

RESEARCH ARTICLE

Open Access



Artificial intelligence in the risk prediction models of cardiovascular disease and development of an independent validation screening tool: a systematic review

Yue Cai^{1†}, Yu-Qing Cai^{1†}, Li-Ying Tang^{1†}, Yi-Han Wang¹, Mengchun Gong², Tian-Ci Jing³, Hui-Jun Li^{4,5}, Jesse Li-Ling⁶, Wei Hu⁷, Zhihua Yin^{8*}, Da-Xin Gong^{3,9*} and Guang-Wei Zhang^{3,9*} 

Abstract

Background A comprehensive overview of artificial intelligence (AI) for cardiovascular disease (CVD) prediction and a screening tool of AI models (AI-Ms) for independent external validation are lacking. This systematic review aims to identify, describe, and appraise AI-Ms of CVD prediction in the general and special populations and develop a new independent validation score (IVS) for AI-Ms replicability evaluation.

Methods PubMed, Web of Science, Embase, and IEEE library were searched up to July 2021. Data extraction and analysis were performed for the populations, distribution, predictors, algorithms, etc. The risk of bias was evaluated with the prediction risk of bias assessment tool (PROBAST). Subsequently, we designed IVS for model replicability evaluation with five steps in five items, including transparency of algorithms, performance of models, feasibility of reproduction, risk of reproduction, and clinical implication, respectively. The review is registered in PROSPERO (No. CRD42021271789).

Results In 20,887 screened references, 79 articles (82.5% in 2017–2021) were included, which contained 114 datasets (67 in Europe and North America, but 0 in Africa). We identified 486 AI-Ms, of which the majority were in development ($n = 380$), but none of them had undergone independent external validation. A total of 66 idiographic algorithms were found; however, 36.4% were used only once and only 39.4% over three times. A large number of different predictors (range 5–52,000, median 21) and large-span sample size (range 80–3,660,000, median 4466) were observed. All models were at high risk of bias according to PROBAST, primarily due to the incorrect use of statistical methods. IVS analysis confirmed only 10 models as “recommended”; however, 281 and 187 were “not recommended” and “warning,” respectively.

[†]Yue Cai, Yu-Qing Cai, and Li-Ying Tang are co-first authors.

*Correspondence:

Zhihua Yin

zhyin@cmu.edu.cn

Da-Xin Gong

gongdx@cmu1h.com

Guang-Wei Zhang

gwzhang@cmu.edu.cn

Full list of author information is available at the end of the article



Conclusion AI has led the digital revolution in the field of CVD prediction, but is still in the early stage of development as the defects of research design, report, and evaluation systems. The IVS we developed may contribute to independent external validation and the development of this field.

Keywords Artificial intelligence, Cardiovascular disease, Machine learning, Risk prediction models

Background

The surge in cardiovascular diseases (CVDs) has become a global challenge with a steadily climbing trend of cardiovascular deaths from 12.1 million in 1990 to 18.6 million in 2019 [1, 2]. Risk prediction, a primary strategy in addressing this worldwide problem, has brought significant benefits to some developed countries through the improvement of the effectiveness of life intervention and reduction of economic burden [3, 4]. Therefore, risk prediction has been expected as an efficient way to achieve World Health Organization (WHO) goals for reducing CVD-related mortality by 25% by 2025, and some classic CVD prediction models (e.g., the Framingham [5] and SCORE [6], referred to as traditional models [T-Ms] in this study) has been incorporated into clinical guidelines by the European Society of Cardiology (ESC) and the American College of Cardiology/American Heart Association (ACC/AHA) [7, 8].

Artificial intelligence (AI), encompassing machine learning (ML) and deep learning (DL), is a field within computer science dedicated to the development of computational systems capable of performing tasks that traditionally necessitate human intelligence, such as learning, reasoning, problem-solving, perception, language comprehension, and decision-making. The application of AI in the healthcare sector, including disease risk prediction, is rapidly advancing and playing an increasingly significant role [9–13]. Alongside the substantial transformations driven by AI in this domain, it also introduces a spectrum of challenges and issues, including concerns related to ethics, legality, data privacy, security, bias, fairness, transparency, and explainability [14–20]. At this critical juncture in the AI field, characterized by a coexistence of challenges and opportunities in the era of big data, AI-driven disease risk prediction stands ready to harness immense potential and address substantial needs [11, 21]. It has demonstrated notable superiority over the T-Ms, owing to its more robust data-processing capability, fewer condition restrictions, and better performance [11], thereby providing a more promising predictive strategy for CVDs.

However, a comprehensive and systematic overview of AI for CVD prediction is still lacking, despite the field has witnessed several recent comparative reviews that tend to emphasize specific aspects. For instance, Suri et al. provided a comprehensive summary of ML

paradigms with a technical emphasis [22]. Azmi et al. focused on emphasis on comparing the predictive performance of various ML-based classification algorithms using medical big data [23]. Infante et al. and Assadi et al. primarily reviewed the contributions of cardiac computed tomography angiography and cardiac magnetic resonance to AI-CVD prediction [24, 25]. Triantafyllidis et al. conducted a review on the impact of DL on the diagnosis, management, and treatment of major chronic diseases, including cardiovascular disease [26]. Zhao et al. only observed social determinants contributing to AI-CVD prediction [27]. Liu et al. compared the ML and traditional approaches for atherosclerotic CVD risk prognostication [28]. These articles provide limited insights for a comprehensive understanding of the current state of this field. Therefore, in reference to previously published reviews that elucidate the development status of T-Ms for CVD prediction [29], we conducted this summarization work and attempted to explore potential solutions to address the current challenges.

Methods

We conducted this systematic review using the CHARMS checklist. This review has been registered in the international prospective register of systematic reviews (PROSPERO), with the registration number CRD42021271789, where all updates of the review will also be recorded. This review followed the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement (Additional file 1). Patients and the public were not involved in setting of the research question, designing or implementing the study, or in interpreting or writing of the results.

Literature search

A literature search was conducted in PubMed, Web of Science, Embase, and IEEE, using search terms to identify primary articles focused on the development and/or validation of AI in predicting incident CVD up to July 2021. A cross-reference check was performed for all reviews on CVD prediction models identified by our search. Search strategies are described in Additional file 2: Text 1.

Eligibility criteria

We included only original research on risk prediction models for humans with full text in English, excluding

studies that (1) are for clustering and outcome classification and (2) in the postoperative or perioperative period of cardiac surgery or non-cardiac surgery. The detailed process and criteria are shown in Fig. 1.

Screening process

Two independent reviewers screened the titles and abstracts. The corresponding full texts were retrieved and reviewed after identifying potentially eligible articles. Any disagreements during this process were resolved

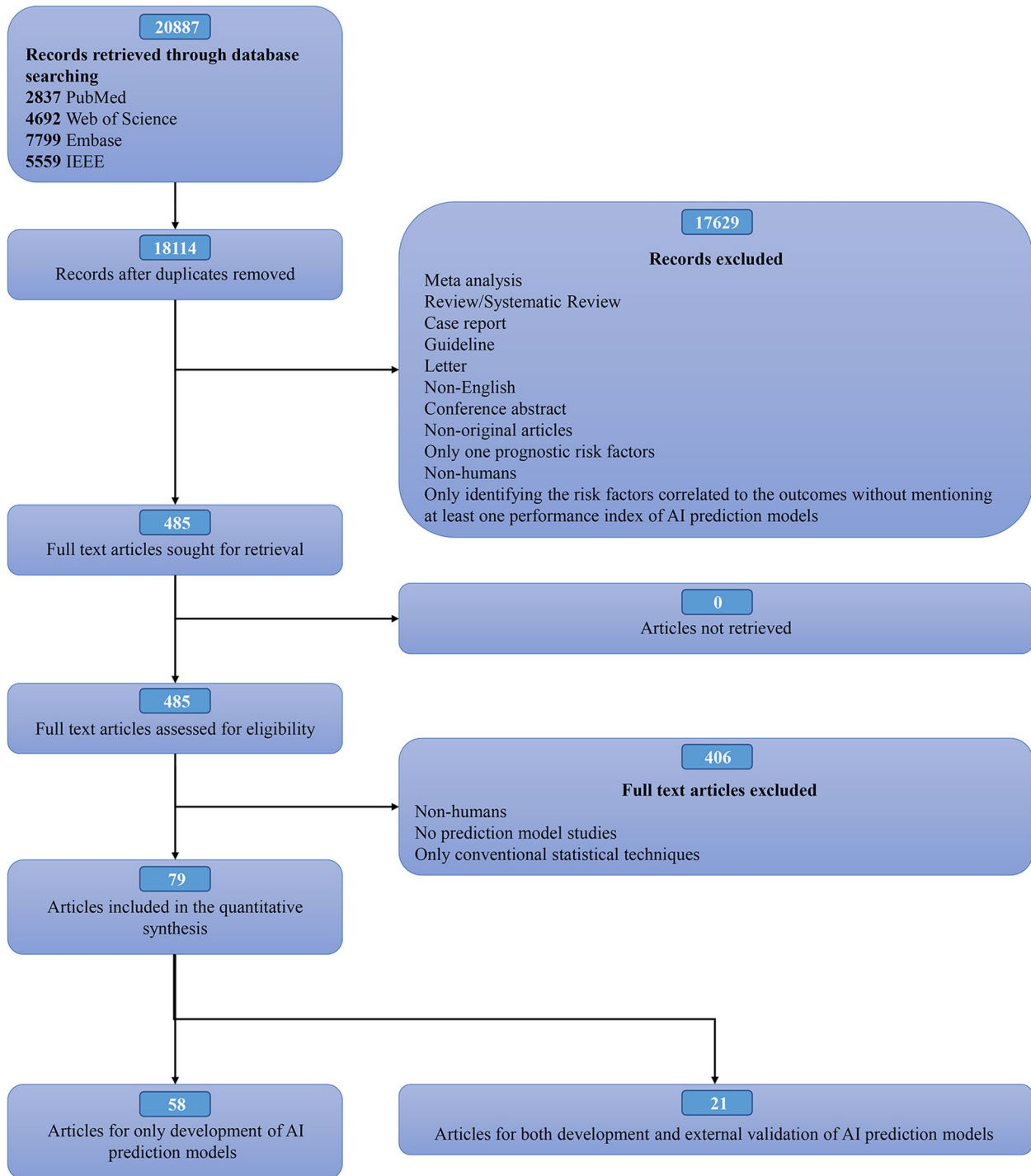


Fig. 1 The flow diagram for the literature search performed in the present study

through discussion among all team members to reach a consensus.

Data extraction and critical appraisal

The list of extracted items was based on the CHARMS checklist. Two independent reviewers extracted the data, with any discrepancies being resolved through discussion by the entire team. The risk of bias was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [30], and the extraction form included four domains: participants, predictors, outcomes, and statistical analysis. Results were summarized using descriptive statistics. Quantitative synthesis of the models was not performed.

