Published in final edited form as:

Crit Care Med. 2017 November ; 45(11): 1946–1947. doi:10.1097/CCM.0000000000002673.

Machine learning and sepsis: on the road to revolution

Vincent Liu, MD, MS¹ and Allan Walkey, MD, MSc²

¹Kaiser Permanente Division of Research, Oakland, CA

²Department of Medicine and Evans Center for Implementation and Improvement Sciences, Boston University; Boston, MA

Keywords

Sepsis; machine learning; risk prediction; early warning scores

With each passing day, the drumbeat of the advancing Big Data and machine learning revolution grows louder.(1) Already having conquered previously intractable tasks like interpreting complex images or voices, driving autonomous vehicles, and winning abstract strategy games, computationally-driven advances are producing disruptions across a wide swath of industries. These advances have arisen from a confluence of three major trends: rapid expansions in the availability of large-scale data; exponential increases in the accessibility of computational power; and critical progress in machine learning methods.

Naturally, many see healthcare as the next major industry ripe for innovation. And, as the National Academy of Medicine has highlighted in numerous landmark reports, advances in the use of data and information technology will be critically important for addressing systemic failures in US healthcare quality, safety, and value.(2) In particular, recent successes in clinical image recognition and ongoing advances in leveraging large-scale – omics data highlight the great promise of machine learning applications in the biomedical and healthcare enterprise.(3, 4)

In this issue of *Critical Care Medicine*, Churpek and colleagues further expand on the promise of machine learning approaches in sepsis: the single most common, costly, and deadly cause of US hospitalization.(5) Extending their prior work in developing and validating the electronic Cardiac Arrest Risk Triage (eCART) score among hospitalized ward patients – the score incorporates granular vital signs, laboratory values, and patient demographics extracted from electronic medical records (EMR) data(6) – they demonstrate that eCART scores uniformly provided the highest discrimination among inpatients with varying degrees of “suspected infection” for predicting a composite adverse outcome of intensive care unit (ICU) transfer or death within 48 hours. As is critical for innovation in this domain, their work capitalized on a large and highly granular dataset including >53,000

Corresponding author: Vincent Liu, MD, MS, 2000 Broadway, Oakland, CA 94612, 510-891-5933.

Copyright form disclosure: Dr. Liu’s institution received funding from the National Institute of General Medical Sciences, and received support for article research from the National Institutes of Health. Dr. Walkey has disclosed that he does not have any potential conflicts of interest.

infected patients as well as access to the computational resources needed to develop and deploy a random forest-based machine learning approach to score development and determination.

Yet, even while the drumbeat of the machine learning revolution portends great promise, it is also accompanied by potential pitfalls.(7, 8) In this case, the discrimination of the eCART score was uniformly better than all other risk scores (median c-statistic, 0.73), however, it was only modestly better than the 7-element, vital sign-based NEWS score (National Early Warning Score; median c-statistic, 0.71). Thus, while the eCART was statistically superior to the NEWS scores, it remains unclear what the clinical impact of this advantage would be particularly since the eCART score also demonstrated modestly lower specificity in some scenarios. Additionally, the comparisons of the eCART versus other scores were repeated in a sample drawn from the same setting in which the score was derived. Whether eCART shows better performance in external validation cohorts remains unclear.

Beyond the statistical development of early warning scores, another major challenge remains identifying the appropriate interventions that should follow a high-risk alert. In a randomized trial deploying a real-time early warning score with excellent discrimination, Kollef and colleagues found that the transmitted alerts in the intervention arm neither decreased mortality nor increased the rate of ICU transfers. The alerts did result in a shorter overall hospital length of stay.(9) Given the impressive performance characteristics of the eCART score, we look forward to a comprehensive evaluation of its impact on clinical outcomes and, in particular, a detailed description of the efferent intervention arm. This will greatly improve our understanding of which elements of implementation are most likely to drive improved outcomes.

Finally, Churpek and colleagues found that eCART performance remained consistent even as the inclusion cohort criteria varied from minimal evidence of suspected infection (i.e., any culture ordered) to considerably stronger evidence (i.e., blood culture orders and intravenous antibiotics for at least four out of seven living days). Their findings confirm that the predictors conferring prognostic risk (i.e., the likelihood of experiencing an adverse event) in hospitalized patients are common across many conditions. At the same time, the determinants of predictive risk (i.e., the likelihood of responding to a specific therapy) are likely to vary across patient subgroups, particularly in heterogeneous syndromes like sepsis. Thus, future work incorporating biomolecular data for predictive enrichment will be critical to target therapies to specific sepsis endotypes who otherwise share the same general risk of adverse events.(10) These factors are certain to gain prominence in future consensus definitions of sepsis and will require carefully balancing the need for increasing predictive validity compared with preserving content validity, reliability, and low measurement burden. (11)

As Churpek and colleagues have shown, the machine learning revolution continues to draw closer, however, more work is needed to ensure that the future is one filled with promise and not pitfalls.

References

1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013; 309(13):1351–1352. [PubMed: 23549579]
2. IOM (Institute of Medicine). Best care at lower cost: The path to continuously learning health care in America. Washington, DC: The National Academies Press; 2012.
3. Beam AL, Kohane IS. Translating Artificial Intelligence Into Clinical Care. *JAMA*. 2016; 316(22): 2368–2369. [PubMed: 27898974]
4. Dzau VJ, Ginsburg GS. Realizing the Full Potential of Precision Medicine in Health and Health Care. *JAMA*. 2016; 316(16):1659–1660. [PubMed: 27669484]
5. Churpek MM, Snyder A, Sokol S, et al. Investigating the impact of different suspicion of infection criteria on the accuracy of qSOFA, SIRS, and early warning scores. *Crit Care Med*. 2017 in press.
6. Churpek MM, Yuen TC, Winslow C, et al. Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med*. 2014; 190(6):649–655. [PubMed: 25089847]
7. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. *JAMA*. 2017
8. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med*. 2017; 376(26):2507–2509. [PubMed: 28657867]
9. Kollef MH, Chen Y, Heard K, et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J Hosp Med*. 2014; 9(7):424–429. [PubMed: 24706596]
10. Prescott HC, Calfee CS, Thompson BT, et al. Toward Smarter Lumping and Smarter Splitting: Rethinking Strategies for Sepsis and Acute Respiratory Distress Syndrome Clinical Trial Design. *Am J Respir Crit Care Med*. 2016; 194(2):147–155. [PubMed: 27244481]
11. Angus DC, Seymour CW, Coopersmith CM, et al. A Framework for the Development and Interpretation of Different Sepsis Definitions and Clinical Criteria. *Crit Care Med*. 2016; 44(3):e113–121. [PubMed: 26901559]