https://doi.org/10.3011/ESARDA.IJNSNP.2023.2

List Mode Inference Using Linear Classifiers for Nuclear Arms Control Verification

Eduardo Padilla [1,2], **Heidi Komkov** [2], **Christopher Siefert** [2], **Adam Hecht** [1], **Ryan Kamm** [2], **Kyle Weinfurther** [2], **Jesus Valencia** [2]

[1] University of New Mexico, Department of Nuclear Engineering, Albuquerque, NM, USA, 87131

[2] Sandia National Laboratories, PO Box 5800, Albuquerque, NM, USA, 87185

Abstract:

In potential future nuclear arms control treaties, methods to confirm the presence or absence of a nuclear warhead or nuclear components are likely to be a central function of a verification regime. Higher confidence in verification methods can be achieved through more rigorous, thus potentially sensitive, analysis of radiation signatures from treaty accountable items. Therefore, methods that protect sensitive information while allowing for rigorous analysis are a critical component of any potential nuclear treaty verification system; these methods are referred to as information barriers. In this paper, we describe the development of a novel radiation analysis method for list-mode (time-stamped pulse heights, pulse-by-pulse) inference using linear classifiers trained on a large set of synthetically generated high resolution gamma spectra. In practice, each detector pulse would be fed into a linear classifier with the applied weight incrementing or decrementing counters for each class. After a set number of pulses, the highest output score determines the classification of the source of radiation. As such, this method serves as both a verification algorithm and information barrier combined. This new method achieves reliable discrimination (83% accuracy) of notional nuclear weapons grade treaty accountable item radiation signatures from those of a diverse, largely unconstrained, set of nuclear, medical and industrial radioisotope combinations. Importantly, this is shown to be achievable without the collection or processing of a potentially sensitive gamma radiation spectrum. This study serves as a proof of concept for the development of an intrinsic information barrier for attribute identification supporting nuclear arms control treaty verification.

Keywords: Information Barrier, Warhead Verification, Machine Learning.

1. Introduction

Potential future nuclear arms control treaties are likely to require much more rigorous and intrusive measures for verification as nuclear weapons states move beyond current absence verification methods such as those employed in New START (Evans, 2021). As such, methods more advanced than neutron detection above a threshold to confirm the presence or absence of a nuclear warhead or component are likely to be a central function of a verification regime. While standard methods for nuclear assay such as gamma ray spectroscopy and/or neutron measurements are largely capable of performing this function, the amount of information revealed during the analysis is likely too high for a nuclear arms control regime; nuclear warhead design information can be inferred from these measurements, thus potentially disclosing sensitive strategic information to a treaty partner. Hence, a critical need for nuclear arms control verification is a method that produces high-confidence assessments without revealing sensitive information such as can be inferred from gamma ray spectra or detailed neutron signatures.

1.1 Background

There are numerous approaches to protecting sensitive information, also referred to as information barriers (IBs). Figure 1 illustrates several types of information barriers which can be used within a verification process to act as information reducers and prevent the passage of sensitive information (red side) while allowing a reduced or transformed subset of non-sensitive information to proceed (black side). The approach described in this paper is a form of an intrinsic information barrier, wherein the pulses from a gamma and/or neutron detector are analyzed immediately and individually, without the creation or storage of accumulated spectra, dose rates, or other potentially sensitive information. This type of information barrier is complementary to other approaches, such as physical encryption, zero knowledge protocols (ZKP) and electronic information barriers.

Yan and Glaser (Yan & Glaser, 2015) provide a comprehensive review of past warhead verification systems incorporating several types of information barriers. Additional systems (Hamel, 2018) (White, 2012) (Wolford & White, 2000) have been included for background consideration in this

Padilla, E., Komkov, H., Siefert, C., Hecht, A., Kamm, R., Weinfurther, K., & Valencia, J. (2023, December). List Mode Inference Using Linear Classifiers for Nuclear Arms Control Verification, ESARDA Bulletin - The International Journal of Nuclear Safeguards and Non-proliferation, 65, 10-18. https://doi.org/10.3011/ESARDA.IJNSNP.2023.2

