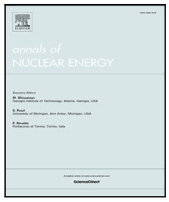
A variational autoencoder for minimally-supervised pulse shape discrimination

Annals of Nuclear Energy 204 (2024) 110496

Contents lists available at ScienceDirect

Annals of Nuclear Energy

journal homepage: www.elsevier.com/locate/anucene



A variational autoencoder for minimally-supervised pulse shape discrimination

Abdullah Abdulaziz [a,*], Jianxin Zhou [b], Ming Fang [b], Stephen McLaughlin [a], Angela Di Fulvio [b], Yoann Altmann [a]

[a] School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK
[b] Department of Nuclear, Plasma, and Radiological Engineering, University of Illinois at Urbana-Champaign, Urbana, USA

ARTICLE INFO

Dataset link: 10.17861/73b6ec57-a691-4337-8781-2087459dd48a

Keywords:
Variational autoencoder
Pulse shape discrimination
Semi-supervised classification
Organic scintillators

ABSTRACT

We propose a novel approach based on variational autoencoder with Gaussian mixture latent space (GMVAE) to address the challenging problem of pulse shape discrimination (PSD) in organic scintillators, in the presence of pile up. Unlike deterministic charge integration, which is very sensitive to pulse-processing parameters, the GMVAE performances are robust against variations of the hyperparameters. When compared to other supervised machine learning methods, GMVAE requires the fewest training pulses (100) to achieve a classification accuracy within 2% of its optimum performance, *i.e.*, 98.3% accuracy. GMVAE exhibited excellent classification despite the difference in energy spectra between the training and test data sets, which were 14.1 MeV neutron pulses and ^{239}PuBe pulses, respectively. While requiring minimum supervision, GMVAE showed superior PSD performances compared to both deterministic and supervised machine learning approaches. GMVAE is hence particularly suitable for real-time pulse classification, where expert labeling is unavailable and fine tuning of the discrimination parameters impractical.

1. Introduction

Ionizing radiation interactions within detectors exhibit time-dependent energy deposition profiles that result in detectable pulses. Pulse rise and decay time constants depend on the energy deposited per unit path, *i.e.*, the linear energy transfer (LET) of the charged recoils produced upon ionizing radiation interaction with the detector materials. Among detectors, the pulse shape of organic scintillators exhibits a strong dependence on particles with different LET (Brooks, 1979; Morishita et al., 2019; Zhou et al., 2023). This property makes them suitable to perform pulse shape discrimination (PSD), enabling the classification of pulses generated by different radiation types. While one can select materials with excellent PSD capabilities, accurate pulse classification remains a challenging task because of the unavoidable degradation of the PSD as the deposited energy decreases and the presence of inherent measurement nuisance sources, including pulse pile-up and electronic noise.

PSD in mixed fields with both neutron and gamma-ray emitters is a typical classification problem with a large number of unlabeled points and only a small fraction of the pulses that can be labeled with high confidence. In this work, we leveraged advancements in inference statistics and machine learning (ML) to maximize the classification accuracy. We enhance our latest work (Fu et al., 2018; Abdulaziz et al., 2022) and propose a novel semi-supervised approach using a variational autoencoder with Gaussian mixture latent space (GMVAE) for solving the PSD problem in the presence of pile-up pulses. We demonstrated it on pulses measured using a stilbene-d_{12} organic scintillation crystal.

Pile up occurs when multiple particles hit the detector in quick succession, causing their signals to overlap in time. Pile-up pulses may contain contributions from multiple particles, making it difficult to distinguish individual signals. Therefore, pile-up pulses can significantly compromise the classification accuracy of a measured data set, particularly when traditional methods such as gated charge integration (CI) (Bourne et al., 2017) are used to analyze the pulses. At high count rates (100,000 detected pulses per second), neutron counts can be overestimated up to two orders of magnitude if pile-up pulses are not identified and rejected, and underestimated by 40% when pile-up pulses are rejected using deterministic methods (Bourne, 2016). The GMVAE (Shu, 0000; Figueroa, 2019; Collier and Urdiales, 2019; Charakorn et al., 2020; Varolgüneş et al., 2020) is particularly suitable for PSD because it performs clustering based on training on a low-dimensional space in which the different types of pulses are characterized by a Gaussian mixture model. Moreover, by adopting a semi-supervised learning approach, the GMVAE can learn discriminant features from both labeled and unlabeled pulses, thus improving the

