Can machine learning algorithms predict lab values?

Charna Albert

February 2020—At Massachusetts General Hospital, machine learning is being used in the laboratories to build next-level clinical decision support, and in the latest phase, it's undergoing trial for use in predicting laboratory results.

"I think this is the new paradigm for cost-effective laboratory medicine. This is an important way we're going to change how we do business," says Anand Dighe, MD, PhD, who spoke about machine learning techniques for labs during a CAP19 presentation last fall and in a recent interview with CAP TODAY.

Dr. Baron

Dr. Dighe, director of clinical informatics and director of the core laboratory at MGH, has been working with other scientists and pathologists to make this vision a reality. He and colleague Jason Baron, MD, a pathologist and clinical informatician within the MGH core laboratory and an assistant professor of pathology at Harvard Medical School, enlisted the help of two computer scientists at Massachusetts Institute of Technology. Together they studied ways to use machine learning to predict laboratory values using the results from other lab tests in the patient's medical record (Luo Y, et al. *Am J Clin Pathol.* 2016;145[6]:778–788; Luo Y, et al. *J Am Med Inform Assoc.* 2018; 25[6]:645–653).

The collaboration with MIT was "particularly fruitful," Dr. Baron tells CAP TODAY, in integrating MGH clinical laboratory and clinical data science expertise with computer science from MIT. "Although many mature machine learning methods developed outside of health care were available for us to use, some were not well suited to clinical data." Existing prediction models required finesse to handle important nuances of clinical data, he says. "For example, no outpatient has a CBC every day. It's not like a stock market ticker." (Finance drove the development of some machine learning algorithms.)

"We had to figure out novel algorithms that could provide useful information, even in the face of the missing data that is so common with laboratory results." The development of these algorithms was a key contribution of their MIT collaborators Peter Szolovits, PhD, professor of computer science and engineering and head of the clinical decision-making group within the MIT computer science and artificial intelligence laboratory, and Yuan Luo, PhD, who is now chief AI scientist and associate professor of preventive medicine at Northwestern University Feinberg School of Medicine.

One target of their work was predicting ferritin results from other laboratory tests. The MIT researchers worked with Dr. Dighe, Dr. Baron, and colleagues to develop imputation algorithms—methods that allowed them to infer the missing lab test values needed to train the model. In stage one of the two-step process, they imputed the results for lab tests that hadn't been performed (other than ferritin). In stage two, they took the measured and imputed values for the predictor tests and used those, in addition to basic patient characteristics, to predict ferritin results.

"When looked at in isolation, ferritin values can lead to misdiagnosis. Ferritin often increases from inflammation, so non-iron-deficient patients undergoing inflammatory responses may have elevated ferritin levels. And normal ferritin values can obscure when a patient is in fact iron-deficient," Dr. Dighe says. One application of the ferritin

algorithm is to look for discrepant results. When predicted and measured ferritin don't agree, "that's almost always an important signal for us."

"In those cases," Dr. Baron says, "the obvious thing to do would be to append a comment to the test result warning the clinician, 'Don't rule out iron deficiency on the basis of a normal ferritin alone.'"

For now, implementation of the algorithm is on hold. "We didn't have an obvious strategy for implementing it within our existing information systems," Dr. Baron says.

Developing predictive models is only part of the solution, Dr. Dighe says. Many types of models will not be useful in improving patient care unless they are implemented as clinical decision support within existing workflows, processes, and health information systems, "and implementation can be challenging," he says. Dr. Dighe and colleagues implemented a relatively straightforward, rule-based interpretive comment intended to flag substantially increasing creatinine values that may indicate acute kidney injury (Baron JM, et al. *Am J Clin Pathol.* 2015;143[1]:42–49).

Dr. Dighe

This AKI flag "was much more difficult to implement than we would have guessed," Dr. Dighe says. Developing the flag required calculating a "baseline" creatinine for each patient and then flagging subsequent creatinine values that were increased from that baseline according to certain rules. "However, there was no straightforward way to calculate the baseline creatinine within the version of the lab information system we were using at the time. We had to develop a complex workaround."

