Review Article

Machine learning in cardiac surgery: a narrative review

Travis J. Miles[1,2]^, Ravi K. Ghanta[1,2]

[1]Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX, USA; [2]Applied Statistics and Machine Learning for the Advancement of Surgery, Department of Surgery, Baylor College of Medicine, Houston, TX, USA

Contributions: (I) Conception and design: Both authors; (II) Administrative support: RK Ghanta; (III) Provision of study materials or patients: Both authors; (IV) Collection and assembly of data: TJ Miles; (V) Data analysis and interpretation: TJ Miles; (VI) Manuscript writing: Both authors; (VII) Final approval of manuscript: Both authors.

Correspondence to: Ravi K. Ghanta, MD. Michael E. DeBakey Department of Surgery, Baylor College of Medicine, 7200 Cambridge Street Floor 7, Houston, TX 77030, USA; Applied Statistics and Machine Learning for the Advancement of Surgery, Department of Surgery, Baylor College of Medicine, Houston, TX, USA. Email: ravi.ghanta@bcm.edu.

Background and Objective: Machine learning (ML) is increasingly being utilized to provide data driven solutions to challenges in medicine. Within the field of cardiac surgery, ML methods have been employed as risk stratification tools to predict a variety of operative outcomes. However, the clinical utility of ML in this domain is unclear. The aim of this review is to provide an overview of ML in cardiac surgery, particularly with regards to its utility in predictive analytics and implications for use in clinical decision support.

Methods: We performed a narrative review of relevant articles indexed in PubMed since 2000 using the MeSH terms "Machine Learning", "Supervised Machine Learning", "Deep Learning", or "Artificial Intelligence" and "Cardiovascular Surgery" or "Thoracic Surgery".

Key Content and Findings: ML methods have been widely used to generate pre-operative risk profiles, consistently resulting in the accurate prediction of clinical outcomes in cardiac surgery. However, improvement in predictive performance over traditional risk metrics has proven modest and current applications in the clinical setting remain limited.

Conclusions: Studies utilizing high volume, multidimensional data such as that derived from electronic health record (EHR) data appear to best demonstrate the advantages of ML methods. Models trained on post cardiac surgery intensive care unit data demonstrate excellent predictive performance and may provide greater clinical utility if incorporated as clinical decision support tools. Further development of ML models and their integration into EHR's may result in dynamic clinical decision support strategies capable of informing clinical care and improving outcomes in cardiac surgery.

Keywords: Cardiac surgery; machine learning (ML); artificial intelligence (AI); critical care; data science

Submitted Oct 31, 2023. Accepted for publication Mar 15, 2024. Published online Apr 24, 2024.
doi: 10.21037/jtd-23-1659
View this article at: https://dx.doi.org/10.21037/jtd-23-1659

Introduction

Data science is a broad field that characterizes and addresses complex problems through the extraction of knowledge from data (1). With the advent of the information age, the amount of data collected and stored has increased exponentially, resulting in an abundance of so-called "Big Data" (2). Characterized by the three V's (extreme volume of data, significant variability of data types, and the high velocity in which data accumulates), big data demand advanced methods for analysis, resulting in the emergence of data science as a field of scientific inquiry (3,4).

Machine learning (ML) refers to a variety of statistical

^ ORCID: 0009-0008-2687-2380.

Table 1 Search strategy summary

Items	Specification
Date of search	Sep 01, 2023
Database	PubMed
Search terms used	"Machine Learning", "Supervised Machine Learning", "Deep Learning", or "Artificial Intelligence" and "Cardiovascular Surgery" or "Thoracic Surgery"
Timeframe	Jan 2000 to Aug 2023
Inclusion and exclusion criteria	Inclusion criteria: English language, adult cardiac surgery
	Exclusion criteria: general thoracic, aortic, congenital cardiac, or thoracic transplant surgery
Selection process	T.J.M. conducted the selection independently. Consensus adjudication was not required given the narrative nature of the study

techniques in which computer algorithms efficiently perform a task by "learning", or optimizing model parameters, on data. Tasks can be predictive such as in supervised ML which involves the prediction of a target feature from data labelled with known attributes. Conversely, in unsupervised ML, descriptive algorithms are designed to find patterns or trends in unstructured data where features are unknown. ML is at the core of modern data science and its efficacy in analysis of big data has contributed to the ongoing artificial intelligence (AI) revolution (5).