Assessment of the feasibility of independent external validation

To evaluate the feasibility of independent external validation of each model, we conducted a literature review of existing assessment guidelines or tools in the field of AI/ML medical research (Additional file 2: Text 2 and Fig. S1) and summarized initially candidate items for designing a screening tool. Subsequently, a preliminary plan that weighs screening efficiency and initiatives for the study of ideal CVD prediction models [29] was further discussed and revised by a panel of experts, including clinicians (G-W Z and D-X G), AI experts (T-C J and MG), clinical epidemiologists (ZY), and information technology specialists (WH), among others. Ultimately, a novel scoring system was further developed through consistent feedback of three independent international experts in AI or CVD domains from ExpertScape™ rank and peer recommendations. It is called the independent validation score (IVS), comprising five steps with five score items as follows: transparency of models, performance of models, feasibility of reproduction, risk of reproduction, and clinical implication sequentially. After the five-step scoring process, five grades of feasibility recommendation were set, including “strongly recommended”, “recommended”, “neutral”, “warning”, and “not recommended”. The detailed definitions and rules are shown in Fig. 2 and Table 1.

Results

Study designs and populations

Overall, 79 articles were finally included from 2000 to 2021 (Additional file 2: Table S1) [49–127], with 65 (82.25%) published between 2017 and 2021 (Fig. 3A). In total, 114 cohorts (datasets) were used, with 27 in Europe, 40 in America (mainly the USA), 27 in Asia (mainly Korea), and 5 in Oceania (Australia), 3 multi-country cohorts, but 0 in Africa, as shown in Fig. 3C. A total of 647 different models were identified, including

161 T-Ms excluded from this bibliometric research and 486 AI-Ms involved in the following analysis. Most models were developed using data from 101 trials, and only a minority were from five case–control studies and eight nested case–control studies. All cohort participants were enrolled consecutively.

We included 63 papers focusing on the general population and 16 that addressed subgroups for specific diseases, including type 2 diabetes ($n=4$, 5%), hypertension ($n=4$, 5%), and kidney diseases ($n=4$, 5%). Forty different age ranges were reported across the cohorts, except for 45 cohorts which did not mention age range. The two most common ones were 40 to 79 years ($n=12$, 15%) and 30 to 77 years ($n=6$, 7%), and the average age ranged from 42 to 78 years. The majority of papers ($n=70$, 89%) were not sex specific or stratified, with only 24 cohorts having roughly equal proportions of males and females (45–55% females or males in numbers).

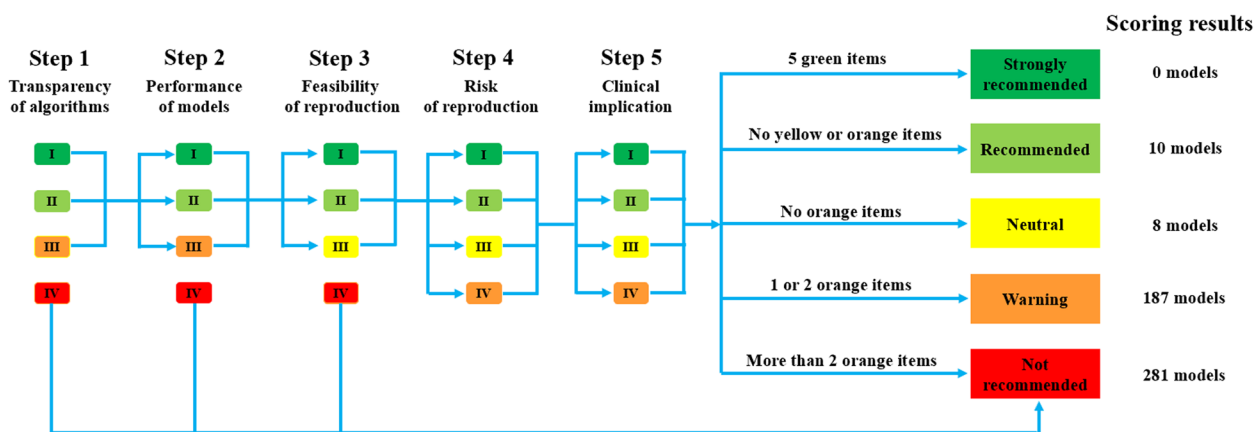
Data sources and research environment

Only a minority of the articles ($n=24$, 30%) used multiple datasets to develop models as indicated in data source analysis, showing an obvious dominance in single dataset-deriving models. Of all 114 datasets, 42 are multi-centered, and 32 are single-centered; however, 40 databases are unknown. In terms of information collection, only 56 were from electronic health record (EHR), 11 from EHR+questionnaire, 1 from questionnaire+personal interview, and 46 did not clearly mention the data sources. Regarding the issue of missing variables problem, only 5 cohorts clearly described the number of participants with missing variables, whereas 94 cohorts did not mention this value. Fifteen cohorts excluded all participants with missing variables.

In the research environment, the largest number came from the hospital scene ($n=44$, 39%), followed by community ($n=20$, 18%), primary health care institutions ($n=5$, 4%), and hospital scene+primary health care institutions ($n=1$, 1%). Forty-four cohorts did not state the environment. The study periods ranged from 1965 to 2019, with 57 cohorts reporting the study period, 26 cohorts reporting only the baseline time, and 31 not mentioning the study period.

Criteria for inclusion and exclusion

Of all 79 articles, only 36 clearly reported the criteria for inclusion, mainly including age restriction, necessary clinical examination and variables, special disease, adequate follow-up time, and number of visits during the period of follow-up. Twenty-two papers did not clearly state the exclusion criteria (Additional file 2: Table S2).



Explanations for this tool

- This tool establishes a five-step screening process for independent external validation research, based on the current reporting trends in CVD-AI prediction models, guidelines for medical AI research, initiatives for the study of ideal CVD prediction models and expert consensus.
- This tool categorizes models into five levels. Models rated in the top two levels may have higher validation value. The “Neutral” level indicates uncertainty about worthiness of validation. The “Warning” suggests lower validation value and a potential risk of resource wastage, it does not imply, however, that validation research cannot be conducted for models in these categories. The classification of “Not recommended” strongly discourages pursuing independent external validation, irrespective of underlying considerations. However, it should be noted that these may not reflect the research’s intrinsic value.
- This tool is a preliminary version based on the current state of affairs. As the field evolves, the assessment process and criteria will undergo changes.
- The objectives and principles of the five assessment steps are outlined as follows:
 - **Step 1** To confirm whether the candidate models can be validated at the algorithm level.
 - **Step 2** To evaluate the worthiness of candidate models for replication from a performance perspective.
 - **Step 3** To assess challenges associated with validating and clinically implementing candidate models, including data acquisition, annotation, and software/hardware requirements, cost effective, and so on, with special consideration for low-income countries or regions.
 - **Step 4** To predict the risk of validation failure for candidate models based on established authoritative standards, such as PROBAST.
 - **Step 5** To appraise the clinical applicability value of candidate models, considering their generalizability, utility, and adherence to AI ethical principles.
- The scoring involves a stepwise evaluation with items stratified within each step, represented by colors ranging from green, light green, yellow, orange, to red, indicating the quality from superior to inferior. The comprehensive output result is formed by synthesizing the outcomes at each step. Two filtering mechanisms are employed: one involves direct negation through three conditions influencing validation in the initial three steps respectively, or direct affirmation based on five strongly recommended conditions in all five steps respectively, facilitating researchers to make rapid decisions. The other involves a judgment system from poor to good based on the cumulative results of difficulty, risk, and clinical implication. If there are two or more “Orange” judgments, it is deemed as “Not recommended”. If not, accumulating one or two “Orange” judgments is labeled as a “Warning”. If there are no “Orange” judgments but there are “Yellow” judgments, it is considered “Neutral”. If none of these conditions are met, it is classified as “Recommended”.
- The final output of this scoring system consists of five recommendation levels, corresponding to:
 - **Strongly recommended** Easily implementable independent external validation with significant clinical implications.
 - **Recommended** Relatively easy to verify with potential clinical significance.
 - **Neutral** Challenging to validate but still clinically meaningful.
 - **Warning** Difficult to validate and may lack clinical significance.
 - **Not recommended** Extremely high risk of implementation or invalid validation.

Fig. 2 The sketch map of the independent validation score procedure and results

Predictors

In all AI-Ms, the median number of predictors was 21 (range 5–52,000), with an unquantifiable total number due to a lack of detailed information in individual articles. These predictors were into two types: traditional factors and new-added ones, according to whether they can be addressed by T-Ms. In addition to traditional factors such as age (in 400 models), sex (in 357 models), total cholesterol (in 276 models), and smoking status (in 266 models), several new-added predictors have emerged in AI-Ms, including electrocardiogram (ECG) image ($n=84$, 17%), ultrasound image ($n=44$, 9%), magnetic resonance imaging (MRI) image ($n=18$, 4%), computed tomography (CT) image ($n=12$, 2%), single nucleotide polymorphisms (SNPs) ($n=9$, 2%), and proteins ($n=4$, 1%) as shown in Fig. 4. Further analysis

showed that 135 models (30.96%) were built using these new-added data.

CVD outcomes and measurement method

We found a large variation in predicted outcomes among different models. A total of 42 single endings and 61 combined endings were confirmed in all models. The most common in all 103 endpoints were complete CVD ($n=40$, 39%) and death ($n=16$, 16%). However, a considerable heterogeneity was identified in the definitions of these outcomes, such as 19 different definitions for CVD. The main origin of definitions is diverse, including disease codes (ICD9 or ICD10, $n=36$, 35%), self-report ($n=4$, 4%), and other international

