Machine-learning approaches might demonstrate greater incremental value in areas such as congenital cardiac surgery, where there are many more complex combinations of diagnoses, procedures, and procedure-specific predictors. It may also be more fruitful to study machine-learning performance for more common, non-fatal complications rather than for mortality, which fortunately is a rare outcome.

References

1. Lippmann RP, Shahian DM. Coronary artery bypass risk prediction using neural networks. *Ann Thorac Surg.* 1997;63:1635-43.
2. Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: systematic review and meta-analysis. *J Thorac Cardiovasc Surg.* 2022;163:2075-87.e9.
3. Jamaati H, Najafi A, Kahe F, Karimi Z, Ahmadi Z, Bolursaz M, et al. Assessment of the EuroSCORE risk scoring system for patients undergoing coronary artery bypass graft surgery in a group of Iranian patients. *Indian J Crit Care Med.* 2015;19:576-9.
4. Mejia OAV, Antunes MJ, Goncharov M, Dallan LRP, Veronese E, Lapenna GA, 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.
5. Nilsson J, Ohlsson M, Thulin L, Höglund P, Nashef SA, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. *J Thorac Cardiovasc Surg.* 2006;132:12-9.
6. Tu JV, Weinstein MC, McNeil BJ, Naylor CD. 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.
7. Rahman HAA, Wah YB, Khairudin Z, Abdullah NN. Comparison of predictive models to predict survival of cardiac surgery patients. In: 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE). 2012:1-5.
8. Mendes RG, de Souza CR, Machado MN, Correa PR, Di Thommazo-Luporini L, Arena R, 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.
9. Allyn J, Allou N, Augustin P, Philip I, Martinet O, Belghiti M, 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.
10. Nouei MT, Kamyad AV, Sarzaem M, Ghazalbash S. Developing a genetic fuzzy system for risk assessment of mortality after cardiac surgery. *J Med Syst.* 2014;38:102.

See Article page 2075.

Commentary: Artificial intelligence to predict mortality: The rise of the machines?

Derrick Y. Tam, MD, PhD,^a Dion Chung, BS, MS,^b and Stephen E. Fremes, MD, MSc^a

In medicine, the words “artificial intelligence” (AI) provoke thoughts of machines that can identify tumors from pathology slides in seconds, parsing through tens of thousands of medical records to find hidden patterns, and detect abnormalities in radiologic imaging that elude trained human eyes.¹ AI—machines leveraging data to perform tasks with human intelligence while not explicitly

CENTRAL MESSAGE

Machine learning may offer improved prediction modeling when used appropriately.

From the ^aDivision of Cardiac Surgery, Schulich Heart Centre, Department of Surgery, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; and ^bAdvanced Data Analytics and Engineering, Naval Information Warfare Center Pacific, San Diego, Calif.

Disclosures: The authors reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Received for publication Aug 17, 2020; revisions received Aug 17, 2020; accepted for publication Aug 18, 2020; available ahead of print Aug 25, 2020.

Address for reprints: Stephen E. Fremes, MSc, MD, FRCSC, FACP, FACC, Schulich Heart Centre, Sunnybrook Health Sciences Centre, 2075 Bayview Ave, Room H4 05, Toronto, ON M4N 3M5, Canada (E-mail: stephen.fremes@sunnybrook.ca).

J Thorac Cardiovasc Surg 2022;163:2092-4

0022-5223/\$36.00

Copyright © 2020 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.08.055>

programmed to do so—has made great strides in the biomedical field in the last decade, particularly in the subdomains of machine learning (ML) and deep learning.² Whether ML can improve upon current models used to predict important outcomes after high-risk surgeries remains uncertain. In this issue of the *Journal*, Benedetto and colleagues’ meta-analysis³ shows machines are certainly getting better, and state-of-the-art supervised ML models are predicting mortality in patients undergoing cardiac surgery more accurately than before.