* Corresponding author.
E-mail address: a.abdulaziz@hw.ac.uk (A. Abdulaziz).

https://doi.org/10.1016/j.anucene.2024.110496
Received 27 September 2023; Received in revised form 16 March 2024; Accepted 19 March 2024
Available online 15 April 2024

classification accuracy in unknown fields (Figueroa, 2019; Figueroa and Rivera, 2017; Abdulaziz et al., 2022). Recently, machine learning approaches have been deployed to address the PSD problem, showing promising results in discriminating between individual, pile-up-free pulses (Esposito et al., 2004; Liu et al., 2009; Tambouratzis et al., 2013; Caldwell et al., 2015). However, when pile-up pulses are included in the test data set, multiple user parameters are needed and the model training needs to be fully supervised (Fu et al., 2018). To minimize the pulse misclassification in PSD, several deterministic and statistical algorithms have been proposed (Zhou et al., 2023; Monterial et al., 2015; Marrone et al., 2002; Yousefi et al., 2009; Nakhostin, 2019). Among these approaches, we selected the most commonly used one, *i.e.*, CI, as the performance comparison term of our method. Our novel GMVAE approach greatly increases the practical feasibility of ML-based PSD by minimizing the need for user supervision while achieving a comparable or better classification accuracy than traditional or ML-based methods.

Our approach can, therefore, be of general interest in many applications where radiation signature analysis relies on PSD, and minimum supervision is possible, especially in the safeguarding of nuclear materials and nuclear security (Di Fulvio et al., 2017, 2018), where fission neutron signatures are often obscured by an intense gamma-ray background.

The remainder of this paper is organized as follows. In Section 2, we introduce the variational autoencoder (VAE) and the adopted semi-supervised learning approach. In Section 3, we describe the experimental setup used to acquire the data needed to train, test, and validate the proposed model. In Section 4, we present the state-of-the-art algorithms used in this study to benchmark the GMVAE-based approach. The results are discussed in Section 5, and finally, we conclude with the key findings and future directions in Section 6.

2. Methodology

2.1. Variational autoencoder

A variational autoencoder (VAE) is a generative model renowned for its ability to learn efficient, compact, and expressive representations of data (Kingma and Welling, 2014, 2019). Its architecture comprises three main components: an encoder, a latent space, and a decoder. The encoder compresses input data into a latent space – a lower-dimensional representation – while the decoder reconstructs the data from this latent representation. The goal of VAE training is to optimize the model such that it maximizes the likelihood of accurately reconstructing the input data, alongside enforcing a regularization that ensures the latent space adheres to a predefined distribution, typically Gaussian. This process allows for the generation of new data samples that are similar to the original data set by sampling from the latent space (Kingma and Welling, 2014, 2019).

A Gaussian mixture VAE (GMVAE) is a variant of VAE that uses a Gaussian mixture model (GMM) as the prior distribution of the latent space (Shu, 0000; Figueroa, 2019; Collier and Urdiales, 2019; Charakorn et al., 2020; Varolgüneş et al., 2020). This prior information allows the GMVAE to model complex data distributions and capture intricate patterns, such as multiple modes or classes, in the data, compared to VAE.

GMVAEs are powerful tools for clustering or semi-supervised learning, as they can identify low-dimensional spaces in which different types/classes of data points are characterized by different Gaussian distributions. Unlike standard VAEs, which contain single continuous (multivariate) latent variables, a GMVAE also contains a discrete latent variable representing the data class or cluster. By assigning different integers/labels to different objects or classes, GMVAEs can provide the so-called membership probabilities of any observed sample for all the pre-defined classes, naturally enabling object classification. In our case, we use three classes: gamma, neutron, and pile-up pulses.

2.2. Semi-supervised learning

In scenarios where no labeled data are available, the GMVAE can divide the set of pulses into different groups by minimizing a cost function. This cost function, referred to as the negative ELBO (Abdulaziz et al., 2022), comprises both the likelihood of the training pulses and a term that structures the latent space using a mixture of Gaussians prior. However, when some labeled data become available, they can be integrated into the model through additional loss functions to enhance classification accuracy (Figueroa, 2019; Figueroa and Rivera, 2017). In essence, these functions quantify the disparity between the predicted class and the actual one for each pulse.

