Thinking Machines and Risk Assessment: On the Path to Precision Medicine

Kirk N. Garratt, MD, MSc; Mark A. Schneider, MD, MBA

A 65-year-old man undergoes percutaneous coronary intervention (PCI) to treat an acute myocardial infarction. He has a nonzero risk of dying; that risk may be predicted using models built chiefly through logistic regression of a sample of variables believed to impact survival after acute myocardial infarction. The number of variables is restricted because integrating the independent (or, worse, interactive) impact of many factors when building risk prediction models is difficult, and the burden of having to enter many variables manually into a prediction tool makes the tool less useful.

Machine learning (ML) is the process of applying statistical algorithms to data sets to model and infer predictions. It is formally a subset of artificial intelligence, but to the nontechnical user, ML may seem like the thinking, reasoning part of artificial intelligence. Good reviews of ML applications in medicine are available to the interested reader; the approachable, clinically oriented summary by Rahul Deo of the University of California, San Francisco, is a good starting point.[1] An algorithm is trained on a corpus of information that contains labeled data on clinical features and defined outcomes (eg, the characteristics of patients with acute myocardial infarction and the frequency of their of dying). Although not explicitly programmed, the algorithm builds a model that predicts linkage between the input variables and the target outcome. This process is referred to as supervised learning. With the right training and access, machines can pluck data from source pools without any human interaction, which greatly enhances their utility. After training, ML models can process huge volumes of data quickly and provide enhanced predictive power compared with conventional risk models. Furthermore, formerly disparate data sets can be combined to improve predictive power. As an example, computers taught to integrate clinical, laboratory, and image data for patients with rheumatoid arthritis have recently been used to generate automated cardiovascular risk prediction models that proved superior to standard risk models for individual patient risk assessments.[2] However, the real magic of ML is in unsupervised learning, that is, the ability to observe trends or patterns in unlabeled data and speed the creation of hypotheses about their meaning. When given the right instructions and access, computers can perform novel work without any direct human interaction. The abilities of a machine to generate plausible hypotheses, test them, and draw conclusions increase with exposure and experience, much like humans. Because of this, ML has the potential to suggest new, independent discoveries when studying large data sets. This aspect of ML has been harnessed thoroughly in some disciplines, like genomic and protein structure basic research, but it is just getting started in clinical medicine. For example, a report earlier this year described how several distinct immune phenotypes were identified in a population of patients with group I pulmonary hypertension when ML algorithms studied markers of systemic inflammation.[3]

Al'Aref and colleagues report in this issue of the *Journal of the American Heart Association* (*JAHA*) on their use of advanced ML methods to develop a tool to predict risk of death after PCI.[4] They used a fairly broad set of clinical and demographic variables drawn from a statewide database of nearly one half million patients undergoing PCI, including factors not always considered in traditional models but that are likely to impact outcome, such as lesion complexity, postprocedure complications, and day of the week that PCI was performed. The authors used several techniques to optimize utility of the data set. For example, when multiple data points were missing, an advanced method for imputing values based on remaining data was used. They used 4 different adaptive learning algorithms to produce prediction models, using the results of each to improve the output of the others. Although the language of this work will be foreign to most clinicians, the methods are well established. The

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Center for Heart and Vascular Health (K.N.G.) and Applied Innovation Group (M.A.S.), Christiana Care Health System, Wilmington, DE.

Correspondence to: Kirk N. Garratt, MD, MSc, Center for Heart and Vascular Health, Christiana Care Health System, 4755 Ogletown-Stanton Rd, Ste 1003, Newark, DE 19718. E-mail: kirk.n.garratt@christianacare.org

J Am Heart Assoc. 2019;8:e011969. DOI: 10.1161/JAHA.119.011969.

resulting final tool produced greater precision in risk estimation than traditional tools.

This is the first known application of ML to predict outcomes in a large sample of patients with PCI. Like similar studies in other patient groups, the current study shows that ML, when given access to rich data sources, can both enhance the precision and simplify the use of risk prediction instruments. The data source for this work was a state-based registry containing many factors that would not necessarily be thought of as essential to a reliable risk model. Although this study demonstrates the value of making more variables available to ML algorithms, it also underscores the limitations imposed with use of any curated data set rather than complete data. The ubiquity of actual or potential data sources (computer-oriented living environments, home-based electronics, wearable sensors, and implanted devices, all with the potential for interconnectivity) creates an infrastructure for collecting comprehensive data. But before overreaching, we should focus on an extremely rich data pool available now: the electronic health record (EHR).

