression tree algorithm (CART), heart disease dataset (IEEE Dataport), Machine Learning based Cardiovascular Disease Diagnosis (MaLCaDD), hybrid random forest with a linear model (HRFLM), accuracy (ACC), precision (Pr), recall (Re), and F1-score (F1).

Table A4. Comparison of different algorithms with different datasets that are used for cardiovascular disease prediction.

Paper	Year	Dataset Used	Algorithms Used	ACC%	Pr%	Re%	F1%
[228]	2023	Cleveland	LR, KNN, DT, XGB, SVM, RF	79.12%	79%	79%	79%
[228]	2023	Comprehensive UCI datasets	LR, KNN, DT, XGB, SVM, RF	99.03%	99%	99%	99%
[29]	2023	Cleveland	Soft voting ensemble based on (RF, KNN, LR, NB, GB, AB)	93.44%	NP	NP	NP
[29]	2023	IEEE Dataport	Soft voting ensemble based on (RF, KNN, LR, NB, GB, AB)	95.00%	NP	NP	NP
[229]	2023	IEEE Dataport	CART	87.25%	88.24%	84.51%	NP
[230]	2023	Cardiovascular Disease dataset	RF, DT, MLP, and XGB	87.28%	88.70%	84.85%	86.71%
[147]	2023	Cleveland	Quine McCluskey Binary Classifier (QMBC) (LR, DT, RF, KNN, NB, SVM, and MLP)	98.36%	100%	97.22%	98.59%

Table A4. *Cont.*

Paper	Year	Dataset Used	Algorithms Used	ACC%	Pr%	Re%	F1%
[147]	2023	Comprehensive UCI datasets	Quine McCluskey Binary Classifier (QMBC) (LR, DT, RF, KNN, NB, SVM, and MLP)	98.31%	96.89%	100%	98.42%
[147]	2023	Cardiovascular Disease dataset	Quine McCluskey Binary Classifier (QMBC) (LR, DT, RF, KNN, NB, SVM, and MLP)	99.95%	100%	99.91%	99.95%
[231]	2023	Cleveland	Deep ANN, LSTM, CNN, and hybrid CNN with LSTM	97.75%	98.57%	97.87%	97.18%
[231]	2023	IEEE Dataport	Deep ANN, LSTM, CNN, and hybrid CNN with LSTM	98.86%	99.13%	99.42%	90.83%
[232]	2022	Cleveland	Stochastic Gradient Descent Classifiers, LR, SVM, NB, ConvSGLV, and ensemble methods	93.00%	NP	NP	NP
[233]	2022	IEEE Dataport	NN, MLPNN, AB, SVM, LR, ANN, RF	93.39%	NP	NP	NP
[234]	2022	Cleveland	NB, SVM, LR, DT, RF, and KNN	94.1%	97.1%	94.8%	90.8%
[235]	2022	Cleveland	NB, DT, LR KNN, SVM, GB, and RF algorithms	85.18%	0.83%	90%	86%
[236]	2022	Cleveland and Statlog	NB with weighted approach, 2 SVMs with XGBoost, an improved SVM (ISVM) based on duality optimization (DO) technique, and an XGBoost	95.9%	97.1%	94.67%	95.35%
[237]	2022	Heart disease dataset (IEEE Dataport)	Stacking-Based Ensemble Learning (XGB, ETs, RF, GB)	92.34%	92.00%	93.49%	92.74%
[238]	2021	PhysioNet's arrhythmia Dataset	SVM, KNN, RF, ETs, Bagging, DT, LR, and Adaptive Boosting	99.8%	100%	100%	100%
[238]	2021	UCI's Arrhythmia Dataset	SVM, KNN, RF, ETs, Bagging, DT, LR, and Adaptive Boosting	95.6%	93%	93%	93%
[239]	2021	Framingham	MaLCaDD using ensemble algorithm (10 fold)	99.1%	NP	NP	NP
[239]	2021	Cardiovascular Disease dataset	MaLCaDD using ensemble algorithm (10 fold)	98.0%	NP	NP	NP
[148]	2021	Cleveland	RF, DT, and hybrid model between RF and DT	88.7%	NP	NP	NP
[240]	2021	Cleveland, Hungary, Switzerland, and VA Long Beach and Statlog	Hybrid classifiers like (DTBM), (RFBM), (KNNBM), (ABBM), (GBBM)	99.05%	99%	98%	99%
[239]	2021	Cleveland	MaLCaDD using ensemble algorithm (10 fold)	95.5%	NP	NP	NP
[142]	2021	Comprehensive datasets (1025)	LR, ABM1, MLP, KNN, DT, RF	100%	100%	100%	100%
[140]	2020	StatLog	Two-tier ensemble PSO-based feature selection	93.55%	NP	NP	91.67%
[140]	2020	Hungarian	Two-tier ensemble PSO-based feature selection	91.18%	NP	NP	90.91%

Table A4. *Cont.*

Paper	Year	Dataset Used	Algorithms Used	ACC%	Pr%	Re%	F1%
[140]	2020	Cleveland	Two-tier ensemble PSO-based feature selection	85.71%	NP	NP	86.49%
[140]	2020	Z-Alizadeh Sani	Two-tier ensemble PSO-based feature selection	98.13%	NP	NP	96.90%
[139]	2020	Cleveland	LR, KNN, DT, SVM, RF	91.80%	93.55%	90.62%	92.06%
[57]	2020	Cardiovascular Disease dataset	DT, NB, LR, RF, SVM, and KNN	73%	75%	68%	73%
[141]	2020	Comprehensive dataset (1025)	RF, SVM, NB, and DT	99%	97.1%	99.7%	99.7%
[80]	2019	Cleveland	HRFLM	88.4%	90.1%	92.8%	90%
[125]	2018	Cleveland and Hungarian	NB, ANN, SVM, RF, LR	98.13%	98.1%	NP	98.1%
[3]	2017	Cleveland	Multi-Layer Perceptron Neural Network (hidden layer size = 8)	95.55%	95.45%	NP	95.45%

