Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

THE EFFECT OF INPUT FEATURES ON ADVANCED GAMMA-RAY SPECTRA ANALYSIS

Aaron Fjeldsted
Department of Nuclear Engineering
Pennsylvania State University

James Bevins
Department of Engineering Physics
Air Force Institute of Technology
Los Alamos National Laboratory

Daren Holland
Department of Engineering Physics
Air Force Institute of Technology

Azaree Lintereur
Department of Nuclear Engineering
Pennsylvania State University

ABSTRACT

Gamma-ray spectroscopy is a powerful, non-destructive analysis technique that can be used to obtain isotopic information from samples of interest. However, signatures associated with certain isotopes can be challenging to identify in complex spectra. Various algorithms and analysis techniques have been developed to obtain information from gamma-ray spectra, but often without generalizable applicability to dissimilar, or changing, measurement environments. Machine learning algorithms have proven successful at developing models that extract information and patterns from complex data. To support the development of a robust gamma-ray spectroscopy analysis method, six different supervised machine learning algorithms were applied to well-controlled simulated data sets. The algorithms' performance was evaluated, and preliminary feature selection studies were conducted to provide insight into the model identified regions of importance. This information was leveraged to develop engineered features, which can be used to guide future algorithm inputs.

INTRODUCTION AND BACKGROUND

Identification of radioactive isotopes is critically important to nonproliferation, arms control, and nuclear security applications. To this end, gamma-ray spectroscopy is a commonly used and non-destructive technique to obtain isotopic information from samples of interest by identifying unique energy signatures, as described by [1]. However, accurate identification of isotopes in complicated spectra can be time-intensive, and under some measurement conditions signatures may be obscured. Thus, the development of an automated and real-time isotope identification technique capable of extracting information difficult to acquire with traditional methods would advance current gamma-ray spectroscopy capabilities. A promising analysis method that has proven successful at extracting information from complex data is machine learning (ML) [2].

The computational capability of ML algorithms have lead to advancements in the medical, financial, and robotics sectors [3]. Recently, application of ML algorithms in nuclear science have shown promise for improved radioisotope identification. Both algorithm comparison [4, 5, 6] and development [7, 8, 2] for spectral analysis have been performed. However, much of the work in this field has focused on single source identification with no, or constant, background, and has relied on the full energy spectrum as the algorithm input [9, 10, 6]. While these studies have provided valuable insight into the capability of ML algorithms for spectroscopic analysis, this technique is still at early stages of implementation.

To advance this analysis method, more complex environments, such as multiple sources and varying background, need to be considered. Additionally, the benefits of feature selection and feature engineering have yet to be thoroughly explored. Feature selection techniques provide insight into the model-identified channels, or energies, which are important for correct data classification. This information can be leveraged to guide feature engineering, which could lead to improved isotope identification capabilities. This report presents a preliminary algorithm performance and feature importance analysis for six different supervised ML algorithms to provide insights into the performance of different algorithms under a variety of scenarios.

METHODS

The selected ML algorithms can be categorized as linear models, decision-tree models, an instance-based model, and a neural network. The training and testing data was created via simulations using the Monte Carlo based radiation transport code MCNP6.2 [11]. Simulations allowed sufficient data to be produced in a reasonable amount of time and enabled the variables to be carefully controlled.

Simulation Setup

A cylindrical high-purity germanium (HPGe) detector with a radius and height of 7.5 cm was located 5 cm above a terrestrial background disk source. The detector and sources were enveloped in a standard air-filled 1 m radius sphere (Figure 1). The Gaussian Energy Broadening (GEB) parameter was applied to a energy deposition tally (f8) and calibrated from [12] providing a resolution of 0.2% for the 661.7 keV ^{137}Cs photopeak. The simulated spectra consisted of 8000 channels covering the energy range of 0 – 3 MeV. Simple physics and non-cascade photon emission were simulated, meaning no bremsstrahlung, Doppler broadening, coincident summing, or coherent scattering events were included.

The terrestrial background source was comprised of the ^{238}U, ^{235}U, and ^{232}Th radioactive decay series and ^{40}K. The initial relative intensities of these sources were 0.096, 0.005, 0.082, 0.817, respectively [13]. For the decay series' progeny, only the prominent gamma-ray emitters were considered, as determined by [14]. Cosmic radiation was not included due to the low germanium cross section for high energy photons [15].

