Machine learning methods for perioperative anesthetic management in cardiac surgery patients: a scoping review

Santino R. Rellum[1,2]^, Jaap Schuurmans[1,2]^, Ward H. van der Ven[1,2]^, Susanne Eberl[1]^, Antoine H. G. Driessen[3]^, Alexander P. J. Vlaar[2]^, Denise P. Veelo[1]^

[1]Department of Anesthesiology, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands; [2]Department of Intensive Care, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands; [3]Department of Cardiothoracic Surgery, Heart Center, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands

Contributions: (I) Conception and design: All authors; (II) Administrative support: SR Rellum, J Schuurmans; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: SR Rellum, J Schuurmans, WH van der Ven; (V) Data analysis and interpretation: SR Rellum, J Schuurmans; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Alexander P. J. Vlaar, MD, PhD, MBA. Department of Intensive Care, Amsterdam UMC, Location AMC, Meibergdreef 9, PO Box 22660, 1105 AZ Amsterdam, The Netherlands. Email: a.p.vlaar@amsterdamumc.nl.

Background: Machine learning (ML) is developing fast with promising prospects within medicine and already has several applications in perioperative care. We conducted a scoping review to examine the extent and potential limitations of ML implementation in perioperative anesthetic care, specifically in cardiac surgery patients.

Methods: We mapped the current literature by searching three databases: MEDLINE (Ovid), EMBASE (Ovid), and Cochrane Library. Articles were eligible if they reported on perioperative ML use in the field of cardiac surgery with relevance to anesthetic practices. Data on the applicability of ML and comparability to conventional statistical methods were extracted.

Results: Forty-six articles on ML relevant to the work of the anesthesiologist in cardiac surgery were identified. Three main categories emerged: (I) event and risk prediction, (II) hemodynamic monitoring, and (III) automation of echocardiography. Prediction models based on ML tend to behave similarly to conventional statistical methods. Using dynamic hemodynamic or ultrasound data in ML models, however, shifts the potential to promising results.

Conclusions: ML in cardiac surgery is increasingly used in perioperative anesthetic management. The majority is used for prediction purposes similar to conventional clinical scores. Remarkable ML model performances are achieved when using real-time dynamic parameters. However, beneficial clinical outcomes of ML integration have yet to be determined. Nonetheless, the first steps introducing ML in perioperative anesthetic care for cardiac surgery have been taken.

Keywords: Cardiac surgery; anesthesiology; perioperative care; artificial intelligence; machine learning

^ ORCID: Santino R. Rellum, 0000-0002-9971-7946; Jaap Schuurmans, 0000-0001-7471-8304; Ward H. van der Ven, 0000-0003-0508-0620; Susanne Eberl, 0000-0002-6609-5888; Antoine H. G. Driessen, 0000-0003-4712-6228; Alexander P. J. Vlaar, 0000-0002-3453-7186; Denise P. Veelo, 0000-0001-6196-1671.

Introduction

Cardiac surgery has gone through many advancements since the first use of cardiopulmonary bypass in 1953 (1). Milestones have been achieved with mechanical circulatory support devices and improved surgical techniques like the introduction of minimally invasive hybrid cardiac surgery. These innovations have enabled the inclusion of elderly and more high-risk patients. In addition, the improvement of perioperative management has further reduced the risk of complications. For example, the assessment of cardiovascular performance has improved with continuous ventricular function monitoring using miniaturized transesophageal echocardiography (TEE) probes (2). Also, the implementation of goal-directed therapy (GDT) has yielded beneficial effects in cardiac surgery (3). These innovations provide crucial diagnostic information needed to aid perioperative supportive and preventive care.

In recent years, machine learning (ML), a subset of artificial intelligence (AI), is the cause for a revolution in several medical fields (4-6), suggesting new possibilities within cardiac surgery. ML's exponential growth in medicine is made possible by the availability of large datasets and improvement in computing power, as it is a computer-controlled technique that automates analytical model building. Three main types of machine learning are distinguished (7): (I) supervised ML is concerned with the training of a model towards a known target variable (outcome). By differing the weighted effect of given labeled inputs (e.g., age, sex, cholesterol level, smoking status), it minimizes the prediction error of the desired output (for example, having cardiovascular disease or not). Most applications in medicine apply this principle of machine learning, using either a classification or regression model. (II) Unsupervised ML is when the algorithm obtains unlabeled data (e.g., large sets of radiological or histological images) and attempts to find patterns. This is a more exploratory method as the algorithm decides what classes and patterns best describe the data. (III) Reinforcement learning is the technique that is perhaps key to surpassing human capability. This method learns what actions lead to the highest possible reward. This reward is predefined and usually custom-tailored to the problem at hand. In this case, a training set is absent, but it is created by the inputs the model receives through interaction with the environment. An example of such a reward is each time an autonomous vehicle stays within its lane. Through positive and negative reinforcement, the self-driving model learns what the required behavior is and what actions lead to that scenario. For now, the use of ML in medicine is mainly limited to supervised methods.

The large quantities of data obtained during the perioperative phases of cardiac surgery are possibly suitable for a versatility of ML applications. Therefore, we have conducted a scoping review with the goal of identifying the full extent of current machine learning applications and their possible limitations for perioperative anesthetic management and risk assessment in cardiac surgery patients.

We present the following article in accordance with the PRISMA-ScR reporting checklist (available at https://dx.doi.org/10.21037/jtd-21-765).

Methods

We performed a scoping review methodology as defined by Arksey and O'Malley (8) to examine the extent and nature of the currently employed machine learning methods within perioperative management of cardiac surgery patients. The preferred reporting items for systematic reviews and meta-analyzes extension for scoping reviews (PRISMA-ScR) checklist (9) was used to guide detailed reporting.

Search strategy

Searches were compiled by a clinical librarian for three databases [MEDLINE (Ovid), EMBASE (Ovid), and Cochrane Library], using the following keywords: cardiothoracic surgery, anesthesiology, and artificial intelligence, including all synonyms (complete list of keywords is included in Appendix 1). Three reviewers (SR, JS, and WV) independently selected the articles, reaching a consensus on all included studies.

Study selection

Articles were included in the review if the following criteria were met: (I) reported on perioperative applications of ML in the field of cardiac surgery with relevance to the work of the anesthetist or the postoperative intensive care unit (ICU) admission; (II) evaluated the performance of the applied AI technique in non-simulated datasets; (III) available in English; (IV) published in the last thirty years (1991 to present). Machine learning application (i.e., advanced method) was considered if a traditional statistical technique (i.e., conventional method) was trained and subsequently
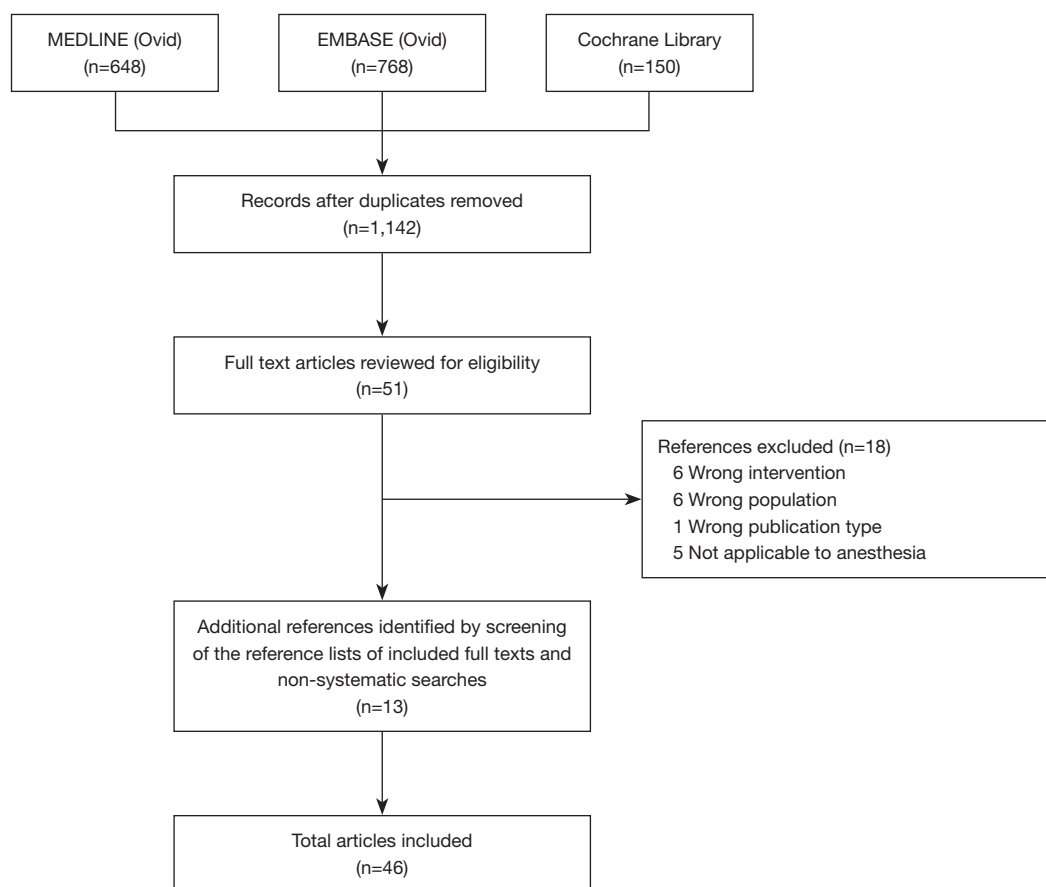
Figure 1 Flow chart of the literature selection process for the present article.

