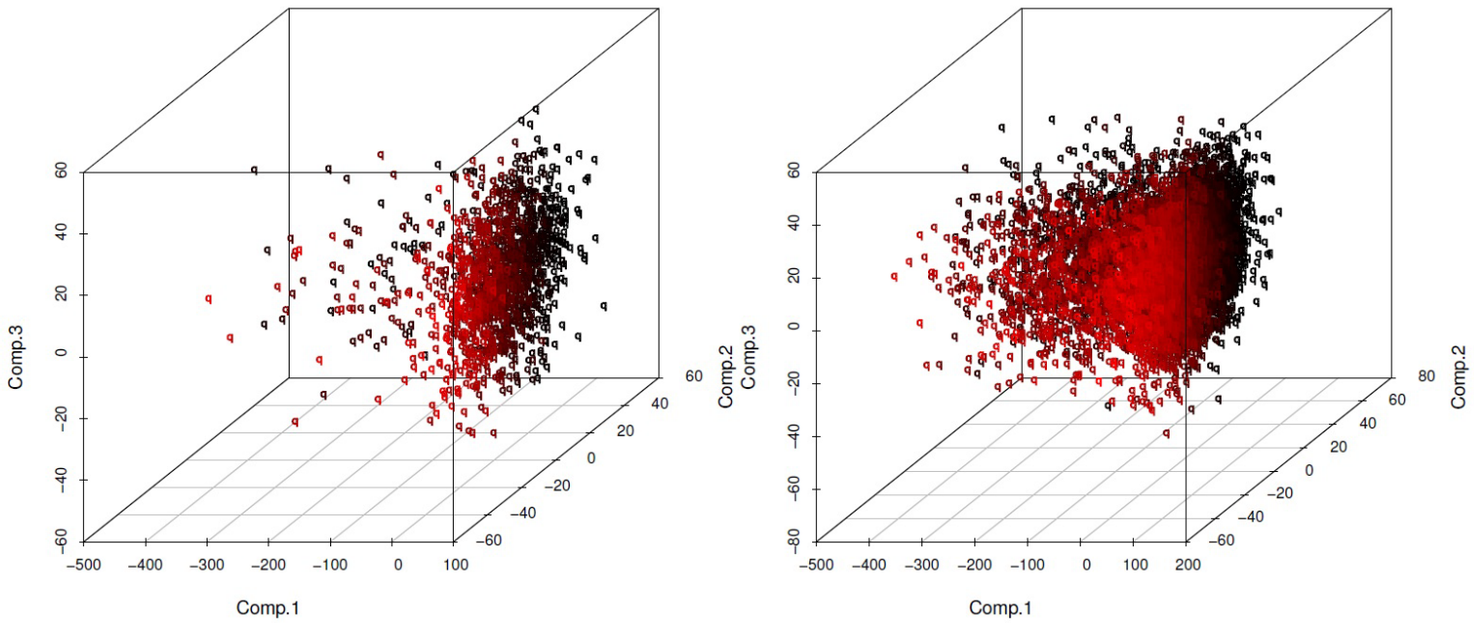
# Machine Learning for Gamma-Ray Spectroscopy

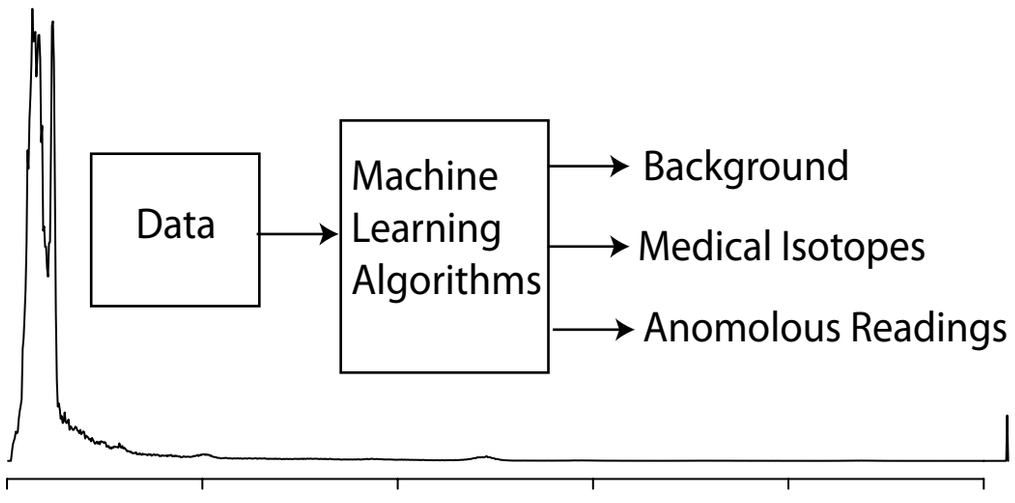
*Colin Bellinger, Shiven Sharma, Nathalie Japkowicz*
*Machine Learning for Defence and Security Group*