Benedetto and colleagues performed a systematic review and meta-analysis of the literature and included only studies that compared both ML (which included tree-based regressions, artificial neural networks, support vector machines, and naive Bayes) and conventional logistic

regression (LR) models.³ They compared the pooled performance of more novel ML methods with conventional models used to predict early mortality after surgery. The concordance statistics (c-statistics), calculated from the area under the receiver-operator curve (AUC), which measures the discriminative performance of a prediction model (where 1 = perfect performance whereas 0.5 is the equivalent of flipping a fair coin), was compared between ML and LR. Overall, in a meta-analysis of 15 studies, the pooled c-statistics of ML models (0.88; 95% credible interval, 0.83-0.93) was greater than that of LR models (0.81; 95% credible interval, 0.77-0.85). They are to be congratulated on conducting such an important and relevant study, given the boon in the use of ML methods in the biomedical literature and the need for accurate mortality prediction models in cardiac surgery. The strength of this study was the rigor of the meta-analytic methods used to address the important question at hand and the set of relevant sensitivity analyses to tease out heterogeneity in the results. There were a few limitations in this study, most of which related to the quality and the risk of bias of the included studies, which are discussed in detail to follow.

In their conclusion, Benedetto and colleagues suggest that ML methods may be preferred over conventional LR models to predict important outcomes like mortality. This may be true; a purported benefit of ML methods is that the human user/programmer does not need to specify model interaction terms between covariates (or features), in contrary to conventional models, where interaction terms have to be specified and incorporated, some of which may be quite complex. That said, there may be advantages to the simplicity of LR models; the odds ratios are easy for clinicians to interpret and can be converted to a risk score for use at the bedside.⁴ In addition, there may be other reasons that the c-statistics were, on average, greater in ML models compared with LR models, and this requires a critical understanding of the studies that were included in this meta-analysis. We note that the sample size of the included studies ranged from a mere 165 patients to more than 80,000 patients. Given that mortality event rates in cardiac surgery are often low, small sample sizes typically preclude the addition of several predictors, a rule of thumb is typically 1 predictor variable for every 10 events, as the inclusion of too many predictor variables in the setting of low event numbers may result in model overfitting, leading to inflated c-statistics. Benedetto and colleagues showed that larger sample sizes were associated with greater c-statistics in ML models; this is consistent with what is known about ML models—they are “data-hungry” models that require up to 10 times the number of events compared with conventional prediction models to outperform their conventional counterparts.⁵ Finally, ML

approaches were superior to LR only when the best ML model was chosen from a series of models when individual studies reported multiple models, which may bias results against LR.

When high c-statistics are seen for models with small sample sizes and low event numbers, it raises concerns for both overfitting and the lack of “external” validation or model testing on a different set of patients. Indeed, typically one way to test the true performance of a model is to derive the model in a training set of patients and see how it performs in a different set of patients in what is known as a test set or a validation cohort in ML and LR respectively.^{1,6} However, there are many ways to test a model outside the traditional derivation versus validation splits, but due to differences in terminology used between ML and LR, the terms for validation may be used indiscriminately. Although Benedetto and colleagues captured the type of validation performed in individual studies, we note that most used some form of internal validation like k-fold cross-validation rather than external validation. A sensitivity analysis in this meta-analysis suggested that studies that used internal validation were able to achieve greater c-statistics. In another meta-analysis of 71 studies that included 282 comparisons between modern ML techniques and LR models for clinical prediction, 68% of studies were at risk for potential bias due to inadequate validation procedures.⁷ When the difference in the AUC were compared between the LR and ML models, those at low risk of bias showed no difference (difference in logit of AUC, 0.00, 95% credible interval, -0.18 to 0.18), whereas studies at high risk of bias showed a significant difference in performance that favored ML (difference in logit of AUC, 0.34; 95% credible interval, 0.20-0.47). This study highlights the importance of both proper validation procedures, full reporting of methods and results, and risk of bias assessment in prediction modeling.

