AsiaIntervention 2024;10:219-232

Accuracy of machine learning in predicting outcomes post-percutaneous coronary intervention: a systematic review

Caitlin Fern Wee[1], MBBS; Claire Jing-Wen Tan[1], MBBS; Chun En Yau[1], MBBS; Yao Hao Teo[1], MBBS; Rachel Go[1], MBBS; Yao Neng Teo[1], MBBS; Benjamin Kye Jyn Tan[1], MBBS; Nicholas L. Syn[1], MBBS; Hui-Wen Sim[2], MBBS, MRCP, MMed; Jason Z. Chen[2], MBBS, MMed, MRCP; Raymond C.C. Wong[1,2], MBBS, MRCP; James W. Yip[1,2], MBBS, MRCP; Huay-Cheem Tan[1,2], MBBS, MMed; Tiong-Cheng Yeo[1,2], MBBS, MRCP; Ping Chai[1,2], MBBS, MMed; Tony Y.W. Li[2], MBBS; Wesley L. Yeung[2], MBBS; Andie H. Djohan[2], MBBS, MRCP, MMed; Ching-Hui Sia[1,2]*, MBBS, MMed, MRCP

1. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 2. Department of Cardiology, National University Heart Centre Singapore, Singapore

This paper also includes supplementary data published online at: https://www.asiaintervention.org/doi/10.4244/AIJ-D-23-00023

KEYWORDS

- coronary artery disease
- prior PCI
- risk stratification

Abstract

Background: Recent studies have shown potential in introducing machine learning (ML) algorithms to predict outcomes post-percutaneous coronary intervention (PCI).

Aims: We aimed to critically appraise current ML models' effectiveness as clinical tools to predict outcomes post-PCI.

Methods: Searches of four databases were conducted for articles published from the database inception date to 29 May 2021. Studies using ML to predict outcomes post-PCI were included. For individual post-PCI outcomes, measures of diagnostic accuracy were extracted. An adapted checklist comprising existing frameworks for new risk markers, diagnostic accuracy, prognostic tools and ML was used to critically appraise the included studies along the stages of the translational pathway: development, validation, and impact. Quality of training data and methods of dealing with missing data were evaluated.

Results: Twelve cohorts from 11 studies were included with a total of 4,943,425 patients. ML models performed with high diagnostic accuracy. However, there are concerns over the development of the ML models. Methods of dealing with missing data were problematic. Four studies did not discuss how missing data were handled. One study removed patients if any of the predictor variable data points were missing. Moreover, at the validation stage, only three studies externally validated the models presented. There could be concerns over the applicability of these models. None of the studies discussed the cost-effectiveness of implementing the models.

Conclusions: ML models show promise as a useful clinical adjunct to traditional risk stratification scores in predicting outcomes post-PCI. However, significant challenges need to be addressed before ML can be integrated into clinical practice.

**Corresponding author: Department of Cardiology, National University Heart Centre Singapore, 1E Kent Ridge Road, NUHS Tower Block Level 9, Singapore, 119228. E-mail: ching_hui_sia@nuhs.edu.sg*

SUBMITTED ON 14/04/2023 - REVISION RECEIVED ON 1st 24/04/2024 / 2nd 13/06/2024 - ACCEPTED ON 14/06/2024

DOI: 10.4244/AIJ-D-23-00023

Abbreviations

AUC	area under the curve
LASSO	least absolute shrinkage and selection operator
ML	machine learning
NPV	negative predictive value
PCI	percutaneous coronary intervention
PPV	positive predictive value

Introduction

Ischaemic heart disease is the greatest cause of mortality and loss of disability-adjusted life years worldwide, accounting for approximately 7 million deaths and 129 million disability-adjusted life years annually[1]. Percutaneous coronary intervention (PCI) is indicated in patients with acute coronary syndrome and has been shown to improve quality of life in those on the maximal tolerated medical therapy[2]. Such intervention may be associated with complications, such as postprocedural acute kidney injury, bleeding, heart failure and others.

Traditional statistical modelling methods have been adopted to predict outcomes post-PCI, involving preselecting and transforming candidate variables based on prior knowledge, applying hierarchical logistic regression to model relationships between variables and outcomes, and reducing the number of variables to create the final model[3]. However, this approach is limited, as it assumes a linear relationship between the variables and logarithmic odds of outcomes, and is weak to collinearity between the variables[4]. Conversely, machine learning (ML) algorithms are free of these linear assumptions and have the additional benefit of being able to control collinearity by regularisation of hyperparameters[5].

ML is a branch of artificial intelligence which uses large datasets to produce algorithms with minimal human intervention, allowing for automated learning. ML learns from examples in training datasets by optimising algorithms according to a loss function. Different ML models exist, including adaptive boosting, k-nearest neighbours, least absolute shrinkage and selection operator (LASSO), random forest, artificial neural network, and support vector machine, amongst others.

In an age of precision medicine, ML has demonstrated its capabilities in sifting through vast amounts of clinical data and reliably predicting outcomes[6], guiding clinicians in efficiently stratifying patients and making individualised treatment decisions[7]. Several studies have also shown significant potential in introducing ML algorithms to predict post-PCI outcomes[8,9]. Nonetheless, other studies have shown no performance benefit of ML over traditional statistical methods for clinical prediction models[10]. Hence, we conducted a systematic review to evaluate the effectiveness and validity of current ML models as a clinical tool to predict outcomes following PCI.

Methods

This systematic review was registered on PROSPERO (International prospective register of systematic reviews; CRD258014) and was reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines[11]. Searches of four databases (PubMed, Embase, Cochrane, and Scopus) were conducted for articles published from the date of inception up to 29 May 2021. A literature search was performed using terms synonymous with "machine learning", "prediction" and "PCI". The full list of search terms can be found in **Supplementary Table 1**.

Table 1 summarises the population, intervention, comparison, outcomes, and inclusion and exclusion criteria used for study selection. Briefly, we included all cohort studies, case-control studies, and randomised controlled trials using ML to predict outcomes post-PCI. Outcomes post-PCI included those relating to mortality (all-cause mortality and in-hospital mortality), the heart (myocardial infarction, heart failure, cardiovascular death, arrhythmia, emergency coronary artery bypass graft, stent thrombosis, and coronary artery restenosis), haemodynamics (bleeding), the kidneys (acute kidney injury, contrast-induced nephropathy, and dialysis) and others (prolonged length of stay ≥7 days and stroke). The range in timeframes for outcome measurement spanned from 72 hours to 1 year.

Three reviewers independently performed the literature search, title and abstract review, full text sieve and data extraction, and all disagreements were resolved by mutual consensus. Baseline demographic information, comorbidities, follow-up duration, medication information and procedural information were collected.

For individual post-PCI outcomes, the number of patients with confirmed disease (N^D), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the curve (AUC), and accuracy were collected for each ML model, when reported. The checklist developed by Banerjee et al[12] was used in this study to critically appraise the included studies, mainly along the stages of the translational pathway: development, validation and impact. Quality of training data and methods of dealing with missing data were evaluated.

Data related to blinding and withdrawals were extracted to assess the risk of bias. Quality control was performed by two independent reviewers using the Newcastle-Ottawa Scale[13] **(Supplementary Table 2)** and the Prediction Risk of Bias ASsessment Tool (PROBAST)[14] **(Supplementary Table 3)**. The Newcastle-Ottawa Scale for cohort studies considers three different domains: selection, comparability, and outcome. PROBAST considers four different domains: participants, predictors, analysis, and outcomes. Studies are graded as having a low, high, or an unclear risk of bias/concern regarding applicability. The Preferred Reporting Items for Systematic reviews and Meta-Analyses checklist[11] is included in **Supplementary Figure 1**.

We included ML models that predicted in-hospital mortality, myocardial infarction, and bleeding. Diagnostic accuracy data for the included models were extracted. The ML models used comprised adaptive boosting, k-nearest neighbours, LASSO, random forest, artificial neural network, support vector machine, multilayer perceptron neural network, Naïve Bayes, extreme gradient boosting, blended model with gradient descent boosting,

Table 1. Population, intervention, comparison, outcomes and study (PICOS) inclusion criteria and exclusion criteria applied to database search.

PICOS	Inclusion criteria	Exclusion criteria
Population	Patients who have undergone PCI	
Intervention	ML model	
Comparison	Traditional risk stratification tools (i.e., CADILLAC risk score, PAMI risk score, Zwolle risk score, GRACE hospital discharge score, dynamic TIMI risk score, RISK-PCI score, APEX AMI risk score, residual SYNTAX score, DAPT Score, GUSTO score, EPICOR prognostic model, and other scores that may be relevant) and statistical modelling	
Outcome	Bleeding, acute kidney injury, contrast-induced nephropathy, dialysis, heart failure, myocardial infarction, cardiovascular deaths, arrhythmias, emergency CABG, stent thrombosis, coronary artery restenosis, all-cause mortality, in-hospital mortality, prolonged length of stay more than or equal to seven days, and stroke	
Study design	Articles in English	Case reports and series, systematic reviews, narrative reviews, qualitative reviews, letters to the editor, non-human studies, abstract only (conference papers), non-peer-reviewed articles
	Cohort studies, case-control studies, randomised controlled trials	
	Year of publication: date of inception-29 May 2021	
	Databases: PubMed, Embase, Cochrane, Scopus	

APEX AMI: Assessment of Pexelizumab in Acute Myocardial Infarction; CABG: coronary artery bypass graft; CADILLAC: Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications; DAPT: dual antiplatelet therapy; EPICOR: long-tErm follow uP of antithrombotic management patterns In acute CORonary syndrome patients; GRACE: Global Registry of Acute Coronary Events; GUSTO: Global Use of Strategies To Open Occluded Coronary Arteries; ML: machine learning; PAMI: Primary Angioplasty in Myocardial Infarction; PCI: percutaneous coronary intervention; PICOS: population, intervention, comparison, outcome, study; TIMI: Thrombolysis in Myocardial Infarction

boosted classification trees algorithm model, and existing simplified risk score with LASSO regression.

Results

The Preferred Reporting Items for Systematic reviews and Meta-Analyses flowchart is presented in **Figure 1**. A literature search of the four databases (PubMed, Embase, Cochrane, Scopus) retrieved 2,546 results. There were 727 duplicates, which were removed. Title and abstract screening excluded a further 1,635 articles as they either did not use ML to predict outcomes post-PCI, did not mention PCI, or had insufficient statistical reporting of post-PCI outcomes. Full text screening excluded 173 articles. Eleven studies were included for the systematic review.

The 11 studies comprised a combined cohort of 4,943,425 patients[3,9,15-23]. Gao 2020 included 2 separate cohorts, comprising 1 retrospective and 1 prospective cohort[17]. Thus, while the flowchart in **Figure 1** shows 11 included studies, 12 cohorts were analysed in total. Across the studies, the reported post-PCI outcomes included in-hospital mortality, myocardial infarction, bleeding, and acute kidney injury. The characteristics of the included studies are shown in **Table 2**. Additional data relating to participant baseline characteristics, including demographics, medications used, and information relating to procedure(s), are presented in **Supplementary Table 4**, **Supplementary Table 5**, and **Supplementary Table 6**, respectively.

The sensitivity, specificity, PPV, NPV, and accuracy for the ML models used to predict in-hospital mortality, myocardial infarction, bleeding, in-hospital mortality and acute kidney injury for each included study are presented in **Table 3**. As seen, the sensitivity, specificity, PPV, NPV and accuracy are consistently high across all models.