In addition to the background, two source isotopes were simulated, ^{60}Co and ^{137}Cs. The isotopes were modeled as point sources located 4.5 cm from the detector. The simulated source emission intensity, I_γ, for each decay gamma-ray was calculated by weighting each decay emission by its given branching ratio, and the results are shown in Table 1.

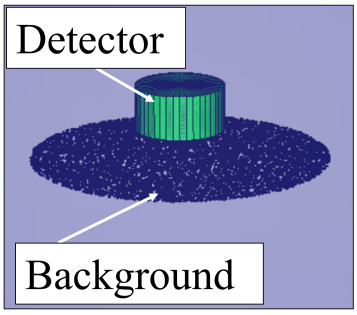


Figure 1. MCNP geometry utilized for simulations.

Table 1. Simulated radionuclides and the primary emission intensities.

Radionuclides	Gamma-ray Emissions[a]			
	Peak 1 (keV)	I_γ	Peak 2 (keV)	I_γ
^{60}Co	1173.2	49.97	1332.5	50.03
^{137}Cs	661.7	100	-	-

a - Data collected from the Evaluated Nuclear Structure Data Files (ENSDF) [16]

Data Sets

Data sets of varying complexity were generated by changing three different simulation parameters: the relative source contributions, the number of particles simulated, and the ratios of the decay series in the background source. Data sets were produced with each of the seven background (BG) plus source combinations (BG, ^{137}Cs, ^{60}Co, BG/^{137}Cs, BG/^{60}Co, ^{137}Cs/^{60}Co, BG/^{137}Cs/^{60}Co). First, each background plus source combination in the data set was simulated 150 times with randomly assigned relative contribution weights, resulting in a data set with 1050 spectra (7x150). Second, the ratios between the background decay series were either kept constant (uniform), or varied for each spectrum (random). Finally, data sets were produced with three different numbers of simulated particles: 10^7, 10^6, and 10^5. This resulted in six different data sets, each containing 1050 different spectra with varying source weights. For a simple scenario with only ^{137}Cs, the 661.7 keV channel had a relative uncertainty of 0.12% for 10^7, 0.37% for 10^6, and 1.18% for 10^5 histories, respectively; the relative uncertainty increased with the number of sources in the scenario and for lower count rate channels in the spectra.

Algorithms

Six supervised ML classification algorithms were used for this preliminary study, shown in Table 2. These algorithms can be grouped into three categories representing diverse approaches to classification: linear models, tree-based models, and a group of ”other” classifiers. The linear models utilize linear combinations of features to predict a value or class. Even though logistic regression utilizes a logistic activation function, it is considered a generalized linear model because the decision boundary of classification is a linear function. A support vector machine (SVM) optimizes the space between hyper-planes for classification.

Decision trees make a classification by learning rules from the input features. Random Forest (RF) and Extreme Gradient Boosting (XGB), in addition to utilizing decision trees, are both ensemble methods, which means they use multiple decision trees to make a classification. However, these algorithms differ in how they train their data. RFs utilize a bagged technique in which multiple decision trees are trained in parallel, while XGB operates with boosted trees, which are trained sequentially.

In addition to these four algorithms, an instance-based K-nearest neighbor (KNN) algorithm and a feed forward neural network (FFNN) were also implemented. These algorithms were selected to add conceptual diversity to the approaches being applied. KNN makes predictions based on the nearest data points to the target point, as defined by the Euclidean distance. Neural networks can learn non-linear functions for classification, and are trained using back-propagation, which modifies the weights given to input features to produce the desired output. Python's scikit-learn library was utilized for the implementation of the machine learning algorithms.

Table 2. List of algorithms used for classification.

Linear	Logistic Regression Support Vector Machine (SVM)
Decision Tree	Ensemble Random Forest (RF) Extreme Gradient Boosting (XGB)
Other	Feed Forward Neural Network (FFNN) K-Nearest Neighbors (KNN)

Testing Procedure

Multi-label classification, rather than multi-class classification, was used as this technique enabled greater insight into misclassifications. Multi-label classification is the task of assigning a set of labels, which are not mutually exclusive, to data points [17]. Prior to training and testing, algorithms without multi-label classification capabilities were converted to support this functionality.

