3 Myths About Machine Learning in Health Care

by Derek A. Haas, Eric C. Makhni, Joseph H. Schwab and John D. Halamka
3 Myths About Machine Learning in Health Care

by Derek A. Haas, Eric C. Makhni, Joseph H. Schwab and John D. Halamka

NOVEMBER 13, 2019

Machine learning will dramatically improve health care. There are already a myriad impactful ML health care applications from imaging to predicting readmissions to the back office. But there are also high-profile, expensive efforts that have not achieved their goals.

In our collective roles as the CEO of a care delivery analytics business, tech-driven clinicians, and the leader of tech innovation at a major health system, we have developed and used dozens of ML applications. Many of these have succeeded, but others have not. From these experiences we have identified three common myths that exist around ML in health care.
**Myth 1: Machine learning can do much of what doctors do.**

The reality is that ML applications can perform some of what doctors do today, but they will not replace most of what doctors do in the foreseeable future (even radiologists). Doctors perform three main duties: (1) help prevent people from getting sick, (2) diagnose what’s wrong when people do, and (3) and then provide care and treatment. ML does have an important contribution to make with the first and second of these functions. For example, ML algorithms have proven especially useful in predicting cancer characteristics from imaging or in diagnosing fractures from x-rays. Unsupervised learning algorithms have demonstrated potential in linking disease risks to genomic biomarkers.

However, even with the further development of these applications, they will not replicate a doctor’s ability to provide care and treatment. The ML output still must be analyzed by someone with domain knowledge; otherwise, trivial data may be interpreted as essential and essential data as trivial. These relationships then have to be translated to actionable clinical management.

There is also a human element in helping patients decide whether and in what way to receive treatment. Patients often have concerns or apprehensions about undergoing treatment. Doctors need to incorporate the patient’s mental state, expectations, past history, and cultural factors into shared decision making with the patient and their family. Patients appreciate this human interaction and not receiving it at sensitive times may be upsetting.

Finally, once the treatment is completed, the recovery process itself requires close monitoring and care. Complications are often detected through clinical observation as opposed to protocol-driven testing or diagnostics.

**Myth 2: “Big data” + brilliant data scientists are always a recipe for success.**

The reality is that they are necessary but not sufficient. More data is better, but only if it is the right data and we fully understand it. We find it helpful to ask the following questions:

- **How was the data gathered?** Consider how the adoption of electronic health records (EHRs) could lead to all the diagnoses made and medications prescribed by different physicians for a patient to be captured in a single record — one that would be more comprehensive than individual physicians’ paper records. Without taking into account this change — which would reduce, if not eliminate, information falling through the cracks — one might wrongly conclude that all of a sudden, patients got sicker.

- **For what purpose was the data gathered?** Consider lab data collected by a hospital. Because the hospital is collecting data of patients who are being treated at the hospital, it will not be representative of the population since sick people are much more likely to have their blood drawn at the hospital.

- **What are potential problems with or limitations of the data?** Consider EHR data collected across many organizations that are all using the same vendor. Even though the organizations may all be using the
same EHR vendor, the data structure, field meanings, and extent of data cleaning likely differ across organizations.

Have circumstances changed? Is the data we used to build the ML model still valid? For instance, gender historically has been represented as male and female. Today gender identity is represented as male, female, other, and unknown. In the near future, it is likely to be represented as answers to a set of questions that enable clinicians to best treat and respect the patient.

Highly skilled data scientists are critical for building sophisticated ML models, but it is also important to have domain experts who understand how to think about the models and output. Imagine using a large dataset to develop an algorithm for predicting the survival rate of patients with cancer. The model identifies multiple factors that are associated with lowest survival rate. One of the factors turns out to be the number of visits a patient makes to the outpatient clinic. Naturally, patients who died would have had more outpatient visits as well as imaging and tests in the weeks or months preceding their deaths than those who were healthier. However, it would be incorrect to identify these factors as a risk factor for death in this patient population.

In some cases, an ML-plus-human approach works best. Often the output of a ML model is not definitive but has a probability associated with it. For instance, Avant-garde Health uses ML to place the hundreds of thousands of medical supplies and drugs used in hospitals into its product categorization system to make it easy for administrators and physicians to interpret. The ML model classifies many products with close to 100% confidence, but there are also many products where it may be right 80% or less of the time. Based on a combination of the dollar value of the products and the probability the ML model is right, Avant-garde will have an expert review the ML output and correct it where appropriate.

**Myth 3: Successful algorithms will be adopted and utilized.**

Unfortunately, many powerful algorithms are not adopted or utilized because they are not integrated into the workflow of potential users. One of the hospitals that we work at built an application to help physicians identify who is the right specialist to whom they should refer a patient with a particular problem. No one used it. Physicians were too busy to exit the EHR, open the app, enter information into it, and then return to using the EHR. Susan Devore, the CEO of Premier, an analytics company that serves health care providers, has noted that “the biggest gap in decision support tools is incorporating them into EHR workflow.”

There are many ways to use ML-derived decision support, including reference materials, order sets, plans of care, after-the-fact reporting, and alerts. Of these, alerts at the point of decision making are usually the most effective. Consider a successful application from Beth Israel Deaconess Medical Center (BIDMC) in Boston. Consents for surgery arrive in a variety of ways — on paper via mail, via fax, and via electronic transmission. Finding them can be a challenge. BIDMC created a ML application that automatically “reads” incoming faxes and files them into the right medical record, adding an alert to the pre-operative checklist. This saves 120 hours of staff time per month.
example concerns the timely hospital discharge of patients, which is critical for optimizing workflow, bed management, and revenues. BIDMC created an application that predicts the discharge date with a high degree of accuracy. Providing this information to clinicians and case managers helps get patients home at the right time, reducing unnecessary hospital days.

These successful examples, as well as many thousands of others within health care, give us great confidence in the ability of ML to meaningfully improve patient care and lower costs. The key is to be thoughtful about what types of problems ML is equipped to solve, who needs to be involved in developing the model and interpreting the output, and how to make it easy for people to utilize and act on the insights.

Derek A. Haas is the CEO of Avant-garde Health, a health care technology and analytics company that improves quality and profitability across the care continuum.

Eric C. Makhni, MD, is a sports medicine orthopedic surgeon at the Henry Ford Health System and a team physician for the Detroit Lions.

Joseph H. Schwab, MD, is the principle investigator of the SORG research group which uses machine learning to develop predictive models in surgery. He is also an associate professor of orthopaedic surgery at Harvard Medical School and the chief of spine surgery at Massachusetts General Hospital.

John D. Halamka, MD, is executive director of the Health Technology Exploration Center of Beth Israel Lahey Health and the International Healthcare Innovation Professor at Harvard Medical School.