Machine Learning: The Next Paradigm Shift in Medical Education
Cornelius A. James, MD, Kevin M. Wheelock, MD, and James O. Woolliscroft, MD

Abstract

Machine learning (ML) algorithms are powerful prediction tools with immense potential in the clinical setting. There are a number of existing clinical tools that use ML, and many more are in development. Physicians are important stakeholders in the health care system, but most are not equipped to make informed decisions regarding deployment and application of ML technologies in patient care. It is of paramount importance that ML concepts are integrated into medical curricula to position physicians to become informed consumers of the emerging tools employing ML. This paradigm shift is similar to the evidence-based medicine (EBM) movement of the 1990s. At that time, EBM was a novel concept; now, EBM is considered an essential component of medical curricula and critical to the provision of high-quality patient care. ML has the potential to have a similar, if not greater, impact on the practice of medicine. As this technology continues its inexorable march forward, educators must continue to evaluate medical curricula to ensure that physicians are trained to be informed stakeholders in the health care of tomorrow.

Machine learning (ML) is a form of artificial intelligence (AI) that learns the relationships between data without the need to define them a priori. ML drives pop-up ads, suggests purchases on Amazon, runs automated stock trading funds, predicts the weather, and performs many other tasks. Not surprisingly, clinical applications of ML are a focus of intense development. For example, recently, the Data Science Institute of the American College of Radiology listed 48 U.S. Food and Drug Administration (FDA)-approved ML algorithms related to medical imaging.1 Indicative of the increasing sophistication of ML clinical programs, in 2018, the FDA approved the first autonomous ML system coupled with a fundus camera, IDx-dr, for the detection of diabetic retinopathy. Clinical applications for ML have expanded beyond image analysis; ClinicalTrials.gov lists over 1,000 studies employing ML across a broad spectrum of applications ranging from diagnosis to enhancing desired healthy behaviors.2 Recently, ML algorithms have been used to monitor the vital signs of COVID-19-positive patients in real time to predict deterioration and to initiate interventions earlier to improve outcomes.3 As ML is increasingly integrated into health care, it is of paramount importance that medical educators equip clinicians with the knowledge to be sophisticated “consumers” of ML products rather than dependent on the word of marketers to determine applicability.

Although growing in prevalence and importance, ML-derived algorithms are far from standard in clinical practice or in medical education. There are questions about the ethics of using ML in medicine: whether physician skills will erode with more reliance on ML as well as the potential for ML to negatively affect the physician–patient relationship or replace some physicians completely.4,5 Concerns related to liability risks, payment mechanisms, and perpetuation of disparities have also contributed to the lack of broader acceptance and use of ML algorithms in practice. Clinicians’ and educators’ uncertainty as to the role and value of ML is often followed by skepticism. Much of this skepticism is due to a lack of understanding of what ML is and the complementary role it should play in clinical decision making. Scrutiny is necessary to provide optimal care for patients; uncritical acceptance or apathy must be avoided.

Data-Driven Patient Care

In charting the future of ML in health care, there are lessons to be learned from the evidence-based medicine (EBM) movement. In the 1990s, clinical epidemiologists at McMaster University called for a paradigm shift that would include changing the practice of medicine into an objective, scientific enterprise.6 They described a method to bring research topics in study design, epidemiology, and biostatistics to the front line of clinical care and equip clinicians to critically assess medical literature. This skill set is designed to help clinicians be informed consumers of published research and avoid relying solely on expert opinion or outdated information acquired during training. This radical idea was met with resistance as it came at a time when hierarchy, and the opinion of the “all-knowing” attending physician, were considered best practice. EBM, once a novel and abstract concept, is now deemed essential for high-quality patient care and is a staple in medical education. We expect ML to follow a similar path as it moves from abstraction to widespread application in clinical practice.

EBM and ML share practical similarities. Consider clinical prediction rules (CPRs), which are predictive tools commonly used to inform diagnostic, prognostic, and therapeutic decisions. An illustrative example is the CHADS<sub>2</sub> score. It was widely used to predict the risk of stroke in patients with atrial fibrillation and guide decisions about anticoagulation. Development of this CPR involved a rigorous, evidence-based, stepwise process.7
• Derivation: identification of diagnostic tests, history, and physical examination factors with predictive power

• Narrow and broad validation: initially, the rule is applied in a setting and population similar to those in derivation, but eventually it is applied in varying clinical settings and populations

• Impact analysis: demonstration that the rule is used by physicians, improves patient outcomes, and/or decreases costs

With time, experts expressed concerns that the CHADS2 score did not consider other important risk factors. This led to development of the CHADS2-VASC (congestive heart failure, hypertension, age ≥75 years, diabetes, prior ischemic stroke, vascular disease, age of 65–74 years, sex) score, which provides a more accurate assessment of stroke risk. This sequence is comparable to the training, tuning, and validation process used to develop ML algorithms. Similarly, ML models must be regularly reevaluated and revised as more information or data are made available, or else they risk becoming outdated, invalid, and useless.