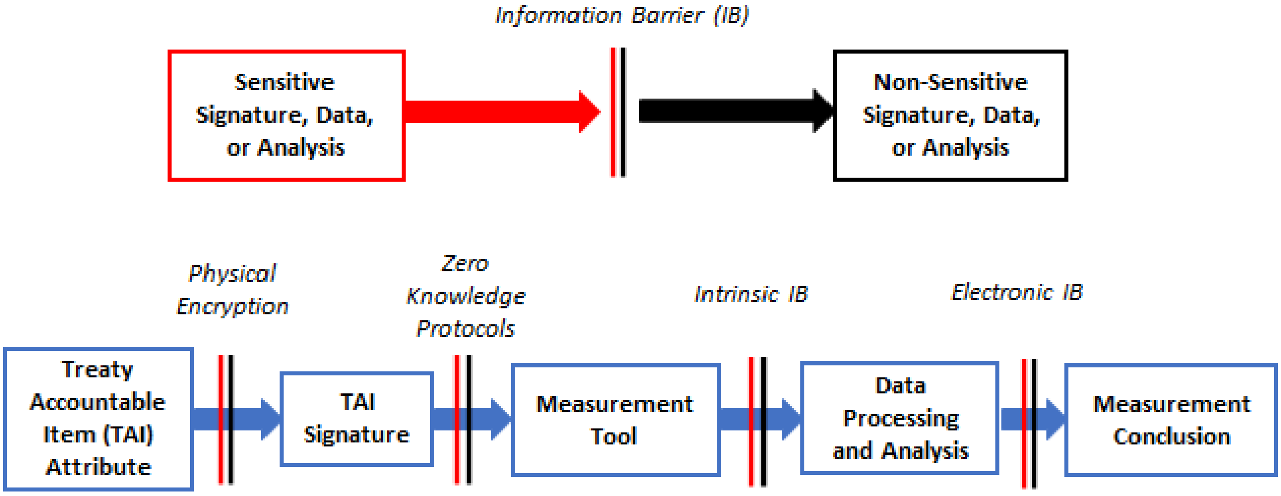
Information Barrier (IB)
Sensitive Signature, Data, or Analysis
Non-Sensitive Signature, Data, or Analysis
Physical Encryption
Zero Knowledge Protocols
Intrinsic IB
Electronic IB
Treaty Accountable Item (TAI) Attribute
TAI Signature
Measurement Tool
Data Processing and Analysis
Measurement Conclusion

Figure 1 – Information Barriers within a Verification Process

paper and a subset are summarized in Table 1, categorizing them according to the proposed conventions in Figure 1.

Table 1 categorizes historically-developed arms control verification concepts and systems by three primary design aspects: 1 – template or attribute verification approach, 2 – active or passive measurement and 3 – the type of information barrier employed. Each of these design aspects have associated strengths and weaknesses and cannot be truly evaluated independent of a well-defined treaty regime. For example, plutonium absence verification can reasonably be expected to be achievable using a much simpler method (gross neutron counting (Harahan, 1993)) than neutron tomographic imaging combined with ultra-high-resolution gamma spectroscopy; the simplest proposed method of performing a specific treaty verification task has a higher likelihood of negotiated implementation.

Compared to attribute verification systems, template verification systems are often considered easier to implement, since these can be designed to be performed behind an information barrier, sometimes requiring little to no a priori knowledge about the treaty accountable item (TAI). All that matters is that a TAI matches a measurement to a reference TAI or “golden copy”. The crux of template-based verification systems (TRIS, NMIS, CIVET, CONFIDANTE,

System	Description (Template/Attribute)	Active Interrogation	Information Barrier
TRIS (Seager, et al., 2001)	Low resolution gamma spectrum **template**	Passive	Electronic IB
TRADS (Mitchell & Tolk, 2000)	HPGe-based Pu **attribute** measurement (minimum mass and enrichment)	Passive	Electronic IB
(F)NMIS (Hamel, 2018)	Fast neutron **template** imaging	Active	N/A
AVNG (Langner, et al., 2002)	Neutron Multiplicity and HPGe-based **attribute** measurement	Passive	Electronic IB
3G-AMS (Dale, et al., 2009)	HPGe and Neutron slab detector based **attribute** measurement	Passive	Electronic IB
UKNI (Chambers, et al., 2010)	HPGe-based plutonium **attribute** measurement	Passive	Electronic IB
INPC (Hamel, 2018)	HPGe-based **attribute** measurement	Passive	Electronic IB
CIVET (Vanier, et al., 2001)	HPGe-based gamma spectrum **template**	Passive	Electronic IB
CONFIDANTE (Marleau & Krentz-Wee, 2020)	Fast neutron coded aperture **template**	Passive	ZKP
Princeton ZKP (Glaser, Barak, & Goldston, 2014)	Neutron radiography **template**	Active	ZKP
Princeton/MIT (Hecla & Danagoulian, 2018) (Engel & Danagoulian, 2019)	Nuclear resonance **template**	Active	Physical Encryption/ZKP

Table 1 – Summary of Previous Arms Control Verification Systems

Princeton ZKP, Princeton/MIT), then, is the authenticity of the golden copy, and how the inspecting party can attain confidence in the item presented as a golden copy. Owing to the difficulty of certifying a golden copy, this huge consideration is often deferred as part of future work.

In an attempt to address the golden copy obstacle, Hecla and Danagoulian (Hecla & Danagoulian, 2018) propose a method by which a golden copy warhead is selected at random and with minimal notice from a fielded system. Even this approach has many potential pitfalls, as this approach could only work for ground-based ICBMs subject to overhead imagery and persistent monitoring; submarine/ship launched warheads, as well as the myriad bombs, cruise missiles and tactical nuclear munitions are more easily moved and not subject to persistent monitoring by design. During the Intermediate Range Nuclear Forces Treaty (INF), inspection notices gave up to six hours of time to the host country to allow an inspection (Harahan, 1993), ample time for golden copy spoofs to be emplaced. Further, the method proposed by Hecla and Danagoulian was to scan only the pit of a warhead, due to the possibility of neutron and x-ray/gamma ray shielding materials being present in a fully assembled nuclear weapon. The disassembly of a nuclear weapon is a highly sensitive operation and would need to be performed in private, thus allowing the host country to modify (e.g., smash, shield, or otherwise obfuscate the true form and signature) the pit before placing in a black box for subsequent golden copy template generation. Combined with undisclosed and host-controlled anti-mask templates, a flattened and shielded pit used as a golden copy could then allow for simple spoofing of warhead dismantlement.

