Newsbytes

Machine learning algorithm shows promise for predicting need for massive transfusion

March 2020—At trauma hospitals, simplicity is considered a virtue. That's why when Jansen Seheult, MD, and his colleagues decided to use machine learning to predict massive transfusion needs, they chose a decision tree algorithm. "It was easy to implement as if/then rules, and it didn't require computational resources to deploy," says Dr. Seheult, clinical assistant professor of pathology at the University of Pittsburgh Medical Center.

Dr. Seheult

Dr. Seheult and his colleagues reported on that algorithm in a 2019 proof-of-concept study (Seheult JN, et al. *Transfusion.* 2019;59[3]:953–964). For the study, they used data from the trauma patient registry at one of UPMC's level one trauma centers to train a supervised machine learning model to predict when patients admitted to the trauma unit would require massive transfusion. The model used data outcomes to train an algorithm to classify new cases.

Most traditional scoring systems for predicting the need for massive transfusion, such as the Assessment of Blood Consumption score and Trauma Associated Severe Hemorrhage score, use univariate analysis to determine which predictor variables are most associated with massive transfusion and then combine those variables using logistic regression, says Dr. Seheult, who is also laboratory medical director at Vitalant Coagulation Laboratory, in Pittsburgh. But when scoring systems rely on this method, "you can have significant interactions between pairs or groups of variables that are not accounted for," he explains. Furthermore, traditional scoring systems tend to be more successful at predicting which patients *won't* need massive transfusion than which ones will need it.

This issue, often called the "accuracy paradox," is a problem with traditional massive transfusion scoring systems, but it's also a trap encountered more generally when building a predictive model using imbalanced data. Because only about five percent of all civilian trauma patients require massive transfusion, the data are highly imbalanced, and it's especially critical to account for this imbalance, Dr. Seheult says.

Data imbalance was one factor that drove Dr. Seheult and his colleagues to choose a recursive partitioning decision tree algorithm to predict massive transfusion at UPMC. With a decision tree, data imbalances can be accounted for by weighting classes differently, says Michelle Stram, MD, clinical instructor in the Department of Forensic Medicine at New York University, who coauthored the *Transfusion* study during her residency training at UPMC. "You can tell the algorithm it's more important for you to identify the group that needs massive transfusion," she says. "You may end up with false-positives, but you won't be missing the people that need [massive transfusion]."

Dr. Stram

The authors of the *Transfusion* study generated two decision trees. The first, called MTPitt, was trained using clinical and demographic parameters, such as vital signs, immediately available upon admission to the emergency department. The second, MTPitt+Labs, was trained using those variables and additional laboratory data.

A training data set comprises a set of inputs used to train the model, explained Dr. Stram during a presentation on machine learning applications in transfusion medicine at the 2019 AABB annual meeting. These inputs "should be representative of the types of cases you intend the algorithm to predict." A validation data set, she added, "is a portion of the data set held back from training the model. This data is used to test the performance of the algorithm."

In the *Transfusion* study, 15 patients from the validation data set received massive transfusion, and the MTPitt decision tree misclassified only two of these cases as false-negatives. The authors also compared the performance of the decision trees with those of the Assessment of Blood Consumption and Trauma Associated Severe Hemorrhage scores, which misclassified four and five cases as false-negatives, respectively. The machine learning algorithm predicted massive transfusion needs with higher sensitivity and balanced accuracy than either of the traditional scoring systems, the authors report. After adding laboratory data, the algorithm achieved significantly higher specificity and balanced accuracy, but the sensitivity did not improve because the initial version of the algorithm predicted too few false-negative cases for further stratification.

Another reason to use a decision tree is interpretability, Dr. Stram says. A decision tree is "human readable" and fits with trauma hospital workflow. And the output from the algorithm could be posted as a flowchart in the emergency department or trauma bay, Dr. Seheult adds. "You only need one individual to determine if it's predicting [a patient will need] massive transfusion." Furthermore, it's easily deployed, he continues. Model implementation and interpretability "were our two primary concerns when we were trying to develop a new algorithm."

Yet developing a predictive machine learning algorithm requires dialing in the correct level of model complexity, Dr. Seheult says. "As you increase the complexity of the decision tree, you reduce the bias [when the model is too simple to represent real-world data] so it starts to fit the training data very well. But you also increase the variance," or how much the model differs when it's trained using another data set. When this happens, he adds, you run the risk of overfitting the training data—that is, the model becomes too complex and begins to pick up irrelevant characteristics, or noise, in the data.

