BMC Cardiovascular Disorders

SYSTEMATIC REVIEW Open Access

# Comparison of machine learning models with conventional statistical methods for prediction of percutaneous coronary intervention outcomes: a systematic review and meta-analysis

Sepehr Nayebirad[1*†], Ali Hassanzadeh[2†], Amir Mohammad Vahdani[3], Aida Mohamadi[4], Shayan Forghani[3], Akbar Shafiee[1,5] and Farzad Masoudkabir[1,5]

## Abstract

**Introduction** Percutaneous coronary intervention (PCI) has been the main treatment of coronary artery disease (CAD). In this review, we aimed to compare the performance of machine learning (ML) vs. logistic regression (LR) models in predicting different outcomes after PCI.

**Methods** Studies using ML or deep learning (DL) models to predict mortality, MACE, in-hospital bleeding, and acute kidney injury (AKI) after PCI or primary PCI were included. Articles were excluded if they did not provide a c-statistic, solely used ML models for feature selection, were not in English, or only used logistic or LASSO regression models. Best-performing ML and LR-based models (LR model or conventional risk score) from the same studies were pooled separately to directly compare the performance of ML versus LR. Risk of bias was assessed using the PROBAST and CHARMS checklists.

**Results** A total of 59 studies were included. Meta-analysis showed that ML models resulted in a higher c-statistic compared to LR in long-term mortality (0.84 vs. 0.79, *P*-value = 0.178), short-term mortality (0.91 vs. 0.85, *P* = 0.149), bleeding (0.81 vs. 0.77 *P* = 0.261), acute kidney injury (AKI; 0.81 vs. 0.75, *P* = 0.373), and major adverse cardiac events (MACE; 0.85 vs. 0.75, *P* = 0.406). PROBAST analysis showed that 93% of long-term mortality, 70% of short-term mortality, 89% of bleeding, 69% of AKI, and 86% of MACE studies had a high risk of bias.

**Conclusion** No statistical significance existed between ML and LR model. In addition, the high risk of bias in ML studies and complexity in interpretation undermines their validity and may impact their adaption in a clinical settings.

†Sepehr Nayebirad and Ali Hassanzadeh contributed equally to this work.

*Correspondence:
Sepehr Nayebirad
nayebisepehr76@gmail.com

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

**Significance**

**What is already known on this topic** Many ML models are available for predicting adverse complications of PCI. However some methodology and performance concerns made it hard to choose between well-established statistical vs. ML models.

**What this study adds** The overall ML and LR models c-statistics were comparable for short- and long-term mortality, bleeding, AKI, and MACE prediction. Our risk of bias assessment (using CHARMS and PROBAST checklists) identified a high risk of bias and applicability concerns.

**How this study might affect research, practice, or policy** Future studies should consider the reporting checklists to improve their methodology.

**Keywords** Percutaneous coronary intervention, Machine learning, Logistic regression, Mortality, Major adverse cardiac events, Acute kidney injury, bleeding

## Introduction

Percutaneous coronary intervention (PCI) has been the mainstay in the treatment of coronary artery disease (CAD) since it was introduced in 1977 [1]. Despite several advances in PCI technology, post-procedural complications such as acute kidney injury, bleeding, and mortality are not uncommon [2]. As such, several prediction models like the United States National Cardiovascular Data Registry Risk Score (NCDR-CathPCI risk score) [3, 4], Mehran Score [5, 6], and New York State Risk Score [7] have been developed to identify high-risk patients. Prediction models have also been used to assess patient prognosis. For instance, SYNTAX score II has been utilized in predicting long-term mortality after PCI [8]. All of these models have been based on conventional statistical methods like logistic regression (LR).

In contrast to statistical methods such as LR, ML refers to a set of computational techniques that automatically learn patterns from data to make predictions or decisions, rather than relying solely on explicitly programmed instructions. Unlike traditional statistical methods that often focus on hypothesis testing and inferring relationships between variables, machine learning is primarily concerned with prediction accuracy and pattern recognition [9].

While LR models the relationship between predictors and a binary outcome using an interpretable formula, ML models such as random forests or neural networks/deep learning models can capture complex, non-linear relationships in the data. This allows these models to potentially identify subtle interactions among variables that might be missed by traditional approaches. However, these advantages come with challenges: ML models may require larger datasets, careful tuning to avoid overfitting, and sometimes yield models that are less immediately interpretable than logistic regression [9]. There is also a widespread, difficult to navigate gap between achieving good performance metrics in internal/external validation (for ML models) and delivering clinical utility. Most studies on ML models in medicine use retrospective data, which limits the validity of their evidence; the adoption of randomized clinical trials as the gold standard for evaluating clinical utility is also lagging behind for ML models [9].

With the increasing adoption of artificial intelligence methods (including ML), the cardiology field as with many other areas of medicine has been promised better predictive accuracy, especially in scenarios with big complex datasets and non-linear relationships between the variables [10]. This has inevitably resulted in a huge surge in the number of papers using machine learning (ML) models to predict post-PCI complications and adverse events [2, 11]. However, as journals may lean towards accepting articles with better predictive performance (higher c-statistic), many ML models may inadvertently have overfitting problems caused by inappropriate methodology [12]. In addition, it is still unclear whether we should move on from well-established statistical models to ML in clinical practice and prognosis assessment of PCI patients. This makes a systematic review and critical appraisal of the literature imperative. Therefore, in the current investigation, we aimed to (1) critically review the available studies that used ML prediction models for post-PCI outcomes and (2) compare the pooled estimates of ML models and conventional risk scores or LR whenever possible.

## Methods

The protocol for this systematic review and meta-analysis was registered in the international prospective register of systematic reviews (CRD42023494659). PRISMA 2020 statement was used for reporting this systemtic review and meta-analysis [13].

### Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

### Eligibility criteria and outcome definition

Studies using ML or deep learning (DL) models to predict mortality, MACE, in-hospital bleeding, and acute kidney injury (AKI) after PCI or primary PCI were deemed eligible. Due to the small number of studies, no exclusion criteria were set for CAD type. However, studies evaluating patients with chronic total occlusion were not included. Articles were also excluded if they did not provide a c-statistic, solely used ML models for feature selection, were not in English, or only used logistic or LASSO regression models. Studies using similar datasets were included for risk of bias assessment, but only the investigation with a higher number of patients was considered in the meta-analysis.

