Taylor & Francis
Taylor & Francis Group

Data Science in Science

ISSN: 2694-1899 (Online) Journal homepage: www.tandfonline.com/journals/udss20

Radioisotope Identification with List-Mode Gamma-Ray Data

Assessing the Value of Temporal Information Applied to Radioisotope Identification

Lekha Patel, Efrain H. Gonzalez, Ryan J. Kamm & Aaron J. Hill

To cite this article: Lekha Patel, Efrain H. Gonzalez, Ryan J. Kamm & Aaron J. Hill (2025) Radioisotope Identification with List-Mode Gamma-Ray Data, Data Science in Science, 4:1, 2597582, DOI: 10.1080/26941899.2025.2597582

To link to this article: https://doi.org/10.1080/26941899.2025.2597582

Published online: 19 Dec 2025.

DATA SCIENCE IN SCIENCE
2025, VOL. 4, NO. 1, 2597582
https://doi.org/10.1080/26941899.2025.2597582

Radioisotope Identification with List-Mode Gamma-Ray Data
Assessing the Value of Temporal Information Applied to Radioisotope Identification

Lekha Patel, Efrain H. Gonzalez, Ryan J. Kamm, and Aaron J. Hill

Sandia National Laboratories, Albuquerque, NM, USA

ABSTRACT

This work explores the potential of utilizing temporal data from gamma-ray detectors, known as list-mode data, to enhance radioisotope identification. Traditional identification methods, which rely on full gamma-ray spectrum analysis, often require long dwell times and struggle with spectra containing similarly spaced spectral peaks. We hypothesize that by leveraging the probabilistic nature of nuclear decay and the time-encoded information from decay sequences and interactions with surrounding materials, we can improve classification accuracy over static spectral analysis. This research examines the temporal content of list-mode data through exploratory data analysis *via* correlation discovery and qualitative distribution analysis. Additionally, we propose a probabilistic classification model that can utilize spectral data, temporal data, or both to determine if the incorporation of temporal information improves radioisotope identification. Our findings suggest that the temporal information present in list-mode gamma-ray data has merit and should be further investigated to develop more robust and optimal methods for utilizing this temporal information in applications requiring radioisotope identification.

ARTICLE HISTORY
Received 27 June 2025
Revised 28 October 2025
Accepted 21 November 2025

KEYWORDS
List-mode data; radioisotope identification; Poisson process; marked point process

Introduction

Radioisotope identification (RIID) is fundamental to nuclear security, safeguards verification, and environmental monitoring, as it enables the rapid and accurate identification of radioactive materials for detecting illicit nuclear material, verifying declared nuclear activities, and responding to radiological emergencies (U.S. Environmental Protection Agency 2025; Pakari et al. 2024). When radioactive nuclei decay, they emit gamma rays at characteristic energies unique to each isotope. A gamma-ray *spectrum*, which is a histogram of detected gamma-ray counts as a function of energy, serves as a fingerprint for identifying the radioactive materials present (Glenn 2010; High resolution gamma-ray spectrometry analyses for normal operations and radiological incident response 2019). These spectra are formed by accumulating individual detection events over time, where each event represents the energy of a gamma ray recorded by the detector (Glenn 2010). Handheld RIID instruments are widely deployed by law enforcement and first responders for these purposes (U.S. Environmental Protection Agency 2025; Mirion Technologies 2025). The challenge lies in conducting successful RIID under diverse operational conditions, including weak signals, high background radiation, multiple overlapping sources, and varying measurement geometries that can render traditional methods unreliable or require impractically long measurement times (Tom and Michael 2009).

Classical identification methods have evolved from simple peak-fitting algorithms to sophisticated statistical approaches over several decades. Region-of-interest (ROI) and peak-matching techniques identify isotopes by detecting characteristic gamma lines within energy tolerance windows. While computationally efficient and intuitive, these methods require well-resolved peaks and sufficient counting statistics, often necessitating extended measurement times in low-count or high-background environments (Tom and Michael 2009). Consequently, full-spectrum template matching has improved upon ROI methods by comparing entire measured spectra against reference libraries using correlation or chi-square metrics, leveraging continuum regions and tolerating minor calibration shifts (Tom and Michael 2009; Jeffcoat et al. 2010; Mitchell et al. 2014). Nevertheless, both approaches struggle with source mixtures and spectrally similar "confuser" nuclides, such as Pu-239 versus I-131 or U-238 versus Ra-226, where overlapping peaks lead to misidentification, often necessitating long dwell times to resolve ambiguities (High resolution gamma-ray spectrometry analyses for normal operations and radiological incident response 2019; Tom and Michael 2009).

Advanced physics-based approaches, exemplified by Sandia National Laboratory's Gamma Detector Response and Analysis Software (GADRAS), employ library fitting

CONTACT Aaron J. Hill ajhill@sandia.gov Sandia National Laboratories, 1515 Eubank SE, Albuquerque, NM, USA

with detector response functions to perform multiple linear regression across candidate isotopes (Jeffcoat et al. 2010; Mitchell et al. 2014). GADRAS accounts for detector-specific energy response characteristics enabling simultaneous identification of multiple isotopes through composite spectral fitting. While more robust to spectral overlaps than peak matching, these methods depend critically on accurate detector models and calibration. Performance typically degrades when deployment conditions including distance geometry, shielding, or background conditions, deviate from library assumptions. Further, the computational demands of fitting large isotope libraries can limit real-time application. To deal with this, statistical decomposition methods developed over the past two decades offer data-driven alternatives. As one example, non-negative matrix factorization (NMF) learns basis spectra representing background components and source signatures, enabling detection and identification of weak or mixed sources through additive reconstruction (Bilton et al. 2019). On the other hand, wavelet-based approaches have been employed for spectral denoising and feature extraction (Sullivan et al. 2006). These methods naturally handle overlapping spectra by attributing counts across multiple components and adapt to background variability better than fixed thresholds, yet their iterative optimization requirements and computational complexity often necessitate significant processing resources that limits their applicability to real-time field deployment.

While these spectral methods have proven valuable, the construction of a spectrum inherently discards temporal information as gamma-ray counts are summed over time across energy bins–a fact that has been theoretically quantified (Clarkson and Kupinski 2020). In fact, nuclear decay is a probabilistic process that follows well-characterized decay sequences and known mean decay times—information that is not present in a spectrum (Gilmore and Joss 2008). Interactions of decay radiation with materials surrounding a source cause subsequent emissions with time-encoded information that can be attributed to geometry and materials. For instance, the interarrival time distribution between detection events provides information orthogonal to spectral features, as nuclear decay follows Poisson processes with isotope-specific rates (Glenn 2010). This temporal information remains discriminative even when spectral peaks overlap or when counting statistics are limited. For this reason, temporal data has been used for Compton imaging with known isotopes and for fissile material characterization *via* active interrogation (Wilderman et al. 2000; Williford 2013). Thus, utilizing the full temporal and sequential data from a detector, known as list-mode data, on its own or in combination with a spectrum, may improve methods for RIID by reducing long dwell time requirements and improving accuracy.

More recently, machine learning approaches have begun exploring temporal dynamics through time-binned representations. For instance, Convolutional neural networks (CNNs) operating on two-dimensional "waterfall" plots, where events are binned in time and energy—have demonstrated improved classification over static spectra, particularly for low-count scenarios (Moore et al. 2019; Moore et al. 2020). However, these approaches still discretize the temporal domain, potentially losing fine-scale temporal structure. To the authors' knowledge, list-mode data has not been used successfully for passive gamma-ray-based RIID.

This work presents a point process framework that operates directly on event-level list-mode data, preserving the full temporal resolution of the detection process. Our approach models the observed decay events as a superposition of isotope-specific Poisson processes, each with unique temporal and spectral signatures. This formulation enables principled likelihood-based classification that systematically compares RIID performance using energy alone, time alone, or both modalities combined. Further, we introduce an adaptive scaling parameter to account for varying signal-to-noise ratios between training and deployment conditions, addressing a critical limitation in field applications. Last, the decomposed estimation strategy achieves deployment-suitable computational efficiency through closed-form rate estimates and convex optimization.

This article is organized as follows. Section 2 provides details on the datasets that were created for this study. Section 3 focuses on the use of data analysis techniques to characterize the temporal content of our datasets. This was attempted by using standard methods for correlation discovery and qualitative distribution analysis. Section 4 develops the marked point process model that jointly leverages spectral and temporal information. Finally, Section 5 discusses the results obtained from using the aforementioned models for RIID.

2. Datasets

Our work utilizes four different datasets[1] (U.S. Environmental Protection Agency 2025), which come from radioactive sources collected with a $2\} \times 4\} \times 16\}$ low resolution inorganic scintillator (Sodium Iodide (NaI) or Cesium Iodide (CsI)) connected to a commercial multichannel analyzer (MCA)—Canberra/Micron Osprey or digiBase. For all datasets, we did not utilize the first 100 s of data due to this period of time being reserved for experiment setup. All collections were performed with a stationary detector and a stationary source. Table 1 details each dataset's sources as well as the date that they were collected. Additionally, relevant details regarding the nature of each collection are included below.

When configuring the equipment, effort was made to achieve a dynamic range of approximately 3 MeV. Because no rigorous energy calibration was implemented in our collection system or post-processing, we here use the resultant raw channel data from the instrument. In Figure 1, we show an example of the raw spectra of the Cs137 source from Dataset A based on 20s worth of events.

2.1. Dataset A

This dataset consists of five collections across four different sources, with the fifth collection being a combination of two

Table 1. The date at which each dataset was collected and the sources in each collection. An empty cell implies that no additional source was collected.

	Datasets			
Dataset name:	A	B	C	D
Date created:	July 2022	April 2024	August 2024	August 2024
Source 1:	Cs-137	Ba-133	Cs-137	Cs-137
Source 2:	Y-88	U-232	Y-88	Y-88
Source 3:	Ba-133	Ba-133 + U-232	Ba-133	Ba-133
Source 4:	U-232		Cs-137 + Ba-133	Cs-137 + Ba-133
Source 5:	Ba-133 + U-232			

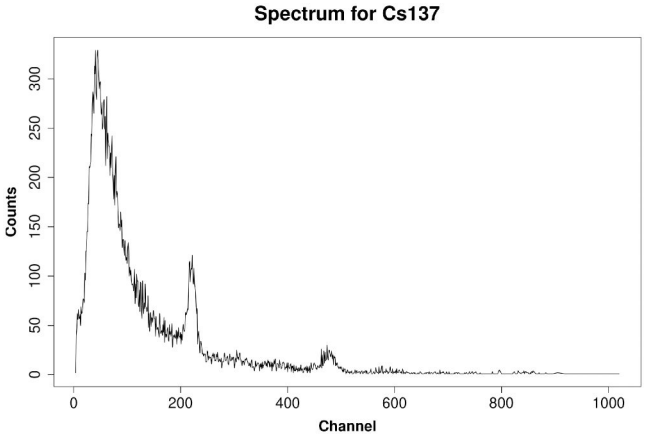


Figure 1. This plot shows the raw and uncalibrated spectrum associated with Cs137. The list-mode data associated with this plots is from Dataset A.

of the sources. The collection was performed with a stationary detector and a stationary source. Collection periods were for approximately three hours. No calibration for drift was performed on this detector, and as such any data beyond 20 min was not used.

2.2. Dataset B

This dataset consists of three collections across two different sources, with the third collection being a combination of the two unique sources. The collection was performed with a stationary detector and a stationary source. Collection periods were not controlled. Calibration of the detector was performed to account for drift from long dwell times.

2.3. Datasets C and D

These datasets were produced as a result of our initial experiments on the first two data sets. Because each source has a distinctly different gross count rate, we were concerned that the classifier may be exploiting the gross count difference as a discriminator. To test this hypothesis, these datasets placed the sources at different distances to make the gross count rate of each collection approximately the same. Table A1 of Appendix A details the distances at which each source was placed from the detector and the corresponding gross count rate achieved. Distance 1 in the table refers to the set of distances necessary in order to achieve a nominal count rate of between 1600 to 1800. The dataset derived from using these distances is referred to as Dataset C. Distance 2 refers to the set of distances necessary in order to achieve a nominal count rate of between 1200 and 1400. The dataset derived from using these distances is referred to as Dataset D. Collection periods were not controlled. For completeness, Table A2 found in Appendix A makes record of the MCA settings used for this collection.