Assume that we have a data set encompassing pulses, of which only a subset are known to be gamma-ray, neutron, or pile-up pulses. In this case, the training data set is partially labeled. We begin by assigning labels to unlabeled data based on the k-nearest neighbors (k-NN) rule (Cover and Hart, 1967). As data points belonging to the same class are expected to share similar features in the latent space, we classify each unlabeled sample by the majority label of its k-NN in the latent space. We use $k = 1$ in our experiments and employ the Euclidean distance as the distance metric (Abdulaziz et al., 2022). In doing so, the first auxiliary loss function, namely the labeling loss, is the cross entropy between the assigned labels and the unsupervised predicted labels (Goodfellow et al., 2016).

The k-NN classification depends on how distances are computed between different samples. Therefore, the effectiveness of k-NN classification can be significantly enhanced by learning an appropriate distance metric from the assigned labels. Here, we adopted the triplet embedding loss as a secondary auxiliary loss function to regularize the latent space (Schroff et al., 2015). This triplet loss penalizes large distances between features belonging to the same class while also penalizing small distances between samples with non-matching labels. The final objective function for the semi-supervised GMVAE is formulated as:

$$\mathcal{L}_{\text{GMVAE}} = \text{ELBO} + \omega\mathcal{L}_{\text{label}} + \gamma\mathcal{L}_{\text{triplet}}, \tag{1}$$

where $\mathcal{L}_{\text{label}}$ is the labeling loss, $\mathcal{L}_{\text{triplet}}$ is the triplet embedding loss and $\omega, \gamma > 0$ are the hyper-parameters controlling the weights of the auxiliary loss functions. Using grid search (Hastie et al., 2009b), we found that $\omega = 50$ and $\gamma = 50$ yield optimal results in our experiments. The same parameters are used for all experiments. However, parameter tuning does not seem critical and setting all hyper-parameters to one does not lead to significant classification performance degradation.

2.3. Proposed architecture

The developed GMVAE architecture consists of an encoder, a latent space, and a decoder. The encoder is designed to understand and learn the distinct features of the input pulses. It is built with three fully-connected layers with respective sizes of L, $L/2$, and $L/4$, each utilizing a rectified linear unit (ReLU) for activation. This helps in translating the intricate patterns in the data into a form that is easier to work with. See Fig. 1 for a detailed schematic of the proposed GMVAE.

In the latent space, there is a classification module with a layer that is as large as the number of classes. This module analyzes the learned features from the encoder to allocate each data point to a predicted class based on its characteristics. Following this classification, it further refines the data representation by determining the mean and variance of the latent features, a process that integrates information from the predicted classes and previous layers.

Finally, the decoder part of the architecture aims to reconstruct the output based on the refined data representations in the latent space. It consists of three fully-connected layers with respective sizes of $L/4$, $L/2$, and L, where the initial two layers use a ReLU activation function, and the last layer employs a logistic (sigmoid) function to deliver the final output vector.

