Archives of Cardiovascular Disease xxx (xxxx) xxx–xxx

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Review

Machine learning approaches for cardiovascular disease prediction: A review

Siming Wan[a,b,c], Feng Wan[c], Xi-jian Dai[a,b,*]

[a] Department of Radiology, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, 330006 Nanchang, China
[b] Jiangxi Provincial Key Laboratory of Intelligent Medical Imaging, 330006 Nanchang, China
[c] Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau, Taipa, Macau, China

ARTICLE INFO

Article history:
Received 21 January 2025
Received in revised form 26 March 2025
Accepted 4 April 2025
Available online xxx

Keywords:
Machine learning
Cardiovascular disease
Classification
Prediction
Feature selection

ABSTRACT

Cardiovascular disease is a leading cause of death worldwide and is associated with significant morbidity and mortality. The use of artificial intelligence techniques, particularly machine learning algorithms, has emerged as a transformative approach for enhancing early diagnostic accuracy of disease compared with conventional diagnostic methods. This systematic review examines three core aspects: (1) comparative analysis of current machine learning algorithms in early diagnosis of cardiovascular disease, (2) operational frameworks for clinical implementation, and (3) critical evaluation of regulatory compliance and ethical implications. It summarizes recent advancements in machine learning-based heart disease prediction, outlines a typical workflow for applying machine learning in clinical settings, and discusses the regulatory and ethical challenges associated with its implementation. Finally, this review explores potential directions for future research in this rapidly evolving field.

1. Background

Cardiovascular disease (CVD) is a leading cause of death worldwide and is associated with significant morbidity and mortality; it not only presents challenges to healthcare systems, but also inflicts economic and social burdens [1]. Therefore, in order to minimize deaths, early detection of CVD is crucial.

However, conventional methods for predicting CVD often suffer from limitations [2], such as high costs, low accessibility and insufficient accuracy. These shortcomings highlight the need for more efficient and cost-effective diagnostic approaches.

In recent years, artificial intelligence (AI), particularly machine learning (ML) algorithms, has shown great potential in enhancing healthcare systems by improving disease identification, diagnosis, classification, prediction, prevention and treatment [3,4]. By leveraging AI-driven models, healthcare providers can achieve more accurate and timely predictions, ultimately reducing CVD-related deaths [5]. Fig. 1 illustrates the workflow of an ML model used in CVD prediction. A glossary of key terms is provided in Table A.1 better understanding of this article.

1.1. Study structure

This review follows a structured framework. In Section 2, we discuss the advances in ML in predicting CVD, and compare it with conventional analytical methods in this field. Section 3 provides an introduction to the innovations in heart disease prediction using ML methods. Data processing methods, the evaluation indicators used most commonly and the cross-validation method are presented in Section 4. Section 5 explores recent trends, discusses regulatory and ethical challenges and provides recommendations for future work in this area. Finally, a summary is provided in Section 6.

1.2. Methodology for paper selection

This review aims to analyse the application of ML models in the prediction of CVD; the aim is to provide a new perspective on how papers in this field explore this topic. The selection and retrieval of papers was conducted using reputable sources, such as the Web of Science and Google Scholar. Only documents written in English are

Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; CVD, cardiovascular disease; EHR, electronic health record; ICD-11, International Classification of Diseases, 11th revision; IoT, internet of things; ML, machine learning; ROC, receiver operating characteristic; SMOTE, Synthetic Minority Oversampling Technique.

* Corresponding author. Department of Radiology, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Minde Road No. 1, 330006 Nanchang (Jiangxi Province), China.
E-mail address: daixjdoctor@126.com (X.-j. Dai).

https://doi.org/10.1016/j.acvd.2025.04.055
1875-2136/

Please cite this article as: S. Wan, F. Wan and X.-j. Dai, Machine learning approaches for cardiovascular disease prediction: A review, Arch Cardiovasc Dis, https://doi.org/10.1016/j.acvd.2025.04.055

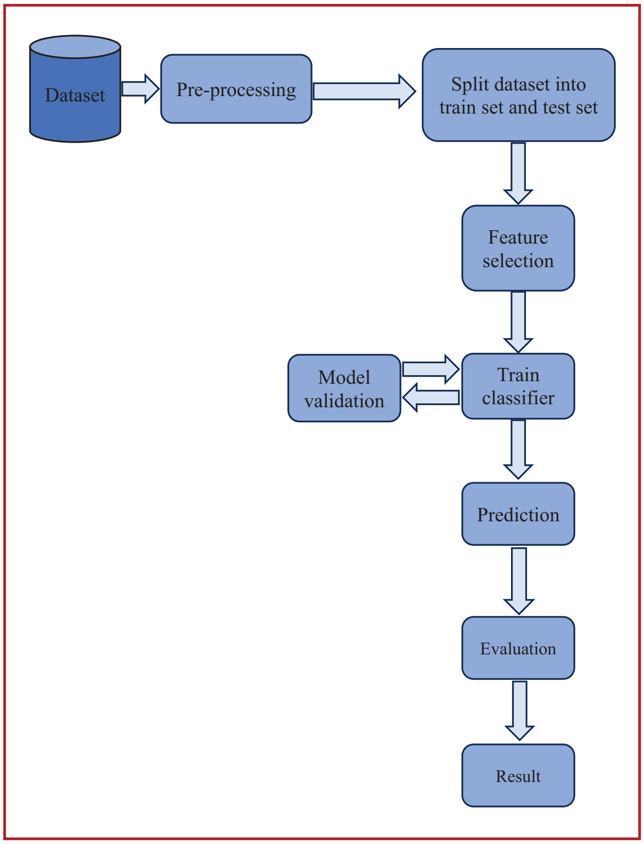


Fig. 1. Overview of machine learning workflow.

included in this study. The search terms were "(Machine learning) AND (heart OR cardio* disease)".

2. Comparison of ML and conventional methods in CVD prediction

This section aims to provide a comparison between ML methods and conventional analytical techniques, to compare the advantages and limitations of the two methods and to review research progress made with ML models in predicting CVD.