3. Exploratory Data Analysis

In this section, we explore the relationship between time and energy in the list-mode data that was collected. Section 3.1 shows the results of correlation analysis on Datasets A and B and Section 3.2 discusses the results of the distribution analysis that was conducted on Dataset A. Datasets C and D were not analyzed because they do not contain any substantial differences over Datasets A and B as it relates to the analysis of this section.

3.1. Correlation

We analyzed a 20s sample from each Datasets A and B. To assess the relationship between interarrival time and energy, we used two correlation measures: Pearson's correlation coefficient and Spearman's rho. Here we define interarrival time as the time in microseconds between consecutive energy readings of gamma particle detections (i.e. consecutive events). Pearson's correlation coefficient quantifies the strength and direction of a linear relationship between two variables, ranging from −1 to 1, with values near zero indicating no linear relationship (Tamhane and Dunlop 2000). As expected, the Pearson correlation between interarrival time and energy was nearly zero for all sources, as shown in Table 2, confirming the absence of a linear relationship in our datasets. Spearman's rho, which measures the strength and direction of a monotonic relationship between two variables using ranks and average ranks (Jay Conover 1999), similarly showed no significant monotonic relationship between interarrival time and energy, as validated by the results in Table 2.

Given that these datasets represent time-series data, we explored two correlation metrics specific to time-series analysis: autocorrelation and cross-correlation. Autocorrelation measures the strength and direction of the linear relationship between a variable at one time point and the same variable at another time point (Shumway and Stoffer 2005). It quantifies how well a linear function can predict the $(j+k)^{th}$ element in a series based on the j^{th} element, with values ranging from −1 to 1, where larger magnitudes indicate stronger linear relationships. We analyzed interarrival

Table 2. The different correlation metrics that were used to analyze the list-mode datasets.

	Relationship between interarrival time and energy			
	Pearson's correlation	Spearman's Rho	Autocorrelation (time, energy)	Cross-correlation
Dataset A				
Background	−0.0028	−0.0007	(0.0151, 0.0083)	0.0147
Y88	−0.0019	−0.0004	(0.0142, 0.0136)	0.0156
Cs137	−0.0029	0.0009	(0.0115, 0.009)	0.0127
U232	−0.0037	−0.0017	(0.0071, 0.0068)	0.0059
Ba133	0.01	0.0085	(0.0077, 0.0089)	0.0118
U232 and Ba133	−0.0003	−0.0044	(0.004, 0.004)	0.0063
Dataset B				
Background	−0.0092	−0.0218	(0.0196, 0.0154)	0.0166
U232	−0.0041	−0.0138	(0.0144, 0.012)	0.0145
Ba133	−0.0077	−0.0107	(0.013, 0.0105)	0.0123
U232 and Ba133	−0.0077	−0.0187	(0.0088, 0.0132)	0.0129

times and energy separately, hypothesizing that autocorrelation values would be near zero due to the lack of linear relationships in list-mode data. As shown in Table 2, the maximum autocorrelation values for interarrival times and energy, calculated over lags $k \in 1, \ldots, 42$, confirmed a lack of correlation across all datasets.

Cross-correlation was used to test for linear relationships between the series of interarrival times and energy channels. The maximum cross-correlation values, calculated between $x[t+k]$ (interarrival time for the t^{th} event with lag $k \in -42, \ldots, 42$) and $y[t]$ (energy channel for the t^{th} event), are presented in Table 2. All values indicate no linear relationship between the two variables. This analysis supports the assumption that interarrival times and energy channels are linearly independent; a critical notion leveraged in the development of our classifier in Section 4.

3.2. Distributions

Although we were unable to establish the existence of a relationship between interarrival time and energy channel in the previous section, we investigated the energy and interarrival time distributions separately in order to evaluate whether or not there are sufficient differences in the distributions of each of these variables for each of the classes. This investigation was conducted by generating box plots from the 20s samples that we had used in the previous section. Box plots are a visual representation of the "five-number summary," which includes the first (Q_1), second (Q_2), and third quartile (Q_3) as well as the upper and lower fence. The second quartile is usually referred to as the median, and the upper and lower fences are calculated by using the following equations:

$$\text{upper fence} : Q_3 + 1.5 \times IQR \tag{1}$$

$$\text{lower fence} : Q_1 - 1.5 \times IQR \tag{2}$$

$$IQR = Q_3 - Q_1 \tag{3}$$

where IQR is called the interquartile range. Values outside of the fences are considered to be outliers.

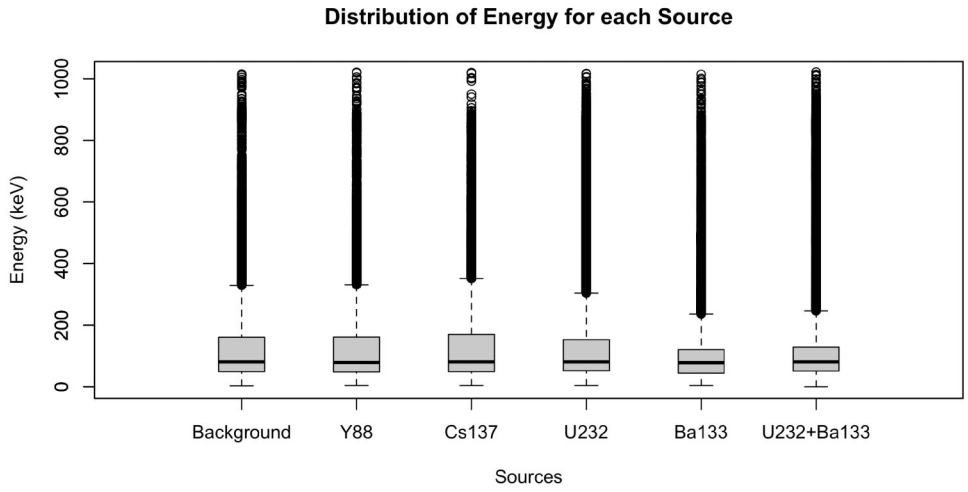
If the minimum value in a dataset exceeds the calculated lower fence, the lower fence is set to the dataset's minimum value. Similarly, if the maximum value is smaller than the calculated upper fence, the upper fence is set to the dataset's maximum value. In our datasets, the lower fence adjustment was common due to the smallest possible interarrival time being 1 microsecond. Figure 2 presents the initial distribution of energy channels and interarrival times for each source. As expected, the energy channel distribution varies by radiological source, which explains why the spectrum is effective for radioisotope identification. Surprisingly, a similar variability is present in the case of interarrival times, suggesting that there is potential utility in using interarrival time data in a model intended for radioisotope identification.

Our study focused on leveraging the time dimension of list-mode data for radioisotope identification, which required examining the stability of interarrival time distributions. We analyzed how the distribution changes across 80 different 20s windows and as the size of the sampling window varies. When varying sampling window sizes, we used the formula $20/d$, where $d \in 1, \ldots, 100$, resulting in window sizes ranging from 20s to 0.2s. Each window size was assigned an interval length index, with $d = 1$ corresponding to 20s and $d = 100$ corresponding to 0.2s. For this analysis, data below channel number 200 were removed to reduce background events. Figure 3 summarizes the analysis that we conducted by presenting the distribution of the median and interquartile range of each source as the 20s sampling window was changed and as the sampling window size was varied. Due to the difference in the range of values across the different sources, the median and interquartile range values were $\log_{10}$-scaled.

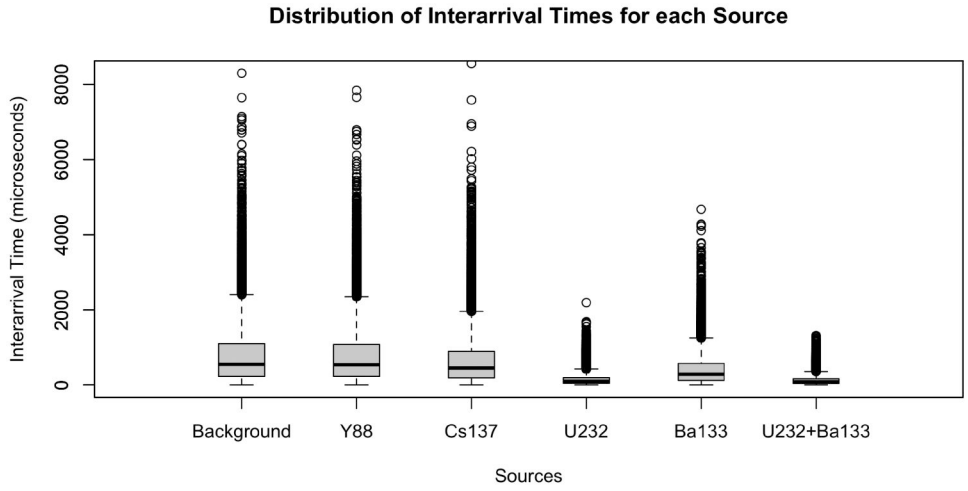
Figures 3(a and b) indicate that the interarrival time distribution remains stable across the 80 20 s windows, whereas Figures 3(c and d) show increased variability as the sampling window size changes. This stability across 20 s windows suggests that variability in interarrival times across sources is consistent regardless of the chosen window.

4. Probabilistic Classifier Model

In this section, we develop a novel probabilistic framework for RIID based on the well-established Poisson process representation of radioactive decay (Gilmore and Joss 2008). The Poisson process provides a natural mathematical model for the physics of radioactive decay, where detectors collect list-mode data recording individual decay events as timestamp-energy pairs. In environments with multiple



(a) Distribution of energy values for each source collected from Dataset A



(b) Distribution of interarrival times for each source collected from Dataset A

Figure 2. (a) and (b) display the distribution of energy and interarrival time on a per source basis over the 20s window that was used for the correlation analysis presented in the previous subsection.

isotopes and background radiation, we model the observed decay events as a superposition of independent Poisson processes, each corresponding to a different source with unique temporal and spectral signatures. By incorporating energy levels as marks within a marked point process approach, we enable likelihood-based classification that can compare RIID performance using (1) energy levels alone, (2) event timestamps alone, or (3) both measurements combined. We further extend our methodology to handle varying signal-to-noise ratios between training and testing environments by introducing a scaling factor on the Poisson rate parameter, learned *via* maximum a posteriori estimation with an appropriate prior. This enables robust RIID across changing measurement conditions.

4.1. Modeling List-Mode Data

In RIID, detectors collect list-mode data containing decay events as tuples: a timestamp paired with a discretized energy measurement. Formally, a gamma-ray detector records data $\{(E_i, T_i)\}_{i=1}^{n}$, where T_i denotes the timestamp of the i-th decay event with $T_i > T_j$ for $i > j$. Given $T_i = t$, the random variable $E_i \equiv E(t)$ represents the discrete energy recorded at time t, taking values in $\mathbb{Z}_+$.

Consider an environment containing $K \geq 0$ isotopes plus background radiation. Each of these $K+1$ sources emits gamma rays randomly and independently. While the detector records decays from all sources, it cannot differentiate between sources at measurement time. We model this scenario using homogeneous Poisson processes (HPPs): let $N_t^0 \sim \text{Poi}(\lambda_0 t)$ denote background decay counts, and $N_t^i \sim \text{Poi}(\lambda_i t)$ denote decay counts from the i-th isotope ($i = 1, ..., K$).

