<u>Viewpoint</u>

Pitfalls in Developing Machine Learning Models for Predicting Cardiovascular Diseases: Challenge and Solutions

Yu-Qing Cai[1*], BD; Da-Xin Gong[2*], MD, PhD; Li-Ying Tang[1*], BD; Yue Cai[1*], BD; Hui-Jun Li[3], BD; Tian-Ci Jing[2], MD; Mengchun Gong[4], MD; Wei Hu[5], MD; Zhen-Wei Zhang[6], MD; Xingang Zhang[7], PhD; Guang-Wei Zhang[2], MD, PhD

[1]The First Hospital of China Medical University, Shenyang, China

[2]Smart Hospital Management Department, The First Hospital of China Medical University, Shenyang, China

[3]Shenyang Medical & Film Science and Technology Co, Ltd, Shenyang, China

[4]Digital Health China Co, Ltd, Beijing, China

[5]Bayi Orthopedic Hospital, Chengdu, China

[6]China Rongtong Medical & Healthcare Co, Ltd, Chengdu, China

[7]Department of Cardiology, The First Hospital of China Medical University, Shenyang, China

[*]these authors contributed equally

Corresponding Author:
Guang-Wei Zhang, MD, PhD
Smart Hospital Management Department
The First Hospital of China Medical University
Shenyang,
China
Phone: 86 24 88283350
Fax: 86 24 83283350
Email: gwzhang@cmu.edu.cn

Abstract

In recent years, there has been explosive development in artificial intelligence (AI), which has been widely applied in the health care field. As a typical AI technology, machine learning models have emerged with great potential in predicting cardiovascular diseases by leveraging large amounts of medical data for training and optimization, which are expected to play a crucial role in reducing the incidence and mortality rates of cardiovascular diseases. Although the field has become a research hot spot, there are still many pitfalls that researchers need to pay close attention to. These pitfalls may affect the predictive performance, credibility, reliability, and reproducibility of the studied models, ultimately reducing the value of the research and affecting the prospects for clinical application. Therefore, identifying and avoiding these pitfalls is a crucial task before implementing the research. However, there is currently a lack of a comprehensive summary on this topic. This viewpoint aims to analyze the existing problems in terms of data quality, data set characteristics, model design, and statistical methods, as well as clinical implications, and provide possible solutions to these problems, such as gathering objective data, improving training, repeating measurements, increasing sample size, preventing overfitting using statistical methods, using specific AI algorithms to address targeted issues, standardizing outcomes and evaluation criteria, and enhancing fairness and replicability, with the goal of offering reference and assistance to researchers, algorithm developers, policy makers, and clinical practitioners.

KEYWORDS

Background

Cardiovascular diseases (CVDs) refer to both ischemic and hemorrhagic diseases that affect the heart, brain, and systemic vasculature, such as heart failure (HF), atrial fibrillation, acute coronary syndrome (ACS), myocardial infarction (MI), coronary heart disease (CHD), stroke, and cerebrovascular disease. As the most common noncommunicable diseases worldwide, CVDs remain a major cause of death in both low- and high-income countries. In 2019, there were 18.6 million deaths due to CVDs [1]. As awareness of the serious threat that CVDs pose to human

health has grown, many studies have focused on developing tools and guidelines for predicting these diseases. These prediction models can help identify individuals at high risk of developing CVDs, enabling preventive measures to be taken in a timely manner and potentially reducing both the human and economic costs of the disease. In this context, effective CVD risk prediction and prevention are critical to addressing this global challenge [2,3]. As a result, developing reliable and feasible prediction models for CVDs has become an ongoing area of exploration and study.

Early in 1976, the first CVD risk prediction equations were developed by the Framingham Heart Study, called the Framingham risk score (FRS). As one of the most classic CVD risk models, this score was widely applied all over the world in the following decades to provide important guidance for public health and clinical practice [4]. With the development of CVD preventive research, other important prediction tools have emerged for regional applicability, such as the Systematic Coronary Risk Evaluation model in Europe, the QRISK in the United Kingdom, the pooled cohort equation (PCE) for atherosclerotic CVDs (ASCVD) reported recently by the American College of Cardiology and American Heart Association guideline, and so on [5,6]. These models have a common feature: they all consist of fixed equations and lack scalability and are thereby defined as traditional models.

Although traditional models remain the most popular tools in the field of CVD prevention, they have proven to be inefficient and inflexible in the face of rapidly expanding amounts and types of data and increasing clinical requirements for precise, comprehensive, and continuous CVD risk prediction and treatment recommendations. The ideal risk prediction models are expected to include the broadest possible range of parameters and clinically relevant outcomes and to provide real-time and continuous support for doctors' decision-making, such as suggestions for smoking cessation, physical activity, diet, and medication use [7]. Obviously, the traditional models are no longer competent for these responsibilities.

Machine learning (ML) involves endowing computers with the ability to simulate or replicate human learning behavior, allowing them to acquire new knowledge or skills, reorganize existing knowledge structures, and continually improve their performance. As an important subset of artificial intelligence (AI), it has emerged as a promising research area in recent years [8-10]. ML models have also led to a significant evolution in the field of CVD risk prediction, allowing for the handling of new features of existing variables, such as nonlinearity and temporal dynamics, as well as novel variables such as electrocardiography results, medical images, and even genomics data [11-21]. Our recent systematic review [22], which included 486 AI-CVD prediction models across 79 articles, identified that AI has initiated a promising digital revolution in CVD risk prediction, characterized by an increase in the number and dimensions of predictors, as well as a notable diversity in applied algorithms, encompassing 66 specific algorithms across 13 categories.

Despite the stronger predictive ability and more promising development prospects of ML models compared with traditional

ones, the development and clinical application of AI prediction technology are still strictly limited by a series of key problems. In the field of CVD prediction, this situation appears more severe than in other disease prediction or prognosis research [22-24]. As our systematic reviews have found [22], all current published AI-CVD prediction models exhibited a high risk of bias, lacked independent external validation, and had no clinical implementation. Moreover, the onset of CVDs is an exceptionally prolonged process, making prospective clinical validation of a flawed model potentially very costly in terms of research resources and possibly harmful to populations. Therefore, summarizing the potential pitfalls in AI-CVD prediction model research is crucial. Providing researchers with adequate warnings and references before initiating their studies is not only necessary but also of significant importance.

Although there are already some criteria in AI research that can be used as references, such as the Prediction Model Risk of Bias Assessment Tool (PROBAST) [25], Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis (TRIPOD) [26], and AI transparent, replicable, ethical, and effective research (AI-TREE) [27], summarizing and analyzing key considerations and strategies during the research process will directly contribute to the practical implementation, replication, dissemination, and clinical application of model studies [28]. This holds significant practical significance not only for researchers but also for algorithm developers, cohort investigators, policy makers, health care providers, and professionals. Russo and Bonassi [29] previously outlined several "pitfalls" encountered in AI studies in nutritional epidemiology, including issues related to measurement methods, confounding factors, nonlinearities, missing data, overfitting, and interpretability, among others. Similarly, Chiarito et al [30] identified a few "pitfalls" in their review of AI and CVD risk prediction [30]. However, there is currently a lack of a comprehensive and systematic summary specific to the AI-CVD prediction field. In this paper [25,27,31], we summarize and analyze these existing problems and their possible solutions, aiming to provide guidance and references during the process of the development of AI models for the prediction of CVDs and other diseases.

Design and Summary of a Framework for Presenting Existing Pitfalls

To develop a framework outlining the key pitfalls, the following process was executed (Figure 1). A literature review of existing assessment guidelines or tools related to AI or ML medical research was performed, with the search strategy of "assessment tool/guideline"+"AI/ML"+"bias" and their synonyms and related terms (details are shown in Multimedia Appendix 1). Full-text articles in English related to human participants were included, while low-quality or irrelevant articles, such as those discussing the current state of ML or AI or guidelines outside the scope of this study, were excluded. Detailed information is provided in Figure 1. The inclusion and exclusion process was conducted by YQC, YC, and GWZ. Ultimately, 31 papers were included in the preliminary summary of candidate items (Multimedia Appendix 2 [25,27,31-60]). Subsequently, incorporating the

risk issues identified in our previous systematic review of AI-CVD prediction models [22], a final framework was further discussed and confirmed by a panel of experts. This panel included AI experts (TCJ and MG), statisticians (ZWZ), clinicians (DXG, XZ, and GWZ), and information technology specialists (WH), among others. As summarized in Multimedia Appendix 3 [15,18,21,28,61-100], the framework encompasses four major categories with 15 subcategories: (1) data quality

(data source, subjective factors of researchers, incomplete data, and parameter acquisition methods); (2) data set characteristics (small sample size, low event rate, characteristics of the data distribution, and multimodal data); (3) model design and statistical methods (outcome definitions, incomplete inclusion of covariates, overfitting, and defects in evaluation criteria); and (4) clinical implications (generalization, interpretability, and AI ethics).

Figure 1. The flowchart for developing a framework for this paper based on the assessment guidelines or tools in the field of medical artificial or machine learning research. CVD: cardiovascular disease; ML: machine learning.
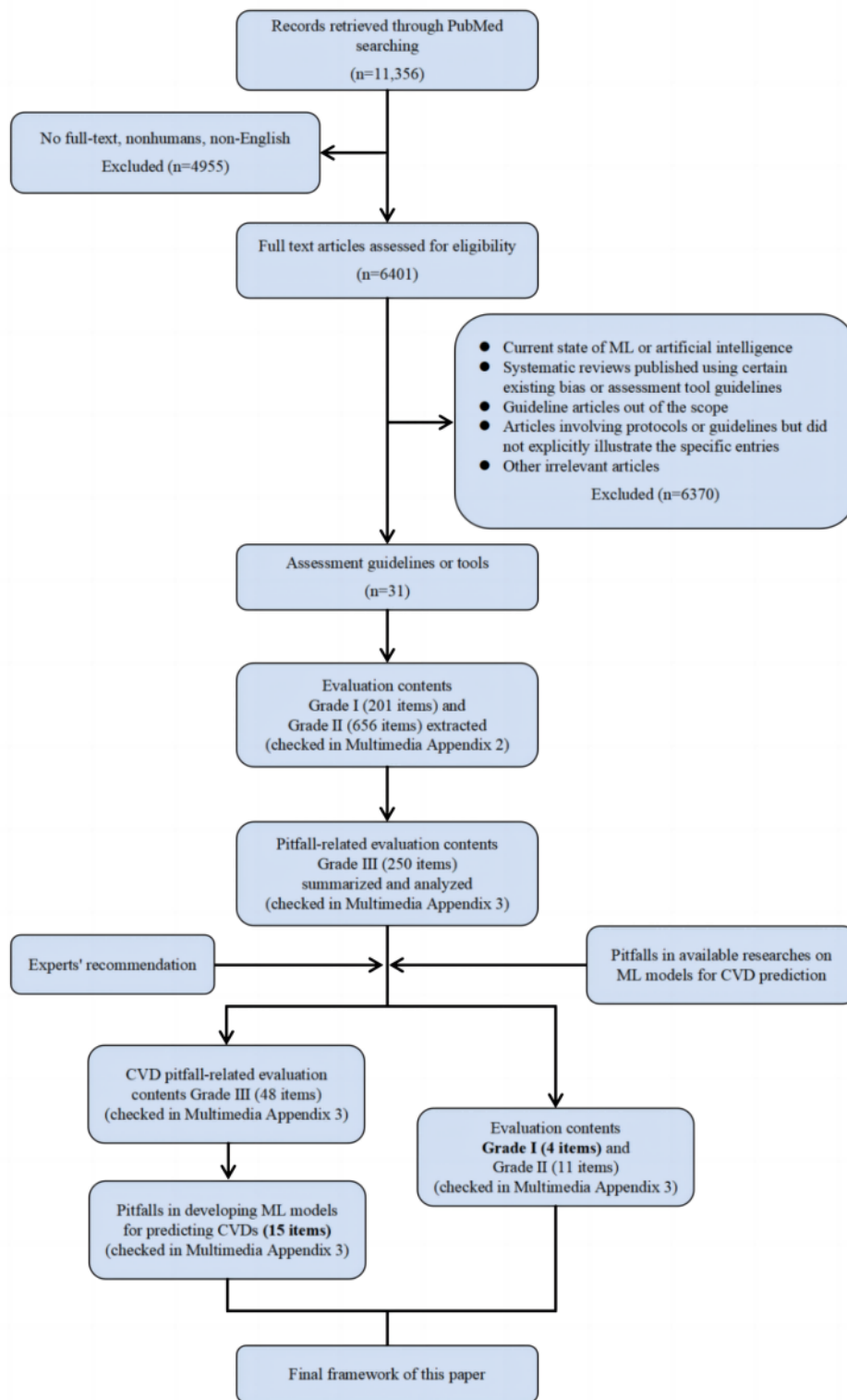
The Effectiveness of the ML Models Still Largely Depends on Data Quality

Overview

Data are undeniably a fundamental element in ML models, which are inherently sensitive to the quality of the data used for training and validation. The adage "garbage in, garbage out" is particularly pertinent here [101]; poor-quality input data can also result in a biased ML model, and even minor errors or biases in training data can lead to unforeseen consequences in a model's predictions. These considerations raise ethical concerns regarding the reliability of decisions based on ML model predictions. Consequently, understanding and mitigating pitfalls in data collection, preprocessing, analysis, and application become critically important. Specific methods involve selecting objective indicators collected through reasonable means based on a thorough evaluation of data authenticity.