2.1. Background

In the International Classification of Diseases, 11th revision (ICD-11) [6], CVDs are primarily categorized under circulatory system diseases. Table 1 provides a summary of the major CVDs based on the ICD-11 classification criteria. The primary objective of CVD prediction is to determine whether a given data point indicates the presence of a cardiovascular condition.

2.2. Conventional methods

Healthcare professionals traditionally assess CVD risk using clinical risk factors, echocardiograms to evaluate heart function [7] and electrocardiograms to detect signs of congestive heart failure [8]. Cardiac catheterization, although invasive, is another diagnostic and treatment method. The term "major acute cerebrovascular and cardiovascular events" (MACCE) is commonly used to describe various CVD-related outcomes [9].

Several conventional risk prediction tools are widely used for assessing CVD risk. The Cox proportional hazards model [10] analyses patient survival times relative to predictor variables, and is valued for adjusting multiple risk factors simultaneously, but it assumes proportional hazards throughout follow-up, which may not always hold. The Framingham Risk Score [11] estimates 10-year coronary heart disease risk based on clinical predictors, but it excludes other events, such as stroke, and is less accurate outside USA-based populations. The Thrombolysis in Myocardial Infarction (TIMI) score [12] predicts ischaemic events in unstable angina or non-ST-segment elevation myocardial infarction, but simplifies risk by assigning equal weight to components, potentially overlooking certain patient characteristics.

The Systematic Coronary Risk Evaluation (SCORE) system [13] predicts 10-year fatal atherosclerotic event risk using limited variables, such as age, cholesterol and blood pressure; although practical in primary care, it focuses only on fatal events, and was developed for European populations, limiting its generalizability. The Global Registry of Acute Coronary Events (GRACE) score [14] estimates 6-month survival after acute coronary syndrome using

Table 1
Cardiovascular disease types in the International Classification of Diseases, 11th revision.

ICD-11 code	Cardiovascular disease type	Description
BA50, BA51, BA52, BA5Y, BA5Z	Chronic ischaemic heart disease	Chronic heart disease is seen as a result of atherosclerosis of coronary arteries; it is characterised by angina pectoris and unstable angina
BA81, BA82, BA83, BA84, BA85, BA86, BA8Y, BA8Z	Diseases of the coronary artery	Conditions affecting the blood perfusion of the heart
8B00, 8B01, 8B02, 8B03, 8B0Z, 8B10, 8B11, 8B1Y, 8B1Z, 8B20, 8B21, 8B22, 8B23, 8B24, 8B25, 8B26, 8B2Z	Cerebrovascular diseases	A group of brain dysfunctions related to disease of the blood vessels supplying the brain, including "stroke", which comprises the following entities: intracerebral haemorrhage; subarachnoid haemorrhage; cerebral ischaemic stroke; and stroke not known if ischaemic or haemorrhagic
BD40, BD41, BD42, BD4Y, BD4Z	Peripheral arterial disease	A disease of blood vessels supplying the arms and legs
BC20	Rheumatic heart disease	Damage to the heart muscle and heart valves from rheumatic fever, caused by streptococcal bacteria
LA80, LA81, LA82, LA83, LA84, LA85, LA86, LA87, LA88, LA89, LA8A, LA8B, LA8 C, LA8D, LA8E, LA8F, LA8G, LA8Y, LA8Z	Structural developmental anomaly of heart or great vessels	A congenital malformation of the heart and/or great vessels or an acquired abnormality unique to the congenitally malformed heart
BD71	Deep vein thrombosis	The process whereby a thrombus (blood clot) forms in the large veins of the peripheral venous system; in addition to obstructing venous return it poses a hazard whereby the thrombus may detach and embolize to the pulmonary circulation
BB00	Pulmonary embolism	Blood clots in the leg veins, which can dislodge and move to the heart and lungs

ICD-11: International Classification of Diseases, 11th revision.

multinational data, but may suffer from selection bias as a result of voluntary participation in the registry.

The QRISK tool [15] predicts 10-year and lifetime CVD risk using a wide range of factors, including body mass index and ethnicity, but overlooks outcomes such as heart failure and the impact of preventive treatments. Lastly, the HEART score [16] helps with risk stratification for chest pain in emergency settings, facilitating early discharge for low-risk patients, although clinical judgment and local validation are essential for its effective use.

2.3. ML methods

The increasing demand for advanced predictive models has driven the adoption of ML in CVD prediction. ML enhances early disease detection, accelerates drug development, enables remote monitoring and facilitates data-driven decision-making [17]. By analysing complex datasets, ML models improve diagnostic accuracy, allowing timely interventions that reduce healthcare costs.

Most studies focus on supervised learning for CVD prediction [18], where models learn from labelled data [19] to predict outcomes. Supervised ML techniques fall into two categories: classification (e.g. disease presence); and regression (e.g. disease risk score). These models use patient features, such as age, cholesterol concentration and blood pressure, to identify disease risks [20].

Various ML techniques are commonly used to evaluate the risk of CVD. A decision tree is a hierarchical model that identifies key risk factors; it is interpretable and computationally efficient, but is prone to overfitting and instability with minor data variations [21]. A random forest is an ensemble of decision trees that improves accuracy and reduces overfitting; it handles high-dimensional data well, but requires substantial computational resources [22]. Logistic regression is a statistical model estimating CVD probability based on input features; it is efficient and interpretable, but assumes linear relationships, limiting its handling of complex patterns [23]. K-nearest neighbours is a distance-based model that classifies new cases based on similar past cases; it is simple and effective for small datasets, but computationally expensive for large-scale applications [24]. Naïve Bayes is a probabilistic model based on Bayes' theorem – efficient for high-dimensional data, but limited by its assumption of feature independence, which may not hold in real-world datasets [25]. AdaBoost is an ensemble method that combines weak classifiers to improve accuracy; it adapts well to difficult cases, but is sensitive to noise and is computationally expensive for large datasets [26]. A support vector machine (SVM) is a powerful model that finds optimal decision boundaries for classification; it performs well in high-dimensional spaces, but requires careful parameter tuning, and is computationally intensive [27]. Deep learning comprises neural networks with multiple layers that automatically extract complex data patterns, excelling in large datasets; however, it requires substantial labelled data and high computational power, and lacks interpretability [28].