The flagging rules provide a solution to the problem of overlooked AKI cases. While their current AKI flag identifies AKI only after the patient already has it, "the longer-term aim is to alert providers in advance that their patient is likely to develop AKI 24 hours or more into the future and perhaps even offer advice regarding actionable steps to take to reduce AKI risk," Dr. Dighe says. One tack the team is taking involves extending their imputation work to forecast creatinine values into the future. "If future creatinine values are expected to increase, that could be a sign of AKI to come," Dr. Dighe says.

The AKI algorithm was implemented at MGH more than five years ago and provider feedback has been positive, with changes in treatment and decision-making resulting from the AKI flagging. "What we found from subsequent surveys one of our hospitalist colleagues did," Dr. Dighe says, "was that more than 50 percent of clinicians had made a change in patient care based on the AKI flag."

"Luckily, our LIS team here is very creative and they were able to implement it," he says of the difficulty. "When you're doing analysis for a paper, you can do all kinds of wonderful things, but you sometimes find yourself limited by technology when you try to implement them."

It helped that the creatinine flag could be reduced to simple if/then rules and that acute kidney injury is a common health problem. "We had a lot of high-level clinical requests to make this go through," Dr. Baron says, noting that the AKI flag affects roughly 10 percent of MGH's inpatients. "As a result, we were willing to put a lot of resources in and spend a lot of IT time, and we had a lot of clinicians helping."

If the AKI algorithm had been based on an artificial neural network or a more complex model, Dr. Baron says, it would have been much more difficult to put into clinical practice at MGH.

Drs. Dighe and Baron collaborated recently with MGH colleagues Aliyah Sohani, MD, director of surgical pathology, and Lisa Zhang, MD, resident in anatomic and clinical pathology, to demonstrate the utility of machine learning models in predicting peripheral blood flow cytometry (PBFC) results, with the aim of optimizing the use of PBFC (Zhang ML, et al. *Am J Clin Pathol.* 2020;153[2]:235–242). Using decision tree and logistic regression models to analyze PBFC samples from MGH's clinical flow cytometry laboratory, the study's authors demonstrate that it's possible to predict PBFC results by looking at the patient's history of hematologic malignancy and CBC/differential parameters.

"The results of multiple pathology and lab tests tend to be associated," Dr. Baron says. "We're finding this is one of many examples where we can predict what a test result will be with some degree of accuracy before we even perform the test."

Dr. Dighe says information gleaned from machine learning can be more easily translated into clinical practice by transforming the machine learning model into a simplified rule-based approach. "It's not like you need some super computer connected to the EHR to run the algorithm," he says. "In many cases you can run the algorithm offline and then implement it using standard EHR tools." The PBFC study is a good example, he says. "You can use machine learning to come up with rules"— in this case, whether the patient has a history of hematologic malignancy, the percentage of neutrophils, and presence or absence of blast cells—and then use those rules to implement standard EHR clinical decision support.

Their study, which included 784 PBFC samples (from 744 patients) with a concurrent or recent CBC/diff order, found that the triaging strategies "could potentially defer 35 to 40 percent of all PBFC (with concurrent or recent CBC/diff)," they and their coauthors write, noting that the deferred tests would be expected to produce no clinically significant findings.

Deciding what to do with the rules comes with a host of practical considerations, Dr. Dighe notes. Laboratory workflow as well as technical, administrative, and economic factors come into play. Put another way, validating an algorithm for a research paper is one thing; implementing the results of that research in a clinical setting is another.

"It requires a whole different standard of clinical evidence," Dr. Baron says. "It requires working with clinicians and having reasonable evidence that this is something safe and good to do for patient care. We have to think not just about what we did for the research paper, but also a practical implementation strategy."