Data Recourse Is the Key Factor

It is well-known that data accuracy is a key influencing factor for data-based prediction models, as inaccurate data can significantly impact the predictive ability of subsequent models. One of the main reasons for data inaccuracy is the collection of subjective data, such as self-reported data on blood pressure (BP), obesity, and family history, as noted by Hippisley-Cox et al [102] and Manuel et al [103] in traditional models, based on the Canadian Community Health Survey (104,219 participants) and QResearch database (3,610,918 participants), respectively. In fact, AI models have encountered the same dilemma [104,105]. As reported by Han et al [61] in their ML research based on the Korea Initiatives on Coronary Calcification (86,155 patients), inaccurate data from self-administered questionnaires had a limited contribution to model development; after feature selection based on information gain ranking (such as the history of diabetes and hypertension, this limitation resulted in an exclusion of important predictive factors from 10 most pertinent variables of a future all-cause mortality (ACM) events. Ultimately, this underestimation of important risk factors may lead to insufficient attention being paid to them, thereby misdirecting primary prevention strategies such as aggressive glycemic control.

Similar to traditional models, the most effective solution in AI models is still to collect objective data, instead of relying solely on self-reported data whenever possible. Even electronic health records (EHR) can contain incorrect data, which may be difficult to detect and therefore easily overlooked [62]. For example, in a study conducted by Rodriguez et al [63], race data, which were obtained from a community-based EHR with 231,622 participants, were mainly self-reported, while other data were inferred through validation methods that were based on PCE, leading to possible misclassification, and may affect the risk prediction for ASCVD. By way of illustration, it is advocated by Teoh [64] that the diagnosis or the value of the check index should be applied to further confirm the disease status of the participants, including hypertension, diabetes, and dyslipidemia reported by the participants themselves, and they found that the area under the curve (AUC) dropped from 0.623 to 0.608 when

the diabetes examination inputs were removed (including blood glucose, glycated hemoglobin [as standardized by Japan Diabetes Society], and glycated hemoglobin [as standardized by National Glycohemoglobin Standardization Program]). These examples clearly demonstrate the impact of data authenticity on the robustness of the models. Unfortunately, even some classic scoring tools, such as PROBAST, TRIPOD, and CREMLS (Consolidated Reporting Guidelines for Prognostic and Diagnostic Machine Learning Models), do not specify methods for collecting objective data [32,33,106,107]. Therefore, we suggest the implementation of appropriate validation measures during model development.

The Subjective Factor of Researchers Is a Potential Interference

In some studies, basic background, sociodemographic, and follow-up information may be collected by local doctors and nurses [65,108,109]. Although these interviewers have been strictly trained, the data collection process is likely to be interfered with by their personal thoughts. Benjamins et al [66] reported that in the Glycometabolic Intervention as Adjunct to Primary Percutaneous Intervention in ST-Segment Elevation Myocardial Infarction-III study, which included 222 people and defined ventricular function dysfunction as an outcome, using data created by local observers to train and validate the U-Net model may result in an unstable model quality. Liu et al [67] analyzed the registry data of 53,213 inpatients in the Cardiovascular Department of Xiangya Hospital; they found that student extractors, who manually filled missing data using a hierarchical mean filling method, subjectively adjusted variable parameters of the extracted data based on experience and introduced a potential risk of bias. Although these reports mention the potential impact of researchers' subjective factors on the models, they lack effective proof methods, highlighting that such factors are often overlooked and difficult to evaluate. Therefore, it is essential to identify subjective factors of researchers before the study.

The most common and effective approaches to address this issue are to improve training and supervision of data collectors. While both traditional and ML models require manual data processing, some studies have shown that AI can potentially replace humans in data processing to reduce the impact of subjective factors. For instance, in the Systolic Blood Pressure Intervention Trial (SPRINT) and the Hong Kong eHealth cohort studies with 8133 and 1094 participants, respectively, Tsoi et al [68] used a K-means clustering ML algorithm to identify variations in BP without the need for human adjustment, which resulted in low clustering similarities (Davies-Bouldin Index: 0.653 in SPRINT and 0.680 in the Hong Kong eHealth cohort). This not only further highlights the importance of controlling subjective factors in model construction but also provides an alternative solution beyond simply enhancing training.

Incomplete Data Also Affect the Effectiveness of ML Models

Incomplete data are a common problem in retrospective studies and may significantly impact the final predictive results of ML studies. For instance, in the Melbourne Collaborative Cohort Study with 32,611 participants, Sajeev et al [69] reported that

32% of high-density lipoprotein cholesterol data were missing and that this resulted in a lower AUC of 0.753 (95% CI 0.729-0.777) in a logistic regression (LR) model, while an AUC of 0.874 (95% CI 0.833-0.915) was achieved in another similar LR model that was developed by the North West Adelaide Health Study, which had only 41 missing data among 3654 participants. It is also argued by Han et al [61] that unmeasured factors could introduce bias in their ACM prediction ML model, which used the Korean Initiatives on Coronary Calcification registry, a retrospective cohort. The severity of this issue lies in its potential to also cause underestimation or neglect of the impact of risk factors in actual clinical practice, thereby affecting CVD prevention strategies.

Both traditional and ML models face the issue of missing data. In traditional models, missing data are often directly excluded, which can impact model stability and weaken its significance [70,108,110]. In contrast, ML models use several statistical methods to address the problem of missing data and optimize the models, such as multiple imputation, multiple imputation with chained equations, logitBoost, multiple imputation by fully conditional specification, Markov chain Monte Carlo method, median imputation, and k-nearest neighbor (KNN) imputation [71,72,111,112]. The application of these statistical methods gives ML models an advantage over traditional models in handling missing data. For instance, in a study that included 423,604 participants from the UK Biobank, Alaa et al [73] used the missForest algorithm to impute missing data in an ML model, achieving a significantly higher AUC of 0.774 (95% CI 0.768-0.780) than with FRS (AUC=0.724, 95% CI 0.720-0.728). In the ATTICA prospective study involving 2020 participants, Dimopoulos et al [112] excluded variables with missing values in >70% of the individuals. In order to improve the performance of the CVD risk estimator, the missing data in part of the remaining 22 variables were then imputed by KNN imputation, resulting in relatively high accuracy rates (KNN: 96%, random forest [RF]: 99%, decision tree classifier: 99%). Therefore, ML models can better manage missing data due to the application of the aforementioned statistical methods. Furthermore, ML models also can register patients with missing necessary variables in traditional models [62], providing an advantage in handling incomplete data. In addition, besides the aforementioned statistical methods, Weng et al [71] used a novel approach in the Clinical Practice Research Datalink (CPRD) study that included 378,256 patients, using missing data as an independent variable to predict the risk of CVDs. Specifically, they created dummy variables to indicate missing continuous variables and a separate category of "unknown" to represent missing categorical variables such as the Townsend deprivation index and race, with an acceptable AUC of 0.764 (95% CI 0.759-0.769) and a specificity of 70.7% in the neural networks model eventually. Although this method needs to be further confirmed, exploring such new approaches provides additional options for ML models in handling missing data, thereby enhancing AI-CVD prediction models.

Appropriate Parameter Acquisition Method Is Another Important Factor

In most studies predicting the risk of CVDs, researchers have typically used one-time measurements of risk factors, ignoring the fact that some factors exhibit significant variability over time, such as systolic BP (SBP), plasma glucose levels, low-density lipoprotein, and serum total cholesterol [113-115]. As demonstrated in the SPRINT study mentioned in The Subjective Factor of Researchers Is a Potential Interference section, BP measured at fixed times (visit-to-visit BP variability) can vary greatly from 24-hour ambulatory BP variability ($r^2<0.026$), for the data that are acquired intermittently with a short follow-up time may introduce bias in the data [68]. These common problems illustrate that variables measured at a certain time point cannot necessarily represent the true level of the variable in the population. In fact, there have already been diagnostic criteria for multiple measurements of some risk factors in guidelines, for single measurements are likely to introduce significant bias within clinical work. For example, Stergiou et al [116] advocated that a minimum of 2 to 3 office visits at 1- to 4-week intervals are frequently necessary for the office BP assessment.

The solving method is to use repeated measurements of risk factors to improve the model prediction [117]. In the National Health Insurance System-National Health Screening Cohort (NHIS-HEALS) study with 361,239 participants selected, Sung et al [74] advocated that by offering discrimination and calibration with repeatedly measured data, the deep learning model takes advantages in CVD risk prediction in the EHR era (female participants: AUC=0.94, 95% CI 0.91-0.97 in; male participants: AUC=0.96, 95% CI 0.95-0.97 in). In another study including 80,964 people, multiple SBP recordings from EHR (provided by UK primary care) were analyzed. It was found that the multiple SBP recordings had a better correlation with CVDs than a single recording, as identified by an increased hazard ratio from 1.22 (95% CI 1.18-1.30) to 1.39 (95% CI 1.31-1.46) with the use of repeated measured factors [75]. Besides, in the NHIS-HEALS study, Cho et al [76] handled continuous variables by using their mean, minimum and maximum values, and SDs in the developing of recurrent neural network models and achieved a high level of discriminative accuracy (female participants: AUC=0.921, 95% CI 0.908-0.934; male participants: AUC=0.896, 95% CI 0.886-0.907), which was highlighted by authors as significant improvement compared with single-measured method [76]. Furthermore, collecting data multiple times can enhance data consistency and to the benefit of exploring the changing trend of data.

Data Set Characteristics Largely Affect the Effectiveness of the ML Models

Overview

The impact of the data set characteristics used for ML on risk prediction is crucial and should not be disregarded. This includes several elements: the impact of the small sample size and low event rates, the alignment of data distribution characteristics with the model, and the challenges associated with the application of multimodal data. These pitfalls can significantly impede ML studies, highlighting the critical need for heightened attention in the various stages such as research design, implementation, data analysis, model construction, paper

preparing, and even in submission and peer review processes. Specific methods include using sufficiently large cohorts, extending follow-up periods, using specific AI algorithms to address targeted issues, and selecting clinical routine examinations as risk factors whenever possible.

Small Sample Size Is a More Serious Problem

It is well known that insufficient sample size will lead to the risk of bias for the study in the traditional model, and the evaluation standard of sample size has been developed by the PROBAST [106]. Therefore, in order to improve the credibility of research conclusions, more and more researchers have used large data sets that are multiethnic cohorts with high quality [77,118-120], although there are still many substandard studies [108,121]. However, in the field of ML prediction models, the problem of insufficient sample size seems to be more serious. In our previous research of 79 CVD risk prediction studies, the number of participants included in the ML model could even be as low as 80 [22]. The lack of a sufficient sample has become the biggest contributor to the high risk of bias [28]. For example, in a retrospective observational study of 420 patients, Ponomartseva et al [78] stated that when the sample size is smaller than the optimal ML method size, the accuracy of the model may change. In another study carried out with 451 consecutive patients from a tertiary hospital, conventional univariate and multivariate analyses were limited because of the small sample size; the median accuracy of artificial neural network (ANN) models in predicting recurrent stroke was only 75%, as reported by Chan et al [79].

