Editorial

The Promise of Machine Learning: When Will it be Delivered?

OGUZ AKBILGIC, PhD,[1,2] AND ROBERT L. DAVIS, MD, MPH[2]

Memphis, TN; and Maywood, IL

ABSTRACT

Background: The real-life applications of machine learning clinical decision making is currently lagging behind its promise. One of the critics on machine learning is that it doesn't outperform more traditional statistical approaches in every problem.

Methods and Results: Authors of "Predictive Abilities of Machine Learning Techniques May Be Limited by Dataset Characteristics: Insights From the UNOS Database" presented in the current issue of the Journal of Cardiac Failure that machine learning approaches do not provide significantly higher performance when compared to more traditional statistical approaches in predicting mortality following heart transplant. In this brief report, we provide an insight on the possible reasons for why machine learning methods do not outperform more traditional approaches for every problem and every dataset.

Conclusions: Most of the performance-focused critics on machine learning are because the bar is set unfairly too high for machine learning. In most cases, machine learning methods provides at least as good results as traditional statistical methods do. It is normal for machine learning models to provide similar performance with linear models if the actual underlying input-outcome relationship is linear. Moreover, machine learning methods outperforms linear statistical models when the underlying input-output relationship is not linear and if the dataset is large enough and include predictors capturing that nonlinear relationship. (*J Cardiac Fail 2019;25:484−485*)

Key Words: Machine learning, Artificial intelligence, Heart transplant, UNOS.

Machine learning techniques are increasingly being developed to address clinical decision-making problems in medicine, yet—at least in clinical practice—real-world applications lag behind the promise. There are several reasons for this state of affairs, one of which is that machine-learning approaches do not necessarily outperform more traditionally used statistical methods for every problem.

In the current issue of the *Journal of Cardiac Failure*, the authors of "Predictive Abilities of Machine Learning Techniques May Be Limited by Dataset Characteristics: Insights From the UNOS Database"[1] present one such example where machine-learning algorithms do not significantly outperform classical linearized statistical models in predicting 1 year survival following heart transplant. Their study used 28 years of data from the United Network for Organ Sharing (UNOS) database and compared standard statistical methods with neural networks and other machine-learning methodologies. Somewhat surprisingly, the best performing approaches for machine learning were only minimally improved compared with more standard statistical methods. Why might this have happened?

Machine learning can model nonlinear predictor-outcome relationships and multi-dimensional interactions between predictors. As a result, the performance of machine learning is highly dependent on the underlying predictor−outcome relationship, which is unknown in most cases. Furthermore, some predictor−outcome relationships may have both linear and nonlinear components; however, if the predictors representing the nonlinear part are not observed (ie, the data are not collected and/or not included in the dataset) or the sample size is not large enough to optimize the parameters of the machine-learning architecture to represent such nonlinearity, machine learning may not outperform linear statistical methods. Such expectations are not so much a failure of machine learning (because, after all,

From the [1]University of Tennessee Health Science Center-Oak Ridge National Laboratory Center for Biomedical Informatics, Memphis, TN 38103 and [2]Department of Health Informatics and Data Science, Loyola University Chicago, Maywood, IL 60153.

Manuscript received April 5, 2019; revised manuscript accepted April 5, 2019.

Reprint requests: Oguz Akbilgic, PhD, Loyola University Chicago, CTRE Rm 127, 2160 S First Avenue, Maywood, IL 60163. E-mail: oakbilgic@luc.edu

See page 485 for disclosure information.

1071-9164/$ - see front matter

© 2019 Elsevier Inc. All rights reserved.

https://doi.org/10.1016/j.cardfail.2019.04.006

the statistical methods provided similar results) as much as the limitations of the data itself. Although the authors characterized the clinical dataset as having limited quality, another reasonable interpretation is that the quality of the data was acceptable, yet simply limited in scope.

Other than clinical factors, survival after heart transplant may also be determined by socioeconomic and environmental factors. However, the track record to date is mixed in demonstrating that inclusion of social determinants of health data significantly improves predictive models. For example, Krumholz et al[2] showed that inclusion of socioeconomic and other nonclinical factors improved model prediction, but regardless, the prediction models performed poorly. Others have described similar results, for heart failure[3] or, for example, when attempting to identify high-risk patients at safety-net hospitals.[4] Similarly, in a study of children with asthma, we found that social determinants of health data improved the ability to predict readmissions,[5] but again only modestly.

Future predictive models may be improved by taking into account more finely detailed data on biologic parameters. Predictive performance might be aided, for example, by inclusion of biomarker or “omic” data (genomic, transcriptomic, metabolomics, etc), imaging data, or longitudinal physiologic data collected prior to transplantation. A recent study (outside of cardiology) by Toffalori et al[6] identified a set of differentially expressed genes (“transcriptional signature”) able to identify patients with leukemia at high risk for relapse after stem cell transplants. In intensive care units, features extracted from streaming physiologic data can predict sepsis more accurately using machine learning compared to logistic regression. [7] Similar findings from other groups[8–12] seem to indicate that the extraordinary volume and granularity of data from these medical environments are especially well suited for examination by machine-learning approaches

It is important to recognize that the application of machine learning and artificial intelligence to clinical medicine is in its infancy. Although it is easy to say that we need more data, including genomics, biomarkers, more granular electronic medical record data, sensor data, etc, to predict future events such as readmission, simply acquiring personal health information from subjects has a whole set of challenges not typically found in other industries more accustomed to using extremely high-volume data for business reasons. The field of medicine is making rapid progress to create the infrastructure(s) that will enable continuous machine-learning using clinical and other data. In time, we will likely see more areas of medical and clinical research in which machine learning lends distinct insight and benefit over more traditionally used linearized statistical methods.

Disclosures

No conflicts of interest to declare.

References

1. Miller P, Pawar S, Vaccaro B, et al. Predictive abilities of machine learning techniques may be limited by dataset characteristics: insights from the UNOS database 2019;25:479–83.
2. Krumholz HM, Chaudhry SI, Spertus JA, Mattera JA, Hodshon B, Herrin J. Do non-clinical factors improve prediction of readmission risk? Results from the Tele-HF study. JACC Hear Fail 2016;4:12–20.
3. Ahmad FS, French B, Bowles KH, Sevilla-Cazes J, Jaskowiak-Barr A, Gallagher TR, et al. Incorporating patient-centered factors into heart failure readmission risk prediction: a mixed-methods study. Am Heart J 2018;200:75–82.
4. Kasthurirathne SN, Vest JR, Menachemi N, Halverson PK, Grannis SJ. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. J Am Med Informatics Assoc 2018;25:47–53.
5. Shin EK, Mahajan R, Akbilgic O, Shaban-Nejad A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. npj Digit Med 2018;1:50.
6. Toffalori C, Zito L, Gambacorta V, Riba M, Oliveira G, Bucci G, et al. Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. Nat Med 2019;25:603–11.
7. Kamaleswaran R, Akbilgic O, Hallman MA, West AN, Davis RL, Shah SH. Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the PICU. Pediatr Crit Care Med 2018;19:e495–503.
8. Kovatchev BP, Farhy LS, Cao H, Griffin MP, Lake DE, Moorman JR. Sample asymmetry analysis of heart rate characteristics with application to neonatal sepsis and systemic inflammatory response syndrome. Pediatr Res 2003;54:892–8.
9. Cao H, Lake DE, Griffin MP, Moorman JR. Increased nonstationarity of neonatal heart rate before the clinical diagnosis of sepsis. Ann Biomed Eng 2004;32:233–44.
10. Lake DE, Richman JS, Griffin MP, Moorman JR. Sample entropy analysis of neonatal heart rate variability. Am J Physiol Regul Integr Comp Physiol 2002;283:R789–97.
11. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015;7. 299ra122.
12. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017;89:248–55.