There are also some emerging techniques. Gradient boosting is an ML learning technique that builds an ensemble of weak learners – typically decision trees – by optimizing pseudo-residuals. When using decision trees, it forms gradient-boosted trees, which often outperform random forests. Unlike traditional boosting, it allows optimization of any differentiable loss function [29]. Reinforcement learning is an ML approach focused on training an agent to make decisions in a dynamic environment to maximize a reward; it differs from supervised learning by not requiring labelled data or explicit corrections, instead balancing exploration and exploitation to optimize long-term rewards [30]. Q-learning, a basic reinforcement learning method, stores data in tables, but becomes impractical for large or continuous state-action spaces. Reinforcement learning problems are often modelled as Markov decision processes, with algorithms leveraging dynamic programming. Unlike classical methods, reinforcement learning does not require a precise model of the environment, making it suitable for complex problems. Automated ML automates the process of applying ML to real-world problems, covering everything from raw data to model deployment. Designed to simplify ML for non-experts, automated ML accelerates model development, and often outperforms manually designed models. Key techniques include hyperparameter optimization, meta-learning and neural architecture search [31]. Transfer learning is an ML technique where knowledge from one task is reused to enhance performance on a related task. For example, learning to recognize cars can help with identifying trucks. This approach improves learning efficiency and relates to cost-sensitive learning and multiobjective optimization [32]. Graph neural networks are specialized neural networks designed for graph-based data; they use message passing, where nodes iteratively update representations by exchanging information with neighbours. Graph neural networks are part of geometric deep learning, with some architectures resembling convolutional or transformer networks. Applications span natural language processing, social networks, biology, chemistry, physics and optimization problems [33].

2.4. Conventional methods versus ML methods

Conventional methods remain widely used today because of their simplicity and rapid calculations, especially in emergency scenarios and settings with limited resources [34]. However, ML methods offer distinct advantages over conventional approaches.

In cardiology, conventional methods typically rely on expert knowledge for feature selection (i.e. manual selection), whereas ML techniques can incorporate both expert input and automated model evaluation. This enables ML-based methods to be more efficient and accurate. Additionally, most conventional approaches are designed based on an "average patient", which may not always account for the complexity of human pathophysiology. As a result, conventional guidelines may not be suitable for all patients. In contrast, ML approaches facilitate the development of standardized predictive models, assisting cardiologists in creating patient-specific guidelines and enhancing personalized decision-making. These models also help clinicians to save time and improve patient–provider interactions. Another key distinction between statistical and ML methods lies in their respective contributions. Conventional statistical methods focus on understanding relationships between a limited number of variables, whereas ML excels at discovering new variables within datasets and enhancing predictive capabilities [35]. Moreover, ML techniques offer advantages in handling non-parametric and non-linear interactions, including addressing missing data [36]. With computational power, ML methods can analyse data significantly faster than conventional approaches.

A detailed comparison of conventional risk score methods and AI-based risk prediction approaches is presented in Table 2.

3. Innovations in heart disease prediction using ML

This section explores key innovations in ML models for heart disease prediction, which fall into two main categories: (1) the development of novel algorithms; and (2) the innovative combination of existing methods. A detailed bibliometric analysis is provided in Appendix A (Table A.2, Figs. A.1–A.3).

3.1. Novel algorithms

Several studies have introduced new algorithms tailored for heart disease prediction, moving beyond the traditional ML approaches discussed in Section 3:

Table 2
Conventional approach versus machine learning approach in cardiovascular disease prediction.

Aspect	Conventional approach	Machine learning approach
Hypothesis	Involves forming and testing a hypothesis	Does not require a predefined hypothesis
Methodology	Focuses on estimating and interpreting data	Uses data-driven predictions
Feature engineering	Requires domain expertise for manually selecting features	Can automatically extract relevant features from raw data
Learning capability	Lacks learning ability	Continuously learns from data
Outcome	Aims to confirm hypotheses and establish causality	Focuses on making predictions and uncovering hidden patterns
Data requirements	Uses appropriately sized data tailored to a specific hypothesis	Handles large and complex datasets
Performance metrics	Evaluated using coefficients and compatibility levels	Assessed using precision, recall, F-measure, accuracy and AUC
Data type	Works with structured and single-type data	Supports diverse data types, including structured and unstructured data
Risk factor consideration	Considers only demographic and clinical factors (e.g. age, sex, smoking, hypertension)	Integrates multiple data sources (e.g. demographics, electrocardiogram signals, gene expression, image-based features)
Model complexity	Simple models with predefined parameters	Complex, flexible and non-parametric models
Bias and fairness	Less prone to bias if assumptions hold	Can inherit biases from data, requiring fairness-aware algorithms
Limitations	Depends on strong assumptions and is limited to low-dimensional data	Prone to overfitting, underfitting and data security concerns

AUC: area under the receiver operating characteristic curve.

Table 3
The variables table for the Cleveland Heart Disease dataset.

Variable name	Data type	Description	Missing value
Age	Integer	Age of patient in years	No
Sex	Categorical	Male = 1; female = 0	No
CP	Categorical	Chest pain level	No
Trestbps	Integer	Resting blood pressure (on admission to the hospital)	No
Chol	Integer	Serum cholesterol	No
Fbs	Categorical	Fasting blood sugar > 120 mg/dL: true = 1; false = 0	No
Restecg	Categorical	Resting electrocardiography results: normal = 0; ST-T wave abnormality = 1; showing probable or definite left ventricular hypertrophy by Estes criteria = 2	No
Thalach	Integer	Maximum heart rate achieved	No
Exang	Categorical	Exercise-induced angina: yes = 1; no = 0	No
Old peak	Integer	ST-segment depression induced by exercise relative to rest	No
Slope	Categorical	Upward sloping = 1; flat = 2; downward sloping = 3	No
Ca	Categorical	Number of major vessels (0–3) coloured by fluoroscopy	Yes
Thal	Categorical	Severity of chest pain or trouble breathing	Yes
Target	Categorical	Yes = 1; no = 0	No

- a polynomial-time algorithm based on game theory was designed to compute optimal explanations for predictive models [37];
- a genetic algorithm-based feature selection combined with a support vector machine was proposed to enhance heart disease prediction accuracy [38];
- a study integrated feature ranking, particle swarm optimization (PSO) and a neural network feedforward backpropagation model, reducing the number of predictive factors from 13 to eight, and achieving an accuracy of 91.94% [39];
- the classification and regression tree (CART) algorithm was employed to rank influential features, extract decision rules and predict heart disease with an accuracy of 87%, providing a cost-effective and time-efficient diagnostic approach [40].