Although a larger population size was prone to the better performance and reliability of the models, as revealed by Alaa et al [73] in an experiment with a series of subpopulations of varying sizes and a fixed number of variables, the sample size of the ML algorithms is dependent on various factors, including the types of algorithms, the number of index, the features of the sample, and so on [122-124]. For example, the classical events per variable method, which often follows the "one in ten rule" for the sample size calculation, may not be applicable to certain ML algorithms due to their specific operation mechanism [22]. Therefore, it is strongly recommended that the sample size calculation be performed before initiating ML research, considering the various factors mentioned above, due to dramatic differences in the mechanisms of sample size calculation between ML and traditional models. Meanwhile, the upcoming PROBAST-AI, a special tool for bias evaluation for AI models, has been expected to provide a reference standard [125-127]. Before this, we are inclined to recommend using the "one in ten rule" as the minimum standard to include as many samples as possible.

Application of the ML Models Is Also Troubled by Low Event Rate

The low incidence rate, a common problem for prediction model construction, may restrict the generalizability and predictive ability of the prediction model in practice [74,80,128]. It may lead to the bad performance of the model in general or cause bias in predictions for a particular segment that has a low incidence rate. For example, on the basis of the data from 13,291 participants from 7 epidemiological cohorts, Jdanov et al [77]

found that the low incidence rate of CVD deaths (<7% in male participants and <1% in female participants) at a young age (aged 45-54 years) affected the prediction outcomes at young ages, with a substantial underestimation of CVD mortality risk by 40% to 45% for male individuals and 3% to 4% for female individuals. Therefore, based on this reasoning, if applied in clinical practice, the low event rate may result in the underestimation of a substantial number of individuals who are at high risk for CVD, thus reducing the role of primary prevention.

The problem of low event rate can be overcome by the following methods, such as filtering of oversampled data using noncooperative game theory [80], synthetic minority oversampling technique [129], random oversampling [130], random undersampling [19], bootstrapping [81], and the stacking paradigm [131]. However, these methods can only improve the performance of the ML from the statistical method and cannot solve the problems with the data themselves. By using the data solely from the DaVita Inc EHRs (124,097 participants), Goldstein et al [82] reported a direct way to increase the incidence rate by broadening the outcome range and showed good training and test set loss at the same time. However, by using this method, the outcome will become more heterogeneous and will decrease the quality of the prediction model, and thus this method may not be an adequate option. The better approach may be to extend the follow-up duration or to include more different populations as mentioned above to improve the reliability of ML [61].

Characteristics of the Data Distribution Underscore the Critical Importance of Model Selection

The cornerstone of constructing predictive models lies in the analysis, mining, and use of data relationships between covariates (risk factors) and gold standards (outcome events). Traditional statistical methods often rely on the assumption of linearity when building predictive models. However, it is essential to acknowledge that in clinical practice, many risk factors and outcome events exhibit nonlinear relationships, although linear relationships may predominantly characterize some factors, such as glycated hemoglobin or smoking pack-years. For instance, Angeli et al [83] and Lip et al [84] reported a J-curve relationship between BP and CVDs or ACM. Similarly, dietary factors such as salt, carbohydrates, and fats have demonstrated U- or J-shaped relationships with CVD outcomes [132,133]. Unfortunately, certain ML algorithms are not well suited to handle nonlinear relationships, including linear regression, LR, and so on. Consequently, if linear algorithms are used to construct predictive models that incorporate nonlinear elements, the result may lead to inaccurate associations and distorted effect estimates. In a study based on the National Health and Nutrition Examination Survey cohort (37,079 participants), Dutta et al [85] developed 6 AI models with nonlinear factors (eg, BP) for predicting CHD, and expectedly, the LR algorithm achieved the worst performance (AUC=0.713). Similarly, Wang et al [70] included 40,711 participants from the Life Risk Pooling Project cohort to build 4 AI models with nonlinear factors (eg, BP) for predicting CVDs and also found the worst performance in linear naive Bayes algorithm (AUC=0.786, 95% CI 0.726-0.735; P=.001) [70]. In addition,

many predictive models aimed to assess long-term risk for adverse events, but commonly used ML models often struggle with handling time-to-event variations and censored patient data, significantly limiting their performance. Li et al [72] demonstrated that LR algorithms that neglect censoring tend to substantially underestimate CVD risk to only 2.2% to 5.8% when compared with Cox models that had a risk of 9.5% to 10.5%, which are better suited for survival data analysis.

Therefore, it is imperative to emphasize that the application of nonlinear AI algorithms plays a pivotal role in the research of constructing predictive models that encompass intricate data relationships. These algorithms have the capability to capture and model intricate patterns, interactions, and dependencies within the data that may not be adequately addressed by linear approaches. As in the 2 examples elaborated in the previous paragraph, support vector machine (SVM) had the best performance with an AUC of 0.776 in the study by Dutta et al [85], and RF had the best performance with an AUC of 0.892 in the study by Wang et al [70]. When it comes to survival data, random survival forest (RSF) appears to exhibit superior performance, owing to their ability to handle complex, time-dependent relationships and censored data more effectively, ultimately resulting in more accurate survival predictions. For instance, in a study based on the Atherosclerosis Risk in Communities cohort (14,842 participants), Zhuang et al [86] developed 4 survival models (Cox proportional hazards model, Akaike information criteria for Cox regression, least absolute shrinkage and selection operator for Cox regression, and RSF), and the RSF algorithm achieved the best performance in predicting CHD (AUC=0.80, 95% CI 0.79-0.81) and ACM (AUC=0.78, 95% CI 0.77-0.78). Another example is meant by Ambale-Venkatesh et al [87], where 6814 participants from the Multi-Ethnic Study of Atherosclerosis (MESA) cohort were used for ACM risk. Eventually, the RSF algorithm (AUC=0.84) performed the best, yielding better results compared to those obtained with other algorithms (eg, Cox least absolute shrinkage and selection operator: AUC=0.80).

Multimodal Data–Based Prediction Model Pitfalls Limit Clinical Applications

The number of studies of multimodal data–based prediction models for CVDs has risen substantially recently, and there have been a variety of data types, such as radiomics [88,134], proteomics [135], and genomics [136]. The multimodal data–based prediction model, which integrates available heterogeneous data into a unified framework, can fully consider the importance of each modality and incorporate information from multiple aspects, thus improving model performance [136]. However, the clinical accessibility of the used multimodal data presents another pitfall. As advocated by Pujadas et al [88], there was a limitation in the multimodal data–based CVD prediction model with cardiovascular magnetic resonance radiomics, because cardiovascular magnetic resonance was not a routine examination. More significantly, according to the AI-TREE criteria, the inclusion of unconventional or difficult-to-obtain examination results is a crucial factor influencing the clinical application of the model [27]. This may be solved by applying more commonly used tests (eg, electrocardiogram [ECG]) or disease-specific imaging

examinations. For instance, it is reported by Lou et al [89] that ECG-based patient characteristics from 2 hospitals in the Tri-Service General System were used for CVD risk prediction, and the deep learning models performed well, yielding excellent results (AUC>0.90). Another example of the solution was carried out by Chao et al [21], who included clinical routine 30,286 low-dose computed tomography data from 2085 patients with lung cancer in the National Lung Cancer Screening Trial to construct a deep learning CVD risk prediction model. Therefore, we suggest that multimodal model studies should focus on the practicality of data acquisition to enhance the feasibility of primary prevention of CVDs.

In addition, studies emphasizing multimodal data often disregard the inclusion of classical risk factors, neglecting essential parameters crucial for comprehensive prediction models. The overarching focus on multimodal data variables can lead to decreased model performance due to the absence of integrated incorporation of traditional risk factors. For instance, in the ECG methods, for the prompt identification of coronary events study (499 patients), Al-Zaiti et al [15] constructed ACS risk prediction models by using 554 temporal-spatial features of the 12-lead ECG without clinical examination factors such as total cholesterol or lifestyle variables such as smoking, which led to unsatisfactory validation performance (LR model: AUC=0.67; gradient boosting machine [GBM] model: ACS=0.71). Therefore, it is strongly recommended that the inclusion of classic predictive factors should form the cornerstone for constructing multimodal models. This is evident in the Singapore epidemiology of eye diseases study (comprised of over 70,000 images), where Cheung et al [134] not only applied imaging data but also comprehensively incorporated multiple types of variables in CVD risk estimators, and the best model achieved a perfect AUC of 0.948 (95% CI 0.941-0.954).

Challenges Arising From Model Design and Statistical Methods

Overview

In AI-CVD prediction research, model design and statistical methods play a critical role not only in aspects of model performance such as discrimination and accuracy but also in reproducibility and generalizability for clinical application. Key pitfalls in this area include issues of outcome definitions, incomplete inclusion of covariates, overfitting, and inadequate evaluation metrics. Some of these pitfalls stem from inherent problems associated with AI algorithms, necessitating researchers to thoroughly understand and mitigate their impact. In addition, other pitfalls can arise from design and implementation choices made by researchers, underscoring the need for careful planning and execution to avoid such issues during the research process. Specific methods include designing studies with standardized outcomes and evaluation criteria, incorporating all necessary variables, and using appropriate algorithms and validation techniques.

The Outcome Definitions May Be Detrimental to the Significance of the Study

The definitions of the CVD outcomes show considerable heterogeneity in both traditional and ML models, concluding from the facts that there have always been some differences more or less in the detailed definitions among almost all studies. All the 3 versions of QRISK set the end points as CVDs, which are defined as a composite outcome of CHD, stroke, and transient ischemic attacks [5,111,137]. Systematic Coronary Risk Evaluation in Europe aimed to predict fatal CVDs, so its end point was CVD mortality [138]. PCE and Prediction for ASCVD Risk in China defined ASCVD as nonfatal MI, CHD death, or fatal or nonfatal stroke [139,140]. The outcomes of adding social deprivation and family history to CVD risk assessments were CVD-cause mortality, CHD, cerebrovascular disease, or interventions for coronary artery (coronary artery bypass grafting or percutaneous transluminal coronary angioplasty) [141]. These differences, to some extent, limit the comparability, reliability, and generalizability of the models, thereby affecting their clinical applicability for CVD prediction.

In studies of ML prediction models, heterogeneity is even more significant, with various origins of outcome definitions, such as disease codes (*International Classification of Diseases [ICD]*, *Ninth Revision* [*ICD-9*] or *ICD-10*), self-reports, and other international guidelines. For example, in a study with 31,466 participants from the UK Biobank, You et al [90] defined CVD events as MI or stroke with *ICD-9* and *ICD-10* codes. Conversely, Cho et al [76] in the NHIS-HEALS study, developed risk prediction models that used death from CVDs, MI, coronary arterial intervention, or bypass surgery stroke as the outcome definitions with only *ICD-10* codes. This heterogeneity in outcome definitions can lead to significant issues such as bias in model performance evaluation, reduced generalizability, challenges in data integration and standardization, difficulties in clinical application, and even misguidance in CVD prevention strategies. Therefore, it is crucial to address this problem adequately. To mitigate these issues, we strongly recommend that risk prediction models be developed with standardized use of *ICD* codes. This standardization would help in achieving consistent outcome definitions, thereby improving the reliability, generalizability, and clinical applicability of the prediction models.