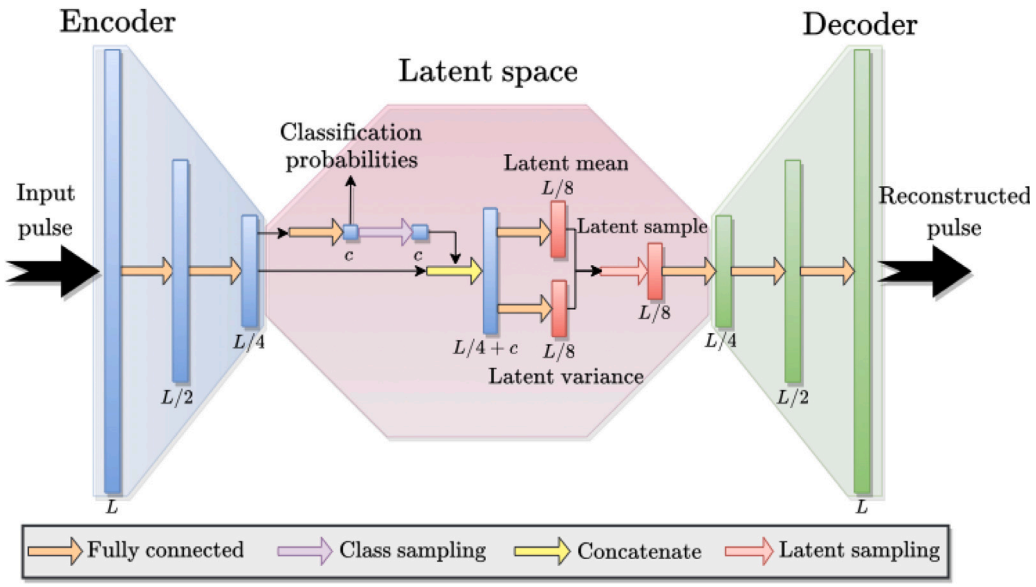


Fig. 1. Schematic of the proposed GMVAE. The color-coded arrows indicate different neural network operations, as detailed in the legend, with output dimensions of each layer specified within the schematic. Here, L denotes the pulse length, and c represents the number of classes. The encoder processes the input pulse down to a lower-dimensional latent space. The classification module at the end of the encoder comprises a fully-connected layer of size c producing logits. Softmax activation applied to these logits yields "Classification probabilities". "Class sampling" utilizes the Gumbel Softmax distribution trick (Jang et al., 2016; Maddison et al., 2017) to select a class, allowing for differentiable sampling."Latent sampling" draws from the latent distribution, defined by the latent mean and variance of dimensionality $L/8$, to generate "latent samples/vectors" used for reconstruction by the decoder.

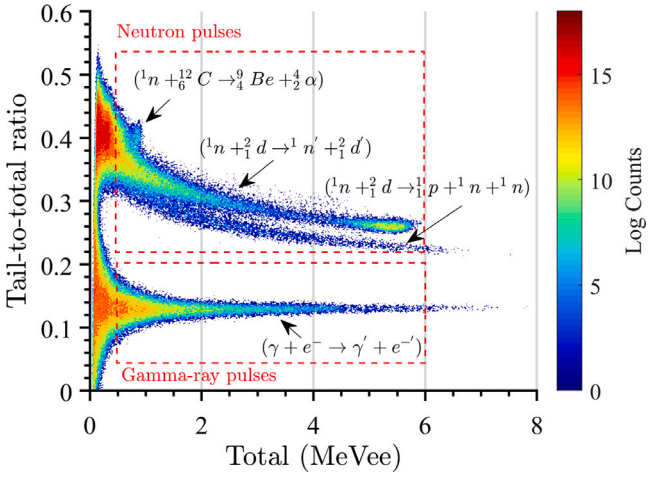


Fig. 2. PSD scatter-density plot produced by the stilbene-d_{12} detector irradiated by a DT source with the detail of the pulse-producing reactions in the detector.

In our experiments, we sampled 592-ns long waveforms at a frequency of 500 MSps, therefore, $L = 296$, however, the proposed architecture can accommodate pulses of any length. The GMVAE was trained over 100 epochs using the Adam optimizer (Kingma and Ba, 2015) with a learning rate 10^{-4} and a batch size 100.

3. Data acquisition

In this section, we briefly describe the experimental acquisition methods and the data sets used to train and test the performance of the proposed algorithm. The data are pulses acquired with a 140 cm^3 deuterated trans-stilbene (stilbene-d_{12}) detector (Zhou et al., 2022). It is a recently developed organic scintillation detector with excellent PSD capability for fast-neutron detection applications (Zhou et al., 2022; Gaughan et al., 2021). The detector is coupled to a HAMAMATSU H6559 photomultiplier assembly, and the pulses are collected by a CAEN DT5730 14-bit 500 MSps desktop digitizer, equipped with the DPP-PSD firmware (SpA, 2020). The digitizer was used in the "waveform" mode, which samples the signal every 2 ns and features an input range of 0–2 V with a low-level threshold of 0.002 V, corresponding to a 15 keVee light output. The total record length for each pulse is 592 ns. In this work, we used two data sets of stilbene-d_{12} pulses to develop and validate our GMVAE. The first data set includes synthetic pulses generated from pulse templates obtained from measured data. The other data set includes measured pulses under controlled irradiation conditions. The creation of these two data sets is discussed in further detail in the following sections. The data sets supporting this work are available at the following DOI: 10.17861/73b6ec57-a691-4337-8781-2087459dd48a.

3.1. Synthesized data from measured pulse templates

We synthesized three types of stilbene-d_{12} pulses: neutron pulses, gamma-ray pulses, and pile-up pulses. The neutron and gamma-ray pulses were generated based on template pulses measured by the stilbene-d_{12} detector, and the pile-up pulses were generated by combining any of these pulse types.