3.2. Innovative combinations of existing methods

Other studies have leveraged ensemble techniques and hybrid models to improve predictive performance:

- an ensemble-based approach combining ML and deep learning models, including random forest for feature extraction, achieved 88.70% accuracy in CVD prediction [41];
- a study employed six ML algorithms and a soft voting ensemble classifier, using hyperparameter optimization and benchmark dataset evaluation to enhance prediction accuracy beyond existing models [42];
- a multiclassifier ensemble model using soft computing tools demonstrated high predictive accuracy and reliability, alongside a user-friendly graphical interface for smart heart disease prediction [43];
- a study applied supervised ML algorithms, with techniques such as mean value imputation and information gain feature selection, achieving a 95.63% classification accuracy using the random forest model, with high precision, recall, F1 score and area under the receiver operating characteristic curve (ROC) values [44];
- another research effort introduced an ML-based diagnostic system, integrating feature selection techniques and evaluation metrics, enabling accurate and timely heart disease prediction, ultimately aiding doctors in effective patient diagnosis [45].

These innovations highlight the continuous advances in ML for heart disease prediction, enhancing accuracy, interpretability and practical applicability in healthcare settings.

4. Data processing and evaluation metrics

4.1. Dataset collection

Data collection in healthcare is essential for monitoring patient health, delivering treatment and supporting clinical decision-making. High-quality data are crucial for achieving reliable research outcomes. Healthcare data are sourced from electronic health records (EHRs), electrocardiography, internet of things (IoT) devices, time-series data, clinical assessments, medical records, patient-reported data and medical imaging.

However, healthcare datasets often contain missing values and errors, making data acquisition costly and time-consuming. Ensuring compliance with healthcare regulations and quality reporting further emphasizes the importance of meticulous data collection.

Table 4
Detecting cardiovascular and heart disease using machine learning techniques.

Data set	Approach	Results
Cleveland (UCI) [51]	ANN, SVM, KNN, linear regression	Accuracy: 88.7%
Cleveland (UCI) [38]	SVM	Accuracy: 88.34%
Kaggle repository's CVD dataset [41]	RF, KNN, DT, XGBoost, DNN, KDNN	Accuracy: 88.70%; precision: 88.02%; recall: 88.02%; F1 score: 88.0%; AUC: 93%
Combination of: Cleveland, Hungarian, Switzerland, Long Beach VA and Statlog (Heart) [40]	Classification and regression trees	Accuracy: 87.25%
Cleveland (UCI) [42]	RF, KNN, logistic regression, NB, gradient boosting, AdaBoost classifier	Accuracy: 93.44%
Cleveland (UCI) [43]	SVM, ANN, NB, regression analysis and RF	Accuracy: 98.17%
Kaggle repository's CVD dataset [44]	KNN, Naïve Bayes and RF	Accuracy: 95.63%; precision: 93%; recall: 92%; F1 score: 92%; ROC: 90%
Cleveland (UCI) [45]	Logistic regression, KNN, ANN, SVM, NB, DT and RF	Accuracy: 89%
Cleveland (UCI) [52]	Logistic regression, KNN, SVM, NB, DT, RF and XGBoost	Accuracy: 88.52%; precision: 82.85%; recall: 96.66%; F1 score: 89.23%
Cleveland (UCI) [53]	Logistic regression, CNN, SVM, gradient boosting, KNN, XGBoost and RF	Accuracy: 96.50%

ANN: artificial neural network; AUC: area under the curve; CNN: convolutional neural network; CVD: cardiovascular disease; DNN: deep neural network; DT: decision tree; KDNN: knowledge distillation neural network; KNN: K-nearest neighbours; NB: Naïve Bayes; RF: random forest; ROC: receiver operating characteristic; SVM: support vector machine; UCI: University of California, Irvine; XGB: extreme gradient boosting.

4.2. Classic datasets

Several publicly available clinical datasets are widely used for CVD prediction, including: the Cleveland Heart Disease dataset (University of California, Irvine) [46]; the Framingham Cardiovascular Disease Prediction dataset [47]; the Cardiovascular Disease dataset [48]; the Physikalisch Technische Bundesanstalt (PTB) Diagnostic Electrocardiography dataset [49]; and the Stroke Prediction dataset [50].

These datasets serve as benchmarks in CVD research. Table 3 lists the attributes of the Cleveland dataset as an example. In Table 4, the findings of different studies using classic datasets in cardiovascular and heart disease prediction are listed. By using ML techniques mentioned in Section 3, the results are good, indicating that these datasets can be useful in developing new methods for CVD prediction.

4.3. Data preprocessing

4.3.1. Handling missing values

Missing values in healthcare datasets can occur as blanks, not a number (NaN) or placeholders. Removing rows or columns with missing data may lead to loss of valuable information. Instead, imputation techniques are preferred in handling missing values [54]: univariate imputation replaces missing values with a constant, mean, median or mode; and multivariable imputation predicts missing values iteratively using regression models trained on available data.

4.3.2. Outlier detection and treatment

Outliers can distort model performance [55]. Common detection methods include [56] visual inspection (using histograms and one-dimensional plots) and statistical methods (transforming variables to achieve normality and setting thresholds for extreme values).

4.3.3. Feature scaling

Many ML algorithms require normalized data for efficient learning: standardization (Z-score normalization) rescales features to have a mean of 0 and standard deviation of 1; and effect on models is essential for algorithms such as support vector machines [radial basis function kernel] and logistic regression, whereas tree-based models remain unaffected.