Incomplete Inclusion of Covariates May Diminish the Study's Value

In many CVD risk prediction studies, several classical risk factors have been omitted. This phenomenon is not limited to retrospective studies using existing databases but is also prevalent in prospective research. By way of illustration, Chua et al [81] included 638 participants from Sandwell and West Birmingham Hospitals NHS Trust (local data set) to build 2 atrial fibrillation risk calculators. They did not include classical factors such as BP and smoking status and achieved a low accuracy performance with an AUC of 0.684 (95% CI 0.62-0.75) and 0.697 (95% CI 0.63-0.76), respectively. In the China Health and Retirement Longitudinal Study, Chen et al [91] included 9821 participants in the development of a stroke risk calculator. However, this classical public data set lacked classic variables

such as total cholesterol, which resulted in a mediocre performance (AUC=0.7388) in the model. Such omissions may potentially lead to inadequate consideration of confounding factors, resulting in model bias and decreased performance or even rendering the model practically inapplicable [142]. From a statistical perspective, these classic risk factors can be considered confounders, and ignoring them may cause confounding, which represents a critical error and is a primary source of systematic errors that can manifest when assessing causality, potentially leading to misinterpretation of the results [143]. In accordance with the AI-TREE criterion [27], a key consideration is whether the available data can effectively address the clinical question at hand. In other words, a data set lacking essential predictors that are known to be relevant to an outcome is unlikely to satisfactorily address related inquiries. Therefore, studies that exclusively focus on model development while disregarding considerations for clinical application may deviate from sound research principles.

One potential approach to mitigate these challenges is to ensure that potentially biased features, such as ethnicity and social determinants of health, are explicitly incorporated into the models [142,144]. As suggested by Suri et al [28] in a study on understanding bias in ML systems for CVD risk assessment, achieving a robust ML-based design for CHD and CVD prediction necessitates the integration of traditional, laboratory, image-based, and medication-based biomarkers. In the realm of clinical research, where there is often a high number of variables and confounding factors, the adoption of high-capacity AI models is warranted [145,146]. For instance, Kakadiaris et al [80] obtained 6459 participants from the MESA cohort to build a CVD risk calculator based on SVM. They included all classic risk factors such as age, sex, smoking status, diabetes, SBP, total cholesterol, and high-density lipoprotein and achieved a rather high AUC of 0.94 (95% CI 0.93-0.95). In addition, in a study conducted with an outpatient health care system, Ward et al [62] enrolled 262,923 participants with risk factors that contained all classic risk factors as mentioned above for ASCVD risk prediction, and the AI model performed well with an AUC of 0.835 (95% CI 0.825-0.846). These also indicated the importance of incorporating complete covariates for accurate modeling, which is more conducive to individualized, precise, systematic, and comprehensive prediction and assessment of CVDs. This, in turn, facilitates targeted prevention and monitoring.

Overfitting Determines the Performance of ML Models

Overfitting, a more severe problem in ML, means that the data can fit relatively perfectly with the derivation queue during training and derivation, but it will yield large biases and unreliable results when applied to other data, consequently leading to difficult generalization of the model [147,148]. The influence factors accounting for overfitting of ML models are summarized as follows: the features of algorithms, the number of variables, cohorts, outcome events, and so on. It is found in the research performed by Weng et al [71] that the more complex the model, the more likely it was prone to overfitting, while Commandeur et al [92] revealed that the few numbers of the derivation or the low event rate will cause the overfitting problem, which can be assessed by comparing the accuracy of

the training set and the test set. These reports suggest that selecting appropriate algorithms and addressing the issue of low event rates have a decisive impact on preventing model overfitting, necessitating thorough consideration by researchers during study design.

There are numerous approaches to reduce the impact of overfitting, and the most commonly used include pretraining, hyperparameter selection, regularization, and cross-validation [149,150]. There are also other approaches, such as data-smoothing techniques, separating training, tuning, testing data, and the filtering of oversampled data using noncooperative game theory algorithm [80,92]. In addition, in the Prospective Cardiovascular Münster Study, which included 5159 participants, Voss et al [93] have proposed 3 solutions to address overfitting for CHD event estimations: cross-validation, stopping training when the errors in validation data sets are at a minimum, and rendering the networks with synthetic data and modifying them until the results are plausible. Eventually, the average negative log likelihood (ANLL) that explains the loss function of the model and the bias value attained comparatively minor numbers in the probabilistic neural networks model for both training (ANLL: 0.1712, bias: 0.0396) and testing sets (ANLL: 0.1807, bias: 0.0396).

Inadequate Evaluation Metrics Still Exist in the Field of ML Model Prediction

The evaluation metrics help researchers assess the performance of models, and thus, defects in them can lead to obstacles in drawing practical conclusions. For example, although the AUC is the most commonly used evaluation criterion for model assessment, it may not accurately reflect the risk of CVDs in the population. This is because a person with a very low risk makes the same contribution to the AUC's value as a person with a very high risk, as noted in a study developed by the CPRD with 3,660,000 participants [72]. Similarly, the Net Reclassification Index has been shown to be insensitive to changes in the model. For example, in a MESA study (5878 participants) reported by Polonsky et al [94], the Net Reclassification Index remained substantially constant even after including or excluding certain participants. These examples indicate that model evaluation cannot be adequately represented by a single metric; rather, it should be a comprehensive, multifaceted, and complex system, especially for AI models, in line with preventive medicine principles.

It is advisable to use a comprehensive evaluation of predictive model capabilities by considering various metrics such as accuracy, sensitivity, precision, specificity, F_1-score, decision curves, and so on [151], which are also required to be reported in ML prediction studies as advocated by several reporting guidelines [32-34]. This is not only crucial for assessing the reliability of the model but also beneficial for its accurate application. Some commercial medical devices, such as ultrasound-assisted ECGs, have already contained criteria with high transparency and high clarity that are not available for predictive models. Thus, marketizing prediction models may be a solution. Moreover, in contrast to the predictive model, the intervention AI evaluation has different criteria, which are more rigorous and systematic. For instance, as the guideline of clinical

trial reports for interventions involving AI, CONSORT (Consolidated Standards of Reporting Trials)-AI recommended presenting absolute and relative effect sizes for binary outcomes [35]. Therefore, we recommend that during model development, comprehensive calculation of these evaluation metrics is essential, as it is crucial for subsequent model dissemination and selection for clinical applications.

Clinical Implication Remains a Significant Challenge in This Field

Overview

In fact, the factors affecting the clinical application of AI-CVD models are complex. Beyond the pitfalls discussed above, 3 directly related and most prevalent factors that researchers should particularly consider are the problem of generalization, the lack of interpretability, and the limitations of AI ethics. These factors not only require researchers to minimize their impact during the study process but also necessitate collaborative efforts from clinicians, policy makers, and even the general public to overcome and address them. This includes clinicians integrating AI tools into their practice, policy makers establishing supportive regulatory frameworks, and the general public engaging in discussions about the ethical implications.

The Problem of Generalization Has Largely Limited ML Algorithms Application

Generalization, which represents the ability to adapt to novel situations, is rather important for the applicability of CVD risk prediction models. However, both traditional and ML models have the problem of generalization, as they are developed based on specific populations. For example, many studies have used cohorts predominantly consisting of White individuals, men, or participants from a single center, even during a relatively narrow socioeconomic range with well-educated and low-risk factor burden compared with the general population [152-155]. Thus, risk prediction models were largely restricted to application of cross-cohorts. The Pooled Cohort Equations to Prevent Heart Failure tool, which was developed with the data of 11,771 individuals from 7 community-based cohorts, was unable to offer precise HF risk estimates for individuals who were not classified to the non-Hispanic White and non-Hispanic Black racial or ethnic groups [7]. The Cox regression models developed by Li et al [156] contained data from 44,869 participants from the Evidence for Cardiovascular Prevention from Observational Cohorts in Japan research group. Models were developed to predict death from CHD, stroke, and CVDs, but they may lack accuracy when predicting risk in women, for the corresponding P values of the Hosmer-Lemeshow test were .27, .002, and .04 in women and .51, .49, and .25 in men, respectively.

In ML, it is still a challenge to generalize models with specific populations, although this problem has been improved to a certain extent due to the elimination of some restrictions by AI algorithms. It has been documented by Bouzid et al [18] that using patients from only 1 region (even multiple hospitals) would significantly influence the promotion of ML models. In terms of improvement, it has been proven by Ward et al [62]

that ML models developed in the United States or Europe could be expanded to Asian and Hispanic populations, and they achieved AUCs of 0.803 (95% CI 0.743-0.863) and 0.768 (95% CI 0.663-0.874), respectively, without the limitation of race in the PCE. In addition, the application of PCE is limited to the population aged from 40 to 79 years, while ML models have eliminated this limitation [62]. Moreover, the LR and SVM algorithms have been confirmed to overcome the age limitation (<79 years), which caused 26% of all cases in FRS not to be classified [95].

The way to solve the problem of generalization is to include more participants in a large data set to improve the models' applicability as advocated by Chang et al [157] and Chen et al [158]. For example, it has been demonstrated that the model trained on the largest training set yielded the best performance in external validation [11,159]. Zarkogianni et al [96] have introduced larger data sets corresponding to patients with different ethnicities and races to other cohorts of patients and consequently extended the applicability of the model. Therefore, it is advocated to conduct multicohort, multicountry, or multiracial studies with adequate cross-validation. In the real world, however, it is difficult to obtain large data with high quality, which will cost lots of manpower and money and take a long time. In response to this point, it may be a more appropriate solution to fully use open-source databases, such as the National Health Insurance Service-National Sample Cohort, NHIS-HEALS, Qresearch, Early Identification of Subclinical Atherosclerosis by Noninvasive Imaging Research, and so on [76,92].

The Lack of Interpretability in ML Algorithms Poses a Significant Challenge

Prediction models play a crucial role in helping doctors make accurate decisions in clinical practice. However, the "black box" effect in many ML algorithms, such as ANNs and RF, creates a challenge for doctors to interpret mechanisms correctly [160]. This limits the clinical utility of the models, as they become complex for doctors to understand the impact of risk factors on the prevalence of CVDs. In contrast to the fixed regression coefficient in traditional models, the complexity of ML algorithms hinders their interpretability, posing a common problem for doctors [95,160].

To overcome or reduce the negative impact of the "black box," Han et al [61] suggested that information gain ranking methods could be used, while Weng et al [71] advocated for the use of data visualization to facilitate interpretation. Besides, in the ECG methods for the prompt identification of coronary events study with 1244 participants, Al-Zaiti et al [15] revealed that by incorporating feature selection and annotation based on clinical knowledge, LR could achieve comparable performance to complex and expensive nonlinear algorithms such as ANN and GBM. This approach also resolved the "black box" issue of ANN and GBM, improving clinical utility by identifying predominant features contributing to ACS. Furthermore, methods such as discriminant analysis, Naive Bayes, logistic and Cox regression, and classification provide relationships between predictors and outcomes, enabling risk scores to be calculated for each predictor in clinical practice [160]. In

addition, there has recently been an increasing focus on explaining AI algorithms, with Xuan et al [97] developing a visual analytics tool for comparing convolutional neural networks (CNNs) to support in-depth inspection and comparative studies of CNN models. Wang et al [98] have reported a CNN explainer, which can show the CNN's structure and provides on-demand dynamic visual explanations. Furthermore, it is exciting that explainable AI techniques have been applied to CVD prediction models and are reasonably expected to promote their application [161,162].

AI-Ethics Pitfalls Also Should Be Considered During the Study Design

Informed model selection is a pivotal decision-making step in AI, particularly when dealing with CVD risk prediction. It is imperative to initiate the process with a rigorous assessment of bias and fairness [144]. While it is acknowledged that most AI algorithms may inherently carry biases, it is equally important to benchmark these biases against those prevalent in the existing systems. AI's propensity to introduce bias often arises from making unequal errors across diverse demographic groups. The degree to which key demographic variables, including sex, age, and ethnicity, are adequately represented in the data set and incorporated during algorithm development significantly impacts the predictive accuracy across subgroups. Consequently, when these AI-derived predictions inform individual decisions, they can either perpetuate or intensify the existing disparities [163]. This issue is exacerbated by data that may not authentically mirror the entire target population, underscoring the ethical imperative of ensuring fairness at every stage of the project life cycle [164]. In the domain of CVD risk prediction, these ethical considerations loom large. For example, Kim et al [99] and Weng et al [71] constructed AI models for CVD risk prediction in the Korea National Health and Nutrition Examination Survey-VI cohort (4244 participants) and the CPRD cohort (378,256 participants), respectively. However, they failed to perform sensitivity analysis for race and gender subgroups, leading to gender and ethnic bias. These issues will significantly impact the application of the models.