Short-term mortality were considered as < 1 year follow-up, while long-term mortality were identified as ≥ 1 year of follow-up. MACE was defined if it was a combination of at least three of the following five components: death, myocardial infarction, coronary revascularization, stroke, and hospitalization because of heart failure. No prior definitions of bleeding and AKI were used to screen articles as studies used different criteria.

### Search strategy

The search for the current study was conducted on PubMed, Embase, Web of Science, and Scopus from inception until December 11th, 2023. The search strategy comprised two components: (1) "machine learning" AND (2) "percutaneous coronary intervention". The full search template is available in online Supplementary File S1.

### Selection process and data gathering

Two independent reviewers (A.V and S.N) screened the articles first based on their title/abstract and then based on the full text. Inconsistencies were solved by consensus. Similarly, data collection was done independently by A.H and A.M.

### Risk of Bias assessment

Critical appraisal and data extraction for systematic reviews of prediction modeling studies (CHARMS) checklist [14] and prediction model risk of bias assessment tool (PROBAST) [15] were used by two independent reviewers (S.N and A.H) to assess all of the included studies. The CHARMS and PROBAST Excel template developed by Fernandez-Felix et al. was utilized for this purpose [16]. Any discrepancies between the reviewers were solved by consensus. The assessment was only done for the best-performing ML model on the validation dataset.

### Data analysis

C-statistics, the area under the receiver operating curves (AUC ROCs), were pooled using random effects meta-analysis. If the corresponding 95% confidence interval was not provided, it was calculated using the number of events and sample size based on the methods proposed by Hanley and McNeil [17]. Best-performing ML and LR-based models (LR model or conventional risk score) from the same studies were pooled separately to directly compare the performance of ML versus LR. This also ensured that heterogeneity stemming from study methodology and population would be limited. Furthermore, we performed secondary comparisons based on whether the LR models had the same number and type of features that ML models had. The pooled estimates were compared using the MedCalc online calculator which is based on the Hanley and McNeil method [17, 18]. All analysis was performed using R statistical software version 4.2.1 and metamisc package [19].

## Results

Overall, 59 studies were included in the current systemic review from which 15 were on long-term mortality [20–34], 25 on short-term mortality [2, 10, 32, 33, 35–55], nine on bleeding [2, 10, 49, 51, 52, 56–59], 16 on AKI [2, 10, 30, 51, 52, 60–70], and seven on MACE [41, 71–76] (Fig. 1). Excluded articles in the full-text screening and reasons for exclusion are provided in the Supplementary Material 1.

### Long-term mortality

Supplementary Table S1-2 (Supplementary Material 2) demonstrate the general characteristics of the studies. In brief, fifteen articles assessed the performance of ML on long-term mortality of which seven (46%) were included in the meta-analysis [20, 23, 24, 27, 29, 32, 34]. 40% (6/15) of the studies did not report an event per variable (EPV) [22–24, 31, 33, 34], while in the others this figure was from 1.1 to 28.8 with only one study [26] having an EPV > 10. No studies used multiple imputations for handling the missing values and 53% (8/15) studies did not report their methods for missing data [22, 24–27, 29, 31, 32]. Only 40% (6/15) studies reported model calibration [20, 25–29], and 26% (4/15) had an external validation dataset [20, 26, 28, 29].

One study (6%) had a low risk of bias [20], whereas all other studies, 93% (14/15), had a high risk of bias [21–34]. The majority of the risk of bias, 93%, was stemming from the analysis domain. Regarding applicability, 40% (6/15) had a low concern [22, 25, 26, 29, 30, 34], 46% (7/15) had a high concern [20, 21, 23, 24, 27, 28, 31], and 13% (2/15) were unclear [32, 33]. The detailed risk of bias assessment using CHARMS and PROBAST is provided in Excel in supplementary material 4 and Supplementary Figure S1 (Supplementary Material 3)

Meta-analysis showed that ML models resulted in a 5% higher c-statistic compared to LR (0.84 vs. 0.79,