4.3.4. Encoding categorical variables

Categorical features must be converted into numerical representations before training ML models; one-hot encoding converts categorical variables into binary vectors, creating a sparse matrix for model compatibility.

4.3.5. Feature selection

Reducing dimensionality improves model performance and computational efficiency. Common techniques include: SelectKBest, which chooses features with the highest scores; the false positive rate test, which selects features with statistically significant *P* values; and least absolute shrinkage and selection operator (LASSO) regression, which performs feature selection by applying L1 regularization.

4.3.6. Handling imbalanced data

CVD datasets often exhibit class imbalance, where healthy patients vastly outnumber those with the disease. This imbalance can substantially impair the performance of standard algorithms, as these models often become biased towards the majority class, thereby neglecting the minority class [57]. Resampling methods address this issue: the synthetic minority oversampling technique (SMOTE) generates synthetic minority class samples by interpolating between existing ones; and undersampling (NearMiss) reduces the majority class to balance the dataset.

4.4. Evaluation metrics

Evaluating model performance requires multiple metrics because of class imbalance in medical datasets. Evaluating CVD prediction models requires appropriate metrics to assess their performance effectively. Given the unique characteristics of each problem, selecting suitable evaluation criteria is crucial [58].

There are some key terms in evaluation metrics: true positive (correctly identifying individuals with CVD); true negative (accurately classifying healthy individuals as disease-free); false positive (misclassifying healthy individuals as having the disease); and false negative (failing to detect the disease in affected individuals).

A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm based on the above key terms. An example of 2×2 confusion matrix is shown in Table A.3.

The common evaluation metrics include accuracy, precision, recall/sensitivity, specificity, F1 score and area under the ROC curve

(AUC). The detailed definitions and formulae for these metrics are presented in Appendix B.

4.5. Model cross-validation

Cross-validation improves model reliability by reducing overfitting. K-fold cross-validation splits the dataset into k subsets, training on $k-1$ folds and testing on the remaining fold. This process repeats k times, ensuring robust performance estimation across different dataset partitions [59].

4.6. Clinical interpretation and application of high F1 score models

Evaluation metrics such as accuracy, F1 score and AUC are essential for assessing model performance, but their clinical significance must be clearly understood for effective medical decision-making. The F1 score is particularly valuable in imbalanced datasets, where accuracy alone can be misleading. For example, in a rare heart disease prediction (1:99 positive-to-negative ratio), a model predicting all cases as negative would achieve 99% accuracy, but would fail to identify any actual cases. A high F1 score, however, ensures a balance between precision (reducing false positives) and recall (reducing false negatives) – critical in healthcare settings where misclassification can have serious consequences.

Clinically, high F1 score models can enhance diagnostic decision support by identifying high-risk patients early (e.g. stroke prediction models aiding patient prioritization), improve treatment recommendations and risk stratification by guiding interventions (e.g. predicting chemotherapy response in oncology) and integrate into clinical workflows via EHRs to provide real-time decision support while minimizing false alarms (e.g. intensive care unit alert systems detecting patient deterioration).

Optimizing the F1 score requires careful data preprocessing, model selection and threshold tuning. Regular clinical validation ensures real-world performance aligns with reported metrics, especially across diverse populations. Additionally, physician training and the use of explainable AI can enhance trust and usability, ensuring that AI-driven insights effectively support clinical decision-making.

5. Current challenges and prospective directions for future research

ML has become an increasingly valuable tool in CVD prediction, particularly with the rise of deep learning methods in medical image analysis and electrocardiogram interpretation. These techniques offer improved accuracy in disease diagnosis and risk assessment [60]. However, several challenges hinder their integration into clinical practice. These challenges include concerns regarding data privacy, the need for extensive and diverse datasets, adherence to regulatory and ethical standards, model robustness and generalization issues. Addressing these limitations is essential for the successful deployment of ML-based CVD prediction models [61].

5.1. Regulatory challenges

ML applications in healthcare must comply with data privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the USA and the General Data Protection Regulation (GDPR) in the European Union. These regulations mandate that patient data must be anonymized and stored securely to prevent breaches and unauthorized access. Additionally, ML models trained on sensitive health data require robust encryption and access control measures.

Although these regulations are crucial for ensuring patient safety and ethical AI use, they also introduce challenges. Strict compliance requirements can slow down innovation and AI deployment, as obtaining regulatory approval from bodies such as the US Food and Drug Administration (FDA) or the European Medicines Agency (EMA) can take years. Moreover, the high costs associated with compliance may present significant barriers for smaller startups and academic institutions.

Another challenge stems from data-sharing restrictions. Privacy laws often limit access to high-quality diverse datasets, which are essential for training unbiased and generalizable ML models. Anonymization is necessary for privacy protection, but can inadvertently reduce the usefulness of data by eliminating clinically relevant correlations. Future research should explore how to strike a balance between ensuring compliance and fostering innovation in AI-driven healthcare, including updating regulations to reflect technological advances while maintaining patient data security.

5.2. Ethical challenges

The integration of AI in healthcare raises several ethical concerns, particularly regarding accountability, fairness and bias. If an ML model makes an incorrect prediction that leads to misdiagnosis or improper treatment, it remains unclear who should bear responsibility – the developers, the healthcare institution or the physician relying on the model. Establishing clear legal and ethical frameworks for accountability is essential to address this issue.

Bias in ML models is another significant ethical concern [62]. These models can inherit biases from the training data, leading to disparities in predictions across different demographic groups, such as race, sex and socioeconomic status. If certain populations are underrepresented in training datasets, the model may perform less accurately for those groups, resulting in unequal healthcare outcomes. Addressing this issue requires proactive measures, including: improving data diversity and representation by collecting balanced datasets that include patients from different demographic backgrounds; and implementing bias detection and mitigation techniques to ensure fairness in model predictions. Future research should focus on developing systematic approaches to address these ethical challenges, ensuring that AI-driven healthcare solutions promote equity and inclusivity.

5.3. Future research directions

Despite significant advances, several research gaps remain in the application of ML for CVD prediction that future studies could focus on.