Table 1 The specific evaluation criteria of IVS

Score items	Grade	Specific evaluation criteria	References
Transparency of algorithms	I	Post the trained models that can be directly loaded by other researchers for a contiguous independent validation or online/mobile user-friendly calculators that can allow batch processing of participant information (e.g., a prediction software or tool)	· APPRAISE-AI [31] · MI-CLAIM [32] · AI-TREE [33]
	II	Apply and report the classic algorithms that can be found in some common tools/platforms OR report complete codes and hyperparameters and required description, allowing independent researchers to run the pipeline end to end	
	III	Report formulas and/or incomplete hyperparameters without required description, leading to difficulties in replication or incomplete reproducibility	
	IV	Incomplete reports that cannot be used for reproduction	
Performance of models	I	At least report the discrimination (preferably c-index) and calibration (preferably calibration plot/table) of the model, and the performance index version is clearly reported and index is excellent (e.g., $0.9 < \text{c-index} < 1.0$; calibration intercept close to 0 and calibration slope close to 1)	TRIPOD [34] · CHARMS checklist [35] · Official statement [36] · AI-TREE [33] · Expert comment [37]
	II	At least report the discrimination (preferably c-index) and calibration (preferably calibration plot/table) of the model, and the performance index version is clearly reported and index is good (e.g., $0.7 < \text{c-index} < 0.9$; calibration intercept deviates moderately from 0, and calibration slope deviates moderately from 1)	
	III	Do not report the discrimination or calibration of the models; OR the performance index version is not clearly reported; OR the value of the index is unknown	
	IV	The model performance is at a low accuracy (e.g., $\text{c-index} < 0.7$; calibration intercept deviates severely from 0 and calibration slope deviates severely from 1)	
Feasibility of reproduction	I	The office-based models without requirement for laboratory and inspection data (also known as non-laboratory models)	· Validation and evaluation framework [38] · AI standardization [39]
	II	The laboratory-based models only requiring routine clinical structured data, which are easy to obtain and do not need secondary operation (e.g., image pre-processing or annotation, etc.)	· AI-TREE [33] · MI-CLAIM [32] · CONSORT-AI [40] · MAIC-10 [41]
	III	Include data derived from unconventional laboratory and inspection, complex gene-related testing, tissue specimen, and other resource-limiting extensive applications, which are hard to obtain or require secondary operation (e.g., labeling)	· SR of validity and clinical utility [11] · WHO laboratory-based and non-laboratory models [42] · Laboratory-based and non-laboratory models [43]
	IV	Do not report the variables	
Risk of reproduction	I	No domain high risk (evaluated by using PROBAST)	· PROBAST [30]
	II	Only one domain is high risk (evaluated by using PROBAST)	
	III	Two domains are high risk (evaluated by using PROBAST)	
	IV	Over two domains are high risk (evaluated by using PROBAST)	
Clinical implication	I	Identified novel risk markers or novel risk standards, which will optimize existing clinical preventive strategies and contribute to patient benefit for the general population and major CVDs, similar to classical T-Ms (e.g., Framingham Score)	· SR of T-Ms [29] · Biomedical research AI guideline [44] · BS30440 [45] · APPRAISE-AI [31]
	II	Do not identify novel risk markers or novel risk standards, but enhance the predictive capacity beyond that of existing methods, which may optimize existing clinical preventive measures or offer additional benefits for the non-rare population and non-rare subset of CVDs (more than 1/2000 of the general population)	· Consolidated AI reporting guideline [46] · AI-TREE [33] · SR of validity and clinical utility [11] · Rare CVD [47, 48]
	III	Only enhance the predictive capacity beyond that of existing methods, but cannot alter the existing preventive interventions or provide additional benefits for the non-rare population and non-rare subset of CVDs (more than 1/2000 of the general population)	
	IV	Do not enhance the predictive performance beyond that of existing methods OR only target a rare population or subset of CVDs (fewer than 1/2000 of the general population, e.g., infiltrative cardiac diseases), leading to inadequate validation and a lack of clinical utility for a broader population	

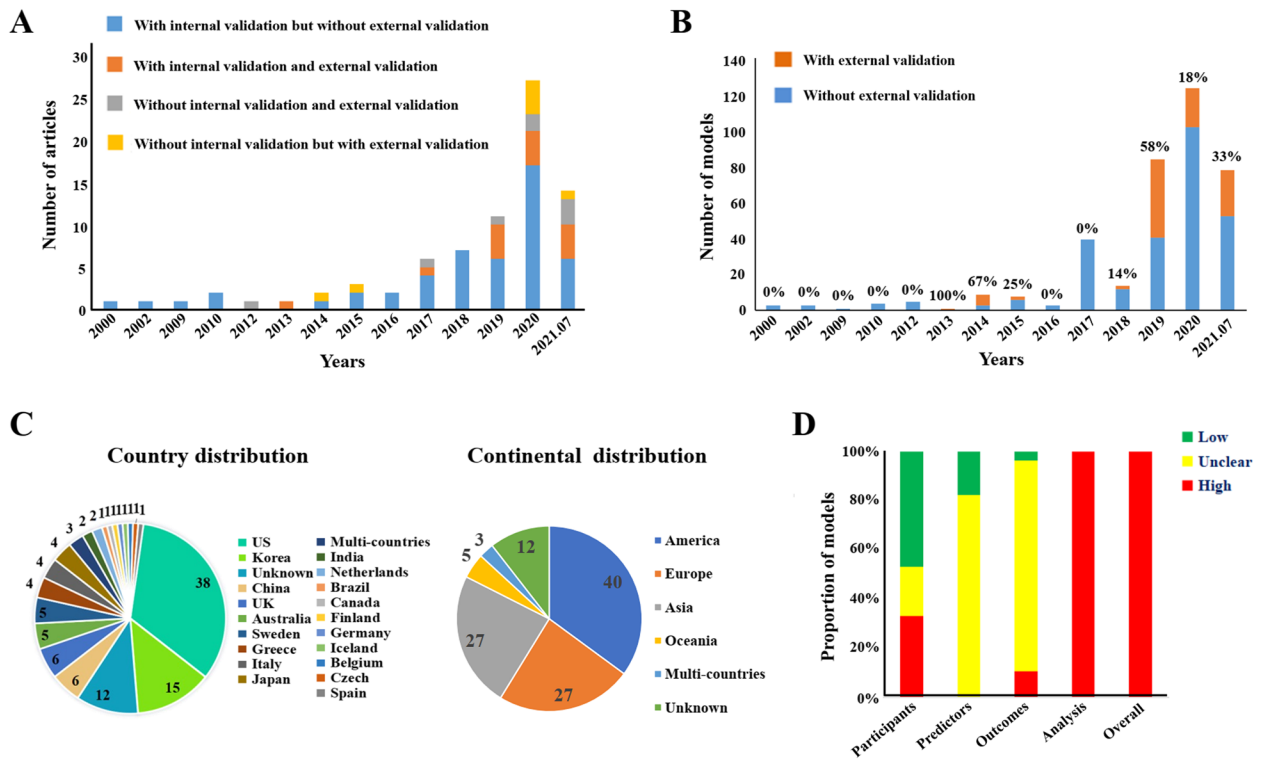


Fig. 3 **A** The bar chart of number analysis for articles per year (up to July 31, 2021). **B** The bar chart of number analysis for validated models per year (up to July 31, 2021). **C** The pie graph of papers' geographical distribution. US, The United States of America; UK, The United Kingdom of Great Britain and Northern Ireland. **D** The bar chart of bias risk analysis with PROBAST

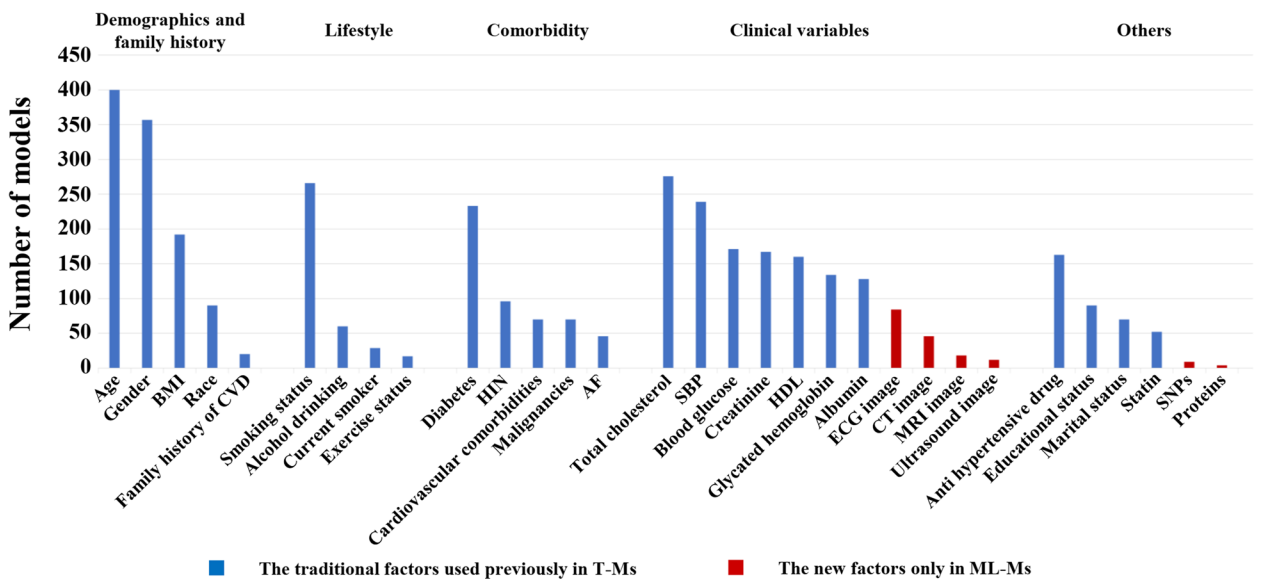


Fig. 4 The bar chart of summary and categories of predictors involved in all models

guidelines ($n=3$, 3%). Additionally, there were 149 models (30.66%) not reporting the definition of the outcomes in 21 papers.

The most common prediction horizons in AI-Ms were 10 ($n=107$, 22%) and 2.5 years ($n=70$, 14%) with a range between 1 day and 15 years. Only 25 papers reported the measurement methods for all included outcomes, which primarily comprised clinical records, national institute statistics, questionnaires, and personal interviews. Only 11 articles reported that the outcome measurement was blinded, and two articles explicitly reported not using the blinding method. Other detailed information is summarized in Additional file 2: Table S3.

Sample size and performance

In total, 4 articles did not report the sample size and 22 articles did not report the number of events. Based on reported data, the number of participants included in AI-Ms ranged between 80 and 3,660,000 (median 4466), and the commonly used order of magnitudes of the number ranged from 1000 to 10,000 ($n=44$). The ending events occurred ranging from 10 to 152,790 (median 504).

In all the articles ($n=79$), at least one measure of predictive performance was reported, which was also one of the inclusion criteria for the article in this system review. C index was mainly reported for 482 models. The calibration plot was for 90 models. Sensitivity/recall was for 312 models. Specificity/true negative rate (TNR) was for 209 models. Precision/positive predictive value (PPV) was for 201 models; accuracy was for 199 models; F1 score was for 137 models; Matthews correlation coefficient (MCC) was for 7 models.

Assessment of algorithms transparency and model reproductivity

Overall, 13 categories of 66 idiographic algorithms were identified based on their operation mechanisms and accepted classification principles. The most frequently applied algorithm in all models is logistic regression ($n=74$, 15.2%), followed by random forest ($n=71$, 14.6%) and neural network ($n=63$, 13.0%) as summarized in Additional file 2: Table S4. Only 26 (39.4%) were used more than 3 times, while 24 (36.4%) appeared only once in all algorithms. In total, 212 models did not report codes, formulas, or hyperparameters, consequently identified as non-reproductive.

Development models and external validation models

Of the 486 models, 380 were development models and 106 were external validation models (validating 103 development models), as reported in their primary papers. Notably, no independent external validations

were found in this field. Additionally, most datasets ($n=17$, 68%) used for external validation were from the same countries as those used for development models in their primary papers; however, most datasets used for external validation were from different research periods ($n=13$, 52%) and different settings ($n=18$, 72%) as those used for development models. The development and external validation of models were conducted by the same investigators in the same article. Our additional exploratory analysis revealed a lower validation propensity in the developed models with new variables (25.24% vs. 43.68%, $P=0.001$) and an AUC < 0.7 (0% vs. 70.45%, $P<0.001$), which provide important information for us to build IVS.