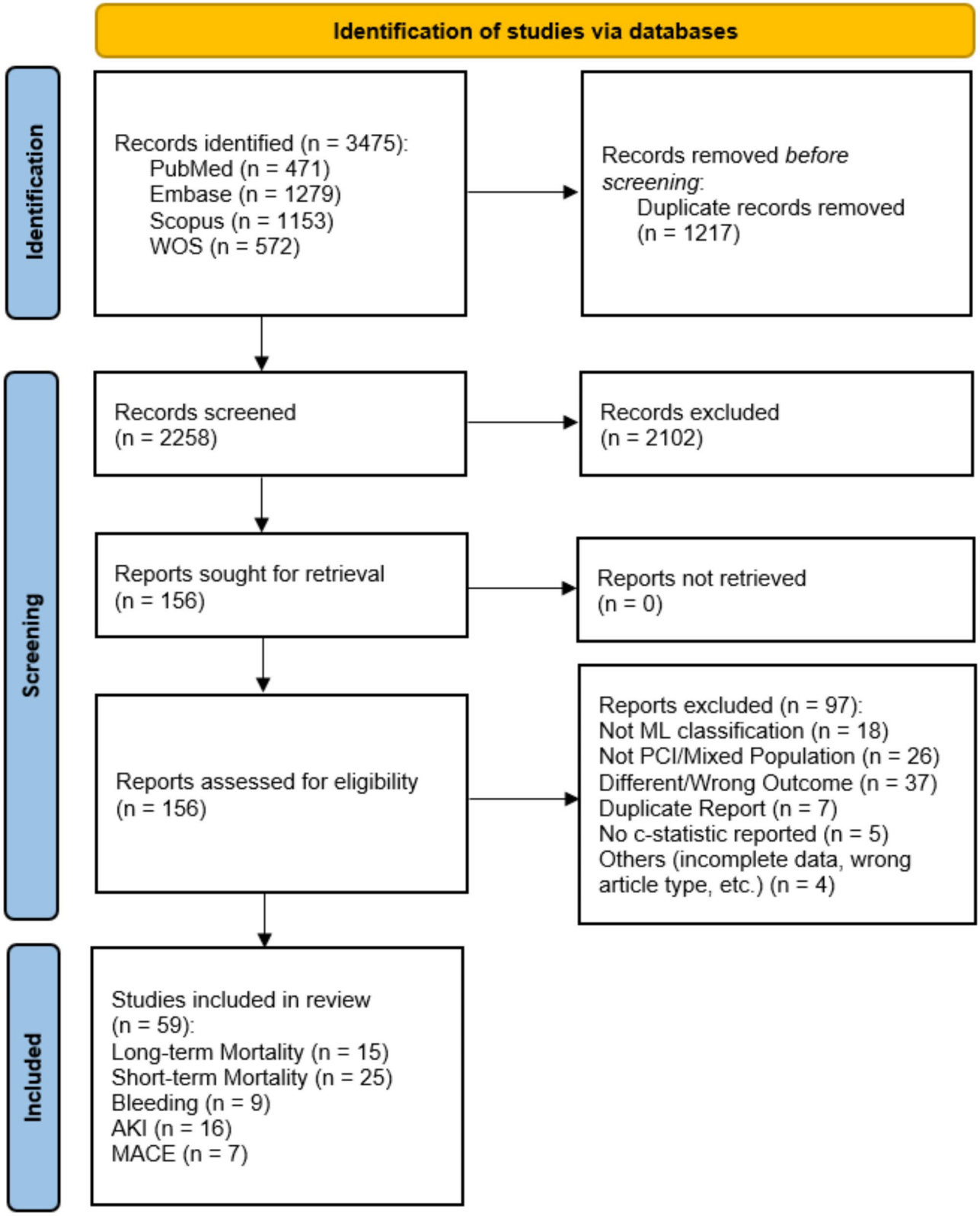


**Fig. 1** PRISMA 2020 flow diagram of study selection

*P*-value = 0.178). This number was 3% when comparing similar features ML and LR, and 6% for different features ML and LR (0.83 vs. 0.77, *P* = 0.230). However, these differences were not statistically significant (Fig. 2, Supplementary Figure S2-3). Assessment of the funnel plots revealed no asymmetry (Supplementary Figure S4).

#### Short-term mortality

Twenty-five studies which had developed models for predicting short-term mortality as an adverse effect of PCI were included in our review, 10 of which were assessed in the meta-analysis [2, 32, 36, 37, 40, 41, 43, 48, 53, 54]. EPV values were in the range 0.4–52, and only 5 studies had an EPV of > 10 [2, 40, 51, 52, 54]. A single study [54] utilized multiple imputation for handling missing data, 3 studies utilized single imputation [39, 50, 52], 8 studies utilized other procedures or provided unclear explanations on their approach to missing data [2, 10, 33, 40, 42, 47, 48, 55], and 13 studies didn't provide any information on their approach to missing data [32, 35–38, 41, 43–46, 49, 51, 53]. (Supplementary Table S3-4).

Risk of bias for these studies is evaluated in supplementary material 5 and Supplementary Figure S4.

Pooled c-statistics were overall 6% higher for ML models vs. statistical models (Fig. 2); specifically 5% higher in the case of models with different numbers of features (Supplementary Figure S5), and 6% higher in the case of models with a similar number of features (Supplementary Figure S6). Notably, none of these differences were statistically significant (Table 1).

#### Bleeding

Nine studies evaluated in-hospital bleeding with an average age of 60 to 77. In 33% (3/9) [49, 56, 57], the EPV was unknown. 11% (1/9) of the investigations used multiple imputations for handling the missing data [56] and 22% (2/9) used single imputation [2, 57]. No studies had an external validation dataset, while only one study (11%) reported results using cross-validation [57]. (Supplementary Table S5-6).

Risk of bias is available at Supplementary Material 6 and Supplementary Figure S7. Meta-analysis results showed a 4% net benefit for the ML models over LR without statistical significance (0.81 vs. 0.77, *P* = 0.261, Fig. 3).

#### AKI

We included 16 ML studies of AKI prediction after PCI (Mean age: 62.5–70). The best-performing models had a c-statistic of 0.74–0.89. One study [61] used multiple and five studies [2, 30, 62, 63, 66] used single imputation methods for handling the missing data, half of the included studies (8/16) did not clearly state their approach to the missing data [51, 52, 60, 64, 65, 68–70]. Five studies (31%) validated their models on external datasets [61, 63, 65–67]. Eight studies examined the models' calibration (50%) [2, 10, 62–67]. Five studies (31%) did not perform feature selection [2, 10, 60, 63, 65], five (31%) had insufficient or unclear information on choosing final variables [51, 52, 61, 64, 69], and one study (6%) used stepwise method [70]. Five studies reported EPV [41, 71–73, 76], and only one study had EPV > 10 [71] (Supplementary Table S7-8).

Ris of bias is demonstrated in Supplementary Material 7 and Supplementary Figure S8. Four studies were included in the meta-analysis to compare the overall ML and LR performance [2, 61, 65, 68], and the pooled c-statistics of ML and LR models were comparable (0.81 vs. 0.75, *P* = 0.373, Fig. 3). Our secondary analysis also reached almost identical results that there was no significant difference between ML and LR models, when different (0.78 vs. 0.73, *P* = 0.337) or similar features (0.81 vs. 0.75, *P* = 0.373) were used in the model development (Supplementary Figure S9-10).

#### MACE

Seven studies developed and validated prediction models of MACE in PCI patients. The c-statistics of the best ML ranged from 0.7 to 0.95 (Mean age: 60–69). Six (86%) studies developed prediction models on one to ten-year MACE [71–76], and one study investigated the in-hospital MACE [41]. Four studies embedded all of the candidate predictors into the final model [41, 71, 73, 74], one used the stepwise selection method [75], and two used random forest-based algorithms [72, 76]. Only two studies used external validation for their models [41, 76], and two measured calibration of the models [41, 76]. Two studies developed survival random forest models for MACE models [72, 73] (Supplementary Table S9-10).

Supplementary Material 8 and Supplementary Figure S11 show the risk of bias evaluation. Four studies were examined in the ML vs. LR meta-analysis [71, 73, 74, 76]. The pooled c-statistic of ML models were comparable to LR models (0.85 vs. 0.75, *P* = 0.406, Fig. 3).

## Discussion

To our knowledge, this was the first systematic review and meta-analysis of the ML models in PCI for CAD patients. Our results revealed that ML models had a net benefit over LR in several outcomes including, mortality, MACE, AKI, and bleeding after PCI, however, there was no statistically significant difference.