Among the 11 studies, different ML models were used, and their methods of derivation varied. Clinical predictors and outcomes for training the ML models utilised in the 11 studies are summarised in **Table 4**. A summary of ML modalities, including the ML model used, software algorithm, training procedure, and optimisation of metrics, is presented in **Table 4**. The quality of training data, including type of study, cohort size, normalisation/standardisation, and validation, is presented in **Table 4** and **Supplementary Table 7**. **Table 5** summarises the studies included for each post-PCI outcome. In all, four studies investigated bleeding outcomes, three studies investigated acute kidney injury outcomes, five studies investigated in-hospital mortality and one study investigated myocardial infarction **(Table 3, Table 5)**. Two studies used artificial neural networks, two used support vector machines, two used random forest algorithms, three used logistic regression models, one used a blended model with gradient descent boosting, two used LASSO techniques, two used adaptive boosting, two used extreme gradient boosting, one used a boosted classification tree algorithm (AI-BR) model, and one used a k-nearest neighbour algorithm. There were concerns about the development of the models. Of the 11 included studies, 10 were studies conducted using data from a single country (seven in the USA, two in China, one in Japan); only one study was a multinational study. The methods of dealing with missing data

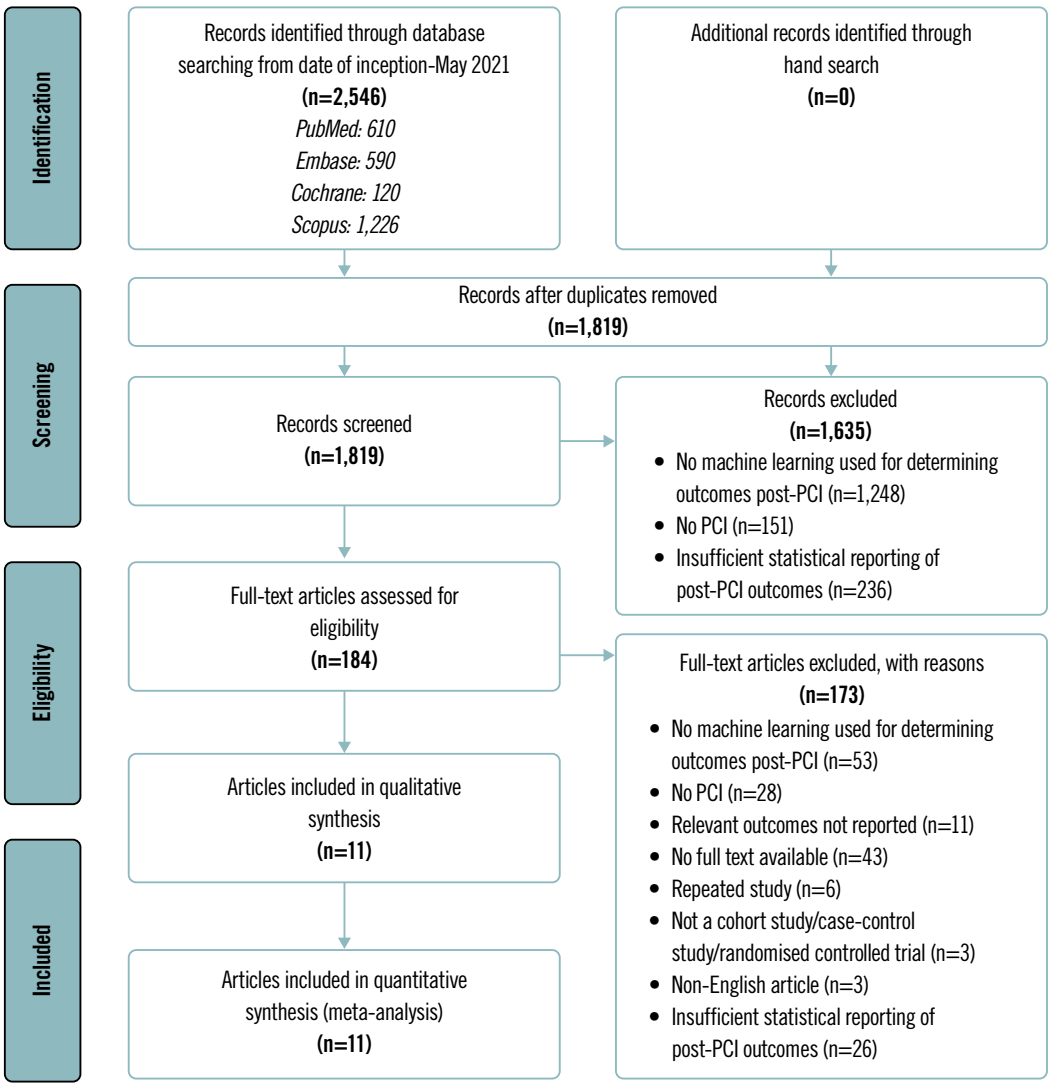

Figure 1. *Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram of study selection. PCI: percutaneous coronary intervention*

were another issue that surfaced. The most common way of dealing with missing data was imputation. However, four studies did not discuss how missing data were handled. One study removed patients if any of the predictor variable data points were missing. In the validation stage, most studies utilised internal validation methods, with four studies using holdout analysis by splitting the dataset into training and test sets, and five studies using N-fold cross-validation. Only three studies externally validated the models presented. There could be concerns over the applicability of the models. While most of the studies presented evidence that the model can be used and interpreted in the clinical context, none of the studies discussed the cost-effectiveness of implementing the model.

Discussion

In this systematic review, we demonstrated that ML models may be useful as an adjunct to existing traditional risk stratification scores in predicting outcomes post-PCI, with moderate to high NPV and AUC.

Traditional risk stratification scores used to predict outcomes post-PCI include the Primary Angioplasty in Myocardial Infarction risk score[24], the RISK-PCI score[25], and the New Mayo Clinic Risk Score[26]. However, such scores are limited by their primary reliance on linear models and diminished ability to explore higher order interactions[27], as they are built on parametric and semiparametric regression scoring systems. Traditional statistical modelling, which is also used to predict outcomes post-PCI, assumes a linear relationship between the variables and logarithmic odds of outcomes[4]. These limitations render traditional risk stratification scores and statistical modelling effective at making predictions at a population level, but less effective at accurately predicting an individual's risk[28].

Compared to the ML models[21], the AUCs for bleeding using traditional scores, such as the Primary Angioplasty in Myocardial Infarction risk score, Thrombolysis in Myocardial Infarction (TIMI) risk score, Global Registry of Acute Coronary Events risk score, and Controlled Abciximab and Device Investigation

Table 2. Characteristics of included studies.

Study name	Study type	Country	Data source	Dates	Inclusion	Exclusion	Sample size	Machine learning model
Al'Aref 2019[19]	Cohort – retrospective	USA	New York PCIRS	1 January 2004 to 31 December 2012	All patients who underwent PCI in the state of New York from 1 January 2004 until 31 December 2012 as documented in the PCIRS database, comprising all elective and emergent cases covering the spectrum of coronary artery disease presentations	Nil	479,804	Adaptive boosting, random forest, XGBoost
D'Ascenzo 2021[16]	Cohort – retrospective	BleeMACS: North and South America, Europe, and Asia; RENAMI: Spain, Italy, Switzerland, Greece, Serbia, United Kingdom	BleeMACS and RENAMI registries	BleeMACS: 1 January 2003 to 31 December 2014 RENAMI: 1 January 2012 to 31 December 2016	BleeMACS: consecutive patients discharged with a diagnosis of ACS undergoing PCI at 1-year follow-up (except death) RENAMI: patients with ACS who underwent PCI and were discharged with DAPT with acetylsalicylic acid plus prasugrel 10 mg once daily or acetylsalicylic acid plus ticagrelor 90 mg twice daily between January 2012 and January 2016	BleeMACS: patients who died during hospitalisation, patients without coronary artery disease, patients who did not undergo PCI (simple balloon angioplasty, stent implantation and/or thromboaspiration). RENAMI: nil	19,826	Adaptive boosting, k-nearest neighbours, Naïve Bayes, random forest
Gao 2020[17]	Cohort – retrospective (training set) – prospective (validation set)	China	Hebei General Hospital, Baoding First Central Hospital, and Cangzhou Central Hospital	Training set: January 2016 to December 2018 Validation set: July 2018 to December 2018	Patients who met the diagnostic criteria of acute STEMI and underwent primary PCI according to current guidelines between the respective time periods for training and validation sets	NR	1,169 (training set); 316 (validation set)	LASSO
Gurm 2014[21]	Cohort – retrospective	USA	BMC2	July 2009 to December 2012	All consecutive patients who underwent PCI between July 2009 and December 2012	Patients who underwent coronary artery bypass grafting during the same hospitalisation	72,328 (training cohort); 30,966 (validation cohort) (PCI procedures)	Random forest
Huang 2018[3]	Cohort – retrospective	USA	NCDR CathPCI Registry	1 June 2009 to 30 June 2011	Patients who underwent PCI procedures	PCIs that were not the first procedure during a single hospitalisation (n=32,999), procedures with same-day discharge (n=41,570), missing serum creatinine before or after the procedure (n=208,158), procedures on patients already on dialysis at the time of their PCI (n=24,271)	947,091	Logistic regression, XGBoost, LASSO regularisation, LASSO regression
Kulkarni 2021[23]	Cohort – retrospective	USA	NCDR CathPCI Registry for 5 BJC HealthCare hospitals	1 July 2009 to 30 April 2018	Patients undergoing PCI at 5 hospitals in the Barnes-Jewish hospital system	NR	28,005 PCIs on 26,784 patients	ANN MLP model

Table 2. Characteristics of included studies (cont'd).

Study name	Study type	Country	Data source	Dates	Inclusion	Exclusion	Sample size	Machine learning model
Kuno 2021[22]	Cohort – prospective	Japan	JCD-KiCS registry	September 2008 to March 2019	Patients undergoing PCI under JCD-KiCS	Patients undergoing chronic dialysis (n=912), patients with missing data on creatinine (n=3,144), haemoglobin (n=3,617) or baseline information e.g., age, sex (n=2,216)	14,273	MLP neural network, logistic model
Matheny 2007[20]	Cohort – retrospective	USA	BWH	1 January 2002 to 31 December 2005	All cases of percutaneous coronary intervention performed at BWH	NR	7,914 PCIs	Support vector machine
Mortazavi 2019[15]	Cohort – retrospective	USA	NCDR CathPCI Registry data, version 4.4	1 July 2009 to 1 April 2015	Patients undergoing the first PCI procedure within same hospitalisation	Not the index PCI of admission, hospital site missing outcome measures, patients who underwent subsequent coronary artery bypass grafting, patients who died in the hospital the same day as the procedure	3,316,465	Blended model with gradient descent boosting, existing simplified risk score with LASSO regularisation
Rayfield 2020[9]	Cohort – retrospective	USA	Mayo Clinic PCI database across 4 sites (La Crosse, Wisconsin; Mankato, Minnesota; Rochester, Minnesota; and Phoenix, Arizona)	January 2006 to December 2017	Patients who had PCI done between January 2006 and December 2017	If any of the 86 variable data points, including bleeding data, were missing	15,603	AI-BR model
Wang 2020[18]	Cohort – retrospective	China	EHR of inpatients who were admitted to the Department of Cardiology at Sir Run Run Shaw Hospital (Hangzhou, Zhejiang, China)	December 2007 to April 2019	1. Inpatients with single coronary artery stenosis (left main artery, left anterior descending artery, left circumflex artery, or right coronary artery); 2. Inpatients with stent implantation during this in-hospital period; 3. From December 2007 to April 2019	1. Myocardial infarction patients or elevated preprocedural cTnI or CK-MB; 2. PCI for more than one artery; 3. Coronary artery with thrombosis; 4. Transluminal extraction-atherectomy therapy for culprit artery; 5. Severe heart failure (EF <45% or NT-proBNP >2,000); 6. Severe valve disease	10,886	Artificial neural networks, support vector machine

ACS: acute coronary syndrome; AI-BR: boosted classification tree algorithm; ANN: artificial neural network; BJC: Barnes-Jewish Corporation; BleeMACS: Bleeding complications in a Multicenter registry of patients discharged with diagnosis of Acute Coronary Syndrome; BMC2: Blue Cross Blue Shield of Michigan Cardiovascular Consortium 2; BWH: Brigham and Women's Hospital; CK-MB: creatinine kinase myocardial band; cTnI: cardiac troponin I; DAPT: dual antiplatelet therapy; EF: ejection fraction; EHR: electronic health record; JCD-KiCS: Japanese Cardiovascular Database-Keio interhospital Cardiovascular Studies; LASSO: least absolute shrinkage and selection operator; MLP: multilayer perceptron; NCDR: National Cardiovascular Data Registry; NR: not reported; NT-proBNP: N-terminal pro B-type natriuretic peptide; PCI: percutaneous coronary intervention; PCIRS: Percutaneous Coronary Interventions Reporting System; RENAMI: REgistry of New Antiplatelets in patients with Myocardial Infarction; STEMI: ST-segment elevation myocardial infarction; XGBoost: eXtreme Gradient Boosting

Table 3. Sensitivity, specificity, PPV, NPV, and accuracy reported by studies that applied an ML method to predict different clinical outcomes post-percutaneous coronary intervention.