References

1. Atun, R. Transitioning health systems for multimorbidity. *Lancet* **2015**, *386*, 721–722. [CrossRef]
2. Panch, T.; Szolovits, P.; Atun, R. Artificial intelligence, machine learning and health systems. *J. Glob. Health* **2018**, *8*, 020303. [CrossRef]
3. Karayılan, T.; Kılıç, Ö. Prediction of heart disease using neural network. In Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017; pp. 719–723.
4. Nagendra, K.; Ussenaiah, M. A study on various data mining techniques used for heart diseases. *Int. J. Recent Sci. Res.* **2018**, *9*, 24350–24354.
5. Ahmed, H.; Younis, E.M.; Hendawi, A.; Ali, A.A. Heart disease identification from patients' social posts, machine learning solution on Spark. *Future Gener. Comput. Syst.* **2020**, *111*, 714–722. [CrossRef]
6. Kelly, B.B.; Fuster, V. (Eds.) *Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health*; National Academies Press: Washington, DC, USA, 2010.
7. Anjaneyulu, M.; Degala, D.P.; Devika, P.; Hema, V. Effective heart disease prediction using hybrid machine learning techniques. In *Proceedings of the AIP Conference Proceedings*; AIP Publishing: Melville, NY, USA, 2023; Volume 2492.
8. Ansarullah, S.I.; Kumar, P. A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method. *Int. J. Recent Technol. Eng.* **2019**, *7*, 1009–1015.
9. Nazir, S.; Shahzad, S.; Mahfooz, S.; Nazir, M. Fuzzy logic based decision support system for component security evaluation. *Int. Arab J. Inf. Technol.* **2018**, *15*, 224–231.
10. Dritsas, E.; Trigka, M. Efficient data-driven machine learning models for cardiovascular diseases risk prediction. *Sensors* **2023**, *23*, 1161. [CrossRef]
11. Dai, W.; Brisimi, T.S.; Adams, W.G.; Mela, T.; Saligrama, V.; Paschalidis, I.C. Prediction of hospitalization due to heart diseases by supervised learning methods. *Int. J. Med. Inform.* **2015**, *84*, 189–197. [CrossRef]
12. Bakar, W.A.W.A.; Josdi, N.L.N.B.; Man, M.B.; Zuhairi, M.A.B. A Review: Heart Disease Prediction in Machine Learning & Deep Learning. In Proceedings of the 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Kedah, Malaysia, 3–4 March 2023; pp. 150–155.
13. Ali, F.; El-Sappagh, S.; Islam, S.R.; Kwak, D.; Ali, A.; Imran, M.; Kwak, K.S. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf. Fusion* **2020**, *63*, 208–222. [CrossRef]
14. Xiao, F.; Miao, Q.; Xie, X.; Sun, L.; Wang, R. Indoor anti-collision alarm system based on wearable Internet of Things for smart healthcare. *IEEE Commun. Mag.* **2018**, *56*, 53–59. [CrossRef]
15. Tian, S.; Yang, W.; Le Grange, J.M.; Wang, P.; Huang, W.; Ye, Z. Smart healthcare: Making medical care more intelligent. *Glob. Health J.* **2019**, *3*, 62–65. [CrossRef]
16. Shah, T.; Yavari, A.; Mitra, K.; Saguna, S.; Jayaraman, P.P.; Rabhi, F.; Ranjan, R. Remote health care cyber-physical system: Quality of service (QoS) challenges and opportunities. *IET Cyber-Phys. Syst. Theory Appl.* **2016**, *1*, 40–48. [CrossRef]
17. Li, J.P.; Haq, A.U.; Din, S.U.; Khan, J.; Khan, A.; Saboor, A. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access* **2020**, *8*, 107562–107582. [CrossRef]
18. Li, W.; Chai, Y.; Khan, F.; Jan, S.R.U.; Verma, S.; Menon, V.G.; Kavita, f.; Li, X. A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mob. Netw. Appl.* **2021**, *26*, 234–252. [CrossRef]

19. Methaila, A.; Kansal, P.; Arya, H.; Kumar, P. Early heart disease prediction using data mining techniques. *Comput. Sci. Inf. Technol. J.* **2014**, *24*, 53–59.
20. Chaithra, N.; Madhu, B. Classification models on cardiovascular disease prediction using data mining techniques. *Cardiovasc. Dis. Diagn.* **2018**, *6*, 1–4.
21. Pazos-López, P.; Peteiro-Vázquez, J.; Carcía-Campos, A.; García-Bueno, L.; de Torres, J.P.A.; Castro-Beiras, A. The causes, consequences, and treatment of left or right heart failure. *Vasc. Health Risk Manag.* **2011**, *7*, 237–254.
22. Sadaka, M.; Aboelela, A.; Arab, S.; Nawar, M. Electrocardiogram as prognostic and diagnostic parameter in follow up of patients with heart failure. *Alex. J. Med.* **2013**, *49*, 145–152. [CrossRef]
23. Steinberg, Z.L.; Singh, H.S. How to plan and perform a diagnostic catheterisation in adult patients with congenital heart disease. *Heart* **2023**, *109*, 151–157. [CrossRef]
24. Ye, G.; Gamage, P.T.; Balasubramanian, V.; Li, J.K.J.; Subasi, E.; Subasi, M.M.; Kaya, M. Short-Term Risk Estimation and Treatment Planning for Cardiovascular Disease Patients after First Diagnostic Catheterizations with Machine Learning Models. *Appl. Sci.* **2023**, *13*, 5191. [CrossRef]
25. Chatterjee, P.; Cymberknop, L.J.; Armentano, R.L. IoT-based decision support system for intelligent healthcare—Applied to cardiovascular diseases. In Proceedings of the 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), Nagpur, India, 11–13 November 2017; pp. 362–366.
26. Alzubi, J.; Nayyar, A.; Kumar, A. Machine learning from theory to algorithms: An overview. *J. Phys.: Conf. Ser.* **2018**, *1142*, 012012. [CrossRef]
27. Elsayed, M.; Erol-Kantarci, M. AI-enabled future wireless networks: Challenges, opportunities, and open issues. *IEEE Veh. Technol. Mag.* **2019**, *14*, 70–77. [CrossRef]
28. Gacanin, H. Autonomous wireless systems with artificial intelligence: A knowledge management perspective. *IEEE Veh. Technol. Mag.* **2019**, *14*, 51–59. [CrossRef]
29. Chandrasekhar, N.; Peddakrishna, S. Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. *Processes* **2023**, *11*, 1210. [CrossRef]
30. Ahsan, M.M.; Mahmud, M.P.; Saha, P.K.; Gupta, K.D.; Siddique, Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies* **2021**, *9*, 52. [CrossRef]
31. Ahsan, M.M.; Luna, S.A.; Siddique, Z. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare* **2022**, *10*, 541. [CrossRef]
32. Yu, K.H.; Beam, A.L.; Kohane, I.S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2018**, *2*, 719–731. [CrossRef]
33. Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W.; Müller, K.R. Unmasking Clever Hans predictors and assessing what machines really learn. *Nat. Commun.* **2019**, *10*, 1096. [CrossRef]
34. Peltonen, E.; Bennis, M.; Capobianco, M.; Debbah, M.; Ding, A.; Gil-Castiñeira, F.; Jurmu, M.; Karvonen, T.; Kelanti, M.; Kliks, A.; et al. 6G White Paper on Edge Intelligence. *arXiv* **2020**, arXiv:2004.14850.
35. Aljanabi, M.; Qutqut, M.H.; Hijjawi, M. Machine learning classification techniques for heart disease prediction: A review. *Int. J. Eng. Technol.* **2018**, *7*, 5373–5379.
36. Xie, R.; Khalil, I.; Badsha, S.; Atiquzzaman, M. An intelligent healthcare system with data priority based on multi vital biosignals. *Comput. Methods Programs Biomed.* **2020**, *185*, 105126. [CrossRef] [PubMed]
37. Pustokhina, I.V.; Pustokhin, D.A.; Gupta, D.; Khanna, A.; Shankar, K.; Nguyen, G.N. An effective training scheme for deep neural network in edge computing enabled Internet of medical things (IoMT) systems. *IEEE Access* **2020**, *8*, 107112–107123. [CrossRef]
38. Vennemann, B.; Obrist, D.; Rösgen, T. Automated diagnosis of heart valve degradation using novelty detection algorithms and machine learning. *PLoS ONE* **2019**, *14*, e0222983. [CrossRef] [PubMed]
39. Martínez-Jiménez, M.A.; Ramirez-GarciaLuna, J.L.; Kolosovas-Machuca, E.S.; Drager, J.; González, F.J. Development and validation of an algorithm to predict the treatment modality of burn wounds using thermographic scans: Prospective cohort study. *PLoS ONE* **2018**, *13*, e0206477. [CrossRef] [PubMed]
40. Wang, W.; Liu, F.; Zhi, X.; Zhang, T.; Huang, C. An integrated deep learning algorithm for detecting lung nodules with low-dose CT and its application in 6G-enabled internet of medical things. *IEEE Internet Things J.* **2020**, *8*, 5274–5284. [CrossRef]
41. Khan, M.A.; Algarni, F. A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. *IEEE Access* **2020**, *8*, 122259–122269. [CrossRef]
42. Wang, S.; Summers, R.M. Machine learning and radiology. *Med. Image Anal.* **2012**, *16*, 933–951. [CrossRef]
43. Ravì, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.Z. Deep learning for health informatics. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 4–21. [CrossRef]
44. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [CrossRef]
45. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
46. Ahsan, M.M.; Alam, T.E.; Trafalis, T.; Huebner, P. Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. *Symmetry* **2020**, *12*, 1526. [CrossRef]
47. Degerli, A.; Zabihi, M.; Kiranyaz, S.; Hamid, T.; Mazhar, R.; Hamila, R.; Gabbouj, M. Early detection of myocardial infarction in low-quality echocardiography. *IEEE Access* **2021**, *9*, 34442–34453. [CrossRef]