The risk of bias analysis of the included studies identified multiple concerns. Several studies did not provide an external validation dataset which could result in overfitting. Overfitting happens when the model has memorized the training data too well including the noise rather than identifying the patterns. As internal validation has a similar source to training data, results in the internal

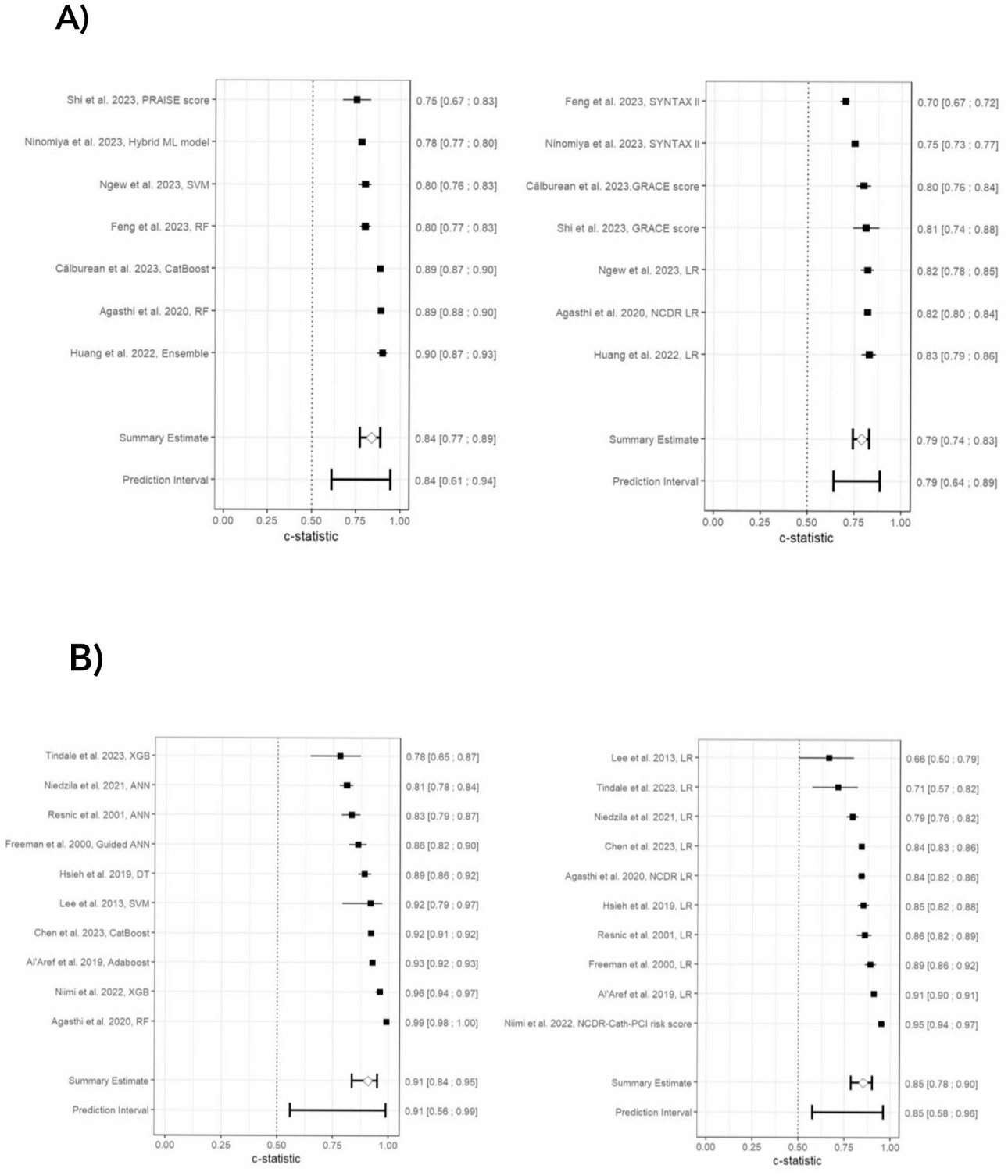


**Fig. 2** (**A**) Pooled c-statistic of the best performing ML (left) vs. LR/ risk scores (right) models for long-term mortality. (**B**) Pooled c-statistic of the best performing ML (left) vs. LR/ risk scores (right) models for short-term mortality

validation may be too optimistic. Therefore, PROBAST recommends against the use of simple data splitting into train and validation sets, while encouraging using cross-validation [77–79]. However, even simple cross-validation in ML prediction studies may be problematic as studies risk data leakage. This is because many ML models require fine-tuning hyperparameters during the cross-validation, resulting in finding the optimal hyperparameters for the

**Table 1** Comparison of the pooled c-statistics of the ML and LR models

| Type of outcome | No. of studies | ML pooled c-statistic | LR pooled c-statistic | Benefit (%) | *P*-value |
|---|---|---|---|---|---|
| **Long-term mortality (≥ 1 year)** | | | | | |
| Overall ML vs. LR | 7 | 0.84 (0.77–0.89) | 0.79 (0.74–0.83) | 5 | 0.178 |
| Similar features ML vs. LR | 4 | 0.81 (0.64–0.91) | 0.78 (0.66–0.87) | 3 | 0.727 |
| Different features ML vs. LR | 5 | 0.83 (0.74–0.90) | 0.77 (0.71–0.83) | 6 | 0.230 |
| **Short-term mortality (< 1 year)** | | | | | |
| Overall ML vs. LR | 10 | 0.91 (0.84–0.95) | 0.85 (0.78–0.90) | 6 | 0.149 |
| Similar features ML vs. LR | 7 | 0.90 (0.83–0.94) | 0.84 (0.72–0.92) | 6 | 0.303 |
| Different features ML vs. LR | 5 | 0.93 (0.80–0.98) | 0.88 (0.73–0.95) | 5 | 0.491 |
| **MACE (Overall ML vs. LR)** | 4 | 0.85 (0.58–0.96) | 0.75 (0.58–0.86) | 10 | 0.406 |
| **Bleeding (Overall ML vs. LR)** | 3 | 0.81 (0.75–0.86) | 0.77 (0.72–0.81) | 4 | 0.261 |
| **Acute Kidney Injury** | | | | | |
| Overall ML vs. LR | 4 | 0.81 (0.75–0.86) | 0.75 (0.61–0.85) | 6 | 0.373 |
| Similar features ML vs. LR | 4 | 0.81 (0.75–0.86) | 0.75 (0.61–0.85) | 6 | 0.373 |
| Different features ML vs. LR | 3 | 0.78 (0.76–0.80) | 0.73 (0.62–0.82) | 5 | 0.337 |

validation dataset that was supposed to be unseen data. The solution is a nest-cross validation approach in which in the outer loops the model performance is assessed, while in the inner loops (training data) the hyperparameters are optimized [80]. This is a critical issue that was routinely neglected in the evaluated studies.

