# Designing Gamma Spectrometry Data Augmentations for Contrastive Machine Learning

Jordan Stomps

Advisor: Paul P.H. Wilson

Laboratory Mentor: Ken Dayman

University of Wisconsin-Madison

stomps@wisc.edu

## Abstract:

The subject matter expertise and/or computational costs required to label enough data to enable cutting-edge machine learning models that currently dominate performance benchmarks can be prohibitive in many applications. This cost is further exacerbated by the high-impact nature of decisions in nuclear nonproliferation. One example is the use of gamma-ray spectrometry to detect and characterize movements of nuclear material in and around nuclear fuel cycle facilities. If the data collected from detectors distributed around the site can be effectively analyzed with machine learning algorithms to an adequate degree of confidence, such a system would greatly enhance the facility's safeguards and security programs with minimal additional burden on operators. Realizing such a capability requires nuclear fuel cycle expertise in relevant gamma-ray measurement technologies and associated material signatures. Therefore, data analytic solutions in nuclear security need to consider coupling models and expertise in a cost-effective manner. One method, contrastive self-supervised learning, learns patterns by perturbing data in ways that do not alter intrinsic labeling information. Subsequently, maximum classification agreement is enforced between pairs of samples, irrespective of augmentation type within reasonable bounds. Six non-exhaustive transformations are designed and implemented in this work for nuclear radiation data, with principles analogous to the augmentation typology used in image classification. These label-preserving transformations should incorporate gamma-ray spectroscopy physics informed by real-world collections to provide context beyond simulated or laboratory observations. A framework that employs transformations in this way could be generalized to active monitoring scenarios. The ideal result is a robust analytical model that reduces the burden on labeling training data while still utilizing the bulk of measurements taken.