There is an undeniable global demand for the development of AI systems that can be relied upon for their trustworthiness [165-168]. This demand is underscored by real-world examples, such as ProPublica's investigation into the Correctional Offender Management Profiling for Alternative Sanctions software, which serves as a stark illustration of how algorithmic performance can exhibit disparities rooted in racial factors. Notably, the field of fairness evaluation has made significant methodological advancements that directly facilitate such analyses [169]. It is imperative for AI developers and health care professionals to actively engage with these tools. Researchers can effectively showcase bias in critical subgroups, such as minority ethnic communities or distinct age groups, through explicit presentation of these findings. This approach not only enhances transparency but also empowers users of the AI algorithm with the knowledge of its strengths and weaknesses in different demographic contexts. This consideration holds particular relevance in the domain of AI-CVD risk prediction. Segar et al [100] included 19,080 participants from Atherosclerosis Risk in Communities, Dallas Heart Study, Jackson Heart Study, and MESA cohort to

build 6 AI models for predicting incident HF. They performed sensitivity analysis for the subgroup of race and sex, and all models performed well.

Conclusions

In conclusion, we highlight that the effectiveness and application of ML models for predicting CVD risk are heavily reliant on data quality, data set characteristics, model design and statistical methods, and clinical implications. To address these challenges, we propose practical solutions, including gathering objective data, improving training, repeating measurements, increasing sample size, preventing overfitting using statistical methods, using specific AI algorithms to address targeted issues, standardizing outcomes and evaluation criteria, and enhancing fairness and replicability. It may provide a helpful reference and assistance to researchers, algorithm developers, policy makers, and clinical practitioners in the ML field of CVD prediction.

Authors' Contributions

GWZ and XZ contributed to the overarching research goals of the paper as well as the manuscript review and editing. YQC, YC, MG, TCJ, and LYT performed the literature search together. They also collected important background information and wrote the manuscript. HJL and ZWZ helped with manuscript preparation. DXG and WH performed supervision and project administration. All authors have read and approved the content of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies of artificial intelligence or machine learning assessment guidelines or tools.
[DOCX File , 15 KB-Multimedia Appendix 1]

Multimedia Appendix 2

The evaluation content of assessment guidelines or tools in the field of medical artificial intelligence or machine learning research.
[DOCX File , 270 KB-Multimedia Appendix 2]

Multimedia Appendix 3

The framework for pitfalls in developing machine learning models for cardiovascular disease prediction.
[DOCX File , 149 KB-Multimedia Appendix 3]

References

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol. Dec 22, 2020;76(25):2982-3021. [FREE Full text] [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
2. Correction to: 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation. Sep 10, 2019;140(11):e649-e650. [doi: 10.1161/CIR.0000000000000725] [Medline: 31498691]
3. Pelliccia A, Sharma S, Gati S, Bäck M, Börjesson M, Caselli S, et al. 2020 ESC guidelines on sports cardiology and exercise in patients with cardiovascular disease. Eur Heart J. Jan 01, 2021;42(1):17-96. [FREE Full text] [doi: 10.1093/eurheartj/ehaa605] [Medline: 32860412]
4. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. Mar 15, 2014;383(9921):999-1008. [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
5. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ. Jul 21, 2007;335(7611):136. [FREE Full text] [doi: 10.1136/bmj.39261.471806.55] [Medline: 17615182]
6. Goff DCJ, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RBS, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force

XSL•FO

RenderX

on Practice Guidelines. J Am Coll Cardiol. Jul 01, 2014;63(25 Pt B):2935-2959. [FREE Full text] [doi: 10.1016/j.jacc.2013.11.005] [Medline: 24239921]

7. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, et al. 10-year risk equations for incident heart failure in the general population. J Am Coll Cardiol. May 21, 2019;73(19):2388-2397. [FREE Full text] [doi: 10.1016/j.jacc.2019.02.057] [Medline: 31097157]

8. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. Dec 31, 2019;188(12):2222-2239. [doi: 10.1093/aje/kwz189] [Medline: 31509183]

9. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. Mar 19, 2019;19(1):64. [FREE Full text] [doi: 10.1186/s12874-019-0681-4] [Medline: 30890124]

10. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. Sep 29, 2016;375(13):1216-1219. [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]

11. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, et al. A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. J Biomed Inform. Aug 2018;84:11-16. [FREE Full text] [doi: 10.1016/j.jbi.2018.06.011] [Medline: 29908902]

12. Du Z, Yang Y, Zheng J, Li Q, Lin D, Li Y, et al. Accurate prediction of coronary heart disease for patients with hypertension from electronic health records with big data and machine-learning methods: model development and performance evaluation. JMIR Med Inform. Jul 06, 2020;8(7):e17257. [FREE Full text] [doi: 10.2196/17257] [Medline: 32628616]

13. Han L, Askari M, Altman RB, Schmitt SK, Fan J, Bentley JP, et al. Atrial fibrillation burden signature and near-term prediction of stroke: a machine learning analysis. Circ Cardiovasc Qual Outcomes. Oct 2019;12(10):e005595. [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.118.005595] [Medline: 31610712]

14. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. The Lancet. Sep 2019;394(10201):861-867. [doi: 10.1016/s0140-6736(19)31721-0]

15. Al-Zaiti S, Besomi L, Bouzid Z, Faramand Z, Frisch S, Martin-Gill C, et al. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. Nat Commun. Aug 07, 2020;11(1):3966. [FREE Full text] [doi: 10.1038/s41467-020-17804-2] [Medline: 32769990]

16. Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. J Am Coll Cardiol. Apr 17, 2018;71(15):1650-1660. [FREE Full text] [doi: 10.1016/j.jacc.2018.02.024] [Medline: 29650121]

17. Kagiyama N, Piccirilli M, Yanamala N, Shrestha S, Farjo PD, Casaclang-Verzosa G, et al. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. J Am Coll Cardiol. Aug 25, 2020;76(8):930-941. [FREE Full text] [doi: 10.1016/j.jacc.2020.06.061] [Medline: 32819467]

18. Bouzid Z, Faramand Z, Gregg RE, Frisch SO, Martin-Gill C, Saba S, et al. In search of an optimal subset of ECG features to augment the diagnosis of acute coronary syndrome at the emergency department. J Am Heart Assoc. Feb 02, 2021;10(3):e017871. [FREE Full text] [doi: 10.1161/JAHA.120.017871] [Medline: 33459029]

19. Dogan MV, Beach SR, Simons RL, Lendasse A, Penaluna B, Philibert RA. Blood-based biomarkers for predicting the risk for five-year incident coronary heart disease in the Framingham Heart Study via machine learning. Genes (Basel). Dec 18, 2018;9(12):641. [FREE Full text] [doi: 10.3390/genes9120641] [Medline: 30567402]

20. Khanna NN, Maindarkar M, Saxena A, Ahluwalia P, Paul S, Srivastava SK, et al. Cardiovascular/stroke risk assessment in patients with erectile dysfunction-a role of carotid wall arterial imaging and plaque tissue characterization using artificial intelligence paradigm: a narrative review. Diagnostics (Basel). May 17, 2022;12(5):1249. [FREE Full text] [doi: 10.3390/diagnostics12051249] [Medline: 35626404]

21. Chao H, Shan H, Homayounieh F, Singh R, Khera RD, Guo H, et al. Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. Nat Commun. May 20, 2021;12(1):2963. [FREE Full text] [doi: 10.1038/s41467-021-23235-4] [Medline: 34017001]

22. Cai Y, Cai YQ, Tang LY, Wang YH, Gong M, Jing TC, et al. Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review. BMC Med. Feb 05, 2024;22(1):56. [FREE Full text] [doi: 10.1186/s12916-024-03273-7] [Medline: 38317226]

23. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. Oct 20, 2021;375:n2281. [FREE Full text] [doi: 10.1136/bmj.n2281] [Medline: 34670780]

24. Li B, Feridooni T, Cuen-Ojeda C, Kishibe T, de Mestral C, Mamdani M, et al. Machine learning in vascular surgery: a systematic review and critical appraisal. NPJ Digit Med. Jan 19, 2022;5(1):7. [FREE Full text] [doi: 10.1038/s41746-021-00552-y] [Medline: 35046493]

25. Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. Jan 01, 2019;170(1):51. [doi: 10.7326/m18-1376]

26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. Jan 06, 2015;162(1):55-63. [FREE Full text] [doi: 10.7326/M14-0697] [Medline: 25560714]

27. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. Apr 01, 2020;369:m1312. [FREE Full text] [doi: 10.1136/bmj.m1312] [Medline: 32238345]

28. Suri JS, Bhagawati M, Paul S, Protogeron A, Sfikakis PP, Kitas GD, et al. Understanding the bias in machine learning systems for cardiovascular disease risk assessment: the first of its kind review. Comput Biol Med. Mar 2022;142:105204. [doi: 10.1016/j.compbiomed.2021.105204] [Medline: 35033879]

29. Russo S, Bonassi S. Prospects and pitfalls of machine learning in nutritional epidemiology. Nutrients. Apr 20, 2022;14(9):1705. [FREE Full text] [doi: 10.3390/nu14091705] [Medline: 35565673]

30. Chiarito M, Luceri L, Oliva A, Stefanini G, Condorelli G. Artificial intelligence and cardiovascular risk prediction: all that glitters is not gold. Eur Cardiol. Feb 20, 2022;17:e29. [FREE Full text] [doi: 10.15420/ecr.2022.11] [Medline: 36845218]

31. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. Mar 15, 2016;3:160018. [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]

32. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res. Aug 31, 2023;25:e48763. [FREE Full text] [doi: 10.2196/48763] [Medline: 37651179]

33. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. Dec 16, 2016;18(12):e323. [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]

34. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. Sep 2020;26(9):1320-1324. [FREE Full text] [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]

35. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. BMJ. Sep 09, 2020;370:m3164. [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]

36. Tanguay W, Acar P, Fine B, Abdolell M, Gong B, Cadrin-Chênevert A, et al. Assessment of radiology artificial intelligence software: a validation and evaluation framework. Can Assoc Radiol J. May 2023;74(2):326-333. [FREE Full text] [doi: 10.1177/08465371221135760] [Medline: 36341574]

37. Lennerz JK, Salgado R, Kim GE, Sirintrapun SJ, Thierauf JC, Singh A, et al. Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML. Clin Chem Lab Med. Jan 25, 2023;61(4):544-557. [FREE Full text] [doi: 10.1515/cclm-2022-1151] [Medline: 36696602]

38. Cerdá-Alberich L, Solana J, Mallol P, Ribas G, García-Junco M, Alberich-Bayarri A, et al. MAIC-10 brief quality checklist for publications using artificial intelligence and medical images. Insights Imaging. Jan 16, 2023;14(1):11. [FREE Full text] [doi: 10.1186/s13244-022-01355-9] [Medline: 36645542]

39. Sujan M, Smith-Frazer C, Malamateniou C, Connor J, Gardner A, Unsworth H, et al. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440. BMJ Health Care Inform. Jun 2023;30(1):e100749. [FREE Full text] [doi: 10.1136/bmjhci-2023-100749] [Medline: 37364922]

40. Mylrea M, Robinson N. Artificial intelligence (AI) trust framework and maturity model: applying an entropy lens to improve security, privacy, and ethical AI. Entropy (Basel). Oct 09, 2023;25(10):1429. [FREE Full text] [doi: 10.3390/e25101429] [Medline: 37895550]

41. Kwong JC, Khondker A, Lajkosz K, McDermott MB, Frigola XB, McCradden MD, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. JAMA Netw Open. Sep 05, 2023;6(9):e2335377. [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.35377] [Medline: 37747733]

42. Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, et al. CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. May 04, 2023;14(1):75. [FREE Full text] [doi: 10.1186/s13244-023-01415-8] [Medline: 37142815]

43. van Smeden M, Heinze G, Van Calster B, Asselbergs FW, Vardas PE, Bruining N, et al. Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. Eur Heart J. Aug 14, 2022;43(31):2921-2930. [FREE Full text] [doi: 10.1093/eurheartj/ehac238] [Medline: 35639667]

44. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, et al. Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR Derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group. JAMA Dermatol. Jan 01, 2022;158(1):90-96. [FREE Full text] [doi: 10.1001/jamadermatol.2021.4915] [Medline: 34851366]

45. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ. May 18, 2022;377:e070904. [FREE Full text] [doi: 10.1136/bmj-2022-070904] [Medline: 35584845]

46. Jha AK, Bradshaw TJ, Buvat I, Hatt M, Kc P, Liu C, et al. Nuclear medicine and artificial intelligence: best practices for evaluation (the RELAINCE guidelines). J Nucl Med. Sep 26, 2022;63(9):1288-1299. [FREE Full text] [doi: 10.2967/jnumed.121.263239] [Medline: 35618476]

47. Banerjee A, Chen S, Fatemifar G, Zeina M, Lumbers RT, Mielke J, et al. Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: systematic review of validity and clinical utility. BMC Med. Apr 06, 2021;19(1):85. [FREE Full text] [doi: 10.1186/s12916-021-01940-7] [Medline: 33820530]

48. Walsh I, Fishman D, Garcia-Gasulla D, Titma T, Pollastri G, ELIXIR Machine Learning Focus Group, et al. DOME: recommendations for supervised machine learning validation in biology. Nat Methods. Oct 27, 2021;18(10):1122-1127. [FREE Full text] [doi: 10.1038/s41592-021-01205-4] [Medline: 34316068]

49. Olczak J, Pavlopoulos J, Prijs J, Ijpma FF, Doornberg JN, Lundström C, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop. Oct 14, 2021;92(5):513-525. [FREE Full text] [doi: 10.1080/17453674.2021.1918389] [Medline: 33988081]

50. Matschinske J, Alcaraz N, Benis A, Golebiewski M, Grimm DG, Heumos L, et al. The AIMe registry for artificial intelligence in biomedical research. Nat Methods. Oct 25, 2021;18(10):1128-1131. [FREE Full text] [doi: 10.1038/s41592-021-01241-0] [Medline: 34433960]

51. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, et al. Artificial intelligence in dental research: checklist for authors, reviewers, readers. J Dent. Apr 2021;107:103610. [doi: 10.1016/j.jdent.2021.103610] [Medline: 33631303]

52. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health Care Inform. Feb 05, 2021;28(1):e100251. [FREE Full text] [doi: 10.1136/bmjhci-2020-100251] [Medline: 33547086]

53. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Lancet Digit Health. Oct 2020;2(10):e537-e548. [FREE Full text] [doi: 10.1016/S2589-7500(20)30218-1] [Medline: 33328048]

54. Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist: reviewed by the American College of Cardiology Healthcare Innovation Council. JACC Cardiovasc Imaging. Sep 2020;13(9):2017-2035. [FREE Full text] [doi: 10.1016/j.jcmg.2020.07.015] [Medline: 32912474]

55. Cruz Rivera S, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group, SPIRIT-AICONSORT-AI Steering Group, et al. SPIRIT-AICONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. Sep 2020;26(9):1351-1363. [FREE Full text] [doi: 10.1038/s41591-020-1037-7] [Medline: 32908284]

56. Kakarmath S, Esteva A, Arnaout R, Harvey H, Kumar S, Muse E, et al. Best practices for authors of healthcare-related artificial intelligence manuscripts. NPJ Digit Med. Oct 16, 2020;3(1):134. [FREE Full text] [doi: 10.1038/s41746-020-00336-w] [Medline: 33083569]

57. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. Circ Cardiovasc Qual Outcomes. Oct 2020;13(10). [doi: 10.1161/circoutcomes.120.006556]

58. Lambin P, Leijenaar RT, Deist TM, Peerlings J, de Jong EE, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. Dec 4, 2017;14(12):749-762. [doi: 10.1038/nrclinonc.2017.141] [Medline: 28975929]

59. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. Jan 06, 2015;162(1):W1-73. [FREE Full text] [doi: 10.7326/M14-0698] [Medline: 25560730]

60. van Royen FS, Asselbergs FW, Alfonso F, Vardas P, van Smeden M. Five critical quality criteria for artificial intelligence-based prediction models. Eur Heart J. Dec 07, 2023;44(46):4831-4834. [FREE Full text] [doi: 10.1093/eurheartj/ehad727] [Medline: 37897346]

61. Han D, Kolli KK, Gransar H, Lee JH, Choi SY, Chun EJ, et al. Machine learning based risk prediction model for asymptomatic individuals who underwent coronary artery calcium score: comparison with traditional risk prediction approaches. J Cardiovasc Comput Tomogr. 2020;14(2):168-176. [doi: 10.1016/j.jcct.2019.09.005] [Medline: 31570323]

62. Ward A, Sarraju A, Chung S, Li J, Harrington R, Heidenreich P, et al. Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. NPJ Digit Med. Sep 23, 2020;3(1):125. [FREE Full text] [doi: 10.1038/s41746-020-00331-1] [Medline: 33043149]

63. Rodriguez F, Chung S, Blum MR, Coulet A, Basu S, Palaniappan LP. Atherosclerotic cardiovascular disease risk prediction in disaggregated Asian and Hispanic subgroups using electronic health records. J Am Heart Assoc. Jul 16, 2019;8(14):e011874. [FREE Full text] [doi: 10.1161/JAHA.118.011874] [Medline: 31291803]

64. Teoh D. Towards stroke prediction using electronic health records. BMC Med Inform Decis Mak. Dec 4, 2018;18(1):127. [FREE Full text] [doi: 10.1186/s12911-018-0702-y] [Medline: 30509279]

65. Nakanishi R, Slomka PJ, Rios R, Betancur J, Blaha MJ, Nasir K, et al. Machine learning adds to clinical and CAC assessments in predicting 10-year CHD and CVD deaths. JACC Cardiovasc Imaging. Mar 2021;14(3):615-625. [FREE Full text] [doi: 10.1016/j.jcmg.2020.08.024] [Medline: 33129741]

66. Benjamins JW, van Leeuwen K, Hofstra L, Rienstra M, Appelman Y, Nijhof W, et al. Enhancing cardiovascular artificial intelligence (AI) research in the Netherlands: CVON-AI consortium. Neth Heart J. Sep 20, 2019;27(9):414-425. [FREE Full text] [doi: 10.1007/s12471-019-1281-y] [Medline: 31111459]

67. Liu L, Tan S, Li Y, Luo J, Zhang W, Li S. An early aortic dissection screening model and applied research based on ensemble learning. Ann Transl Med. Dec 2020;8(23):1578. [FREE Full text] [doi: 10.21037/atm-20-1475] [Medline: 33437777]

68. Tsoi KK, Chan NB, Yiu KK, Poon SK, Lin B, Ho K. Machine learning clustering for blood pressure variability applied to systolic blood pressure intervention trial (SPRINT) and the Hong Kong community cohort. Hypertension. Aug 2020;76(2):569-576. [doi: 10.1161/hypertensionaha.119.14213]

69. Sajeev S, Champion S, Beleigoli A, Chew D, Reed RL, Magliano DJ, et al. Predicting Australian adults at high risk of cardiovascular disease mortality using standard risk factors and machine learning. Int J Environ Res Public Health. Mar 19, 2021;18(6):3187. [FREE Full text] [doi: 10.3390/ijerph18063187] [Medline: 33808743]

70. Wang H, Li Y, Ning H, Wilkins J, Lloyd-Jones D, Luo Y. Using machine learning to integrate socio-behavioral factors in predicting cardiovascular-related mortality risk. Stud Health Technol Inform. Aug 21, 2019;264:433-437. [doi: 10.3233/SHTI190258] [Medline: 31437960]

71. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. Apr 4, 2017;12(4):e0174944. [FREE Full text] [doi: 10.1371/journal.pone.0174944] [Medline: 28376093]

72. Li Y, Sperrin M, Ashcroft DM, van Staa TP. Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar. BMJ. Nov 04, 2020;371:m3919. [FREE Full text] [doi: 10.1136/bmj.m3919] [Medline: 33148619]

73. Alaa AM, Bolton T, Di Angelantonio E, Rudd JH, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. PLoS One. May 15, 2019;14(5):e0213653. [FREE Full text] [doi: 10.1371/journal.pone.0213653] [Medline: 31091238]

74. Sung JM, Cho IJ, Sung D, Kim S, Kim HC, Chae MH, et al. Development and verification of prediction models for preventing cardiovascular diseases. PLoS One. Sep 19, 2019;14(9):e0222809. [FREE Full text] [doi: 10.1371/journal.pone.0222809] [Medline: 31536581]

75. Ayala Solares JR, Canoy D, Raimondi FE, Zhu Y, Hassaine A, Salimi-Khorshidi G, et al. Long-term exposure to elevated systolic blood pressure in predicting incident cardiovascular disease: evidence from large-scale routine electronic health records. J Am Heart Assoc. Jun 18, 2019;8(12):e012129. [FREE Full text] [doi: 10.1161/JAHA.119.012129] [Medline: 31164039]

76. Cho IJ, Sung JM, Kim HC, Lee SE, Chae MH, Kavousi M, et al. Development and external validation of a deep learning algorithm for prognostication of cardiovascular outcomes. Korean Circ J. Jan 2020;50(1):72-84. [FREE Full text] [doi: 10.4070/kcj.2019.0105] [Medline: 31456363]

77. Jdanov DA, Deev AD, Jasilionis D, Shalnova SA, Shkolnikova MA, Shkolnikov VM. Recalibration of the SCORE risk chart for the Russian population. Eur J Epidemiol. Sep 2, 2014;29(9):621-628. [FREE Full text] [doi: 10.1007/s10654-014-9947-7] [Medline: 25179794]

78. Ponomartseva DA, Derevitskii IV, Kovalchuk SV, Babenko AY. Prediction model for thyrotoxic atrial fibrillation: a retrospective study. BMC Endocr Disord. Jul 11, 2021;21(1):150. [FREE Full text] [doi: 10.1186/s12902-021-00809-3] [Medline: 34246271]

79. Chan KL, Leng X, Zhang W, Dong W, Qiu Q, Yang J, et al. Early identification of high-risk TIA or minor stroke using artificial neural network. Front Neurol. Mar 1, 2019;10:171. [FREE Full text] [doi: 10.3389/fneur.2019.00171] [Medline: 30881336]

80. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. J Am Heart Assoc. Nov 20, 2018;7(22). [doi: 10.1161/jaha.118.009476]

81. Chua W, Purmah Y, Cardoso VR, Gkoutos GV, Tull SP, Neculau G, et al. Data-driven discovery and validation of circulating blood-based biomarkers associated with prevalent atrial fibrillation. Eur Heart J. Apr 21, 2019;40(16):1268-1276. [FREE Full text] [doi: 10.1093/eurheartj/ehy815] [Medline: 30615112]

82. Goldstein BA, Chang TI, Mitani AA, Assimes TL, Winkelmayer WC. Near-term prediction of sudden cardiac death in older hemodialysis patients using electronic health records. Clin J Am Soc Nephrol. Jan 2014;9(1):82-91. [FREE Full text] [doi: 10.2215/CJN.03050313] [Medline: 24178968]

83. Angeli F, Reboldi G, Verdecchia P. Hypertension and the J-curve phenomenon: implications for tight blood pressure control. Hypertens Res. Feb 15, 2013;36(2):109-111. [doi: 10.1038/hr.2012.165] [Medline: 23154585]

84. Lip S, Tan LE, Jeemon P, McCallum L, Dominiczak AF, Padmanabhan S. Diastolic blood pressure J-curve phenomenon in a tertiary-care hypertension clinic. Hypertension. Oct 2019;74(4):767-775. [doi: 10.1161/hypertensionaha.119.12787]

85. Dutta A, Batabyal T, Basu M, Acton ST. An efficient convolutional neural network for coronary heart disease prediction. Expert Syst Appl. Nov 2020;159:113408. [doi: 10.1016/j.eswa.2020.113408]

86. Zhuang XD, Tian T, Liao LZ, Dong YH, Zhou HJ, Zhang SZ, et al. Deep phenotyping and prediction of long-term cardiovascular disease: optimized by machine learning. Can J Cardiol. Jun 2022;38(6):774-782. [doi: 10.1016/j.cjca.2022.02.008] [Medline: 35157988]

87. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ Res. Oct 13, 2017;121(9):1092-1101. [doi: 10.1161/CIRCRESAHA.117.311312] [Medline: 28794054]

88. Pujadas ER, Raisi-Estabragh Z, Szabo L, McCracken C, Morcillo CI, Campello VM, et al. Prediction of incident cardiovascular events using machine learning and CMR radiomics. Eur Radiol. May 2023;33(5):3488-3500. [FREE Full text] [doi: 10.1007/s00330-022-09323-z] [Medline: 36512045]

89. Lou YS, Lin CS, Fang WH, Lee CC, Lin C. Extensive deep learning model to enhance electrocardiogram application via latent cardiovascular feature extraction from identity identification. Comput Methods Programs Biomed. Apr 2023;231:107359. [doi: 10.1016/j.cmpb.2023.107359] [Medline: 36738606]

90. You J, Guo Y, Kang JJ, Wang HF, Yang M, Feng JF, et al. Development of machine learning-based models to predict 10-year risk of cardiovascular disease: a prospective cohort study. Stroke Vasc Neurol. Dec 29, 2023;8(6):475-485. [FREE Full text] [doi: 10.1136/svn-2023-002332] [Medline: 37105576]

91. Chen X, Liu S, Chu J, Hu W, Sun N, Shen Y. Joint effect of elevated-c-reactive protein level and hypertension on new-onset stroke: a nationwide prospective cohort study of CHARLS. Front Public Health. Oct 3, 2022;10:919506. [FREE Full text] [doi: 10.3389/fpubh.2022.919506] [Medline: 36262245]

92. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, et al. Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. Cardiovasc Res. Dec 01, 2020;116(14):2216-2225. [FREE Full text] [doi: 10.1093/cvr/cvz321] [Medline: 31853543]

93. Voss R, Cullen P, Schulte H, Assmann G. Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Münster Study (PROCAM) using neural networks. Int J Epidemiol. Dec 2002;31(6):1253-62; discussion 1262-64. [doi: 10.1093/ije/31.6.1253] [Medline: 12540731]

94. Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD, et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA. Apr 28, 2010;303(16):1610-1616. [FREE Full text] [doi: 10.1001/jama.2010.461] [Medline: 20424251]

95. Unnikrishnan P, Kumar DK, Poosapadi Arjunan S, Kumar H, Mitchell P, Kawasaki R. Development of health parameter model for risk prediction of CVD using SVM. Comput Math Methods Med. 2016;2016:3016245. [FREE Full text] [doi: 10.1155/2016/3016245] [Medline: 27594895]

96. Zarkogianni K, Athanasiou M, Thanopoulou AC, Nikita KS. Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication. IEEE J Biomed Health Inform. Sep 2018;22(5):1637-1647. [doi: 10.1109/jbhi.2017.2765639]

97. Xuan X, Zhang X, Kwon OH, Ma KL. VAC-CNN: a visual analytics system for comparative studies of deep convolutional neural networks. IEEE Trans Visual Comput Graphics. Jun 2022;28(6):2326-2337. [doi: 10.1109/tvcg.2022.3165347]

98. Wang ZJ, Turko R, Shaikh O, Park H, Das N, Hohman F, et al. CNN explainer: learning convolutional neural networks with interactive visualization. IEEE Trans Visual Comput Graphics. Feb 2021;27(2):1396-1406. [doi: 10.1109/tvcg.2020.3030418]

99. Kim J, Kang U, Lee Y. Statistics and deep belief network-based cardiovascular risk prediction. Healthc Inform Res. Jul 2017;23(3):169-175. [FREE Full text] [doi: 10.4258/hir.2017.23.3.169] [Medline: 28875051]

100. Segar MW, Jaeger BC, Patel KV, Nambi V, Ndumele CE, Correa A, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. Circulation. Jun 15, 2021;143(24):2370-2383. [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.120.053134] [Medline: 33845593]

101. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. Obstet Gynecol. 2010;116(5):1018-1019. [doi: 10.1097/aog.0b013e3181f98300]

102. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ. Dec 09, 2010;341(dec09 1):c6624. [FREE Full text] [doi: 10.1136/bmj.c6624] [Medline: 21148212]

103. Manuel DG, Tuna M, Bennett C, Hennessy D, Rosella L, Sanmartin C, et al. Development and validation of a cardiovascular disease risk-prediction model using population health surveys: the Cardiovascular Disease Population Risk Tool (CVDPoRT). CMAJ. Jul 23, 2018;190(29):E871-E882. [FREE Full text] [doi: 10.1503/cmaj.170914] [Medline: 30037888]

104. Inoue K, Seeman TE, Horwich T, Budoff MJ, Watson KE. Heterogeneity in the association between the presence of coronary artery calcium and cardiovascular events: a machine-learning approach in the MESA study. Circulation. Jan 10, 2023;147(2):132-141. [doi: 10.1161/circulationaha.122.062626]

105. Li JX, Li L, Zhong X, Fan SJ, Cen T, Wang J, et al. Machine learning identifies prominent factors associated with cardiovascular disease: findings from two million adults in the Kashgar Prospective Cohort Study (KPCS). Glob Health Res Policy. Dec 06, 2022;7(1):48. [FREE Full text] [doi: 10.1186/s41256-022-00282-y] [Medline: 36474302]

106. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. Jan 01, 2019;170(1):W1-33. [FREE Full text] [doi: 10.7326/M18-1377] [Medline: 30596876]

107. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. Jan 07, 2015;350:g7594. [FREE Full text] [doi: 10.1136/bmj.g7594] [Medline: 25569120]

108. Johansson JK, Puukka PJ, Niiranen TJ, Varis J, Peltonen M, Salomaa V, et al. Health 2000 score - development and validation of a novel cardiovascular risk score. Ann Med. Sep 2016;48(6):403-409. [FREE Full text] [doi: 10.1080/07853890.2016.1183257] [Medline: 27187608]

109. Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation. Nov 21, 2006;114(21):2217-2225. [doi: 10.1161/circulationaha.105.607499]

110. Carresi C, Scicchitano M, Scarano F, Macrì R, Bosco F, Nucera S, et al. The potential properties of natural compounds in cardiac stem cell activation: their role in myocardial regeneration. Nutrients. Jan 19, 2021;13(1):275. [FREE Full text] [doi: 10.3390/nu13010275] [Medline: 33477916]

111. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ. May 23, 2017;357:j2099. [FREE Full text] [doi: 10.1136/bmj.j2099] [Medline: 28536104]

112. Dimopoulos AC, Nikolaidou M, Caballero FF, Engchuan W, Sanchez-Niubo A, Arndt H, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. BMC Med Res Methodol. Dec 29, 2018;18(1):179. [FREE Full text] [doi: 10.1186/s12874-018-0644-1] [Medline: 30594138]

113. Saeed A, Nambi V, Sun W, Virani SS, Taffet GE, Deswal A, et al. Short-term global cardiovascular disease risk prediction in older adults. J Am Coll Cardiol. Jun 05, 2018;71(22):2527-2536. [FREE Full text] [doi: 10.1016/j.jacc.2018.02.050] [Medline: 29535064]

114. Xu J, Li L, Huang S, Song H, Gao J, Ni H, et al. Impact of visit-to-visit fasting plasma glucose variability on the development of diabetes: the mediation by insulin resistance. J Diabetes. Mar 16, 2022;14(3):205-215. [FREE Full text] [doi: 10.1111/1753-0407.13253] [Medline: 35170854]

115. Hukportie DN, Li FR, Zhou R, Zheng JZ, Wu XX, Zou MC, et al. Lipid variability and risk of microvascular complications in Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: a post hoc analysis. J Diabetes. Jun 06, 2022;14(6):365-376. [FREE Full text] [doi: 10.1111/1753-0407.13273] [Medline: 35668633]

116. Stergiou GS, Palatini P, Parati G, O'Brien E, Januszewicz A, Lurbe E, et al. 2021 European Society of Hypertension practice guidelines for office and out-of-office blood pressure measurement. J Hypertens. Jul 01, 2021;39(7):1293-1302. [doi: 10.1097/HJH.0000000000002843] [Medline: 33710173]

117. Paige E, Barrett J, Pennells L, Sweeting M, Willeit P, Di Angelantonio E, et al. Use of repeated blood pressure and cholesterol measurements to improve cardiovascular disease risk prediction: an individual-participant-data meta-analysis. Am J Epidemiol. Oct 15, 2017;186(8):899-907. [FREE Full text] [doi: 10.1093/aje/kwx149] [Medline: 28549073]

118. WHO CVD Risk Chart Working Group. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Health. Oct 2019;7(10):e1332-e1345. [FREE Full text] [doi: 10.1016/S2214-109X(19)30318-3] [Medline: 31488387]

119. Hajifathalian K, Ueda P, Lu Y, Woodward M, Ahmadvand A, Aguilar-Salinas CA, et al. A novel risk score to predict cardiovascular disease risk in national populations (Globorisk): a pooled analysis of prospective cohorts and health examination surveys. Lancet Diabetes Endocrinol. May 2015;3(5):339-355. [FREE Full text] [doi: 10.1016/S2213-8587(15)00081-9] [Medline: 25819778]

120. McClelland RL, Jorgensen NW, Budoff M, Blaha MJ, Post WS, Kronmal RA, et al. 10-Year coronary heart disease risk prediction using coronary artery calcium and traditional risk factors: derivation in the MESA (multi-ethnic study of atherosclerosis) with validation in the HNR (Heinz Nixdorf Recall) study and the DHS (Dallas Heart Study). J Am Coll Cardiol. Oct 13, 2015;66(15):1643-1653. [FREE Full text] [doi: 10.1016/j.jacc.2015.08.035] [Medline: 26449133]

121. Parikh NI, Jeppson RP, Berger JS, Eaton CB, Kroenke CH, LeBlanc ES, et al. Reproductive risk factors and coronary heart disease in the women's health initiative observational study. Circulation. May 31, 2016;133(22):2149-2158. [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017854] [Medline: 27143682]

122. Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. Neuroimage. Sep 2018;178:622-637. [doi: 10.1016/j.neuroimage.2018.06.001] [Medline: 29870817]

123. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. Can Assoc Radiol J. Nov 29, 2019;70(4):344-353. [FREE Full text] [doi: 10.1016/j.carj.2019.06.002] [Medline: 31522841]

124. Ma J, Fong SH, Luo Y, Bakkenist CJ, Shen JP, Mourragui S, et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. Nat Cancer. Feb 25, 2021;2(2):233-244. [FREE Full text] [doi: 10.1038/s43018-020-00169-2] [Medline: 34223192]

125. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. Jul 09, 2021;11(7):e048008. [FREE Full text] [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]

126. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. The Lancet. Apr 2019;393(10181):1577-1579. [doi: 10.1016/s0140-6736(19)30037-6]

127. Andaur Navarro CL, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. BMC Med Res Methodol. Jan 13, 2022;22(1):12. [FREE Full text] [doi: 10.1186/s12874-021-01469-6] [Medline: 35026997]

128. Backholer K, Hirakawa Y, Tonkin A, Giles G, Magliano DJ, Colagiuri S, et al. Development of an Australian cardiovascular disease mortality risk score using multiple imputation and recalibration from national statistics. BMC Cardiovasc Disord. Jan 06, 2017;17(1):17. [FREE Full text] [doi: 10.1186/s12872-016-0462-5] [Medline: 28061760]

129. Dritsas E, Trigka M. Efficient data-driven machine learning models for cardiovascular diseases risk prediction. Sensors (Basel). Jan 19, 2023;23(3):1161. [FREE Full text] [doi: 10.3390/s23031161] [Medline: 36772201]

130. Mandair D, Tiwari P, Simon S, Colborn KL, Rosenberg MA. Prediction of incident myocardial infarction using machine learning applied to harmonized electronic health record data. BMC Med Inform Decis Mak. Oct 02, 2020;20(1):252. [FREE Full text] [doi: 10.1186/s12911-020-01268-x] [Medline: 33008368]