validated in different datasets. We excluded studies that solely focus on conventional methods, studies performed in children (age <18 years), or involving animals. In addition, editorials, commentary letters, and case reports were excluded.

Data extraction

For each article, data were extracted regarding: (I) perioperative phase; (II) type of cardiac surgery; (III) size of datasets; (IV) type of ML methods used; (V) area under the receiver operating characteristic curve (AUC). AUC, as a generalization referred to as the C-index, is used in this paper to provide insight into the mutual relationships of different models. An AUC value between 0.7 and 0.8 is considered satisfactory in this scoping review (10). A meta-analysis was not deemed appropriate given the wide variety of included subjects and ML techniques in the included studies.

Results

A total of 1,566 articles were identified, with a remainder of 1,142 articles after deduplication. Of these, 51 full-text studies were assessed for inclusion after the screening of titles and abstracts. Eighteen full-text articles were excluded for various reasons (*Figure 1*). We found thirteen additional articles through citation tracking and non-systematic searches, resulting in 46 included articles. We identified three distinguishable categories: (I) event and risk prediction, (II) hemodynamic monitoring, and (III) automation of echocardiography.

Preoperative

Risk scores enable the assessment of preoperative risks to help in the stratification of patients. Additionally, they inform and guide patients and their relatives in shared-decision making and are used in cost-benefit analyzes. However, a

known drawback of widely used risk scores in cardiac surgery is that they do not fit all (sub)populations, especially causing underperformance in high-risk patients (11).

Prediction of mortality

The established European System for Cardiac Operative Risk Evaluation (EuroSCORE) and the Society of Thoracic Surgeons (STS) score are based on conventional logistic regression analysis. These two clinical scores are among the most commonly used mortality risk scores, with AUCs ranging from 0.74 to 0.80 (*Table 1*). Important to note are the predictive discrepancies that persist for these scores in a few cardiothoracic operations and some subpopulations, especially in high-risk patients (33-38). Contrary to this underperformance, several advanced machine learning models demonstrated their added value in elderly and rheumatic heart disease (RHD) subpopulations. Within the elderly population, six perioperative variables (not further specified by the authors) were found to be strongly correlated with mortality. Based on those variables, a logistic regression (LR) model, Bayesian network (BN), and an artificial neural network (ANN) produced AUCs of respectively 0.854, 0.931, and 0.941, clearly outperforming the EuroSCORE that had an AUC of 0.648 in this population (19). Overall, the main mortality predictors in RHD were found to be left atrium size, high creatinine, tricuspid procedure, reoperation, and pulmonary hypertension. Using a random forest (RF) model, a new clinical score, the RheSCORE, was built on those predictors. With an area of 0.98, it outperforms the EuroSCORE II, which produces an AUC of 0.857 based on essentially the same predictors (25).

However, in a mixture of cardiac surgery procedures, the two aforementioned clinical scores perform similarly or slightly less than advanced models (12,13,20). An ANN yielded comparable predictive properties to the EuroSCORE (AUC 0.80 *vs.* 0.79), with only a small advantage in the case of valve procedures (AUC 0.76 *vs.* 0.72, P value =0.0001) (13). Assembling four ML models [gradient boosting machines (GBM), RF, support vector machines (SVM), and Naïve Bayes (NB)] created a significant but modest benefit with an AUC of 0.795 versus 0.737 for the EuroSCORE II (12). Similarly, modest advantages in accuracy and AUC were seen comparing an advanced ML model [extreme gradient boosting machine (XGBoost)] to the STS clinical score. Interestingly, despite both the STS score and the XGBoost being well-calibrated and having a high area under the curve (respectively 0.808 and 0.795),

they identified a large proportion of different patients as being at risk (20). Even one of the first clinical scores, the Parsonnet score, still holds value in predicting in-hospital mortality with a comparable AUC to an advanced LR and ANN model (0.829, 0.852, and 0.873, respectively) (14).

Also, when comparing advanced ML methods, little difference in predicting performance is seen (14-16,21,22,39,40), with only a slight advantage for nonlinear models [ANN, BN, and multilayer sigmoid perceptron (MLP)] over linear LR models (13,19,24). The majority of these studies use a set of preoperative data, including demographic characteristics, medical history, and type of surgery performed. Adding intraoperative hypotension as a dynamic parameter to these preoperative data showed improved AUCs for advanced LR, RF, and XGBoost models. At the same time, an SVM and ANN did not benefit from this added parameter, outputting AUCs of 0.66 and 0.70, respectively (17).

Risk survival scores in heart transplantation

There are currently three main risk scores for heart transplant patients based on conventional logistic regression methods: the Donor Risk Index (DRI) (41), the Index for Mortality Prediction After Cardiac Transplantation (IMPACT) (42), and the Risk-Stratification Score (RSS) (43). These produce C-indices ranging from 0.55 to 0.57 for overall survival (*Table 1*). We identified two studies that compared an advanced model directly to these risk scores, obtaining slightly better performances (AUCs between 0.62 to 0.66) (26,27). Comparing only advanced models in their ability to predict 1-year mortality, both linear and nonlinear models show similar results with moderate AUCs consistently ≤0.66 (28-31). Only in predicting 5-year mortality after heart transplantation, an advanced GBM model transcended other machine learning models, generating an area of 0.717 (32).

A different application of advanced modeling was used to stratify patients on a heart transplant waiting list. The applied neural network only moderately determined the most likely patient status: still waiting, transplanted, or deceased at three different time points (44).