Since distinct isotopes have unique decay rates (half-lives), each of the $K+1$ sources possesses a unique temporal signature, implying $\lambda_0 \neq \lambda_1 \neq \cdots \neq \lambda_K$. The total decay count recorded by the detector follows

$$N_t = \sum_{i=0}^{K} N_t^i \sim \text{Poi}\left(\sum_{i=0}^{K} \lambda_i \; t\right), \tag{4}$$

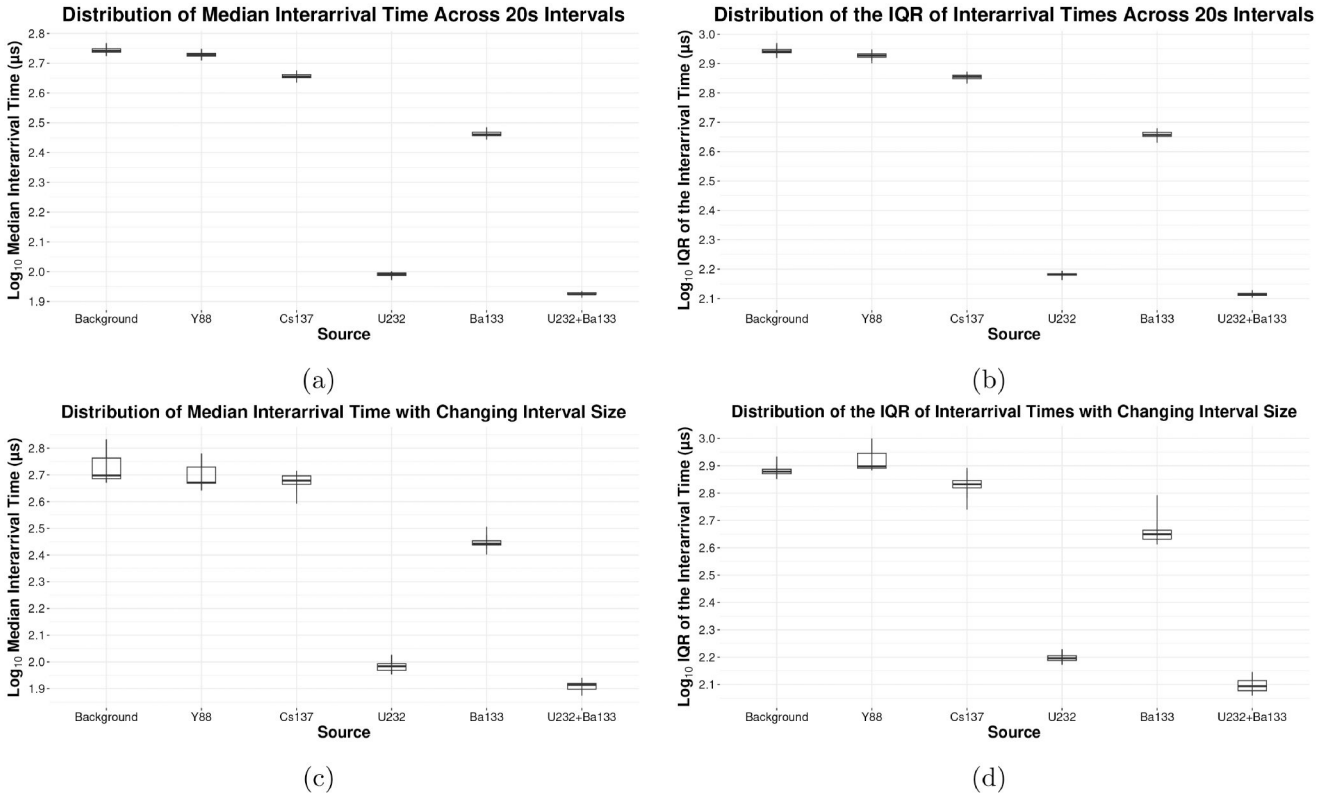


Figure 3. (a) and (c) display the distribution of the $\log_{10}$-scaled median interarrival times as the sampling window changes and as the size of the sampling window changes, respectively. (b) and (d) display the change in the $\log_{10}$-scaled interquartile range of the interarrival times as the sampling window changes and as the size of the sampling window changes, respectively. The list-mode data associated with these plots is from Dataset A.

by the superposition property of Poisson processes. This property also implies that a count from process N_t^i is seen in N_t with mixing probability $\pi_i = \lambda_i / \sum_{j=0}^{K} \lambda_j$.

With this, one can show that the interarrival times $X_1, \ldots, X_n$ between successive decays follow exponential distributions:

$$X_i \sim \text{Exp}\left(\sum_{k=0}^{K} \lambda_k\right), \tag{5}$$

where $X_i = T_i - T_{i-1}$ and $T_0 = 0$.

4.1.1. Marked Point Process Setup

Due to the tuple of observations acquired at each observation time point, a marked point process setup is natural for list-mode RIID. Specifically, let $(\mathcal{E}, \mathcal{T}, \mathcal{P})$ denote a marked point process with random counting measure $\mathcal{N}$ defined on $\mathcal{E} \times \mathcal{T}$, where: $\mathcal{T}$ is the underlying point process representing the locations of the points, $\mathcal{E}$ is the set of possible marks, $\mathcal{P}$ is the probability distribution of the marks; $\mathcal{N}(B \times A)$ denotes the number of points of $\mathcal{T}$ in $A \subset \mathcal{T}$ with marks in $B \subset \mathcal{E}$, and for any finite collection of pairwise disjoint sets $\{(B_i, A_i) | i \in \mathbb{N}\}$, the collection of random variables $\{\mathcal{N}(B_i \times A_i)\}_{i \in \mathbb{N}}$ are mutually independent.

A marked point process on $\mathbb{R}_{\geq 0} \times \mathbb{Z}_+$ is characterized by its *conditional intensity function* (CIF) $\lambda^*(e, t)$, such that for each $t \in \mathbb{R}_{\geq 0}$ and $e \in \mathbb{Z}_+$, $\lambda^*(e, t) > 0$ represents the intensity at temporal location t with energy e. It satisfies

$$\lambda^*(e,t) = \lim_{\substack{\delta t \to 0 \\ \delta e \to 0}} \frac{\mathbb{E}\left[\mathcal{N}(e + \delta e, t + \delta t) - \mathcal{N}(e,t) | \mathcal{H}_t\right]}{\delta e \delta t},$$

where $\mathcal{H}_t$ denotes the history up to time t. Specifically, the CIF specifies the expected number of points in an infinitesimal neighborhood of t with mark e given $\mathcal{H}_t$.

To define the CIF of the list-mode marked process, we note that

$$\lambda^*(e,t) = p_E(e;t)\bar{\lambda}(t),$$

where $p_E(e;t)$ denotes the probability of observing mark e at time t, and $\bar{\lambda}(t)$ is the (temporal) CIF of the underlying point process. Based on our correlation analysis in Section 3.1, which found no temporal-spectral dependence, we assume $p_E(e;t) \equiv p_E(e)$ for the modeling herein.

4.2. Probabilistic Classification Framework

Building upon the Poisson process foundation, we now present a probabilistic classification framework for radioisotope identification from list-mode data. Our approach leverages the marked point process representation to perform likelihood-based classification using temporal information, spectral information, or both modalities combined.

For parameter estimation, we assume access to training data from each isotope of interest. Specifically, for K isotopes, we require $K + 1$ datasets:

$$\{\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_K\}, \tag{6}$$

where $\mathcal{D}_0$ contains background-only measurements and $\mathcal{D}_k$ contains measurements from background plus isotope k. Each dataset is defined as:

$$\mathcal{D}_k = \{(e_i^{(k)}, t_i^{(k)})\}_{i=1}^{n_k}, \tag{7}$$

where n_k denotes the number of events, $t_i^{(k)}$ represents time-stamps, and $e_i^{(k)} \in \mathbb{Z}_+$ represents discretized energy measurements, with (k) denoting observations taken from dataset $\mathcal{D}_k$.

4.2.1. Joint Log-Likelihood Decomposition

For data $\mathcal{D}_k$ with $|\mathcal{D}_k| = n_k$ containing background and isotope k, observed over interval $[0, T]$, the likelihood under parameters $\boldsymbol{\theta} = \{(\lambda_i, \theta_{p_{E_i}})\}_{i\in\{0,k\}}$, where $\theta_{p_{E_i}}$ denotes the parameters (or nonparametric specification) of the mark distribution p_{E_i} is:

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}_k) = \left(\prod_{i=1}^{n_k} \lambda^*(e_i^{(k)}, t_i^{(k)})\right) \exp\left(-\int_0^T \bar{\lambda}(u)du\right), \qquad \text{with } \bar{\lambda}(t) \equiv \lambda_0 + \lambda_k, \tag{8}$$

as per the standard marked point process likelihood [see, e.g. Daley and Vere-Jones 2003, Equation (7.3.5)]. Under our independence assumption,

$$\begin{aligned} \lambda^*(e_i^{(k)}, t_i^{(k)}) &= (\lambda_0 + \lambda_k) p_E^{(k)}\left(e_i^{(k)}\right), \\ p_E^{(k)}(e) &:= \pi_0\ p_{E_0}(e) + \pi_k\ p_{E_k}(e), \end{aligned} \tag{9}$$

where $\pi_0 := \frac{\lambda_0}{\lambda_0+\lambda_k}$, $\pi_k := \frac{\lambda_k}{\lambda_0+\lambda_k}$ are the mixture weights. Therefore, the likelihood simplifies to

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{D}_k) = \left(\prod_{i=1}^{n_k} (\lambda_0 + \lambda_k)\ p_E^{(k)}(e_i^{(k)})\right) \exp\left[-T(\lambda_0 + \lambda_k)\right]. \tag{10}$$

Taking logs gives

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathcal{D}_k) &= n_k \log(\lambda_0 + \lambda_k) + \sum_{i=1}^{n_k} \log p_E^{(k)}(e_i^{(k)}) \\ &\quad - T(\lambda_0 + \lambda_k). \end{aligned} \tag{11}$$

This decomposition naturally separates temporal and spectral components:

$$\ell_{\text{temporal}}(\lambda_0, \lambda_k \,|\, \mathcal{D}_k) = n_k \log(\lambda_0 + \lambda_k) - T(\lambda_0 + \lambda_k), \tag{12}$$

$$\ell_{\text{spectral}}(p_{E_0}, p_{E_k} \,|\, \mathcal{D}_k) = \sum_{i=1}^{n_k} \log p_E^{(k)}(e_i^{(k)}). \tag{13}$$

For computational efficiency, we estimate the temporal rates in closed form and the spectra *via* density estimation, rather than performing a joint high-dimensional optimization.

4.2.2. Temporal Parameter Estimation

The maximum likelihood estimate for the combined rate $\lambda_{\{k,0\}} = \lambda_0 + \lambda_k$ is:

$$\hat{\lambda}_{\{k,0\}} = \frac{n_k}{\sum_{i=1}^{n_k} x_i^{(k)}}, \tag{14}$$

where $x_i^{(k)} = t_i^{(k)} - t_{i-1}^{(k)}$ are interarrival times and $t_0^{(k)} = 0$. From the background-only dataset $\mathcal{D}_0$:

$$\hat{\lambda}_0 = \frac{n_0}{\sum_{i=1}^{n_0} x_i^{(0)}}. \tag{15}$$

For each isotope k, we then obtain:

$$\hat{\lambda}_k = \hat{\lambda}_{\{k,0\}} - \hat{\lambda}_0. \tag{16}$$

4.2.3. Spectral Parameter Estimation

For each dataset $\mathcal{D}_k$, we estimate the energy distribution using kernel density estimation:

$$\hat{p}_E^{(k)}(e) = \frac{1}{n_k h} \sum_{i=1}^{n_k} K\left(\frac{e - e_i^{(k)}}{h}\right), \tag{17}$$

where $K(\cdot)$ is a kernel function and h is the bandwidth parameter. Note that for $k > 0$, this represents the mixture distribution:

$$p_E^{(k)}(e) = \frac{\lambda_0}{\lambda_0 + \lambda_k}\ p_{E_0}(e) + \frac{\lambda_k}{\lambda_0 + \lambda_k}\ p_{E_k}(e), \tag{18}$$

where p_{E_0} is the background spectrum and p_{E_k} is the isotope spectrum.

4.2.4. Treatment of Isotope Mixtures

In our experimental validation, datasets containing multiple isotopes (e.g. U-232 + Ba-133) are treated as distinct classes with their own characteristic temporal and spectral signatures, rather than attempting spectral unmixing. This simplification allows us to focus on validating the core hypothesis about temporal information's value for RIID.

4.3. Classification of Unknown Sequences

Given the estimated parameters $\hat{\boldsymbol{\theta}} = \{\hat{\lambda}_k, \hat{p}_E^{(k)}\}_{k=0}^K$ from the training phase, we classify an unknown sequence $\mathcal{D}_{\text{test}} = \{(e_i', t_i')\}_{i=1}^{n'}$ into one of $K+1$ classes: background only $(k = 0)$ or background plus isotope k $(k \in \{1, \ldots, K\})$.