## Introduction

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



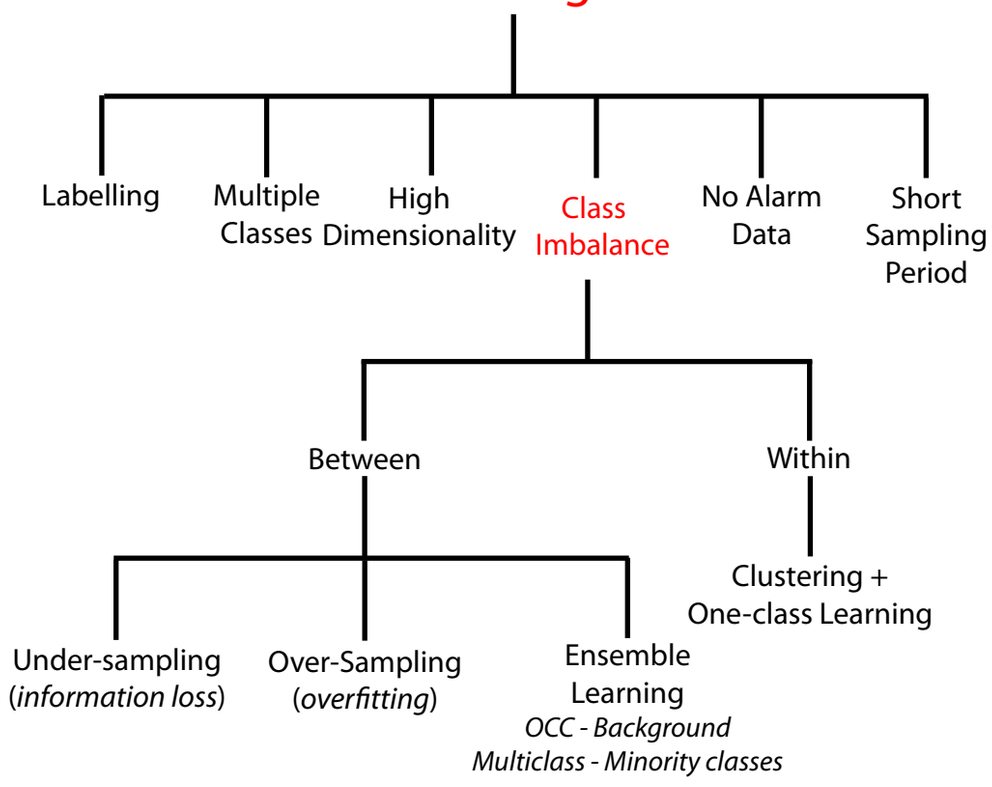

PCA Plots of first three principle components indicating separability for rain




## Why Machine Learning?

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

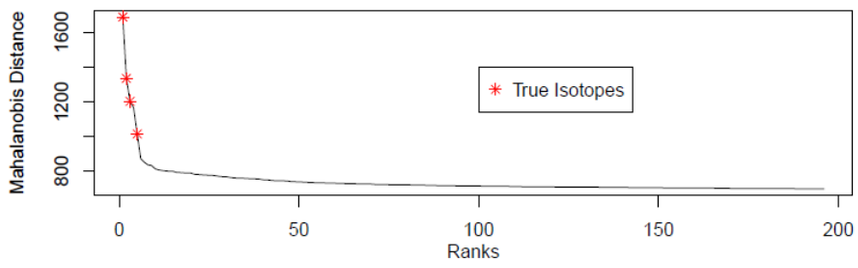
## Challenges



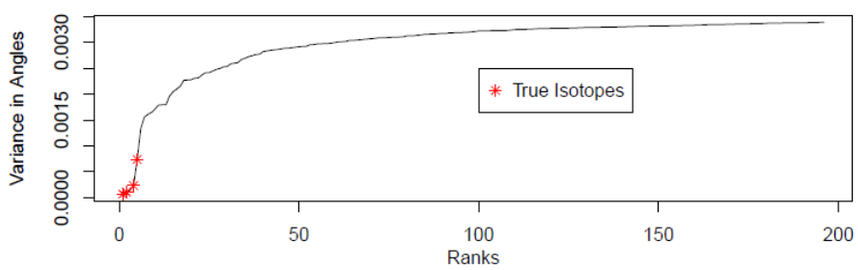

Health Canada deployed gamma-ray spectroscopes produced one reading per minute. The vast majority of these reperesent natural background noise. Less than 0.01% of the current data contains medical isotopes, and none are of the threat class.

## 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.




Progression of Mahalanobis Distance across ranks




Progression of Variance in Angles across ranks

## 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.