This inherent difficulty in golden copy certification demonstrates the value of attribute verification systems. Instead of blindly comparing two items, these systems seek to verify one or multiple signatures consistent with various characteristics (attributes) of a warhead, such as the presence of weapons grade nuclear material, certain isotopic ratios, geometric extent of intrinsically radioactive material, minimum mass of fissile material, etc. To achieve this, attribute verification systems generally require the measurement and analysis of more sensitive information, such as gamma spectra, neutron multiplicity and/or radiographic imaging. Therefore, this analysis is generally performed behind an electronic information barrier to protect against the release of sensitive measurement data to an inspector.

While template verification systems can more easily limit the generation of sensitive information, they are completely reliant on the veracity of the golden copy template, creating a single point failure. On the other hand, attribute verification systems can be designed to confirm the veracity of TAIs (or even golden copies themselves), while relying on electronic information barriers and more rigorous authentication and certification needs. Depending on the specific treaty regime and agreed upon implementation protocols, either a template, attribute or combined approach may be most effective.

When considering passive measurements versus active interrogation, the simplest proposed solution to address the needs of the verification regime is more likely to result in successful implementation negotiations, as seen in INF negotiations (Harahan, 1993). As nuclear arms control reduction treaties progress from New START-like treaty regimes (absence verification), more intrusive inspection approaches are likely to be necessary. If the nature of nuclear arms control treaties follows a progressive track towards complete global nuclear disarmament, solutions spanning multiple levels of intrusiveness and complexity will be required. It follows that the complexity of system hardware is directly proportional to the level of intrusiveness of the inspection technology, and also to the difficulty of performing authentication and certification on inspection equipment. Active interrogation systems will need to be authenticated by the inspection team and certified by the hosts as whole, meaning additional effort for developing trust in imaging sources (linacs, nuclear reactors, x-ray generators, etc.) will have additive effort and the potential for reduced trust as they introduce more attack vectors (each piece of hardware must be authenticated and certified down to individual electronic components. (Greenberg, 2019))

When designing information barriers, having the IB further to the left (Figure 1) lowers the number of potential vectors for sensitive host information exfiltration. Once the sensitive information is stripped out it cannot be regenerated. Thus, from a host perspective, pushing the IB as far to the left as

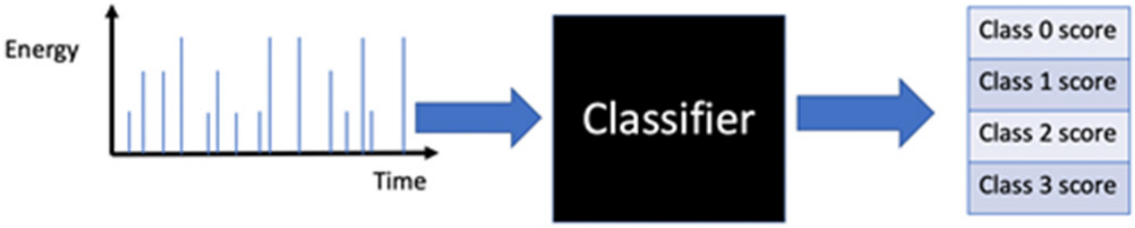

Figure 2 – List-Mode Linear Classifier Architecture

the verification process allows is desirable, and incorporating redundant IB's of independent design will add trust.

In contrast, an inspector may gain higher confidence in a measurement by performing rigorous analysis on the raw signatures of TAI's with an IB as far to the right as possible, depending on the verification technologies involved. These competing design constraints result in the development of vastly different approaches to the challenge of nuclear warhead verification.

Instead of using an electronic information barrier to separate sensitive data from an output display (far right in Figure 1), the method we propose is inherently limited in the amount of information it collects. Individual gamma ray detection pulses from a detector are input into the linear classifier individually, and only four floating point values are saved (Figure 2). The gamma ray spectrum, which is sensitive information, is never collected.

Figure 2 illustrates our linear classifier system architecture, which will ingest a pre-defined number of pulses in list-mode, storing only running scores for a small number of classes.

1.2 Scope

The concept of operation for this method in a treaty verification scenario is that a spectroscopic gamma detector system would be developed to run exclusively in list-mode operation and set to process a pre-defined number of pulse events sequentially. Ingesting a set number of pulses is a key normalization function allowing for source strength information to be largely obviated and relevant radiation signatures appropriately weighted. However, administrative controls for minimum and maximum count rates would be necessary to guard against highly shielded sources or detector saturation, respectively. During a verification process, the detector system would be set up to measure the treaty accountable item, and at the end of collection the highest class score would be used to determine the type of item being measured (Figure 2).

This paper does not directly address authentication and certification concerns, as that will be done in future work. The primary goal of this paper is to present a novel information barrier and algorithmic approach to warhead verification. As discussed in the previous section, there has never been a complete, end-to-end verification technology solution to the many problems posed by nuclear arms reduction treaties; many systems have been developed to address specific issues at various points in a more broadly comprehensive nuclear arms control treaty. This system is envisioned as a flexible option capable of tailored attribute measurement.