4.3.1. Likelihood Computation

For each isotope hypothesis $k \in \{1, \ldots, K\}$, we compute the log-likelihood of the test sequence under the model that includes background and isotope k:

$$\begin{aligned} \ell_k(\mathcal{D}_{\text{test}}|\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n'} \left[\log(\hat{\lambda}_0 + \hat{\lambda}_k) + \log \hat{p}_E^{(k)}(e_i')\right] \\ &\quad - (\hat{\lambda}_0 + \hat{\lambda}_k)T', \end{aligned} \tag{19}$$

where T' is the observation period and

$$\hat{p}_E^{(k)}(e) = \frac{\hat{\lambda}_0}{\hat{\lambda}_0 + \hat{\lambda}_k}\ \hat{p}_{E_0}(e) + \frac{\hat{\lambda}_k}{\hat{\lambda}_0 + \hat{\lambda}_k}\ \hat{p}_{E_k}(e). \tag{20}$$

This can be equivalently written in terms of interarrival times:

$$\ell_k(\mathcal{D}_{\text{test}}|\hat{\boldsymbol{\theta}}) = n' \log(\hat{\lambda}_0 + \hat{\lambda}_k) + \sum_{i=1}^{n'} \log \hat{p}_E^{(k)}(e_i') - (\hat{\lambda}_0 + \hat{\lambda}_k) \sum_{i=1}^{n'} x_i', \tag{21}$$

where $x_i' = t_i' - t_{i-1}'$ with $t_0' = 0$.

4.3.2. Decision Rule

The maximum likelihood classification rule assigns the test sequence to isotope k^* where:

$$k^* = \arg \max_{k \in \{1, \ldots, K\}} \ell_k(\mathcal{D}_{\text{test}}|\hat{\boldsymbol{\theta}}). \tag{22}$$

This decision rule minimizes the Bayes risk under 0–1 loss when the prior probabilities for each isotope are equal. Specifically, under the 0–1 loss function $L(k, k') = 1_{k \neq k'}$, the Bayes-optimal decision rule is:

$$k^*_{\text{Bayes}} = \arg \max_k P(k|\mathcal{D}_{\text{test}}) = \arg \max_k p(\mathcal{D}_{\text{test}}|k) P(k). \tag{23}$$

With uniform priors $P(k) = 1/K$, this reduces to maximum likelihood classification.

4.3.3. Modal Classification

To assess the relative importance of temporal and spectral information, we can perform classification using each modality independently. For temporal-only classification:

$$k_t^* = \arg \max_{k \in \{1, \ldots, K\}} \left[n' \log(\hat{\lambda}_0 + \hat{\lambda}_k) - (\hat{\lambda}_0 + \hat{\lambda}_k) \sum_{i=1}^{n'} x_i' \right]. \tag{24}$$

For spectral-only classification:

$$k_e^* = \arg \max_{k \in \{1, \ldots, K\}} \sum_{i=1}^{n'} \log \hat{p}_E^{(k)}(e_i'). \tag{25}$$

The relative performance of these modal classifiers provides insight into the information content of different data types for RIID.

4.4. Adaptive Signal-to-Noise Ratio Modeling

In operational settings, the signal-to-noise ratio between training and deployment conditions often varies due to factors such as source-detector distance, shielding, and environmental conditions. We address this challenge by introducing a multiplicative scaling factor $\alpha > 0$ that modulates the observed isotope intensities while maintaining the background rate constant.

4.4.1. Modified Observation Model

Under varying SNR conditions, the observed isotope rate for isotope k becomes:

$$\lambda_k^{\text{obs}} = \alpha\ \lambda_k^{\text{train}}, \tag{26}$$

where $\alpha > 0$ is an unknown parameter denoting the proportion of the training rates uncorrupted by noise, while the background rate remains unchanged: $\lambda_0^{\text{obs}} = \lambda_0^{\text{train}}$. The modified mixture energy distribution therefore becomes:

$$p_E^{(k)}(e|\alpha) = \frac{\lambda_0}{\lambda_0 + \alpha\lambda_k}\ p_{E_0}(e) + \frac{\alpha\lambda_k}{\lambda_0 + \alpha\lambda_k}\ p_{E_k}(e). \tag{27}$$

4.4.2. Bayesian Estimation of α

We adopt a Bayesian framework for estimating α, placing a log-normal prior that encodes our expectation of moderate deviations from training conditions:

$$\alpha \sim \text{LogNormal}(\mu_\alpha, \sigma_\alpha^2), \tag{28}$$

where typically $\mu_\alpha = 0$ (corresponding to mean 1) and σ_α is chosen based on expected operational variability. In the results that follow, we fix $\mu_\alpha = 0$ and $\sigma_\alpha = 1$.

For each isotope hypothesis k and test sequence $\mathcal{D}_{\text{test}}$, with $|\mathcal{D}_{\text{test}}| = n'$ over time $[0, T']$, the posterior distribution of α is:

$$p(\alpha|\mathcal{D}_{\text{test}}, k) \propto p(\mathcal{D}_{\text{test}}|k, \alpha)\pi(\alpha), \tag{29}$$

where the likelihood term is:

$$p(\mathcal{D}_{\text{test}}|k, \alpha) = (\lambda_0 + \alpha\lambda_k)^{n'} \exp\left[-(\lambda_0 + \alpha\lambda_k)T'\right] \prod_{i=1}^{n'} p_E^{(k)}(e_i'|\alpha). \tag{30}$$

4.4.3. MAP Estimation and Optimization

To enable probabilistic classification, we find the maximum a posteriori (MAP) estimate of α:

$$\hat{\alpha}_k = \arg \max_{\alpha > 0} \log p(\alpha|\mathcal{D}_{\text{test}}, k). \tag{31}$$

Here, the log-posterior objective function is:

$$\log p(\alpha|\mathcal{D}_{\text{test}}, k) = n' \log(\lambda_0 + \alpha\lambda_k) - (\lambda_0 + \alpha\lambda_k)T' + \sum_{i=1}^{n'} \log p_E^{(k)}(e_i'|\alpha) - \frac{(\log\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2} - \log\alpha + \text{const}. \tag{32}$$

Under reasonable conditions (sufficient data and well-separated isotope signatures), a unique maximum exists. We employ Newton–Raphson iteration for efficient optimization.

4.4.4. Classification with Adaptive SNR

The final classification incorporates both the data likelihood and the prior on α:

$$k^* = \arg \max_{k \in \{1, \ldots, K\}} \left[\log p(\mathcal{D}_{\text{test}}|k, \hat{\alpha}_k) + \log \pi(\hat{\alpha}_k)\right]. \tag{33}$$

Table 3. The accuracy results that were obtained for each experiment. The presence of α after the experimental number indicates that the extended α model was tested using the experimental setup associated with that number.

Accuracy results			
Experiment:	Time model	Energy model	Energy and time model
1	0.7106	0.6917	0.7767
1-α	0.5966	0.6917	0.7288
2	0.8773	0.4782	0.6258
2-α	0.6850	0.4782	0.5390
3	0.4841	0.8570	0.8763
3-α	0.4186	0.8571	0.8752
4	0.3428	0.6412	0.7304
4-α	0.4131	0.6421	0.8160
5	0.3977	0.8883	0.9036
5-α	0.3716	0.8883	0.8989
6	0.5157	0.7939	0.8512
6-α	0.4463	0.7938	0.8457
7	0.5038	0.8012	0.8312
7-α	0.4352	0.8012	0.8314

This adaptive approach automatically adjusts for varying measurement conditions while maintaining discriminative power between isotopes. Further, the inclusion of the prior term $\log \pi(\hat{\alpha}_k)$ penalizes extreme values of α, promoting robust classification when limited data makes α estimation uncertain.

5. Classifier Analysis

Using the probabilistic classification models from Section 4 and the datasets described in Section 2, we conducted experiments to evaluate whether temporal information, in the form of interarrival times, provides an advantage over using energy information alone for classifying radiological source material. For each experiment, 100,000 consecutive events per source were selected with 70% (70,000 events) used for training. The test sets, while varying by experiment, consistently contained 3,333 samples per source, with each sample comprising 100 events. Accuracy results, shown in Table 3, reflect the model's performance when the predicted class corresponds to the class with the highest likelihood.

5.1. Experimental Framework

Each experiment utilized a unique combination of training and test sets. For combination sources such as U-232 + Ba-133 and Cs-137 + Ba-133, these were treated as single isotopes[2] (2), allowing for individual isotope classification as described in Section 4.2. Experiments 1 and 2 tested the model using training and test sets from the same distribution, with the detector positioned at the same distance from each source, serving as an initial evaluation of the utility of incorporating temporal information into radioisotope identification. Experiments 3 and 6 assessed the model's dependency on gross count rate for source identification. Experiments 4 and 5 evaluated the impact of variations in signal-to-noise ratio on model performance. Finally, Experiment 7 examined the effect of added variability by combining different distributions for the training and test sets. We note that for all experiments that relied on Datasets A and B, the source–detector standoff remained the same for all sources throughout acquisition; as such, distance is not a factor in the A-only and B-only analyses of Experiments 1 and 2. The only experiments which addressed standoff effects used Datasets C and D. Further details on all experiments and their results are provided below:

5.1.1. Experiment 1

In this experiment, Dataset A was used for training and testing. The test set for each source was derived from the 30,000 events that were not used in the training set. Therefore, the test set and the training set come from the same data distribution.

5.1.2. Experiment 2

In this experiment, Dataset B was used for training and testing. The test set was derived in a similar way as that described in Experiment 1. Therefore, the test set and the training set come from the same data distribution.

5.1.3. Experiment 3

In this experiment, Dataset C was used for training and testing. The test set was derived in a similar way as that described in Experiment 1. Therefore, the test set and the training set come from the same data distribution which means that the distances at which the sources are placed is remaining constant from the training set to the test set.

5.1.4. Experiment 4

In this experiment, Dataset C was used as the training set whereas Dataset D was used as the test set. The test set for each source was derived from 30,000 events found in Dataset D. Therefore, the test set and the training set come from two different data distributions and the distances at which the sources are placed is increasing from the training set to the test set.

5.1.5. Experiment 5

In this experiment, Dataset D was used as the training set whereas Dataset C was used as the test set. The test set for each source was derived from 30,000 events found in Dataset C. Therefore, the test set and the training set come from two different data distributions and the distances at which the sources are placed is decreasing from the training set to the test set.

5.1.6. Experiment 6

In this experiment, Dataset D was used for training and testing. The test set was derived in a similar way as that described in Experiment 1. Therefore, the test set and the training set come from the same data distribution which means that the distances at which the sources are placed is remaining constant from the training set to the test set.

5.1.7. Experiment 7

In this experiment, Datasets C and D were combined and then used to create the training and test set. The test set for each source was derived from 30,000 events found in the combined dataset for each source. Therefore, the test set and the training set come from a combination of two different data distributions. This arrangement also means that the training and test sets are representative of the two different distances at which events were collected for each source.

5.2. Accuracy of Probabilistic Models

In Table 3, we present experimental results of all datasets when trained and tested on the initial probabilistic classifier developed, and the extended α model that incorporates variabilities in the SNRs between training and testing datasets.

The results demonstrate that, across all experiments, models incorporating both energy and time consistently outperformed energy-only models (i.e. spectral analysis) in terms of accuracy. However, the benefit of adding temporal information varied by experiment, largely due to differences in isotopes' decay parameters. For instance, the improved performance of time-based models in Experiments 1 and 2 may be attributed to their reliance on count rates for class differentiation. In Experiments 3 and 6, count rates were controlled by positioning sources at distances that ensured similar rates across sources, reducing the model's ability to rely on count rate as a distinguishing feature. This likely explains the lower performance of time-based models in these experiments compared to Experiments 1 and 2.

Experiments 4 and 5 introduced additional performance drops, likely due to differences in data distributions between training and test sets. In these experiments, count rates were controlled similarly to Experiments 3 and 6, but Dataset D represented sources positioned farther from the detector than those in Dataset C. Consequently, Experiment 4's training set had a higher source-to-background ratio than its test set, while Experiment 5′s training set had a lower ratio. These shifts in data distribution likely contributed to the observed performance differences between the models in Experiments 4 and 5.

The extended α model, designed to account for uncertainties in signal-to-noise ratios (SNRs) between training and test sets, significantly improved classifier performance, as evidenced by Experiment 4-α. Moreover, the extended α model often achieved comparable performance to the original model, with only minor decreases in accuracy, demonstrating its robustness and potential for handling complex and varied data collection scenarios.