The pulse templates were obtained by irradiating the detector with a 14.1 MeV deuterium-tritium (DT) neutron source, *i.e.*, the ThermoFisher MP320 DT neutron generator. Then, we applied the PSD-CI method to classify the pulses. In CI, the pulses are represented in a scatter-density plot, with the PSD parameter as a function of the pulse integral. By examining the distribution of points on the DT scatter-density plot, it is possible to identify and classify different radiation types based on their distinct patterns of the PSD parameters. Further details on the PSD-CI are described in Section 4.1. In the stilbene-d_{12} detector, 14.1 MeV neutrons can undergo three main interactions that generate detectable pulses: neutron–deuteron elastic scattering, neutron-induced deuteron break-up, and neutron-induced carbon break-up (Zhou et al., 2022). The D-T neutrons can also interact with the materials of the generator and produce high-energy gamma rays. These gamma-ray pulses were used to produce gamma-ray templates. Fig. 2 shows the PSD scatter-density plot of the pulses measured by the stilbene-d_{12} detector irradiated by the DT generator

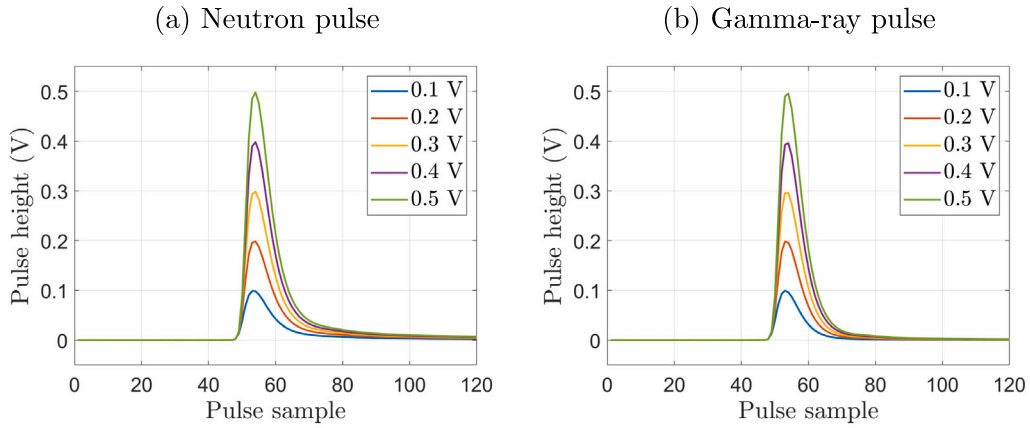


Fig. 3. Neutron (a) and gamma-ray (b) pulse templates with amplitudes from 0.05 V to 0.5 V.

and relative reactions occurring in the detector. Fig. 2 shows that the neutron and gamma-ray pulses with a light output higher than 500 keVee are generally well separated and easier to classify compared to lower amplitude pulses.

The selection of the data to be used for ground truth in machine learning algorithms is crucial because it needs to represent the domain of the data well without overfitting it. One additional concern in PSD is the practical challenge of obtaining "pure neutron" data sets, owing to the nature of neutron interactions with matter, such as inelastic scattering reactions that produce gamma rays. Even data sets based on time-selected neutron pulses, which are classified based on their slower speed compared to gamma rays in time-of-flight experiments, cannot be assumed free of spurious gamma rays. Considering these challenges and the excellent inherent PSD of stilbene d_{12}, we applied the CI-based discrimination with a relatively high light output threshold to create the template data set.

We classified neutron and gamma-ray pulses in the 500 keVee to 6 MeVee light output range. We sorted and binned them by amplitude in ten bins (bin width 0.045 V), ranging from 0.05 to 0.5 V. Pulses in each energy bin were averaged to obtain a pulse template. Through this method, we generated ten pulse templates with discrete amplitudes from 0.05 V to 0.5 V for both neutron and gamma-ray pulse types. Fig. 3 shows the ten neutron pulse templates. From the linear interpolation of templates with discrete amplitudes, we synthesized neutron pulses with continuous amplitudes in the range of 0.05 V to 0.5 V. The gamma-ray pulses were generated using the same method.