5.3.1. Enhancing data preprocessing techniques

Limited research has examined the impact of data preprocessing methods on ML model performance in CVD prediction. However, preprocessing plays a critical role in model training, influencing predictive accuracy and generalizability. Techniques such as data cleaning, normalization, feature selection and dimensionality reduction can significantly impact model outcomes. Future research should systematically evaluate various preprocessing strategies to determine optimal workflows that enhance model robustness and reliability.

5.3.2. Strengthening security and privacy protection

Maintaining data security and privacy is paramount in healthcare ML applications. Future research should explore advanced encryption techniques, secure federated learning and privacy-preserving ML approaches that enable data sharing while ensuring compliance with regulations. Collaboration among healthcare institutions is essential for building extensive and diverse datasets, which can improve the accuracy and generalizability of medical ML models [63].

5.3.3. *Integrating IoT and clinical data for personalized healthcare*

The IoT has become increasingly prevalent in healthcare, generating vast amounts of real-time health data through wearable devices and mobile applications. However, few studies have explored the integration of IoT-generated health data with conventional clinical imaging and EHRs. Future research should focus on designing ML pipelines that merge IoT data with clinical datasets to enable more personalized and precise CVD prediction. This integration could enhance early detection, continuous monitoring and individualized treatment recommendations.

5.3.4. *Enhancing model generalization for clinical deployment*

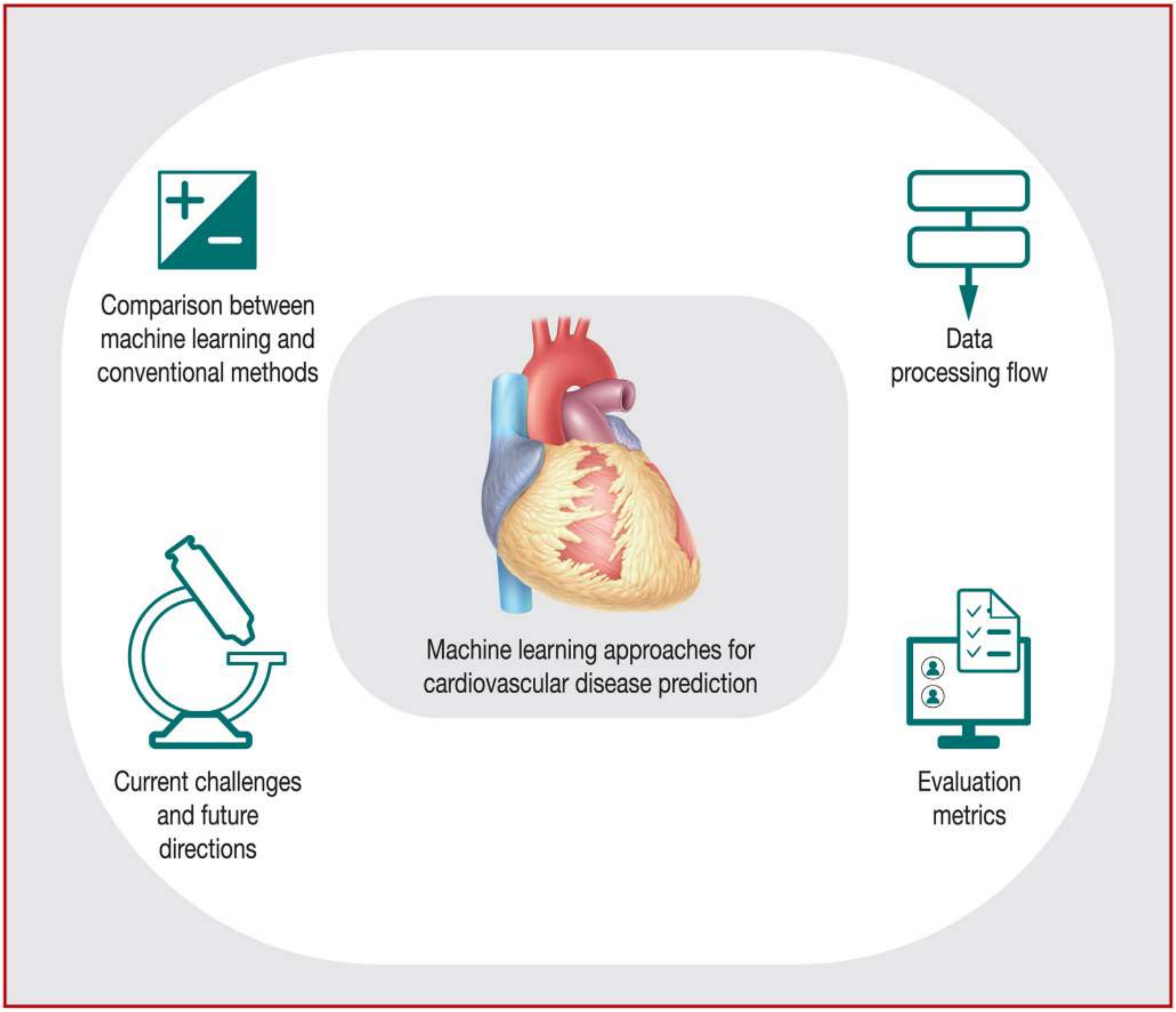
To enhance model generalization for CVD prediction, future research must address critical challenges such as dataset bias, class imbalance and temporal dynamics. Whereas many datasets skew toward urban, older populations or underrepresent women and minorities, synthetic data generation (e.g. generative adversarial networks [GANs]) and federated learning could mitigate these gaps by pooling diverse privacy-preserving data across institutions. Class imbalance in rare events, such as myocardial infarction, requires advanced techniques, such as SMOTE focal loss or cost-sensitive learning, to prioritize recall and reduce false negatives. Robust validation frameworks – such as adversarial cross-validation on multicentre datasets with varying demographics, imaging protocols and EHR formats–are essential to ensure reliability in heterogeneous clinical settings. Additionally, dynamic models capable of continuous adaptation via IoT/EHR data streams and concept drift detection will remain relevant amid evolving risk factors, such as aging populations or pandemics. Transfer learning architectures pretrained on diverse datasets (e.g. ImageNet) could standardize features across imaging vendors, improving interoperability, while adaptive regulatory frameworks would accelerate deployment by balancing innovation with safety accountability. Addressing these challenges collectively will foster equitable, scalable and clinically actionable ML tools for global cardiovascular care.