Risk of bias

All models were at high risk of bias ($n=486$, 100%) according to the assessment using PROBAST, as shown in Fig. 3D. The most common reasons were as follows: 1) inappropriate data sources or inappropriate enrolment strategy in the participant domain ($n=161$, 33%); 2) not mentioning the definition and measurement of the predictors, or not mentioning whether the predictor's assessments were blinded to outcome knowledge in the predictor domain ($n=401$, 83%); 3) inappropriate outcome classification method, outcome definition was not the same for all participants, predictors included in the outcome definition, or the determination of outcomes with the knowledge of predictors in the outcome domain ($n=52$, 11%); 4) not accounting for the complexities of data, not evaluating the performance of models appropriately, or not accounting for model overfitting and optimism in the statistical analysis domain ($n=486$, 100%). The details are shown in Additional file 2: Table S5.

Summary of existing assessment guidelines or tools

Overall, a total 29 of guidelines or tools related to quality assessment or control in the past decade (mainly in the last four years), with 5 for developing quality, 14 for reporting quality, and 10 for both (Additional file 2: Table S6) [11, 30–32, 34, 38, 40, 41, 44–46, 128–145]. In addition to the study design, statistical methods, model performance, risk of bias, AI ethics risk, replicability, as well as clinical implementation, application, and implication in both developing and reporting assessments, the complexity and standardization of data acquisition and processing, required resources (such as software platforms, hardware, or technical professionals), and cost-effectiveness are also focal points in many developing assessments. These provide a core framework for the construction of IVS.

Independent validation score

Most models were identified as “not recommended” ($n=281$, 58%) or given a “warning” ($n=187$, 38%). Only 10 (2%) were classified as “recommended,” and none were identified as “strongly recommended” as revealed by our IVS for all 486 models in Fig. 2. The recommended models are displayed in Additional file 2: Table S7. Insufficient transparency of models contributed the largest number of “not recommended” ($n=212$), followed in turn by performance ($n=56$), feasibility of reproduction ($n=12$), and comprehensive reasons ($n=1$).

Discussion

This systematic review is the first to encompass global AI studies of CVD prediction in the general population for more than 20 years, starting from the first article published in 2000 [72]. It presents the current status and broad trends in this field through a comprehensive search and careful selection of studies. We performed an extensive data extraction and thorough analysis of key characteristics in publications, including the predictors, populations, algorithms, performance, and bias. On top of this, we have developed a tool for evaluating replicability and applicability, to screen appropriate AI-Ms for independent external validation, addressing the key issues currently hindering the development of this field. The findings and conclusions are expected to provide references and help for algorithm developers, cohort researchers, healthcare professionals, and policy makers.

Principal findings

Our results revealed significant inefficiency in external validations and a lack of independent external validation for the existing models, indicating that researchers in the field of AI risk prediction were more inclined to put emphasis on new models developing, instead of validating, although validation is crucial in determining clinical decisions [146]. According to the experience in the field of T-Ms research, these may lead to a large number of useless prediction models, thereby suggesting that more attention should be paid to external validation to avoid research waste and facilitate the translation of high-performing predictive models into clinical practice [147–149]. Based on the facts that most studies used data from only one cohort, we conjecture that limited data source may be one of the main reasons that restrict the implementation of external validations. Therefore, the multi-centers studies, especially multi-countries studies (only three were found in our review), should be encouraged to establish multi-source databases.

It is found that the majority of studies were conducted in Europe and North America, with only a few in the

developing countries from Asia and South America, and unfortunately none in Africa. The similar geographical trends have been confirmed in the conventional CVD prediction models through previous literature reviews [29, 150]. However, the prevalence of the CVD is dramatically increasing in those low- or middle-income countries, consequently contributing over three quarters of CVD deaths all over the world and causing great burden to the local medical system [151–154]. Considering the influence of ethnic heterogeneity on the prediction model [155], native AI-Ms tailored to these countries should be developed for local prevention of CVD.

Four classic indexes, age, sex, total cholesterol, and smoking status, were more frequently used in AI-Ms in all presented predictors (some papers not fully representing the used predictors), similar to T-Ms. However, more importantly, the following summary demonstrates that AI-Ms have triggered a profound revolution to predictors owing to its strong data computing capability. First, the median number of predictors in the AI-Ms was approximately 3 times greater than that in T-Ms as collated by Damen et al. [29]. Second, except for the classic predictors (e.g., demographics and family history, lifestyle, and laboratory measures), several new indexes have been involved in AI-Ms, mainly consisting of some multimode data that cannot be recognized and utilized by T-Ms at all (e.g., image factors and gene- or protein-related information). Third, the limitation of data range has been eliminated, as proven by the no fixed age range and sex-specific equation for the development of AI-Ms, which were important concerns in classic T-Ms. Fourth, AI models allow data re-input and utility. Researchers gathered data many times in the follow-up procedure in recurrent neural network (RNN) models, and these time series data were used to retrain the AI-Ms for further improvement of performance [55, 112]. Another interesting improvement is that the screening of predictors could be executed automatically by AI instead of classic log calculation [50, 52].

The systematic review of specific models is imperative for the head-to-head comparison of these models and the design of the relevant clinical trials [156, 157]. Our analysis of report quality was performed through reference to the TRIPOD statement and CHARMS-CHECKLIST, to inform readers regarding how the study was carried out [158]. Worryingly, we found that many articles did not report important research information, which not only significantly restrict the readability of articles largely but also may lead to the unwarranted neglect for the previous evidence through subsequent researches [159–162]. Therefore, we have to strongly recommend that each study should upload a statement of TRIPOD or upcoming TRIPOD-AI designed specifically for AI prediction

models when the manuscripts were submitted [12, 163, 164].

According to PROBAST, a common evaluation method of risk of bias for traditional prediction models [165], all included AI-Ms were judged as high risk in our summary, mainly owing to ignorance or failure to report competing risk in the item of statistical analysis. Similar trends of high risk have been confirmed in many previous systematic reviews regarding AI-Ms for other diseases, although there are some differences in specific reasons, which involved more frequently sample size, calibration, missing data handling, and so on [12, 166–168]. This could potentially be another significant constraint on the independent external validation of models, in addition to the various issues mentioned earlier, which currently hinder the widespread adoption of AI-Ms for CVD clinical practice. Therefore, it is strongly suggested again that more attention should be focussed on statistical analysis, not only for authors in the research and writing process, but also for reviewers and editors during review and publication. Meanwhile, these widely high-risk judgment ratios prompt us to raise question whether the current criteria are too harsh for AI-Ms, because it is unclear whether some algorithms may offset competing risk due to their “black box” effect, and it should not be ignored that the classic method of EPV may not be suitable for the sample size calculation in some ML algorithms owing to their specific operation mechanism [169–171].

Best practice guidance and specific pathways for the translation of AI-healthcare research into routine clinical applications have been developed. Holmes et al. summarized the AI-TREE criteria [33], while Banerjee et al. created a pragmatic framework for assessing the validity and clinical utility of ML studies [11]. Building on this prior work and the experiences reported in studies involving AI risk prediction models for various diseases [75, 172–174], our insights gained during the validation process of existing AI models, as well as a combination of summary of existing AI research assessment guidelines or tools and experts’ suggestions, we have developed an IVS for screening independent external validation models. This tool is primarily intended for researchers involved in the validation process rather than developers during the implementation phase. In this scoring system, in addition to the two recognized criteria of transparency and risk assessment, the performance and clinical implication were included to determine their suitability for independent external validation, which to some extent, align with factors typically considered during the model development process, such as impact, cost-effectiveness, and AI-ethics [11, 33]. In assessing performance, we opted for the two most widely reported and strongly recommended indices for discrimination and calibration, namely the c

index and calibration plot/table, instead of specificity or sensitivity, as they are not recommended by the TRIPOD and checklist guidelines [34, 35, 158]. Furthermore, the consistency of retrospective validation datasets and the challenges in acquiring prospective study data are key factors influencing external validation [75, 172–174], especially in the case of factors like imaging, biomarkers, genomics, which may also encounter issues such as lack of standardization and biased reporting [33]. Building upon the WHO’s principles of model utility [42], the acquisition and handling of laboratory-based and emerging multimodal predictive factors’ acquisition and handling are essential assessment components in evaluating the feasibility of independent external validation.

Our IVS results have indicated that more than 95% of the models may not be suitable for independent external validation by other researchers, and as a result, may not provide any useful help for the following clinical application. Therefore, it is rather reasonable to explain why there have been no independent external validation researches in the field of CVD-AI prediction for over 20 years. In addition to the problem of model transparency, the following other four reasons also are considered to account for irreproducibility of the models, including increased difficulty in parameter acquisition and processing, uncertain expected performance, and low reliability owing to high risk. Therefore, it is strongly suggested that the assessment of model replicability should be performed in the process of project research, and a statement of IVS should be reported at the time of submission. However, even after screening, it is still necessary to comprehensively consider other factors, such as unquantifiable AI ethics issues, due to the emphasis on assessing technical feasibility and impact in the scoring system. It is also important to emphasize that the current scoring system remains theoretical and requires practical validation and adjustment, necessitating input and refinement from numerous scholars.

Challenges and opportunities

Despite over 20 years of development, the AI field of CVD prediction experienced a surge of articles in the past 5 years, accompanied by the aforementioned phenomena regarding the emphasis on development but validation, no independent validation studies, and a large number of new algorithms studied only once. This field has been concluded as being in an early stage of development, similar to the traditional Framingham model from the 1970s to 1990s [175, 176]. Different from T-Ms, however, the AI ones are quite hard to comprehend and implement for clinical researchers owing to their complexity and “black box”. Meanwhile, there appear continually new algorithms or new combinations of the existing

(such as model averaging and stacked regressions), even there may be rather different ranking indexes in the same algorithm [160, 177]. Therefore, it is reasonable to speculate that new exploratory research will continue to dominate for the foreseeable future, which may be the inherent demand for this field, although the external validation of existing models was necessary to avoid research waste, as advocated strongly by many researchers [10, 11, 164].

Several pivotal problems limiting the development of this field still require to be emphasized again. First, the solution to study design and reporting defects, including insufficient external validation, geographical imbalance, inappropriate data sources, and deficiency in algorithm details, largely depends on improving scientific research consciousness and level of all researchers in this industry, which is a gradual process, and thereby uneven development and research waste will be difficult to stop in a short time. Second, another grim situation is how to improve model intelligibility, reproducibility, and replicability, which may far outweigh our understanding concluded in the studies of T-MS, although some researchers have been making great efforts to explore underlying mechanisms of AI operation, with the increasingly intense expectations of a revolutionary breakthrough as soon as possible [178]. Additionally, it is urgent to establish an integral system of quality control and performance evaluation for the studies in this field. However, this requires a gradual development process, although the World Health Organization (WHO) and International Telecommunication Union (ITU) have established a Focus Group on Artificial Intelligence for Health (FG-AI4H), which has begun shaping guidelines and benchmarking process for health AI models through an international, independent, and standard evaluation framework to guide and standardize the industry development [179].