48. Alloghani, M.; Al-Jumeily, D.; Mustafina, J.; Hussain, A.; Aljaaf, A.J. A systematic review on supervised and unsupervised machine learning algorithms for data science. In *Supervised and Unsupervised Learning for Data Science*; Springer: Cham, Switzerland, 2020; pp. 3–21.
49. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*, 3–24.
50. Libbrecht, M.W.; Noble, W.S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **2015**, *16*, 321–332. [CrossRef]
51. Fatima, M.; Pasha, M. Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* **2017**, *9*, 1–16. [CrossRef]
52. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
53. De Lannoy, G.; François, D.; Delbeke, J.; Verleysen, M. Weighted conditional random fields for supervised interpatient heartbeat classification. *IEEE Trans. Biomed. Eng.* **2011**, *59*, 241–247. [CrossRef] [PubMed]
54. Dutta, A.; Batabyal, T.; Basu, M.; Acton, S.T. An efficient convolutional neural network for coronary heart disease prediction. *Expert Syst. Appl.* **2020**, *159*, 113408. [CrossRef]
55. Banerjee, A.; Mohanta, B.K.; Panda, S.S.; Jena, D.; Sobhanayak, S. A secure IoT-fog enabled smart decision making system using machine learning for intensive care unit. In Proceedings of the 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 10–12 January 2020; pp. 1–6.
56. Mamun, M.; Farjana, A.; Al Mamun, M.; Ahammed, M.S.; Rahman, M.M. Heart failure survival prediction using machine learning algorithm: Am I safe from heart failure? In Proceedings of the 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 6–9 June 2022; pp. 194–200.
57. Princy, R.J.P.; Parthasarathy, S.; Jose, P.S.H.; Lakshminarayanan, A.R.; Jeganathan, S. Prediction of cardiac disease using supervised machine learning algorithms. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 13–15 May 2020; pp. 570–575.
58. Choudhury, R.P.; Akbar, N. Beyond diabetes: A relationship between cardiovascular outcomes and glycaemic index. *Cardiovasc. Res.* **2021**, *117*, e97–e98. [CrossRef] [PubMed]
59. Khan, A.; Qureshi, M.; Daniyal, M.; Tawiah, K. A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction. *Health Soc. Care Community* **2023**, *2023*, 1406060. [CrossRef]
60. Zhou, X.; Liang, W.; Kevin, I.; Wang, K.; Wang, H.; Yang, L.T.; Jin, Q. Deep-learning-enhanced human activity recognition for Internet of healthcare things. *IEEE Internet Things J.* **2020**, *7*, 6429–6438. [CrossRef]
61. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
62. Haq, A.U.; Li, J.P.; Memon, M.H.; Nazir, S.; Sun, R. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mob. Inf. Syst.* **2018**, *2018*, 3860146. [CrossRef]
63. Khemphila, A.; Boonjing, V. Heart disease classification using neural network and feature selection. In Proceedings of the 2011 21st International Conference on Systems Engineering, Las Vegas, NV, USA, 16–18 August 2011; pp. 406–409.
64. Takci, H. Improvement of heart attack prediction by the feature selection methods. *Turk. J. Electr. Eng. Comput. Sci.* **2018**, *26*, 1–10. [CrossRef]
65. Awan, S.M.; Riaz, M.U.; Khan, A.G. Prediction of heart disease using artificial neural network. *VFAST Trans. Softw. Eng.* **2018**, *6*, 51–61.
66. Gárate-Escamila, A.K.; El Hassani, A.H.; Andrès, E. Classification models for heart disease prediction using feature selection and PCA. *Inform. Med. Unlocked* **2020**, *19*, 100330. [CrossRef]
67. Martis, R.J.; Acharya, U.R.; Mandana, K.; Ray, A.K.; Chakraborty, C. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health. *Expert Syst. Appl.* **2012**, *39*, 11792–11800. [CrossRef]
68. Dolatabadi, A.D.; Khadem, S.E.Z.; Asl, B.M. Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. *Comput. Methods Programs Biomed.* **2017**, *138*, 117–126. [CrossRef] [PubMed]
69. Arabasadi, Z.; Alizadehsani, R.; Roshanzamir, M.; Moosaei, H.; Yarifard, A.A. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput. Methods Programs Biomed.* **2017**, *141*, 19–26. [CrossRef] [PubMed]
70. Long, N.C.; Meesad, P.; Unger, H. A highly accurate firefly based algorithm for heart disease prediction. *Expert Syst. Appl.* **2015**, *42*, 8221–8231. [CrossRef]
71. Sudarshan, V.K.; Acharya, U.R.; Oh, S.L.; Adam, M.; Tan, J.H.; Chua, C.K.; Chua, K.P.; San Tan, R. Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals. *Comput. Biol. Med.* **2017**, *83*, 48–58. [CrossRef]
72. Ramalingam, V.; Dandapath, A.; Raja, M.K. Heart disease prediction using machine learning techniques: A survey. *Int. J. Eng. Technol.* **2018**, *7*, 684–687. [CrossRef]
73. Cao, Y.; Wang, Y.; Peng, J.; Zhang, L.; Xu, L.; Yan, K.; Li, L. DML-GANR: Deep metric learning with generative adversarial network regularization for high spatial resolution remote sensing image retrieval. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8888–8904. [CrossRef]