"You don't want to train the tree [to the point that] it predicts every single case because that's going to result in overfitting," Dr. Seheult continues. He and his colleagues avoided overfitting by "pruning the tree," or removing peripheral branches at four splits. In other words, they prevented the decision tree from splitting into additional subpopulations after four splits because they determined that at that complexity parameter, the model achieved the greatest balanced accuracy.

"Working with incomplete data is one of the fundamental challenges of machine learning," Dr. Seheult says. Therefore, he and his coauthors chose predictor variables with a minimal amount of missing data to train the algorithm for the MTPitt prediction tool. "It's important to stress to laboratories that they should . . . adopt storage practices that try to retain as much of the granular detail of the data as possible." It's also important, he adds, to think about the pattern of missing data. "If it's missing at random, multiple imputation techniques can be used to come up with a predictive model of the missing data."

Data aggregation and storage are other common challenges. "Ideally, you want to come up with an algorithm that is robust to different data sets, using data from multiple centers and sites," he says, noting that the 2019 study queried data from only one trauma center. Training an algorithm with data from only one institution "makes it difficult for other institutions to generalize your algorithm."

Interoperability also comes into play, says Dr. Stram. Storing data in discrete fields and in a manner in which you know you're aggregating like data is a necessity. It depends on the scenario, she says, but factors such as reference range and laboratory instrument may be important for pooling data correctly.

The MTPitt prediction tool is "more of a blueprint for developing an ML model than the ideal state," says Dr. Seheult. "We made it simple for a reason, and that's to make it interpretable. The infrastructure for deploying [complex] algorithms is not there right now in most laboratory and blood bank information systems. [But] you could envision a system where you have a complex neural network deployed with the electronic medical record and it would flag a patient as high risk for needing massive transfusion support. That's the long-term goal for these algorithms." —*Charna Albert*

Xifin debuts latest version of RCM platform

Xifin recently introduced Xifin RPM 11, a revenue cycle management platform that features robust automation, enhanced portals, advanced analytics, and artificial intelligence options.

The software uses automation to automatically identify errors and clean dirty claims, reducing denials and improving reimbursement. It provides a robust workflow configuration at multiple levels, including facility, client, and payer, and automatically generates payer-specific appeal forms and appeal letters.

Other features of Xifin RPM 11 include a patient portal that allows consumers to determine insurance eligibility and estimate out-of-pocket costs and a physician portal for submitting supporting documentation and enhancing other forms of information exchange. The software also offers such billing and payment-collection options as paper invoice suppression, patient notification letters, and interactive statements.

"Xifin RPM 11 users can choose to extend the enterprise-grade business intelligence that comes standard with Xifin RPM with new advanced analytics options that offer additional scalability, flexibility, and speed," according to a press release from the company. "The AI capabilities paired with Xifin Business Intelligence provide financial insights faster, greater long-term trend analysis, and can pinpoint revenue cycle workflow efficacy opportunities."

[Xifin](), 866-999-4346

Orchard and Epic receive vendor honors from KLAS Enterprises

KLAS Enterprises recognized Orchard Software and Epic Systems as leaders in the laboratory marketplace in its 2020 Best in KLAS awards.

Orchard was named the category leader in the laboratory (small/ambulatory) division for its Orchard Harvest laboratory information system, while Epic was named the category leader in the laboratory (large hospital/integrated delivery network) division for its Epic Beaker LIS.

The "Best in KLAS 2020: Software & Services" report lists the top-performing health care information technology companies within various market segments based on feedback from health care providers.

A list of 2020 Best in KLAS awardees is available at www.j.mp/KLAS-honors.

Intermountain Healthcare and Cerner recommit to partnership

Cerner and the Salt Lake City-based nonprofit Intermountain Healthcare system have announced that they are extending the health technology and innovation agreement they formed in 2013.

Over the years, Intermountain has adopted a range of Cerner products, including the company's EHR and revenue cycle solutions, across its 24 hospitals and approximately 210 clinics as part of a joint effort to improve care delivery, standardize processes, and control costs.

"Through this [multiyear] expansion, Cerner will continue to increase the speed of innovation and help effectively address business and patient needs today and in the future," the company reported. "Intermountain will continue to provide expert guidance that helps direct developments to improve experiences for caregivers, system operators, and patients across the globe."

[Cerner,](#) 816-201-1024

Dr. Aller practices clinical informatics in Southern California. He can be reached at raller@usc.edu.