



Can AI Fix Medical Records?

Digitized patient charts were supposed to revolutionize medical practice. Artificial intelligence could help unlock their potential

By Cassandra Willyard

A YOUNG MAN, let's call him Roger, arrives at the emergency department complaining of belly pain and nausea. A physical exam reveals that the pain is focused in the lower right portion of his abdomen. The doctor worries that it could be appendicitis. But by the time the imaging results come back, Roger is feeling better, and the scan shows that his appendix appears normal. The doctor turns to the computer to prescribe two medications, one for nausea and Tylenol for pain, before discharging him.

This is one of the fictitious scenarios presented to 55 physicians around the country as part of a study to look at the usability of electronic health records (EHRs). To prescribe medications, a doctor has to locate them in the EHR system. At one hospital a simple search for Tylenol brings up a list of more than 80 options. Roger is a 26-year-old man, but the list includes Tylenol for children and infants, as well as Tylenol for menstrual cramps. The doctor tries to winnow the list by typing the desired dose—500 milligrams—into the search window, but now she gets zero hits. So she returns to the main list and finally selects the 68th option—Tylenol Extra Strength (500 mg), the most commonly prescribed dose of Tylenol. What should have been a simple task has taken precious minutes and far more brainpower than it deserved. This is just one example of the countless agonizing frustrations that physicians deal with every day when they use EHRs.

These EHRs—digital versions of the paper charts in which doctors used to record patients' visits, laboratory results and other important medical information—were supposed to transform the practice of medicine. The Health Information Technology for Economic and Clinical Health (HITECH) Act, passed in 2009, has provided \$36 billion in financial incentives to drive hospitals and clinics to transition from paper charts to EHRs. Then president Barack Obama said the shift would “cut waste, eliminate red tape and reduce the need to repeat expensive medical tests.” He added that it would “save lives by reducing the deadly but preventable medical errors that pervade our health care system.”

When HITECH was adopted, 48 percent of physicians used EHRs. By 2017 that number had climbed to 85 percent, but the transformative power of EHRs has yet to be realized. Physicians complain about clunky interfaces and time-consuming data entry. Polls suggest that they spend more time interacting with a patient's file than with the actual patient. As a result, burnout is on the rise. Even Obama observed that the rollout did not go as planned. “It's proven to be harder than we expected,” he told Vox in 2017.

Yet EHRs do have the potential to deliver insights and efficiencies, according to physicians and data scientists. Artificial intelligence in the form of machine learning—which allows computers to identify patterns in data and draw conclusions on their own—might be able to help overcome the obstacles encountered with EHRs and unlock their potential for making predictions and improving patient care.

DIGITAL DEBACLE

IN 2016 the American Medical Association teamed up with MedStar Health, a health care organization that operates 10 hospitals in the Baltimore-Washington area, to examine the usability of two of the largest EHR systems, developed by Cerner, based in North Kansas City, Mo., and Epic, based in Verona, Wis., respectively. Together these two companies account for 54 percent of the acute care hospital market. The

team recruited emergency physicians at four hospitals and gave them fictitious patient data and six scenarios, including the one about Roger, who presented with what seemed like appendicitis. These scenarios asked the physicians to perform common duties such as prescribing medications and ordering tests. The researchers assessed how long it took the physicians to complete each task, how many clicks were required and how accurately they performed.

What they found was disheartening. The time and the number of clicks required varied widely from site to site and even between sites using the same system. And some tasks, such as tapering the dose of a steroid, proved exceptionally tricky across the board. Physicians had to manually calculate the taper doses, which took anywhere from two to three minutes and required 20 to 42 clicks. These design flaws were not benign. The physicians often made dosage mistakes. At one site the error rate reached 50 percent. “We've seen patients being harmed and even patients dying because of errors or issues that arise from usability of the system,” says Raj Ratwani, director of MedStar Health's National Center for Human Factors in Healthcare.

But clunky interfaces are just part of the problem with EHRs. Another stumbling block is that information still does not flow easily between providers. The system lacks “the ability to seamlessly and automatically deliver data when and where it is needed under a trusted network without political, technical, or financial blocking,” according to a 2018 report from the National Academy of Medicine. If a patient changes doctors, visits urgent care or moves across the country, her records might or might not follow. “Connected care is the goal; disconnected care is the reality,” the authors wrote.

In March 2018 the Harris Poll conducted an online survey on behalf of Stanford Medicine that examined physicians' attitudes about EHRs. The results were sobering. Doctors reported spending, on average, about half an hour on each patient. More than 60 percent of that time was spent interacting with the patient's EHR. Half of office-based primary care physicians think using an EHR actually diminishes their clinical effectiveness. Isaac Kohane, a computer scientist and chair of the department of biomedical informatics at Harvard Medical School, puts it bluntly: “Medical records suck.”

Yet despite the considerable drawbacks of existing EHR systems, most physicians agree that electronic records are a vast improvement over paper charts. Getting patients' data digitized means that they are now accessible for analysis using the power of AI. “There's huge potential to use artificial intelligence and machine learning to develop predictive models and better understand health outcomes,” Ratwani says. “I think that's absolutely the future.”