For initial training and testing, the simulated energy spectra were used. By way of preprocessing, the raw pulse height tally (F8 tally) was normalized to a probability density function (PDF) and standardized, a common requirement for algorithms to improve sensitivity to variance. The data was then randomly partitioned with an 80/20 train/test split and fit to a model for classification. To reduce potential data selection bias effects, this was repeated five times and the average performance was recorded.The F1-score, which is a measure of a model's predictive capability, was used as a first evaluation of classification performance. Data analysis was performed using an Intel(R) Core(TM) i9-10980HK CPU @2.4 GHz with 32 GB RAM, as well as Penn State's Roar supercomputer.

RESULTS

The model performance results were analyzed to identify trends, and extract preliminary feature selection and engineering information to guide further development. The results shown in all figures and tables are for 10^7 simulated particles and uniform background with the logistic regression algorithm. Trends from the other data sets and algorithms are discussed throughout the text. It should be noted that for visualization purposes the training and testing repetition was increased from five to 100 for

Figures 2- 5.

Misclassification Analysis

Multi-label confusion matrices were examined for different algorithms and data sets to identify macroscopic trends. Figure 2 shows an example of three confusion matrices, produced with the logistic regression algorithm, which represent the model's average predictions for each label (BG, ^{137}Cs, ^{60}Co). For this representative data set, each of the three labels were present in four of the seven possible label combinations, and thus the results are weighed in favor of both false and true positives. All algorithms, with the exception of KNN, reported more false negatives than false positives despite this bias. For this work, a false negative is when a source is present but not identified. Conversely, a false positive is the identification of a source that is not present.

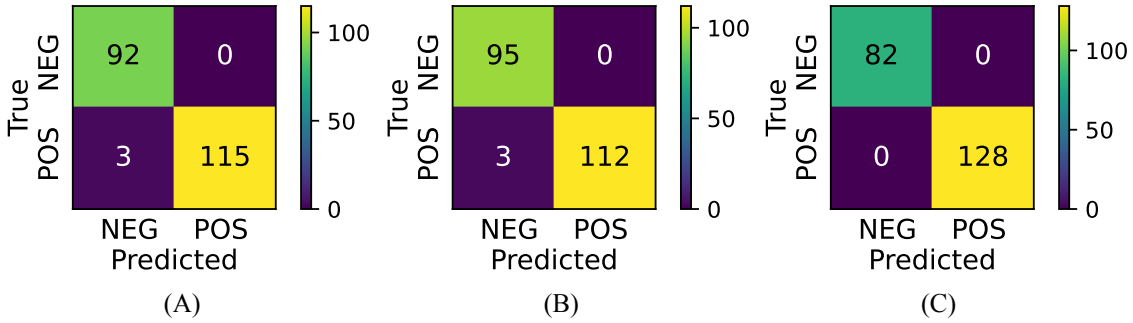


Figure 2. Confusion matrices for (A) ^{137}Cs, (B) ^{60}Co, and (C) BG.

To extract additional information, classifications were plotted as the fraction of ^{137}Cs or ^{60}Co's total contribution to the counts in the spectra, as shown in Figure 3. It can be seen that spectra are misclassified when there were low source contributions to the total counts. Misclassifications increased with spectra complexity, as specific signatures were more challenging to identify. Single source scenarios (BG only, ^{137}Cs only, ^{60}Co) were not included in this figure as they have a ratio of one for their contributions and zero for their misclassified spectra, which is not visible on the log scale.

A quantitative analysis of Figure 3 was performed for all of the algorithms in Table 3. This table compares the greatest source to total contribution ratio that was misclassified (Max Ratio) for each algorithm. Interestingly, points below the Max Ratio were occasionally identified correctly. To quantify this, the percent of correctly classified sources whose contributions fell below the Max Ratio was calculated. Table 3 shows as the Max Ratio decreased, so did the number of sources correctly classified below this threshold.

Table 3. The maximum source / total spectral contribution ratio misclassified (Max Ratio), and the percentage of sources correctly classified below the Max Ratio.