In addition to the challenges posed by the “black box” issue leading to non-interpretable problems, biases and fairness, technical safety, preservation of human autonomy, privacy, and data security are significant AI ethics concerns within this field [20, 180]. The development of trustworthy AI in healthcare has become a crucial responsibility worldwide [181]. For instance, the European Commission has enacted both the “Ethics Guidelines for Trustworthy AI” and the “Artificial Intelligence Act” [182, 183]. Similarly, in the USA, the creation of the National AI Initiative Office aims to promote the development and utilization of trustworthy AI in both the public and private sectors [184]. Although the articles in this review have devoted limited discussion to these topics, it is essential to note that the aforementioned aspects (including improvement of model transparency and interpretability, reduction in bias risk, enhancement of reproducibility, as well as placing additional emphasis on

data and privacy protection), in addition to their scientific research roles, also play a crucial role in addressing AI ethics concerns. These efforts are beneficial for alleviating public concerns about AI ethics issues related to predictive models, thereby increasing trust and acceptance of the models. These aspects improve the balance between AI-assisted decision-making and the preservation of human autonomy, facilitating the clinical application and dissemination of the models. Therefore, we strongly recommend that AI ethics considerations be thoroughly integrated into the model development and validation processes.

For AI intervention studies, the relatively excellent guidelines for the design, implementation, reporting, and evaluation have been developed by the EQUATOR-network, including STARD-AI, CONSORT-AI, and SPIRIT-AI, as well as different scientific journals and associations [139, 142, 185–187]. These guidelines will also serve as a roadmap for the development of predictive AI. In practice, Banerjee et al. have designed a seven-domain, AI-specific checklist based on AHA QUADAS-2 CHARMS PROGRESS TRIPOD AI-TREE and Christodoulou, to evaluate the clinical utility and validity of predictive AI algorithms [11]. Oala et al. are building a tool of AI algorithm auditing and quality control for more effective and reliable application of ML systems in healthcare, helping to manage dynamic workflows that may vary through used case and ML technology [188]. Collins et al. have begun to develop TRIPOD-AI and PROBAST-AI for AI prediction models [21, 163, 164]. Additionally, based on the results of our IVS analysis, we are planning an independent external validation study with multiple datasets to fill the gap in AI field of CVD prediction. These will be expected to propel this field into a new and mature stage of development.

Recommendations

Despite the increasing recommendations by healthcare providers and policymakers for the use of prediction models within clinical practice guidelines to inform decision-making at various stages in the clinical pathway [161, 189], we still suggest that experts in this field should put more emphasis on establishment and implementation of scientific research guidelines, for example, promoting ML4H supervision and management for AI prediction models [188]. Additionally, referring to the requirements for intervention AI statement, some AI-relevant information should be added into TRIPOD-AI, such as algorithm formulas, hyperparameter tuning, predictive performance, interpretability, sample size determination, and so on [186, 190]. Certain items in PROBAST need to be modified for AI prediction models, especially 2.3, 4.1, and 4.9, due to inappropriate standards or nonexistent

coefficients in some algorithms. Items 4.6–4.8 should be renegotiated on the premise of fully considering the algorithm characteristics. Furthermore, algorithm auditing, overfitting control, sample size calculation, and identification of variables in image data should be added into PROBAST-AI.

In light of studies on conventional models, a greater responsibility falls upon AI algorithm developers, which include improving the transparency in reporting to facilitate model reproduction, and heightening the comprehensibility and enforceability of algorithms to users for wider clinical practice [191]. Furthermore, we should improve the transparency of reporting not only at the time of publication but also in the process of pre-submission, reviewing, or post-publication stages. Meanwhile, editors and reviewers should also play a key role in improving the quality of reporting.

Study limitations

The systematic review has several limitations. Firstly, similar to other studies [10, 11, 29], the papers not in English, without available full text, or published in other forms (for example, conferences, workshops, news reports, even the unpublished) were also excluded in our review, which may lead to an underestimation of the number of models and an imbalance in geographical contribution as mentioned above. Second, the potential impact of AI on healthcare might still be overestimated during the present procedure of retrospective literature analysis, owing to unavoidable publication bias and reporting bias, despite some measures that have been performed to reduce the omission of included literature [11, 192]. Furthermore, we did not evaluate the clinical usefulness aspects such as net benefit or impact study [159, 193, 194], which are outside our scope and require further investigation.

Conclusions

In summary, AI has triggered a promising digital revolution for CVD risk prediction. However, this field is still in its early stage, characterized by geographical imbalance, low reproducibility, a lack of independent external validation, a high risk of bias, a low standard-reaching rate of report quality, and an imperfect evaluation system. Additionally, the IVS method we designed may provide a practical tool for assessing model replicability. It is expected to contribute to independent external validation research and subsequent extensive clinical application. The development of AI CVD risk prediction may depend largely on the collaborative efforts of researchers, health policymakers, editors, reviewers, as well as quality controllers.

Abbreviations

ACC/AHA	American College of Cardiology/American Heart Association
AI-Ms	Artificial intelligence models
CT	Computed tomography
CVD	Cardiovascular disease
DL	Deep learning
ECG	Electrocardiogram
EHR	Electronic health record
ESC	European Society of Cardiology
FG-AI4H	Focus Group on Artificial Intelligence for Health
ITU	International Telecommunication Union
IVS	Independent validation score
MCC	Mattews correlation coefficient
ML	Machine learning
MRI	Magnetic resonance imaging
PPV	Positive predictive value
PROBAST	Prediction risk of bias assessment tool
RNN	Recurrent neural network
SNPs	Single nucleotide polymorphisms
T-Ms	Traditional models
TNR	True negative rate
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-024-03273-7>.

Additional file 1. PRISMA 2020 Checklist.

Additional file 2: Text 1. Search strategies for AI-Ms of CVD prediction. **Table S1.** Characteristics of the included studies. **Table S2.** Inclusion and exclusion criteria of the included studies. **Table S3.** The definition and measurement of outcomes. **Table S4.** The counting and characteristics of algorithms. **Table S5.** Risk of bias assessment of prediction models. **Text 2.** Search strategies of AI/ML assessment guidelines or tools. **Fig. S1.** The flow diagram for literature search in the assessment guidelines or tools in the field of medical AI/ML research. **Table S6.** The characteristics of assessment guidelines or tools in the field of medical AI/ML research. **Table S7.** The characteristics of 10 recommended models.

Acknowledgements

Not applicable.

Authors' contributions

YC, Y-Q C, L-YT, YH W, and WH drafted the manuscript and analyzed the data. H-J L and T-C J helped with manuscript preparation. MG, J L-L, and WH provided feedback and suggestions for improving the article. G-W Z, D-X G, and ZY participated in the final review and approval of the article. All authors have read and approved the content of the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China [No.2020YFC2006401 and 2020YFC2006406], Science and Technology Projects in Liaoning Province [2023JH2/20200056], and the National College Students' innovation and entrepreneurship training program: [No.202110159005].

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in PubMed, Web of Science, Embase, and IEEE library up to July 2021.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹China Medical University, Shenyang 110122, China. ²Digital Health China Co. Ltd, Beijing 100089, China. ³Smart Hospital Management Department, the First Hospital of China Medical University, Shenyang 110001, China. ⁴Shenyang Medical & Film Science and Technology Co. Ltd., Shenyang 110001, China. ⁵Enduring Medicine Smart Innovation Research Institute, Shenyang 110001, China. ⁶Institute of Genetic Medicine, School of Life Science, State Key Laboratory of Biotherapy, Sichuan University, Chengdu 610065, China. ⁷Bayi Orthopedic Hospital, Chengdu 610017, China. ⁸Department of Epidemiology, School of Public Health, China Medical University, Shenyang 110122, China. ⁹The Internet Hospital Branch of the Chinese Research Hospital Association, Beijing 100006, China.