EBM has facilitated fundamental changes in the way clinical medicine is practiced. We anticipate that ML-derived algorithms will also drive dramatic changes in health care. The coupling of genomic and physiologic biomarker determinations with powerful computational programs has ushered in the era of precision medicine, care tailored to the individual. While arguably most advanced in oncology, where it has had a significant impact on diagnosis and the choice of therapeutic options, this approach is being applied across an ever-growing spectrum of diseases. Increasingly, the potential of truly individualized, precision medicine is becoming a reality. It is being incorporated in a new model that is envisioned as personalized, predictive, preventive, and participatory, or “P4 medicine.”

Central to the development of precision and predictive medicine is ML, an array of mathematical approaches to identify complex relationships based on hundreds or thousands of variables and massive datasets. While the potential of precision and predictive medicine is gradually being realized, ML has already been used to develop hundreds or thousands of algorithms for specific applications. ML has the potential to augment decision making in a highly individualized way and play a complementary role in a variety of clinical situations.

The effective practice of EBM requires an understanding of how it should be employed when caring for patients. Likewise, if there is not an adequate understanding of the role and applicability of ML algorithms in clinical care, this could negatively affect patient outcomes. A prior criticism of EBM was that it was “cookbook” medicine. In other words, critics were concerned about a one-size-fits-all approach to medicine that would devalue, or even ignore, clinical expertise or patient circumstances. Similarly, an understanding of the appropriate use of ML is required to avoid problems such as social bias, which could perpetuate disparities when using ML applications.

Uncritical application of ML algorithms should be avoided to avoid suboptimal patient care.

Skills for the Data-Aware Physician

Few physicians will ever develop a therapeutic intervention, design a clinical trial, or build an ML model. However, all physicians should be able to effectively acquire, appraise, and apply literature that is relevant to their patients; this includes studies of ML-derived algorithms. As with any intervention that may affect patient care, ML models must be evaluated with empirical research studies to show evidence of efficacy and safety. Basic tenets of EBM are applicable to classic clinical studies and studies including ML-derived models. Are the results of the study valid? Are the results applicable/generalizable to my patient(s)? Is the study design (e.g., randomized controlled trial, cohort study, case–control study) appropriate for the type of question asked (e.g., diagnosis, prognosis, therapy)? Studies using ML come with additional questions: is the amount of testing/training data adequate? Is there any overlap between those datasets? Was the model further validated on an external dataset? To what gold standard is the algorithm being compared? These are but a few of the questions clinicians need to be prepared to ask when evaluating clinical applications for ML.

To have a voice in the design and deployment of ML algorithms clinicians will need to communicate effectively with data scientists. As with medicine, the field of ML is packed with language that is puzzling to the uninitiated. For instance, “deep learning,” which involves artificial neural networks with many hidden layers, can offer significant performance benefits over other ML approaches in the correct clinical setting. At first glance, it may seem that an algorithm using deep learning is always preferable to one without it. However, this performance improvement comes at the cost of turning the algorithm into a true “black box.” Complex neural networks with many hidden layers make it difficult, if not impossible, to fully understand how a given set of inputs influences the outputs of a model.

Understandingly, some clinicians and patients may not trust a model when they are unable to comprehend how it works, especially when the physician disagrees with the recommendation of the model.

Physicians must also appreciate what makes an ML model valid and useful. For example, the previously mentioned IDx-dr model met the following standards:

• Abundance of training data: Before evaluation in a clinical trial, the IDx-dr algorithm was trained and validated using over 1 million photographs of diabetic retinopathy lesions. Generally, ML algorithms need to be trained on thousands of data points to be reliable.

• Well-defined inputs: The input to the algorithm, fundus images, is highly standardized across patients and institutions. When algorithms rely on inputs gathered from a subjective data source, the validity of the algorithm may suffer.

• Presence of a clear gold standard: The algorithm was tested against the interpretation of a group of experts using a standardized scoring system, to ensure it agreed with widely accepted clinical standards.

• Purpose beyond prediction: The result of the algorithm, presence or absence of diabetic retinopathy, is clinically meaningful.

Finally, physicians must allow the performance of an algorithm to influence
how they integrate it into clinical decision making. To conceptually understand the performance of ML models, learners need a foundation in statistics. ML literature contains some statistical concepts that are slightly different from, but conceptually related to, those commonly used in EBM. For instance, positive predictive value is mathematically equivalent to precision in ML parlance. Similarly, sensitivity is equivalent to recall. ML model performance is often assessed using the F1 statistic, which is derived from precision and recall. This is analogous to the area under the receiver operating characteristic curve, which is derived from sensitivity and specificity. These statistics measure performance in unique ways that can have a profound impact on clinical decision making.