A second source of data leakage occurs when data pre-processing, such as missing data imputation, feature selection, or data normalization, is performed before the data is split. Conducting these steps prior to partitioning the validation data may lead to data leakage, potentially resulting in overfitting. However, This too is difficult to assess in ML studies as it is rarely discussed in detail.

A recurrent issue identified in the reviewed literature was the inadequacy of the EPV. PROBAST guideline advises a minimum EPV of 10 for traditional modeling approaches; however, this figure may be insufficient for machine learning (ML) techniques [81]. Research indicates that ML models, including RF, SVM, and ANN, might require an EPV that is at least 10 times higher than that of the LR [81]. This was often overlooked in the reviewed articles with numerous studies reporting EPVs below 10, far from the suggested threshold of 200 [77, 81].

The objective of many articles was to compare ML models with LR or traditional risk scores [2, 20, 34, 61, 66]. However, a discrepancy often existed in the number of features used; LR models typically included fewer features compared to the more extensive feature set selected for ML models. This disparity led to overly favorable results for ML models. Additionally, numerous studies benchmarked the performance of ML models against established risk scores such as the GRACE or SYNTAX scores. Given that ML models are developed datasets similar to those they are tested against, it is plausible to anticipate superior performance over traditional risk scores, which are derived from different datasets and incorporate a more limited number of features. To address this issue, we ensured that when possible the ML models were compared using a feature set analogous to that of the LR models. This was done alongside the broader comparison encompassing all models to provide a balanced evaluation. Neither of the analyses showed statistical significance, however, in long-term mortality, the net benefit of ML models with a similar feature set to LR was only 3% in comparison to 6% when comparing ML models with more features than LR.

Another significant issue identified in the reviewed articles was the absence of model calibration reporting. Calibration refers to the alignment of model outcomes with the actual likelihood of an event's occurrence. This is crucial, especially when precise predicted probabilities are needed alongside a binary outcome. Such detailed information can greatly aid in clinical decision-making, as it allows a clinician to understand the likelihood of an event occurring, rather than simply receiving a binary yes or no answer [82].

In our study, outcomes such as MACE and long-term mortality were associated with time-to-event data. This implies that, in addition to determining the probability of an outcome, it is crucial to predict the timing of its occurrence. Despite this importance, our review found that none of the long-term mortality studies and only two MACE studies employed time-to-event models. To remedy this gap, various machine learning and deep learning-based time-to-event models, such as random survival forests and DeepSurv [83] exist.

Our findings are in line with the previous research. A study by Dhiman et al. identified that the methodological conduct of ML studies in oncology was substandard in several domains including, sample size, handling of missing data, model development, and model availability for evaluation [84]. A study by Mortazavi et al. suggested that ML models may only improve performance when

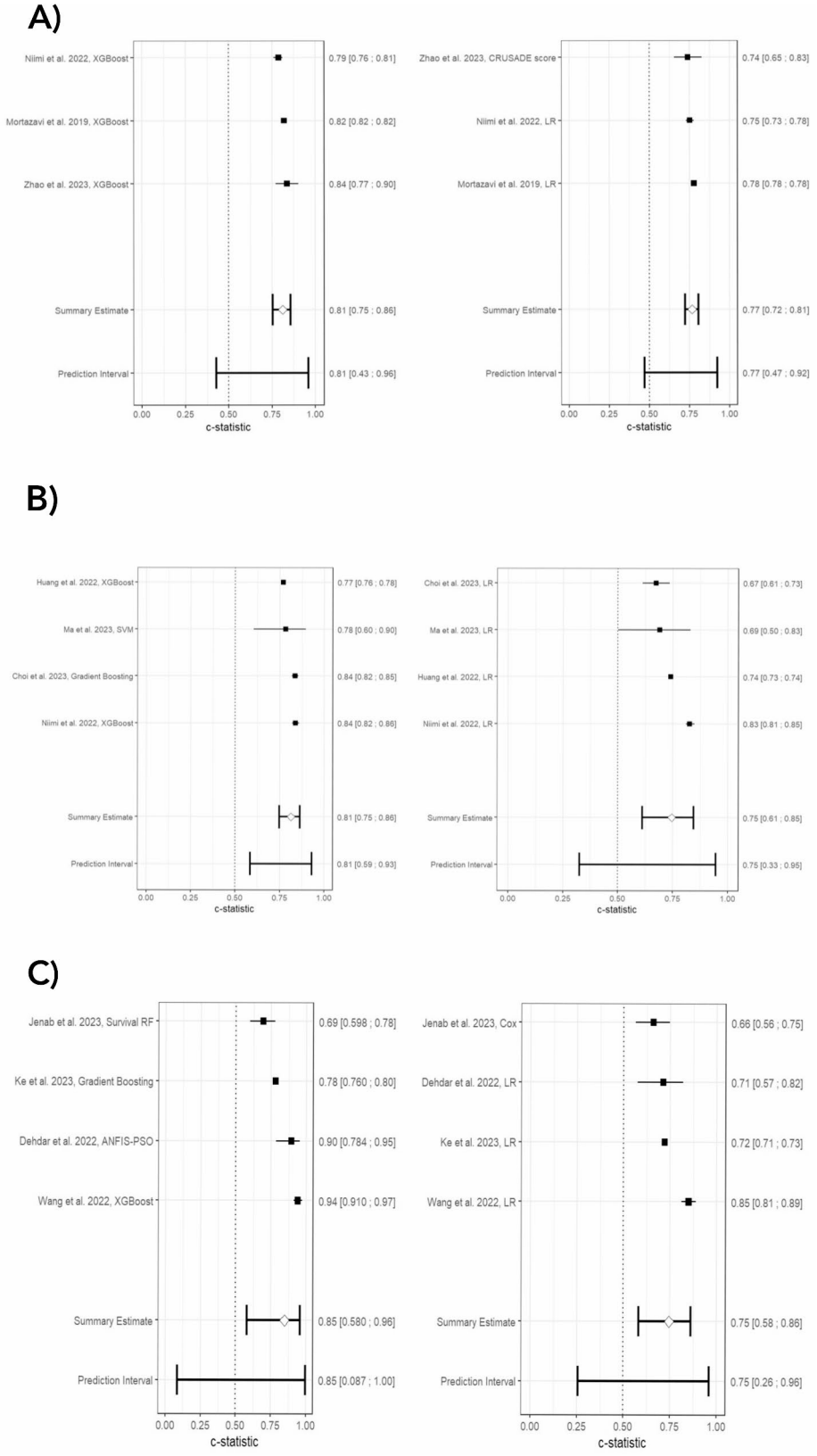


**Fig. 3** (**A**) Pooled c-statistic of the best performing ML (left) vs. LR/ risk scores (right) models for bleeding. (**B**) Pooled c-statistic of the best performing ML (left) vs. LR (right) models for AKI. (**C**) Pooled c-statistic of the best performing ML (left) vs. LR/ Cox proportional hazard (right) models for MACE