The advent of the electronic health record (EHR) and subsequent digital transformation of healthcare have prompted an increased interest in leveraging data science and ML to enhance quality of medical care (6). Applications of ML in healthcare are extensive and include clinical diagnostics, medical imaging, biomedical research, and clinical trial design (7-10). The potential of ML in the field of cardiac surgery is increasingly being recognized (11,12), with promising results in the analysis of chest radiographs (13), detection of arrhythmias from electrocardiograms (14), and assessment of pre-operative risk (15).

Predictive analytics, the modeling of risk profiles from patient data, has long been a focus in cardiac surgery and has proven an exciting area of ML research. While ML methods have demonstrated promising results, the successful deployment of these models in clinical practice has thus far been limited (16). Furthermore, clinicians, whose domain expertise is vital to the successful implementation of AI in healthcare, may be unfamiliar with the technical aspects of ML methodology (17,18). The aim of this article is to provide an overview of the current state of ML in cardiac surgery by reviewing model development and interpretation, summarizing performance in predictive analytics, and describing challenges facing clinical implementation. We present this article in accordance with the Narrative Review reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1659/rc).

Methods

A literature review of peer-reviewed research articles involving ML in adult cardiac surgery between January 2000 to August 2023 was performed. A PubMed literature search using the MeSH terms "Machine Learning", "Supervised Machine Learning", "Deep Learning", or "Artificial Intelligence" and "Cardiovascular Surgery" or "Thoracic Surgery" was performed. Given the narrative nature of this paper, studies were carefully reviewed by abstract and title to provide a general understanding of the topic. This review exclusively considered studies published in English. Further exclusion criteria included studies concentrating on congenital heart surgery, general thoracic surgery, minimally invasive cardiac surgery, or cardiac transplant (*Table 1*). The concordance index (C-index) was the primary performance metric reported in this review.

ML methodology

A brief overview of ML methodology is warranted as algorithm selection and development have implications in model performance and interpretation (19). Most existing ML models in cardiac surgery involve supervised learning, in which models are trained to recognize patterns from datasets containing input and output (target) features, ultimately providing a predicted target parameter. Target outputs can be continuous variables (regression tasks) or categorical

Table 2 Description of common machine learning methods

Algorithm	Description	Strengths	Limitations
Generalized linear models	Parametric statistical methods effective in both regression or classification tasks. Linear and logistic regression traditionally have represented the most common method for predictive analytics in cardiac surgery	Algorithm structure provides quantitative estimates characterizing associations between features and outputs making model results easily interpretable (e.g., odds ratio)	Learning bias may prevent modeling of non-linear relationships in the data. Collinearity between independent variables can adversely impact model performance
Classification and regression decision trees	Provides prediction of output by grouping values of features into non-overlapping regions through a process known as recursive partitioning. The data is split based on a sequence of attribute tests producing a series of branches and nodes that result in classification rule for prediction tasks	Highly flexible algorithms capable of accounting for non-linear associations Individual trees are visually interpretable	Individual decision trees are prone to overfitting which may limit generalizability to unseen data. Often demonstrate worse performance compared to ensemble methods
Random forest	Type of ensemble decision tree model composed of a randomized collection, or "forest", of individual decision trees	Combining decision trees and aggregating an average of predictions enhances accuracy and reduces overfitting	While capable of reporting feature importance, ensemble models sacrifice interpretability when compared to individual trees or generalized linear models
Gradient boosting machine	Family of additive models in which decision trees are sequentially incorporated into an ensemble, with iterative optimization of each additional tree through a loss function. Gradient boosting enhances performance by identifying and learning from weak performing trees as opposed to aggregating trees randomly	Flexible, precise, and efficient making it suitable for a variety of prediction tasks with large datasets	Difficult to assess associations between variables influencing prediction and is often considered a "black box" algorithm
Artificial neural network	Model inspired by the biologic nervous system composed of interconnected layers of functions, or neurons, that receive inputs from other neurons and compute outputs that are propagated through the network and processed through an activation function to produce a prediction. Neural networks learn by adjusting weights and thresholds between connections to minimize a loss function thus optimizing predictive performance	Capable of modeling complex patterns from high dimensional data leading to excellent results in natural language processing, image processing, and voice recognition	Complex system that requires sophisticated computational resources for analysis of larger datasets. Artificial neural network based models are prone to overfitting particularly with smaller datasets. Similar to random forest and gradient boosting machines, artificial neural networks are considered "black box" algorithms