The Time Is Now
ML in health care is moving forward at a rapid pace, and it appears to be doing so with or without physicians. Numerous startups have developed ML-driven products promising to improve the health of patients. This has led to concerns about commercialization and potential ethical issues that may follow as developers attempt to profit. As private companies collect ever-increasing amounts of patient data, some patients and physicians have questioned who actually owns these data and determines what they are used for, and who is at fault when databases are hacked by perpetrators. Information technologists must be held accountable for the key role that they play in the protection of patient information. Similar to the pharmaceutical industry, there are many stakeholders involved in ML in health care. The values of different stakeholders are not always aligned, which makes some form of governance and regulation necessary. Perhaps we will begin to see government oversight agencies developed to ensure that harmful algorithms are not approved for use in practice. Physicians as educators, policy makers, and frontline clinicians are uniquely positioned to serve as safeguards against the ethical concerns related to ML in medicine. We must take up this mantle if we are to fulfill our responsibility to protect our patients.

The medical community, on an individual level and a systems level, is slow to adapt to change or put new recommendations into practice even when they are based on strong evidence. Too frequently medical education similarly lags. Learners, and worse, patients have undoubtedly suffered because of this. Countless physicians continue to practice with an inability to effectively acquire, interpret, and apply evidence. Current learners entering the wards and clinics are receiving instruction from physicians who may not have a solid foundation in EBM. Because EBM has a firm footing in medical education, this is likely to improve as future generations of physicians assume teaching roles. We can avoid this same fate for ML in medicine by proactively integrating ML into medical curricula before adoption of these technologies becomes more widespread. This will enable newer physicians to be ready to engage with and critically evaluate ML algorithms. They will be active participants in the process of integrating ML into health care rather than passive recipients.

Organizations such as the Accreditation Council for Graduate Medical Education, Liaison Committee on Medical Education, and American Board of Medical Specialties must begin to develop competencies to ensure that physicians are capable of appropriately using ML algorithms in practice. The Association of American Medical Colleges and the American Medical Association have called on medical educators to provide AI educational programming to current and future physicians. The onus is on us to answer this call and reimagine existing curricula. Decisions about how to best educate and train physicians to practice in a data-driven environment must be made.

Integration of ML Into Curricula
Given the significant mathematical and technical complexity of designing ML models, it is unrealistic to expect all physicians to become ML experts. Instead, focus should be on high-level principles that help learners understand and incorporate the outputs of ML algorithms into clinical decision making. To do so, learners should be conversant with and understand the jargon used in the ML field, recognize the types of clinical problems ML is most useful for solving, and be able to identify the performance trade-offs of different types of models.

To that end, we propose thoughtful integration of ML content into existing curricula and educational programming. Given the inherent overlap and similarities between EBM and ML, ML content would fit nicely into EBM curricula. For example, as with any tool used to aid in diagnostic and, ultimately, treatment decisions, an ML diagnostic algorithm must be compared with a gold standard in a prospective study, and this study must be critically appraised. Also, when learning about CPRs such as the ASCVD risk score, it would be equally beneficial to learn about the use of ML algorithms that aid in predicting risk; for example, predicting the risk of a positive COVID-19 test. ML content should be horizontally integrated with “Doctoring” and clinical skills courses. This would keep the content clinically based and learner centered to train learners to ignore the hype that can accompany ML, recognize the appropriate use of ML, and view it as a tool in their toolkit to care for patients. Too often, concepts are deemed important in the classroom, but utilization of these concepts is not demonstrated on the wards and clinics. In other words, the hidden curriculum takes hold, and learners devalue certain aspects of medicine because this is what they see “real doctors” do (consciously or unconsciously) in practice. As the old adage states: “Actions speak louder than words.” This makes clear the importance of intentionally training physicians to be capable of applying ML in practice and effectively teaching it. Such vertical integration is likely to lead to reinforcement of the value and importance of ML in patient care and further development of skills necessary to effectively use ML programs in practice. Vertical integration will only occur with appropriate longitudinal integration across physicians’ careers, which should come in the form of continuing medical education and faculty development programs.

Concluding Remarks
Physicians and medical educators need to be key stakeholders as the use of ML in health care increases. How deeply we drive our stakes into the ground remains to be seen. Will physicians be an active voice in the development and implementation of ML algorithms? Who will we rely upon to teach physicians to apply ML in practice? We are in the midst of another paradigm shift in medicine.
Medical educators must embrace the call to deliver educational programs conducive to training evidence-based, data-conscious, and patient-centered physicians. Ignoring this call will prove detrimental to current and future physicians, and more importantly the patients for whom we care.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

C.A. James is assistant professor, Departments of Internal Medicine and Pediatrics, University of Michigan Medical School, Ann Arbor, Michigan.

K.M. Wheelock is an internal medicine house officer, Yale School of Medicine, New Haven, Connecticut.

J.O. Woolliscroft is professor, Departments of Internal Medicine and Learning Health Sciences, and Lyle C. Roll Professor of Medicine, University of Michigan Medical School, Ann Arbor, Michigan.

References