The fact that an electronic learning instrument could theoretically reside within a hospital's computing system raises interesting possibilities for risk prediction tools. Why rely on models built from abstracted registry data when ML can exploit everything in the EHR? Using an institution's own data may produce results more applicable to that organization's patient population. In addition to an expanded set of input variables, more potential outcome measures should be available, providing there is reasonably good entry of follow-up data. Removing the barriers imposed through use of abstracted outcome measures would support the supervised and unsupervised functions of an embedded ML algorithm. Theoretically, linkage between hospital systems using the same EHR (or, in a perfect future state, all EHRs) could create a vast, comprehensive medical data set suitable for automated production of near-perfect outputs. Early work using EHR-embedded ML shows that greatly improved risk analytics are possible for cardiology patients.[5] Imagine being able to quantify and estimate the consequences of every key clinical decision for our theoretical patient with acute myocardial infarction, based on experience in tens of millions of patients and hundreds of millions of interactions, before executing an order.

Risk assessment would be just one aspect of care to benefit. Just as registry data combine the efforts of many institutions, linked EHRs with ML capabilities would allow continuous learning at a system level. It is easy to imagine a computer advising on medical therapies: not just a reminder about guideline recommendations, but a digital assistant that surveys cardiac monitors for rhythm patterns, radiologic studies for changes in pulmonary congestion, and laboratory results for metabolic patterns, and then guides subsequent decisions. Integration of newer medical information, like genomics, will advance abilities further, as will EHR capture of data from nonstandard sources, such as wearable electronics. This is a new way of thinking that requires acceptance and thoughtful application, but as pointed out by Weintraub and colleagues, incorporation of ML into the EHR has the potential to speed us toward our collective goal of personalized, precision medicine perhaps faster than any other single action.[6] As a result of this potential, consulting firms like Accenture forecast explosive growth in use of ML and artificial intelligence, predicting its commercial value to approach $7 billion by 2021.[7]

Enhanced predictive analytics through ML may be an important stepping stone toward integration of big data processing intelligence into clinical care environments. Al'Aref and colleagues[4] are congratulated for demonstrating the utility of ML in predicting outcomes in a large sample of patients undergoing PCI, a population that involves both high risk and high costs, personally and for health systems. Improving risk assessment facilitates efficient management of this population. We must acknowledge, however, that implementation of ML instruments into clinical practice to enhance mortality risk estimation is neither simple nor inexpensive. Modest improvements in predictive accuracy may hardly seem worth the effort. The real payoff may well come when the more advanced abilities of ML are put to work in the clinical environment. Next steps for ML, including full integration into EHRs, comprehensive support for clinical decision making, and establishment of structures to search for clinically relevant new discoveries, could come quickly, but this means clearing some big hurdles. Some challenges ahead are technical, certainly, but the greatest challenges are likely to be nontechnical: administrators, corporate interests, and regulators must find ways to share data in novel ways, whereas physicians and other providers must learn to trust that machines making millions of observations can offer good advice to augment usual clinical judgement.

Disclosures

None.

References

1. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920–1930.
2. Khanna NN, Jamthikar AD, Gupta D, Piga M, Saba L, Carcassi C, Giannopoulos AA, Nicolaides A, Laird JR, Suri HS, Mavrogeni S, Protogerou AD, Sfikakis P, Kitas GD, Suri JS. Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning-based tissue characterization. *Curr Atheroscler Rep*. 2019;21:7.
3. Sweatt AJ, Hedlin HK, Balasubramanian V, Hsi A, Blum LK, Robinson WH, Haddad F, Hickey PM, Condliffe RA, Lawrie A, Nicolls MR, Rabinovitch M, Khatri P, Zamanian RT. Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ Res*. 2019. Available at: https://doi.org/10.1161/CIRCRESAHA.118.313911. Accessed March 1, 2019.

4. Al'Aref SJ, Singh G, van Rosendael AR, Kolli KK, Ma X, Maliakal G, Pandey M, Lee BC, Wang J, Xu Z, Zhang Y, Min JK, Wong SC, Minutello RM. Determinants of in-hospital mortality after percutaneous coronary intervention: a machine learning approach. *J Am Heart Assoc*. 2019;8:e011160. DOI: 10.1161/JAHA.118.011160.
5. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, Denny JC, Wei WQ. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep*. 2019;9:717.
6. Weintraub WS, Fahed AC, Rumsfeld JS. Translational medicine in the era of big data and machine learning. *Circ Res*. 2018;123:1202–1204.
7. Accenture, April 2017. Artificial intelligence: Healthcare's new nervous system. Available at: https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare. Accessed February 10, 2019.

Key Words: Editorials • artificial intelligence • computer-based model • machine learning • percutaneous coronary intervention • risk prediction