Overall, this study highlights several important caveats in the era of using AI in biomedical research. We must critically evaluate novel ML models by asking whether the research question warrants the use of ML methods and offers particular advantages over conventional models⁶ and whether all relevant measures of model performance are appropriately reported. ML methods have been often criticized as complex “black boxes,” as both the input layers may be difficult to explain and the findings are harder to interpret due to the nonlinear nature of the models compared with generalized linear regression models. To this end, much research and work is being conducted in the area of “explainable AI” that shows promising results in both improving model understandability and result interpretation. An area of active study includes the use of deep learning algorithms to analyze the novel continuous physiological data from wearable technology which, when combined with patient

characteristics data, may be used to personalize treatment decisions and subsequent management in advanced cardiovascular disease.⁸ As we await developments in explainable AI, the use of ML will continue to rise exponentially in the biomedical literature; therefore, it is prudent that clinicians have a cursory understanding of the advantages and disadvantages of AI methods to understand the benefits and limitations of prediction modeling using these novel techniques.

References

- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018;284:603-19.
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347-58.
- Benedetto U, Dimagli A, Sinha S, Cocomello L, Gibbison B, Caputo M, et al. Machine learning improves mortality risk prediction after cardiac surgery: systematic review and meta-analysis. *J Thorac Cardiovasc Surg*. 2022;163:2075-87.e9.
- Tam DY, Fang J, Tran A, Tu JV, Ko DT, Deb S, et al. A clinical risk scoring tool to predict readmission after cardiac surgery: an Ontario administrative and clinical population database study. *Can J Cardiol*. 2018;34:1655-64.
- van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol*. 2014;14:137.
- Chen P-HC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater*. 2019;18:410-4.
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12-22.
- Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J*. 2019;40:2058-73.

See Article page 2075.

Commentary: Machine learning in cardiothoracic surgery: From evidence-based to intelligence-based practice

Arman Arghami, MD, and Joseph A. Dearani, MD

One of the marvels of modern medical practice is the shift from clinical decision making based-on “intuition, unsystematic clinical experience, and pathophysiologic rationale”¹ to scientific, clinically relevant evidence-based decision making. For the past decades, common statistical methods (regressions) have allowed researchers to extract meaningful outcome predictions based on set of known variables. In more recent years, the exponential growth in information and advancement in “big data” storage and sharing and improvements in computational power have matured this field for application of machine learning

(ML). Benedetto and colleagues² report

Arman Arghami, MD (left), and Joseph A. Dearani, MD (right)

CENTRAL MESSAGE

Machine learning helps improve prediction models of mortality after cardiac surgery and will transform today's evidence-based-practice to tomorrow's practice assisted with artificial intelligence.

a systemic review of predictive modeling with ML (also referred to as artificial in-

telligence). They reviewed published reports from 1966 to 2019 of ML modeling to predict mortality in cardiac surgical patients and compared them with the traditional logistic regression (LR) models. Despite their limitations, which

From the Department of Cardiovascular Surgery, Mayo Clinic, Rochester, Minn. Disclosures: The authors reported no conflicts of interest.

The *Journal* policy requires editors and reviewers to disclose conflicts of interest and to decline handling or reviewing manuscripts for which they may have a conflict of interest. The editors and reviewers of this article have no conflicts of interest.

Received for publication Aug 28, 2020; revisions received Aug 28, 2020; accepted for publication Aug 31, 2020; available ahead of print Sept 3, 2020.

Address for reprints: Joseph A. Dearani, MD, Department of Cardiovascular Surgery, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (E-mail: jdearani@mayo.edu).

J Thorac Cardiovasc Surg 2022;163:2094-5

0022-5223/\$36.00

Copyright © 2020 by The American Association for Thoracic Surgery

<https://doi.org/10.1016/j.jtcvs.2020.08.103>