COMMENTARY

Predictive analytics by deep machine learning: A call for next-gen tools to improve health care

Philip Wells MD, MSc, FRCPC

Department of Medicine, Ottawa Hospital Research Institute, The Ottawa Hospital, University of Ottawa, Ottawa, ON, Canada

This is a commentary on *Nafee T et al* [2020]: https://doi.org/10.1002/rth2.12292

Correspondence
Philip Wells, Department of Medicine, Ottawa Hospital Research Institute, The Ottawa Hospital, University of Ottawa, General Campus, 501 Smyth Road, Box 206, Ottawa, ON K1H 8L6, Canada.
Email: pwells@toh.ca

Handling Editor: Mary Cushman.

With the massive growth in computing capacity and the connectivity that exists in the world today, medicine has changed drastically. These changes bring opportunity for more change to improve health care. Medical information is now available at our fingertips and can be acquired rapidly. As such, the massive memorizing tasks and learning skills of the past are less relevant. Robust clinical pathways are available in apps, online medical care tools, and health information systems, and in some cases are proven to improve patient care, with failure to follow them resulting in worse patient outcomes.[1] Thirty years ago, there were only a handful of clinical prediction algorithms/models to assist physicians in decision making. Now there are hundreds. Despite this, the uptake of algorithms in clinical practice has been slow, sporadic, and fraught with skepticism.[2] This uptake, or lack thereof, has been justified by arguments that predictive algorithms were developed in populations of patients that were not necessarily applicable to "the patient in front of me." In other words, studies were not generalizable. Yet physicians have become all too quick to order diagnostic tests without following algorithms or predictive tools under the assumption that this is best, and ignoring the possibility of harm, including radiation exposure, false-positive tests, and the economic burden on society. It is estimated that 5% of the US gross domestic product is spent on diagnostic tests and procedures that do not result in any improvement in patient outcomes.[3]

The time for broader application and use of predictive algorithms for diagnosis, treatment, prevention, and prognosticating is here. Technological advances enabling the use of big data enable the possibility of developing more accurate prediction models; continuous variables previously categorized as binary for ease of use and prediction rules can now be used as continuous variables in algorithms created with deep machine learning, providing more predictive power.

Artificial neural networks, the modeling used in deep machine learning, identifies complex relationships in large data sets and can apply this to newly added data to continually improve algorithms; this will provide more individualized assessments for our patients.[4] These algorithms can be web or app based. In this issue of *Research and Practice in Thrombosis and Haemostasis*, Nafee and colleagues applied machine learning to predict the risk of developing of venous thromboembolism (VTE) in acutely ill patients considered to be at higher risk of VTE. They used data from 7513 acutely ill medical patients who were enrolled in the APEX (Acute Medically Ill Venous Thromboembolism Prevention With Extended Duration Betrixaban) trial, which studied extended-duration betrixaban vs. shorter-duration enoxaparin.[5] Their super learner model and their "reduced" model both outperformed the previously validated IMPROVE (International Medical Prevention Registry on Venous Thromboembolism) score in predicting VTE. However, the study illustrated that machine learning is not enough. The c-statistic to predict thromboembolism, even utilizing super computing power, was only 0.69, so the proposed model is not ideal.

Why did the approach of Nafee and colleagues fall short? Perhaps the answer lies in the quality of the analyzed data. Although health information systems allow us to enter massive amounts of

information and potentially share that data across centers (a promise not yet realized due to fragmentation of data sources), the quality of the data is often poor. In prospective studies or trials designing and testing predictive algorithms and prediction tools, a key design component is strict definition of variables that are collected, testing of interobserver agreement during data variable collection, and using objective data whenever possible to verify these variables. A recent study demonstrated that neural network analysis can improve diagnostic pretest probability of a diagnosis of deep vein thrombosis (DVT) stratification of patients with suspected DVT; the algorithm significantly reduced the proportion of patients requiring ultrasound to exclude DVTs, but the study was limited by the single-center design and limited number of observers inputting data.[6]