2. Approach

The necessarily transparent nature of nuclear arms control verification research and development often requires the use of publicly available and non-sensitive datasets. While more constrained (and thus potentially more sensitive) datasets might yield better algorithm performance, the ability to co-develop and share methods and approaches is highly prioritized in the arms control verification research community. For this initial proof of concept, our team used an algorithmic approach to generate synthetic spectra, which were fed into a linear classifier described in the following sections.

2.1 Data Generation

GADRAS, a software suite developed to perform detector response modeling, is used to generate realistic gamma-ray spectra for a multitude of potential detectors to nearly any radiological source of interest (Thoreson, et al., 2019). With ongoing development for over three decades, the built-in library of radioisotopic sources is robust, and rapid radiation transport modeling allows users to generate

Material	Very Highly Enriched Uranium	Highly Enriched Uranium (20-85%)	Weapons Grade Pu	Reactor Grade Pu 33 MWd/kg	Reactor Grade Pu 65 MWd/kg	^{233}U	Am	Np
Composition (weight %)	^{234}U, 0.70 ^{235}U, 85-92 ^{236}U, 0.3 ^{238}U, rest	^{234}U, 0.70 ^{235}U, 20-85 ^{236}U, 0.3 ^{238}U, rest	^{236}Pu, 5e-9 ^{238}Pu, 0.015 ^{239}Pu, 93.63 ^{240}Pu, 6.0 ^{241}Pu, 0.355	^{236}Pu, 3e-8 ^{238}Pu, 1.2 ^{239}Pu, 59.0 ^{240}Pu, 24.0 ^{241}Pu, 11.8 ^{242}Pu, 4.0	^{236}Pu, 4e-8 ^{238}Pu, 4.6 ^{239}Pu, 49.36 ^{240}Pu, 23.92 ^{241}Pu, 12.49 ^{242}Pu, 9.63	^{232}U, 3e-4 ^{233}U, rest	Am	Np
Age (y)	0 - 65	0 - 65	0 - 20	0 - 20	0 - 20	0 - 5	0 - 20	0 - 20
Mass (kg)	1 - *	1 - *	0.5 - 10	1 - 13	1 - 13	1 - 16	1 - 60	1 - 60
Density (g/cc)	18.95	18.95	15.75	15.75	15.75	18.95	12.0	20.45

Table 2 – Fissile materials and their associated parameters (Nelson & Sokkappa, 2008)

simulated spectra for fairly complex sources. Users can specify radiation emitting materials as well as shielding material layers in arbitrary configurations. The catalogue of training data used in this study comprises two principal classes of simulated sources: nuclear material and nuisance sources.

A method for generating random sources containing various forms of nuclear material is described by Nelson and Sokkappa (Nelson & Sokkappa, 2008). Following the algorithm for generating nuclear threat objects in this "Spanning Set" paper, tens of thousands of randomly generated fissile and fissionable material objects were created as GADRAS 1D models and transported to produce simulated gamma ray spectra. Material age and isotopic ratios were sampled as prescribed by the algorithm and outlined in Table 2. Some targeted model generation was performed to allow for class balanced training, e.g., the branching ratio specified for models containing two layers of fissile material was 10%, and of these many were supercritical and therefore not usable.Table 2 – Fissile materials and their associated parameters (Nelson & Sokkappa, 2008)

Nuisance sources encompass 184 radionuclides contained in GADRAS's built-in library; most commonly-known medical, industrial, and natural radioisotopes are available for simulating detector responses. These radionuclides were randomly selected and grouped up to three at a time, in varying activities from 10 μCi to 1 mCi. Modeled as point sources, these mixed isotope "cocktails" were then placed inside randomly generated layers of shielding as prescribed in the "Spanning Set" paper (Nelson & Sokkappa, 2008). The product of this data generation process is a continuously growing library (over 90,000 spectra at the time of writing this paper) of highly-realistic gamma ray spectra representing a very diverse set of medical, industrial and nuclear radiological sources of varying strength and shielding configurations.

2.2 Linear Classifier

The requirement to not store a full spectrum, even temporarily, necessitates processing each pulse as it arrives to the classifier. Instead of constructing a spectrum – summing the data in energy bins before it enters the classifier – we instead apply classifier weights to each pulse, keeping a running sum of the classifier's output. Notably, this is incompatible with typical classification algorithms such as neural networks with nonlinear activations, because for a nonlinear function σ, the function of a sum is not necessarily equal to the sum of the function for scalar inputs $x_i : \sigma\left(\sum_i x_i\right) \neq \sum_i \sigma(x_i)\ \forall\, x_i$ for scalar inputs x_i. List-mode processing can be done with models that have no nonlinear elements, such as linear classifiers.

3. Theory

A linear classifier is a linear mapping of inputs x to output scores y, which can be described in terms of vectors representing sets of data and outputs as:

$$\vec{y} = \boldsymbol{W}\vec{x} + \vec{b} \tag{1}$$

where $\vec{x}$ is a vector of inputs, $\boldsymbol{W}$ is a matrix of weights, $\vec{b}$ is a vector of biases, and $\vec{y}$ is a vector of output scores. The weights and biases are tunable parameters, which are trained using an optimization algorithm such as stochastic gradient descent. During inference, the predicted output class is determined by the index of the maximum value in the vector of output scores, $\vec{y}$. The desired output ($\hat{\vec{y}}$), also called the ground truth, is represented by a vector of zeros with 1 in the index of the true class.