trained using appropriate features that do not reduce the information [57]. The validation study by Shi et al. demonstrated that the ML-based PRAISE score overestimated the risk of 1-year mortality, Bleeding, and recurrent acute myocardial infarction. Furthermore, the AUC for GRACE 2.0 score was 0.81 compared to 0.75 for PRAISE [29].

### Limitations

The current study has some limitations. Different ML studies have various methodologies which may lead to heterogeneity. To overcome this hurdle, we only compared studies that provided both LR and ML models. In addition, when possible we provided secondary analysis of studies that used an analogous feature set for both ML and LR models. Additionally, For some outcomes like bleeding, only a limited number of articles provided the relevant data for meta-analysis which could lead to lower statistical power. Finally, we included only English articles which may introduce some publication bias to the current review. Nevertheless, it was the first study evaluating ML investigations in PCI and we provided a critical review of the articles in addition to statistical analysis.

## Conclusion

No statistical significance was observed between ML and LR models. Methodological assessment of the articles revealed concerns such as small sample size, lack of external validation, possible data leakage, and overfitting. While ML models may perform better with much larger datasets, there was the black-box nature of ML models may make the LR models more useful for clinical adaption for now.

We recommend that future studies ensure clearer reporting of methodologies, adhere to PROBAST and CHARMS guidelines, employ nested cross-validation, achieve high values, utilize appropriate methods for handling missing data (such as multiple imputation), and incorporate external validation cohorts. These steps will enable a more robust and reliable comparison between ML and LR models.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12872-025-04746-0.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

**Author contributions**
S.N and A.H wrote the manuscript. S.N and A.V helped in screening. A.H and A.M contributed in data collection. A.S and F.M conceptualized the article.

**Funding**
None.

**Data availability**
The data of the paper are presented in the main text and the supplementary files.

## Declarations

**Conflict of interest**
None.

**Ethics approval statement**
Not applicable.

**Patient consent statement**
Not applicable.

**Permission to reproduce material from other sources**
Not applicable.

**Registration**
PROSPERO (CRD42023494659).

**Consent for publication**
Not applicable.

**Author details**
[1]Tehran Heart Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran
[2]Diabetes Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran
[3]School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[4]Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran
[5]Cardiac Primary Prevention Research Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran

Received: 4 March 2025 / Accepted: 8 April 2025
Published online: 23 April 2025

**References**

1. Serruys PW, Rutherford JD. The birth, and evolution, of percutaneous coronary interventions: A conversation with Patrick Serruys, MD, phd. Circulation. 2016;134(2):97–100.
2. Niimi N, Shiraishi Y, Sawano M, et al. Machine learning models for prediction of adverse events after percutaneous coronary intervention. Sci Rep. 2022;12(1):6262.
3. Rao SV, McCoy LA, Spertus JA, et al. An updated bleeding model to predict the risk of post-procedure bleeding among patients undergoing percutaneous coronary intervention: a report using an expanded bleeding definition from the National cardiovascular data registry CathPCI registry. JACC Cardiovasc Interv. 2013;6(9):897–904.
4. Peterson ED, Dai D, DeLong ER, et al. Contemporary mortality risk prediction for percutaneous coronary intervention: results from 588,398 procedures in the National cardiovascular data registry. J Am Coll Cardiol. 2010;55(18):1923–32.
5. Mehran R, Aymong ED, Nikolsky E, et al. A simple risk score for prediction of contrast-induced nephropathy after percutaneous coronary intervention: development and initial validation. J Am Coll Cardiol. 2004;44(7):1393–9.
6. Nikolsky E, Mehran R, Lasic Z, et al. Low hematocrit predicts contrast-induced nephropathy after percutaneous coronary interventions. Kidney Int. 2005;67(2):706–13.