6. Conclusions

CVD remains a leading cause of global deaths, particularly in developing countries, highlighting the urgent need for effective early detection strategies. AI techniques have demonstrated significant potential in identifying heart disease at an early stage because of their speed, automation and high predictive accuracy. These advantages are particularly valuable in underdeveloped regions where access to healthcare professionals is limited.

Despite these benefits, AI-based CVD prediction also presents certain challenges. The reliability of ML models is highly dependent on data quality, requiring extensive preprocessing to handle issues such as missing values, imbalanced datasets and potential biases. Additionally, although AI can enhance diagnostic precision, its integration into clinical practice necessitates careful validation to ensure robustness and generalizability across diverse populations. Ethical considerations, including patient data privacy and the accountability of AI-driven decisions, must also be addressed to foster trust among healthcare providers and patients.

To maximize the clinical impact of AI in CVD detection, future research should focus on improving model transparency and integrating AI systems with existing healthcare workflows. Collaborative efforts between data scientists, clinicians and policymakers will be essential in translating AI advances into practical clinical tools. Based on the findings discussed in this review, AI-driven CVD prediction holds great promise for enhancing early diagnosis and improving patient outcomes, provided that its limitations are carefully managed, and its implementation is guided by rigorous validation protocols (Central illustration).



Central illustration. ML vs Conventional Methods: Evaluates strengths and limitations of machine learning against conventional statistical models. Innovations and Challenges: summarized key innovation points in current research and discussed regulatory and ethical challenges in ML use in medical applications.

Sources of funding

This study was supported by the National Natural Science Foundation of China (grant No. 82460341), the Science Foundation for Distinguished Young Scholars of Jiangxi Province (grant number 20242BAB23086), the GanPo Talent plan of Jiangxi Province (grant number gpyc20240213), the Natural Science Foundation of Jiangxi Province (grant number 20224BAB216077) and Hunan Province (grant numbers 2022JJ40417 and 2021SK52203), and by open funding from the Key Laboratory of Medical Imaging and Artificial Intelligence of Hunan Province (YXZN2021003).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used OpenAI's ChatGPT in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed, and take full responsibility for the content of the publication.

Appendix A. Supplementary data [64–72] for

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.acvd.2025.04.055.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Negesa LB, et al. Cardiovascular risk behaviour is an emerging health issue in developing countries: a cross-sectional study. Eur J Cardiovasc Nurs 2019;18(8):679–90.
[2] Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. J Am Coll Cardiol 2019;74(20):2529–32.
[3] Li JP, et al. Heart disease identification method using machine learning classification in e-healthcare. IEEE Access 2020;8:107562–82.
[4] Durga S, et al. Survey on Machine Learning and Deep Learning Algorithm used in Internet of Things (IoT) Healthcare. In: 3rd International Conference on Computing Methodologies and Communication (ICCMC). India: Erode; 2019.