It is already happening to some extent. In 2015 Epic began offering its clients machine-learning models. To develop these models, computer scientists start with algorithms and train them using real-world examples with known outcomes. For example, if the goal is to predict which

“Health data is like crude oil. It is useless unless it is refined.” —Leo Anthony Celi, M.I.T. Laboratory for Computational Physiology

patients are at greatest risk of developing the life-threatening blood condition known as sepsis, which is caused by infection, the algorithm might incorporate data routinely collected in the intensive care unit, such as blood pressure, pulse and temperature. The better the data, the better the model will perform.

Epic now has a library of models that its customers can purchase. “We have over 300 organizations either running or implementing models from the library today,” says Seth Hain, director of analytics and machine learning at Epic. The company’s sepsis-prediction model, which scans patients’ information every 15 minutes and monitors more than 80 variables, is one of its most popular. The North Oaks Health System in Hammond, La., implemented the model in 2017. If a patient’s score reaches a certain threshold, the physicians receive a warning, which signals them to monitor the patient more closely and provide antibiotics if needed. Since the health system implemented the model, mortality caused by sepsis has fallen by 18 percent.

But building and implementing these kinds of models is trickier than it might first appear. Most rely solely on an EHR’s structured data—data that are collected and formatted in the same way. Those data might consist of a blood-pressure reading, lab results, a diagnosis or a drug allergy. But EHRs include a wide variety of unstructured data, too, such as a clinician’s notes about a visit, e-mails and x-ray images. “There is information there, but it’s really hard for a computer to extract it,” says Finale Doshi-Velez, a computer scientist at Harvard University. Ignoring this free text means losing valuable information, such as whether the patient has improved. “There isn’t really a code for doing better,” she says. Moreover, Ratwani points out that because of poor usability, data often end up in the wrong spot. For example, a strawberry allergy might end up documented in the clinical notes rather than being listed in the allergies box. In such cases, a model that looks for allergies only in the allergy section of the EHR “is built off of inaccurate data,” he adds. “That is probably one of the biggest challenges we’re facing right now.”

Leo Anthony Celi, an intensive care specialist and clinical research director at the Massachusetts Institute of Technology’s Laboratory for Computational Physiology, agrees. Most of the data found in EHRs are not ready to be fed into an algorithm. A massive amount of curation has to occur first. For example, say you want to design an algorithm to help patients in the intensive care unit avoid low blood glucose, a common problem. That sounds straightforward, Celi says. But it turns out that blood sugar is measured in different ways, with blood drawn from either a finger prick or a vein. Insulin is administered in different ways, too. When Celi and his colleagues examined all the data on insulin and blood sugar from patients at one hospital, “there were literally thousands of different ways they were entered in the EHR.” These data have to be manually sorted and clustered by type before one can even design an algorithm. “Health data is like crude oil,” Celi says. “It is useless unless it is refined.”

AN INTELLIGENT FIX

THE CURRENT PITFALLS of EHRs hamper efforts to use artificial intelligence to glean important insights, but AI might itself provide a possible solution. One of the main drawbacks of the existing EHR

systems, doctors say, is the time it takes to document a visit—everything from the patient’s complaint to the physician’s analysis and recommendation. Many physicians believe that much of the therapeutic value of a doctor visit is in the interactions, Kohane says. But EHRs have “literally taken the doctor from facing the patient to facing the computer.” Doctors have to type up their narrative of the visit, but they also enter much of the same information when they order lab tests, prescribe medications and enter billing codes, says Paul Brient, chief product officer at athenahealth, another EHR vendor. This kind of duplicate work contributes to physician frustration and burnout.

As a stopgap measure, some hospitals now have scribes sit in on appointments to document the visit while the physician interacts with the patient. But several companies are working on digital scribes, machine-learning algorithms that can take a conversation between a doctor and a patient, parse the text and use it to fill in the relevant information in the patient’s EHR.

Indeed, some such systems are already available. In 2017 Saykara, a Seattle-based start-up, launched a virtual assistant named Kara. The iOS app uses machine learning, voice recognition and language processing to capture conversations between patients and physicians and turn them into notes, diagnoses and orders in the EHR. Previous versions of the app required prompts from the physician—much like Apple’s Siri—but the current version can be put in “ambient mode,” in which it simply listens to the entire conversation and then selects the relevant information. EHRs turned physicians into data-entry clerks, Kohane says. But apps like Kara could serve as intelligent, knowledgeable co-workers. And Saykara is just one of a host of start-ups developing such tools. Athenahealth’s latest mobile app allows physicians to dictate their documentation. The app then translates that text into the appropriate billing and diagnostic codes. But “it’s not perfect by any stretch of the imagination,” Brient says. The physician still has to check for errors. The app does reduce the workload, however. The systems that Robert Wachter, chair of the department of medicine at the University of California, San Francisco, has seen are “probably not quite ready for prime time,” he says, but they should be in a couple of years.

Artificial intelligence might also help clinicians make better, more sophisticated decisions. “We think of the decision support in a computer system as an alert,” says Jacob Reider, a physician and CEO at Alliance for Better Health, a New York-based health care system that works to improve the health of communities. That alert might be a box that pops up to warn of a drug allergy. But a more sophisticated system might list the likelihood of a side effect with drug option A versus drug option B and provide a cost comparison. From a technological standpoint, developing such a feature is “no different from Amazon putting an advertisement or making you aware of a purchasing opportunity,” he says.

Wachter sees at least one encouraging sign that progress is coming. In the past few years the behemoths of the tech world—Google, Amazon, Microsoft—have developed a strong interest in health care. Google, for example, partnered with researchers from U.C.S.F., Stanford University and the University of Chicago to develop models aimed at predicting events relevant to hospitalized patients, such as mortality and unexpected readmission.

To deal with the messy data problem, the researchers first translated data from two EHR systems into a standardized format called Fast Healthcare Interoperability Resources, or FHIR (pronounced “fire”). Then, rather than hand-selecting a set of variables such as blood pressure and heart rate, they had the model read patients’ entire charts as they unfolded over time up until the point of hospitalization. The data unspooled into a total of 46,864,534,945 data points, including clinical notes. “What’s interesting about that approach is every single prediction uses the exact same data to make the prediction,” says Alvin Rajkomar, a physician and AI researcher at Google who led the effort. That element both simplifies data entry and enhances performance.

But the involvement of massive corporations also raises serious privacy concerns. In mid-November 2019 the *Wall Street Journal* reported that Google, through a partnership with Ascension, the country’s second-largest health care system, had gained access to the records of tens of millions of people without their knowledge or consent. The company planned to use the data to develop machine-learning tools to make it easier for doctors to access patient data.

This type of data sharing is not unprecedented or illegal. Tariq Shaukat, Google Cloud’s president of industry products and solutions, wrote that the data “cannot be used for any other purpose than for providing these services we’re offering under the agreement, and patient data cannot and will not be combined with any Google consumer data.” But those assurances did not stop the Department of Health and Human Services from opening an inquiry to determine whether Google/Ascension complied with Health Insurance Portability and Accountability Act regulations. As of press time, the inquiry was ongoing.

But privacy concerns should not halt the quest for better, smarter, more responsive electronic health records, according to Reider. There are ways to develop these systems that maintain privacy and security, he says.

Ultimately real transformation of medical practice may require an entirely new kind of EHR, one that is not simply a digital file folder. All the major EHRs are built on top of database-type architecture that is 20 to 30 years old, Reider observes. “It’s rows and columns of information.” He likens these systems to the software used to record inventory at a brick-and-mortar bookstore: “It would know which books it bought, and it would know which books it sold.” Now envision how Amazon uses algorithms to predict what a customer might buy tomorrow and to anticipate demand. “They’ve engineered their systems so that they can learn in this way, and then they can autonomously take action,” Reider says. Health care needs the same kind of transformative leap.

.....
Cassandra Willyard is a science writer based in Madison, Wis.

Wiring Minds

Successfully applying AI to biomedicine requires innovators trained in contrasting cultures

By Amit Kaushal and Russ B. Altman

From the popular press to the largest health care conferences, promises of artificial intelligence revolutionizing biomedicine are ubiquitous. It often seems as if we are on the cusp of AI systems that can remotely identify a person about to get sick, make a diagnosis (no doctor needed!), select a custom AI-designed pharmaceutical and deliver it to the patient just in time—in an AI-powered self-driving car, of course.

If indeed this is the future, we are far from reaching it. To be sure, the pace of change has been rapid. Deep learning—the fast-growing subfield of AI that enables machines to diagnose pneumonia from chest x-rays or predict health deterioration from medical records—was unfamiliar even to most computer scientists a decade ago. And we do not know what evolutionary or revolutionary advances will drive AI in the coming decades. What we do know is that the success of biomedical AI depends not just on developing the technology but also on developing the people behind it.

Translating algorithmic advances to biomedical breakthroughs requires critically considering both realms of knowledge and endeavor on many levels. What, for example, are the true capabilities of a new technology, and what is simply hype? What problems in biomedicine are most likely to benefit from emerging computational capabilities? And how do we go from an interesting biomedical application of a new technology to the implementation of systems that actually improve human health? These challenging, multifaceted questions will need to be answered by interdisciplinary teams. The teams will require experts in AI, experts in biology and medicine, and, most important, leaders who can motivate and guide individuals with such diverse talents.

Unlike some domains in which AI has been applied, in biomedicine the consequences of failure are weighty. For a social media company, an AI model that is ineffective at increasing ad clicks can be detected and rolled back the same day. When it comes to medicine, however, human lives are at stake. Inadequately informed uses of AI can lead to obvious harm, such as inaccurate diagnostic or therapeutic recommendations, but also to more insidious failures, such as an algorithm that gives racially biased recommendations because it was trained with subtly biased data. Given the complexities of biomedicine and the inscrutable nature of many AI algorithms, it might be years before such a flaw is uncovered. Group leaders—whether in academia, pharmaceutical laboratories or start-ups—must not only understand the technical and scientific issues but also anticipate and articulate the potential risks, benefits and implications of the projects they undertake.

We need men and women who can build AI systems in med-