Received: 16 July 2023 Accepted: 23 January 2024

Published online: 05 February 2024

References

- Group WCRWC. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health*. 2019;7(10):e1332–45.
- Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, Barengo NC, Beaton AZ, Benjamin EJ, Benziger CP, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 Study. *J Am Coll Cardiol*. 2020;76(25):2982–3021.
- Zhao D, Liu J, Xie W, Qi Y. Cardiovascular risk assessment: a global perspective. *Nat Rev Cardiol*. 2015;12(5):301–11.
- Usher-Smith JA, Silarova B, Schuit E, Moons KG, Griffin SJ. Impact of provision of cardiovascular disease risk estimates to healthcare professionals and patients: a systematic review. *BMJ Open*. 2015;5(10):e008717.
- Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837–47.
- Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24(11):987–1003.
- Roffi M, Patrono C, Collet JP, Mueller C, Valgimigli M, Andreotti F, Bax JJ, Borger MA, Brotons C, Chew DP, et al. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). *Eur Heart J*. 2016;37(3):267–315.
- Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, Himmelfarb CD, Khera A, Lloyd-Jones D, McEvoy JW, et al. 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*. 2019;140(11):e596–646.
- Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: Current status and future challenges - a systematic review. *Artif Intell Med*. 2021;120:102164.
- Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA, Collins GS, Maruthappu M. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368:m689.
- Banerjee A, Chen S, Fatemifar G, Zeina M, Lumbers RT, Mielke J, Gill S, Kotecha D, Freitag DF, Denaxas S, 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*. 2021;19(1):85.
- Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. 2021;375:n2281.
- Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
- Forcier MB, Gallois H, Mullan S, Joly Y. Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers? *J Law Biosci*. 2019;6(1):317–35.
- White DJ, Skorburg JA. Why Canada's Artificial Intelligence and Data Act Needs "Mental Data." *AJOB Neurosci*. 2023;14(2):101–3.
- Currie G, Hawk KE. Ethical and legal challenges of artificial intelligence in nuclear medicine. *Semin Nucl Med*. 2021;51(2):120–5.
- Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Comput Biol Med*. 2023;158:106848.
- Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, Matsui Y, Nozaki T, Nakaura T, Fujima N, et al. Fairness of artificial intelligence in healthcare: review and recommendations. *Jpn J Radiol*. 2024;42(1):3–15.
- Ferryman K, Mackintosh M, Ghassemi M. Considering Biased Data as Informative Artifacts in AI-Assisted Health Care. *N Engl J Med*. 2023;389(9):833–8.
- Ng MY, Kapur S, Blizinsky KD, Hernandez-Boussard T. The AI life cycle: a holistic approach to creating ethical AI for health decisions. *Nat Med*. 2022;28(11):2247–9.
- Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. *BMC Med Res Methodol*. 2022;22(1):12.
- Suri JS, Bhagawati M, Paul S, Protogerou AD, Sfrikakis PP, Kitas GD, Khanna NN, Ruzsa Z, Sharma AM, Saxena S, et al. A Powerful paradigm for cardiovascular risk stratification using multiclass, multi-label, and ensemble-based machine learning paradigms: a narrative review. *Diagnostics (Basel)*. 2022;12(3):722.
- Azmi J, Arif M, Nafis MT, Alam MA, Tanweer S, Wang G. A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Med Eng Phys*. 2022;105:103825.
- Assadi H, Alabed S, Maiter A, Salehi M, Li R, Ripley DP, Van der Geest RJ, Zhong Y, Zhong L, Swift AJ, et al. The role of artificial intelligence in predicting outcomes by cardiovascular magnetic resonance: a comprehensive systematic review. *Medicina (Kaunas)*. 2022;58(8):1087.
- Infante T, Cavaliere C, Punzo B, Grimaldi V, Salvatore M, Napoli C. Radiogenomics and artificial intelligence approaches applied to cardiac computed tomography angiography and cardiac magnetic resonance for precision medicine in coronary heart disease: a systematic review. *Circ Cardiovasc Imaging*. 2021;14(12):1133–46.
- Triantafyllidis A, Kondylakis H, Katehakis D, Kouroubali A, Koumakis L, Marias K, Alexiadis A, Votis K, Tzovaras D. Deep learning in mhealth for cardiovascular disease, diabetes, and cancer: systematic review. *JMIR Mhealth Uhealth*. 2022;10(4):e32344.
- Zhao Y, Wood EP, Mirin N, Cook SH, Chunara R. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. *Am J Prev Med*. 2021;61(4):596–605.
- Liu W, Laranjo L, Klimis H, Chiang J, Yue J, Marschner S, Quiroz JC, Jorm L, Chow CK. Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: a systematic review and meta-analysis. *Eur Heart J Qual Care Clin Outcomes*. 2023;9(4):310–22.
- Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, Lassale CM, Siontis GC, Chiochia V, Roberts C, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ*. 2016;353:i2416.
- Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, Groupdagger P. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;170(1):51–8.
- Kwong JCC, Khondker A, Lajkosz K, McDermott MBA, Frigola XB, McCadden MD, Mamdani M, Kulkarni GS, Johnson AEW. APPRAISE-AI Tool for Quantitative Evaluation of AI Studies for Clinical Decision Support. *JAMA Netw Open*. 2023;6(9):e2335377.
- Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, Arnaout R, Kohane IS, Saria S, Topol E, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med*. 2020;26(9):1320–4.

33. Vollmer S, Mateen BA, Bohner G, Kiraly FJ, Ghani R, Jonsson P, Cumbers S, Jonas A, McAllister KSL, Myles P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*. 2020;368:16927.
34. 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*. 2015;350:g7594.
35. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744.
36. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, Go AS, Harrell FE Jr, Hong Y, Howard BV, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119(17):2408–16.
37. Corbanese U. Assessing the performance of the HAS-BLED score: is the C statistic sufficient? *Chest*. 2011;139(5):1247–8.
38. Tanguay W, Acar P, Fine B, Abdollell M, Gong B, Cadrin-Chenevert A, Chartrand-Lefebvre C, Chalaoui J, Gorgos A, Chin AS, et al. Assessment of radiology artificial intelligence software: a validation and evaluation framework. *Can Assoc Radiol J*. 2023;74(2):326–33.
39. de Biase A, Sourlos N, van Ooijen PMA. Standardization of Artificial Intelligence Development in Radiotherapy. *Semin Radiat Oncol*. 2022;32(4):415–20.
40. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Spirit AI. Group C-AW: Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ*. 2020;370:m3164.
41. Cerda-Alberich L, Solana J, Mallol P, Ribas G, Garcia-Junco M, Alberich-Bayarri A, Marti-Bonmati L. MAIC-10 brief quality checklist for publications using artificial intelligence and medical images. *Insights Imaging*. 2023;14(1):11.
42. Dehghan A, Rayatinejad A, Khezri R, Aune D, Rezaei F. Laboratory-based versus non-laboratory-based World Health Organization risk equations for assessment of cardiovascular disease risk. *BMC Med Res Methodol*. 2023;23(1):141.
43. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort. *Lancet*. 2008;371(9616):923–31.
44. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, Shilton A, Yearwood J, Dimitrova N, Ho TB, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res*. 2016;18(12):e323.
45. Sujjan M, Smith-Frazier C, Malamateniou C, Connor J, Gardner A, Unsworth H, Husain H. Validation framework for the use of AI in healthcare: overview of the new British standard BS30440. *BMJ Health Care Inform*. 2023;30(1):e100749.
46. Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. *J Med Internet Res*. 2023;25:e48763.
47. Majid Akhtar M, Elliott PM. Rare Disease in Cardiovascular Medicine I. *Eur Heart J*. 2017;38(21):1625–8.
48. Majid Akhtar M, Elliott PM. Rare Diseases in Cardiovascular Medicine II. *Eur Heart J*. 2017;38(21):1629–31.
49. Perez MV, Dewey FE, Tan SY, Myers J, Froelicher VF. Added value of a resting ECG neural network that predicts cardiovascular mortality. *Ann Noninvasive Electrocardiol*. 2009;14(1):26–34.
50. Han D, Kolli KK, Gransar H, Lee JH, Choi SY, Chun EJ, Han HW, Park SH, Sung J, Jung HO, 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):168–76.
51. Ward A, Sarraju A, Chung S, Li J, Harrington R, Heidenreich P, Palaniappan L, Scheinker D, Rodriguez F. Machine learning and atherosclerotic cardiovascular disease risk prediction in a multi-ethnic population. *NPJ Digit Med*. 2020;3:125.
52. Nakanishi R, Slomka PJ, Rios R, Betancur J, Blaha MJ, Nasir K, Miedema MD, Rumberger JA, Gransar H, Shaw LJ, et al. Machine Learning Adds to Clinical and CAC Assessments in Predicting 10-Year CHD and CVD Deaths. *JACC Cardiovasc Imaging*. 2021;14(3):615–25.
53. 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*. 2018;7(22):e009476.
54. Kim J, Kang U, Lee Y. Statistics and Deep Belief Network-Based Cardiovascular Risk Prediction. *Healthc Inform Res*. 2017;23(3):169–75.
55. Cho IJ, Sung JM, Kim HC, Lee SE, Chae MH, Kavousi M, Rueda-Ochoa OL, Ikram MA, Franco OH, Min JK, et al. Development and External Validation of a Deep Learning Algorithm for Prognostication of Cardiovascular Outcomes. *Korean Circ J*. 2020;50(1):72–84.
56. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, Gomes AS, Folsom AR, Shea S, Guallar E, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res*. 2017;121(9):1092–101.
57. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS One*. 2019;14(5):e0213653.
58. 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*. 2020;371:m3919.
59. Weng SF, Reys J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12(4):e0174944.
60. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, McElhinney P, Gransar H, Cantu S, Miller RUH, 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*. 2020;116(14):2216–25.
61. Apostolopoulos ID, Groumpos PP. Non - invasive modelling methodology for the diagnosis of coronary artery disease using fuzzy cognitive maps. *Comput Methods Biomech Biomed Engin*. 2020;23(12):879–87.
62. Dogan MV, Beach SRH, 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)*. 2018;9(12):641.
63. Du Z, Yang Y, Zheng J, Li Q, Lin D, Li Y, Fan J, Cheng W, Chen XH, Cai Y. 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*. 2020;8(7):e17257.
64. Tay D, Poh CL, Kitney RI. A novel neural-inspired learning algorithm with application to clinical risk prediction. *J Biomed Inform*. 2015;54:305–14.
65. Raghu A, Praveen D, Peiris D, Tarassenko L, Clifford G. Implications of cardiovascular disease risk assessment using the WHO/ISH risk prediction charts in rural India. *PLoS One*. 2015;10(8):e0133618.
66. Bundy JD, Heckbert SR, Chen LY, Lloyd-Jones DM, Greenland P. Evaluation of Risk Prediction Models of Atrial Fibrillation (from the Multi-Ethnic Study of Atherosclerosis [MESA]). *Am J Cardiol*. 2020;125(1):55–62.
67. 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.
68. Bouzid Z, Faramand Z, Gregg RE, Frisch SO, Martin-Gill C, Saba S, Callaway C, Sejdic E, Al-Zaiti S. 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*. 2021;10(3):e017871.
69. Al-Zaiti S, Besomi L, Bouzid Z, Faramand Z, Frisch S, Martin-Gill C, Gregg R, Saba S, Callaway C, Sejdic E. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat Commun*. 2020;11(1):3966.
70. Ricciardi C, Edmunds KJ, Recenti M, Sigurdsson S, Gudnason V, Carraro U, Gargiulo P. Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. *Sci Rep*. 2020;10(1):2863.
71. Okser S, Lehtimäki T, Elo LL, Mononen N, Peltonen N, Kahonen M, Juonala M, Fan YM, Hernesniemi JA, Laitinen T, 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*. 2010;6(9):e1001146.