Intra-operative

A staggering eighty percent of all intra-operative alarms in cardiac surgery, mainly hemodynamic warnings, do not require therapeutic intervention (45). Many redundant alarms involve artifacts or expected procedure-specific events. This is not fully acceptable as it can cause distraction

Table 1 Area under the curve values in validation datasets for mortality prediction at different time-points

Study	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]	Definition mortality
		Training	Test		Category	Subtype		
Mixed surgical population								
Allyn et al. (12)	Mix	N/A		Preoperative	Conventional	LR (EuroSCORE II)	0.737	Postoperative, time point not specified
		4,564	1,956	Preoperative	Advanced	LR	0.742	
						RF + NB + GBM + SVM	0.795	
Nilsson et al. (13)	Mix	N/A		Preoperative	Conventional	LR (EuroSCORE I)	0.79	Death during hospitalization or ≤30 days after cardiac surgery
		13,771	4,591	Preoperative	Advanced	LR	0.78	
						ANN	0.80	
Peng, Peng (14)	Mix	N/A		Preoperative	Conventional	LR (Parsonnet)	0.829	Postoperative, time point not specified
		637	315	Pre-, and postoperative	Advanced	LR	0.852	
						ANN	0.873	
Orr (15)	Mix	732	380	Pre-, and postoperative	Advanced	PNN	0.81	Not specified
Benedetto et al. (16)	Mix	N/A		Preoperative	Conventional	LR (EuroSCORE I)	0.76	Postoperative, in-hospital
						LR (EuroSCORE II)	0.77	
		20,133	8,628	Preoperative	Advanced	LR	0.80	
						RF	0.80	
						Naïve Bayes	0.77	
						ANN	0.77	
Fernandes et al. (17)	Mix	3,761	1,254	Pre-, and intraoperative	Advanced	LR	0.80	Postoperative, time point not specified
						RF	0.83	
						XGB	0.85	
						SVM	0.66	
						ANN	0.70	
Zhong et al. (18)	Mix	5,475	1,369	Pre-, intra-, postoperative	Advanced	LR	0.86	30-day mortality
						RF	0.88	
						XGBoost	0.90	
						ANN	0.64	
Celi et al. (19)	Mix in elderly^	N/A		Preoperative	Conventional	LR (EuroSCORE I)	0.648	In-hospital, time point not specified
		116	49	Pre-, intra-, postoperative	Advanced	LR	0.854	
						BN	0.931	
						ANN	0.941	
CABG and/or valve surgery								
Kilic et al. (20)	CABG + valve	N/A		Preoperative	Conventional	LR (STS PROM)	0.795	Death during hospitalization or ≤30 days after cardiac surgery
		10,071	1,119	Preoperative	Advanced	XGBoost	0.808	

Table 1 (*continued*)

Table 1 (*continued*)

Study	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]	Definition mortality
		Training	Test		Category	Subtype		
Lippmann, Shahian (21)	CABG	40,480	40,126	Preoperative	Advanced	LR	0.762	Not specified
						Bayesian model	0.748	
						Committee classifier	0.764	
						Single-layer MLP	0.754	
						Two-layer MLP	0.761	
						Three-layer MLP	0.761	
Mendes et al. (22)	CABG	1,053	262	Pre-, intra-, postoperative	Advanced	LR	0.86	Death 30-day after CABG
						ANN	0.85	
Tu, Guerriere (23)	CABG	4,782	5,517	Preoperative	Advanced	LR	0.77	Postoperative, time point not specified
						ANN	0.78	
Lippmann (24)	CABG	1,257[†]		Pre-, intra-, postoperative	Advanced	LR	0.705[‡]	Not specified
						Single-layer MLP	0.760[‡]	
						MLP		
						MLP-Committee		
Mejia et al. (25)	Valve in RHD	N/A		Preoperative	Conventional	LR (B-Parsonnet)	0.876	Death during hospitalization or ≤30 days after cardiac surgery
						LR (EuroSCORE II)	0.857	
						LR (InsCor)	0.835	
						LR (AmblerSCORE)	0.831	
						LR (Guaragna)	0.816	
						LR (New York)	0.834	
		2,919[†]		Preoperative	Advanced	RheSCORE[1]	0.98	
Heart transplantation								
Yoon et al. (26)	HTx	N/A		Preoperative	Conventional	LR (DRI)	0.529	Generalization of four time point at 3-month, 1-, 3-, and 10-year
						LR (IMPACT)	0.527	
						LR (RSS)	0.544	
		66,306	16,576	Preoperative	Advanced	ToPs/R[2]	0.577	
Nilsson et al. (27)	HTx	N/A		Preoperative	Conventional	LR (DRI)	0.56	1-year mortality
						LR (IMPACT)	0.61	
						LR (RSS)	0.61	
		41,780	8,569	Preoperative	Advanced	IHTSA[3]	0.650	
Shah et al. (28)	HTx	4,054[†]		Preoperative	Advanced	LR	0.60	1-year mortality or retransplantation
						ML model not specified	0.64	

Table 1 (*continued*)

6982

Rellum et al. Machine learning for perioperative care in CTS patients

Table 1 (*continued*)

Study	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]	Definition mortality
		Training	Test		Category	Subtype		
Villela *et al.* (29)	HTx	18,612[†]		Preoperative	Advanced	LR	0.62	1-year mortality or retransplantation
						Stacking of GBM	0.66	
Bravo *et al.* (30)	HTx after LVAD	7,700[†]		Preoperative	Advanced	LR	0.63	1-year mortality or retransplantation
						ML model not specified	0.61	
Miller *et al.* (31)	HTx	45,182	11,295	Preoperative	Advanced	LR	0.65	1-year mortality
						Ridge regression	0.65	
						Regression LASSO	0.65	
						RF	0.63	
						NB	0.61	
						TA-NB	0.62	
						SVM	0.52	
						SGB	0.64	
						ANN	0.66	
Agasthi *et al.* (32)	HTx	12,189	3,047	Pre-, intra-, postoperative	Advanced	GBM	0.717	5-year mortality

[a], perioperative phase: pre-, intra, postoperative used variables in prediction models; [b], distinction between conventional and advanced models is explained in the methods section; [c], definitions of both the AUC and C-index is given in the methods section. [1], ensemble of thirteen advanced models; [2], trees of predictors based on three regression methods (cox regression, linear perceptron, and logistic regression); [3], international Heart Transplant Survival Algorithm based on an artificial neural network model. [†], ratio between training and validation set not reported; [‡], not all values are extractable as they are mainly displayed in bar graphs; [^], ≥80 years. ANN (1, 2, etc.), artificial neural network (one-layer, two-layer, etc.); AUC, area under the receiving operating characteristics curve for the validation sets; BN, Bayesian network; B-Parsonnet, 2000 Bernstein-Parsonnet score; CABG, coronary artery bypass graft surgery; GBM, gradient-boosted machine; HTx, heart transplantation; LASSO, least absolute shrinkage and selection operator; LVAD, left ventricular assist device; LR, logistic regression; Mix, various cardiac surgery patients with/without heart transplantation; ML, machine learning model; MLP, multilayer sigmoid neural network; TA-NB, tree-augmented NB; NB, Naïve Bayes; PNN, probabilistic neural network; RF, random forest; RHD, rheumatic heart disease; SGB, stochastic gradient boosting; SVM, support-vector machines; Valve, heart valve surgery; XGBoost, extreme gradient boosting.

or alarm fatigue (46). Advancements can be made to reduce the multitude of distracting alarms, as shown in a few AI applications in this chapter.

Predictions of hemodynamic instability

In 1997, Becker *et al.* (47) developed a monitoring system based on fuzzy logic to provide a continuous intuitive descriptive overview of a patient's hemodynamic status (e.g., 'preload is too high'). These hemodynamic interpretations were based on vital parameters and administered anesthetics. The validation process demonstrated promising results with a predictability of 99.5%. Compared to a simple threshold alarm, this system can help the physician to interpret changes quickly.

The hypotension prediction index (HPI) (48) is another monitoring application. It is an advanced logistic regression-based model that can predict a hypotensive event (mean arterial pressure <65 mmHg for at least one minute), regardless of current blood pressure, up to 15 minutes in advance (48). The model was developed using large datasets, including cardiac surgery patients. A recently published study demonstrated the high predictive capability of the HPI solely in cardiac surgery (49). ML can also be used to identify relationships between risks, as demonstrated

with three advanced RF models that adequately found cardiopulmonary bypass associated factors contributing to a reduction in right ventricular (RV) function (50).

Automation of intraoperative echocardiography (IOE)

Two articles were identified on the automation of ultrasound assessments that have the potential to enable a more efficient intraoperative workflow (51,52). As RV function analysis is both challenging and time-consuming, an AI-based automated RV strain assessment was compared with the most commonly used parameters: tricuspid annular plane systolic excursion (TAPSE), tissue Doppler-derived systolic tricuspid annulus motion velocity (S'), and RV fractional area change (FAC). A strong correlation was found between FAC and global longitudinal strain (GLS) over various RV function measurements on three different ultrasound machines (51).