(classification tasks) (20). There are many different techniques for modeling patterns in data, each of which has its own characteristic strengths and weaknesses (21). *Table 2* contains information on commonly used ML models including regression, decision trees, and neural networks.

Each ML method relies on assumptions made regarding the underlying nature of the data. For example, the logistic regression method traditionally used to model risk assumes linear relationships within the data and therefore fails to capture non-linear trends. These learning biases affect performance and render different models particularly well or poorly suited for analysis based on the pattern of the data (22). As a result, the "no free lunch" theorem states that no one individual model demonstrates better performance than others across all situations (23). Therefore, standard practice is to train multiple models on the data and identify one that is best suited to the data and task of interest.

Evaluation of model performance is important for model selection, comparison, and assessment of generalizability. Typically, this is accomplished by dividing data into a training set and test set (*Figure 1*). After fitting the model on training data, model performance is assessed on the yet unseen test dataset to provide an overall estimate of model performance. Often the training set is further

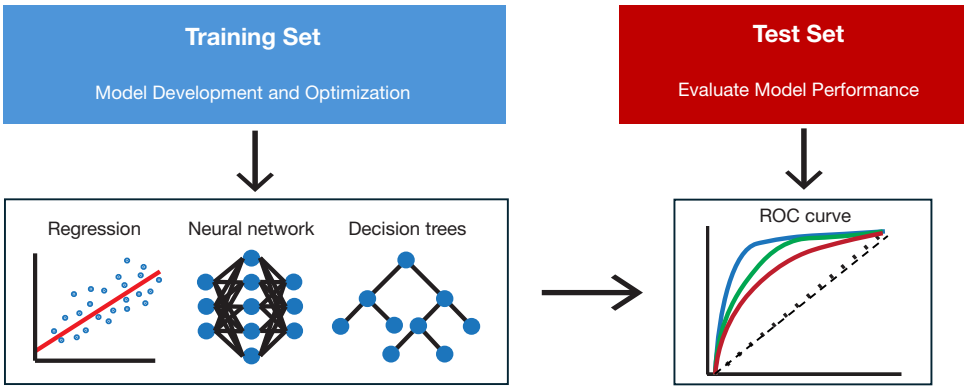

Figure 1 Representation of typical machine learning workflow where data is split into a separate training set for model development and a test set for model evaluation and comparison. ROC, receiver operating characteristic.

divided into a training set to fit models and validation set to optimize, or tune, parameters. A crucial component of model development is ensuring models are not tested on any data used for training, as this so called "data leak" will result in poor generalizability and an overestimate of performance (24). The importance of proper data handling and model development, particularly when more advanced methods like K-fold cross validation and bootstrapping are implemented, has led experts in the field to call for a standardized guideline for reporting predictive models (25,26).

Methods to assess and compare model performance are similar to those used in traditional methods like logistic regression (27). Namely, for classification models the discriminatory power of ML models are typically evaluated with the area under the receiver operating characteristic curve (AUC-ROC) or C-index. Additional ML metrics include precision and recall which represent the positive predictive value and sensitivity of the model respectively. The F-1 score is the harmonic mean of precision and recall and thus provides a representation of both measures in a single metric.