Metric	Logistic Regression	SVM	RF	XGB	KNN	FFNN
Max Ratio	0.021	0.016	0.019	0.012	0.139	0.278
Below Max Ratio (%)	1.02	0.98	1.04	0.75	7.46	20.37

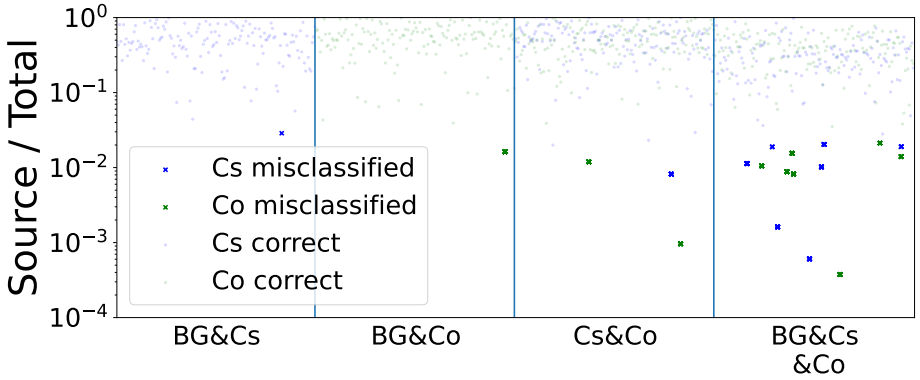
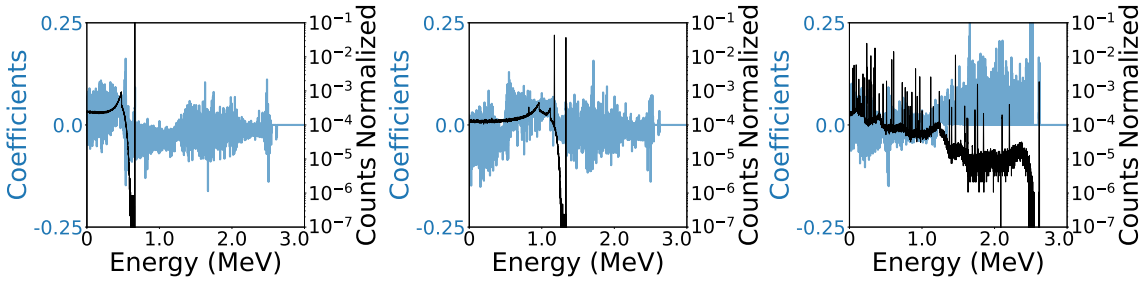


Figure 3. Misclassifications as a function of total source contribution to the spectra. The labels along the bottom of the plot show the combination simulated for those spectra. The y-axis represents the fraction of ^{137}Cs or ^{60}Co in the spectrum while the x-axis represents all of the data points.

Feature Selection

The misclassification analysis illustrated that the source contribution ratio had a significant impact on algorithm performance. To determine the spectral regions that influence algorithm performance, the importance of individual energy channels were considered. For this initial study, the coefficients of the linear models were leveraged to identify important channels, whereas the mean decrease in impurity was utilized for the decision tree models. Positive coefficients are associated with useful channels for source prediction, while negative coefficients represent the contrary.

An example of the coefficients for each label in the data set, overlaid on the corresponding spectrum, are shown in Figure 4 for the logistic regression model. The single-label binary classifier model was trained using data from the entire train/test split, i.e. training of the ^{60}Co classifier included data containing ^{137}Cs and BG sources. This was significant because it means the model was mutually inclusive.



(A) (B) (C)

Figure 4. Source spectra (A) ^{137}Cs, (B) ^{60}Co, and (C) BG overlaid on coefficients for the logistic regression algorithm.

Utilizing the coefficients from logistic regression or SVM models was an initial approach to feature selection, and shows that there are unique regions important for classification. As can be seen in

Figure 4, due to the mutually inclusive nature of the data, the model identifies coefficients that are unique to a specific source. For example, the Compton region for the ^{137}Cs source, the upper end of the ^{60}Co Compton region, and the high energy channels for the BG. This will be further explored in future work.