74. Ghorbani, R.; Ghousi, R.; Makui, A.; Atashi, A. A new hybrid predictive model to predict the early mortality risk in intensive care units on a highly imbalanced dataset. *IEEE Access* **2020**, *8*, 141066–141079. [CrossRef]
75. Ding, J.; Errapotu, S.M.; Guo, Y.; Zhang, H.; Yuan, D.; Pan, M. Private empirical risk minimization with analytic gaussian mechanism for healthcare system. *IEEE Trans. Big Data* **2020**, *8*, 1107–1117. [CrossRef]
76. Jabeen, F.; Maqsood, M.; Ghazanfar, M.A.; Aadil, F.; Khan, S.; Khan, M.F.; Mehmood, I. An IoT based efficient hybrid recommender system for cardiovascular disease. *Peer-to-Peer Netw. Appl.* **2019**, *12*, 1263–1276. [CrossRef]
77. Muzammal, M.; Talat, R.; Sodhro, A.H.; Pirbhulal, S. A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. *Inf. Fusion* **2020**, *53*, 155–164. [CrossRef]
78. Paul, A.K.; Shill, P.C.; Rabin, M.R.I.; Murase, K. Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. *Appl. Intell.* **2018**, *48*, 1739–1756. [CrossRef]
79. Ahmad, W.; Ahmad, A.; Iqbal, A.; Hamayun, M.; Hussain, A.; Rehman, G.; Khan, S.; Khan, U.U.; Khan, D.; Huang, L. Intelligent hepatitis diagnosis using adaptive neuro-fuzzy inference system and information gain method. *Soft Comput.* **2019**, *23*, 10931–10938. [CrossRef]
80. Mohan, S.; Thirumalai, C.; Srivastava, G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access* **2019**, *7*, 81542–81554. [CrossRef]
81. Jothi Prakash, V.; Karthikeyan, N. Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*, 389–412. [CrossRef] [PubMed]
82. Samir, A.A.; Rashwan, A.R.; Sallam, K.M.; Chakrabortty, R.K.; Ryan, M.J.; Abohany, A.A. Evolutionary algorithm-based convolutional neural network for predicting heart diseases. *Comput. Ind. Eng.* **2021**, *161*, 107651. [CrossRef]
83. Srinivas, P.; Katarya, R. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomed. Signal Process. Control* **2022**, *73*, 103456. [CrossRef]
84. Ay, Ş.; Ekinci, E.; Garip, Z. A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases. *J. Supercomput.* **2023**, *79*, 11797–11826. [CrossRef] [PubMed]
85. Cocianu, C.L.; Uscatu, C.R.; Kofidis, K.; Muraru, S.; Văduva, A.G. Classical, Evolutionary, and Deep Learning Approaches of Automated Heart Disease Prediction: A Case Study. *Electronics* **2023**, *12*, 1663. [CrossRef]
86. Villamil, H.C.; Espitia, H.E.; Bejarano, L.A. Multiobjective Optimization of Fuzzy System for Cardiovascular Risk Classification. *Computation* **2023**, *11*, 147. [CrossRef]
87. Karthikeyan, G.; Komarasamy, G.; Daniel Madan Raja, S. Design of an efficient decision support system using evolutionary deep forward network model. *J. Intell. Fuzzy Syst.* **2023**, *44*, 7027–7042. [CrossRef]
88. Al Bataineh, A.; Manacek, S. MLP-PSO hybrid algorithm for heart disease prediction. *J. Pers. Med.* **2022**, *12*, 1208. [CrossRef]
89. Ahmad, Z.; Li, J.; Mahmood, T. Adaptive Hyperparameter Fine-Tuning for Boosting the Robustness and Quality of the Particle Swarm Optimization Algorithm for Non-Linear RBF Neural Network Modelling and Its Applications. *Mathematics* **2023**, *11*, 242. [CrossRef]
90. Bashir, S.; Khan, Z.S.; Khan, F.H.; Anjum, A.; Bashir, K. Improving heart disease prediction using feature selection approaches. In Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019; pp. 619–623.
91. Fassnacht, F.E.; Neumann, C.; Förster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2547–2561. [CrossRef]
92. Xue, B.; Zhang, M.; Browne, W.N. A comprehensive comparison on evolutionary feature selection approaches to classification. *Int. J. Comput. Intell. Appl.* **2015**, *14*, 1550008. [CrossRef]
93. Tan, F.; Fu, X.; Zhang, Y.; Bourgeois, A.G. A genetic algorithm-based method for feature subset selection. *Soft Comput.* **2008**, *12*, 111–120. [CrossRef]
94. Nyathi, T.; Pillay, N. Comparison of a genetic algorithm to grammatical evolution for automated design of genetic programming classification algorithms. *Expert Syst. Appl.* **2018**, *104*, 213–234. [CrossRef]
95. Gokulnath, C.B.; Shantharajah, S. An optimized feature selection based on genetic approach and support vector machine for heart disease. *Clust. Comput.* **2019**, *22*, 14777–14787. [CrossRef]
96. Kalamkar, S.; A, G.M. Clinical data fusion and machine learning techniques for smart healthcare. In Proceedings of the 2020 International Conference on Industry 4.0 Technology (I4Tech), Pune, India, 13–15 February 2020; pp. 211–216.
97. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **2021**, *13*, 218. [CrossRef]
98. Nusinovici, S.; Tham, Y.C.; Yan, M.Y.C.; Ting, D.S.W.; Li, J.; Sabanayagam, C.; Wong, T.Y.; Cheng, C.Y. Logistic regression was as good as machine learning for predicting major chronic diseases. *J. Clin. Epidemiol.* **2020**, *122*, 56–69. [CrossRef]
99. Wu, X.; Kumar, V.; Ross Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
100. Haq, A.U.; Li, J.; Khan, J.; Memon, M.H.; Parveen, S.; Raji, M.F.; Akbar, W.; Ahmad, T.; Ullah, S.; Shoista, L.; et al. Identifying the predictive capability of machine learning classifiers for designing heart disease detection system. In Proceedings of the 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, Chengdu, China, 14–15 December 2019; pp. 130–138.