After obtaining the neutron and gamma-ray pulses, we generated the pile-up pulses by randomly combining pairs of pulses. The amplitudes of the two pulses were randomly chosen between 0.05 V and 0.5 V, and the time interval between the two pulses was sampled according to the probability $P(r,t)$ of detecting a pile-up pulse within time interval t and detector count rate r, as in Eq. (2) (Knoll, 2010).

$$P(r,t) = e^{(-rt)} \times (1 - e^{(-rt)}) \tag{2}$$

We trained and tested our model using the synthetic pulses normalized to their peak values, with each pulse's maximum amplitude normalized to one. This normalization aids in analyzing results across different noise levels, specifically examined at $\sigma = 0.001, 0.01, 0.05$. Fig. 4 illustrates a pile-up pulse with no noise alongside one with the highest noise level ($\sigma = 0.05$). We generated 120,000 pulses (40,000 per class) for training and 480,000 pulses (160,000 per class) for testing. As the pulse types for these synthetic data sets were known – serving as ground truth labels – we leveraged this data to quantitatively evaluate classification accuracy, comparing GMVAE predictions to the ground truth.

3.2. Measured data

Besides the template-generated pulses, we measured pulses from a 1-Ci neutron-gamma ^{239}PuBe source (neutron energy up to 12 MeV) and a 30 μCi gamma-ray ^{137}Cs source with the stilbene-d_{12} detector to test the classification model. Fig. 5 shows the experimental setup. The digital pulse acquisition chain for this experiment has been described earlier in this section. In this measurement, the low-level energy threshold was 0.002 V (15 LO), the calibration gain was 758.7 keVee/V. The same gain was used for the ^{239}PuBe measurement and the DT measurement described in the previous section. We placed the detector 120 cm away from the ^{239}PuBe source and placed the ^{137}Cs source at nine different locations: 30 cm, 40 cm, 50 cm, 60 cm, 70 cm, 80 cm, 90 cm, 100 cm, and 110 cm away from the detector front surface. At each ^{137}Cs position, we recorded pulses for ten minutes. Table 1 reports the total number of counts as a function of the source-to-detector distances. The neutron count rate is expected to remain constant because the distance between the detector and the ^{239}PuBe source is fixed. However, as the gamma-ray ^{137}Cs source is closer to the detector, the probability of pile-up pulses is expected to increase and so does the misclassification rate of pile-up pulses. Pile-up pulses are likely to be misclassified as neutron pulses because the second pulse within the acquisition window can be misinterpreted as an increase of the pulse slow-decay component, which characterizes neutron-induced pulses. The neutron count rate can be used to compare the robustness of the tested algorithms against increased pile-up probability.

4. Competing algorithms

The proposed GMVAE's accuracy is compared to the PSD-CI method, Gaussian mixture model (GMM), and three top-performing supervised classifiers: k-nearest neighbor (k-NN), support vector machine (SVM), and random forest. It is worth mentioning that other classifiers, such as logistic regression (Svensén and Bishop, 2007), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) (Hastie et al., 2008), decision tree (Loh, 2011), AdaBoost (Hastie et al., 2009a), and Gaussian naive Bayes (Schütze et al., 2008), were also studied, but their results were not as good as those of k-NN, SVM, and random forest, so they were excluded from the analysis reported here.

PSD-CI can be applied directly to raw (non-normalized) pulses, as demonstrated with the measured data set. For the synthetic data, however, PSD-CI is applied to pulses normalized to their peak values, ensuring the maximum amplitude of each pulse is normalized to one. This normalization facilitates the analysis of results at various noise levels. The other algorithms – k-NN, SVM, random forest, GMM, and the proposed GMVAE – are used on pulses normalized to a peak of one for both the measured and synthetic data sets. While most algorithms process pulses of length $L = 296$ samples, GMM uniquely operates on a 3D projection of the pulses using principal component analysis (PCA).