Model	ML model	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
In-hospital mortality (best)							
D'Ascenzo 2021[16]	K-nearest neighbour	0.57 (0.53, 0.61)					
D'Ascenzo 2021[16]	Adaptive boosting		0.91 (0.91, 0.91)	0.21 (0.19, 0.23)	0.98 (0.98, 0.98)	0.89 (0.89, 0.90)	0.82 (0.79, 0.85)
Gao 2020[17] (training set)	LASSO	0.98 (0.93, 0.99)					
Gao 2020[17] (validation set)	LASSO		0.95 (0.92, 0.97)	0.63 (0.47, 0.77)	1.00 (0.98, 1.00)	0.95 (0.92, 0.97)	0.99 (0.98, 1.00)
Al'Aref 2019[19]	Adaptive boosting						0.93 (0.92, 0.93)
Matheny 2007[20]	SVM						0.92 (0.91, 0.92)
Kulkarni 2021[23]	ANN						0.92 (0.90, 0.94)
In-hospital mortality (worst)							
D'Ascenzo 2021[16]	K-nearest neighbour		0.88 (0.87, 0.89)	0.17 (0.16, 0.19)	0.98 (0.98, 0.98)	0.87 (0.86, 0.87)	
D'Ascenzo 2021[16]	Adaptive boosting	0.55 (0.51, 0.59)					0.82 (0.79, 0.85)
Gao 2020[17] (training set)	LASSO		0.92 (0.90, 0.93)	0.51 (0.44, 0.58)	1.00 (0.99, 1.00)	0.92 (0.90, 0.94)	0.99 (0.98, 0.99)
Gao 2020[17] (validation set)	LASSO	0.96 (0.80, 0.99)					
Al'Aref 2019[19]	Random forest						0.89 (0.89, 0.90)
Matheny 2007[20]	SVM						0.88 (0.87, 0.88)
Kulkarni 2021[23]	ANN						0.81 (0.76, 0.86)
Myocardial infarction (best)							
D'Ascenzo 2021[16]	Random forest	0.67 (0.63, 0.71)					
D'Ascenzo 2021[16]	Adaptive boosting		0.79 (0.78, 0.80)	0.10 (0.09, 0.11)	0.98 (0.98, 0.98)	0.78 (0.78, 0.79)	
Wang 2020[18]	SVM	0.73 (0.71, 0.75)					
Wang 2020[18]	ANN		0.72 (0.70, 0.74)	0.71 (0.69, 0.73)	0.73 (0.71, 0.75)	0.72 (0.71, 0.73)	
Myocardial infarction (worst)							
D'Ascenzo 2021[16]	Random forest		0.63 (0.62, 0.64)	0.07 (0.06, 0.07)	0.98 (0.98, 0.98)	0.63 (0.62, 0.64)	
D'Ascenzo 2021[16]	Adaptive boosting	0.58 (0.54, 0.62)					
Wang 2020[18]	SVM		0.65 (0.63, 0.67)	0.67 (0.65, 0.69)	0.71 (0.69, 0.73)	0.69 (0.68, 0.70)	
Wang 2020[18]	ANN	0.72 (0.70, 0.74)					
Bleeding (best)							
Mortazavi 2019[15]	Blended model with gradient descent boosting	0.37 (0.37, 0.37)	0.95 (0.95, 0.95)	0.27 (0.26, 0.27)	0.97 (0.97, 0.97)	0.93 (0.93, 0.93)	
Rayfield 2020[9]	Boosted classification tree algorithm	0.77 (0.72, 0.82)	0.81 (0.80, 0.82)	0.07 (0.06, 0.08)	0.99 (0.99, 1.00)	0.81 (0.80, 0.81)	
Gurm 2014[21]	Random forest						0.89 (0.88, 0.90)
Kulkarni 2021[23]	ANN						0.80 (0.86, 0.89)
Bleeding (worst)							
Mortazavi 2019[15]	Existing simplified risk score with LASSO regularisation	0.35 (0.35, 0.35)	0.93 (0.93, 0.93)	0.20 (0.20, 0.20)	0.97 (0.97, 0.97)	0.91 (0.91, 0.91)	
Rayfield 2020[9]	Boosted classification tree algorithm	0.77 (0.72, 0.82)	0.81 (0.80, 0.82)	0.07 (0.06, 0.08)	0.99 (0.99, 1.00)	0.81 (0.80, 0.81)	
Gurm 2014[21]	Random forest						0.88 (0.87, 0.89)
Kulkarni 2021[23]	ANN						0.73 (0.71, 0.76)
Acute kidney injury (best)							
Huang 2018[3]	XGBoost						0.76 (0.76, 0.76)
Kulkarni 2021[23]	ANN						0.82 (0.81, 0.83)
Kuno 2021[22]	Logistic regression						0.83 (0.81, 0.84)
Acute kidney injury (worst)							
Huang 2018[3]	Logistic regression						0.71 (0.71, 0.71)
Kulkarni 2021[23]	ANN						0.63 (0.59, 0.66)
Kuno 2021[22]	Logistic regression						0.81 (0.80, 0.83)

Values in parentheses are 95% confidence intervals. ANN: artificial neural network; AUC: area under the curve; LASSO: least absolute shrinkage and selection operator; ML: machine learning; NPV: negative predictive value; PPV: positive predictive value; SVM: support vector machine; XGBoost: eXtreme Gradient Boosting

to Lower Late Angioplasty Complications risk score (AUC=0.60, 0.62, 0.58, and 0.79, respectively)[29], demonstrated lower values. This suggests a better performance of ML models, compared to traditional predictive models, in prognosticating patients for bleeding risk post-PCI. Compared to that of the best ML models[22,23], the AUC for predicting acute kidney injury using the Primary Angioplasty in Myocardial Infarction risk score (AUC=0.71)[29] demonstrated a lower value, whilst ML models were outperformed by other traditional risk models such as the TIMI risk score, Global Registry of Acute Coronary Events risk score, and Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications risk score (AUC=0.83, 0.78, and 0.98, respectively)[29]. Several studies have also shown traditional statistical methods to have a similar performance to ML in clinical prediction situations[10,30]. Hence, traditional risk stratification scores and statistical modelling are still crucial in clinical practice, but ML models, which are free of linear assumptions and have the additional benefit of being able to control collinearity by optimising hyperparameters[5], may be used as an adjunctive tool to augment clinicians' decision-making regarding personalised risk-benefit analysis[31,32] on whether or not a patient should undergo elective PCI.

In contrast to traditional statistical methods, ML models tend to incorporate a diverse range and greater number of clinically relevant key variables in the training process, comprising demographic characteristics, medical history, preprocedural imaging characteristics, and procedural characteristics, as well as postprocedural complications and outcomes **(Supplementary**

Table 4. Systematic review and quality assessment of included studies.

Author	Al'Aref[19]	D'Ascenzo[16]	Gao[17]	Gurm[21]	Huang[3]
Type of study	Cohort – retrospective	Cohort – retrospective	Cohort – retrospective (training set) Cohort – prospective (validation set)	Cohort – retrospective	Cohort – retrospective
Cohort size	479,804	19,826	316	30,985	947,091
Cohort country	USA	15 tertiary hospitals in North and South America, Europe, and Asia+12 European hospitals	China	USA	USA
Development					
Cohort population	PCIRS database	BleeMACS registry (ClinicalTrials.gov: NCT02466854) and the RENAMI registry+RENAMI	Hebei General Hospital, Baoding First Central Hospital, and Cangzhou Central Hospital	BMC2: all non-federal hospitals in the state of Michigan	NCDR CathPCI
Normalisation/ standardisation	Yes – done before use in model training and validation	Not reported	Yes – all data were normalised by transforming the data into new scores (z-score transformation) with a mean of 0 and a standard deviation of 1	Not reported	Yes – may be performed during feature engineering step
Validation	Yes (5-fold cross-validation)	Yes (internal validation, external validation)	Yes (internal validation, external validation)	Yes (independent validation)	Yes (temporal validation performed on a more contemporary cohort of PCI patients from the NCDR CathPCI registry)
Machine learning model	Adaptive boosting, random forest, XGBoost, logistic regression	Adaptive boosting, k-nearest neighbour	LASSO	Random forest	Logistic regression, XGBoost
Software algorithm	Not reported	SPSS Statistics, version 24.0 (IBM)	R software, version 3.3.0 (R Foundation for Statistical Computing) and Glmnet R package was used for the LASSO regression model	R software, version 2.14.1, using freely distributed contributed packages	All analyses were developed in R. LASSO regularisation with logistic regression was performed using the Glmnet R package. XGBoost was performed using the XGBoost R package. Brier score, reliability, and resolution were calculated with the SpecsVerification R package

Table 8). This facilitates the development of a more robust algorithm, guiding the prediction of post-PCI outcomes in clinical practice in a more precise manner.

Moreover, ML models, especially deep learning models, are adept in handling high-dimensional and complex data. This is particularly beneficial in healthcare systems, where a vast amount of data is constantly generated from diverse sources. While traditional methods can capture non-linear relationships, ML models can do so in a more flexible manner and without need for explicit specification of polynomial terms and interaction variables. In addition, techniques like cross-validation and regularisation in ML can facilitate the development of models that generalise better on unseen data, a key consideration in clinical applications.

It is worthwhile to note that Greenhalgh et al previously published a multilevel non-adoption, abandonment, scale-up, spread, and sustainability (NASSS) framework for studying the diffusion of innovations and promoting technology adoption in healthcare systems[33]. This framework takes into account key factors including the condition, technology, value proposition, adopters, organisation, the wider system, and adaptation over time. Application of this framework to ML models in PCI could potentially aid in the translation of algorithmic success to patient benefit.

The high NPVs using the ML models for in-hospital mortality, myocardial infarction, and bleeding, of 100%, 99%, and 98%, respectively, demonstrate that patients who were predicted not to have poor outcomes post-PCI indeed did not suffer from such complications, thus guiding risk-benefit analysis for PCI. Poor

Table 4. Systematic review and quality assessment of included studies (cont'd).

Kulkarni[23]	Kuno[22]	Matheny[20]	Mortazavi[15]	Rayfield[9]	Wang[18]
Cohort – retrospective	Cohort – retrospective	Cohort – retrospective	Cohort – retrospective	Cohort – retrospective	Cohort – retrospective
26,784	14,273	7,914 PCIs	3,316,465	15,604	10,886
USA	Japan	USA	USA	USA	China
Seven hospitals – Alton Memorial Hospital, Alton, IL; Barnes-Jewish Hospital, St. Louis, MO; Barnes-Jewish St. Peters Hospital, St. Peters, MO; Boone Hospital Center, Columbia, MO; Christian Hospital, St Louis, MO; Missouri Baptist Medical Center, St. Louis, MO; and Progress West HealthCare, O'Fallon, MO	JCD-KiCS registry	BWH (Boston, MA) containing all cases (7,914) of PCI performed at the institution from 1 January 2002 to 31 December 2005	NCDR CathPCI	Mayo Clinic CathPCI registry data	Sir Run Run Shaw hospital (Hangzhou, Zhejiang, China)
Yes – normalisation done for continuous variables before use in model training and validation	Not reported	Not reported	Not reported	Not reported	Not reported
Yes (validation with a separate retrospective dataset)	Yes (automatic system validation)	Yes (3-fold cross-validation inner and outer loop method)	Yes (5-fold cross-validation)	Yes (10-fold cross-validation)	Yes (4-fold cross-validation)
ANN MLP model	Logistic model	Support vector machine-P (CEE)-optimised, support vector machine-R (MSE)-optimised	Blended model with gradient descent boosting, existing simplified risk score with LASSO regression	AI-BR model	Artificial neural networks, support vector machine
All analyses were carried out on R statistical software or Stata (StataCorp)	Statistical calculations and analyses performed using SPSS Statistics, version 24, R 3.5.3 and Python 3.7 (Python Software Foundation)	SVM models were developed using GIST (Columbia University, New York, NY, USA) 2.2.1. LR models were developed using SAS, version 9.1 (SAS Institute)	All analyses were conducted in R (version 3.3.2), with Glmnet used for LASSO regularisation, XGBoost for gradient descent boosting and pROC for C statistics; mgcv and sandwich were used for the continuous calibration curves and SpecsVerification was used for the Brier score	R software, version 3.5.1	Python 3.x software+SPSS Statistics for macOS, version 23

Table 4. Systematic review and quality assessment of included studies (cont'd).