101. Enriko, I.K.A.; Suryanegara, M.; Gunawan, D. Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters. *J. Telecommun. Electron. Comput. Eng. (JTEC)* **2016**, *8*, 59–65.
102. Palimkar, P.; Shaw, R.N.; Ghosh, A. Machine learning technique to prognosis diabetes disease: Random forest classifier approach. In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021*; Springer: Singapore, 2022; pp. 219–244.
103. Haq, A.U.; Li, J.; Memon, M.H.; Memon, M.H.; Khan, J.; Marium, S.M. Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 29–31 March 2019; pp. 1–4.
104. Cunningham, P.; Delany, S.J. k-Nearest neighbour classifiers—A Tutorial. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–25. [CrossRef]
105. Schapire, R.E. Explaining adaboost. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.
106. Schölkopf, B.; Luo, Z.; Vovk, V. *Empirical Inference*; Springer: Berlin/Heidelberg, Germany, 2016.
107. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [CrossRef]
108. Berrar, D. Bayes' theorem and naive Bayes classifier. *Encycl. Bioinform. Comput. Biol. ABC Bioinform.* **2018**, *403*, 412.
109. Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.J.; Sandhu, S.; Guppy, K.H.; Lee, S.; Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am. J. Cardiol.* **1989**, *64*, 304–310. [CrossRef]
110. Gudadhe, M.; Wankhade, K.; Dongre, S. Decision support system for heart disease based on support vector machine and artificial neural network. In Proceedings of the 2010 International Conference on Computer and Communication Technology (ICCCT), Allahabad, India, 17–19 September 2010; pp. 741–745.
111. Kadhim, M.A.; Radhi, A.M. Heart disease classification using optimized Machine learning algorithms. *Iraqi J. Comput. Sci. Math.* **2023**, *4*, 31–42. [CrossRef]
112. Kahramanli, H.; Allahverdi, N. Design of a hybrid system for the diabetes and heart diseases. *Expert Syst. Appl.* **2008**, *35*, 82–89. [CrossRef]
113. Das, R.; Turkoglu, I.; Sengur, A. Effective diagnosis of heart disease through neural networks ensembles. *Expert Syst. Appl.* **2009**, *36*, 7675–7680. [CrossRef]
114. Jabbar, M.A.; Deekshatulu, B.; Chandra, P. Classification of heart disease using artificial neural network and feature subset selection. *Glob. J. Comput. Sci. Technol. Neural Artif. Intell.* **2013**, *13*, 4–8.
115. Sa, S. Intelligent heart disease prediction system using data mining techniques. *Int. J. Healthc. Biomed. Res.* **2013**, *1*, 94–101.
116. Olaniyi, E.O.; Oyedotun, O.K.; Adnan, K. Heart diseases diagnosis using neural networks arbitration. *Int. J. Intell. Syst. Appl.* **2015**, *7*, 72. [CrossRef]
117. Samuel, O.W.; Asogbon, G.M.; Sangaiah, A.K.; Fang, P.; Li, G. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst. Appl.* **2017**, *68*, 163–172. [CrossRef]
118. Liu, X.; Wang, X.; Su, Q.; Zhang, M.; Zhu, Y.; Wang, Q.; Wang, Q. A hybrid classification system for heart disease diagnosis based on the RFRS method. *Comput. Math. Methods Med.* **2017**, *2017*, 8272091. [CrossRef]
119. Geweid, G.G.; Abdallah, M.A. A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique. *IEEE Access* **2019**, *7*, 149595–149611. [CrossRef]
120. Mondéjar-Guerra, V.; Novo, J.; Rouco, J.; Penedo, M.G.; Ortega, M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomed. Signal Process. Control* **2019**, *47*, 41–48. [CrossRef]
121. Yıldırım, Ö.; Pławiak, P.; Tan, R.S.; Acharya, U.R. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* **2018**, *102*, 411–420. [CrossRef] [PubMed]
122. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 664–675. [CrossRef] [PubMed]
123. Dixit, S.; Kala, R. Early detection of heart diseases using a low-cost compact ECG sensor. *Multimed. Tools Appl.* **2021**, *80*, 32615–32637. [CrossRef]
124. Bemando, C.; Miranda, E.; Aryuni, M. Machine-learning-based prediction models of coronary heart disease using naïve bayes and random forest algorithms. In Proceedings of the 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), Pekan, Malaysia, 24–26 August 2021; pp. 232–237.
125. Jan, M.; Awan, A.A.; Khalid, M.S.; Nisar, S. Ensemble approach for developing a smart heart disease prediction system using classification algorithms. *Res. Rep. Clin. Cardiol.* **2018**, *9*, 33–45. [CrossRef]
126. Ram Kumar, R.; Polepaka, S. Performance comparison of random forest classifier and convolution neural network in predicting heart diseases. In *Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018*; Springer: Singapore, 2020; pp. 683–691.
127. Singh, H.; Navaneeth, N.; Pillai, G.N. Multisurface proximal SVM based decision trees for heart disease classification. In Proceedings of the TENCON 2019–2019 IEEE Region 10 Conference (TENCON), Kochi, India, 17–20 October 2019; pp. 13–18.
128. Desai, S.D.; Giraddi, S.; Narayankar, P.; Pudakalakatti, N.R.; Sulegaon, S. Back-propagation neural network versus logistic regression in heart disease classification. In *Proceedings of the Advanced Computing and Communication Technologies: Proceedings of the 11th ICACCT 2018*; Springer: Singapore, 2019; pp. 133–144.