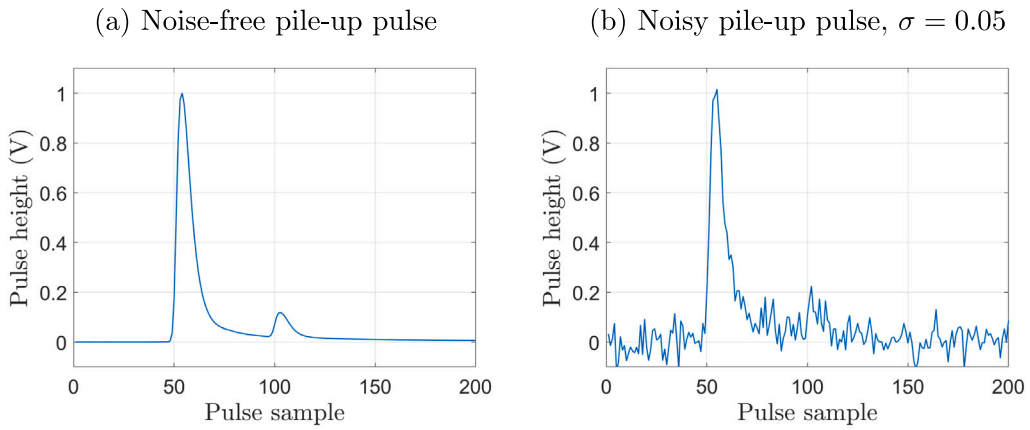


Fig. 4. Synthetic pile-up pulses normalized to a maximum amplitude of one, showcasing (a) a noise-free scenario and (b) a scenario with noise level $\sigma = 0.05$.

Table 1
Number of pulses in the measured data set as a function of the source-to-detector distance.

Distance	30 cm	40 cm	50 cm	60 cm	70 cm	80 cm	90 cm	100 cm	110 cm
#pulses	439 777	312 440	249 665	211 799	187 733	171 725	158 792	152053	146 250

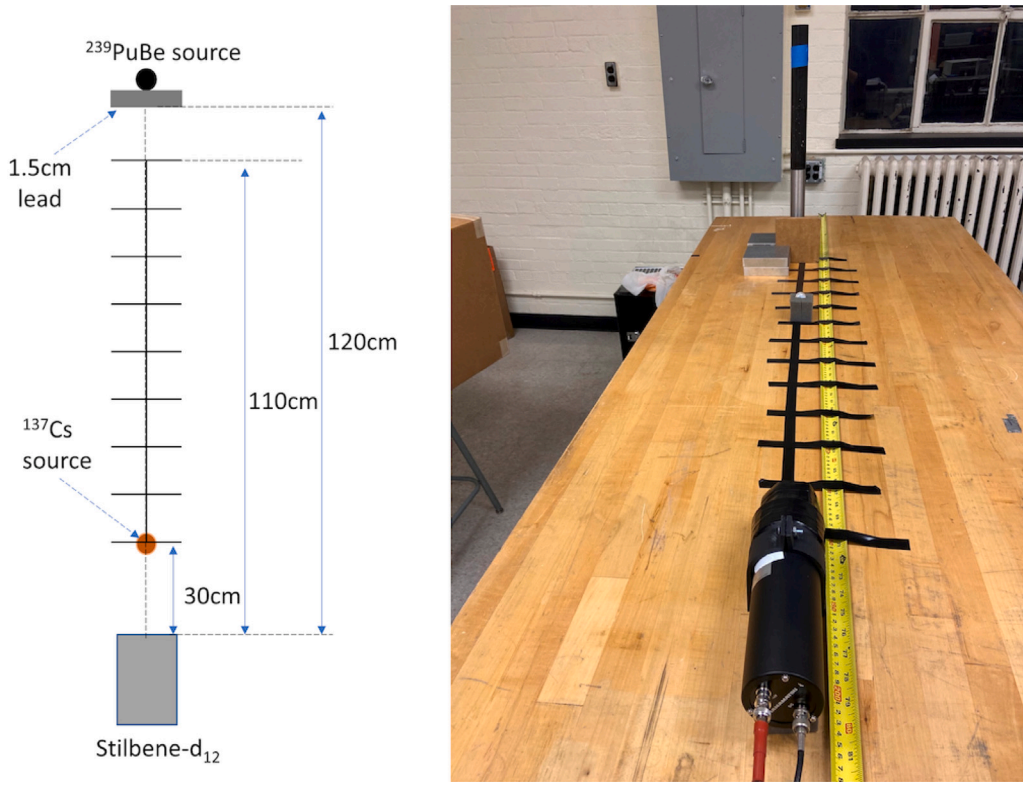


Fig. 5. Schematic of the experimental setup (left) and picture of the setup (right) with the stilbene-d_{12} irradiated by the ^{239}PuBe and ^{137}Cs sources.

4.1. Charge-integration pile-up-rejection PSD

The CI pile-up-rejection PSD, hereafter referred to as PSD-CI, consists of two steps: pile-up rejection followed by standard CI classification (Brooks, 1959). CI is described first to clarify the impact that pile-up has on the computed pulse charges.