7. Hannan EL, Farrell LS, Walford G, et al. The new York state risk score for predicting in-hospital/30-day mortality following percutaneous coronary intervention. JACC Cardiovasc Interv. 2013;6(6):614–22.
8. Vroegindewey MM, Schuurman AS, Oemrawsingh RM, et al. SYNTAX score II predicts long-term mortality in patients with one- or two-vessel disease. PLoS ONE. 2018;13(7):e0200076.
9. Ellis RJ, Sander RM, Limon A. Twelve key challenges in medical machine learning and solutions. Intelligence-Based Med. 2022;6:100068.
10. Kulkarni H, Amin AP. Artificial intelligence in percutaneous coronary intervention: improved risk prediction of PCI-related complications using an artificial neural network. BMJ Innovations. 2021;7(3):564.
11. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC State-of-the-Art review. J Am Coll Cardiol. 2021;77(3):300–13.
12. Singh K, Beam AL, Nallamothu BK. Machine learning in clinical journals: moving from inscrutable to informative. Circ Cardiovasc Qual Outcomes. 2020;13(10):e007491.
13. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
14. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
15. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
16. Fernandez-Felix BM, López-Alcalde J, Roqué M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. BMC Med Res Methodol. 2023;23(1):44.
17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
18. MedCalc Software Ltd Comparison of AUC of independent ROC curves. April 23. 2024; 22.023:https://www.medcalc.org/calc/comparison_of_independentROCtest.php
19. R Core Team R: A Language and Environment for Statistical Computing. 2022; https://www.R-project.org/
20. Ninomiya K, Kageyama S, Shiomi H, et al. Can machine learning aid the selection of percutaneous vs surgical revascularization?? J Am Coll Cardiol. 2023;82(22):2113–24.
21. Bai Z, Lu J, Li T, et al. Clinical Feature-Based machine learning model for 1-Year mortality risk prediction of ST-Segment elevation myocardial infarction in patients with hyperuricemia: A retrospective study. Comput Math Methods Med. 2021;2021:7252280.
22. Liu N, Liu Q. A comparative study of the prognostic value of independent SYNTAX score, GRACE score and a new scoring system combining SYNTAX and GRACE for patients with ACS undergoing PCI. J Arrhythmia. 2019;35:601.
23. Huang YC, Chen KY, Li SJ, Liu CK, Lin YC, Chen M. Implementing an ensemble learning model with feature selection to predict mortality among patients who underwent Three-Vessel percutaneous coronary intervention. Appl SCIENCES-BASEL 2022;12(16).
24. Feng XX, Zhang C, Huang X, et al. Machine learning improves mortality prediction in three-vessel disease. ATHEROSCLEROSIS. 2023;367:1–7.
25. Liu SY, Yang SW, Xing AL, et al. Machine learning-based long-term outcome prediction in patients undergoing percutaneous coronary intervention. Cardiovasc DIAGNOSIS THERAPY. 2021;11(3):736–43.
26. D'Ascenzo F, De Filippo O, Gallone G, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. Lancet. 2021;397(10270):199–207.
27. Călburean PA, Grebenişan P, Nistor IA, et al. Prediction of 3-year all-cause and cardiovascular cause mortality in a prospective percutaneous coronary intervention registry: machine learning model outperforms conventional clinical risk scores. Atherosclerosis. 2022;350:33–40.
28. Burrello J, Gallone G, Burrello A et al. Prediction of All-Cause mortality following percutaneous coronary intervention in bifurcation lesions using machine learning algorithms. J Pers Med 2022;12(6).
29. Shi B, Wang HY, Liu J, et al. Prognostic value of Machine-Learning-Based PRAISE score for ischemic and bleeding events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. J Am Heart Assoc. 2023;12(7):e025812.
30. Song L, Li Y, Nie S et al. Using machine learning to predict adverse events in acute coronary syndrome: A retrospective study. Clin Cardiol. 2023.
31. Li YH, Lee IT, Chen YW, Lin YK, Liu YH, Lai FP. Using text content from coronary catheterization reports to predict 5-Year mortality among patients undergoing coronary angiography: A deep learning approach. Front Cardiovasc Med. 2022;9:800864.
32. Agasthi P, Ashraf H, Chao CJ et al. Machine learning helps predict short and Intermediate-term risk for all cause mortality in patients undergoing percutaneous coronary intervention. Circulation 2020;142(SUPPL 3).
33. Iftikhar A, Bond RR, Rjoob K et al. Machine Learning to Predict 30 Days and 1-Year Mortality in STEMI and Turndown Patients. Paper presented at: Computing in Cardiology2020.
34. Ngew KY, Tay HZ, Yusof AKM. Development and validation of a predictive models for predicting the cardiac events within one year for patients underwent percutaneous coronary intervention procedure at IJN. BMC Cardiovasc Disord. 2023;23(1):545.
35. Tourassi GD, Xenopoulos NP. An artificial neural network to predict mortality in patients who undergo percutaneous coronary interventions. Paper presented at: IEEE International Conference on Neural Networks - Conference Proceedings1997.
36. Freeman RV, Eagle KA, Bates ER, et al. Comparison of artificial neural networks with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty. Am Heart J. 2000;140(3):511–20.
37. Tindale A, Cretu I, Meng H, Panoulas V. Complete revascularization is associated with higher mortality in patients with ST-elevation myocardial infarction, multi-vessel disease and shock defined by hyperlactataemia: results from the Harefield shock registry incorporating explainable machine learning. Eur Heart J Acute Cardiovasc Care. 2023;12(9):615–23.
38. Sladojevic M, Tadic S, Pavlovic K, et al. Data mining approach for in-hospital mortality prediction of patients presented with STEMI after primary PCI. Eur J Prev Cardiol. 2014;21(1):S125.
39. Limprasert S, Phu-Ang A. Data modeling using vital sign dynamics for In-hospital mortality classification in patients with acute coronary syndrome. Healthc Inf Res. 2023;29(2):120–31.
40. Al'Aref SJ, Singh G, van Rosendael AR, et al. Determinants of In-Hospital mortality after percutaneous coronary intervention: A machine learning approach. J Am Heart Assoc. 2019;8(5):e011160.
41. Resnic FS, Ohno-Machado L, Selwyn A, Simon DI, Popma JJ. Simplified risk score models accurately predict the risk of major in-hospital complications following percutaneous coronary intervention. Am J Cardiol. 2001;88(1):5–9.
42. Resnic FS, Popma JJ, Ohno-Machado L. Development and evaluation of models to predict death and myocardial infarction following coronary angioplasty and stenting. *Proc AMIA Symp*. 2000:690–693.
43. Hsieh MH, Lin SY, Lin CL, et al. A fitting machine learning prediction model for short-term mortality following percutaneous catheterization intervention: a nationwide population-based study. Ann Transl Med. 2019;7(23):732.
44. Sladojevic M, Pavlovic K, Velicki L, et al. In-hospital mortality prediction for STEMI patients submitted to primary PCI. Eur Heart J. 2013;34:85.
45. Sladojevic M, Sladojevic S, Dejanovic J, et al. Predictive model for in-hospital outcome in ST elevated myocardial infarction complicated with acute heart failure. Eur J Heart Fail. 2016;18:131–2.
46. Sladojevic M, Sladojevic S, Tadic S, Stefanovic M. In-hospital outcome predictions for acute coronary sindrome patients after coronry angioplasty by mining echocardiography parameters data. J Hypertens. 2017;35:e118.
47. Senecal C, Zack CJ, Kinar Y et al. Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention. JACC Cardiovasc Interv 2016;134.
48. Chen P, Wang B, Zhao L, et al. Machine learning for predicting intrahospital mortality in ST-elevation myocardial infarction patients with type 2 diabetes mellitus. BMC Cardiovasc Disord. 2023;23(1):585.
49. Bansal A, Kapadia S, Ellis S, MACHINE LEARNING TECHNIQUES TO PREDICT IN-HOSPITAL CARDIOVASCULAR OUTCOMES IN ELDERLY PATIENTS PRESENTING WITH ACUTE MYOCARDIAL INFARCTION, et al. J Am Coll Cardiol. 2020;75(11):3603.
50. Deng L, Zhao X, Su X, Zhou M, Huang D, Zeng X. Machine learning to predict no reflow and in-hospital mortality in patients with ST-segment elevation myocardial infarction that underwent primary percutaneous coronary intervention. BMC Med Inf Decis Mak. 2022;22(1):109.
51. Hamilton D, Seth M, Sukul D, Gurm HS. A novel tool for highly reliable and accurate prediction of multiple complications in patients undergoing percutaneous coronary intervention. *Circulation*. 2021;144(SUPPL 1).
52. Hamilton DE, Albright J, Seth M, Sukul D, Gurm HS. Merging Machine Learning and Patient Preference: A Contemporary, Comprehensive,