129. Patil, D.D.; Singh, R.; Thakare, V.M.; Gulve, A.K. Analysis of ECG arrhythmia for heart disease detection using SVM and cuckoo search optimized neural network. *Int. J. Eng. Technol.* **2018**, *7*, 27–33. [CrossRef]
130. Liu, N.; Lin, Z.; Cao, J.; Koh, Z.; Zhang, T.; Huang, G.B.; Ser, W.; Ong, M.E.H. An intelligent scoring system and its application to cardiac arrest prediction. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 1324–1331. [CrossRef] [PubMed]
131. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M.; Gertych, A.; San Tan, R. A deep convolutional neural network model to classify heartbeats. *Comput. Biol. Med.* **2017**, *89*, 389–396. [CrossRef] [PubMed]
132. Yang, W.; Si, Y.; Wang, D.; Guo, B. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Comput. Biol. Med.* **2018**, *101*, 22–32. [CrossRef]
133. Saxena, R.; Johri, A.; Deep, V.; Sharma, P. Heart diseases prediction system using CHC-TSS Evolutionary, KNN, and decision tree classification algorithm. In *Proceedings of the Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018*; Springer: Singapore, 2019; Volume 2, pp. 809–819.
134. Basheer, S.; Mathew, R.M.; Devi, M.S. Ensembling coalesce of logistic regression classifier for heart disease prediction using machine learning. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 127–133. [CrossRef]
135. Soni, J.; Ansari, U.; Sharma, D.; Soni, S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. *Int. J. Comput. Appl.* **2011**, *17*, 43–48. [CrossRef]
136. Hinchliffe, R.; Brownrigg, J.; Apelqvist, J.; Boyko, E.; Fitridge, R.; Mills, J.; Reekers, J.; Shearman, C.; Zierler, R.; Schaper, N.; et al. IWGDF guidance on the diagnosis, prognosis and management of peripheral artery disease in patients with foot ulcers in diabetes. *Diabetes/Metab. Res. Rev.* **2016**, *32*, 37–44. [CrossRef]
137. Al-Milli, N. Backpropogation neural network for prediction of heart disease. *J. Theor. Appl. Inf. Technol.* **2013**, *56*, 131–135.
138. Singh, A.; Kumar, R. Heart disease prediction using machine learning algorithms. In Proceedings of the 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 14–15 February 2020; pp. 452–457.
139. Hashi, E.K.; Zaman, M.S.U. Developing a hyperparameter tuning based machine learning approach of heart disease prediction. *J. Appl. Sci. Process Eng.* **2020**, *7*, 631–647. [CrossRef]
140. Tama, B.A.; Im, S.; Lee, S. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *BioMed Res. Int.* **2020**, *2020*, 9816142. [CrossRef]
141. Shah, D.; Patel, S.; Bharti, S.K. Heart disease prediction using machine learning techniques. *SN Comput. Sci.* **2020**, *1*, 1–6. [CrossRef]
142. Ali, M.M.; Paul, B.K.; Ahmed, K.; Bui, F.M.; Quinn, J.M.; Moni, M.A. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Comput. Biol. Med.* **2021**, *136*, 104672. [CrossRef] [PubMed]
143. Maji, S.; Arora, S. Decision tree algorithms for prediction of heart disease. In *Proceedings of the Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017*; Springer: Singapore, 2019; pp. 447–454.
144. Nikhar, S.; Karandikar, A. Prediction of heart disease using machine learning algorithms. *Int. J. Adv. Eng. Manag. Sci.* **2016**, *2*, 239484.
145. Patel, S.B.; Yadav, P.K.; Shukla, D. Predict the diagnosis of heart disease patients using classification mining techniques. *IOSR J. Agric. Vet. Sci. (IOSR-JAVS)* **2013**, *4*, 61–64.
146. Nguyen, T.; Khosravi, A.; Creighton, D.; Nahavandi, S. Classification of healthcare data using genetic fuzzy logic system and wavelets. *Expert Syst. Appl.* **2015**, *42*, 2184–2197. [CrossRef]
147. Kapila, R.; Ragunathan, T.; Saleti, S.; Lakshmi, T.J.; Ahmad, M.W. Heart Disease Prediction using Novel Quine McCluskey Binary Classifier (QMBC). *IEEE Access* **2023**, *11*, 64324–64347. [CrossRef]
148. Kavitha, M.; Gnaneswar, G.; Dinesh, R.; Sai, Y.R.; Suraj, R.S. Heart disease prediction using hybrid machine learning model. In Proceedings of the 2021 6th international conference on inventive computation technologies (ICICT), Coimbatore, India, 20–22 January 2021; pp. 1329–1333.
149. Mehmood, A.; Iqbal, M.; Mehmood, Z.; Irtaza, A.; Nawaz, M.; Nazir, T.; Masood, M. Prediction of heart disease using deep convolutional neural networks. *Arab. J. Sci. Eng.* **2021**, *46*, 3409–3422. [CrossRef]
150. Ramprakash, P.; Sarumathi, R.; Mowriya, R.; Nithyavishnupriya, S. Heart disease prediction using deep neural network. In Proceedings of the 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–28 February 2020; pp. 666–670.
151. Hoodbhoy, Z.; Jiwani, U.; Sattar, S.; Salam, R.; Hasan, B.; Das, J.K. Diagnostic accuracy of machine learning models to identify congenital heart disease: A meta-analysis. *Front. Artif. Intell.* **2021**, *4*, 708365. [CrossRef]
152. Jin, B.; Che, C.; Liu, Z.; Zhang, S.; Yin, X.; Wei, X. Predicting the risk of heart failure with EHR sequential data modeling. *IEEE Access* **2018**, *6*, 9256–9261. [CrossRef]
153. McPhee, S.J.; Papadakis, M.A.; Rabow, M.W. (Eds.) *Current Medical Diagnosis & Treatment 2010*; McGraw-Hill Medical: New York, NY, USA, 2010.
154. Beyer, M.A.; Laney, D. *The Importance of "Big Data": A Definition*; G00235055; Gartner: Stamford, CT, USA, 2012.
155. Forkan, A.R.M.; Khalil, I. PEACE-Home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring. *Pervasive Mob. Comput.* **2017**, *38*, 296–311. [CrossRef]
156. Jiang, P.; Winkley, J.; Zhao, C.; Munnoch, R.; Min, G.; Yang, L.T. An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. *IEEE Syst. J.* **2014**, *10*, 1147–1159. [CrossRef]

157. Salem, O.; Liu, Y.; Mehaoua, A.; Boutaba, R. Online anomaly detection in wireless body area networks for reliable healthcare monitoring. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1541–1551. [CrossRef]
158. Teich, J.M.; Osheroff, J.A.; Pifer, E.A.; Sittig, D.F.; Jenders, R.A. Clinical decision support in electronic prescribing: Recommendations and an action plan: Report of the joint clinical decision support workgroup. *J. Am. Med. Inform. Assoc.* **2005**, *12*, 365–376. [CrossRef]
159. UCI. *Heart Disease*; UCI Machine Learning Repository: Irvine, CA, USA, 1988. [CrossRef]
160. Framingham. Framingham Heart Disease Study. 2023. Available Online: https://www.framinghamheartstudy.org/fhs-for-researchers/ (accessed on 22 May 2023).
161. Cardiovascular. Cardiovascular Disease Dataset. 2023. Available Online: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset. (accessed on 22 May 2023).
162. ECG. Physikalisch Technische Bundesanstalt Diagnostic ECG Dataset. Available Online: https://www.physionet.org/content/ptbdb/1.0.0/ (accessed on 22 May 2023).
163. kaggle. Stroke Prediction Dataset. 2023. Available Online: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset (accessed on 22 May 2023).
164. Ras, G.; van Gerven, M.; Haselager, P. Explanation methods in deep learning: Users, values, concerns and challenges. In *Explainable and Interpretable Models in Computer Vision and Machine Learning*; Springer: Cham, Switzerland, 2018; pp. 19–36.
165. General Data Protection Regulation. General data protection regulation (GDPR). In *Intersoft Consulting*; European Union: Brussels, Belgium, 2018; Volume 24.
166. Alabdulatif, A.; Khalil, I.; Forkan, A.R.M.; Atiquzzaman, M. Real-time secure health surveillance for smarter health communities. *IEEE Commun. Mag.* **2018**, *57*, 122–129. [CrossRef]
167. Narwal, B.; Mohapatra, A.K. A survey on security and authentication in wireless body area networks. *J. Syst. Archit.* **2021**, *113*, 101883. [CrossRef]
168. Benhar, H.; Idri, A.; Fernández-Alemán, J. Data preprocessing for heart disease classification: A systematic literature review. *Comput. Methods Programs Biomed.* **2020**, *195*, 105635. [CrossRef]
169. Ambarwari, A.; Adrian, Q.J.; Herdiyeni, Y. Analysis of the effect of data scaling on the performance of the machine learning algorithm for plant identification. *J. RESTI (Rekayasa Sistem Dan Teknologi Informasi)* **2020**, *4*, 117–122. [CrossRef]
170. Ilyas, I.F.; Chu, X. *Data Cleaning*; Morgan & Claypool: San Rafael, CA, USA, 2019.
171. Grampurohit, S.; Sagarnal, C. Disease prediction using machine learning algorithms. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 5–7 June 2020; pp. 1–7.
172. Ting, D.S.W.; Pasquale, L.R.; Peng, L.; Campbell, J.P.; Lee, A.Y.; Raman, R.; Tan, G.S.W.; Schmetterer, L.; Keane, P.A.; Wong, T.Y. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **2019**, *103*, 167–175. [CrossRef]
173. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
174. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 1322–1328.
175. Yahaya, L.; Oye, N.D.; Garba, E.J. A comprehensive review on heart disease prediction using data mining and machine learning techniques. *Am. J. Artif. Intell.* **2020**, *4*, 20–29. [CrossRef]
176. Haq, A.U.; Li, J.P.; Khan, J.; Memon, M.H.; Nazir, S.; Ahmad, S.; Khan, G.A.; Ali, A. Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data. *Sensors* **2020**, *20*, 2649. [CrossRef]
177. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]
178. Idrees, F.; Rajarajan, M.; Conti, M.; Chen, T.M.; Rahulamathavan, Y. PIndroid: A novel Android malware detection system using ensemble learning methods. *Comput. Secur.* **2017**, *68*, 36–46. [CrossRef]
179. Saqlain, M.; Jargalsaikhan, B.; Lee, J.Y. A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Trans. Semicond. Manuf.* **2019**, *32*, 171–182. [CrossRef]
180. Romdhane, T.F.; Pr, M.A. Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. *Comput. Biol. Med.* **2020**, *123*, 103866. [CrossRef]
181. Pouriyeh, S.; Vahid, S.; Sannino, G.; De Pietro, G.; Arabnia, H.; Gutierrez, J. A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 3–6 July 2017; pp. 204–207.
182. Normawati, D.; Ismi, D.P. K-fold cross validation for selection of cardiovascular disease diagnosis features by applying rule-based datamining. *Signal Image Process. Lett.* **2019**, *1*, 62–72. [CrossRef]
183. Pires, I.M.; Marques, G.; Garcia, N.M.; Ponciano, V. Machine learning for the evaluation of the presence of heart disease. *Procedia Comput. Sci.* **2020**, *177*, 432–437. [CrossRef]
184. Ananey-Obiri, D.; Sarku, E. Predicting the presence of heart diseases using comparative data mining and machine learning algorithms. *Int. J. Comput. Appl.* **2020**, *176*, 17–21. [CrossRef]