[5] Aversano L, et al. Using Machine Learning for early prediction of Heart Disease. In: 14th IEEE Conference on Evolving and Adaptive Intelligent Systems (IEEE EAIS). Larnaca, Cyprus: Ieee; 2022.
[6] (WHO), W.H.O. the eleventh revision of the International Classification of Diseases. 2022; Available from: https://icd.who.int/browse/2024-01/mms/en#426429380. (Accessed on 13 May).
[7] Chaithra N, Madhu B. Classification models on cardiovascular disease prediction using data mining techniques. J Biodivers Endang Species 2018;6(6):1–4.
[8] Pazos-Lopez P, et al. The causes, consequences, and treatment of left or right heart failure. Vasc Health Risk Manag 2011;7:237–54.
[9] Serruys PW, et al. Coronary bypass surgery guided by computed tomography in a low-risk population. Eur Heart J 2024;45(20):1804–15.
[10] Zhang L, et al. Risk factors and outcomes in patients who switched from peritoneal dialysis to physician-oriented or patient-oriented kidney replacement therapy. Ren Fail 2024;46(1):2337286.
[11] Nelson MR, et al. Australian guideline for assessing and managing cardiovascular disease risk. Med J Aust 2024;220(9):482–90.
[12] Small AM, et al. Novel polygenic risk score and established clinical risk factors for risk estimation of aortic stenosis. JAMA Cardiol 2024;9(4):357–66.
[13] Yesiloglu C, Emiroglu C, Aypak C. The relationship between glycated hemoglobin (HbA1c), hematocrit, mean platelet volume, total white blood cell counts, Visceral Adiposity Index, and Systematic Coronary Risk Evaluation 2 (SCORE2) in patients without diabetes. Int J Diabetes Dev Count 2025;45(1):33–9.
[14] Cheng YL, et al. Long-term all-cause death prediction by coronary, aortic, and valvular calcification in patients with acute ST-segment elevation myocardial infarction. Bmc Cardiovasc Disord 2024;24(1):117.
[15] Koletsos N, et al. Accumulation of microvascular target organ damage in systemic lupus erythematosus patients is associated with increased cardiovascular risk. J Clin Med 2024;13(7):2140.
[16] Backus BE, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. Int J Cardiol 2013;168(3):2153–8.
[17] Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2(10):719–31.
[18] Alloghani M, et al. A systematic review on supervised and unsupervised machine learning algorithms for data science. Superv Unsuperv Learn Data Sci 2020:3–21.
[19] Mamun M, et al. Heart failure survival prediction using machine learning algorithm: am I safe from heart failure? In: 2022 IEEE world AI IoT congress (AIIoT). IEEE; 2022.
[20] Ren L, et al. Development of a Clinical Risk Score Prediction Tool for 5-, 9-, and 13-year risk of dementia. JAMA Network Open 2022;5(11).
[21] Kaminski B, Jakubczyk M, Szufel P. A framework for sensitivity analysis of decision trees. Cent Eur J Oper Res 2018;26(1):135–59.
[22] Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learn 2000;40(2):139–57.
[23] Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. JAMA 2016;316(5):533–4.
[24] Toussaint G. Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. Int J Comput Geom Appl 2005;15(02):101–50.
[25] John GH, Langley P. Estimating continuous distributions in Bayesian classifiers. In: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence. Montréal, Qué, Canada: Morgan Kaufmann Publishers Inc; 1995. p. 338–45.
[26] Zhang T. Statistical behavior and consistency of classification methods based on convex risk minimization. Ann Stat 2004;32(1), 56–85, 30.
[27] Vapnik VN. The support vector method. In: Artificial Neural Networks — ICANN'97. Berlin, Heidelberg: Springer Berlin Heidelberg; 1997.
[28] Arnold L, et al. An introduction to deep learning. In: The European Symposium on Artificial Neural Networks; 2011.
[29] Biau G, Cadre B, Rouvière L. Accelerated gradient boosting. Machine Learn 2019;108(6):971–92.
[30] Jonsson A. Deep reinforcement learning in medicine. Kidney Dis 2018;5(1):18–22.
[31] Waring J, Lindvall C, Umeton R. Automated machine learning: review of the state-of-the-art and opportunities for healthcare. Artif Intell Med 2020;104:101822.
[32] Chato L, Regentova E. Survey of transfer learning approaches in the machine learning of digital health sensing data. J Pers Med 2023;13(12):1703.
[33] Zhou J, et al. Graph neural networks: a review of methods and applications. AI Open 2020;1:57–81.
[34] Knuuti J, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J 2020;41(3):407–77.
[35] Shameer K, et al. Machine learning in cardiovascular medicine: are we there yet? Heart 2018;104(14):1156–64.
[36] Steele AJ, et al. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. PLoS One 2018;13(8):e0202344.
[37] Lundberg SM, et al. From local explanations to global understanding with explainable AI for Trees. Nat Mach Intell 2020;2(1):56–67.
[38] Gokulnath CB, Shantharajah SP. An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Comput 2018;22(S6):14777–87.
[39] Feshki MG, Shijani OS. Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network. In: 2016 Artificial Intelligence and Robotics (IRANOPEN); 2016.
[40] Ozcan M, Peker S. A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Anal 2023;3:100130.
[41] Alqahtani A, et al. Cardiovascular disease detection using ensemble learning. Comput Intell Neurosci 2022;2022:5267498.
[42] Chandrasekhar N, Peddakrishna S. Enhancing heart disease prediction accuracy through machine learning techniques and optimization. Processes 2023;11(4):1210.
[43] Jan M, et al. Ensemble approach for developing a smart heart disease prediction system using classification algorithms. Res Rep Clin Cardiol 2018;9:33–45.
[44] Nayeem MJN, Rana S, Islam MR. Prediction of heart disease using machine learning algorithms. Eur J Artif Intell Machine Learn 2022;1(3):22–6.
[45] Haq AU, et al. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Inform Syst 2018;2018:1–21.
[46] Janosi, A., Steinbrunn, W, Pfisterer M, Detrano R. Heart Disease. 1988; Available from: https://doi.org/10.24432/C52P4X. (Accessed on 9 May 2024).
[47] Study, F.H. Available from: https://www.framinghamheartstudy.org/fhs-for-researchers/data-available-overview/. (Accessed on 9 May).
[48] Ulianova, S. Cardiovascular Disease dataset. Available from: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset. (Accessed on 9 May).
[49] Bousseljot, R.-D. PTB Diagnostic ECG Database. 2004; Available from: https://www.physionet.org/content/ptbdb/1.0.0/. (Accessed on 04 April 2024).
[50] Fedesoriano. Stroke Prediction Dataset. 2021; Available from: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. (Accessed on 9 May).
[51] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 2019;7:81542–54.
[52] Kumar Sahoo G, et al. Machine Learning-Based Heart Disease Prediction: A Study for Home Personalized Care. In: 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP); 2022. p. 01–6.
[53] Ogunpola A, et al. Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics (Basel) 2024;14(2):144.
[54] van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1–67.
[55] Hawkins DM. Identification of outliers, 11. Springer; 1980.
[56] Deaton A, Tarozzi A. Prices and poverty in India. Princeton; 2000.
[57] Stando A, Cavus M, Biecek P. The effect of balancing methods on model behavior in imbalanced classification problems; 2023, http://dx.doi.org/10.48550/arXiv.2307.00157.
[58] Haq AU, et al. Intelligent machine learning approach for effective recognition of diabetes in e-healthcare using clinical data. Sensors 2020;20(9):2649.
[59] Gonsalves AH, et al. Prediction of Coronary Heart Disease using Machine Learning. In: Proceedings of the 2019 3rd International Conference on Deep Learning Technologies; 2019. p. 51–6.
[60] Wang ZQ, Zhu KH. Research progress of deep learning in ECG. in International Conference on Intelligent Traffic Systems and Smart City (ITSSC). Zhengzhou, Peoples R China; 2021.
[61] Zhang B, et al. Opportunities and challenges: classification of skin disease based on deep learning. Chin J Mech Eng 2021;34:1–14.
[62] Obermeyer Z, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–53.
[63] Briguglio W, et al. Machine learning in precision medicine to preserve privacy via encryption. Patt Recogn Lett 2021;151:148–54.
[64] Groselj C, et al. Machine learning improves the accuracy of coronary artery disease diagnostic methods. Comput Cardiol 1997;1997:57–60.
[65] Jiang F, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230–43.
[66] Weng SF, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One 2017;12(4):e0174944.
[67] Johnson KW, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71(23):2668–79.
[68] Krittanawong C, et al. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol 2017;69(21):2657–64.
[69] Motwani M, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. Eur Heart J 2017;38(7):500–7.
[70] Zhang J, et al. Fully automated echocardiogram interpretation in clinical practice. Circulation 2018;138(16):1623–35.
[71] Bai W, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson 2018;20(1):65.
[72] Ambale-Venkatesh B, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ Res 2017;121(9):1092–101.

Further reading

Preferred reporting items for systematic reviews and meta-analyses the PRISMA Statement Ann Int Med 2009;151(4):264–269.