Predictive analytics

Mortality

Risk assessment and prognostication are important components of cardiac surgery. Several risk models such as the EuroScore II and Society of Thoracic Surgery Predicted Risk of Mortality (STS PROM) are currently utilized to inform surgical decision making, assess quality, and measure performance (28-30). Derived from national registry data, these models are based on traditional linear methods. While these models demonstrate fair predictive performance, they are restrained by a significant learning bias. As a result, more advanced ML methods have garnered considerable interest as alternative techniques for modelling risk in cardiac surgery.

Initial studies comparing societal risk models with those based on ML have shown modest but absolute improvement in predictive performance. Allyn *et al.* demonstrated decision tree based algorithms could more accurately predict operative mortality when compared to EuroScore I and II (C-index 0.795 *vs.* 0.737, P<0.0001) for patients undergoing elective cardiac surgery (31). However, these findings were not replicated in a similar study comparing ML to EuroScore II which failed to demonstrate a significant improvement in predictive performance (32). Comparison to STS models have been favorable, with extreme gradient boosting models demonstrating a modest benefit in predicting operative mortality when compared to STS PROM (33). A meta-analysis composed of 15 studies comparing ML to generalized linear models confirmed superior performance (C-index 0.88 *vs.* 0.81, P=0.03) with ML (34). Of note, significant benefit was only demonstrated when comparing the best performing model from each study to logistic regression. No single model showed superiority across all studies, illustrating the "no free lunch" theorem.

While these results are encouraging, some question whether the marginal gain in predictive performance seen with ML is mitigated by the reduced interpretability of many of these models. Data quality has significant implications in model performance and limitations in the

input data for these models likely diminish their accuracy. While ML models perform best when exposed to granular, multidimensional data most comparison studies to date utilize single center registry data.

Developing models on datasets more reflective of "big data", such as that extracted from EHR platforms or medical imaging, have demonstrated the relative strength of ML more effectively. Gradient boosted models trained on routinely collected EHR data have resulted in significantly improved mortality prediction (C-index =0.978) (35). Neural network models predicting operative mortality solely from a single pre-operative chest radiograph have demonstrated comparable performance to STS PROM, highlighting the power of ML methods when developed on appropriate datasets (36).

Outcomes

Outcomes beyond mortality that influence operative decision making are now being modeled using ML. In a nationwide analysis of patients undergoing valve replacement, Kilic *et al.* showed ML to be superior to traditional models in predicting renal failure, prolonged ventilation, and re-operation (37). Our group found gradient boosted methods highly effective in predicting major morbidity in a cohort containing all cardiac surgery patients within our institution, including cases without existing risk metrics (38). Additionally, our models were able to accurately predict hospitalization cost, a metric that has traditionally been challenging to predict from pre-operative data (39). Given outcomes of cardiac surgery are influenced by parameters aside from pre-operative data, our group assessed the accuracy of ML methods by phase of care. We found that incorporation of intra- and post-operative data improved the efficacy of ML models in predicting mortality and major morbidity after aortocoronary bypass (40). This variability in model performance over the course of the hospitalization has been confirmed in other studies (41) and likely reflects the evolving nature of patient risks profiles.

Additionally, studies have successfully utilized ML methods to predict outcomes not currently incorporated in societal models such as atrial fibrillation, readmission, and acute kidney injury (AKI) (40,42-44). Lee *at al.* demonstrated decision tree based models to be superior to generalized linear models in predicting AKI (45). Building on this work, Tseng *et al.* trained models on both pre-operative and intra-operative data to accurately predict AKI (46). Importantly, they demonstrated that incorporation of granular intraoperative hemodynamic data increased model performance. This trend of improved model performance when incorporating granular perioperative data has stimulated interest in the potential utility of EHR data for the development of dynamic risk models.