5.3. Receiver Operating Characteristic Curves

Receiver operating characteristic (ROC) curves are a tool used for assessing the performance of a binary classifier, comparing the performance of binary classifiers, and establishing a threshold likelihood value at which a data point is said to belong to a specific class (Francis Sahngun 2022). The ROC curve typically has sensitivity on the y-axis and specificity or the false positive rate (FPR) on the x-axis. Sensitivity is defined as the ratio of the number of observations that were correctly associated to the class of interest by the model to the total number of observations that belong to the class of interest. Specificity is defined as the ratio of the number of observations that were correctly associated to the class that was not of interest by the model to the total number of observations that do not belong to the class of interest (Ho Park et al. 2004).

Nonparametric ROC curves are generated by calculating sensitivity and specificity at various threshold values. Higher thresholds require larger likelihood values for classification into the class of interest, typically resulting in higher specificity and lower sensitivity. Points corresponding to higher thresholds appear closer to the lower left corner of the ROC curve, while lower thresholds result in the opposite.

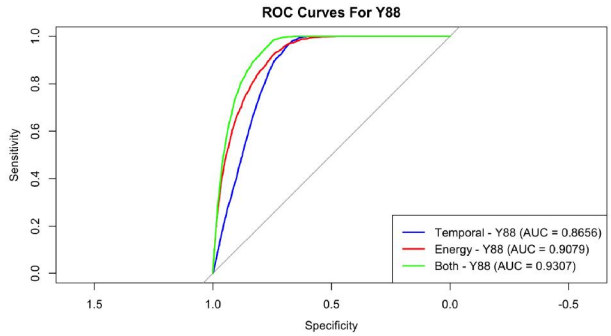
The area under the ROC curve (AUC) is often used to summarize the curve and compare classifiers. Unlike accuracy or error rate, AUC is independent of the threshold value (Hand and Till 2001). Let $\hat{p} = P(\hat{x}_i = 0|x_i = 1)$ be the estimated probability of incorrectly assigning a randomly chosen member of one class to another class and $\hat{q} = P(\hat{x}_i = 0|x_i = 0)$ be the estimated probability of correctly assigning a randomly chosen member of the other class, then the AUC can be defined as the probability that $\hat{p}$ is smaller than $\hat{q}$ (Hand and Till 2001). Therefore, better performing models are associated with larger AUC values, with AUC values ranging between 0 and 1 (Ho Park et al. 2004).

We applied ROC curve analysis to our experiments. Experiment 1's ROC curves and associated AUC values are shown in Figure 4. Appendix B contains the AUC values for every source across each experiment. These ROC curves and AUC values were calculated by treating the class of interest as one class and grouping all other classes into a second class, ensuring a binary classification scenario.

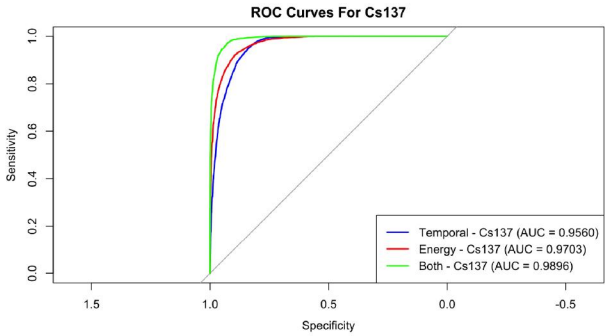
Since AUC is only applicable to situations where there are only two classes, additional analysis was conducted using the M measure presented in (Hand and Till 2001), which is the average of all pairwise AUCs and can be calculated as shown in Equation (34). In Equation (34), s represents the number of classes and AUC_{ij} represents the pairwise AUC between the i^{th} and j^{th} classes. It is important to note that these pairwise AUCs are different from the AUCs shown in Figure 4 and those found in Appendix B. The M measure is also referred to as the "multi-class AUC" (Robin et al., 2011).

$$M = \frac{\sum_{i \neq j} AUC_{ij}}{s(s-1)} \tag{34}$$

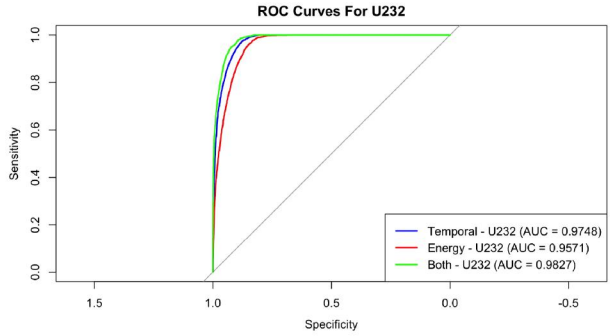
The results presented in Table 4 and the ROC curves and AUC values found in Appendix B indicate that the energy and time model outperforms the energy model in nearly every experiment. The extended α model seemed to perform reliably under all conditions. Our results highlight the flaw in the original model that we used, which exploits the proportion of background that is present in the signal for testing. When the proportion of background present in the signal increases from the training set to the test set, the



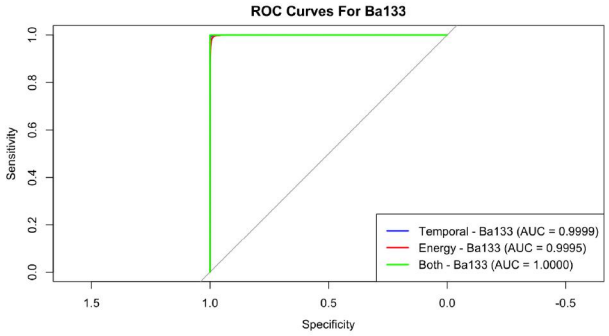
(a) Experiment 1: ROC for model performance on Y88



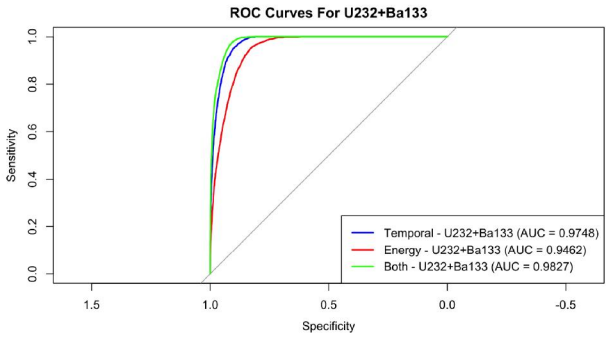
(b) Experiment 1: ROC for model performance on Cs137



(c) Experiment 1: ROC for model performance on U232



(d) Experiment 1: ROC for model performance on Ba133



(e) Experiment 1: ROC for model performance on U232+Ba133

Figure 4. (a–e) display the ROCs associated with the performance of the time, energy, and energy & time models on the different sources. The list-mode data associated with these plots is from Dataset A.

Table 4. The multi-class AUC results computed for each experiment and model type.

Multi-class AUC results			
Experiment:	Time model	Energy model	Energy and time model
1	0.9454	0.9501	0.9713
1-α	0.9081	0.9501	0.9616
2	0.9742	0.7902	0.9870
2-α	0.9099	0.7902	0.8527
3	0.7882	0.9832	0.9857
3-α	0.7830	0.9832	0.9858
4	0.7426	0.9516	0.9574
4-α	0.8275	0.9516	0.9706
5	0.7457	0.9836	0.9848
5-α	0.7765	0.9836	0.9846
6	0.8281	0.9621	0.9718
6-α	0.8215	0.9621	0.9704
7	0.8049	0.9750	0.9801
7-α	0.7983	0.9750	0.9801

extended α model, which accounts for the SNR (different distances), has seemed to overcome the flaws of the original model and provides accurate classification.

To study the impact of distance on the AUC results, we estimate a smooth performance surface by fitting Gaussian-process (GP) regression to the logit of AUC as a function of inverse-squared distances, shown in Figure 5. For each experiment i, we set $y_i = \text{logit}(\text{AUC}_i)$ and inputs $\mathbf{x}_i = (1/r^2_{\text{train},i}, 1/r^2_{\text{test},i})$, rescaled each coordinate to $[0,1]$, and used a squared-exponential kernel

$$k(\mathbf{x},\mathbf{x}') = \sigma^2 \exp\left\{-\sum_j (x_j - x'_j)^2/\theta_j\right\}$$

with a homoscedastic nugget on the logit scale. Kernel hyperparameters $(\theta_1, \theta_2, \sigma^2, \sigma^2_{\text{nugget}})$ are estimated by

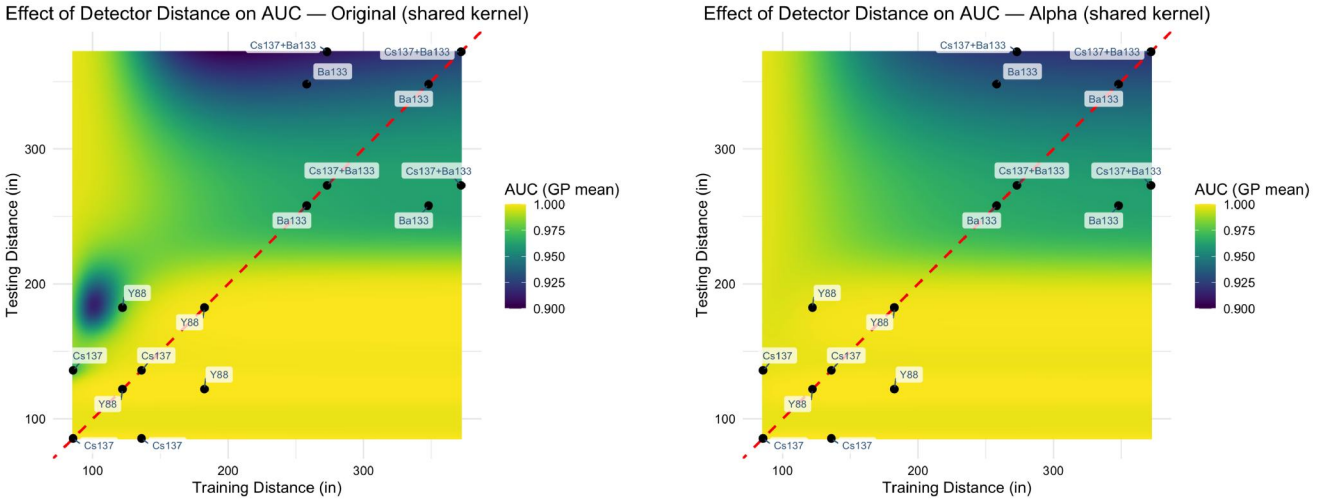


Figure 5. Gaussian-process smoothed AUC surfaces for the original classifier (left panel) and extended–α classifier (right panel) as functions of training and testing distances for each isotope measured. We model logit(AUC) with a squared-exponential Gaussian process in the inverse-square features $(1/r_{\text{train}}^2, 1/r_{\text{test}}^2)$. Kernel hyperparameters are estimated on the original data and held fixed for the extended-α panel. Heatmaps show Gaussian-process posterior means on the same color scale (AUC 0.90–1.00). Black dots indicate the observed experiments, and the red dashed line marks equal training and testing distance.

maximum likelihood on the original model and then held fixed when predicting the extended-α model (to facilitate a shared-kernel comparison) so that any differences in the maps reflect the data, not different smoothness assumptions. This formulation encodes the physical inverse-square law, assumed to hold for signal strength in this application. Consequently, the GP surface recovers the expected monotone degradation with increasing distance and allows us to assess robustness to train–test distance mismatch.

The GP fit uses only 16 points from 4 different experiments, hence the heatmaps are trend visualizations rather than precise estimates far from the sampled region. As such, predictions near the boundaries of the transformed space can revert toward the GP mean. Further, results depend on the logit link and the inverse-square feature choice, and alternative links or kernels are not explored here. Uncertainty bands are additionally not shown but are wider where sampling is sparse, emphasizing our focus on qualitative comparisons between models under a shared kernel.

Under a shared kernel and identical color scale (AUCs between 0.90–1.00), the extended-α model maintains near-unity AUC across a wider range of $(r_{\text{train}}, r_{\text{test}})$ pairs, particularly off the $r_{\text{train}} = r_{\text{test}}$ diagonal, indicating greater robustness to distance misspecification while better preserving the expected inverse-square pattern. On the other hand, the original model struggles to maintain the same pattern as a result of differing training and testing distances, with the greater testing distance to training distance scenario causing AUC jumps due to poor model calibration without accounting for SNR degradation. It is well established that source intensity decreases proportionally to the inverse square of the distance from the detector. Consequently, model performance is expected to decline as the source distance from the detector increases. Our sparse sampling results support this expectation.