131. Subramani S, Varshney N, Anand MV, Soudagar ME, Al-Keridis LA, Upadhyay TK, et al. Cardiovascular diseases prediction by machine learning incorporation with deep learning. Front Med (Lausanne). Apr 17, 2023;10:1150933. [doi: 10.3389/fmed.2023.1150933] [Medline: 37138750]

132. Kong YW, Baqar S, Jerums G, Ekinci EI. Sodium and its role in cardiovascular disease - the debate continues. Front Endocrinol (Lausanne). Dec 23, 2016;7:164. [FREE Full text] [doi: 10.3389/fendo.2016.00164] [Medline: 28066329]

133. Dehghan M, Mente A, Zhang X, Swaminathan S, Li W, Mohan V, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. Lancet. Nov 04, 2017;390(10107):2050-2062. [doi: 10.1016/S0140-6736(17)32252-3] [Medline: 28864332]

134. Cheung CY, Xu D, Cheng CY, Sabanayagam C, Tham YC, Yu M, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. Nat Biomed Eng. Jun 2021;5(6):498-508. [doi: 10.1038/s41551-020-00626-4] [Medline: 33046867]

135. Unterhuber M, Kresoja KP, Rommel KP, Besler C, Baragetti A, Klöting N, et al. Proteomics-enabled deep learning machine algorithms can enhance prediction of mortality. J Am Coll Cardiol. Oct 19, 2021;78(16):1621-1631. [FREE Full text] [doi: 10.1016/j.jacc.2021.08.018] [Medline: 34649700]

136. Xia X, Zhu C, Zhong F, Liu L. MDTips: a multimodal-data-based drug-target interaction prediction system fusing knowledge, gene expression profile, and structural data. Bioinformatics. Jul 01, 2023;39(7):btad411. [FREE Full text] [doi: 10.1093/bioinformatics/btad411] [Medline: 37379157]

137. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ. Jun 28, 2008;336(7659):1475-1482. [FREE Full text] [doi: 10.1136/bmj.39609.449676.25] [Medline: 18573856]

138. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J. Jun 2003;24(11):987-1003. [doi: 10.1016/s0195-668x(03)00114-3] [Medline: 12788299]

139. Goff DCJ, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. Jun 24, 2014;129(25 Suppl 2):S49-S73. [FREE Full text] [doi: 10.1161/01.cir.0000437741.48606.98] [Medline: 24222018]

140. Yang X, Li J, Hu D, Chen J, Li Y, Huang J, et al. Predicting the 10-year risks of atherosclerotic cardiovascular disease in Chinese population: the China-PAR project (prediction for ASCVD risk in China). Circulation. Nov 08, 2016;134(19):1430-1440. [doi: 10.1161/CIRCULATIONAHA.116.022367] [Medline: 27682885]

141. Woodward M, Brindle P, Tunstall-Pedoe H, SIGN group on risk estimation. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart. Feb 2007;93(2):172-176. [FREE Full text] [doi: 10.1136/hrt.2006.108167] [Medline: 17090561]

142. Zeraatkar D, Cheung K, Milio K, Zworth M, Gupta A, Bhasin A, et al. Methods for the selection of covariates in nutritional epidemiology studies: a meta-epidemiological review. Curr Dev Nutr. Sep 17, 2019;3(10):nzz104. [FREE Full text] [doi: 10.1093/cdn/nzz104] [Medline: 31598577]

143. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189-212. [doi: 10.1146/annurev.publhealth.22.1.189] [Medline: 11274518]

144. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med. Nov 01, 2018;178(11):1544-1547. [FREE Full text] [doi: 10.1001/jamainternmed.2018.3763] [Medline: 30128552]

XSL•FO

RenderX

145. Brisk R, Bond R, Finlay D, McLaughlin J, Piadlo A, Leslie SJ, et al. The effect of confounding data features on a deep learning algorithm to predict complete coronary occlusion in a retrospective observational setting. Eur Heart J Digit Health. Feb 20, 2021;2(1):127-134. [FREE Full text] [doi: 10.1093/ehjdh/ztab002] [Medline: 36711180]

146. Badgeley MA, Zech JR, Oakden-Rayner L, Glicksberg BS, Liu M, Gale W, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. NPJ Digit Med. Apr 30, 2019;2:31. [FREE Full text] [doi: 10.1038/s41746-019-0105-1] [Medline: 31304378]

147. Blackshear JL, McRee CW, Safford RE, Pollak PM, Stark ME, Thomas CS, et al. von Willebrand factor abnormalities and Heyde syndrome in dysfunctional heart valve prostheses. JAMA Cardiol. May 01, 2016;1(2):198-204. [doi: 10.1001/jamacardio.2016.0075] [Medline: 27437891]

148. Kwiendacz H, Wijata AM, Nalepa J, Piaśnik J, Kulpa J, Herba M, et al. Machine learning profiles of cardiovascular risk in patients with diabetes mellitus: the Silesia diabetes-heart project. Cardiovasc Diabetol. Aug 24, 2023;22(1):218. [FREE Full text] [doi: 10.1186/s12933-023-01938-w] [Medline: 37620935]

149. Okser S, Lehtimäki T, Elo LL, Mononen N, Peltonen N, Kähönen M, et al. Genetic variants and their interactions in the prediction of increased pre-clinical carotid atherosclerosis: the cardiovascular risk in young Finns study. PLoS Genet. Sep 30, 2010;6(9):e1001146. [FREE Full text] [doi: 10.1371/journal.pgen.1001146] [Medline: 20941391]

150. Tesche C, Bauer MJ, Baquet M, Hedels B, Straube F, Hartl S, et al. Improved long-term prognostic value of coronary CT angiography-derived plaque measures and clinical parameters on adverse cardiac outcome using machine learning. Eur Radiol. Jan 2021;31(1):486-493. [doi: 10.1007/s00330-020-07083-2] [Medline: 32725337]

151. Zhang PI, Hsu CC, Kao Y, Chen CJ, Kuo YW, Hsu SL, et al. Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. Scand J Trauma Resusc Emerg Med. Sep 11, 2020;28(1):93. [FREE Full text] [doi: 10.1186/s13049-020-00786-x] [Medline: 32917261]

152. Pencina MJ, D'Agostino RBS, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: the Framingham Heart Study. Circulation. Jun 23, 2009;119(24):3078-3084. [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.816694] [Medline: 19506114]

153. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PW, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. Circulation. Feb 14, 2006;113(6):791-798. [doi: 10.1161/CIRCULATIONAHA.105.548206] [Medline: 16461820]

154. Wickramasinghe CD, Ayers CR, Das S, de Lemos JA, Willis BL, Berry JD. Prediction of 30-year risk for cardiovascular mortality by fitness and risk factor levels: the Cooper Center Longitudinal Study. Circ Cardiovasc Qual Outcomes. Jul 2014;7(4):597-602. [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000531] [Medline: 24987054]

155. Nakai M, Watanabe M, Kokubo Y, Nishimura K, Higashiyama A, Takegami M, et al. Development of a cardiovascular disease risk prediction model using the Suita study, a population-based prospective cohort study in Japan. J Atheroscler Thromb. Nov 01, 2020;27(11):1160-1175. [FREE Full text] [doi: 10.5551/jat.48843] [Medline: 32023562]

156. Li Y, Yatsuya H, Tanaka S, Iso H, Okayama A, Tsuji I, et al. Estimation of 10-year risk of death from coronary heart disease, stroke, and cardiovascular disease in a pooled analysis of Japanese cohorts: EPOCH-JAPAN. J Atheroscler Thromb. Aug 01, 2021;28(8):816-825. [FREE Full text] [doi: 10.5551/jat.58958] [Medline: 33041313]

157. Chang W, Liu Y, Wu X, Xiao Y, Zhou S, Cao W. A new hybrid XGBSVM model: application for hypertensive heart disease. IEEE Access. Dec 03, 2019;7:175248-175258. [doi: 10.1109/access.2019.2957367]

158. Chen Y, Mao Y, Pan X, Jin W, Qiu T. Verification and comparison of three prediction models of ischemic stroke in young adults based on the back propagation neural networks. Medicine (Baltimore). Mar 19, 2021;100(11):e25081. [doi: 10.1097/MD.0000000000025081] [Medline: 33725985]

159. Barbieri S, Mehta S, Wu B, Bharat C, Poppe K, Jorm L, et al. Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach. Int J Epidemiol. Jun 13, 2022;51(3):931-944. [FREE Full text] [doi: 10.1093/ije/dyab258] [Medline: 34910160]

160. Quesada JA, Lopez-Pineda A, Gil-Guillén VF, Durazo-Arvizu R, Orozco-Beltrán D, López-Domenech A, et al. Machine learning to predict cardiovascular risk. Int J Clin Pract. Oct 2019;73(10):e13389. [doi: 10.1111/ijcp.13389] [Medline: 31264310]

161. Westerlund AM, Hawe JS, Heinig M, Schunkert H. Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence. Int J Mol Sci. Sep 24, 2021;22(19):10291. [FREE Full text] [doi: 10.3390/ijms221910291] [Medline: 34638627]

162. Rao S, Li Y, Ramakrishnan R, Hassaine A, Canoy D, Cleland J, et al. An explainable transformer-based deep learning model for the prediction of incident heart failure. IEEE J Biomed Health Inform. Jul 2022;26(7):3362-3372. [doi: 10.1109/jbhi.2022.3148820]

163. Curtis D. Polygenic risk score for schizophrenia is not strongly associated with the expression of specific genes or gene sets. Psychiatr Genet. Aug 2018;28(4):59-65. [doi: 10.1097/YPG.0000000000000197] [Medline: 29672343]

164. Morales J, Welter D, Bowler EH, Cerezo M, Harris LW, McMahon AC, et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol. Feb 15, 2018;19(1):21. [FREE Full text] [doi: 10.1186/s13059-018-1396-2] [Medline: 29448949]

165. Advancing governance, innovation, and risk management for agency use of artificial intelligence. Executive Office of the President, United States. Mar 28, 2024. URL: https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf [accessed 2024-07-10]

166. Ethics guidelines for trustworthy AI. Publications Office of the European Union. URL: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en [accessed 2024-07-03]

167. Rees C, Müller B. All that glitters is not gold: trustworthy and ethical AI principles. AI Ethics. Nov 16, 2022;3(4):1-14. [FREE Full text] [doi: 10.1007/s43681-022-00232-x] [Medline: 36406882]

168. Radclyffe C, Ribeiro M, Wortham RH. The assessment list for trustworthy artificial intelligence: a review and recommendations. Front Artif Intell. Mar 9, 2023;6:1020592. [FREE Full text] [doi: 10.3389/frai.2023.1020592] [Medline: 36967834]

169. Chen IY, Szolovits P, Ghassemi M. Can AI help reduce disparities in general medical and mental health care? AMA J Ethics. Feb 01, 2019;21(2):E167-E179. [FREE Full text] [doi: 10.1001/amajethics.2019.167] [Medline: 30794127]

Abbreviations

ACM: all-cause mortality
ACS: acute coronary syndrome
AI-TREE: AI transparent, replicable, ethical, and effective research
AI: artificial intelligence
ANLL: average negative log likelihood
ANN: artificial neural network
ASCVD: atherosclerotic cardiovascular diseases
AUC: area under the curve
BP: blood pressure
CHD: coronary heart disease
CNN: convolutional neural network
CONSORT: Consolidated Standards of Reporting Trials
CPRD: Clinical Practice Research Datalink
CVD: cardiovascular disease
ECG: electrocardiogram
EHR: electronic health record
FRS: Framingham risk score
GBM: gradient boosting machine
HF: heart failure
ICD: International Classification of Diseases
KNN: k-nearest neighbor
LR: logistic regression
MESA: Multi-Ethnic Study of Atherosclerosis
MI: myocardial infarction
ML: machine learning
NHIS-HEALS: National Health Insurance System-National Health Screening Cohort
PCE: pooled cohort equation
PROBAST: Prediction Model Risk of Bias Assessment Tool
RF: random forest
RSF: random survival forest
SBP: systolic blood pressure
SPRINT: Systolic Blood Pressure Intervention Trial
SVM: support vector machine
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis

XSL•FO

RenderX