Linear classifiers have the advantage of being highly interpretable, which is useful in an arms verification context. The input-output mapping is plainly shown by the weights. There are several complementary interpretations of the weights: the first is that they define templates onto which inputs are projected. A dot product of an input onto a template that is similar to it results in large output magnitude. The second interpretation is that the weights and biases are slopes and intercepts of decision planes in feature space, in which every input is a point. The planes make binary separations of the points into classes.

To train machine learning models, data is separated into a training set with which the model's weights and biases are adjusted, a validation set used to monitor the model's performance during training, and a test set used to measure the model's final accuracy. The weights and biases are randomly initialized. The output of each training example is computed, and a loss function quantifies the error between the computed output scores and the ground truth (the desired output). A common choice of cost function in a classification task is categorical cross-entropy loss. First, the softmax function, s, (also called the normalized exponential function) is computed over the scores, y, converting them to normalized probabilities:

$$s_i = \frac{e^{y_i}}{\sum_{j=1}^{C} e^{y_j}} \tag{2}$$

where the sum is taken over $\mathtt{C}$ output classes. Then the cross-entropy (CE) between the softmax output and the ground truth is computed:

$$CE = -\sum_{i=1}^{c} \hat{\vec{y}}_i \log(s_i). \tag{3}$$

Because $\hat{\vec{y}}$ is a vector in which there is a single nonzero entry of value 1, the sum can effectively be removed from the equation, thereby making equation (4) the full categorical cross-entropy loss function, where y_{true} is the output in the index of the true class as given by $\hat{\vec{y}} \cdot \vec{y}$.

$$L = -\log\left(\frac{e^{y_{true}}}{\sum_{j=1}^{C} e^{y_j}}\right) \tag{4}$$

In training, the weights and biases are adjusted to minimize the loss function by taking steps in the direction of the downward gradient of the loss function with respect to each tunable parameter. Despite the simplicity of our model, optimization is difficult because the inputs to the linear classifier are poorly-conditioned: gamma-ray spectra have significant differences in the orders of magnitude of their input features, and due to constraints of our algorithm, no nonlinear pre-processing transformations are permissible. The Adam optimizer (a form of gradient descent optimization) (Kingma & Ba, 2014) was selected for this application because of its individual adaptive learning rates for every parameter, with an initial learning rate of 0.1, for 100,000 epochs.

3.1 Equivalence of Linear Classifier Inference on Binned and List Mode Data

Assume that at some time $t > 0$ we have recorded p pulses. Let $v_i \in [1, N]$ be the bin associated with pulse i, for $i = 1, \ldots, P$. Let e_{v_i} be a vector of length N that is zero except for the vi-th entry, which is one. Let our energy spectrum x, be defined as $x = \sum_{i=1}^{P} e_{v_i}$ which is the count of the pulses in each bin. Finally, let our linear model be defined as $y = Wx + b$. Then,

$$y = Wx + b, \tag{5}$$

$$= \quad \sum_{i=1}^{P} e_{v_i} + b_{,}, \tag{6}$$

$$= \sum_{i=1}^{P} W e_{v_i} + b_{,}, \tag{7}$$

which means we can apply the W portion of the linear model to each individual pulse, rather than the whole accumulated x vector and get the same answer.

Therefore, a linear classifier trained on spectra may perform inference on a spectrum, or inference on list-mode data while keeping a running sum of the outputs, and the results will be the same.

4. Experiment

For this initial study, the standard detector response function for an ORTEC Detective EX-100 HPGe was used, with all spectra including default Albuquerque, NM natural background radiation. All spectra generated for model training were ideal, without Poisson noise; the impacts of varying the background and the counting statistics were not considered as part of this study. In general, this effect can be mitigated through administrative controls requiring the object of interest count rate to exceed a minimum threshold value based on background count rates (3σ is a commonly used multiplier).

To test our approach, we compared models containing weapons-grade material, defined for this study as 94% Pu-239 and greater than or equal to 90% enriched U-235 (weight percentages) to models containing reactor-grade material to determine if we could discriminate the different material types. Weapons-grade-containing models were discriminated from models containing reactor-grade material, highly enriched uranium (HEU) material just below the arbitrary threshold of weapons-grade used in this study, and standard radiological sources such as industrial and medical isotopes. Class 1 contains the 90%+ U-235 samples, class 2 contains the 94% Pu-239 samples, and class 3 contains samples with a combination of uranium and plutonium layers, where at least one shell layer is weapons grade. All other samples, whether sub-threshold or containing only industrial and/or medical isotopes are defined as class 0.

Our training data consists of 41,595 samples, of which 10% are used for validation, and our testing set consists of 5,136 samples. Every spectrum in the dataset is normalized so that the features (in this case counts in each channel bin) sum to 1. There are 8127 features, spanning energies from 20keV – 3.27 MeV (the default bin structure of the Detective EX-100 is 8192 channels, though 65 were below the lower-level discriminator and thus excluded from our optimization and training processes.

After the training and loss curves fully converged, the validation set accuracy was compared to other commonly available machine learning models included in MATLAB's classification learner (The MathWorks, Inc., 2022).