The second AI application in ultrasound automation relates to the analysis of the mitral valve (MV) (52). Patients with a normal biventricular function who underwent an elective CABG surgery were included for ultrasound imaging to evaluate the clinical applicability and accuracy of an AI-based MV analysis software. An experienced echocardiographer captured three end-systolic frames of the MV in each patient. Postoperatively, these frames were analyzed with the AI software. The software automatically traced the valves, and three experienced examiners independently verified the valve tracings. Thus, creating three separate datasets for all frames, as the examiners could administer minor manual adjustments when deemed necessary. Subsequently, the software's six clinically relevant geometric parameters were calculated from the verified MV tracings (annulus anterolateral posteromedial diameter, annulus anteroposterior diameter, annular area, annulus nonplanarity angle, annulus total perimeter, and anterior and posterior leaflet areas). Statistical analyses showed a high precision for the calculated parameters in corresponding end-systolic frames in which only the valve tracings were verified by different examiners. Meaning that the latter did not affect the outcome (52).

Postoperative

Traditionally, risk scores were developed for mortality prediction alone. Only recently, morbidity has been incorporated in these models as they provide a marker for the quality of life (53). So far, most risk scores for postoperative outcomes are based on preoperative inputs and lack incorporation of intraoperative variables to improve on performance (54).

Morbidity in the ICU

The previously mentioned Parsonnet score initially developed for mortality prediction also generates acceptable AUCs in morbidity prediction concerning cardiovascular, respiratory, and neurological complications. Addressing these same outcomes, two advanced models (LR and ANN) show even better predictive capability in comparison, with the most significant advantage in predictive power for the ANN model with an AUC of 0.85 (14) (*Table 2*). In a recent study comparing advanced models reciprocally, an XGBoost model had the upper hand over ANN (18). However, these are outliers in terms of morbidity prediction. Most comparative studies in cardiac surgery show a reasonable predictive value for all advanced ML models with AUCs around 0.77 (55,56,62).

Acute kidney injury (AKI) is a common complication after cardiac surgery (63). Isolating patients at risk for AKI or renal replacement therapy (RRT) could guide perioperative treatment. Advanced predictive models based on GBM, LR, and ANN showed superior ability in identifying patients at risk for AKI and RRT as opposed to conventional risk scores based on LR (64,65). Overall, reasonable predictions in AKI prediction are seen for conventional and advanced models in most articles with AUCs ranging from 0.66 to 0.84 (61). Looking at individual studies in which various advanced models are directly compared with each other, it is noticeable that both MLP and XGBoost models are often better (24,58-60). Lastly, promising results have been found in evaluating the need for early continuous venovenous hemofiltration (CVVH) after cardiac surgery with comparable and accurate predictive results for both an ANN and an advanced LR model (66).

Prevention and early recognition of delirium are essential as it is associated with poor outcomes (67). We identified one study on this topic. It cross-examined seven advanced models comparing their performance in an imbalanced dataset (integral dataset) to their performance in a balanced dataset (i.e., 10-fold cross-validation applied). This was done in order to reduce overestimation. In line with their expectation, the predictive values of the models showed better performance in the balanced sets, with the best predictions for an LR and RF model and the least for a BN model that still performed sufficiently with

6984

Rellum et al. Machine learning for perioperative care in CTS patients

Table 2 Area under the curve values in validation datasets for postoperative morbidity prediction

	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]
		Training	Test		Category	Subtype	
Miscellaneous[1]							
Cevenini et al. (55)	CABG	545	545	Pre-, intra-, postoperative	Advanced	LR	0.781
						BL	0.778
						BQ	0.785
						HS	0.768
						DS	0.779
						k-NN	0.772
						ANN1	0.776
						ANN2	0.778
Chong et al. (56)	CABG	N/A		Preoperative	Conventional	LR (QMMI score)	0.752
		423	140	Preoperative	Advanced	LR	0.807
						ANN	0.886
Peng, Peng (14)	Mix	N/A		Preoperative	Conventional	LR (Parsonnet)	0.727
		637	315	Pre-, and postoperative	Advanced	LR	0.789
						ANN	0.852
Secluded morbidities							
Zhong et al. (18)	Mix	5,475	1,369		Septic shock		
				Pre-, intra-, postoperative	Advanced	LR	0.93
						RF	0.81
						XGBoost	0.96
						ANN	0.88
					Thrombocytopenia		
				Pre-, intra-, postoperative	Advanced	LR	0.87
						RF	0.89
						XGBoost	0.89
						ANN	0.83
					Liver dysfunction		
				Pre-, intra-, postoperative	Advanced	LR	0.82
						RF	0.89
						XGBoost	0.89
						ANN	0.70
Mufti et al. (57)	Mix	4,476	1,117		Agitated delirium		
				Pre-, intra-, postoperative	Advanced	LR	0.814
						RF	0.813
						NB	0.799
						BN	0.774
						SVM	0.811
						DT	0.772
						ANN	0.804

Table 2 (*continued*)

Table 2 (*continued*)

	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]
		Training	Test		Category	Subtype	
Acute kidney injury							
Lei *et al.* (58)	Aortic arch	627	270	Pre-, intra-, postoperative	Advanced	LR	0.65
						RF	0.71
						SVM	0.64
						LGM	0.80
Tseng *et al.* (59)	Mix	470	201	Pre-, and intraoperative	Advanced	LR	0.806
						RF	0.839
						DT	0.781
						XGboost	0.837
						SVM	0.825
						RF+XGBoost	0.843
Lee *et al.* (60)	Mix	1,005	1,005	Pre-, intra-, postoperative	Advanced	LR	0.70
						RF	0.68
						DT	0.71
						XGBoost	0.78
						SVM	0.69
						NN classifier	0.64
						Deep learning	0.55
Penny-Dimri *et al.* (61)	Mix	N/A		Preoperative	Conventional	LR (Cleveland Clinic)	0.71
						LR (Risk score)	0.74
						LR (Risk score)	0.75
		77,322	19,331	Preoperative	Advanced	LR	0.76
						GBM	0.76
						k-NN	0.66
						ANN	0.76
				Pre-, and intraoperative	Advanced	LR	0.77
						GBM	0.78
						k-NN	0.67
						ANN	0.77

[a], perioperative phase: pre-, intra, postoperative used variables in prediction models; [b], distinction between conventional and advanced models is explained in the methods section; [c], definitions of both the AUC and C-index is given in the methods section. [1], Mix of cardiovascular, respiratory, neurological, renal, infectious, and hemorrhagic complications. ANN (1, 2, etc.), artificial neural network (one-layer, two-layer, etc.). AUC, area under the receiving operating characteristics curve for the validation sets; BL, Bayes linear; BN, Bayesian network; BQ, Bayes quadratic; CABG, coronary artery bypass graft surgery; DS, direct score; DT, decision trees; GBM, gradient-boosted machine; HS, Higgins score; k-NN, k-nearest neighbor; LGM, light gradient machine; LR, logistic regression; Mix, various cardiac surgery patients with/without heart transplantation; NN, neural network; NB, Naïve Bayes; RF, random forest; SVM, support-vector machines; XGBoost, extreme gradient boosting.

an AUC of 0.77 (57).

Length-of-stay

Accurate estimation of ICU length-of-stay (LOS) is not only advantageous in the counseling of patients and their families but even so in the organization of the bed capacity and scheduling of the operating rooms. More so in recent times, with the increased scarcity of ICU beds due to the

6986

Rellum et al. Machine learning for perioperative care in CTS patients

ongoing COVID-19 pandemic (68).

The conventional EuroSCORE I is positively correlated with prolonged LOS, making it a suitable predicting tool for LOS (69). We identified one article demonstrating the superiority of an advanced ML model to the EuroSCORE I. It outperformed other advanced models as well and showed similar distinctiveness to physicians' LOS predictions (70) (*Table 3*). Other comparing data suggest that ANNs outperform other advanced models regarding LOS (76). By itself, an ANN developed in 1993 showed that it successfully stratified cardiothoracic surgery patients at risk of extended stay (>2 days) with an AUC of 0.70 (23). These promising results are even outperformed when ANNs are ensembled (72). Although slightly more modest in performance, advanced regression models still produce acceptable LOS predictions with AUCs ranging from 0.83 to 0.87 (73).