An example of the feature selection results for tree-based models, which carry multi-label functionality and therefore do not need to be cast into a binary relevance task as was done for the linear models, is shown in Figure 5 for the RF algorithm. In this example, the algorithm clearly identifies the ^{137}Cs photopeak and the two ^{60}Co photopeaks as important regions. Several high energy channels (1.4, 1.75, and 2.6 MeV) associated with the background were also identified as significant. Additionally, the regions corresponding to the ^{137}Cs and ^{60}Co Compton edges were important for classification. The trends observed in logistic regression and RF models were also seen in SVM and XGB models and for the other data sets. This analysis will be continued for the other algorithms and more complex spectra.

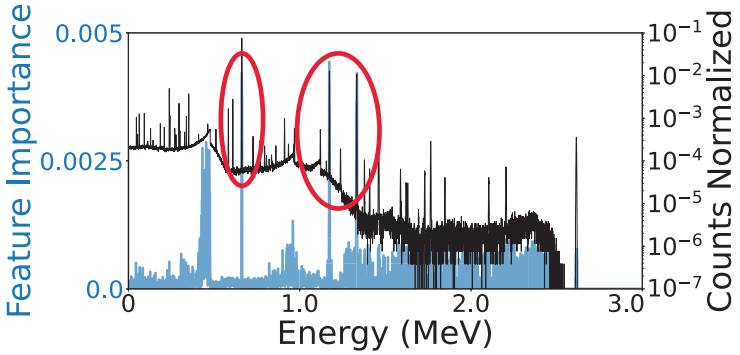
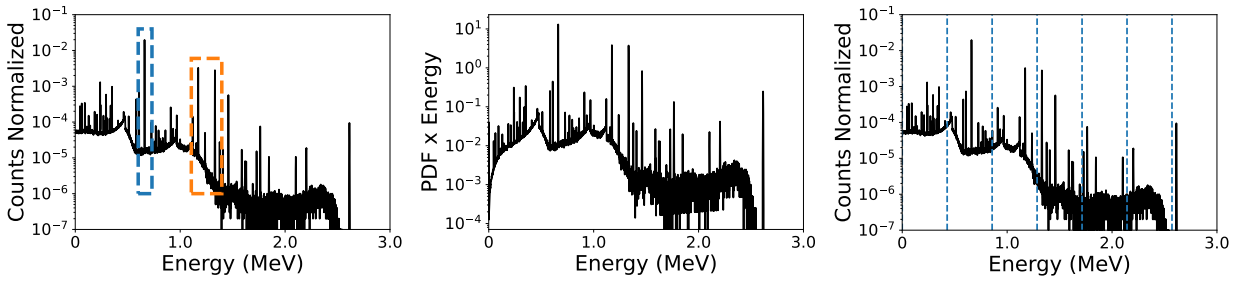


Figure 5. Feature selection (blue) performed with the RF algorithm. This specific spectrum (overlaid in black) consists of 0.29 ^{137}Cs, 0.18 ^{60}Co, and 0.53 BG

Feature Engineering

The misclassification analysis and feature selection results were leveraged to guide feature engineering. The coefficients and feature importances suggested that when multiple sources are present unique energy regions have a strong contribution to the algorithms' ability to correctly classify the spectra. Thus, three initial engineered features were selected to emphasize unique energy regions and explore the effects of reduced input dimensions. The three input features were photofractions, energy weighted spectra, and segmented spectra, as shown in Figure 6. The photofractions (ratio of photopeak area to the entire response function [18]) reduced the input dimensions, and provided features specific to the isotopes. However, this input did not highlight features from the higher energy BG regions. These regions were emphasized with the energy-weighted input, where the number of counts recorded for a specific channel was multiplied by the corresponding energy thereby emphasizing higher energy features. Finally, the segmented spectra feature provided a reduction in dimensionality while retaining some spectral information. For this feature the counts recorded in groups of 16 neighboring channels across the entire spectrum were averaged to reduce the dimensions from 8000 to 64. Sixty-four segments were utilized for this work as prominent spectral signatures could still be distinguished. Additional segments were also considered (2, 8, 64, 512, 1024, and 2048), and these results and

algorithm-specific optimization will be presented in future efforts.