5. Results

Primary emissions from HEU (blue) and Pu-239 (green) are illustrated in Figure 3, which are representative, ideal, plots from GADRAS of both weapons grade plutonium (WGPu) and HEU sources as would be measured with standard Albuquerque, NM USA terrestrial and cosmic background.

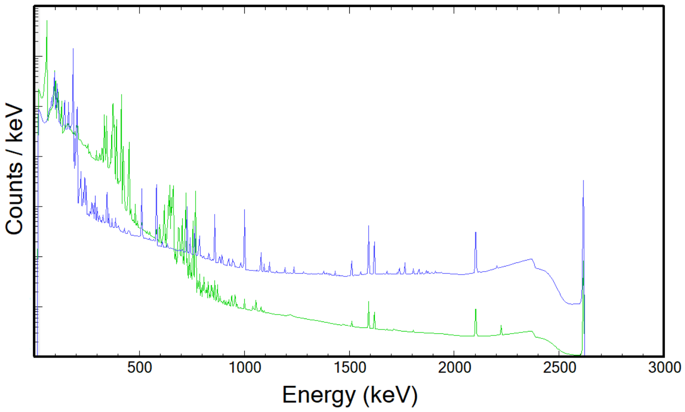

Figure 3 – Example Gamma Spectra from HEU (Blue) and WGPu (Green)

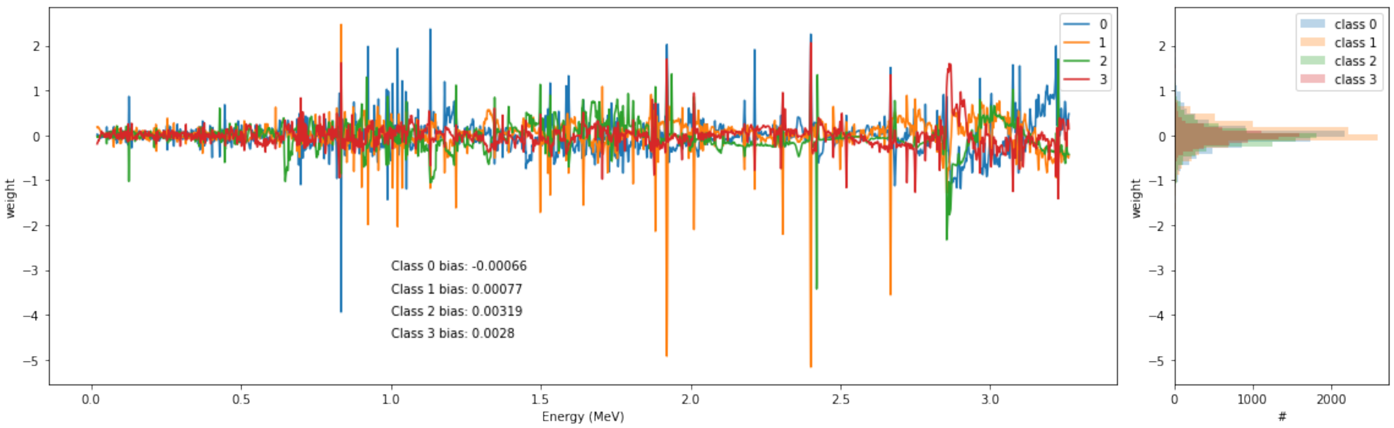

Figure 4 – Weights and Biases of Linear Classifiers

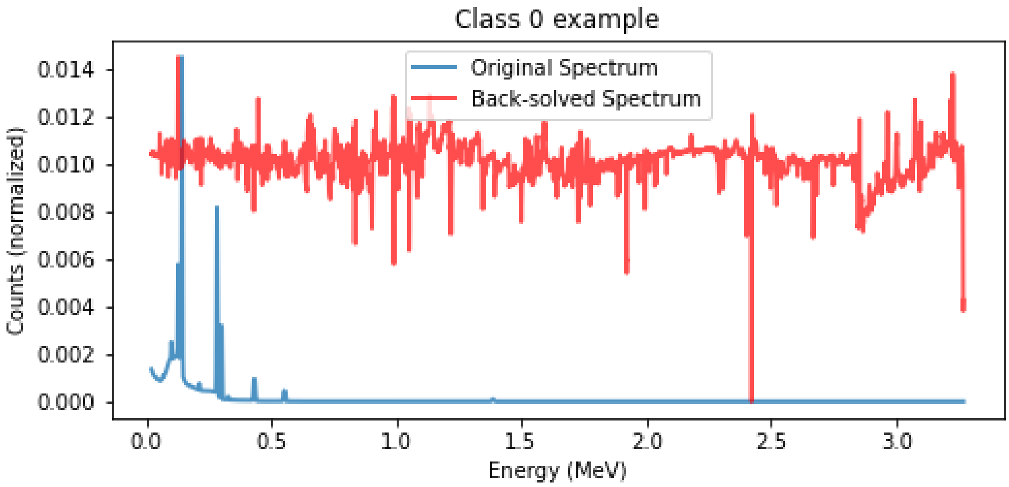

Figure 5 – Example Reconstructed Gamma Energ Spectrum

The methods described in the previous section were used to optimize the weights and biases for this dataset, shown in Figure 4. Weights are plotted against energy bins for each class on the left, with the distribution of weights plotted on the right.

