Model improves prediction of mortality risk in ICU patients

August 30 2018, by Rob Matheson

MIT researchers have developed a machine-learning model that groups patients into subpopulations by health status to better predict a patient's risk of dying during their stay in the ICU. This technique outperforms "global" mortality-prediction models and reveals performance disparities of those models across specific patient subpopulations. Credit: Massachusetts Institute of Technology

In intensive care units, where patients come in with a wide range of health conditions, triaging relies heavily on clinical judgment. ICU staff run numerous physiological tests, such as bloodwork and checking vital signs, to determine if patients are at immediate risk of dying if not treated aggressively.

Enter: machine learning. Numerous models have been developed in recent years to help predict patient mortality in the ICU, based on various health factors during their stay. These models, however, have performance drawbacks. One common type of "global" model is trained on a single large patient population. These might work well on average, but poorly on some patient subpopulations. On the other hand, another type of model analyzes different subpopulations—for instance, those grouped by similar conditions, patient ages, or hospital departments—but often have limited data for training and testing.

In a paper recently presented at the Proceedings of Knowledge Discovery and Data Mining conference, MIT researchers describe a machine-learning model that functions as the best of both worlds: It trains specifically on patient subpopulations, but also shares data across all subpopulations to get better predictions. In doing so, the model can better predict a patient's risk of mortality during their first two days in the ICU, compared to strictly global and other models.

The model first crunches physiological data in electronic health records of previously admitted ICU patients, some who had died during their stay. In doing so, it learns high predictors of mortality, such as low heart rate, high blood pressure, and various lab test results—high glucose levels and white blood cell count, among others—over the first few days and breaks the patients into subpopulations based on their health status. Given a new patient, the model can look at that patient's physiological data from the first 24 hours and, using what it's learned through analyzing those patient subpopulations, better estimate the likelihood

that the new patient will also die in the following 48 hours.

Moreover, the researchers found that evaluating (testing and validating) the model by specific subpopulations also highlights performance disparities of global models in predicting mortality across patient subpopulations. This is important information for developing models that can more accurately work with specific patients.

"ICUs are very high-bandwidth, with a lot of patients," says first author Harini Suresh, a graduate student in the Computer Science and Artificial Intelligence Laboratory (CSAIL). "It's important to figure out well ahead of time which patients are actually at risk and in more need of immediate attention."

Co-authors on the paper are CSAIL graduate student Jen Gong, and John Guttag, the Dugald C. Jackson Professor in Electrical Engineering.

Multitasking and patient subpopulations

A key innovation of the work is that, during training, the model separates patients into distinct subpopulations, which captures aspects of a patient's overall state of health and mortality risks. It does so by calculating a combination of physiological data, broken down by the hour. Physiological data include, for example, levels of glucose, potassium, and nitrogen, as well as heart rate, blood pH, oxygen saturation, and respiratory rate. Increases in blood pressure and potassium levels—a sign of a heart failure—may indicate health problems over other subpopulations.

Next, the model employs a multitasking method of learning to build predictive models. When the patients are broken into subpopulations, differently tuned models are assigned to each subpopulation. Each variant model can then more accurately make predictions for its

personalized group of patients. This approach also allows the model to share data across all subpopulations when it's making predictions. When given a new patient, it will match the patient's [physiological data](#) to all subpopulations, find the best fit, and then better estimate the mortality risk from there.

"We're using all the patient data and sharing information across populations where it's relevant," Suresh says. "In this way, we're able to … not suffer from data scarcity problems, while taking into account the differences between the different patient subpopulations."

"Patients admitted to the ICU often differ in why they're there and what their health status is like. Because of this, they'll be treated very differently," Gong adds. Clinical decision-making aids "should account for the heterogeneity of these patient populations … and make sure there is enough data for accurate predictions."

A key insight from this method, Gong says, came from using a multitasking approach to also evaluate a model's performance on specific subpopulations. Global models are often evaluated in overall performance, across entire patient populations. But the researchers' experiments showed these models actually underperform on subpopulations. The global model tested in the paper predicted mortality fairly accurately overall, but dropped several percentage points in accuracy when tested on individual subpopulations.

Such performance disparities are difficult to measure without evaluating by subpopulations, Gong says: "We want to evaluate how well our model does, not just on a whole cohort of patients, but also when we break it down for each cohort with different medical characteristics. That can help researchers in better predictive model training and evaluation."

Getting results

The researchers tested their model using data from the MIMIC Critical Care Database, which contains scores of data on heterogeneous patient populations. Of around 32,000 patients in the dataset, more than 2,200 died in the hospital. They used 80 percent of the dataset to train, and 20 percent to test the model.

In using data from the first 24 hours, the model clustered the patients into subpopulations with important clinical differences. Two subpopulations, for instance, contained patients with elevated blood pressure over the first several hours—but one decreased over time, while the other maintained the elevation throughout the day. This subpopulation had the highest mortality rate.

Using those subpopulations, the model predicted the mortality of the patients over the following 48 hours with high specificity and sensitivity, and various other metrics. The multitasking model significantly outperformed a global model by several percentage points.

Next, the researchers aim to use more data from electronic health records, such as treatments the patients are receiving. They also hope, in the future, to train the model to extract keywords from digitized clinical notes and other information.

More information: Learning Tasks for Multitask Learning: Heterogenous Patient Populations in the ICU: arxiv.org/pdf/1806.02878.pdf

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology