Automated Interpretation of Serum Protein Electrophoresis

Andrew E.O. Hughes[1]; Christopher W. Farnsworth[1]; Ann M. Gronowski[1]
[1]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO

Formatted: Left: 0.75", Right: 0.75", Top: 0.75", Bottom: 0.75"

Background: Serum protein electrophoresis (SPE) is a common laboratory test that plays a critical role in diagnosing and monitoring patients with clonal plasma cell disorders. SPE data consist of one-dimensional traces with characteristic peaks corresponding to defined populations of proteins. Interpreting these profiles requires manual inspection to differentiate normal versus abnormal patterns, which is often labor-intensive and subjective. To address this, we developed a machine learning model to predict the diagnostic labels assigned to SPE traces from clinical samples.

Methods: 6,737 traces from SPE performed at Barnes Jewish Hospital on a Sebia Capillarys 3 were used for model development. The seven target labels were: no apparent monoclonal peak, abnormal alpha-2, abnormal beta-1, abnormal/possible abnormal beta-2, and abnormal/possible abnormal gamma. For each trace, candidate peaks were identified, and 107 morphological features were extracted. Samples were split 80/20 into training/testing sets, and extracted features were used to train the following models on binary (normal vs. abnormal) and multiclass (specific label) classification tasks: k-nearest neighbors, penalized logistic regression, random forest, and gradient boosting machine. Hyperparameters were tuned using repeated cross-validation or Bayesian optimization. Area under the receiver operating characteristic curve (AUC-ROC) and precision recall curve (AUC-PR) were calculated on the test set. Data processing and modeling were implemented in R (v4.0.3) using tidymodels (v0.1.2).

Results: The best binary classification was obtained with logistic regression, with an AUC-ROC of 0.985 and AUC-PR of 0.993 (Table 1). At fixed sensitivities of 0.90, 0.95, or 0.99, the corresponding specificities of logistic regression were 0.97, 0.92, or 0.73, respectively. The best multiclass classification was obtained with the gradient boosting machine, with an AUC-ROC of 0.978 and AUC-PR of 0.895. Classification errors were predominantly traces with possible abnormal peaks that were predicted as normal (or vice versa).

Conclusions: The features used by laboratorians to interpret SPE traces can be readily extracted and used by standard machine learning models to accurately label SPE traces. In practice, these tools have the potential to reduce the time required for manual review and standardize interpretations across reviewers and laboratories. In addition, this approach is applicable to one-dimensional data produced by different instruments or for other applications.

Classification	Model	AUC-ROC	AUC-PR	Log Loss
Binary	K-nearest Neighbors	0.948	0.976	0.333
	Penalized Logistic Regression	0.985	0.993	0.152
	Random forest	0.981	0.991	0.170
	Gradient Boosting Machine	0.985	0.992	0.154
Multiclass	K-Nearest Neighbors	0.938	0.799	0.823
	Penalized Logistic Regression	0.972	0.847	0.381
	Random Forest	0.974	0.867	0.384
	Gradient boosting machine	0.978	0.895	0.314