

1086-62-246

**T. Vance\*** (tia\_1\_vance@hotmail.com), Princess Anne, MD, **D. Pokrajac**, Dover, DE, **A. Lazarevic**, Hartford, CT, **N. Reljin**, Dover, DE, **N. Melikechi**, Dover, DE, **A. Marciano**, Dover, DE, and **Y. Markushin**, Dover, DE. *Data Mining for Complex Biochemistry Spectroscopy Data*.

In recent years, the spectroscopy community has increasingly been using various techniques for automatic computer assisted quantitative and qualitative evaluation of specimen based on spectroscopy data. Laser induced breakdown spectroscopy (LIBS) surges as a fast, versatile and powerful analytical technique with the ability to make remote measurements in field environments. We perform multi-class classification of LIBS data of four proteins: BSA (most abundant protein in blood plasma), Osteopontin, Leptin and IGF II (potential biomarkers for ovarian cancer). Principal Component Analysis (PCA) is applied on the data for feature extraction. Classification is performed using K-nearest neighbor, classification and regression trees, neural networks, support vector machines, and adaptive local hyperplane. Discrimination of potential cancer biomarkers vs. a common blood plasma protein can lead to the identification of elemental fingerprints of biological and chemical components that are vital in the early detection of cancer based on spectroscopy data. Our approach demonstrates that highly accurate automatic classification of complex protein samples is possible on LIBS data, using PCA with sufficiently large number of extracted features and appropriate classification techniques. (Received September 22, 2012)