Machine Learning for Gamma-Ray Spectroscopy
Colin Bellinger, Shiven Sharma, Nathalie Japkowicz
Machine Learning for Defence and Security Group

Introduction
Health Canada has significant amounts of spectral data, that requires manual examination to identify anomalous radionuclides. This study examines the use and development of modern machine learning computer algorithms in order to alleviate the human effort.

Why Machine Learning?
Machine Learning automates identification and enables analysts to spend more time examining possible threats.

Challenges
Labelling
Multiple Classes
High Dimensionality
No Alarm Data
Short Sampling Period

Between
Within
Clustering + One-class Learning

Under-sampling (information loss)
Over-Sampling (overfitting)
OCC - Background
Multiclass - Minority classes

Algorithms Employed
The algorithms employed are the autoassociator, the Mahalanobis distance and the Variance in Angle Spectrum. We examine the rankings produced by these algorithms, expecting to find medical isotopes to be ranked at the top, followed by background. The data is split into rain and non-rain, since rain, being exceptionally distinct, will always rank at the top, whether it be background or medical isotopes.

Conclusions and Future Work
The initial rankings obtained show that, due to the data distribution, Mahalanobis distance offers the best rankings, outperforming the experts at Health Canada. In order to further refine the system, we will explore Time series analysis, feature selection, the automated separation of rain, and ensembles of learners.