5.4. McNemar's Test

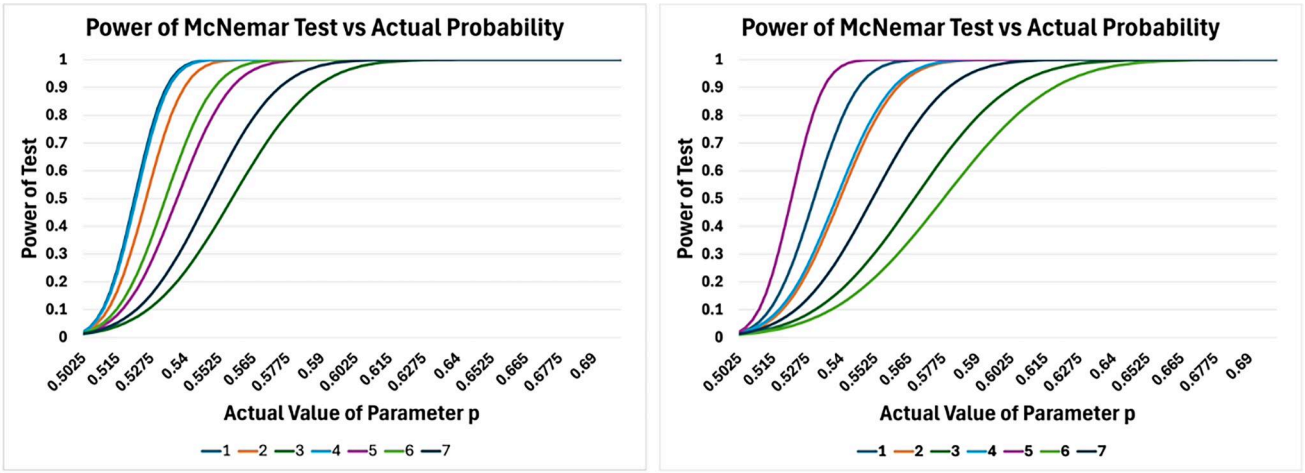
The classifier results seem to indicate that a model that uses both interarrival time and energy channels in order to classify the sources outperforms a model that only uses energy channels. In order to determine whether the two classifiers are significantly different from each other, McNemar's Test is used. McNemar's test has been used and is frequently recommended as a nonparametric test for comparing the accuracy or error rate of two classifiers (Katarzyna Sta¸por 2018; Dietterich 1998).

In order to stick to convention, we decided to set the level of significance, α, to 0.05 (Tamhane and Dunlop 2000). The *P*-values that were obtained for each of the respective experiments that were conducted were nearly zero. In this case, a P-value can be interpreted as the probability of observing at least x samples that were incorrectly classified by the energy model and correctly classified by the model that includes energy and temporal information given that $p = \frac{1}{2}$. Therefore, under the hypothesis test that we conducted, the smaller the P-value the more significant the difference between the two classifiers. This difference would indicate that the energy model performs significantly worse than the model that includes temporal and energy data.

The results for this test indicate that under each experiment the error rate for the energy model is significantly larger than the error rate for the energy & time model. These results seem to indicate that it is beneficial to include temporal information for the classification of radioactive material.

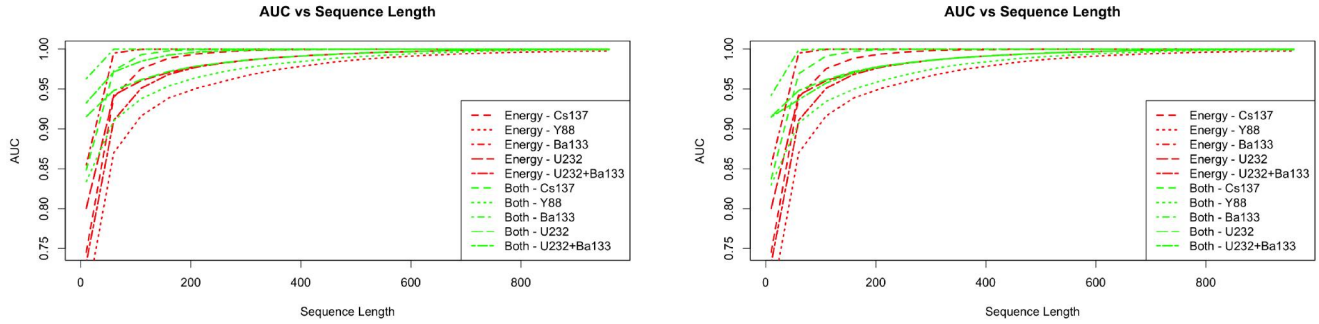
5.4.1. Power Analysis

In order to ensure that the McNemar's tests were able to properly discriminate between the null and alternative hypotheses, we analyzed the power function for each test (Tamhane and Dunlop 2000). The plots of the power function show how the probability of rejecting the null



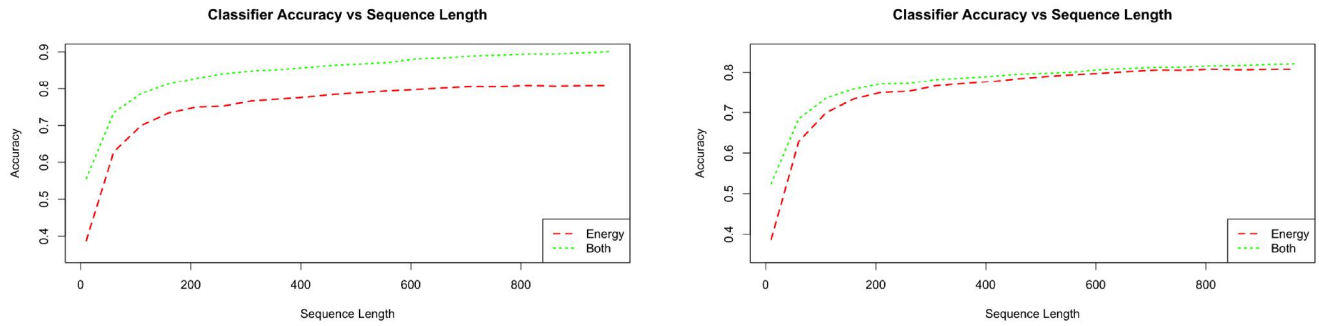
(a) Original Model: Power At Several Values For p (b) α-Model: Power At Several Values For p

Figure 6. (a and b) display the power of the McNemar's test as the value of the p parameter is increased. The colors in each figure are related to the experimental number. (a) shows the results when the original model is used whereas, (b) shows the results when the extended α model is used.



(a) Experiment 1 (b) Experiment 1-α

Figure 7. (a and b) display the impact that sequence length has on the AUC values for each class in Dataset A.



(a) Experiment 1 (b) Experiment 1-α

Figure 8. (a and b) display the impact that sequence length has on the accuracy of each classifier. The data in these figures is from Dataset A.

hypothesis changes as the number of observations that were incorrectly labeled by the energy model and correctly labeled by the energy & time model increases. An increase in p represents an increase in the number of observations that were incorrectly labeled by the energy model and correctly labeled by the energy and time model. Figure 6 displays all plots of power as a function of p. The power plots indicate that the McNemar's test is fairly sensitive to the change in p. These results provide evidence to our claim that the tests were able to effectively discriminate between the null and alternative hypotheses.

5.5. Sequence Length

In the above experiments the test sequences that needed to be classified all had a sequence length of 100. However, it is important to investigate the impact that varying the sequence length has on the performance of the different

models because this relates to dwell time requirements. In order to evaluate the impact that sequence length has on model performance we looked at the effect that sequence length had on the AUC values for each class under each model and the effect that it had on the accuracy of each classifier. Figures 7 and 8 are example plots that display the change in AUC or classifier accuracy as the sequence length is increased. As expected, model performance improves as the sequence length increases. Additionally, the model that relies on both energy and time always outperforms the energy model. Similar results are observed across all of the plots found in Appendix C. In each plot there is some variation due to the dataset and some variation due to whether the extended α model is in use, but the general trend is that the energy and time models outperform the energy models when the sequence length is smaller.

6. Conclusion

In this work, we asked two simple questions, what is the nature of the temporal domain information of list-mode gamma-ray data and is that temporal domain information beneficial to the application of radioisotope identification. The nature of the time domain was first explored through the evaluation of different correlation metrics. The different metrics indicated that there is no evidence of correlations between the time and energy space of our radiation datasets. Additionally, a distributional analysis of our data sets was conducted by inspecting the box plots to identify characteristics of the data sets that may lead to discernible discriminators between radioactive sources. This analysis indicated that the distributions of interarrival times are relatively stable across consistent collection windows and that these distributions show meaningful differences across radioactive sources. Little variation in the distributions for the same source occurred when collection windows increased or decreased.

To evaluate the usefulness of temporal information for RIID, we developed a probabilistic classification method capable of using temporal data, spectral data, or the fusion of both. The classifier assumes independence between the spectral and time models, enabling us to test whether incorporating temporal information improves RIID performance. Using distributional estimates of temporal and spectral domain data, we rigorously assessed the classifier's performance with interarrival time data alone, spectral data alone, and both combined. Performance was evaluated using accuracy, ROC curves, and McNemar tests. Results showed that, in all cases, including temporal information improved classification performance.

A critical aspect of our framework is that the marked point process model naturally highlights the interconnectedness of spectral and temporal components. For computational efficiency, however, we estimated their parameters independently. Our results show that fuzing spectral and temporal information far exceeds the performance of either modality alone. Moreover, our choice of a kernel density estimator for the spectral model is not essential; it can be substituted with more sophisticated approaches such as full-spectrum matching, non-negative matrix factorization, or neural network classifiers. Regardless of the spectral estimator chosen, this article has shown that temporal information provides additional classification value beyond what is currently available in the field. This underscores the robustness of our approach and its potential for integration with both classical and modern RIID methodologies.

Several avenues for future research emerge from this work. Extending the probabilistic classifier to address source separation, applying these methods to non-stationary sources, determining efficacy of temporal information on sources with spectrally similar peaks, and exploring alternative metrics beyond interarrival time to identify discriminative temporal features for classification are promising directions. Finally, extending the classifier framework to higher dimensional representations of list-mode data—such as interarrival time distributions within individual energy bins, between pairs of bins, or across larger groupings—could further enhance the discrimination of temporal information content of the source.

The results of this study are promising and highlight the potential for further exploration of temporal domain information in RIID applications. GADRAS (Mitchell et al. 2014) is the *defacto* standard for RIID algorithms, with nearly 40 years of continuous development and refinement, and we do not claim any comparison or advancements over GADRAS. Instead, we have provided evidence supporting the potential benefits of incorporating temporal information into RIID algorithms. Our findings warrant further investigation and offer a foundation for developing new methods or enhancing existing ones with temporal data to improve RIID for the broader community.

Notes

1. The data that support the findings of this study are available from the corresponding author, AJH, upon reasonable request.
2. Classifying multiple isotopes within a single sample is a promising extension of the probabilistic classifier but is beyond the scope of this preliminary work.

Acknowledgments

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Aaron J. Hill http://orcid.org/0000-0003-3573-7435

References

Bilton KJ et al. 2019. Non-negative matrix factorization of gamma-ray spectra for background modeling, detection, and source identification. IEEE Trans Nucl Sci. 66(5):827–837. https://doi.org/10.1109/TNS.2019.2907267

Clarkson E, Kupinski M. 2020. Quantifying the loss of information from binning list-mode data. J Opt Soc Am A Opt Image Sci Vis. 37(3):450–457. https://doi.org/10.1364/JOSAA.375317

Daley DJ, Vere-Jones D. 2003. An introduction to the theory of point processes: volume I: elementary theory and methods. Springer.

Dietterich TG. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput. 10(7):1895–1923. https://doi.org/10.1162/089976698300017197

Francis Sahngun N. 2022. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 75(1):25–36.

Gilmore G, Joss D. 2008. Practical gamma-ray spectrometry. John Wiley & Sons.

Glenn FK. 2010. Radiation detection and measurement. John Wiley & Sons.

Hand DJ, Till RJ. 2001. A simple generalisation of the area under the roc curve for multiple class classification problems. Mach Learn. 45:171–186.

