Generalizability of machine learning models for autoverification of mass spectrometry assays in the clinical laboratory

James H. Harrison, Jr., MD, PhD and Zacary Abushmaies
Department of Pathology, University of Virginia

Background

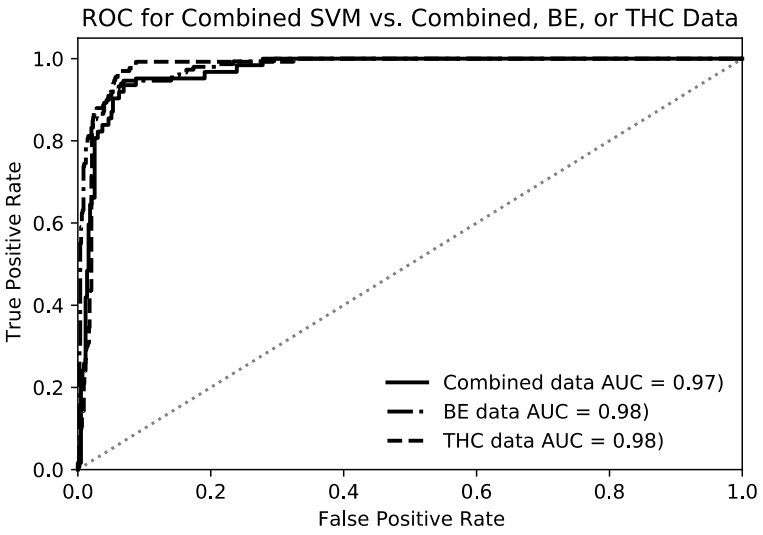
Generalizable machine learning models retain utility across contexts such as location or task. We previously reported a support vector model for autoverification of tetrahydrocannabinol (THC) by mass spectrometry. We now assess the generalizability of this type of model to other mass spectrometry assays using cocaine/benzoylecgonine (CBE) as a test case.

Methods

Retrospective data from routine urine THC (n=1267) and CBE (n=982) analyses were exported from Agilent instrumentation (Santa Clara, CA, USA). Samples repeated based on operator judgement were targets for identification by machine learning. Data analysis and machine learning used Python 3.8.3 (Anaconda, Inc., Austin, Texas, USA) with the Scikit-Learn library (v. 0.23.2). Run parameters were normalized to the middle calibrator, data was scaled to the interquartile range, and features were pruned based on correlation and importance ranking. The data for each assay were divided into training (67%) and test (33%) sets. The training data were concatenated to create an additional combined training/test data set (67%/33%). Separate support vector classifiers were built using each training set and predictive performance was evaluated within and between the test data sets.

Results

Ten percent of THC and 15% of BE assays were repeated. Six features were highest ranked and used in the analysis: sample height/area, sample peak shape, sample qualifier ion 1 & 2 ratios, internal standard peak area ratio, and internal standard qualifier ion 1 ratio. Recall, precision, and ROC AUC of the THC model for repeated assays in THC and CBE test data were 0.91/0.69/0.97 and 0.58/0.87/0.96, respectively. The BE model yielded 0.94/0.68/0.98 (CBE data) and 0.98/0.50/0.93 (THC data). The model trained on combined data (Figure 1) yielded recall, precision, and AUC values of 0.94/0.64/0.97 (combined), 0.98/0.59/0.98 (THC), and 0.90/0.79/0.98 (CBE).



Conclusions

The CBE-trained model was generalizable to THC data and the combined model was reasonably generalizable to both individual assays. The THC model was not generalizable to CBE data. These results suggest that common models may be possible for mass spectrometry autovalidation if they are carefully constructed and validated.