For one thing, the bulk of their machine learning research analyzes certain lab test results to make predictions about the results of other lab tests. But in the real world of the clinic, the predictor tests aren't always ordered first.

"With the flow cytometry project we're trying to decide if a physician should move forward with flow cytometry," Dr. Baron says. But the algorithm is predicated on knowing a patient's CBC value. If the physician orders a CBC and flow cytometry in parallel, the prediction algorithm won't work. This problem could potentially be solved with a reflex protocol, he says, where "we first perform a CBC, and then depending on the results of the CBC we would reflex to flow cytometry. Or we could say based on the CBC results that flow cytometry isn't needed for this patient."

Alert fatigue is another consideration when implementing decision support. It's a well-known concern in health care. "It's important if you're going to stop a provider's workflow," Dr. Dighe says, "that you do it only when absolutely necessary and helpful."

Making clear to clinicians that clinical decision support is based on carefully researched and validated rules is critical, he says. "It's important for these machine learning algorithms not to be complete black boxes all of the time. We want clinicians to change their behavior, so we have to explain why we're alerting them."

The clinical version of the flow cytometry algorithm is now in what Dr. Dighe calls "silent mode," a trial period during which the algorithm runs in the background while the system collects data about when an alert would have

fired, without triggering alerts to clinicians. “You need a system to test these things out and look into those patients to make sure if it would have fired that it would have been appropriate,” Dr. Dighe says.

With the movement toward algorithms that are ever more complex, Dr. Baron says, “we need to think about how we’re going to leverage native LIS or EHR functionality, or how we’re going to build systems that can easily interface with existing health information systems.”

“If we had the full toolbox” for the AKI alert, Dr. Dighe says, “we could implement a very complex imputation method and a prediction algorithm. We would be able to look at not just the last or baseline creatinine but the whole picture of the patient.” Those approaches, however, would be almost impossible to implement within the current generations of LIS, he says, because none of the major companies permit external calculation engines.

“It isn’t hopeless, though,” Dr. Dighe adds. “You could potentially have a data repository and an external request to a clinical decision support engine, have all your computation occur somewhere else, and then bring the results back into your lab system or EHR.” Some EHRs now permit native machine learning implementation; algorithms that determine readmission risk and perform sepsis scoring in real-time are examples. “That same approach can work for lab tests too.”

“I think it’s very encouraging and a sign of recognition of the value of machine learning that EHRs have begun to create machine learning modules within the EHR build,” he says.

LIS and EHR functionality aren’t the only obstacles, Dr. Baron notes. Administrative and economic barriers also play a large role. One solution, he says, is to find a scalable model for shared clinical decision support. “It’s hard for an individual hospital to justify the resources to push these over the goal line by itself,” he says. “Let’s say it would take $2 million to build out a highly robust machine-learning–based solution for flagging of AKI. If you could build a solution that could be plugged into hospitals all over the country, then it could easily justify a few million dollar investment.”

Standards and the application of standards, like LOINC, SNOMED, and ICD-10, are holding things back too, Dr. Dighe says. In many organizations, “they’re typically not well applied, so even the basics like identifying a lab test can be a challenge. Now that we’re aggregating all our lab results from many EHRs in the New England area, we can build decision support inclusive of the entirety of the patient’s record, but we first have to manually and carefully map virtually all of those tests together for the decision support to work.”

“You can make this wonderful model that can look at all these parameters,” he continues, “but if you can’t identify and use a CBC result from an external organization that was deposited into your EHR, then it’s not as useful.”

Then, too, there is the tension around data sharing, Dr. Baron says. “In general, technology companies themselves don’t have direct access to patient data, so they try to partner with academic and nonacademic centers to collaborate on projects and get data.” Working with companies may be a solution, he says, to help future patients and build scalable models for decision support. “If we’re going to make this a reality, we’re going to need to develop these collaborations between health systems and industry.”

Charna Albert is CAP TODAY associate contributing editor.