Author	Al'Aref[19]	D'Ascenzo[16]	Gao[17]	Gurm[21]	Huang[3]
Development					
Training procedure	5-fold cross-validation on the dataset for each model. Attribute selection was done after fine-tuning of the hyperparameter – defined as the model parameters that are given an arbitrary value before the initiation of the learning process. Attribute selection was performed using the information gain ranking method that aims at ranking features based on high information gain entropy. The attributes with information gain >0 were only used for the ML approach.	The derivation cohort was randomly split into 2 datasets: a training (80%) cohort, which was used to train the 4 ML models and tune their parameters, and an internal validation (20%) cohort, which was used to test the developed models on unseen data and to fine-tune the hyperparameters. To determine the major predictors of each study outcome in our patient population, the importance of each permutation feature was measured from the final model. Permutation feature importance computes the value of each feature included in the model by calculating the increase in the model's prediction error after permuting its values. A feature is considered important if permuting its values decreases the model's discriminative capability, as the model relies heavily on that feature for the prediction.	The LASSO method was used to select the features that were the most significantly associated with the outcome (in-hospital mortality). Then, a regression model was built using the selected variables. The λ value was selected for which the cross-validation error was the smallest. Finally, the model was refitted using all available observations and the selected λ. Thus, most of the coefficients of the covariates were reduced to 0, and the remaining non-zero coefficients were selected by LASSO.	The study cohort was divided randomly into training and validation datasets, with 70% of procedures assigned to training, and the remaining 30% utilised for validation. A random forest regression model was trained for predicting transfusion using 45 baseline clinical variables including preprocedural medications, with missing predictors imputed to be the overall median for continuous values and mode for categorical variables. The transfusion outcome was entered as a continuous variable coded as 1 in patients who were transfused, and 0 for those not meeting the criteria to facilitate regression rather than classification modelling, so that estimated means (leaf node probabilities of transfusion) assigned to a given observation were then aggregated in the ensemble. To facilitate the development of an easy-to-use bedside tool, a reduced model was also trained using only the 14 most important predictors as assessed in the full model by the incremental decrease in node impurity (residual sum of squares) associated with splitting on the predictor averaged over all trees in the ensemble.	9 prediction models were developed, with combinations of the following 3 categories: (1) preprocessing models (strategy A vs strategy B), (2) variable selection (stepwise backward selection with logistic regression vs LASSO regularisation with logistic regression vs permutation-based selection with XGBoost) and (3) relationship modelling: (logistic regression model vs ML method XGBoost). Analytic cohort was randomly split into a training set (70% of the cohort) and a test set (30% of the cohort). The 9 models were built using data from the training set only, and the corresponding selected variables were recorded. Finally, the performance of the models was assessed on the internal test set.
Optimising metrics	AUC	AUC	AUC	AUC	AUC, Brier score, resolution, reliability

AI-BR: boosted classification tree algorithm; AKI: acute kidney injury; ANN: artificial neural network; AUC: area under the curve; BleeMACS: Bleeding complications in a Multicenter registry of patients discharged with diagnosis of Acute Coronary Syndrome; BMC2: Blue Cross Blue Shield of Michigan Cardiovascular Consortium 2; BWH: Brigham and Women's Hospital; CEE: cross-entropy error; HL: Hosmer-Lemeshow; JCD-KiCS: Japanese Cardiovascular Database-Keio interhospital Cardiovascular Studies; LASSO: least absolute shrinkage and selection operator;

outcomes such as in-hospital mortality, myocardial infarction, and bleeding, might diminish the overall utility of PCI. The high discriminatory value serves as a good adjunctive clinical tool to allow clinicians to weigh the risks and benefits of PCI for their patients.

We have also critically appraised the studies along the key elements of the translational pathway. Development is hampered by the population in each cohort. Of the 12 cohorts included, seven cohorts analysed populations in the USA[3,9,15,19-21,23], three

Table 4. Systematic review and quality assessment of included studies (cont'd).

Kulkarni[23]	Kuno[22]	Matheny[20]	Mortazavi[15]	Rayfield[9]	Wang[18]
Randomly shuffled dataset was split into a derivation set (n=21,004) and a validation dataset (n=7,001). All training for ML algorithms used data from the derivation set, while all models were validated on data from the validation set. Data preprocessing was undertaken using variable encoding. The 2 generated datasets were used to develop 2 separate learning models for each outcome – one incorporating baseline and pre-PCI variables, and the other incorporating variables related to the PCI procedure. Predictions from these two models were then finally combined into a single prediction model using logistic regression. For each training epoch, the estimated best fitting model was independently applied to the test set (the encoded dataset obtained from the validation set) to trace the classification accuracy. Model training continued as long as there was improvement in the classification accuracy for both the training and the independently assessed test set. If the model only showed accuracy improvement in the training set but showed a decreased accuracy for the test set, then a potential overfitting was interpreted, and model training was stopped.	Restricted cubic spline with multivariate logistic regression models were used to assess the association between absolute/ relative decrease in haemoglobin and AKI. ML was constructed with a neural network to evaluate the association between periprocedural haemoglobin reduction and AKI and for risk stratification of AKI, by comparing the effect of NCDR variables versus NCDR variables plus haemoglobin absolute change (continuous value) versus NCDR variables plus haemoglobin relative change (continuous value) and with logistic models.	The cases were used to generate 100 random datasets. All cases were used in each set, and 5,540 were allocated for training and 2,374 were allocated for testing. For SVM evaluation, each training set was randomly divided into 3,957 kernel training and 1,583 sigmoid training portions. The parameter of each kernel type (d and w for the polynomial and Gaussian kernels, respectively) and the magnitude of the constant applied to the soft margin were optimised on the kernel training set separately for AUC, HL χ^2, MSE, and CEE indices by a grid search method, using 3-fold cross-validation. The sigmoid training set was used to convert SVM results into probabilities. Using the training set cross-validation results for each of the performance measures, the best set of parameters for the radial and polynomial kernels were used to generate a model on the entire kernel training set, and a sigmoid for discriminant conversion was generated using the sigmoid training set. Each of the models was then evaluated using the respective test dataset. Logistic regression was chosen to provide the benchmark for SVM comparisons, with similar 3-fold cross-validation performed on each training dataset to optimise feature selection threshold for AUC, HL χ^2, MSE, and CEE performance measures.	Derivation and validation cohorts were created using stratified 5-fold cross-validation. Each variable set was divided randomly into 5 equal subsets, preserving the same event rate in each subset, by first randomly dividing bleeding cases and then non-bleeding cases. Each bleeding subset was then paired with 1 non-bleeding subset. The derivation cohort combined 4 (80%) of the subsets; the remaining subset (20%) was reserved as a validation set. This process was repeated 5 times, such that each of the subsets served as the validation set. Two methods were used to train models in the analysis: logistic regression with LASSO regularisation and gradient descent boosting – XGBoost. The final model used 1,000 trees, a learning rate of 0.1, and a maximum depth of each tree of 6, and it was trained with an objective function aimed at minimising errors similar to logistic regression for binary classification (bleed vs non-bleed).	All recorded variables were considered candidate variables. The variables, once scaled, were fed into an AI-BR. This model trained the base estimator on the training set and observed the training data samples that the base estimator misclassified and created a weighted coefficient for these samples. A second base estimator was then trained, applying the above weight coefficient, to samples when calculating the entropy measure of homogeneity. Boosting was performed to create successive base classifiers that were programmed to place greater emphasis on the misclassified samples from the training data. Finally, a probability of class membership was calculated based on the sum of the individual tree results for each patient. If the sum was >50% probability of bleeding, the patient was predicted to have bled.	Feature selection by information gain measured how much information an attribute gave researchers about the outcome to be predicted. Class-balanced oversampling method was another approach to balance the imbalanced dataset. Drop imputation and mean imputation were individually applied in the dataset to build ML models.
AUC	AUC	AUC, mean squared error, mean CEE, HL goodness-of-fit test	AUC	ROC curve	AUC of ROC curve

LR: logistic regression; ML: machine learning; MLP: multilayer perceptron; MSE: mean squared error; NCDR: National Cardiovascular Data Registry; PCI: percutaneous coronary intervention; PCIRS: Percutaneous Coronary Interventions Reporting System; RENAMI: REgistry of New Antiplatelets in patients with Myocardial Infarction; ROC: receiver operating characteristic; SVM: support vector machine; XGBoost: eXtreme Gradient Boosting

cohorts analysed populations in China[17,18], one cohort analysed populations in Japan[22], and one cohort analysed populations across North America, South America, Europe, and Asia[16]. The small number of countries where these ML models have been developed could limit the generalisability of the results to other potentially underinvestigated, underserved populations. The applicability of the results could also be reduced by the lack of external validation. To date, only one study[16] externally validated

Table 5. Summary table of studies included for each outcome.

Outcome	Number of studies	Studies included
Bleeding AUC	2	Gurm 2014[21] Kulkarni 2021[23]
Acute kidney injury AUC	3	Huang 2018[3] Kulkarni 2021[23] Kuno 2021[22]
In-hospital mortality AUC	5	D'Ascenzo 2021[16] Gao 2020[17] Al'Aref 2019[19] Matheny 2007[20] Kulkarni 2021[23]
Bleeding sensitivity, specificity, PPV, NPV, and accuracy	2	Mortazavi 2019[15] Rayfield 2020[9]
Myocardial infarction sensitivity, specificity, PPV, NPV, and accuracy	1	D'Ascenzo 2021[16]
In-hospital mortality sensitivity, specificity, PPV, NPV, and accuracy	2	D'Ascenzo 2021[16] Gao 2020[17]

AUC: area under the curve; NPV: negative predictive value; PPV: positive predictive value

the model in a multinational cohort. More resources should be allocated to validate the model and apply the results in more diverse patient populations. Another issue of missing data surfaced in our analysis. Four studies did not discuss how missing data were handled. One study conducted complete case analysis by removing patients with missing predictor variable data points. Unclear methods of handling missing data, or complete case analysis, may lead to underpowered studies or bias, especially if the data are not missing at random[34].

Limitations

To the best of our knowledge, this is the first review to critically appraise and review the accuracy of ML models used in predicting outcomes post-PCI. Comprehensive data comprising baseline clinical characteristics, training procedures for ML models, quality of training data and ML outcomes were retrieved, analysed, and synthesised from individual studies to evaluate the accuracy of ML models in predicting pertinent post-PCI outcomes.

Nonetheless, this study should be interpreted in the context of known and potential limitations. Firstly, there existed significant heterogeneity among the studies included in this systematic review. For the clinical predictors reported, while the categories of predictors used were largely similar, the individual predictors included in each category differed across the studies. The baseline demographics of study populations also differed, and the duration of follow-up for post-PCI outcomes was not reported in the majority of the included studies. Most studies examined supervised machine learning techniques such as LASSO and random forest models **(Table 4)**. Also, the performance between different models, particularly that of deep learning networks and traditional supervised ML models, was not reported. Further studies should be conducted to explore the different ML models and to determine which ML models have the best predictive performance.

Secondly, while the quality of training data was overall high, the majority of the studies (n=10) were retrospective in nature, which may further introduce bias into the training of ML models. Moreover, software algorithms and training procedures employed for ML models across studies were not standardised. Also, ML models can be very sensitive to the optimisation model chosen[20]. Thus, caution should be exercised before declaring any model to be superior to other risk prediction tools.

Thirdly, the "black box" technology of ML models leads to these models being complex and unpredictable because of a lack of transparency about the underlying decision-making processes. Input data may undergo complex transformations in multiple layers of the algorithm, with the relationship between individual clinical predictors and contribution of each predictor to the outcome unknown to the user[35]. The complex datasets utilised in ML models may also be prone to missing data, unmeasured confounding, and systemic errors, all of which may further compromise the validity of the models' predictions[35]. Also, ML models with low sensitivity may miss patients at risk of adverse outcomes post-PCI. This may impact clinicians' ability to accurately weigh the risks and benefits of elective PCI, affect preprocedural counselling, and may potentially lead to medico-legal issues. To mitigate this issue, the developers of ML algorithms should define the purpose (screening vs diagnosis) of the ML models and choose a binary threshold in the validation set to derive appropriate sensitivities. In the usage of low-sensitivity ML models, outcome predictions made using ML models must ultimately still be interpreted cautiously in appropriate clinical contexts, which should be done by experienced clinicians.

Lastly, while the findings of our research are informative and useful for understanding PCI outcomes, it is important to acknowledge that they may not be universally applicable to all scenarios. This is due to the fact that all of the included studies are single-centre studies, four of them have unclear data handling strategies, and only three externally validated the models presented. This significantly increases the risk of overfitting to training data, limiting the interpretation of good model performance. Thus, it is challenging to comment on the definitive benefit of real-world effectiveness. The majority of the studies also focused on the USA (seven studies), with two studies focused on China, but not other countries, limiting generalisability. In light of the fact that the robustness and generalisability may be overstated, PROBAST was performed. Ultimately, outcome predictions by ML models must still be interpreted judiciously and contextualised to each case.