(A) (B) (C)

Figure 6. Input features: (A) photofraction, (B) energy weighted, and (C) segmented spectra. The segmented spectra is being presented with seven segments for viewing purposes, but was implemented with 64 segments. These three spectra represent the same simulation in which there was 0.29 ^{137}Cs , 0.18 ^{60}Co, and 0.53 BG.

The F1 scores were used to provide initial information about the algorithms' performance with the different input features (Figure 7). The linear models (logistic regression, SVM) and the tree-based models (RF, XGB) performed well with the high-dimensional input features (normalized, energy weighted) but saw a slight decrease in performance with low-dimensional data (photofraction, segmented). The KNN and FFNN algorithms performed better with low-dimensional than high-dimensional input features. These insights will be leveraged to develop features for more complex data sets. For example, while some of the algorithms saw a decrease in performance with the low-dimensional inputs, as the number of sources increases, additional benefit may be obtained. Also, to leverage more of the unique regions, features which utilize both the photopeak and the Compton continuum will be considered.

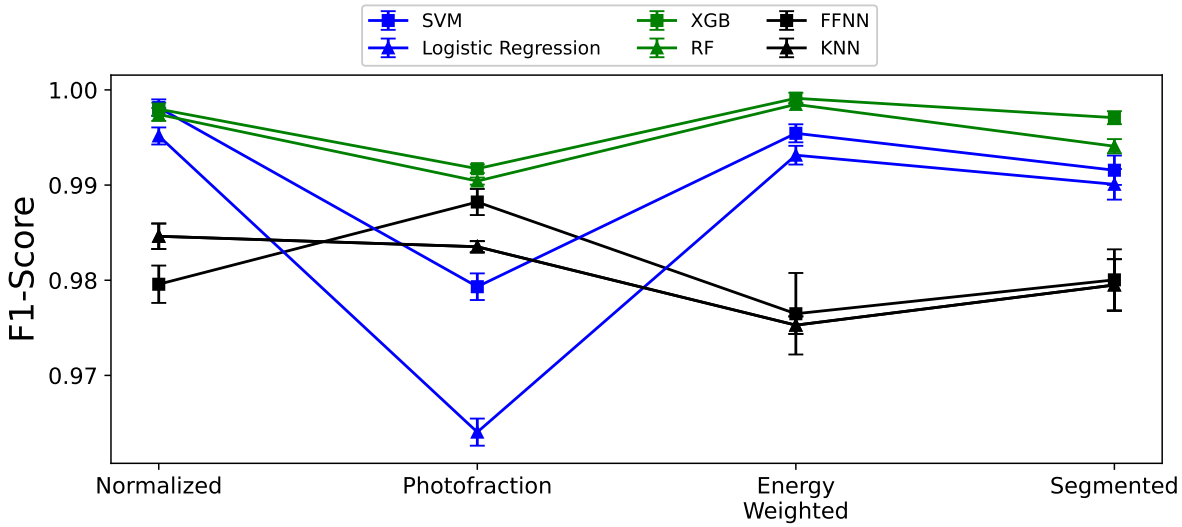


Figure 7. Comparison of the F1-score for the algorithms as a function of the various input features.

CONCLUSION

Six supervised ML algorithms were applied to the task of multi-label classification for isotope identification. The algorithms were trained and tested on simulated gamma-ray spectra with varying contributions of ^{137}Cs and/or ^{60}Co sources, varying backgrounds, and different numbers of simulated particles. A misclassification analysis showed that most incorrect labels were assigned to spectra with low source contributions (false negative). To gain additional insight into the features with the most importance, a preliminary feature selection search using both linear models (logistic regression, SVM) and tree-based models (RF, XGB) was performed. This feature selection led to the development of an initial set of engineered features, which considered different input dimensions and emphasized different spectral regions. While the F1-score for some of the algorithms decreased with fewer input dimensions, it remained relatively constant for others. This information will be used to guide feature engineering and algorithm selection for more complex data sets, including those with more sources. Additional future work will include further refinement of the feature-selection approach for multi-label classification, and engineering features that utilize more information from the regions identified as important by specific models.

