B89 Forensic Body Fluid Identification and Differentiation by Raman Spectroscopy

Claire Muro, BS, University at Albany, Life Sciences Research, Bldg 01113, 1400 Washington Avenue, Albany, NY 12222; Kyle C. Doty, BS, 2165 Robinwood Avenue, Schenectady, NY 12306; Luciana de Souza Fernandes, BS, University at Albany, 1400 Washington Avenue, Albany, NY 12222; and Igor K. Lednev, PhD, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222*

The goal of this presentation is to educate attendees regarding how multivariate data analysis can be leveraged to extract information from Raman spectra and build predictive models.

This presentation will impact the forensic science community by offering a new approach for body fluid analysis, which, in many respects, is an improvement over the current methods.

The ability to identify body fluid traces at crime scenes and preserve any DNA present is critically important in forensic science. Identification can be difficult because many of the current techniques are specific to one body fluid, and typical biochemical methods are destructive, which prevents any further analysis. To develop a universal, confirmatory, non-destructive approach that can be used to differentiate and identify body fluids, the specificity of Raman spectroscopy was combined with the analytical power of statistical modeling.

Raman spectra were collected from 75 body fluid samples, including peripheral blood, saliva, semen, sweat, and vaginal fluid. After preprocessing the experimental spectra, the samples were split into calibration and validation datasets. Several chemometric analysis techniques were trained and tested to find the best model. These included Partial Least Squares Discriminant Analysis (PLSDA), Support Vector Machine Discriminant Analysis (SVMDA) modeling, and variable selection by interval PLSDA (iPLSDA) and Genetic Algorithm (GA). By exploring so many different combinations of classification algorithms and variable selection methods, this research was able to study patterns in the data, the effects of various modeling parameters, and to ultimately determine the most robust method for differentiation. All of the models were internally cross-validated during calibration and externally validated with a test dataset.

The first PLSDA model, which used the entire spectral range, misclassified 2.3% and 2.1% of the calibration and validation spectra, respectively. When an SVMDA model was trained by the same dataset, it misclassified 0.9% and 0.5% of the calibration and validation spectra, respectively. When iPLSDA was employed for variable selection, the rate of error in calibration predictions decreased; however, the error rate of validation predictions increased. Lastly, the dataset produced through variable selection by GA was used to train a final PLSDA and SVMDA model. The rate of misclassifications in the calibration dataset by the PLSDA model decreased to 2.2%, while the misclassifications in the validation dataset dropped to 1.7%, both slightly lower than the first PLSDA model. The final SVMDA model built on a dataset produced by GA performed the best. This model accurately predicted the identity of 99.9% of the spectra from the calibration dataset. More importantly, it correctly predicted the identity of 100% of the spectra in the external validation dataset.

All five body fluids were successfully discriminated by coupling Raman spectroscopy and chemometrics. This technique is reliable, non-destructive, and is not specific to one body fluid, offering substantial advantages over the current techniques used to identify body fluids.

Forensics, Body Fluids, Chemometrics