REVOLUTIONARY DIAGNOSTICS – NONTRADITIONAL APPROACHES FOR DEVELOPING BREAKTHROUGH CAPABILITIES AGAINST EMERGING THREATS

Machine Learning Techniques For Threat Classification On Ion Mobility Spectroscopy: A Survey Of Methods

John Cooper Pacific Northwest National Laboratory **Prativa Hartnett** Pacific Northwest National Laboratory **Gabe Anderson** Pacific Northwest National Laboratory **Robert Finedore** Pacific Northwest National Laboratory **Samuel Dixon** Pacific Northwest National Laboratory **Carolyn Cramer** Pacific Northwest National Laboratory **Lauren Charles** Pacific Northwest National Laboratory

Portable Ion Mobility Spectroscopy (IMS) devices are often used in field operations to verify whether warfighters are in the presence of a chemical warfare agent (CWA). Because these detectors operate in complex environments, they often encounter overlapping chemical signatures and false alarms. In the era of 'big data', it is easy to overlook the pervasive, real-world challenges of small datasets. Many times, collecting large, labeled datasets can be costly, inefficient, and as is the case with warfighters in the field, harmful to collect in large amounts. The problem of accurate classification is further magnified when dealing with small datasets with imbalanced classes, such as those containing novel agents, and has a wide range of applications, from healthcare to environmental safety. In this work, we perform a variety of experiments with different machine learning architectures on a class imbalanced, IMS dataset of limited size. We use models with varying levels of complexity: Twin Convolutional Networks with Contrastive Loss, Autoencoders for anomaly detection, 2D CNNs, 1D CNNs, and Compact Convolutional Transformers to classify chemical threats in real time, and to serve as an early warning system for warfighters in the field. We also experiment with signal-to-image conversion in our data pre-processing to extract richer spatio-temporal features via recurrence plots. Lastly, we test the robustness of our most successful model architecture by only training it on ten samples per class (N = 110) and achieving excellent results.

Approved for public release; distribution is unlimited.