Regardless of the potential of artificial intelligence and deep machine learning, we must remember to follow basic rules in the development of prediction rules, as outlined on the Tripod website (http://www.tripod-statement.org), and endorsed by the EQUATOR Network (http://www.equator-network.org). There will always be limitations to quality of clinical information collected by health information systems. As such, prediction tools must focus on including variables known to be accurate and reproducible using this source of information. As an example, thrombophilia and immobility are unlikely to be useful variables in a prediction model, yet they are included in IMPROVE.[7] It is simply not feasible or economically viable to test all patients for thrombophilia. Furthermore, thrombophilia tests have their own limitations and are not necessarily accurate.[8] If artificial intelligence creates clinical tools through analysis of highly curated data from real-world applications, where data are less accurate, results may be limited.[9] This is not to say that I disagree with machine learning models and the use of predictive algorithms, in fact, I am a major advocate and have publicly stated such (https://www.youtube.com/watch?v=QWps8A-hljw). Certainly, the machine learning model was superior to the original IMPROVE score in Nafee's study. As the authors illustrate and indicate in their discussion, traditional risk assessment models do have many limitations. The authors provide useful insight by noting that it is also important to simultaneously apply a tool that evaluates the risk of bleeding, as all decisions in preventing and treating venous thrombosis come down to balancing risk versus benefit (similar to any other disease). With respect to risk, it is important for those of us in the venous thrombosis community to come to agreement on what the appropriate end points are in trials of prophylaxis. It may be argued that the outcome asymptomatic DVT, included in the APEX trial, is less relevant than symptomatic DVT.[10] One must be cautious to not include clinically irrelevant diseases in outcomes and in decision tools, or this will lead to systematic overtreatment of patients, and no reduction of relevant patient outcomes. Perhaps the most relevant outcomes in venous thrombosis are quality of life and death. Finally, it goes without saying that any new predictive algorithm must undergo rigorous assessment in prospective multicenter trials, and with appropriate outcomes.

In conclusion, I encourage physicians to have an open mind about artificial intelligence and deep machine learning, and to embrace the application and use of predictive algorithms that undoubtedly will unfold over the next decade. This is one of the key pathways to cost-effective, efficient, and safe health care. We should overcome fear of the black box concept of articifial intelligence, and physicians need to trust that large, well-managed data sets can produce tools that will improve patient care. We must always guard against the all-too-easy default position of practicing anecdotal medicine. It is human nature to do such, but it does not improve patient care, especially from a societal perspective.

RELATIONSHIP DISCLOSURE

ORCID

Philip Wells https://orcid.org/0000-0002-8657-8326

TWITTER

Philip Wells @PhilWellsMD1

REFERENCES

1. Roy P-M, Meyer G, Vielle B, et al. Appropriateness of diagnostic management and outcomes of suspected pulmonary embolism. Ann Intern Med. 2006;144(3):157–164.
2. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21):2199.
3. Laine C. High-value testing begins with a few simple questions. Ann Intern Med. 2012;156(2):162–163.
4. Tsega S, Cho HJ. Prediction and prevention using deep learning. JAMA Netw Open. 2019;2(7):e197447.
5. Cohen AT, Harrington RA, Goldhaber SZ, et al. Extended thromboprophylaxis with betrixaban in acutely ill medical patients. N Engl J Med. 2016;375(6):534–544.
6. Willan J, Katz H, Keeling D. The use of artificial neural network analysis can improve the riskâ stratification of patients presenting with suspected deep vein thrombosis. Br J Haematol. 2019;185(2): 289–296.
7. Rosenberg D, Eichorn A, Alarcon M, McCullagh L, McGinn T, Spyropoulos AC. External validation of the risk assessment model of the International Medical Prevention Registry on Venous Thromboembolism (IMPROVE) for medical patients in a tertiary health system. J Am Heart Assoc. 2014;3(6):e001152.
8. Lane D, Mannucci P, Bauer K, et al. Inherited thrombophilia: part 2. Thromb Haemost. 1996;76(6):824–834.
9. Nsoesie EO. Evaluating artificial intelligence applications in clinical settings. JAMA Netw Open. 2018;1(5):e182658.
10. Heit JA, Elliot G, Trowbridge AA, Morrey BF, Gent M, Hirsh J. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement. A randomized, double-blind, placebo-controlled trial. Ann Intern Med. 2000;132(11):853–861.