Mechanical ventilation

We identified two studies that elaborate on the prediction of prolonged mechanical ventilation and the chance of re-intubation. Both studies performed in a CABG subpopulation show minor differences in accuracy, sensitivity, and specificity in favor of an ANN over an advanced LR model (22,71).

Readmission

Given the high costs associated with readmission after hospital discharge, the ability to stratify the risk is essential for preventive measures. Improving upon existing conventional LR prediction models solely based on time-independent variables (e.g., 1-point lab values only postoperative) (77-79), an advanced XGBoost algorithm incorporating time-dependent factors (e.g., lab values at several time-points) demonstrated a better accuracy in predictive ability (74). Another but more complex ML tool called genetic programs performed equally well in accuracy to an advanced LR method. In contrast, an advanced ANN model in the same study showed a significantly worse predictive ability (75).

Discussion

This scoping review includes forty-six articles describing various ML techniques in cardiac surgery patients with relevance to perioperative anesthetic management and risk assessment. We identified three specific applications with the majority (n=41) on prediction analyses (e.g., mortality,

AKI, readmission), three articles on hemodynamic monitoring, including a form of prediction, and two studies that elaborate on ultrasound guidance. The combined overall data suggest that the current applications of ML techniques on stationary variables (e.g., hemodynamic parameter at one time-point) in the cardiac surgery population perform similar to conventional statistical methods (not using a training and validation set) concerning predictive capability.

In between ML methods, complex or straightforward in construction, only GBM more often shows superior outcomes than others. For one study, that can in part be attributed to the relatively updated registry it used (32). Major differences, however, are absent, with also high correlations in between these models, suggesting that they find similar patterns. Although major predictive improvements are not seen for single ANNs, it is beneficial to use them in an ensemble (27,72). However, the true power of ML seems to be triumphant when applied to more complex data such as full dynamic arterial waveforms or complex ultrasound images as opposed to stationary perioperative variables. Using these parameters yields real-time clinically insightful results (48,51,52) that are a valuable addition to current dynamic parameters (e.g., heart rate, stroke volume variation). Contemporary literature lacks data on clinical outcomes from ML implementation in the cardiac surgery population. Still, beneficial results are probably not long in coming, given the effectiveness seen in a different surgical population using real-time dynamic data in an ML model (80).

In contrast, comparable performances or modest improvements in prediction models are similar to other medical fields (81,82). An explanatory factor in this may be that these implementations are often based on manageable datasets that do not use an uncountable number of variables. While the strength of some advanced ML models is attributed to their ability to establish nonlinear relationships between variables in complex datasets that conventional methods have not previously demonstrated. In line with this, one of our included studies showed that the predicted risk correlation between an advanced and conventional model was very low. Although comparable in prediction, this suggests that they did not assign their prediction to the same features. Besides, the high complexity of ML models in small datasets poses a risk for overfitting (83). This occurs when irrelevant characteristics in the training data are marked as predictive parameters, causing underperformance in the test set (84). This problem might be avoided by

Table 3 Area under the curve values in validation datasets for prediction of additional-, prolonged-, or re-intervention and/or care

	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]
		Training	Test		Category	Subtype	
Renal replacement and CVVH							
Penny-Dimri *et al.* (61)	Mix	N/A		Preoperative	Conventional	LR (Cleveland Clinic)	0.80[d]
						LR (Risk score)	0.80[d]
						LR (Risk score)	0.81[d]
		77,322	19,331	Preoperative	Advanced	LR	0.82[d]
						GBM	0.83[d]
						k-NN	0.68[d]
						ANN	0.82[d]
				Pre-, and intraoperative	Advanced	LR	0.84[d]
						GBM	0.85[d]
						k-NN	0.69[d]
						ANN	0.84[d]
Bent *et al.* (66)	CABG + valve surgery	30	35	Perioperative	Advanced	LR	0.89[e]
						ANN	0.90[e]
Prolonged mechanical ventilation and reintubation							
Wise *et al.* (71)	CABG	N/A		Preoperative	Conventional	LR	0.698[f]
		590	148	Preoperative	Advanced	ANN	0.714[f]
Mendes *et al.* (22)	CABG	1,053	262	Pre-, intra-, postoperative	Advanced	LR	0.67[f]
						ANN	0.72[f]
						LR	0.62[g]
						ANN	0.65[g]
Length of stay							
Rowan *et al.* (72)	Mix	480	240	Pre-, intra-, postoperative	Advanced	Ensemble ANNs	0.901
Barbini *et al.* (73)	CABG + valve surgery	2,605	651	Pre-, intra-, postoperative	Advanced	NB	0.859
Meyfroidt *et al.* (70)	Mix	N/A		Preoperative	Conventional	LR (EuroSCORE I)	0.726
				Pre-, intra-, postoperative		Nurses' prediction	0.695
						Physician's prediction	0.758
		461	499		Advanced	Gaussian processes	0.758
30-day readmission							
Manyam *et al.* (74)	CABG	1,042	261	Time-independent[1]	Advanced	XGBoost	0.627
				Time-dependent + time-independent[1]	Advanced	XGBoost	0.868

Table 3 (*continued*)

6988

Rellum et al. Machine learning for perioperative care in CTS patients

Table 3 (*continued*)

	Surgery	Datasets		Phase[a]	Model type[b] (clinical score)		AUC[c]
		Training	Test		Category	Subtype	
Engoren *et al.* (75)	CABG	2,644	2,711	Pre-, intra-, postoperative	Advanced	LR	0.644
						Genetic programs	0.654
						ANN	0.537
Graft failure at 5 years							
Agasthi *et al.* (32)	HTx	12,189	3,047	Pre-, intra-, postoperative	Advanced	GBM	0.716

[a], perioperative phase: pre-, intra, postoperative used variables in prediction models; [b], distinction between conventional and advanced models is explained in the methods section; [c], definitions of both the AUC and C-index is given in the methods section; [d], need for renal replacement therapy; [e], need for early continuous venovenous hemofiltration; [f], prolonged mechanical ventilation; [g], reintubation. [1], perioperative variables. ANN (1, 2, etc.), artificial neural network (one-layer, two-layer, etc.). AUC, area under the receiving operating characteristics curve for the validation sets; CABG, coronary artery bypass graft surgery; GBM, gradient-boosted machine; HTx, heart transplantation; k-NN, k-nearest neighbor; LGM, light gradient machine; LR, logistic regression; Mix, various cardiac surgery patients with/without heart transplantation; NB, Naïve Bayes; XGBoost, extreme gradient boosting.

implementing cross-validation as demonstrated for delirium in cardiac surgery (57). Still, traditional linear statistics may be more suitable for risk prediction models (85), as advanced models perform similarly but are more complex to develop.

Although not convincing in risk prediction models, ML does excel in datasets consisting of dynamic parameters. Future research should focus on these real-time applications of ML exploring patterns in complex datasets. Then promising results can be achieved, as demonstrated by the effective hypotension early warning system by Hatib *et al.* (48,49) and the automation of echocardiography in two other studies (51,52). Not only aimed at the development and validation of such models but also their clinical effectiveness in randomized controlled trials should be addressed.

Future directions and challenges

Anesthesia is pre-eminently the field where many dynamic physiological data can be collected digitally, especially in cardiac surgery, where the mean operative time is about three hours (86). The current application of anesthesia information management systems (AIMS) is expected to be well above 80% in academic centers (87). Future incorporation of machine learning into AIMS could facilitate the continuous development of these models, unlocking their full potential based on a regularly updating and expanding dataset. Still, it might be safer only to update

ML models in controlled research settings. Especially as neural networks, in particular, are not transparent and, at best opaque in how and what variables are processed in these algorithms (7,88). Not without reason, there is growing interest in explainable artificial intelligence, an AI in which decisions made by the model are transparent and better interpretable (89).