Due to the large variation in magnitudes of the features and the inability to apply any nonlinear preprocessing techniques (incompatible with list-mode data), convergence of the model was extremely slow; 100,000 epochs were run to achieve the results presented in this paper.

An important feature of an information barrier is the concept of irreversibility, wherein the sensitive input signatures cannot be reconstructed given information to which an inspector may have access. In this proposed system, that would include the classifier's outputs and linear classifier weights. Figure 5 shows an example spectrum that was reconstructed by multiplying the classifier's output by the weights:

$$\vec{x} = \boldsymbol{W}^{+}(\vec{y} - \vec{b}) \tag{8}$$

Where $\boldsymbol{W}^{+}$ is the Moore-Penrose pseudo-inverse of the weight matrix. The back-solved spectrum is then rescaled to match the maximum and minimum of the original spectrum for convenient visual comparison. The results shown in Figure 5 are representative of all examples visualized; the input spectrum or pulse train is unrecognizable from the backwards reconstruction.

Accuracy results on the test set are summarized in Table 3, with our linear classifier confusion matrix shown in Figure 6. We show class-weighted accuracy measures as well as two measures specific to our dataset. The first study-specific accuracy, red/green, measures the binary classification accuracy of natural, industrial, medical, and sub-threshold special nuclear material (SNM) sources of radiation (class 0) versus all weapons grade nuclear material as defined in this study (classes 1, 2 and 3). The second study-specific accuracy, Class 3f, is a 4-class classification which "forgives" any misclassification of class 3 material (containing layers of both uranium and plutonium with at least one of them weapons grade) as class 1 or class 2. The logic here is that a sample which contains WGPu nested outside of HEU may preferentially self-shield the emissions from uranium and therefore appear to contain only plutonium. The class 3f accuracy measure considers such a classification as correct instead of erroneous. These caveated accuracy results are relevant to a notional treaty verification regime in which the inspector may only care whether an object contains weapons grade nuclear material or not.

The Tree and k-nearest neighbors (KNN) models are computed using MATLAB's classification learner app (The MathWorks, Inc., 2022) and consists of all of the "quick to train" models available in the app. None of the MATLAB models are compatible with list mode data, but we have included them for the sake of comparison to illustrate relative performance of our linear classifier to existing mature classifiers without the additional self-imposed limitations of this application.

Accuracy:	Class-weighted	Red-Green	Class 3f
fineTree	70.40	83.07	86.87
medTree	65.28	78.62	84.82
coarseTree	54.66	74.15	76.92
fineKNN	65.94	79.87	77.22
medKNN	67.16	81.63	82.46
coarseKNN	66.07	80.51	84.70
cosineKNN	67.14	81.82	82.15
cubicKNN	67.21	82.04	82.65
weightedKNN	68.97	82.09	83.11
linear classifier	73.03	83.13	82.85

Table 3 – Accuracy Results for Associated Classifiers

The linear classifier presented in this paper achieved the highest class-weighted and Red-Green accuracy scores, while achieving slightly above average for the class 3 forgiving score.

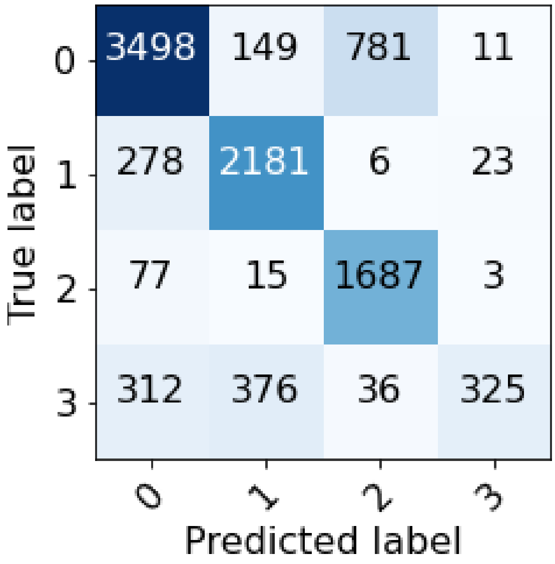

Figure 6 – Linear Classifier Confusion Matrix

From the confusion matrix (Figure 6), it is apparent that the largest single source of error in our linear classifier comes from misclassifying "other" sources (class 0) as 90%+ highly enriched uranium (class 2). This is somewhat expected, in that HEU is a relatively low-intensity source with most gamma emissions in the sub-200 keV energy range; with a minimal amount of shielding, HEU can be very hard to detect and therefore identify reliably. Further, due to the prescribed structure of the Spanning Set data generation algorithm, HEU enrichment was varied linearly from 20-92%, meaning much of the class 0 data is sub-threshold HEU (89% or less) with nearly identical signatures to 90-92% enriched HEU.

6. Conclusions and Future Work

We have shown an inherently information-limited method to classify radioactive sources using a linear classifier performing inference on list-mode gamma ray data. A sensitive spectrum is never collected, and the input cannot be reconstructed from the values that are stored.