Conclusions

In this systematic review, we demonstrated that ML models may be a valuable clinical adjunct to existing traditional risk stratification scores in predicting outcomes post-PCI, with moderate to high NPV and AUC. Such a clinical tool may one day guide clinicians in prognostication of complications and the selection of patients with

the most optimal risk-benefit profile to undergo the procedure. The limitations of the findings are difficult to address in the near future, as the data and technological needs to incorporate ML models into daily clinical practice would require some time to develop. Given the heterogeneity and retrospective design of the studies analysed, future prospective studies are required to investigate the accuracy of ML models more consistently. Employment of larger datasets to train ML models, and refinement of existing ML algorithms via improvements in development and validation may also help to improve the sensitivity, specificity, predictive values, and accuracy of ML models to facilitate their meaningful use in clinical practice.

Impact on daily practice

We suggest that machine learning (ML) can be used as an adjunct to help clinicians weigh the risks and benefits of percutaneous coronary intervention (PCI) versus continued medical therapy in elderly patients with multiple comorbidities who are at higher risk of complications. When a patient presents for elective PCI, clinicians can extract demographic data and past medical history from the electronic health records and enter them into the ML algorithm. Following a targeted history, physical examination, and investigations, clinicians can input further relevant data, including preprocedural imaging data, into the ML algorithm, to determine the potential benefit and personalised risk, so that patients can make a better-informed decision. By selecting the most suitable patients with precision medicine, morbidity, mortality, and healthcare burden can be decreased.

Availability of data and materials

Data used for this study can be accessed upon request from the principal investigator (Dr Ching-Hui Sia) at: ching_hui_sia@nuhs.edu.sg

Funding

C.-H. Sia was supported by the National University of Singapore Yong Loo Lin School of Medicine's Junior Academic Fellowship Scheme.

Conflict of interest statement

H.C. Tan is a deputy editor at AsiaIntervention. The other authors have no conflicts of interest to declare.

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, Mabweijano J, MacIntyre MF, Mallinger L, March L, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGrath J, Mensah GA, Merriman TR, Michaud C, Miller M, Miller TR, Mock C, Mocumbi AO, Mokdad AA, Moran A, Mulholland K, Nair MN, Naldi L, Narayan KM, Nasseri K, Norman P, O'Donnell M, Omer SB, Ortblad K, Osborne R, Ozgediz D, Pahari B, Pandian JD, Rivero AP, Padilla RP, Perez-Ruiz F, Perico N, Phillips D, Pierce K, Pope CA 3rd, Porrini E, Pourmalek F, Raju M, Ranganathan D, Rehm JT, Rein DB, Remuzzi G, Rivara FP, Roberts T, De León FR, Rosenfeld LC, Rushton L, Sacco RL, Salomon JA, Sampson U, Sanman E, Schwebel DC, Segui-Gomez M, Shepard DS, Singh D, Singleton J, Sliwa K, Smith E, Steer A, Taylor JA, Thomas B, Tleyjeh IM, Towbin JA, Truelsen T, Undurraga EA, Venketasubramanian N, Vijayakumar L, Vos T, Wagner GR, Wang M, Wang W, Watt K, Weinstock MA, Weintraub R, Wilkinson JD, Woolf AD, Wulf S, Yeh PH, Yip P, Zabetian A, Zheng ZJ, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380:2095-128.

2. Maron DJ, Hochman JS, Reynolds HR, Bangalore S, O'Brien SM, Boden WE, Chaitman BR, Senior R, López-Sendón J, Alexander KP, Lopes RD, Shaw LJ, Berger JS, Newman JD, Sidhu MS, Goodman SG, Ruzyllo W, Gosselin G, Maggioni AP, White HD, Bhargava B, Min JK, Mancini GBJ, Berman DS, Picard MH, Kwong RY, Ali ZA, Mark DB, Spertus JA, Krishnan MN, Elghamaz A, Moorthy N, Hueb WA, Demkow M, Mavromatis K, Bockeria O, Peteiro J, Miller TD, Szwed H, Doerr R, Keltai M, Selvanayagam JB, Steg PG, Held C, Kohsaka S, Mavromichalis S, Kirby R, Jeffries NO, Harrell FE Jr, Rockhold FW, Broderick S, Ferguson TB Jr, Williams DO, Harrington RA, Stone GW, Rosenberg Y; ISCHEMIA Research Group. Initial Invasive or Conservative Strategy for Stable Coronary Disease. *N Engl J Med.* 2020;382:1395-407.

3. Huang C, Murugiah K, Mahajan S, Li SX, Dhruva SS, Haimovich JS, Wang Y, Schulz WL, Testani JM, Wilson FP, Mena CI, Masoudi FA, Rumsfeld JS, Spertus JA, Mortazavi BJ, Krumholz HM. Enhancing the prediction of acute kidney injury risk after percutaneous coronary intervention using machine learning techniques: A retrospective cohort study. *PLoS Med.* 2018;15:e1002703.

4. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. *Perspect Clin Res.* 2017;8:148-51.

5. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, Chihara H, Fukumitsu R, Okawa M, Yamana N, Imamura H, Sadamasa N, Hatano T, Nakahara I, Sakai N, Miyamoto S. Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning. *Stroke.* 2019;50:2379-88.

6. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.* 2016;375:1216-9.

7. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database* (Oxford). 2020 Jan 1;2020:baaa010.

8. Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar-Sinai Y, Widmer RJ, Lennon R, Singh M, Bell MR, Lerman A, Gulati R. Leveraging Machine Learning Techniques to Forecast Patient Prognosis After Percutaneous Coronary Intervention. *JACC Cardiovasc Interv.* 2019;12:1304-11.

9. Rayfield C, Agasthi P, Mookadam F, Yang EH, Venepally NR, Ramakrishna H, Slomka P, Holmes DR Jr, Arsanjani R. Machine Learning on High-Dimensional Data to Predict Bleeding Post Percutaneous Coronary Intervention. *J Invasive Cardiol.* 2020;32:E122-9.

10. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12-22.

11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.

12. Banerjee A, Chen S, Fatemifar G, Zeina M, Lumbers RT, Mielke J, Gill S, Kotecha D, Freitag DF, Denaxas S, Hemingway H. 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;19:85.

13. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol.* 2010;25:603-5.

14. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med.* 2019;170:51-8.

15. Mortazavi BJ, Bucholz EM, Desai NR, Huang C, Curtis JP, Masoudi FA, Shaw RE, Negahban SN, Krumholz HM. 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:e196835.

16. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, Ariza-Solé A, Liebetrau C, Manzano-Fernández S, Quadri G, Kinnaird T, Campo G, Simao Henriques JP, Hughes JM, Dominguez-Rodriguez A, Aldinucci M, Morbiducci U, Patti G, Raposeiras-Roubin S, Abu-Assi E, De Ferrari GM; PRAISE study group. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. *Lancet.* 2021;397:199-207.

17. Gao N, Qi X, Dang Y, Li Y, Wang G, Liu X, Zhu N, Fu J. Establishment and validation of a risk model for prediction of in-hospital mortality in patients with acute ST-elevation myocardial infarction after primary PCI. *BMC Cardiovasc Disord.* 2020;20:513.

18. Wang Y, Zhu K, Li Y, Lv Q, Fu G, Zhang W. A machine learning-based approach for the prediction of periprocedural myocardial infarction by using routine data. *Cardiovasc Diagn Ther.* 2020;10:1313-24.

19. Al'Aref SJ, Singh G, van Rosendael AR, Kolli KK, Ma X, Maliakal G, Pandey M, Lee BC, Wang J, Xu Z, Zhang Y, Min JK, Wong SC, Minutello RM. Determinants of In-Hospital Mortality After Percutaneous Coronary Intervention: A Machine Learning Approach. *J Am Heart Assoc.* 2019;8:e011160.

20. Matheny ME, Resnic FS, Arora N, Ohno-Machado L. Effects of SVM parameter optimization on discrimination and calibration for post-procedural PCI mortality. *J Biomed Inform*. 2007;40:688-97.

21. Gurm HS, Kooiman J, LaLonde T, Grines C, Share D, Seth M. A random forest based risk model for reliable and accurate prediction of receipt of transfusion in patients undergoing percutaneous coronary intervention. *PLoS One.* 2014;9:e96385.

22. Kuno T, Numasawa Y, Mikami T, Niimi N, Sawano M, Kodaira M, Suzuki M, Ueno K, Ueda I, Fukuda K, Kohsaka S. Association of decreasing hemoglobin levels with the incidence of acute kidney injury after percutaneous coronary intervention: a prospective multi-center study. *Heart Vessels.* 2021;36:330-6.

23. 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;7:564-79.

24. Addala S, Grines CL, Dixon SR, Stone GW, Boura JA, Ochoa AB, Pellizzon G, O'Neill WW, Kahn JK. Predicting mortality in patients with ST-elevation myocardial infarction treated with primary percutaneous coronary intervention (PAMI risk score). *Am J Cardiol.* 2004;93:629-32.

25. Savic L, Mrdovic I, Asanin M, Stankovic S, Krljanac G, Lasica R. Using the RISK-PCI Score in the Long-Term Prediction of Major Adverse Cardiovascular Events and Mortality after Primary Percutaneous Coronary Intervention. *J Interv Cardiol.* 2019;2019:2679791.

26. Amin LZ, Amin HZ, Nasution SA, Panggabean M, Shatri H. The New Mayo Clinic Risk Score Characteristics in Acute Coronary Syndrome in Patients Following Percutaneous Coronary Intervention. *J Tehran Heart Cent.* 2017;12:149-54.

27. Gibson WJ, Nafee T, Travis R, Yee M, Kerneis M, Ohman M, Gibson CM. Machine learning versus traditional risk stratification methods in acute coronary syndrome: a pooled randomized clinical trial analysis. *J Thromb Thrombolysis.* 2020;49:1-9.

28. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15:233-4.

29. Kao YT, Hsieh YC, Hsu CY, Huang CY, Hsieh MH, Lin YK, Yeh JS. Comparison of the TIMI, GRACE, PAMI and CADILLAC risk scores for prediction of long-term cardiovascular outcomes in Taiwanese diabetic patients with ST-segment elevation myocardial infarction: From the registry of the Taiwan Society of Cardiology. *PLoS One.* 2020;15:e0229186.

30. 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.

31. Patel JL, Goyal RK. Applications of artificial neural networks in medical science. *Curr Clin Pharmacol*. 2007;2:217-26.

32. Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach. *Artif Intell Med*. 2013;57: 9-19.

33. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, Hinder S, Fahy N, Procter R, Shaw S. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res.* 2017;19:e367.

34. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol.* 2017;17:162.

35. Gavan SP, Thompson AJ, Payne K. The economic case for precision medicine. *Expert Rev Precis Med Drug Dev.* 2018;3:1-9.

Supplementary data

Supplementary Table 1. Search terms.

Supplementary Table 2. Evaluation of risk of bias using the Newcastle-Ottawa Scale (NOS).

Supplementary Table 3. Evaluation of risk of bias using the Prediction Risk of Bias ASsessment Tool (PROBAST).

Supplementary Table 4. Additional data on participant baseline characteristics (demographics).

Supplementary Table 5. Additional data on participant baseline characteristics (medications).

Supplementary Table 6. Additional data on participant baseline characteristics (procedure).

Supplementary Table 7. Quality assessment of included studies.

Supplementary Table 8. Clinical predictors and outcomes involved in the training of different ML models.

Supplementary Figure 1. PRISMA 2020 checklist.

The supplementary data are published online at: https://www.asiaintervention.org/ doi/10.4244/AIJ-D-23-00023

Supplementary data

Supplementary Table 1. Search terms.

Concept	Search Terms
Machine Learning	"machine learning" OR "convolutional network" OR "deep network" OR "neural network" OR "neural networks" OR "bayesian network" OR "classification tree" OR "regression tree" OR "probability tree" OR "multilayer perceptron" OR "artificial intelligence" OR "deep learning" OR "decision trees" OR "random forest" OR "support vector machine" OR "SVM" OR "elastic net" OR "ridge" OR "lasso"
Prediction	"predictive modelling" OR "predictive model" OR "predict" OR "prediction" OR "forecast" OR "learning algorithm" OR "learning algorithms" OR "bayesian logistic regression"
Percutaneous Coronary Intervention	"percutaneous coronary intervention" OR "PCI" OR "drug eluting stent" OR "coronary stent" OR "coronary angioplasty" OR "angioplasty with stent" OR "stent" OR "angina" OR "myocardial infarction" OR "acute coronary syndrome" OR "ACS"

Supplementary Table 2. Evaluation of risk of bias using the Newcastle-Ottawa Scale (NOS).

Study	Representativeness of cohort	Selection of the non-exposed cohort	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts on the basis of the design or analysis controlled for confounders	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow-up of cohorts	Final score	Risk of bias
Al'Aref 2019	*		*	*	*	*			5	Moderate
D'Ascenzo 2021	*		*		*	*	*		5	Moderate
Gao 2020	*		*	*	*	*			5	Moderate
Gurm 2014	*		*	*	**	*			6	Moderate
Kulkarni 2021	*		*	*	*	*			5	Moderate
Kuno 2020	*		*	*	**	*			6	Moderate
Matheny 2007	*		*	*	**	*			6	Moderate
Mortazavi 2019	*		*		**	*	*		6	Moderate
Rayfield 2020	*	*	*	*	**	*			7	Moderate

Supplementary Table 3. Evaluation of risk of bias using the Prediction Risk of Bias ASsessment Tool (PROBAST).

Study	Participants	Predictors	Outcomes	Analysis	Overall
Al'Aref 2019	+	-	+	+	-
D'Ascenzo 2021	+	?	+	+	+
Gao 2020	+	-	+	+	-
Gurm 2014	+	+	+	-	-
Kulkarni 2021	+	-	+	+	+
Kuno 2020	+	-	+	+	-
Matheny 2007	+	+	+	-	+
Mortazavi 2019	+	-	+	?	?
Rayfield 2020	+	-	+	-	-

*+ indicates low ROB/low concern regarding applicability; - indicates high ROB/high concern regarding applicability; and ? indicates unclear ROB/unclear concern regarding applicability

Supplementary Table 4. Additional data on participant baseline characteristics (demographics).

STUDY NAME	AGE (mean years ± SD, (range))	MALES (%)	PRESENT SMOKER (%)	BMI (kg/m2)	DM (%)	HTN (%)	HLD (%)	AF (%)	CAD (%)	Prior MI (%)	Prior CVA/TIA (%)	Prior PCI (%)	Prior CABG (%)	LVEF (%)
Al'Aref 2019	65.2±11.9	31.5	NR	29.4±5.9	33.7	NR	NR	NR	NR	NR	NR	22.1	16.5	50.6±14.5
D'Ascenzo 2021	64 (54-73)	78.0	NR	NR	24.8	55.9	51.0	NR	NR	12.6	5.6	12.7	2.7	55 (39-61)
Gao 2020	25.40±3.45	71.2	40.8	25.40±3.45	20.9	47.2	NR	NR	NR	NR	NR	NR	NR	53.94±7.62
Gurm 2014	64.91±12.08	65.6	29.7	30.51±7.54	37.1	85.2	83.2	NR	NR	35.4	NR	45.3	18.7	52.08±12.67
Huang 2018	64.8±12.2	32.8	NR	30.1±11.8	35.8	81.8	NR	NR	NR	29.8	12.2	39.7	18.6	NR
Kulkarni 2021	65.6	65.3	27.0	NR	40.4	83.8	84.5	NR	NR	34.8	16.7	49.6	22.5	52.1
Kuno 2021	68.4±11.6	79.0	NR	NR	40.4	72.8	63.8	NR	NR	20.9	8.9	31.9	4.3	NR
Matheny 2007	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Mortazavi 2019	65	68.1	NR	29	37.0	82.1	NR	NR	NR	NR	NR	41.2	18.1	NR
Rayfield 2020	67±12.7	70.0	NR	29.75 (26.33-33.85)	25.9	64.2	NR	NR	NR	7.0	NR	30.5	15.6	NR
Wang 2020	64.51±18.3	67.0	39.0	23.44±10.81	27.0	71.0	NR	NR	NR	NR	23.0	25.0	NR	NR
	62.45±21.32	69.0	42.0	23.98±6.11	24.0	54.0	NR	NR	NR	NR	19.0	33.0	NR	NR
	67.85±10.05	69.0	39.0	24.54±9.52	27.0	73.0	NR	NR	NR	NR	23.0	25.0	NR	NR
	67.71±9.88	75.0	42.0	24.7±4.4	24.0	69.0	NR	NR	NR	NR	19.02	33.0	NR	NR

Abbreviations: AF: Atrial fibrillation; BMI: Body mass index; CABG: Coronary artery bypass graft; CAD: Coronary artery disease; CVA: Cerebrovascular accident; DM: Diabetes mellitus; HLD: Hyperlipidaemia; HTN: Hypertension; LVEF: Left ventricular ejection fraction; MI: Myocardial infarction; PCI: Percutaneous coronary intervention; TIA: Transient ischemic attack

Supplementary Table 5. Additional data on participant baseline characteristics (medications).

Study	ANTIPLATELETS (%)	ANTICOAGULANTS (%)	STATINS (%)	ACE INHIBITORS / ARBs (%)	BETA BLOCKERS (%)	CALCIUM CHANNEL BLOCKERS (%)
Al'Aref 2019	NR	NR	NR	NR	NR	NR
D'Ascenzo 2021	Clopidogrel: 68.4; Prasugrel: 11.8; Ticagrelor: 16.9	4.2	80.4	63.5	68.2	NR
Gao 2020	NR	NR	NR	56.8 (Training set); 60.4 (Validation set)	72.6 (Training set); 76.6 (Validation set)	NR
Gurm 2014	NR	NR	NR	NR	NR	NR
Huang 2018	NR	NR	NR	NR	NR	NR
Kulkarni 2021	Aspirin: 99.1; Bivalirudin: 67.4; Clopidogrel: 77.5; Ticlopidine: 0.3; Prasugrel: 10.5; Ticagrelor: 14.8	Fondaparinux: 0.1; Low molecular weight heparin: 8.8; Unfractionated heparin: 50.1	NR	NR	67.2	21.7
Kuno 2021	NR	NR	NR	NR	NR	NR
Matheny 2007	NR	NR	NR	NR	NR	NR
Mortazavi 2019	NR	NR	NR	NR	NR	NR
Rayfield 2020	NR	NR	NR	NR	NR	NR
Wang 2020	83.6	91	81.9	NR	NR	NR

Abbreviations: ACE: Angiotensin-Converting Enzyme; ARB: Angiotensin Receptor Blocker; NR: Not Reported

Supplementary Table 6. Additional data on participant baseline characteristics (procedure).

Study	TIME TO REPERFUSION	CULPRIT LESION / VESSEL RELATED TO INFARCTION e.g. LAD, LCx, RCA	MULTI-VESSEL DISEASE (%)	NUMBER OF DISEASED VESSELS (%)	PRIORITY (ELECTIVE/ URGENT/ EMERGENT/ SALVAGE ETC.) (%)	TIMI FLOW (%)	STENOSIS MORPHOLOGY	TYPE OF STENT e.g. Drug-eluting, Bare metal	NUMBER OF STENTS	LENGTH OF STENTS (mm)	DIAMETER OF STENTS (mm)	ARTERY ACCESS (RADIAL OR FEMORAL) (%)	CALCIFICATION (%)	FRACTIONAL FLOW RESERVE, INTRAVASCULAR ULTRASOUND, OPTICAL COHERENCE TOMOGRAPHY (%)	CHRONIC TOTAL OCCLUSIONS (%)	PCI WITHOUT DILATATION (%)	DURATION OF FOLLOW-UP
Al Aref 2019	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
D'Ascenzo 2021	NR	NR	58.2	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	1 year
Gao 2020	NR	NR	NR	NR	NR	0: 1.3, 1: 1.0, 2: 5.9, 3: 91.6 (Training set); 0: 1.9, 1: 1.3, 2: 5.4, 3: 91.5 (Validation set)	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Gurm 2014	NR	NR	NR	NR	Elective: 41.2, Urgent: 43.0, Emergency: 15.6. Salvage: 0.2 (Training cohort); Elective: 41.6, Urgent: 42.8, Emergency: 15.4, Salvage: 0.2 (Validation cohort)	NR	NR	NR	NR	NR	NR	Femoral: 89.7; Radial: 9.9 (Training cohort); Femoral: 89.7, Radial: 9.9 (Validation cohort)	NR	NR	NR	NR	NR	NR
Huang 2018	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Kulkarni 2021	Door to balloon time: 74.5 min; Symptom action time: 395.6 min	Left main stem, proximal LAD, mid/distal LAD, circumflex artery, ramus, RCA, proximal LAD graft, mid/distal LAD graft, circumflex artery graft, RCA graft, ramus graft	50.5	0: 1.2, 1: 48.2; 2: 30, 3: 20.5	Elective: 38.4, Urgent: 47., Emergency: 13., Salvage: 0.17	NR	NR	Drug-eluting stents, bare-metal stents	Average number of DES (n = 1.45); Average number of BMS (n = 0.18)	31.6	Average minimum stent diameter: 2.87	NR	NR	NR	NR	NR	NR
Kuno 2021	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	Radial: n = 7,092, Femoral: n = 7,181	NR	NR	NR	NR	NR
Matheny 2007	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Mortazavi 2019	NR	NR	NR	NR	Elective: 41.5, Urgent: 39.9, Emergent: 18.3, Salvage: 0.3	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Rayfield 2020	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

Wang 2020	NR	Left coronary artery: 26.0	NR	NR	NR	NR	ACC/AHA TypeB2C: 887	NR	1.71±0.83	49.66±24.92	≥ 2.5mm: 2047	NR	16.0	13.0	11.0	11.0	NR
	NR	Left coronary artery: 30.0	NR	NR	NR	NR	ACC/AHA TypeB2C: 409	NR	1.39±0.67	33.16±20.15	≥ 2.5mm: 2115	NR	7.0	8.0	6.0	13.0	NR
	NR	Left coronary artery:26.0	NR	NR	NR	NR	ACC/AHA TypeB2C: 887	NR	1.71±0.83	49.66±24.92	≥ 2.5mm: 2047	NR	16.0	13.0	11.0	11.0	NR
	NR	Left coronary artery: 30.0	NR	NR	NR	NR	ACC/AHA TypeB2C: 409	NR	1.39±0.67	33.16±20.15	≥ 2.5mm: 2115	NR	7.0	8.0	6.0	13.0	NR

Abbreviations: LAD: Left Anterior Descending; LCx: Left Circumflex; NR: Not Reported; PCI: Percutaneous Coronary Intervention; RCA: Right Coronary Artery; TIMI: Thrombolysis in Myocardial Infarction

Supplementary Table 7. Quality assessment of included studies.

	Author	Al'Aref	D'Ascenzo	Gao	Gurm	Huang	Kulkarni	Kuno	Matheny	Mortazavi	Rayfield	Wang
Question Relates To Patient Benefit?	**Is There A Health Question Relating To Patient Benefit?**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Target Condition Applicability?	**Is There Concern That The Target Condition As Defined Does Not Match The Research Question?**	No	No	No	No	No	No	No	No	No	No	No
Data Suitable For Clinical Question?	**Is The Data Suitable To Answer The Clinical Question, I.E. Does It Capture The Relevant Real-World Heterogeneity, And Is It Of Sufficient Detail And Quality?**	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Patient Applicability?	**Is There Concern That The Included Patients Do Not Match The Research Question? Y/N**	No	No	No	No	No	No	No	No	No	No	No

Patient Selection Bias?	**Could The Selection Of Patients Have Introduced Bias?**	No	No	The exclusion criteria were: (1) STEMI but No primary PCI; or (2) acute non-STEMI or unstable angina	Patients who underwent coronary artery bypass grafting during the same hospitalizatio n were excluded from the analysis since a post –operative transfusion could Not be distinguished from post PCI transfusion. The choice of vascular access, procedural anticoagulati on and decision to transfuse was as per the operator preference guided by institutional policy and practice.	Possibly: We excluded PCIs that were Not the first during a single hospitalizati on (n = 32,999), procedures with same-day discharge (n = 41,570), missing serum creatinine before or after the procedure (n = 208,158), and procedures on patients already on dialysis at the time of their PCI (n = 24,271). T	Unclear	Excluded patients whose pre- and post-procedural creatinine and haemoglobi n data were missing. Although creatinine levels in relatively stable patients were Not consistently assessed, these exclusions could have created a bias in our results.	Unclear	We only included the first PCI procedure within the same episode because we have unique coded identifiers per admission and procedure identifiers linked to this. If a patient had a second PCI in a different admission, we treated this as an independe nt procedure because we did Not have patient identifiers. We added an exclusion for patients who	The patients in the data set might have skewed demographic s, as they were predominant ly white, which has implications with external validity.	Possibly: The excluded criteria as follows: myocardia l infarction patients or elevated pre-procedural cardiac troponin I (cTnI) or creatine kinase-MB fraction (CK-MB), PCI for more than one artery, coronary artery with thrombosis , translumin al extraction-atherectom y therapy for culprit artery, severe heart failure (EF <45% or NT-pro BNP >2,0 00), severe valve

										underwent coronary artery bypass grafting (CABG) because the high risk of bleeding after CABG may obscure the bleeding risk attributable to PCI alone; these cases were Not excluded in the primary prior model.		diseases.