ACKNOWLEDGMENTS

This research is sponsored by the Defense Threat Reduction Agency (DTRA) as part of the Interaction of Ionizing Radiation with Matter University Research Alliance (IIRM-URA) under contract number HDTRA1-20-2-0002.The content of the information does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

REFERENCES

[1] M. W. Rawool-Sullivan, J. A. Bounds, and L. Prasad, "Steps Toward Automated Gamma Ray Spectroscopy Steps Toward Automated Gamma Ray Spectroscopy: How a Spectroscopist Deciphers an Unknown Spectrum to Reveal the Radioactive Source Prototype Compton Imager (PCI) View project Species ID View project," no. June 2014, 2010. [Online]. Available: https://www.researchgate.net/publication/244395084

[2] S. Sharma, C. Bellinger *et al.*, "Anomaly detection in gamma ray spectra: A machine learning perspective," *2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications, CISDA 2012*, vol. 1, no. i, 2012.

[3] S. Das, A. Dey *et al.*, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect," *International Journal of Computer Applications*, vol. 115, no. 9, pp. 31–41, 2015.

[4] M. Kamuda, J. Zhao, and K. Huff, "A comparison of machine learning methods for automated gamma-ray spectroscopy," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, no. July, 2020.

[5] K. Peter, H. Ladislav, and B. Juraj, "Machine learning in radioactive nuclides identification," *SISY 2013 - IEEE 11th International Symposium on Intelligent Systems and Informatics, Proceedings*, pp. 57–61, 2013.

[6] M. Gomez-Fernandez, W. K. Wong *et al.*, "Isotope identification using deep learning: An explanation," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators,*

Spectrometers, Detectors and Associated Equipment, vol. 988, no. December, p. 164925, 2021. [Online]. Available: https://doi.org/10.1016/j.nima.2020.164925

[7] S. Jhung, S. Hur *et al.*, "A neural network approach for identification of gamma-ray spectrum obtained from silicon photomultipliers," 2 2020.

[8] H. Hata, K. Yokoyama *et al.*, "Application of support vector machine to rapid classification of uranium waste drums using low-resolution γ-ray spectra," *Applied Radiation and Isotopes*, vol. 104, pp. 143–146, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.apradiso.2015.06.030

[9] M. Kamuda, J. Stinnett, and C. J. Sullivan, "Using Artificial Neural Networks," vol. 64, no. 7, pp. 1858–1864, 2017.

[10] B. T. Koo, H. C. Lee *et al.*, "Development of a radionuclide identification algorithm based on a convolutional neural network for radiation portal monitoring system," *Radiation Physics and Chemistry*, vol. 180, no. May 2020, p. 109300, 2021. [Online]. Available: https://doi.org/10.1016/j.radphyschem.2020.109300

[11] C. J. Werner, J. Armstrong *et al.*, "MCNP User's Manual Code Version 6.2," *Los Alamos National Laboratory*, p. 746, 2017.

[12] E. Eftekhari Zadeh, S. A. H. Feghhi *et al.*, "Gaussian Energy Broadening Function of an HPGe Detector in the Range of 40 keV to 1.46 MeV," *Journal of Experimental Physics*, vol. 2014, pp. 1–4, 2014.

[13] F. Radiological, "MONITORING AND ASSESSMENT CENTER FRMAC Gamma Spectroscopist Knowledge Guide," no. August, 2019.

[14] G. Skinner, *Practical gamma-ray spectrometry*, 2nd ed. John Wiley & Sons, Inc., 1996, vol. 52, no. 3.

[15] M. J. Berger, J. H. Hubbell *et al.*, "XCOM: Photon Cross Sections Database," *NIST Standard Reference Database 8 (XGAM)*, vol. Version 1., 2010. [Online]. Available: https://www.nist.gov/pml/xcom-photon-cross-sections-database

[16] "Evaluated Nuclear Structure Data File." [Online]. Available: https://www.nndc.bnl.gov/ensdf/

[17] G. Doquire and M. Verleysen, "Feature selection for multi-label classification problems," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6691 LNCS, no. PART 1, pp. 9–16, 2011.

[18] G. Knoll, *Radiation Detection and Measurement Fourth Edition*. John Wiley & Sons, Inc., 2010.