Our results show 83% classification accuracy in distinguishing weapons grade nuclear material (as defined here) from nuisance sources, which include special nuclear material and thousands of combinations of medical and industrial isotopes. This initial result is a promising indicator that our algorithm will perform well with further refinement. Particularly interesting would be a closed-loop data generation method to maximize generation of spectra on the decision boundaries, therefore generating data that maximally improves the model.

Beyond additional data generation, there are opportunities to add complexity to the linear model to enable a more complicated decision surface defining class boundaries. One option is to include additional output classes, possibly by an unsupervised clustering of existing data into similar groups. More sophisticated non-invertible list-mode-compatible architectures also hold promise, such as an autoencoder with list-mode encoder and nonlinear decoder, storing intermediate values between them.

From a signature verification perspective, increased performance is expected when adding other radiation detection modalities such as neutron counting or multiplicity, potentially in addition to further constraining the class definitions to include attributes such as minimum mass of weapons grade material or the presence of high explosives. Substituting gamma-only scintillators with lithium loaded neutron-sensitive inorganic detectors such as Cs_2LiYCl_6:Ce (CLYC) or $Cs_2LiLaBr_{6-x}Cl_x$:Ce (CLLBC) is also of interest. Parametric studies investigating background variation, statistical sampling in list mode (accuracy vs. total counts), minimum mass of SNM and detector resolution are also being pursued.

7. Acknowledgements

This work was supported by Laboratory Directed Research and Development funding at Sandia National Laboratories.

8. References

1. Chambers, D. M., & al, e. (2010). UK-Norway Initiative: Research into Information Barriers to all warhead attribute verification without release of sensitive or proliferative information. 51st Annual Meeting of the Institute of Nuclear Materials Management. Baltimore, MD, USA.

2. Dale, C., & al, e. (2009). Third-Generation Attribute Measurement System Conceptual Design Report. Los Alamos, NM, USA: Los Alamos National Laboratory.

3. Engel, E. M., & Danagoulian, A. (2019). A physically cryptographic warhead verification system using neutron induced nuclear resonances. Nature Communications(10).

4. Evans, D. (2021). Strategic Arms Control Beyond New START - Lessons from Prior Treaties and Recent Developments. Johns Hopkins University Applied Physics Laboratory LLC.

5. Glaser, A., Barak, B., & Goldston, R. J. (2014). A zero-knowledge protocol for nuclear warhead verification. Nature, 510, 497-502.

6. Greenberg, A. (2019, October 10). WIRED. (Condé Nast) Retrieved January 23, 2023, from https://www.wired.com/story/plant-spy-chips-hardware-supermicro-cheap-proof-of-concept/

7. Hamel, M. (2018). Next-Generation Arms-Control Agreements Based on Emerging Radiation Detection Technologies. Institute of Nuclear Materials Management 59th Annual Meeting. Baltimore, MD, USA.

8. Harahan, J. P. (1993). On-Site Inspections Under the INF Treaty, A History of the On-Site Inspection Agency and Treaty Implementation, 1988-1991. Washington, DC USA: Library of Congress Cataloging-in-Publishing Data.

9. Hecla, J. J., & Danagoulian, A. (2018). Nuclear disarmament verification via resonant phenomena. Nature Communications, 9.

10. Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations. San Diego.

11. Langner, D., & al, e. (2002). Attribute Measurement Equipment for the Verification of Plutonium in Classified Forms for the Trilateral Initiative. Vienna, Austria: International Atomic Energy Agency.

12. Marleau, P., & Krentz-Wee, R. (2020). CONFIDANTE Demonstration Prototype Report. Albuquerque, NM, USA: Sandia National Laboratories.

13. Mitchell, D. J., & Tolk, K. M. (2000). Trusted Radiation Attribute Demonstration System. INMM 41st Annual Meeting. New Orleans, LA, USA.

14. Nelson, K., & Sokkappa, P. (2008). LLNL-TR-408407 A Statistical Model for Generating a Population of Unclassified Objects and Radiation Signatures Spanning Nuclear Threats. Livermore, CA: Lawrence Livermore National Laboratory.

15. Seager, K. D., & al, e. (2001). Trusted Radiation Identification System. Albuquerque, NM, USA: Sandia National Laboratories.

16. The MathWorks, Inc. (2022). Classification Learner. Retrieved April 4, 2022, from https://www.mathworks.com/help/stats/classificationlearner-app.html

17. Thoreson, G., Horne, S., Theisen, L., Mitchell, D., Harding, L., & Amai, W. (2019). SAND2019-14305 GADRAS Version 18 User's Manual. Albuquerque, NM: Sandia National Laboratories.

18. Vanier, P. E., & al, e. (2001). Study of the CIVET Design of a Trusted Processor for Non-intrusive Measurements. Annual Meeting of the Institue of Nuclear Materials Management. Indian Wells, CA, USA.

19. White, G. (2012). Review of Prior U.S. Attribute Measurement Systems. Livermore, CA, USA: Lawrence Livermore National Laboratory.

20. Wolford, J. K., & White, G. K. (2000). Progress in Gamma Ray Measurement Information Barriers for Nuclear Material Transparency Monitoring. Institute of Nuclear Materials Management 41st Annual Meeting. New Orleans, LA, USA.

21. Yan, J., & Glaser, A. (2015). Nuclear Warhead Verification: A Review of Attribute and Template Systems. Science & Global Security, 23, 157-170.