Accounting for diagnostic uncertainty when training a machine learning algorithm to detect patients with the Acute Respiratory Distress Syndrome

Narathip Reamaroon, Michael W. Sjoding, Kayvan Najarian
University of Michigan, Ann Arbor

When training a supervised machine learning algorithm for a clinical application, in many cases a classification label may be confidently assigned to each patient. However, in other applications, there may be less certainty when assigning labels to patients due to clinical ambiguity. For example, when determining if patients have a specific medical condition, there may be difficult cases where even medical experts may have low confidence in determining if the condition was present or absent. As a result, the assigned labels for these patients may not accurately reflect their underlying state. Inaccuracy in training data could adversely affect a machine learning algorithm's ability to detect the medical condition.

We implemented a support vector machine (SVM) algorithm that accounted for diagnostic uncertainty in the training data when detecting patients with the Acute Respiratory Distress Syndrome (ARDS). ARDS is a critical illness syndrome among patients in intensive care units with high mortality. However, patients with ARDS are not always recognized by clinicians and fail to receive evidence-based treatment. A machine learning algorithm using electronic health record (EHR) data could potentially help clinicians recognize and treat ARDS earlier, improving patient outcomes.

We trained and tested an algorithm for ARDS detection on 401 patients (2/3 training, 1/3 testing). Variables used in the model were chosen based on guidance from a clinical expert and included 24 vital signs and laboratory values, sampled every 2 hours. Previous values were carried forward until a new value was recorded, and missing data was imputed as normal. All patients were independently reviewed by 1 – 3 clinical experts to determine whether ARDS developed, the time of onset, and their diagnostic confidence on a scale of 1 – 4. This confidence was converted into a weighting that directly influenced the C-parameter in the linear SVM model during 5-fold cross validation, rescaling the cost of misclassification based on the uncertainty in the diagnosis. We also implemented a time-series sampling method, guided by the theory of mixing in stochastic processes, to reduce inter-correlation of data points sampled on the same patient in the training data. During testing, we found the SVM classifier that accounted for label uncertainty had a 9% improvement in performance compared to a standard SVM model (AUROC = 0.84 versus AUROC = 0.75). These results demonstrate how accounting for label uncertainty when training a machine-learning algorithm to detect a medical condition may improve training efficiency and model accuracy.