Nowadays, the application of ML models is approved by the US Food & Drugs Administration (FDA) and the European Commission when the algorithm in patients cannot improve on its capabilities (so-called locked models). This is done to ensure consistent, reproducible results and safety from the algorithm (90). The FDA is currently assessing regulatory modification possibilities (91) that enable the use of "unlocked" ML, taking into account potential safety issues.

There is still plenty of work to be done in the application and clinical evaluation of promising "locked" ML methods based on various perioperative dynamic variables. We suggest that future clinical trials implementing ML models address the following three primary outcomes: (I) will patient outcomes improve with ML-based diagnostic and treatment guidance, (II) does it improve workflow efficiency, and (III) is it cost-effective.

Limitations

Although we conducted a systematic search, we might have missed articles due to the broad range of included

topics and acronyms in the literature. This may have led to the incorrect exclusion of studies from the initial selection. Another limitation of our article is that we only descriptively summarized the data without a meta-analysis. Therefore, definitive conclusions cannot be drawn about AUC differences across different methodologies. Nevertheless, this article provides an overview of the current ML applications per perioperative phase in cardiac surgery, showcasing where research is still needed.

Conclusion

Machine learning in cardiac surgery is being applied in perioperative anesthetic management and risk assessment. They are generally yielding comparable predictive outcomes to existing clinical scores. With the exception that models implementing dynamic variables obtain promising results. However, there is still a need for data on clinical outcomes after using ML-based models for diagnostic and treating guidance.

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Jianxing He and Hengrui Liang) for the series "Artificial Intelligence in Thoracic Disease: from Bench to Bed" published in *Journal of Thoracic Disease*. The article has undergone external peer review.

Reporting Checklist: The authors have completed the PRISMA-ScR reporting checklist. Available at: https://dx.doi.org/10.21037/jtd-21-765

Peer Review File: Available at https://dx.doi.org/10.21037/jtd-21-765

Conflicts of Interest: The authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/jtd-21-765). The series "Artificial Intelligence

in Thoracic Disease: from Bench to Bed" was commissioned by the editorial office without any funding or sponsorship. APJV reports having received unrestricted research grants from Edwards Lifesciences and Philips. DPV reports having received research grants from Philips as well as consultancy, lecture, and travel expenses fees from Edwards Lifesciences. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Gibbon JH Jr. Application of a mechanical heart and lung apparatus to cardiac surgery. Minn Med 1954;37:171-85; passim.

2. Fletcher N, Geisen M, Meeran H, et al. Initial clinical experience with a miniaturized transesophageal echocardiography probe in a cardiac intensive care unit. J Cardiothorac Vasc Anesth 2015;29:582-7.

3. Li P, Qu LP, Qi D, et al. Significance of perioperative goal-directed hemodynamic approach in preventing postoperative complications in patients after cardiac surgery: a meta-analysis and systematic review. Ann Med 2017;49:343-51.

4. Raita Y, Camargo CA Jr, Macias CG, et al. Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: a multicenter prospective study. Sci Rep 2020;10:10979.

5. Liu H, Shen Y, Sun L, et al. Effects of response gene to complement 32 as a new biomarker in children with acute kidney injury. Zhonghua Er Ke Za Zhi 2014;52:494-9.

6. Baxt WG. Application of artificial neural networks to clinical medicine. Lancet 1995;346:1135-8.

7. Hashimoto DA, Witkowski E, Gao L, et al. Artificial Intelligence in Anesthesiology: Current Techniques,

6990

Rellum et al. Machine learning for perioperative care in CTS patients

Clinical Applications, and Limitations. Anesthesiology 2020;132:379-94.

8. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005;8:19-32.

9. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018;169:467-73.

10. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315-6.

11. Hote M. Cardiac surgery risk scoring systems: In quest for the best. Heart Asia 2018;10:e011017.

12. Allyn J, Allou N, Augustin P, et al. A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis. PLoS One 2017;12:e0169772.

13. Nilsson J, Ohlsson M, Thulin L, et al. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. J Thorac Cardiovasc Surg 2006;132:12-9.

14. Peng SY, Peng SK. Predicting adverse outcomes of cardiac surgery with the application of artificial neural networks. Anaesthesia 2008;63:705-13.

15. Orr RK. Use of a probabilistic neural network to estimate the risk of mortality after cardiac surgery. Med Decis Making 1997;17:178-85.

16. Benedetto U, Sinha S, Lyon M, et al. Can machine learning improve mortality prediction following cardiac surgery? Eur J Cardiothorac Surg 2020;58:1130-6.

17. Fernandes MPB, Armengol de la Hoz M, Rangasamy V, et al. Machine Learning Models with Preoperative Risk Factors and Intraoperative Hypotension Parameters Predict Mortality After Cardiac Surgery. J Cardiothorac Vasc Anesth 2021;35:857-65.

18. Zhong Z, Yuan X, Liu S, et al. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. Sci Rep 2021;11:3384.

19. Celi LA, Galvin S, Davidzon G, et al. A Database-driven Decision Support System: Customized Mortality Prediction. J Pers Med 2012;2:138-48.

20. Kilic A, Goyal A, Miller JK, et al. Predictive Utility of a Machine Learning Algorithm in Estimating Mortality Risk in Cardiac Surgery. Ann Thorac Surg 2020;109:1811-9.

21. Lippmann RP, Shahian DM. Coronary artery bypass risk prediction using neural networks. Ann Thorac Surg 1997;63:1635-43.

22. Mendes RG, de Souza CR, Machado MN, et al. Predicting reintubation, prolonged mechanical ventilation and death

in post-coronary artery bypass graft surgery: a comparison between artificial neural networks and logistic regression models. Arch Med Sci 2015;11:756-63.

23. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput Biomed Res 1993;26:220-9.

24. Lippmann RP, Kukolich L, Shahian D. Predicting the Risk of Complications in Coronary Artery Bypass Operations using Neural Networks. In: Tesaukro GTD, Leen T, editor. Advances in neural information processing systems. A Bradford Book; 1995. p. 1055-62.

25. Mejia OAV, Antunes MJ, Goncharov M, et al. Predictive performance of six mortality risk scores and the development of a novel model in a prospective cohort of patients undergoing valve surgery secondary to rheumatic fever. PLoS One 2018;13:e0199277.

26. Yoon J, Zame WR, Banerjee A, et al. Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation. PLoS One 2018;13:e0194985.

27. Nilsson J, Ohlsson M, Höglund P, et al. The International Heart Transplant Survival Algorithm (IHTSA): a new model to improve organ sharing and survival. PLoS One 2015;10:e0118644.

28. Shah M, Villela MA, Bravo C, et al. Impact of Donor Hemodynamics on Heart Transplant Outcomes: Using Machine Learning Techniques. J Heart Lung Transplant 2020;39:S295.

29. Villela MA, Bravo CA, Shah M, et al. Prediction of Outcomes after Heart Transplantation Using Machine Learning Techniques. J Heart Lung Transplant 2020;39:S295-6.

30. Bravo CA, Villela MA, Shah M, et al. Risk Factors for Post-Transplant Outcomes in Patients with LVAD Support: A Machine Learning and Logistic Regression of the UNOS Database. J Heart Lung Transplant 2020;39:S410.

31. Miller PE, Pawar S, Vaccaro B, et al. Predictive Abilities of Machine Learning Techniques May Be Limited by Dataset Characteristics: Insights From the UNOS Database. J Card Fail 2019;25:479-83.

32. Agasthi P, Buras MR, Smith SD, et al. Machine learning helps predict long-term mortality and graft failure in patients undergoing heart transplant. Gen Thorac Cardiovasc Surg 2020;68:1369-76.

33. Kacila M, K Tiwari K, Granov N, et al. Assessment of the Initial and Modified Parsonnet score in mortality prediction of the patients operated in the Sarajevo Heart

center. Bosn J Basic Med Sci 2010;10:165-8.

34. Nashef SA, Sharples LD, Roques F, et al. EuroSCORE II and the art and science of risk modelling. Eur J Cardiothorac Surg 2013;43:695-6.