185. Wong, T.T.; Yeh, P.Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [CrossRef]
186. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 569–575. [CrossRef] [PubMed]
187. Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4793–4813. [CrossRef] [PubMed]
188. Wijesinghe, I.; Gamage, C.; Perera, I.; Chitraranjan, C. A smart telemedicine system with deep learning to manage diabetic retinopathy and foot ulcers. In Proceedings of the 2019 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 3–5 July 2019; pp. 686–691.
189. Shaw, J.; Rudzicz, F.; Jamieson, T.; Goldfarb, A. Artificial intelligence and the implementation challenge. *J. Med. Internet Res.* **2019**, *21*, e13659. [CrossRef] [PubMed]
190. Ghoshal, B.; Tucker, A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. *arXiv* **2020**, arXiv:2003.10769.
191. Huynh-Thu, V.A.; Saeys, Y.; Wehenkel, L.; Geurts, P. Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. *Bioinformatics* **2012**, *28*, 1766–1774. [CrossRef]
192. Kumarakulasinghe, N.B.; Blomberg, T.; Liu, J.; Leao, A.S.; Papapetrou, P. Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; pp. 7–12.
193. Neves, I.; Folgado, D.; Santos, S.; Barandas, M.; Campagner, A.; Ronzio, L.; Cabitza, F.; Gamboa, H. Interpretable heartbeat classification using local model-agnostic explanations on ECGs. *Comput. Biol. Med.* **2021**, *133*, 104393. [CrossRef]
194. Assegie, T.A. Evaluation of Local Interpretable Model-Agnostic Explanation and Shapley Additive Explanation for Chronic Heart Disease Detection. *Proc. Eng. Technol. Innov.* **2023**, *23*, 48–59.
195. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1379. [CrossRef]
196. Sixian, L.; Imamura, Y.; Ahmed, A. Application of Shapley Additive Explanation towards Determining Personalized Triage from Health Checkup Data. In *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare*; Springer: Cham, Switzerland, 2022; pp. 496–509.
197. Kim, Y.; Kim, Y. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustain. Cities Soc.* **2022**, *79*, 103677. [CrossRef]
198. Miranda, E.; Adiarto, S.; Bhatti, F.M.; Zakiyyah, A.Y.; Aryuni, M.; Bernando, C. Understanding Arteriosclerotic Heart Disease Patients Using Electronic Health Records: A Machine Learning and Shapley Additive exPlanations Approach. *Healthc. Inform. Res.* **2023**, *29*, 228–238. [CrossRef]
199. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
200. Arumugam, K.; Naved, M.; Shinde, P.P.; Leiva-Chauca, O.; Huaman-Osorio, A.; Gonzales-Yanac, T. Multiple disease prediction using Machine learning algorithms. *Mater. Today Proc.* **2023**, *80*, 3682–3685. [CrossRef]
201. Albahri, O.S.; Albahri, A.S.; Zaidan, A.; Zaidan, B.; Alsalem, M.; Mohsin, A.H.; Mohammed, K.; Alamoodi, A.H.; Nidhal, S.; Enaizan, O.; et al. Fault-tolerant mHealth framework in the context of IoT-based real-time wearable health data sensors. *IEEE Access* **2019**, *7*, 50052–50080. [CrossRef]
202. Marcinkevičs, R.; Vogt, J.E. Interpretability and explainability: A machine learning zoo mini-tour. *arXiv* **2020**, arXiv:2012.01805.
203. Buettner, R.; Kuri, T.; Feist, A.; Hudak, J. Overview of Machine Learning Approaches Applied in Disease Profiling. In Proceedings of the 2020 IEEE Symposium on Industrial Electronics & Applications (ISIEA), Kuala Lumpur, Malaysia, 17–18 July 2020; pp. 1–5.
204. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18. [CrossRef]
205. Nazir, S.; Ali, Y.; Ullah, N.; García-Magariño, I. Internet of things for healthcare using effects of mobile computing: A systematic literature review. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 1–20. [CrossRef]
206. Mani, N.; Singh, A.; Nimmagadda, S.L. An IoT guided healthcare monitoring system for managing real-time notifications by fog computing services. *Procedia Comput. Sci.* **2020**, *167*, 850–859. [CrossRef]
207. Selvaraj, S.; Sundaravaradhan, S. Challenges and opportunities in IoT healthcare systems: A systematic review. *SN Appl. Sci.* **2020**, *2*, 139. [CrossRef]
208. Seetharam, K.; Kagiyama, N.; Sengupta, P.P. Application of mobile health, telemedicine and artificial intelligence to echocardiography. *Echo Res. Pract.* **2019**, *6*, R41–R52. [CrossRef] [PubMed]
209. Jagpal, A.; Navarro-Millán, I. Cardiovascular co-morbidity in patients with rheumatoid arthritis: A narrative review of risk factors, cardiovascular risk assessment and treatment. *BMC Rheumatol.* **2018**, *2*, 1–14. [CrossRef] [PubMed]
210. Casalino, G.; Castellano, G.; Kaymak, U.; Zaza, G. Balancing accuracy and interpretability through neuro-fuzzy models for cardiovascular risk assessment. In Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 5–7 December 2021; pp. 1–8.