High resolution gamma-ray spectrometry analyses for normal operations and radiological incident response. Technical Report EPA 402-B-17-001, U.S. Environmental Protection Agency, Office of Radiation and Indoor Air (ORIA), National Analytical Radiation Environmental Laboratory (NAREL), Montgomery, AL, October 2019. Revision 0.

Ho Park S, Goo JM, Jo C-H. 2004. Receiver operating characteristic (roc) curve: practical review for radiologists. Korean J Radiol. 5(1):11–18.

Jay Conover W. 1999. Practical nonparametric statistics, volume 350. John Wiley & Sons.

Jeffcoat R, Salaymeh S, Clare A. 2010. A comparison of GADRAS-simulated and measured gamma-ray spectra. In Proceedings of the 51st Annual Meeting of the Institute of Nuclear Materials Management (INMM). Institute of Nuclear Materials Management.

Sta¸por K. 2018. Evaluating and comparing classifiers: review, some recommendations and limitations. In Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017 10, pp. 12–21. Springer.

Mirion Technologies. 2025. Spir-ace radionuclide identification device (RIID) with quantitative assessment capability; [accessed 2025 Oct 20]. https://www.mirion.com/products/technologies/defense-security-systems/security-search-systems/handheld-radioisotope-identification-devices/spir-ace-radionuclide-identification-device-riid-with-quantitative-assessment-capability

Mitchell DJ, Harding L, Thoreson GG, Horne SM. 2014. Gadras detector response function. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States)

Moore ET, Ford WP, Hague EJ, Turk J. 2019. An application of cnns to time sequenced one dimensional data in radiation detection. In Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXV, volume 10986, pp. 398–408. SPIE. https://doi.org/10.1117/12.2519037

Moore ET, Turk JL, Ford WP, Hoteling NJ, McLean LS. 2020. Transfer learning in automated gamma spectral identification. arXiv Preprint arXiv:2003.10524.

Pakari OV et al. 2024. Gamma-ray spectroscopy in low-power nuclear research reactors. J Nucl Eng. 5(1):26–43.

Robin X et al. 2011. Proc: an open-source package for r and s + to analyze and compare roc curves. BMC Bioinf. 12:1–8.

Shumway RH, Stoffer DS. 2005. Time series analysis and its applications (springer texts in statistics). Springer-Verlag.

Sullivan CJ, Martinez ME, Garner SE. 2006. Wavelet analysis of sodium iodide spectra. IEEE Trans Nucl Sci. 53(5):2916–2922. https://doi.org/10.1109/TNS.2006.881909

Tamhane A, Dunlop D. 2000. Statistics and data analysis: from elementary to intermediate. Prentice-Hall.

Tom B, Michael H. 2009. Radio-isotope identification algorithms for nai γ spectra. Algorithms. 2(1):339–360.

U.S. Environmental Protection Agency. 2025. Radiation and shipping port security. [accessed 2025 Oct 20]. https://www.epa.gov/radtown/radiation-and-shipping-port-security.

Wilderman SJ, Clinthorne NH, Fessler JA, Hua C-H, Leslie Rogers W. 2000. List Mode em Reconstruction of Compton Scatter Camera Images in 3-d. In 2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No. 00CH37149), Volume 2, pp. 15–292. IEEE.

Williford RS. 2013. *Temporal gamma-ray spectrometry to quantify relative fissile material content*. [PhD thesis]. University of Tennessee.

Appendix A. Dataset details

Table A1. Distances used for each source to achieve specific count rate targets.

	Distance 1		Distance 2	
Source	Distance	Gross count rate	Distance	Gross count rate
Cs-137	10ft 2in	1598	15ft 2.5in	1181
Y-88	7ft 1.5in	1738	11ft 4in	1230
Ba-133	21ft 6in	1827	29ft	1381
Cs-137 & Ba-133	22ft 9in	1849	31ft	1350
Background		751		

Table A2. MCA settings for the August 2024 collection.

Amplifier	Gain	0.4
	Shaping Time	2 μ sec
ADC	Lower Level Disk	17
	Upper Level Disk	1023
High voltage	Target	666

Appendix B. AUC values

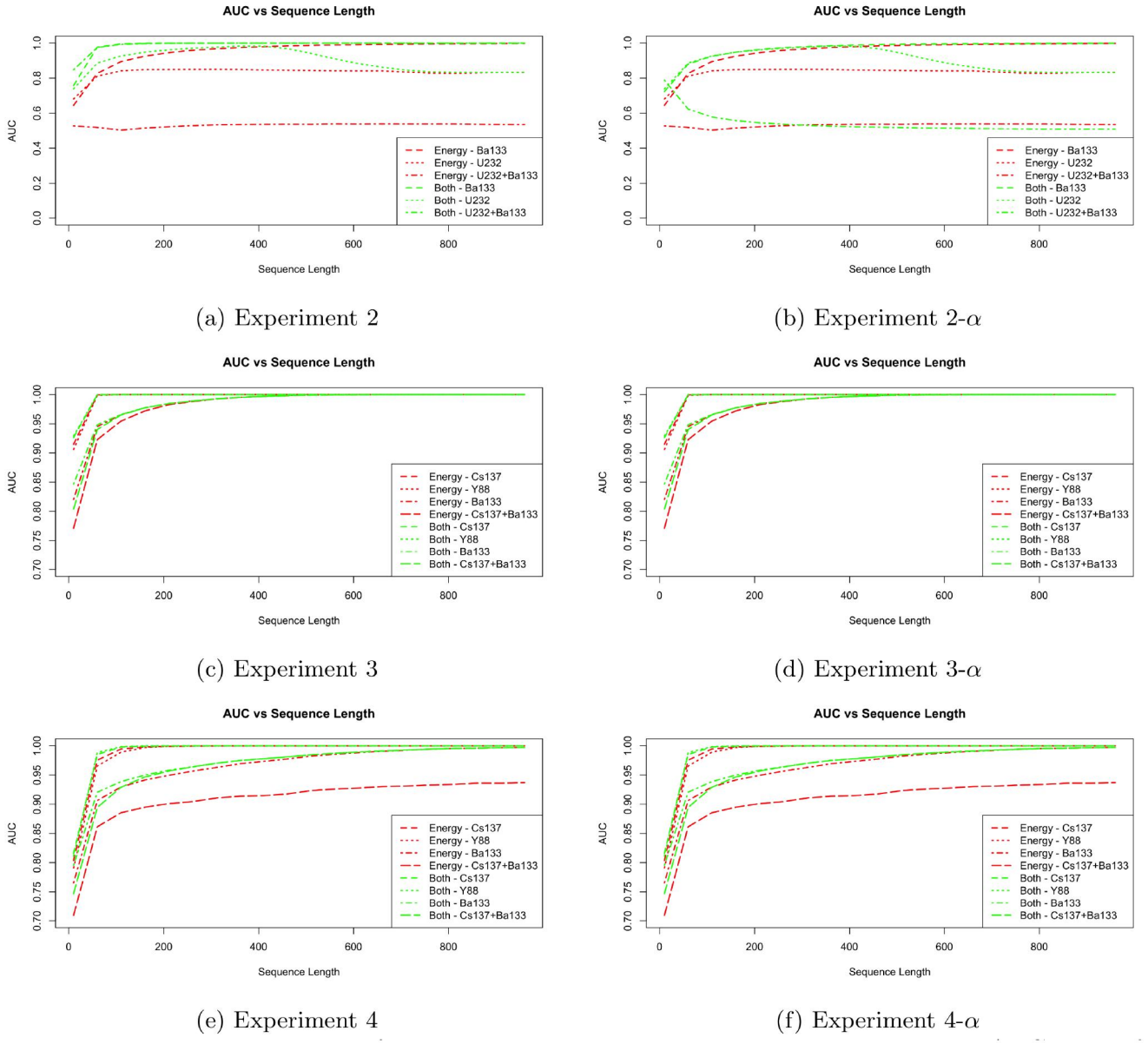
Table B1. The table displays the AUC results that were obtained for each experiment where the distribution of the training set and the distribution of the test set were the same. The presence of α after the experimental number indicates that the extended α model was tested using the experimental setup associated with that number.

		AUC results		
Experiment:	Source	Time model	Energy model	Energy and time model
1	Y88	0.8656	0.9079	0.9307
	Cs137	0.9560	0.9703	0.9896
	U232	0.9748	0.9571	0.9827
	Ba133	0.9999	0.9995	1
	U232 + Ba133	0.9748	0.9462	0.9827
1-α	Y88	0.8601	0.9079	0.9274
	Cs137	0.9375	0.9703	0.9877
	U232	0.8252	0.9571	0.9593
	Ba133	0.9495	0.9995	0.9999
	U232 + Ba133	0.9748	0.9462	0.9541
2	U232	0.9390	0.8328	0.9628
	Ba133	0.9619	0.8755	0.9934
	U232 + Ba133	0.9958	0.5051	0.9918
2-α	U232	0.7168	0.8328	0.9182
	Ba133	0.9268	0.8755	0.9117
	U232 + Ba133	0.9958	0.5051	0.5811
3	Y88	0.6664	1	1
	Cs137	0.8119	1	1
	Ba133	0.7278	0.9644	0.9650
	Cs137 + Ba133	0.7351	0.9525	0.9636
3-α	Y88	0.6500	1	1
	Cs137	0.8022	1	1
	Ba133	0.7278	0.9643	0.9650
	Cs137 + Ba133	0.7351	0.9524	0.9636
6	Y88	0.7418	0.9947	0.9993
	Cs137	0.8276	0.9964	0.9990
	Ba133	0.7879	0.9314	0.9364
	Cs137 + Ba133	0.7834	0.9050	0.9245
6-α	Y88	0.7296	0.9947	0.9990
	Cs137	0.8071	0.9964	0.9987
	Ba133	0.7879	0.9314	0.9359
	Cs137 + Ba133	0.7829	0.9050	0.9183
7	Y88	0.6974	0.9990	1
	Cs137	0.8207	0.9995	0.9999
	Ba133	0.7620	0.9487	0.9529
	Cs137 + Ba133	0.7444	0.9332	0.9477
7-α	Y88	0.6811	0.9990	1
	Cs137	0.8039	0.9995	0.9999
	Ba133	0.7620	0.9487	0.9526
	Cs137 + Ba133	0.7444	0.9332	0.9481

Table B2. The AUC results that were obtained for each experiment where the distribution of the training set was different from that of the test set. The presence of α after the experimental number indicates that the extended α model was tested using the experimental setup associated with that number.

AUC results				
Experiment:	Source	Time model	Energy model	Energy and time model
4	Y88	0.4871	0.9862	0.9760
	Cs137	0.6268	0.9920	0.9713
	Ba133	0.7896	0.9256	0.9322
	Cs137 + Ba133	0.7836	0.8810	0.9074
4-α	Y88	0.7349	0.9862	0.9977
	Cs137	0.8295	0.9920	0.9972
	Ba133	0.7896	0.9257	0.9352
	Cs137 + Ba133	0.7836	0.8811	0.9233
5	Y88	0.6480	1	1
	Cs137	0.7897	1	1
	Ba133	0.7313	0.9627	0.9622
	Cs137 + Ba133	0.5594	0.9560	0.9616
5-α	Y88	0.6427	1	1
	Cs137	0.7722	1	1
	Ba133	0.7314	0.9627	0.9627
	Cs137 + Ba133	0.7361	0.9560	0.9601

Appendix C. Effects of sequence length



(a) Experiment 2

(b) Experiment 2-α

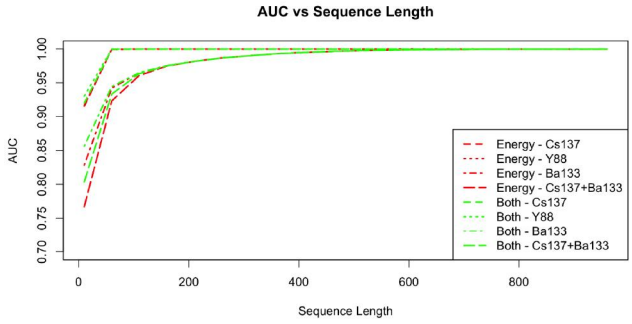
(c) Experiment 3

(d) Experiment 3-α

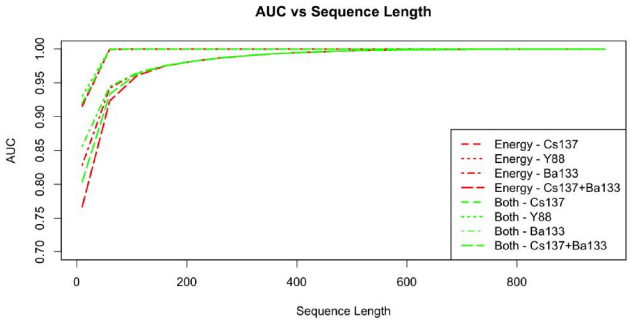
(e) Experiment 4

(f) Experiment 4-α

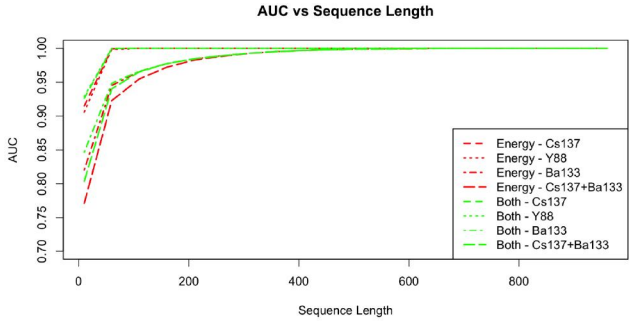
Figure C1. (a–f) display the impact that sequence length has on the AUC values for each class in the respective datasets.