Dynamic prediction models

Despite an abundance of evidence suggesting superior predictive performance of ML models over generalized linear methods, their implementation in the clinical setting has thus far been limited. For advanced analytics to be incorporated into clinical workflows, models must not only be accurate but also provide actionable insights to the clinician in a timely fashion. Static pre-operative risk profiles are heavily influenced by non-modifiable parameters and have limited utility in the post-operative setting. However, the post-operative cardiac intensive care setting represents a data rich environment ripe for the application of data science. A growing area of research involves leveraging high-volume, multidimensional intensive care unit (ICU) data to produce dynamic risk profiles (*Figure 2*) (47,48). Such models, capable of mapping complex patterns in clinical data, could provide clinicians with risk profiles in real time thereby facilitating action to prevent or mitigate adverse outcomes. Development of real-time, dynamic models resulting in personalized, data-driven risk profiles represents an innovative opportunity to enhance clinical decision making.

Our group has shown that ML models trained on routinely collected ICU EHR data accurately predict outcomes after cardiac surgery (49). We developed an extreme gradient boosted model capable of predicting AKI at various time points after cardiac surgery in data from the publicly available Medical Information Mart for Intensive Care IV (MIMIC IV) database. Training models on data containing both pre-operative demographic information as well as post-operative parameters such as vital signs, medication doses, and intake/output measures, resulted in the accurate prediction of AKI over the subsequent 48 hours (C-index =0.95). Importantly, the model predicted development of AKI before clinical evidence of renal injury in 89.7% of patients and did so a median of 13 hours prior to clinical detection. Given AKI occurs in approximately 20% of cardiac surgery patients and carries a five-fold increased risk of operative mortality, the dynamic and early prediction of AKI represents an exciting opportunity to leverage data science to improve outcomes

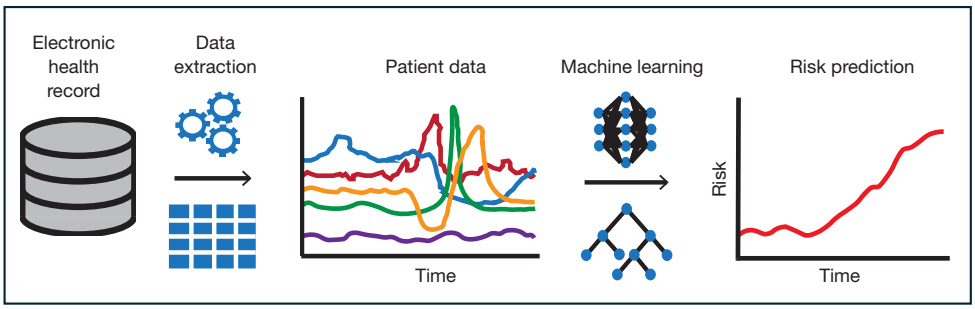

Figure 2 Overview of process wherein dynamic risk models are formulated by applying machine learning methods to analyze health data extracted from the electronic health record.

in cardiac surgery (50,51).

Models based on artificial neural networks have likewise been encouraging, predicting mortality (C-index =0.95), renal failure (C-index =0.96), and reoperation for bleeding (C-index =0.87) from post-operative EHR data (52). A follow-up study demonstrated neural network-based prediction of AKI to be superior to clinician judgement in predicting subsequent renal injury, highlighting the clinical utility of these models (53). The accurate and early prediction of complications after cardiac surgery can inform clinical decision making by alerting providers of patients at risk for complication and guiding appropriate clinical management.

Future directions

ML models have consistently demonstrated the ability to generate accurate risk profiles for cardiac surgery patients, often better than currently used conventional methods. However, the translation of this technology into clinical practice poses a significant challenge for both the medical and data science communities. To date, models with the most promise for decision support are based on EHR data. EHR databases have been designed primarily for billing and liability purposes. Data analytics, while an exciting opportunity, represents a secondary function in these datasets. From a data science perspective, issues with compartmentalization, complexity, and corruption complicate model development and EHR integration (47). Moreover, existing ML risk models rely primarily on structured EHR data (labs, medications, vitals). Advances in natural language processing through large language models present a promising opportunity to enrich model inputs through incorporation of unstructured data (clinical notes, radiology reports) (54). Prospective clinical implementation will require a multidisciplinary collaboration between data scientists, healthcare informaticists, and clinicians to overcome these challenges.