Algorithm Applicability?	**Is There Concern That The Algorithm, Its Conduct, Or Interpretation Differ From The Research Question?**	No	No	No	No	No	No	No	No	No	No	No
Bias In Algorithm? 2.1 Were Predictors Defined And Assessed In A Similar Way For All Participants ? 2.2 Were Predictor Assessments Made Without KNowledge Of Outcome Data? 2.3 Are All Predictors Available At The Time The Model Is Intended To Be Used?	**Could The Variable Selection, Predictor Selection Or Interpretation Of The Machine Learning Have Introduced Bias?**	No	No	No	No	No	No	No	No	No	No	No

	Treatment Of Missing Data	Multiple imputations by chained equations	Missing data with imputation	Not mentioned	Missing predictors imputed to be the overall median for continuous values and mode for categorical variables	Following the same strategy used in the baseline model development, missing variables were imputed by the most common value for categorical variables and median for continuous variables	The second step included data pre-processing using variable encoding. We aimed to maximize the information contained within a variable and therefore did Not discard any records with missing values. Rather, we coded all missing values for all variables as -1 to include missing information as a separate	No mention of missing data	Not stated	Simple data imputation strategy due to the low rate of missing values in the data se	After the entire cohort was obtained, patients were removed from the patient cohort if any of the 86 variable data points, including bleeding data, were missing.	Mean imputation

							category					
	Hold-Out= Simplest Cross-Validation Where The Dataset Is Split Into A 'Training' And 'Testing' Set.	No	Yes	No	Yes	Yes	Yes	No	No	No	No	No

	Leave-One-Out Cross-Validation=When Number Of Folds Equals The Number Of Instances In The Data Set.	No	No	No	No	No	No	No	No	No	No	No
	N-Fold Cross-Validation=When The Train Dataset Is Split Into "N" Folds.	Yes	No	No	No	No	No	No	Yes	Yes	Yes	Yes
	External Validation Done	No	Yes	Yes	No	Yes	No	No	No	No	No	No
Source And Size Of External Validation Dataset	**Randomised Controlled Trial**		The best-performing model for each study outcome (the PRAISE score) was tested in an external validation cohort of 3444 patients with ACS pooled from a randomised			Temporally validated with more contemporary dataset						

			controlled trial and three prospective registries.									
	Prospective Cohort	No	No	Yes	No	No	No	Yes	No	No	No	No
	Registry	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
	Method Of Validation	5-fold cross validation	20-80 test train split + external validation	prospective observational study	30:70 split	split + external validation	25-75 split	prospective observational study	3-fold cross validation	5-fold cross validation	10-fold cross validation	
Size Of Validation Dataset (N)		95961	7409	316	30966	93902	21004	14273	2638	663293	3900	2177

Improved Outcome Prediction	**Is There Evidence Of Improved Risk Prediction?**	A boosted ensemble algorithm (AdaBoost) had optimal discriminati on with AUC of 0.927 (95% CI 0.923–0.929) compared with AUC of 0.913 for XGBoost (95% CI 0.906–0.919, P=0.02), AUC of 0.892 for Random Forest (95% CI 0.889–0.896, P	The PRAISE score showed an AUC of 0·82 (95% CI 0·78–0·85) in the internal validation cohort and 0·92 (0·90–0·93) in the external validation cohort for 1-year all-cause death; an AUC of 0·74 (0·70–0·78) in the internal validation cohort and 0·81 (0·76–0·85) in the external validation cohort for 1-year myocardia	The mortality risk prediction Nomogram achieved good discriminatio n for in-hospital mortality (training set: C-statistic=0.98 7; model calibration: P=0.722; validation set: C-statistic=0.98 4, model calibration: P=0.669). Area under the curve (AUC) values for the training and validation sets are 0.987 (95% CI: 0.981–0.994, P=0.003) and 0.990 (95% CI: 0.987–0.998, P=0.007)	AUC: full model = 0.888 (95% CI 0.877–0.899), reduced model AUC = 0.880 (95% CI, 0.868–0.892), p for difference 0.003, NRI = 2.77%, p = 0.007)	Compared with the baseline model that uses 11 variables, the best model used 13 variables and achieved a significantly better area under the receiver operating characteristi c curve (AUC) of 0.752 (95% confidence interval [CI] 0.749–0.754) versus 0.711 (95% CI 0.708–0.714), a significantly better Brier score of 0.0617 (95% CI 0.0615–0.0618) versus 0.0636 (95% CI 0.0634–0.0638), and	Compared to the currently used models for AKI, bleeding and death prediction , our models showed a significant ly higher AUC (range 1.6% – 5.6%), IDI (range 4.9% – 7.2%) and NRI (range 0.07 – 0.61).	Neural network performed similarly to logistic reg	While the logistic regression results in this study were similar to those found in the past for this clinical domain [25–31,47], the optimizati on process was limited to backward variable selection using each of the four optimizati on methods. This limitation may have contribute d to the insensitivi ty of the LR models to the optimizati on processes,	Logistic reg and xgradient boost, x gradient boost improved performan ce of logistic reg	The AI-BR model accurately predicts bleeding post and accurately predicts bleeding post	AUC moderate, group which dropped data performed best

			l infarction; and an AUC of 0·70 (0·66–0·75) in the internal validation cohort and 0·86 (0·82–0·89) in the external validation cohort for 1-year major bleeding.			a better calibration slope of observed versus predicted rate of 1.008 (95% CI 0.988–1.028) versus 1.036 (95% CI 1.015–1.056). The best model also had a significantly wider predictive range (25.3% versus 21.6%, p < 0.001) and was more accurate in stratifying AKI risk for patients			and may have biased the findings that SVM models were superior to LR models			

Methods Available?	**Are The Different Parts Of The Prediction Modelling Pipeline Available To Others To Allow For Methods Reproducibility, Including: The Statistical Code For 'Pre-Processing', And The Modelling Workflow (Including The Methods, Parameters, Random Seeds, Etc. Utilised)?**	Yes, code is available	Yes, No code	Yes, No code	No	Yes, code Not available	Yes, pipeline clear but No code	No	Yes, code Not available	Yes, pipeline clear but No code	Yes, code Not available	No statistical code for 'pre-processing', but workflow is clear
Metrics Clinically Relevant?	**Are The Reported Performance Metrics Relevant For The Clinical Context In Which The Model Will Be Used?**	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Interpretable By Clinicians?	**Is There Evidence That Clinicians And Patients Find The Model And Its Output (Reasonably) Interpretable?**	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Results Clinically Justified?	**Is The Reported Gain In Statistical Performance With The Ml/Ai Algorithm Clinically Justified In The Context Of Any Trade-Offs?**	Yes	Yes	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes	Yes	Unclear
Real World Effectiveness	**Is There Evidence Of Real World Model Effectiveness In The Proposed Clinical Setting?**	Yes	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes	Yes
Cost Effectiveness	**Is There Evidence Of Cost Effectiveness In The Proposed Clinical Setting?**	No	No	No	No	No	No	No	No	No	No	No

Supplementary Table 8. Clinical predictors and outcomes involved in the training of different ML models.

Article	Clinical predictors and outcomes included in ML models	
Mortazavi 2019	Demographic	Demographic characteristics and medical history: Age (age > 70y, age ≤ 70y), body mass index (BMI ≤ 30), chronic lung disease, chronic kidney disease (no, mild, moderate or severe), GFR, sex, diabetes (composite, non-insulin treatment, insulin-requiring), currently receiving dialysis, NYHA (composite, 1, 2, 3 or 4), history of cerebrovascular disease, history of peripheral arterial disease, previous PCI, pre-procedural haemoglobin (Hb ≤ 13g/dL, Hb > 13g/dL), pre-procedural creatinine
	Pre-procedural imaging	Pre-procedural TIMI flow grade, pre-procedural LV ejection fraction
	Intervention	Procedural characteristics: PCI lesion composite (1: Proximal right, mid-LAD, or proximal circumflex, 2: Proximal LAD, 3: Left main, 0: Other), proximal LAD PCI, left main PCI, vessel disease composite, 2-vessel or 3-vessel disease, lesion complexity, SCAI lesion class,
	Intervention (time-specific)	CAD presentation, STEMI, stenosis % before treatment)
	Procedural or post-procedural complications	PCI status (elective, urgent, emergency or salvage)
		Cardiogenic shock (at start of PCI, within 24h, at start of PCI or within 24h, composite), cardiac arrest within 24h, PCI status and shock (composite, 1: Salvage and shock (within 24 h and at start of PCI), 2: Salvage or shock (within 24 h and at start of PCI), 3: Shock within 24 h or at start of PCI, 4: Emergent procedure, 5: Urgent procedure, 6: Elective procedure),
	Outcomes	subacute stent thrombosis
		In-hospital major bleeding within 72 hours after PCI
Rayfield 2020	Demographic	

	Age, gender, recent myocardial infarction, presence of cardiogenic shock, presenting symptoms, presence of angina, presence of acute coronary artery disease symptoms, presence of unstable angina, presence of non-ST segment elevation myocardial infarction, other symptoms (respiratory, abdominal, etc), Canadian Cardiovascular Society grading score for angina, New York Heart Association classification of congestive heart failure symptoms, presence of diabetes, presence of hypertension, body mass index, hyperlipidaemia, family history of coronary artery disease, current smoking status, history of prior myocardial infarction, prior PCI, prior coronary artery bypass grafting, presence of peripheral arterial disease, cerebrovascular disease, dialysis status, history of chronic lung disease, peptic ulcer disease, presence of cancer diagnosis, metastatic disease status, cardiac arrest within 24 hours, pre-PCI, left ventricular ejection fraction, indication for PCI, presence of shock at the start of PCI, thrombolytic administration, diastolic blood pressure, systolic blood pressure, heart rate, troponin T level prior to PCI, serum creatinine prior to PCI, glomerular filtration rate, pre-PCI haemoglobin, presence of intra-aortic balloon pump, presence of other ventricular support devices, access site femoral, access site brachial, access site radial, left main disease >50%, proximal left anterior descending artery stenosis ≥70%, middle to distal left anterior descending ≥70%, right coronary artery stenosis ≥70%, left circumflex artery stenosis ≥70%, right acute marginal artery stenosis ≥70%, number of diseased vessels, PCI performed on culprit lesion, PCI performed on non-culprit lesion, PCI of chronic total occlusion performed, number of segments treated, number of vessels treated, number of lesions treated, number of native lesions treated, worst pre-PCI TIMI flow of treated lesions, any complex lesions treated, presence of thrombus in the lesion, any bifurcation lesion treated, worst post-PCI TIMI flow of

		treated lesion, any treated lesion ≤20% post-PCI stenosis, number of bare-metal stents used, number of drug-eluting stents used, total number of stents, maximum device diameter (mm), left main intervention performed, left anterior descending intervention performed, left circumflex intervention performed, right coronary artery intervention performed, use of fondaparinux, use of low-molecular-weight heparin, use of unfractionated heparin, use of aspirin, use of bivalirudin, use of other direct thrombin inhibitor, use of glycoprotein IIb/IIIa inhibitor, use of clopidogrel, use of ticlopidine, use of prasugrel, use of ticagrelor
	Pre-procedural imaging	Any intravascular ultrasound performed, Any fractional flow reserve performed
	Intervention	PCI
	Intervention (time-specific)	NA
	Procedural or post-procedural complications	NA
	Outcomes	Bleeding within 72h of PCI and prior to hospital discharge. Bleeding was defined according to the National Cardiovascular Data Registry (NCDR), which considers retroperitoneal, gastrointestinal, genitourinary, and intracranial bleeding, as well as access-site hematoma, as bleeding events.
D'Ascenzo 2021	Demographic	Clinical variables (including age, sex, diabetes, hypertension, hyperlipidaemia, peripheral artery disease, estimated glomerular filtration rate, previous myocardial infarction, previous percutaneous coronary intervention, previous coronary artery bypass graft, previous stroke, previous bleeding, malignancy, ST-segment elevation myocardial infarction presentation, haemoglobin, left ventricular ejection fraction), therapeutic variables (including treatment with

		beta blockers, angiotensin-converting enzyme inhibitors or angiotensin receptor blockers, statins, oral anticoagulation, and proton pump inhibitors)
	Pre-procedural imaging	Multivessel disease and complete revascularisation
	Intervention	Vascular access and percutaneous coronary intervention with drug-eluting stent
	Intervention (time-specific)	NA
	Procedural or post-procedural complications	NA
	Outcomes	1-year all-cause mortality, 1-year re-myocardial infarction, 1-year major bleeding
Wang 2020	Demographic	General information and history: Gender, Age, BMI, kg/m2, SBP, DBP, UAP (unstable angina previously), Hypertension, DM, P-CVD, P-PCI, Smoking, Drinking, F-CVD (family history of CVD), Biochemistry results: TC, mmol/L HDL-C, mmol/L LDL-C, mmol/L VLDL-C, mmol/L TG, mmol/L LPa, mg/dL TB, μmol/L UB, μmol/L CB, μmol/L UA, μmol/L Cr, μmol/L BUN, mmol/L eGFR, mL/min, Blood routine examinations: WBC, ×109 Lymphocyte, % Neutrophil, % Plt, ×109 MPV, fL CRP, mg/L CKMB, IU FBG, mg/L
	Pre-procedural imaging	NA
	Intervention	PCI + Procedure factors: FFR, IVUS, OCT, CTO, ACC/AHA TypeB2C, Left coronary artery, Total length of stents, Number of stents, Diameter of stent ≥2.5 mm, Calcification, PCI without dilation, Medications: anti-Hyper Med, Statins, anti-Plt Med, Trimetazidine, Fibrates, Cilostazol, Warfarin, PPI, Ezetimibe
	Intervention (time-specific)	NA
		NA