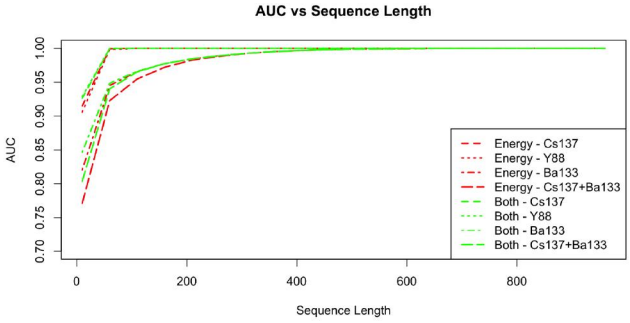
(a) Experiment 5



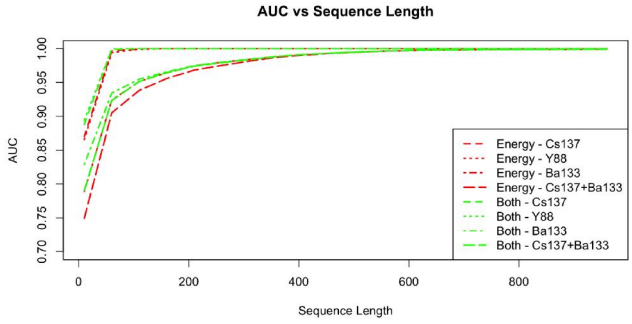
(b) Experiment 5-α



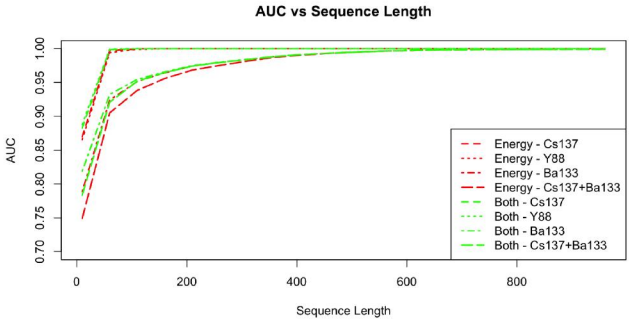
(c) Experiment 6



(d) Experiment 6-α

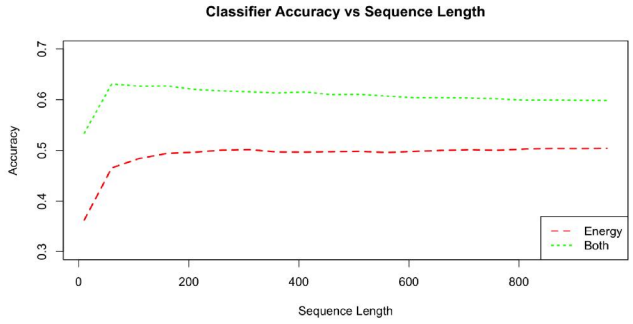


(e) Experiment 7

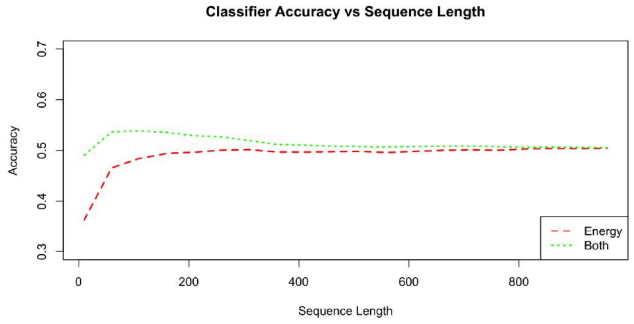


(f) Experiment 7-α

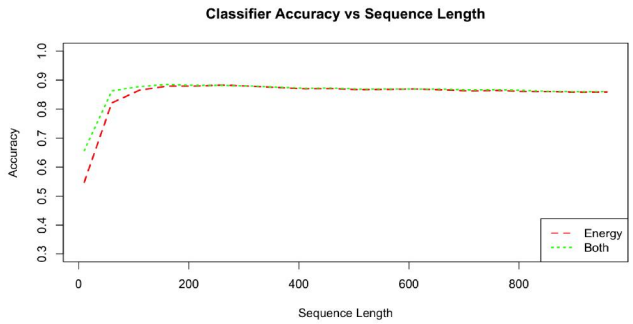
Figure C2. (a–f) display the impact that sequence length has on the AUC values for each class in the respective datasets.



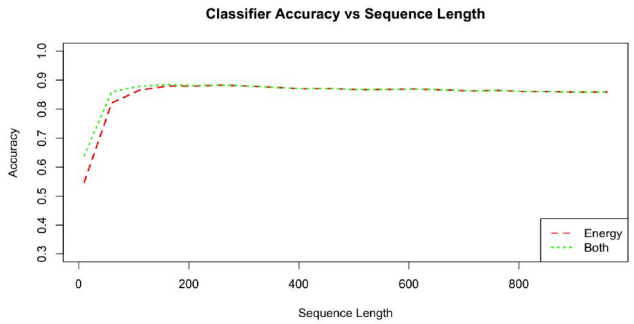
(a) Experiment 2



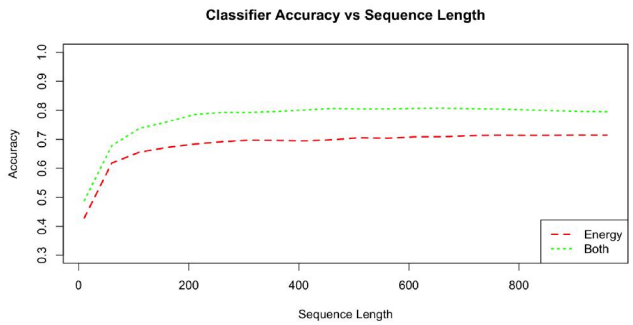
(b) Experiment 2-α



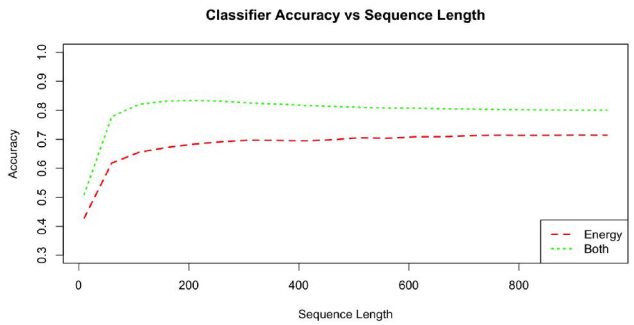
(c) Experiment 3



(d) Experiment 3-α

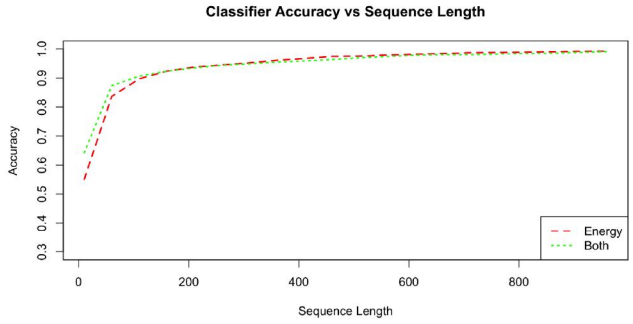


(e) Experiment 4

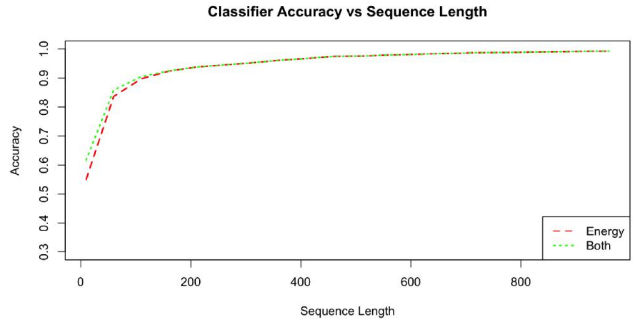


(f) Experiment 4-α

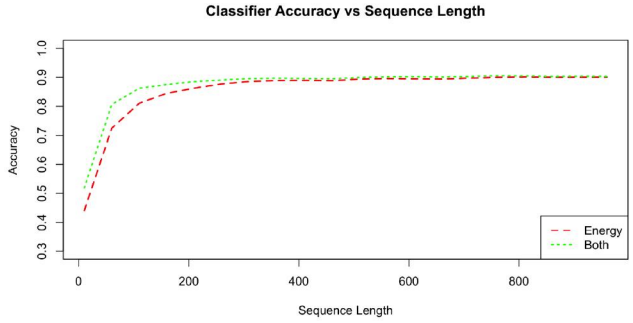
Figure C3. (a–f) display the impact that sequence length has on the accuracy of each classifier.



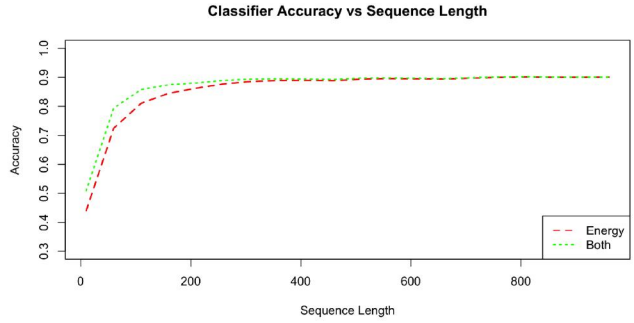
(a) Experiment 5



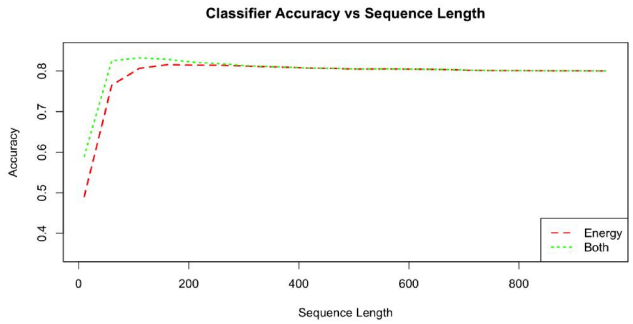
(b) Experiment 5-α



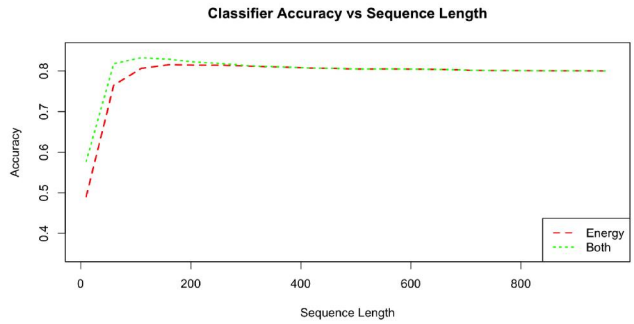
(c) Experiment 6



(d) Experiment 6-α



(e) Experiment 7



(f) Experiment 7-α

Figure C4. (a–f) display the impact that sequence length has on the accuracy of each classifier.