While results of ML methods *in silico* have been encouraging, little is known about how they will fair when transitioning from internal, to external, to prospective validation. Studies outside of cardiac surgery have shown diminished model performance with this transition (55), suggesting an element of overfitting inherent to these models. While generalizability is a concern, a paradigm where individual cardiac centers develop models from their own institutional EHR data to produce models uniquely suited to their patients and practice may optimize performance when translated into real world clinical settings.

An additional criticism of ML has centered on the lack of transparency regarding the relationships within data that affect outcome prediction (56). While predictive performance may be excellent, the "black box" nature of these algorithms has traditionally resulted in physician skepticism (57). In their meta-analysis evaluating factors associated with end-user trust in AI, Kaplan *et al.* found that operator understanding of the AI system was highly associated with trust (58). This would suggest that clinicians would be more likely to incorporate risk predictions into their clinical decision making if the risk factors influencing the model are interpretable to the end-user. Interpretability science is a major focus of ongoing research in the data science community and advances in this domain will likely facilitate clinical adoption (59).

As with any intervention or diagnostic test, rigorous evaluation and prospective validation should occur prior to adoption of these models into clinical practice (60). Recognizing this, the bioinformatics community has developed a comprehensive framework for implementing AI

based decision support systems that parallels clinical trials for drugs and devices (61). By subjecting AI based solutions to rigorous, multiphase assessment, safety and efficacy can be confirmed prior to clinical implementation.

Lastly, the ethical implications of ML-enabled clinical tools must be considered (62). While ML applications hold promise in enhancing patient care, concerns regarding transparency, fairness, accountability, and bias exist (63-65). A balance between innovation and ethical principles is essential to ensure ML-enabled clinical tools maintain the highest standards of medical ethics (66). As ML continues to advance, close collaboration between clinicians, data scientists, and regulatory agencies is necessary to address ethical considerations and ensure responsible implementation of ML in healthcare.

Conclusions

ML represents a promising technology with the potential to improve outcomes and quality of care in cardiac surgery. Despite robust performance across many prediction tasks, several challenges must be addressed prior to adoption in clinical practice. Surgeons should become familiar with the development, interpretation, and applications of ML as their domain expertise is vital to the effective implementation of AI-based solutions in healthcare.

Acknowledgments

Funding: This work was supported by the National Institutes of Health (NIH), National Heart Lung and Blood Institute (No. T32HL139430 to T.J.M.).

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1659/rc

Peer Review File: Available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1659/prf

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1659/coif). T.J.M. is a post-doctoral research fellow participating in the Baylor College of Medicine T32 Research training program in cardiovascular surgery funded through the National Institutes of Health National Heart Lung and Blood Institute (No. T32HL139430) (received as salary support). The other author has no 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. Payne PRO, Bernstam EV, Starren JB. Biomedical informatics meets data science: current state and future directions for interaction. JAMIA Open 2018;1:136-41.
2. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. N Engl J Med 2023;388:1201-8.
3. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform 2014;9:8-13.
4. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015;349:255-60.
5. Xu Y, Liu X, Cao X, et al. Artificial intelligence: A powerful paradigm for scientific research. Innovation (Camb) 2021;2:100179.
6. Orfanoudaki A, Dearani JA, Shahian DM, et al. Improving Quality in Cardiothoracic Surgery: Exploiting the Untapped Potential of Machine Learning. Ann Thorac Surg 2022;114:1995-2000.
7. Javaid M, Haleem A, Pratap Singh R, et al. Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks 2022;3:58-73.
8. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. NPJ Digit Med 2018;1:40.
9. Habehh H, Gohel S. Machine Learning in Healthcare. Curr Genomics 2021;22:291-300.
10. Shailaja K, Seetharamulu B, Jabbar MA. Machine Learning in Healthcare: A Review. 2018 Second