211. Dhanvijay, M.M.; Patil, S.C. Internet of Things: A survey of enabling technologies in healthcare and its applications. *Comput. Netw.* **2019**, *153*, 113–131. [CrossRef]
212. Mahdavinejad, M.S.; Rezvan, M.; Barekatain, M.; Adibi, P.; Barnaghi, P.; Sheth, A.P. Machine learning for Internet of Things data analysis: A survey. *Digit. Commun. Netw.* **2018**, *4*, 161–175. [CrossRef]
213. Ding, Y.; Wu, G.; Chen, D.; Zhang, N.; Gong, L.; Cao, M.; Qin, Z. DeepEDN: A deep-learning-based image encryption and decryption network for internet of medical things. *IEEE Internet Things J.* **2020**, *8*, 1504–1518. [CrossRef]
214. Puspitasari, R.D.I.; Ma'sum, M.A.; Alhamidi, M.R.; Jatmiko, W. Generative adversarial networks for unbalanced fetal heart rate signal classification. *ICT Express* **2022**, *8*, 239–243. [CrossRef]
215. Qadri, Y.A.; Nauman, A.; Zikria, Y.B.; Vasilakos, A.V.; Kim, S.W. The future of healthcare internet of things: A survey of emerging technologies. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1121–1167. [CrossRef]
216. Rieke, N.; Hancox, J.; Li, W.; Milletari, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The future of digital health with federated learning. *NPJ Digit. Med.* **2020**, *3*, 119. [CrossRef]
217. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. Advances and open problems in federated learning. *Found. Trends® Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]
218. Taha, Z.K.; Yaw, C.T.; Koh, S.P.; Tiong, S.K.; Kadirgama, K.; Benedict, F.; Tan, J.D.; Balasubramaniam, Y.A. A Survey of Federated Learning from Data Perspective in the Healthcare Domain: Challenges, Methods, and Future Directions. *IEEE Access* **2023**, *11*, 45711–45735. [CrossRef]
219. Linardos, A.; Kushibar, K.; Walsh, S.; Gkontra, P.; Lekadir, K. Federated learning for multi-center imaging diagnostics: A simulation study in cardiovascular disease. *Sci. Rep.* **2022**, *12*, 3551. [CrossRef] [PubMed]
220. Sahoo, P.K.; Mohapatra, S.K.; Wu, S.L. SLA based healthcare big data analysis and computing in cloud network. *J. Parallel Distrib. Comput.* **2018**, *119*, 121–135. [CrossRef]
221. Thanigaivasan, V.; Narayanan, S.J.; Iyengar, S.N.; Ch, N. Analysis of parallel SVM based classification technique on healthcare using big data management in cloud storage. *Recent Pat. Comput. Sci.* **2018**, *11*, 169–178. [CrossRef]
222. Wang, Y.; Kung, L.; Wang, W.Y.C.; Cegielski, C.G. An integrated big data analytics-enabled transformation model: Application to health care. *Inf. Manag.* **2018**, *55*, 64–79. [CrossRef]
223. Alsabah, M.; Naser, M.A.; Mahmmod, B.M.; Abdulhussain, S.H.; Eissa, M.R.; Al-Baidhani, A.; Noordin, N.K.; Sait, S.M.; Al-Utaibi, K.A.; Hashim, F. 6G wireless communications networks: A comprehensive survey. *IEEE Access* **2021**, *9*, 148191–148243. [CrossRef]
224. Wenzel, M.; Wiegand, T. Toward global validation standards for health AI. *IEEE Commun. Stand. Mag.* **2020**, *4*, 64–69. [CrossRef]
225. Li, X.; Zhang, L.; Wu, Z.; Liu, Z.; Zhao, L.; Yuan, Y.; Liu, J.; Li, G.; Zhu, D.; Yan, P.; et al. Artificial General Intelligence for Medical Imaging. *arXiv* **2023**, arXiv:2306.05480.
226. Obaid, O.I. From Machine Learning to Artificial General Intelligence: A Roadmap and Implications. *Mesopotamian J. Big Data* **2023**, *2023*, 81–91. [CrossRef]
227. Kuusi, O.; Heinonen, S. Scenarios from artificial narrow intelligence to artificial general intelligence—Reviewing the results of the international work/technology 2050 study. *World Futures Rev.* **2022**, *14*, 65–79. [CrossRef]
228. Ahamad, G.N.; Shafiullah; Fatima, H.; Imdadullah.; Zakariya, S.; Abbas, M.; Alqahtani, M.S.; Usman, M. Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease. *Processes* **2023**, *11*, 734. [CrossRef]
229. Ozcan, M.; Peker, S. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthc. Anal.* **2023**, *3*, 100130. [CrossRef]
230. Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective heart disease prediction using machine learning techniques. *Algorithms* **2023**, *16*, 88. [CrossRef]
231. Al Reshan, M.S.; Amin, S.; Zeb, M.A.; Sulaiman, A.; Alshahrani, H.; Shaikh, A. A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks. *IEEE Access* **2023**, *11*, 121574–121591. [CrossRef]
232. Rustam, F.; Ishaq, A.; Munir, K.; Almutairi, M.; Aslam, N.; Ashraf, I. Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction. *Diagnostics* **2022**, *12*, 1474. [CrossRef] [PubMed]
233. Doppala, B.P.; Bhattacharyya, D.; Janarthanan, M.; Baik, N. A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. *J. Healthc. Eng.* **2022**, *2022*, 2585235. [CrossRef] [PubMed]
234. Ramesh, T.; Lilhore, U.K.; Poongodi, M.; Simaiya, S.; Kaur, A.; Hamdi, M. Predictive analysis of heart diseases with machine learning approaches. *Malays. J. Comput. Sci.* **2022**, 132–148. [CrossRef]
235. Boukhatem, C.; Youssef, H.Y.; Nassif, A.B. Heart disease prediction using machine learning. In Proceedings of the 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 21–24 February 2022; pp. 1–6.
236. Nagavelli, U.; Samanta, D.; Chakraborty, P. Machine learning technology-based heart disease detection models. *J. Healthc. Eng.* **2022**, *2022*, 7351061. [CrossRef] [PubMed]
237. Tiwari, A.; Chugh, A.; Sharma, A. Ensemble framework for cardiovascular disease prediction. *Comput. Biol. Med.* **2022**, *146*, 105624. [CrossRef] [PubMed]
238. Ketu, S.; Mishra, P.K. Empirical analysis of machine learning algorithms on imbalance electrocardiogram based arrhythmia dataset for heart disease detection. *Arab. J. Sci. Eng.* **2022**, 47, 1447–1469. [CrossRef]

239. Rahim, A.; Rasheed, Y.; Azam, F.; Anwar, M.W.; Rahim, M.A.; Muzaffar, A.W. An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access* **2021**, *9*, 106575–106588. [CrossRef]
240. Ghosh, P.; Azam, S.; Jonkman, M.; Karim, A.; Shamrat, F.J.M.; Ignatious, E.; Shultana, S.; Beeravolu, A.R.; De Boer, F. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access* **2021**, *9*, 19304–19326. [CrossRef]