35. Kunt AG, Kurtcephe M, Hidiroglu M, et al. Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort. Interact Cardiovasc Thorac Surg 2013;16:625-9.

36. Puskas JD, Kilgo PD, Thourani VH, et al. The society of thoracic surgeons 30-day predicted risk of mortality score also predicts long-term survival. Ann Thorac Surg 2012;93:26-33; discussion 33-5.

37. O'Brien SM, Feng L, He X, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. Ann Thorac Surg 2018;105:1419-28.

38. Ad N, Holmes SD, Patel J, et al. Comparison of EuroSCORE II, Original EuroSCORE, and The Society of Thoracic Surgeons Risk Score in Cardiac Surgery Patients. Ann Thorac Surg 2016;102:573-9.

39. Buzatu DA, Taylor KK, Peret DC, et al. The determination of cardiac surgical risk using artificial neural networks. J Surg Res 2001;95:61-6.

40. Tu JV, Weinstein MC, McNeil BJ, et al. Predicting mortality after coronary artery bypass surgery: what do artificial neural networks learn? The Steering Committee of the Cardiac Care Network of Ontario. Med Decis Making 1998;18:229-35.

41. Weiss ES, Allen JG, Kilic A, et al. Development of a quantitative donor risk index to predict short-term mortality in orthotopic heart transplantation. J Heart Lung Transplant 2012;31:266-73.

42. Weiss ES, Allen JG, Arnaoutakis GJ, et al. Creation of a quantitative recipient risk index for mortality prediction after cardiac transplantation (IMPACT). Ann Thorac Surg 2011;92:914-21; discussion 921-2.

43. Smits JM, De Pauw M, de Vries E, et al. Donor scoring system for heart transplantation and the impact on patient survival. J Heart Lung Transplant 2012;31:387-97.

44. Medved D, Nugues P, Nilsson J. Predicting the outcome for patients in a heart transplantation queue using deep learning. Annu Int Conf IEEE Eng Med Biol Soc 2017;2017:74-7.

45. Schmid F, Goepfert MS, Kuhnt D, et al. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg 2011;112:78-83.

46. Solet J, Barach P. Managing alarm fatigue in cardiac care.

Prog Pediatr Cardiol 2012;33:85-90.

47. Becker K, Thull B, Käsmacher-Leidinger H, et al. Design and validation of an intelligent patient monitoring and alarm system based on a fuzzy logic process model. Artif Intell Med 1997;11:33-53.

48. Hatib F, Jian Z, Buddi S, et al. Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology 2018;129:663-74.

49. Shin B, Maler SA, Reddy K, et al. Use of the Hypotension Prediction Index During Cardiac Surgery. J Cardiothorac Vasc Anesth 2021;35:1769-75.

50. Lang AL, Huang X, Alfirevic A, et al. Patient characteristics and surgical variables associated with intraoperative reduced right ventricular function. J Thorac Cardiovasc Surg 2020. [Epub ahead of print]. doi: 10.1016/j.jtcvs.2020.11.075.

51. Liu S, Bose R, Ahmed A, et al. Artificial Intelligence-Based Assessment of Indices of Right Ventricular Function. J Cardiothorac Vasc Anesth 2020;34:2698-702.

52. Jeganathan J, Knio Z, Amador Y, et al. Artificial intelligence in mitral valve analysis. Ann Card Anaesth 2017;20:129-34.

53. Higgins TL. Quantifying risk and assessing outcome in cardiac surgery. J Cardiothorac Vasc Anesth 1998;12:330-40.

54. Lamarche Y, Elmi-Sarabi M, Ding L, et al. A score to estimate 30-day mortality after intensive care admission after cardiac surgery. J Thorac Cardiovasc Surg 2017;153:1118-1125.e4.

55. Cevenini G, Barbini E, Scolletta S, et al. A comparative analysis of predictive models of morbidity in intensive care unit after cardiac surgery - part II: an illustrative example. BMC Med Inform Decis Mak 2007;7:36.

56. Chong CF, Li YC, Wang TL, et al. Stratification of adverse outcomes by preoperative risk factors in coronary artery bypass graft patients: an artificial neural network prediction model. AMIA Annu Symp Proc 2003;2003:160-4.

57. Mufti HN, Hirsch GM, Abidi SR, et al. Exploiting Machine Learning Algorithms and Methods for the Prediction of Agitated Delirium After Cardiac Surgery: Models Development and Validation Study. JMIR Med Inform 2019;7:e14993.

58. Lei G, Wang G, Zhang C, et al. Using Machine Learning to Predict Acute Kidney Injury After Aortic Arch Surgery. J Cardiothorac Vasc Anesth 2020;34:3321-8.

59. Tseng PY, Chen YT, Wang CH, et al. Prediction of the development of acute kidney injury following cardiac

6992

Rellum et al. Machine learning for perioperative care in CTS patients

surgery by machine learning. Crit Care 2020;24:478.

60. Lee HC, Yoon HK, Nam K, et al. Derivation and Validation of Machine Learning Approaches to Predict Acute Kidney Injury after Cardiac Surgery. J Clin Med 2018;7:322.

61. Penny-Dimri JC, Bergmeir C, Reid CM, et al. Machine Learning Algorithms for Predicting and Risk Profiling of Cardiac Surgery-Associated Acute Kidney Injury. Semin Thorac Cardiovasc Surg 2021;33:735-45.

62. Barbini E, Cevenini G, Scolletta S, et al. A comparative analysis of predictive models of morbidity in intensive care unit after cardiac surgery - part I: model planning. BMC Med Inform Decis Mak 2007;7:35.

63. Vives M, Hernandez A, Parramon F, et al. Acute kidney injury after cardiac surgery: prevalence, impact and management challenges. Int J Nephrol Renovasc Dis 2019;12:153-66.

64. Thakar CV, Arrigain S, Worley S, et al. A clinical score to predict acute renal failure after cardiac surgery. J Am Soc Nephrol 2005;16:162-8.

65. Ng SY, Sanagou M, Wolfe R, et al. Prediction of acute kidney injury within 30 days of cardiac surgery. J Thorac Cardiovasc Surg 2014;147:1875-83, 1883.e1.

66. Bent P, Tan HK, Bellomo R, et al. Early and intensive continuous hemofiltration for severe renal failure after cardiac surgery. Ann Thorac Surg 2001;71:832-7.

67. Koster S, Hensens AG, Schuurmans MJ, et al. Risk factors of delirium after cardiac surgery: a systematic review. Eur J Cardiovasc Nurs 2011;10:197-204.

68. Ma X, Vervoort D. Critical care capacity during the COVID-19 pandemic: Global availability of intensive care beds. J Crit Care 2020;58:96-7.

69. Messaoudi N, De Cocker J, Stockman BA, et al. Is EuroSCORE useful in the prediction of extended intensive care unit stay after cardiac surgery? Eur J Cardiothorac Surg 2009;36:35-9.

70. Meyfroidt G, Güiza F, Cottem D, et al. Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a Gaussian processes model. BMC Med Inform Decis Mak 2011;11:64.

71. Wise ES, Stonko DP, Glaser ZA, et al. Prediction of Prolonged Ventilation after Coronary Artery Bypass Grafting: Data from an Artificial Neural Network. Heart Surg Forum 2017;20:E007-14.

72. Rowan M, Ryan T, Hegarty F, et al. The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors. Artif Intell Med 2007;40:211-21.

73. Barbini P, Barbini E, Furini S, et al. A straightforward approach to designing a scoring system for predicting length-of-stay of cardiac surgery patients. BMC Med Inform Decis Mak 2014;14:89.

74. Manyam R, Zhang Y, Carter S, et al. Unraveling the impact of time-dependent perioperative variables on 30-day readmission after coronary artery bypass surgery. J Thorac Cardiovasc Surg 2020. [Epub ahead of print]. doi: 10.1016/j.jtcvs.2020.09.076.

75. Engoren M, Habib RH, Dooner JJ, et al. Use of genetic programming, logistic regression, and artificial neural nets to predict readmission after coronary artery bypass surgery. J Clin Monit Comput 2013;27:455-64.