International Conference on Electronics, Communication and Aerospace Technology (ICECA). Coimbatore: IEEE; 2018:910-4.
11. Baxter RD, Fann JI, DiMaio JM, et al. Digital Health Primer for Cardiothoracic Surgeons. Ann Thorac Surg 2020;110:364-72.
12. Salna M. The Promise of Artificial Intelligence in Cardiothoracic Surgery. J Chest Surg 2022;55:429-34.
13. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15:e1002686.
14. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65-9.
15. Kilic A. Artificial Intelligence and Machine Learning in Cardiovascular Health Care. Ann Thorac Surg 2020;109:1323-9.
16. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health 2021;3:e195-203.
17. Petersson L, Larsson I, Nygren JM, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022;22:850.
18. Chen M, Zhang B, Cai Z, et al. Acceptance of clinical artificial intelligence among physicians and medical students: A systematic review with cross-sectional survey. Front Med (Lausanne) 2022;9:990604.
19. Raschka S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. arXiv 2020. Available online: http://arxiv.org/abs/1811.12808
20. Bzdok D, Krzywinski M, Altman N. Points of Significance: Machine learning: a primer. Nat Methods 2017;14:1119-20.
21. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nat Methods 2016;13:703-4.
22. van Giffen B, Herhausen D, Fahse T. Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. Journal of Business Research 2022;144:93-106.
23. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1997;1:67-82.
24. Kapoor S, Narayanan A. Leakage and the Reproducibility Crisis in ML-based Science. arXiv 2022. Available online: http://arxiv.org/abs/2207.07048
25. Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Machine Learning Models in Health Care. JAMA 2020;323:305-6.
26. Stevens LM, Mortazavi BJ, Deo RC, et al. Recommendations for Reporting Machine Learning Analyses in Clinical Research. Circ Cardiovasc Qual Outcomes 2020;13:e006556.
27. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep 2022;12:5979.
28. Nashef SA, Roques F, Sharples LD, et al. EuroSCORE II. Eur J Cardiothorac Surg 2012;41:734-44; discussion 744-5.
29. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development. Ann Thorac Surg 2018;105:1411-8.
30. 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.
31. 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.
32. Molina RS, Molina-Rodríguez MA, Rincón FM, et al. Cardiac Operative Risk in Latin America: A Comparison of Machine Learning Models vs EuroSCORE-II. Ann Thorac Surg 2022;113:92-9.
33. 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.
34. Benedetto U, Dimagli A, Sinha S, et al. Machine learning improves mortality risk prediction after cardiac surgery: Systematic review and meta-analysis. J Thorac Cardiovasc Surg 2022;163:2075-2087.e9.
35. Weiss AJ, Yadaw AS, Meretzky DL, et al. Machine learning using institution-specific multi-modal electronic health records improves mortality risk prediction for cardiac surgery patients. JTCVS Open 2023;14:214-51.
36. Raghu VK, Moonsamy P, Sundt TM, et al. Deep Learning to Predict Mortality After Cardiothoracic Surgery Using Preoperative Chest Radiographs. Ann Thorac Surg 2023;115:257-64.
37. Kilic A, Goyal A, Miller JK, et al. Performance of a Machine Learning Algorithm in Predicting Outcomes of Aortic Valve Replacement. Ann Thorac Surg 2021;111:503-10.