Patient-Centered Tool for Risk Prediction Prior to Percutaneous Coronary Intervention. *CIRCULATION*. 2022;146.
53. Lee G, Gurm HS, Syed Z. Predicting complications of percutaneous coronary intervention using a novel support vector method. J Am Med Inf Assoc. 2013;20(4):778–86.
54. Niedziela JT, Cieśla D, Wojakowski W, et al. Is neural network better than logistic regression in death prediction in patients after ST-segment elevation myocardial infarction? Kardiol Pol. 2021;79(12):1353–61.
55. Bhattacharya A, Sadasivuni S, Chao CJ et al. Multi-modal fusion model for predicting adverse cardiovascular outcome post percutaneous coronary intervention. Physiol Meas 2022;43(12).
56. Hurley NC, Desai N, Dhruva SS et al. A dynamic model to estimate evolving risk of major bleeding after percutaneous coronary intervention. In:2021.
57. Mortazavi BJ, Bucholz EM, Desai NR, et al. Comparison of machine learning methods with National cardiovascular data registry models for prediction of risk of bleeding after percutaneous coronary intervention. JAMA Netw Open. 2019;2(7):e196835.
58. Rayfield C, Agasthi P, Mookadam F, et al. Machine learning on high-dimensional data to predict bleeding post percutaneous coronary intervention. J Invasive Cardiol. 2020;32(5):E122–9.
59. Zhao XY, Wang JM, Yang JA et al. Machine learning for prediction of bleeding in acute myocardial infarction patients after percutaneous coronary intervention. THERAPEUTIC Adv CHRONIC DISEASE 2023;14.
60. Chen P, Liu Y, Chen S, Xian Y, Chen J, Tan N. A novel tool for pre-procedural risk stratification for contrast-induced nephropathy and associations between hydration volume and clinical outcomes following coronary angiography at different risk levels. J Am Coll Cardiol 2018;71(11).
61. Choi H, Choi B, Han S et al. Applicable machine learning model for predicting Contrast-induced nephropathy based on Pre-catheterization variables. Intern Med 2023.
62. Gurm HS, Seth M, Kooiman J, Share D. A novel tool for reliable and accurate prediction of renal complications in patients undergoing percutaneous coronary intervention. J Am Coll Cardiol. 2013;61(22):2242–8.
63. Huang C, Li SX, Mahajan S, et al. Development and validation of a model for predicting the risk of acute kidney injury associated with contrast volume levels during percutaneous coronary intervention. JAMA Netw Open. 2019;2(11):e1916021.
64. Huang C, Murugiah K, Annapureddy A et al. Automating risk assessment of acute kidney injury for patients undergoing percutaneous coronary intervention. *Circulation*. 2021;144(SUPPL 1).
65. Huang C, Murugiah K, Li X et al. Effect of removing race correction factor in glomerular filtration rate Estimation on predicting acute kidney injury after percutaneous coronary intervention. In:2022.
66. Huang C, Murugiah K, Mahajan S, et al. Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: A retrospective cohort study. PLoS Med. 2018;15(11):e1002703.
67. Kuno T, Mikami T, Sahashi Y, et al. Machine learning prediction model of acute kidney injury after percutaneous coronary intervention. Sci Rep. 2022;12(1):749.
68. Ma X, Mo C, Li Y, Chen X, Gui C. Prediction of the development of contrast–induced nephropathy following percutaneous coronary artery intervention by machine learning. Acta Cardiol. 2023;78(8):912–21.
69. Tsutsui R, Johnston J, Felix C, et al. A supervised machine learning approach for predicting acute kidney injury following percutaneous coronary intervention. J Am Coll Cardiol. 2019;74(13):B604–604.
70. Yuan N, Ebinger J. A new multivariate model for safe contrast limits to prevent contrast induced nephropathy after percutaneous coronary intervention. Circulation 2019;140.
71. Dehdar Karsidani S, Farhadian M, Mahjub H, Mozayanimonfared A. Intelligent prediction of major adverse cardiovascular events (MACCE) following percutaneous coronary intervention using ANFIS-PSO model. BMC Cardiovasc Disord 2022;22(1).
72. Farhadian M, Dehdar Karsidani S, Mozayanimonfared A, Mahjub H. Risk factors associated with major adverse cardiac and cerebrovascular events following percutaneous coronary intervention: a 10-year follow-up comparing random survival forest and Cox proportional-hazards model. BMC Cardiovasc Disord. 2021;21(1):38.
73. Jenab Y, Hedayat B, Karimi A, Taaghi S, Ghorashi SM, Ekhtiari H. Effects of opium use on one-year major adverse cardiovascular events (MACE) in the patients with ST-segment elevation MI undergoing primary PCI: a propensity score matched - machine learning based study. BMC Complement Med Ther. 2023;23(1):16.
74. Ke B, Gong R, Shen A, et al. Risk stratification algorithm for clinical outcomes in anemic patients undergoing percutaneous coronary intervention. Ann Med. 2023;55(2):2249200.
75. Park JS. Machine learning for risk prediction of future clinical events in patients with acute coronary syndrome undergone percutaneous coronary intervention. Circulation 2019;140.
76. Wang XB, Cui NH, Liu X. A novel 6-metabolite signature for prediction of clinical outcomes in type 2 diabetic patients undergoing percutaneous coronary intervention. Cardiovasc Diabetol. 2022;21(1):126.
77. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1–33.
78. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49–58.
79. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
80. Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11:2079–107.
81. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014;14:137.
82. Malleson N. Calibration of simulation models. In: Bruinsma G, Weisburd D, editors. Encyclopedia of criminology and criminal justice. New York, NY: Springer New York; 2014. pp. 243–52.
83. Wang Z, Lee JW, Chakraborty T et al. Survival modeling using deep learning, machine learning and statistical methods: A comparative analysis for predicting mortality after hospital admission. arXiv preprint arXiv:240306999. 2024.
84. Dhiman P, Ma J, Andaur Navarro CL, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol. 2022;22(1):101.

## Publisher's note