	Procedural or post-procedural complications Outcomes	Periprocedural myocardial infarction (PMI) after PCI *PMI def: [PMI3: cTnI >3-fold upper reference limit (URL); PMI5: cTnI >5-fold URL]
Gurm 2014	Demographic	History and risk factors (including current smoker (within 1 year), former smoker, hypertension, dyslipidaemia, family history of premature coronary artery disease, prior myocardial infarction, prior heart failure, prior valve surgery/procedure, prior peripheral artery disease, prior percutaneous coronary intervention, prior coronary artery bypass graft, prior implantable cardioverter-defibrillator, height, weight, age, currently on dialysis, cerebrovascular disease, chronic lung disease, diabetes/diabetes therapy (diet, oral rx, insulin), gastro-intestinal bleeding, valve disease, surgery within the prior 7 days, atrial fibrillation, cardiac transplant, cardiac arrest), clinical presentation (including pre-operative evaluation prior to non-cardiac surgery, cardiogenic shock within 24 hours prior to presentation, cardiac arrest within 24 hours prior to presentation, coronary artery disease presentation, angina classification within 2 weeks, NYHA class within 2 weeks, exercise stress test results), pre-procedural lab values (including creatine kinase-MB, troponin I, troponin T, creatinine, haemoglobin)
	Pre-procedural imaging	Cardiomyopathy/left ventricular systolic dysfunction, stress/imaging study performed, stress echo imaging results, cardiac CTA performed, cardiac CTA results, coronary calcium score
	Intervention	NA
	Intervention (time-specific)	NA
		NA

	Procedural or post-procedural complications	NA
Kulkarni 2021	Outcomes	
	Demographic	Age, gender, race, body mass index, current smoker, diabetes, hypertension, dyslipidaemia, diabetes therapy, chronic lung disease, chronic kidney disease, current dialysis, anaemia, family history of CAD, past history of myocardial infarction, past history of heart failure, past history of peripheral arterial disease, past history of valve surgery, past history of PCI, past history of CABG, past history of cerebrovascular disease, past history of heart failure within 2 weeks, cardiogenic shock within past 24 hours, cardiac arrest within past 24 hours, NYHA class within past 2 weeks, past history of other major surgery, time elapsed since last CABG (days), time elapsed since last PCI, time since onset of symptoms, anginal classification within 2 weeks, cardiomyopathy or LV dysfunction, CAD presentation, insurance (medicare/medicaid only or multiple), medications (thrombolytics, anti-anginal - beta-blockers, calcium channel blockers, long-acting nitrates, ranolazine, other), laboratory investigations (pre-PCI CKMB, pre-PCI TnI, pre-PCI TnT, pre-PCI serum creatinine, pre-PCI haemoglobin), estimated glomerular filtration rate
	Pre-procedural imaging	Stress echocardiogram, SPECT stress test, exercise stress test, stress test with CMR, coronary calcium score, calcium score, cardiac CTA, degree of vessel stenosis (left main stem, proximal LAD, mid/distal LAD, circumflex artery, ramus, RCA, proximal LAD graft, mid/distal LAD graft, circumflex artery graft, RCA graft, ramus graft), dominance (left, right or co-dominant), LV ejection fraction, number of diseased vessels, diagnostic catheterisation done, other procedure with diagnostic catheterisation, fluoroscopy time, fluoroscopy dose, contrast volume

	Intervention	Hospital status (outpatient, outpatient converted to inpatient or inpatient), admit source (emergency department, transfer from another acute care facility or other), inpatient for current episode, medications (glycoprotein IIb/IIa inhibitors, fondaparinux, low molecular weight heparin, unfractionated heparin, aspirin, bivalirudin, clopidogrel, ticlopidine, prasugrel, ticagrelor), number of drug-eluting stents, number of bare metal stents, minimum stent diameter, total stent length, number of lesions, transradial access, vascular closure advice, intra-aortic balloon pump, other mechanical ventricular support
	Intervention (time-specific)	PCI status (urgent, emergency or salvage), door to balloon time, symptom action time, time of PCI start, day of PCI
	Procedural or post-procedural complications	Cardiogenic shock at start of PCI
	Outcomes	Acute kidney injury (AKI), bleeding, stroke, death, at least one adverse outcome
Gao 2020		
	Demographic	Sex, Killip classification, administration of beta-blocker, ACEi/ARB, CK-MB peak
	Pre-procedural imaging	Left main coronary artery disease, grading of thrombus, TIMI classification, slow flow, syntax score, left ventricular ejection fraction
	Intervention	Application of IABP
	Intervention (time-specific)	Symptom-to-door time, symptom-to-balloon time
	Procedural or post-procedural complications	NA
	Outcomes	In-hospital mortality

Al'Aref 2019	Demographic	Baseline demographics and clinical characteristics (including age, gender, ethnicity body mass index, median Canadian Cardiovascular Society class, previous PCI - 1,2, 3 or more, cerebrovascular disease, peripheral vascular disease, heart failure, malignant ventricular arrhythmia, COPD, diabetes mellitus, renal failure on dialysis, previous CABG, hemodynamic stability, ST-segment elevation on ECG, time in days since onset of myocardial ischemia/infarction), baseline chemistry values (including serum creatinine levels)
	Pre-procedural imaging	Ejection fraction
	Intervention	Periprocedural therapy and equipment used, hemodynamic instability, invasive coronary angiographic findings (including stenosis severity within coronary vasculature), day of the week PCI was performed, and facility type
	Intervention (time-specific)	Reperfusion time intervals in acute myocardial infarction patients
	Procedural or post-procedural complications	Periprocedural complications and outcomes, the occurrence of postprocedural complications was defined as the occurrence of stroke, Q-wave myocardial infarction, acute occlusion in the target lesion or in a significant side branch, vascular injury at the access site requiring intervention, renal failure, emergency cardiac surgery, stent thrombosis, and coronary perforation or the need to emergently return to the catheterization laboratory for PCI.
	Outcomes	In-hospital mortality
Matheny 2007	Demographic	Age, acute heart attack, body mass index, CHF class, CHF on presentation, creatinine >2.0mg/dL, diabetes, family history of heart disease, heart rate, history of COPD,

		history of peripheral vascular disease, history of stroke, hyperlipidaemia, hypertension, prior PCI
	Pre-procedural imaging	NA
	Intervention	Intra-aortic balloon pump (IABP)
	Intervention (time-specific)	Elective, emergent or urgent case
	Procedural or post-procedural complications	Shock, unstable angina
Huang 2018	Outcomes	Post-procedural in-hospital mortality
	Demographic	Age, sex, race (White, Black or African American, Asian, American Indian or Alaskan Native, Native Hawaiian or Pacific Islander), ethnicity (Hispanic or Latino ethnicity), current/recent smoker, hypertension, dyslipidaemia, family history of premature CAD, prior MI, prior heart failure, prior valve surgery/procedure, prior PCI, most recent PCI date, prior CABG, most recent CABG date, height, weight, cerebrovascular disease, peripheral arterial disease, chronic lung disease, diabetes mellitus, diabetes therapy, CAD presentation, anginal classification within 2 weeks, anti-anginal medication within 2 weeks, beta blockers, calcium channel blockers, long-acting nitrates, ranolazine, other anti-anginal agent, heart failure within 2 weeks, cardiomyopathy or left ventricular systolic function, NYHA class within 2 weeks, pre-procedure creatinine, pre-procedure GFR, pre-procedure haemoglobin
	Pre-procedural imaging	Stress or imaging studies (i.e. if an exercise stress test, stress echocardiogram, stress testing with SPECT MPI, stress testing with CMR, cardiac CTA or coronary calcium scoring was performed), pre-PCI left ventricular ejection fraction

	Intervention	Admit source (emergency department, transfer in from another acute care facility or other), thrombolytics, IABP, other mechanical ventricular support
	Intervention (time-specific)	PCI status (elective, urgent, emergency or salvage), IABP timing, other mechanical ventricular support timing
	Procedural or post-procedural complications	Cardiogenic shock within 24 hours, cardiac arrest within 24 hours,
	Outcomes	Acute kidney injury (AKI)
Kuno 2021		
	Demographic	Age, chronic kidney disease, previous heart failure, diabetes mellitus, cerebrovascular disease, heart failure at admission, cardiogenic shock at admission, cardiopulmonary arrest at admission, ST elevation myocardial infarction, non-ST elevation myocardial infarction/unstable angina, pre-procedural haemoglobin (<10g/dL), >3g/dL decrease in haemoglobin level versus relative decrease of 20% in haemoglobin
	Pre-procedural imaging	NA
	Intervention	Use of intra-aortic balloon bump
	Intervention (time-specific)	NA
	Procedural or post-procedural complications	NA
	Outcomes	Acute kidney injury (AKI)

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Abstract	2	PRISMA 2020 Abstract checklist: ● Identify the report as a systematic review. ● Provide an explicit statement of the main objective(s) or question(s) the review addresses. ● Specify the inclusion and exclusion criteria for the review. ● Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched. ● Specify the methods used to assess risk of bias in the included studies. ● Specify the methods used to present and synthesise results. ● Give the total number of included studies and participants and summarise relevant characteristics of studies. ● Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured). ● Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision). ● Provide a general interpretation of the results and important implications. ● Specify the primary source of funding for the review. ● Provide the register name and registration number.	3-4
INTRODUCTION			

Rationale	3	Describe the rationale for the review in the context of existing knowledge.	6-7
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	6-7
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	7-8
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	7
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	7
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	7-8
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	8-9
Data items	10	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	8-9
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	9-10
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	9

Synthesis methods	13	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. Describe any methods used to tabulate or visually display results of individual studies and syntheses. Describe any methods used to synthesise results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). Describe any sensitivity analyses conducted to assess robustness of the synthesised results.	9-10
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	NIL
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	NIL

RESULTS			
Study selection	16	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	11
Study characteristics	17	Cite each included study and present its characteristics.	11-12

Risk of bias in studies	18	Present assessments of risk of bias for each included study.	NIL
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	11-12
Results of syntheses	20	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. Present results of all investigations of possible causes of heterogeneity among study results. Present results of all sensitivity analyses conducted to assess the robustness of the synthesised results.	11-12
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NIL
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	11-12
DISCUSSION			
Discussion	23	Provide a general interpretation of the results in the context of other evidence. Discuss any limitations of the evidence included in the review. Discuss any limitations of the review processes used. Discuss implications of the results for practice, policy, and future research.	13-19
OTHER INFORMATION			
Registration and protocol	24	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	7

		Indicate where the review protocol can be accessed, or state that a protocol was not prepared. Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	21
Competing interest	26	Declare any competing interests of review authors.	21
Availability of data, code, and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	21

Supplementary Figure 1. PRISMA 2020 checklist.