76. LaFaro RJ, Pothula S, Kubal KP, et al. Neural Network Prediction of ICU Length of Stay Following Cardiac Surgery Based on Pre-Incision Variables. PLoS One 2015;10:e0145395.

77. Benuzillo J, Caine W, Evans RS, et al. Predicting readmission risk shortly after admission for CABG surgery. J Card Surg 2018;33:163-70.

78. Zywot A, Lau CSM, Glass N, et al. Preoperative Scale to Determine All-Cause Readmission After Coronary Artery Bypass Operations. Ann Thorac Surg 2018;105:1086-93.

79. Chiorino CDRN, Santos VB, Lopes JL, et al. Predictors of Hospital Readmission within 30 Days after Coronary Artery Bypass Grafting: Data Analysis of 2,272 Brazilian Patients. Braz J Cardiovasc Surg 2020;35:884-90.

80. Wijnberge M, Geerts BF, Hol L, et al. Effect of a Machine Learning-Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial. JAMA 2020;323:1052-60.

81. Lynch CM, Abdollahi B, Fuqua JD, et al. Prediction of lung cancer patient survival via supervised machine learning classification techniques. Int J Med Inform 2017;108:1-8.

82. Kendale S, Kulkarni P, Rosenberg AD, et al. Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension. Anesthesiology 2018;129:675-88.

83. Connor CW. Artificial Intelligence and Machine Learning in Anesthesiology. Anesthesiology 2019;131:1346-59.

84. Mehta P, Wang CH, Day AGR, et al. A high-bias, low-variance introduction to Machine Learning for physicists. Phys Rep 2019;810:1-124.

85. Saeb S, Lonini L, Jayaraman A, et al. The need to

approximate the use-case in clinical machine learning. Gigascience 2017;6:1-9.

86. Tolis G Jr, Spencer PJ, Bloom JP, et al. Teaching operative cardiac surgery in the era of increasing patient complexity: Can it still be done? J Thorac Cardiovasc Surg 2018;155:2058-65.

87. Simpao AF, Rehman MA. Anesthesia Information Management Systems. Anesth Analg 2018;127:90-4.

88. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.

89. Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9:e1312.

90. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020;3:118.

91. Smith JA, Abhari RE, Hussain Z, et al. Industry ties and evidence in public comments on the FDA framework for modifications to artificial intelligence/machine learning-based medical devices: a cross sectional study. BMJ Open 2020;10:e039969.

Appendix 1 Search summary

Performed on: February 25, 2021
With the help of: Faridi Jamaludin, Amsterdam UMC, Medical Library AMC

Databases:		
MEDLINE (Ovid), EMBASE (Ovid), Cochrane Library	Before deduplication	After deduplication
Total	1,562	1,142

MEDLINE (Ovid)
Database(s):   Ovid MEDLINE(R) ALL; 1947 to
Search Strategy:

#	Searches	Results
1	exp Thoracic Surgical Procedures/ or Thoracic Surgery/ or exp Cardiovascular Surgical Procedures/ or exp Cardiac Surgical Procedures/ or exp Thoracic Diseases/	551,140
2	((cardi* or heart or vascular or thoracic* or thorax) adj3 (surg* or operat*)).ti,ab,kw,kf.	137,200
3	1 or 2	612,707
4	exp Artificial Intelligence/	107,748
5	(artificial intelligence or machine learning or deep learning or deep neural network or neural network* or machine intelligence or computational intelligence or computer reasoning or predictive analytics or cognitive comput* or computer vision system* or naive bayes or bayesian learning or support vector* or automated or automatic or random forest* or boosting or automatic anesthe* or automatic anaesthe* or automated cardiac anesthe* or automated cardiac anaesthe*).ti,ab,kw,kf.	309,465
6	4 or 5	365,261
7	exp Anesthesia/ or Anesthesiology/	209,295
8	(anesth* or anaesthes* or intraoperat*).ti,ab,kw,kf.	479,483
9	(undergo* adj3 (surg* or operat* or replac*)).ti,ab,kw,kf.	107,044
10	7 or 8 or 9	632,036
11	3 and 6 and 10	680
12	(exp Animals/ or exp Animal Experimentation/ or exp models, animal/ or (rat or rats or mice or mouse or dog or dogs or pig or pigs or cow or cows or monkey or monkeys or goat or goats or horse or horses or ape or apes or gorilla or gorillas).ti,ab,kw.) not (Humans/ or human*.ti,ab,kw.)	4,735,867
13	11 not 12	653
14	comment/ or editorial/ or letter/ or (letter or comment* or editorial).ti.	2,019,154
15	13 not 14	648

EMBASE (Ovid)
Database(s):   EMBASE Classic + EMBASE; 1947 to February 25, 2021
Search Strategy:

#	Searches	Results
1	exp thorax surgery/ or exp cardiovascular surgery/ or exp heart surgery/ or exp thorax disease/	1,154,605
2	((cardi* or heart or vascular or thoracic* or thorax) adj3 (surg* or operat*)).ti,ab,kw.	220,373
3	1 or 2	1,212,463
4	exp artificial intelligence/ or exp machine learning/	253,856
5	(artificial intelligence or machine learning or deep learning or deep neural network or neural network* or machine intelligence or computational intelligence or computer reasoning or predictive analytics or cognitive comput* or computer vision system* or naive bayes or bayesian learning or support vector* or automated or automatic or random forest* or boosting or automatic anesthe* or automatic anaesthe* or automated cardiac anesthe* or automated cardiac anaesthe*).ti,ab,kw.	418,467
6	4 or 5	552,770
7	exp anesthesia/ or cardiac anesthesia/ or anesthesiology/ or receiver operating characteristic/	552,219
8	(anesth* or anaesthes* or intraoperat*).ti,ab,kw.	682,092
9	(undergo* adj3 (surg* or operat* or replac*)).ti,ab,kw.	161,477
10	7 or 8 or 9	1,079,643
11	3 and 6 and 10	1,231
12	(exp animal/ or exp animal experiment/ or exp animal model/ or (rat or rats or mice or mouse or dog or dogs or pig or pigs or cow or cows or monkey or monkeys or goat or goats or horse or horses or ape or apes or gorilla or gorillas).ti,ab,kw.) not (human/ or human*.ti,ab,kw.)	5,723,322
13	11 not 12	1,183
14	letter/ or editorial/ or note/ or (letter or comment* or editorial).ti.	2,684,107
15	13 not 14	1,169
16	limit 15 to conference abstract status	401
17	15 not 16	768

Cochrane Library
Database(s):   Cochrane Central Register of Controlled Trials
Search Strategy:

#	Searches	Results
1	MeSH descriptor: [Thoracic Surgical Procedures] explode all trees	15,672
2	MeSH descriptor: [Thoracic Surgery] explode all trees	164
3	MeSH descriptor: [Cardiovascular Surgical Procedures] explode all trees	19,951
4	MeSH descriptor: [Cardiac Surgical Procedures] explode all trees	12,945
5	MeSH descriptor: [Thoracic Diseases] explode all trees	172
6	((cardi* or heart or vascular or thoracic* or thorax) near/3 (surg* or operat*)):ti,ab,kw	22,803
7	#1 or #2 or #3 or #4 or #5 or #6	39,277
8	MeSH descriptor: [Artificial Intelligence] explode all trees	1,062
9	(artificial intelligence or machine learning or deep learning or deep neural network or neural network* or machine intelligence or computational intelligence or computer reasoning or predictive analytics or cognitive comput* or computer vision system* or naive bayes or bayesian learning or support vector* or automated or automatic or random forest* or boosting or automatic anesthe* or automatic anaesthe* or automated cardiac anesthe* or automated cardiac anaesthe*):ti,ab,kw	23,612
10	#8 or #9	24,277
11	(anesth* or anaesthes* or intraoperat*):ti,ab,kw	98,381
12	(undergo* near/3 (surg* or operat* or replac*)):ti,ab,kw	29,516
13	#11 or #12	114,026
14	#7 and #10 and #13	150