72. Colombet I, Ruelland A, Chatellier G, Gueyffier F, Degoulet P, Jaulent MC. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. *Proc AMIA Symp.* 2000;156–60. <https://pubmed.ncbi.nlm.nih.gov/11079864/>.
73. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. *Med Care.* 2010;48(6 Suppl):S106–113.
74. Voss R, Cullen P, Schulte H, Assmann G. Prediction of risk of coronary events in middle-aged men in the Prospective Cardiovascular Munster Study (PROCAM) using neural networks. *Int J Epidemiol.* 2002;31(6):1253–62 discussion 1262–1264.
75. Segar MW, Jaeger BC, Patel KV, Nambi V, Ndumele CE, Correa A, Butler J, Chandra A, Ayers C, Rao S, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation.* 2021;143(24):2370–83.
76. Ayala Solares JR, Canoy D, Raimondi FED, Zhu Y, Hassaine A, Salimi-Khorshidi G, Tran J, Copland E, Zottoli M, Pinho-Gomes AC, 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.* 2019;8(12):e012129.
77. Lacson RC, Baker B, Suresh H, Andriole K, Szolovits P, Lacson E Jr. Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients. *Clin Kidney J.* 2019;12(2):206–12.
78. Chang W, Liu Y, Wu X, Xiao Y, Zhou S, Cao W. A new hybrid XGBSVM model: application for hypertensive heart disease. *IEEE Access.* 2019;7:175248–58.
79. Joo G, Song Y, Im H, Park J. Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea). *IEEE Access.* 2020;8:157643–53.
80. Rao VS, Kumar MN. Novel approaches for predicting risk factors of atherosclerosis. *IEEE J Biomed Health Inform.* 2013;17(1):183–9.
81. Johri AM, Mantella LE, Jamthikar AD, Saba L, Laird JR, Suri JS. Role of artificial intelligence in cardiovascular risk prediction and outcomes: comparison of machine-learning and conventional statistical approaches for the analysis of carotid ultrasound features and intra-plaque neovascularization. *Int J Cardiovasc Imaging.* 2021;37(11):3145–56.
82. Chun M, Clarke R, Cairns BJ, Clifton D, Bennett D, Chen Y, Guo Y, Pei P, Lv J, Yu C, et al. Stroke risk prediction using machine learning: a prospective cohort study of 0.5 million Chinese adults. *J Am Med Inform Assoc.* 2021;28(8):1719–27.
83. Zhang PI, Hsu CC, Kao Y, Chen CJ, Kuo YW, Hsu SL, Liu TL, Lin HJ, Wang JJ, Liu CF, et al. Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. *Scand J Trauma Resusc Emerg Med.* 2020;28(1):93.
84. Lindholm D, Fukaya E, Leeper NJ, Ingelsson E. Bioimpedance and New-Onset Heart Failure: A Longitudinal Study of >500 000 Individuals From the General Population. *J Am Heart Assoc.* 2018;7(13):e008970.
85. Zarkogianni K, Athanasidou M, Thanopoulou AC. Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication. *IEEE J Biomed Health Inform.* 2018;22(5):1637–47.
86. Lee AK, Katz R, Jotwani V, Garimella PS, Ambrosius WT, Cheung AK, Gren LH, Neyra JA, Punzi H, Raphael KL, et al. Distinct dimensions of kidney health and risk of cardiovascular disease, heart failure, and mortality. *Hypertension.* 2019;74(4):872–9.
87. Bello GA, Dumanas GG, Gennings C. Development and validation of a clinical risk-assessment tool predictive of all-cause mortality. *Bioinform Biol Insights.* 2015;9(Suppl 3):1–10.
88. Andy AU, Guntuku SC, Adusumalli S, Asch DA, Groeneveld PW, Ungar LH, Merchant RM. Predicting cardiovascular risk using social media data: performance evaluation of machine-learning models. *JMIR Cardio.* 2021;5(1): e24473.
89. Dalakleidi K, Zarkogianni K, Thanopoulou A, et al. Comparative assessment of statistical and machine learning techniques towards estimating the risk of developing type 2 diabetes and cardiovascular complications. *Expert Syst.* 2017:e12214. <https://doi.org/10.1111/esyx.12214>.
90. Cho SY, Kim SH, Kang SH, Lee KJ, Choi D, Kang S, Park SJ, Kim T, Yoon CH, Youn TJ, et al. Pre-existing and machine learning-based models for cardiovascular risk prediction. *Sci Rep.* 2021;11(1):8886.
91. Lin A, Wong ND, Razipour A, McElhinney PA, Commandeur F, Cadet SJ, Gransar H, Chen X, Cantu S, Miller RJH, et al. Metabolic syndrome, fatty liver, and artificial intelligence-based epicardial adipose tissue measures predict long-term risk of cardiac events: a prospective study. *Cardiovasc Diabetol.* 2021;20(1):27.
92. Tesche C, Bauer MJ, Baquet M, Hedels B, Straube F, Hartl S, Gray HN, Jochheim D, Aschauer T, Rogowski 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.* 2021;31(1):486–93.
93. Priyanga P, Pattankar VV, Sridevi S. A hybrid recurrent neural network-logistic chaos-based whale optimization framework for heart disease prediction with electronic health records. *Comput Intell.* 2020;37:315–43. <https://api.semanticscholar.org/CorpusID:224845329>.
94. Dutta A, Batabyal T, Basu M, Acton ST. An efficient convolutional neural network for coronary heart disease prediction. *Expert Syst Appl.* 2020;159:113408.
95. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. *JAMA Netw Open.* 2020;3(1):e1919396.
96. Jiang Y, Zhang X, Ma R, Wang X, Liu J, Keerman M, Yan Y, Ma J, Song Y, Zhang J, et al. Cardiovascular disease prediction by machine learning algorithms based on cytokines in Kazakhs of China. *Clin Epidemiol.* 2021;13:417–28.
97. Jamthikar A, Gupta D, Saba L, Khanna NN, Araki T, Viskovic K, Mavrogeni S, Laird JR, Pareek G, Miner M, et al. Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovasc Diagn Ther.* 2020;10(4):919–38.
98. Ngufor C, Caraballo PJ, O'Byrne TJ, Chen D, Shah ND, Pruinelli L, Steinbach M, Simon G. Development and validation of a risk stratification model using disease severity hierarchy for mortality or major cardiovascular event. *JAMA Netw Open.* 2020;3(7):e208270.
99. Zhang Y, Han Y, Gao P, Mo Y, Hao S, Huang J, Ye F, Li Z, Zheng L, Yao X, et al. Electronic health record-based prediction of 1-year risk of incident cardiac dysrhythmia: prospective case-finding algorithm development and validation study. *JMIR Med Inform.* 2021;9(2):e23606.
100. Hong D, Fort D, Shi L, Price-Haywood EG. Electronic Medical Record Risk Modeling of Cardiovascular Outcomes Among Patients with Type 2 Diabetes. *Diabetes Ther.* 2021;12(7):2007–17.
101. Ogata K, Miyamoto T, Adachi H, Hirai Y, Enomoto M, Fukami A, Yokoi K, Kasahara A, Tsukagawa E, Yoshimura A, et al. New computer model for prediction of individual 10-year mortality on the basis of conventional atherosclerotic risk factors. *Atherosclerosis.* 2013;227(1):159–64.
102. 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.* 2014;9(1):82–91.
103. Puddu PE, Menotti A. Artificial neural networks versus proportional hazards Cox models to predict 45-year all-cause mortality in the Italian Rural Areas of the Seven Countries Study. *BMC Med Res Methodol.* 2012;12:100.
104. Betancur J, Otaki Y, Motwani M, Fish MB, Lemley M, Dey D, Gransar H, Tamarappoo B, Germano G, Sharir T, et al. Prognostic Value of Combined Clinical and Myocardial Perfusion Imaging Data Using Machine Learning. *JACC Cardiovasc Imaging.* 2018;11(7):1000–9.
105. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* 2017;38(7):500–7.
106. VanHouten JP, Starmer JM, Lorenzi NM, Maron DJ, Lasko TA. Machine learning for risk prediction of acute coronary syndrome. *AMIA Annu Symp Proc.* 2014;2014:1940–9.
107. Sanchez-Cabo F, Rossello X, Fuster V, Benito F, Manzano JP, Silla JC, Fernandez-Alvira JM, Oliva B, Fernandez-Friera L, Lopez-Melgar B, et al. Machine learning improves cardiovascular risk definition for young, asymptomatic individuals. *J Am Coll Cardiol.* 2020;76(14):1674–85.
108. Wu Y, Fang Y. Stroke prediction with machine learning methods among older Chinese. *Int J Environ Res Public Health.* 2020;17(6):1828.
109. Christopoulos G, Graff-Radford J, Lopez CL, Yao X, Attia ZI, Rabinstein AA, Petersen RC, Knopman DS, Mielke MM, Kremers W, et al. Artificial

- intelligence-electrocardiography to predict incident atrial fibrillation: a population-based study. *Circ Arrhythm Electrophysiol.* 2020;13(12):e009355.
110. Hoogeveen RM, Pereira JPB, Nurmohamed NS, Zampoleri V, Bom MJ, Baragetti A, Boekholdt SM, Knaapen P, Khaw KT, Wareham NJ, et al. Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention. *Eur Heart J.* 2020;41(41):3998–4007.
 111. Orfanoudaki A, Chesley E, Cadisch C, Stein B, Nouh A, Alberts MJ, Bertsimas D. Machine learning provides evidence that stroke risk is not linear: the non-linear framingham stroke risk score. *PLoS ONE.* 2020;15(5):e0232414.
 112. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, Wu H, Xu H, Zhi D. 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.* 2018;84:11–6.
 113. Nowak C, Carlsson AC, Ostgren CJ, Nystrom FH, Alam M, Feldreich T, Sundstrom J, Carrero JJ, Leppert J, Hedberg P, et al. Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia.* 2018;61(8):1748–57.
 114. Dimopoulos AC, Nikolaidou M, Caballero FF, Engchuan W, Sanchez-Niubo A, Arndt H, Ayuso-Mateos JL, Haro JM, Chatterji S, Georgousopoulou EN, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Med Res Methodol.* 2018;18(1):179.
 115. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, Denny JC, Wei WQ. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep.* 2019;9(1):717.
 116. Suzuki S, Yamashita T, Sakama T, Arita T, Yagi N, Otsuka T, Semba H, Kano H, Matsuno S, Kato Y, et al. Comparison of risk models for mortality and cardiovascular events between machine learning and conventional logistic regression analysis. *PLoS One.* 2019;14(9):e0221911.
 117. Mezzatesta S, Torino C, Meo P, Fiumara G, Vilasi A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Comput Methods Programs Biomed.* 2019;177:9–15.
 118. Sung JM, Cho IJ, Sung D, Kim S, Kim HC, Chae MH, Kavousi M, Rueda-Ochoa OL, Ikram MA, Franco OH, et al. Development and verification of prediction models for preventing cardiovascular diseases. *PLoS One.* 2019;14(9):e0222809.
 119. Quesada JA, Lopez-Pineda A, Gil-Guillen VF, Durazo-Arvizu R, Orozco-Beltran D, Lopez-Domenech A, Carratala-Munuera C. Machine learning to predict cardiovascular risk. *Int J Clin Pract.* 2019;73(10):e13389.
 120. Grout RW, Hui SL, Imler TD, El-Azab S, Baker J, Sands GH, Ateya M, Pike F. Development, validation, and proof-of-concept implementation of a two-year risk prediction model for undiagnosed atrial fibrillation using common electronic health data (UNAFIED). *BMC Med Inform Decis Mak.* 2021;21(1):112.
 121. Sajeev S, Champion S, Beleigoli A, Chew D, Reed RL, Magliano DJ, Shaw JE, Milne RL, Appleton S, Gill TK, 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.* 2021;18(6):3187.
 122. de Gonzalo-Calvo D, Martinez-Cambolor P, Bar C, Duarte K, Girerd N, Fellstrom B, Schmieder RE, Jardine AG, Massy ZA, Holdaas H, et al. Improved cardiovascular risk prediction in patients with end-stage renal disease on hemodialysis using machine learning modeling and circulating microribonucleic acids. *Theranostics.* 2020;10(19):8665–76.
 123. Kim IS, Yang PS, Jang E, Jung H, You SC, Yu HT, Kim TH, Uhm JS, Pak HN, Lee MH, et al. Long-term PM(2.5) exposure and the clinical application of machine learning for predicting incident atrial fibrillation. *Sci Rep.* 2020;10(1):16324.
 124. Schrepf M, Kramer D, Jauk S, Veeranki SPK, Leodolter W, Rainer PP. Machine learning based risk prediction for major adverse cardiovascular events. *Stud Health Technol Inform.* 2021;279:136–43.
 125. Nusinovici S, Tham YC, Chak Yan MY, Wei Ting DS, Li J, Sabanayagam C, Wong TY, Cheng CY. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol.* 2020;122:56–69.
 126. Navarini L, Sperti M, Currado D, Costa L, Deriu MA, Margiotta DPE, Tasso M, Scarpa R, Afeltra A, Caso F. A machine-learning approach to cardiovascular risk prediction in psoriatic arthritis. *Rheumatology (Oxford).* 2020;59(7):1767–9.
 127. 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.* 2020;20(1):252.
 128. Lennerz JK, Salgado R, Kim GE, Sirintrapun SJ, Thierauf JC, Singh A, Indave I, Bard A, Weissinger SE, Heher YK, et al. Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML. *Clin Chem Lab Med.* 2023;61(4):544–57.
 129. 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).* 2023;25(10):1427.
 130. Kocak B, Baessler B, Bakas S, Cuocolo R, Fedorov A, Maier-Hein L, Mercado N, Muller H, Orhac F, Pinto Dos Santos D, et al. CheckList for Evaluation of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMI. *Insights Imaging.* 2023;14(1):75.
 131. van Smeden M, Heinze G, Van Calster B, Asselbergs FW, Vardas PE, Bruining N, de Jaegere P, Moore JH, Denaxas S, Boulesteix AL, et al. Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease. *Eur Heart J.* 2022;43(31):2921–30.
 132. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, Guitera P, Gutman D, Halpern A, Helba B, 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.* 2022;158(1):90–6.
 133. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, Faes L, Geerts B, Ibrahim M, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med.* 2022;28(5):924–33.
 134. Jha AK, Bradshaw TJ, Buvat I, Hatt M, Kc P, Liu C, Obuchowski NF, Saboury B, Slomka PJ, Sunderland JJ, et al. Nuclear Medicine and Artificial Intelligence: Best Practices for Evaluation (the RELAINCE Guidelines). *J Nucl Med.* 2022;63(9):1288–99.
 135. Walsh I, Fishman D, Garcia-Gasulla D, Titma T, Pollastri G, Group EMLF, Harrow J, Psomopoulos FE, Tosatto SCE. DOME: recommendations for supervised machine learning validation in biology. *Nat Methods.* 2021;18(10):1122–7.
 136. Olczak J, Pavlopoulos J, Prijs J, Ijpmma FA, Doornberg JN, Lundstrom C, Hedlund J, Gordon M. 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.* 2021;92(5):513–25.
 137. Matschinske J, Alcaraz N, Benis A, Golebiewski M, Grimm DG, Heumos L, Kacprowski T, Lazareva O, List M, Louadi Z, et al. The AIME registry for artificial intelligence in biomedical research. *Nat Methods.* 2021;18(10):1128–31.
 138. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J. network le-oh, the ITUWHOfgAlfH: Artificial intelligence in dental research: Checklist for authors, reviewers, readers. *J Dent.* 2021;107:103610.
 139. Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health Care Inform.* 2021;28(1):e100251.
 140. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, Cumbers S, Jonas A, McAllister KSL, Myles P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ.* 2020;368:l6927. <https://doi.org/10.1136/bmj.l6927>.
 141. Sengupta PP, Shrestha S, Berthon B, Messas E, Donal E, Tison GH, Min JK, D'Hooge J, Voigt JU, Dudley J, 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.* 2020;13(9):2017–35.
 142. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, Spirit AI. Group C-AW: Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ.* 2020;370:m3210.
 143. Kakarmath S, Esteva A, Arnaout R, Harvey H, Kumar S, Muse E, Dong F, Wedlund L, Kvedar J. Best practices for authors of healthcare-related artificial intelligence manuscripts. *NPJ Digit Med.* 2020;3:134.

144. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes*. 2020;13(10):e006556.
145. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749–62.
146. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
147. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103.
148. Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest*. 2009;27(3):235–43.
149. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak*. 2006;6:38.
150. Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ*. 2012;344:e3318.
151. Qureshi NQ, Mufarrih SH, Bloomfield GS, Tariq W, Almas A, Mokdad AH, Bartlett J, Nisar I, Siddiqi S, Bhutta Z, et al. Disparities in Cardiovascular Research Output and Disease Outcomes among High-, Middle- and Low-Income Countries - An Analysis of Global Cardiovascular Publications over the Last Decade (2008–2017). *Glob Heart*. 2021;16(1):4.
152. Timmis A, Vardas P, Townsend N, Torbica A, Katus H, De Smedt D, Gale CP, Maggioni AP, Petersen SE, Huculeci R, et al. European Society of Cardiology: cardiovascular disease statistics 2021. *Eur Heart J*. 2022;43(8):716–99.
153. Peiris D, Ghosh A, Manne-Goehler J, Jaacks LM, Theilmann M, Marcus ME, Zhumadilov Z, Tsabedze L, Supiyev A, Silver BK, et al. Cardiovascular disease risk profile and management practices in 45 low-income and middle-income countries: a cross-sectional study of nationally representative individual-level survey data. *PLoS Med*. 2021;18(3):e1003485.
154. Cardiovascular diseases (CVDs) <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>.
155. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery CL, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570(7762):514–8.
156. Collins GS, Moons KG. Comparing risk prediction models. *BMJ*. 2012;344:e3186.
157. Carresi C, Scicchitano M, Scarano F, Macri R, Bosco F, Nucera S, Ruga S, Zito MC, Mollace R, Guarnieri L, et al. The Potential Properties of Natural Compounds in Cardiac Stem Cell Activation: Their Role in Myocardial Regeneration. *Nutrients*. 2021;13(1):275.
158. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–73.
159. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
160. Debray TP, Koffijberg H, Nieboer D, Vergouwe Y, Steyerberg EW, Moons KG. Meta-analysis and aggregation of multiple published prediction models. *Stat Med*. 2014;33(14):2341–62.
161. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76–86.
162. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567–86.
163. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, 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*. 2021;11(7):e048008.
164. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. *Lancet*. 2019;393(10181):1577–9.
165. Akyea RK, Leonardi-Bee J, Asselbergs FW, Patel RS, Durrington P, Wierzbicki AS, Ibiwoye OH, Kai J, Qureshi N, Weng SF. Predicting major adverse cardiovascular events for secondary prevention: protocol for a systematic review and meta-analysis of risk prediction models. *BMJ Open*. 2020;10(7):e034564.
166. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med*. 2021;47(7):750–60.
167. Gallifant J, Zhang J, Del Pilar Arias Lopez M, Zhu T, Camporota L, Celi LA, Formenti F. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. *Br J Anaesth*. 2022;128(2):343–51.
168. Li B, Feridooni T, Cuen-Ojeda C, Kishibe T, de Mestral C, Mamdani M, Al-Omran M. Machine learning in vascular surgery: a systematic review and critical appraisal. *NPJ Digit Med*. 2022;5(1):7.
169. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, Garcia-Pedrero A, Ramirez SC, Kong D, Moody AR, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Can Assoc Radiol J*. 2019;70(4):344–53.
170. Ma J, Fong SH, Luo Y, Bakkenist CJ, Shen JP, Mourragui S, Wessels LFA, Hafner M, Sharan R, Peng J, et al. Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nat Cancer*. 2021;2(2):233–44.
171. Su X, Xu Y, Tan Z, Wang X, Yang P, Su Y, Jiang Y, Qin S, Shang L. Prediction for cardiovascular diseases based on laboratory data: An analysis of random forest model. *J Clin Lab Anal*. 2020;34(9):e23421.
172. Nunez JJ, Nguyen TT, Zhou Y, Cao B, Ng RT, Chen J, Frey BN, Milev R, Muller DJ, Rotzinger S, et al. Replication of machine learning methods to predict treatment outcome with antidepressant medications in patients with major depressive disorder from STAR*D and CAN-BIND-1. *PLoS One*. 2021;16(6):e0253023.
173. Pan Z, Zhang R, Shen S, Lin Y, Zhang L, Wang X, Ye Q, Wang X, Chen J, Zhao Y, et al. OWL: an optimized and independently validated machine learning prediction model for lung cancer screening based on the UK Biobank, PLCO, and NLSST populations. *EBioMedicine*. 2023;88:104443.
174. Alfieri F, Ancona A, Tripepi G, Randazzo V, Paviglianiti A, Pasero E, Vecchi L, Politi C, Cauda V, Fagugli RM. External validation of a deep-learning model to predict severe acute kidney injury based on urine output changes in critically ill patients. *J Nephrol*. 2022;35(8):2047–56.
175. Sheridan S, Pignone M, Mulrow C. Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians. *J Gen Intern Med*. 2003;18(12):1039–52.
176. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*. 2014;383(9921):999–1008.
177. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG. Cochrane IPDm-aMg: Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med*. 2015;12(10):e1001886.
178. Wang ZJ, Turko R, Shaikh O, Park H, Das N, Hohman F, Kahng M, Polo Chau DH. CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Trans Vis Comput Graph*. 2021;27(2):1396–406.
179. Wiegand T, Krishnamurthy R, Kuglitsch M, Lee N, Pujari S, Salathe M, Wenzel M, Xu S. WHO and ITU establish benchmarking process for artificial intelligence in health. *Lancet*. 2019;394(10192):9–11.
180. Karimian G, Petelos E, Evers SMAA. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI Ethics*. 2022;2:539–51.
181. Radclyffe C, Ribeiro M, Wortham RH. The assessment list for trustworthy artificial intelligence: A review and recommendations. *Front Artif Intell*. 2023;6:1020592.
182. Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. <https://www.eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A52021PC0206&from=EN>.
183. Directorate-General for Communications Networks, Content and Technology (2019) Ethics guidelines for trustworthy AI <https://www.data.europa.eu/doi/10.2759/177365>.

184. Advancing Trustworthy AI Initiative. <https://www.ai.gov/strategic-pillars/advancing-trustworthy-ai/>.
185. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, Karthikesalingam A, King D, Liu X, Markar SR, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med*. 2020;26(6):807–8.
186. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Spirit AI. Group C-AW: Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364–74.
187. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc*. 2020;27(12):2011–5.
188. Oala L, Murchison AG, Balachandran P, Choudhary S, Fehr J, Leite AW, Goldschmidt PG, Johnner C, Schorverth EDM, Nakasi R, et al. Machine Learning for Health: Algorithm Auditing & Quality Control. *J Med Syst*. 2021;45(12):105.
189. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res*. 2008;14(19):5977–83.
190. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? *J Am Med Inform Assoc*. 2019;26(12):1651–4.
191. Pencina MJ, Goldstein BA, D'Agostino RB. Prediction models - development, evaluation, and clinical application. *N Engl J Med*. 2020;382(17):1583–6.
192. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, Hayden J, Collins GS, Debray TPA. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ*. 2019;364:k4597.
193. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–74.
194. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.