38. Zea-Vera R, Ryan CT, Navarro SM, et al. Development of a Machine Learning Model to Predict Outcomes and Cost After Cardiac Surgery. Ann Thorac Surg 2023;115:1533-42.
39. Osnabrugge RL, Speir AM, Head SJ, et al. Prediction of costs and length of stay in coronary artery bypass grafting. Ann Thorac Surg 2014;98:1286-93.
40. Zea-Vera R, Ryan CT, Havelka J, et al. Machine Learning to Predict Outcomes and Cost by Phase of Care After Coronary Artery Bypass Grafting. Ann Thorac Surg 2022;114:711-9.
41. Castela Forte J, Yeshmagambetova G, van der Grinten ML, et al. Comparison of Machine Learning Models Including Preoperative, Intraoperative, and Postoperative Data and Mortality After Cardiac Surgery. JAMA Netw Open 2022;5:e2237970.
42. Sherman E, Alejo D, Wood-Doughty Z, et al. Leveraging Machine Learning to Predict 30-Day Hospital Readmission After Cardiac Surgery. Ann Thorac Surg 2022;114:2173-9.
43. Karri R, Kawai A, Thong YJ, et al. Machine Learning Outperforms Existing Clinical Scoring Tools in the Prediction of Postoperative Atrial Fibrillation During Intensive Care Unit Admission After Cardiac Surgery. Heart Lung Circ 2021;30:1929-37.
44. Li Q, Lv H, Chen Y, et al. Development and Validation of a Machine Learning Predictive Model for Cardiac Surgery-Associated Acute Kidney Injury. J Clin Med 2023;12:1166.
45. 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.
46. Tseng PY, Chen YT, Wang CH, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. Crit Care 2020;24:478.
47. Johnson AE, Ghassemi MM, Nemati S, et al. Machine Learning and Decision Support in Critical Care. Proc IEEE Inst Electr Electron Eng 2016;104:444-66.
48. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. Chest 2018;154:1239-48.
49. Ryan CT, Zeng Z, Chatterjee S, et al. Machine learning for dynamic and early prediction of acute kidney injury after cardiac surgery. J Thorac Cardiovasc Surg 2023;166:e551-64.
50. Chen JJ, Chang CH, Wu VC, et al. Long-Term Outcomes of Acute Kidney Injury After Different Types of Cardiac Surgeries: A Population-Based Study. J Am Heart Assoc 2021;10:e019718.
51. Lopez-Delgado JC, Esteve F, Torrado H, et al. Influence of acute kidney injury on short- and long-term outcomes in patients undergoing cardiac surgery: risk factors and prognostic value of a modified RIFLE classification. Crit Care 2013;17:R293.
52. Meyer A, Zverinski D, Pfahringer B, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. Lancet Respir Med 2018;6:905-14.
53. Rank N, Pfahringer B, Kempfert J, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. NPJ Digit Med 2020;3:139.
54. Boonstra MJ, Weissenbacher D, Moore JH, et al. Artificial intelligence: revolutionizing cardiology with large language models. Eur Heart J 2024;45:332-45.
55. Lupei MI, Li D, Ingraham NE, et al. A 12-hospital prospective evaluation of a clinical decision support prognostic algorithm based on logistic regression as a form of machine learning to facilitate decision making for patients with suspected COVID-19. PLoS One 2022;17:e0262193.
56. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell 2019;1:206-15.
57. Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. JAMA 2018;319:19-20.
58. Kaplan AD, Kessler TT, Brill JC, et al. Trust in Artificial Intelligence: Meta-Analytic Findings. Hum Factors 2023;65:337-59.
59. Murdoch WJ, Singh C, Kumbier K, et al. Definitions, methods, and applications in interpretable machine learning. Proc Natl Acad Sci U S A 2019;116:22071-80.
60. Chomutare T, Tejedor M, Svenning TO, et al. Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. Int J Environ Res Public Health 2022;19:16359.
61. Park Y, Jackson GP, Foreman MA, et al. Evaluating artificial intelligence in medicine: phases of clinical research. JAMIA Open 2020;3:326-31.
62. Christodoulou KC, Tsoucalas G. Artificial Intelligence-Oriented Heart Surgery: A Complex Bioethical Concept. Cureus 2023;15:e41911.
63. Li F, Ruijs N, Lu Y. Ethics & AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare. AI 2023;4:28-53.

64. Rogers WA, Draper H, Carter SM. Evaluation of artificial intelligence clinical applications: Detailed case analyses show value of healthcare ethics approach in identifying patient care issues. Bioethics 2021;35:623-33.
65. Jeyaraman M, Balaji S, Jeyaraman N, et al. Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare. Cureus 2023;15:e43262.
66. Karimian G, Petelos E, Evers SMAA. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI Ethics 2022;2:539-51.

Cite this article as: Miles TJ, Ghanta RK. Machine learning in cardiac surgery: a narrative review. J Thorac Dis 2024;16(4):2644-2653. doi: 10.21037/jtd-23-1659