4.1.1. Charge integration

Radiation-induced pulses in organic scintillators result mostly from prompt and delayed fluorescence produced by the molecular excitation. Because of their higher ionization density, recoil protons produced by neutron collisions induce a relatively larger delayed component, compared to recoil electrons from gamma-ray interactions, for an equivalent overall energy deposited (Brooks, 1959). This phenomenon results in neutron pulses with a larger tail area, compared to gamma-ray pulses, as shown in Fig. 6 for neutron and gamma-ray pulses acquired by the stilbene-d_{12} detector and normalized to their respective peak value. The CI algorithm exploits this physical principle by classifying neutron and gamma-ray pulses based on the area under the tail portion of each pulse relative to the total pulse area. The most common PSD-CI parameter is the tail-to-total ratio, *i.e.*, the ratio between the area under the "tail" of the pulse, which begins several nanoseconds after

Table 2
Optimized PSD-CI discrimination parameters for each data set.

Data set	σ	a (V^{-1})	b	s (V)
Measured	N.A.	−0.03261	0.32424	0.00248
	0.05	0	0.22	0
Synthetic	0.01	0	0.22	0
	0.001	0	0.22	0

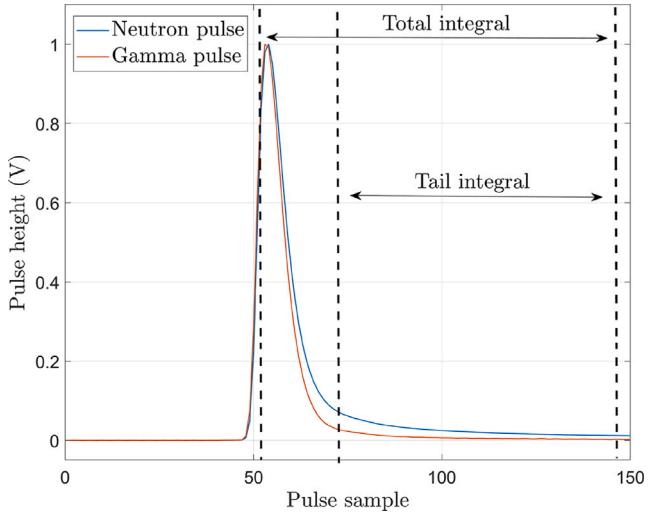


Fig. 6. Neutron and gamma-ray pulses measured by the stilbene-d_{12} detector normalized to their respective peak value. Time gates used to calculate the "total" and "tail" integrals are also shown.

the pulse maximum and extends to the end of the pulse, and the "total" area, which includes the entire pulse. In this work, we calculated the "total" area over 374 ns, starting 4 ns before the time stamp of the pulse maximum, while the "tail" gate started 22 ns after the pulse maximum and extended for 352 ns. These parameters were optimized in our previous work (Zhou et al., 2022). Based on the total integral (total) and tail-to-total ratio (TTR), we generated a discrimination line to classify the particle types. The equation of this line is: $\mathrm{TTR} = a \times \mathrm{total} + b + \frac{s}{\mathrm{total}}$, where a, b and s are constants. This line divides pulses into two groups: neutron and gamma-ray pulses. Optimized values of a, b and s were selected for each data set to have the minimum number of misclassified pulses. Table 2 shows the optimized a, b and s values for DT measured and synthetic pulses. For the synthesized pulses, as their pulse heights were normalized to their peak values, neutron and gamma-ray pulses post-normalization can be discriminated using a straight PSD-CI line ($a, s = 0$).

4.1.2. Pile-up rejection

Pulse pile up increases the tail area of the first pulse and hence degrades the PSD performance. Here we implemented a pile-up rejection algorithm by counting the number of rising edges in the pulse. Given a pulse $\boldsymbol{x} \in \mathbb{R}^L$ with peak height $x_{\max}$ located at the sample i^*, we move a window of width Δ along the pulse and calculate the difference between the last and first samples in the window

$$dx[i] = x[i + \Delta] - x[i] \tag{3}$$

where Δ is the pulse rising edge width. If dx exceeds a fixed fraction f of pulse height and the noise level τ more than once, the pulse has two or more rising edges and is classified as "pile up". The pile-up algorithm is outlined in Algorithm 1. The user-defined parameters of this method are Δ, f and τ, which in our study were optimized for each data set, as shown in Table 3.

Table 3
Optimized pile-up rejection parameters for each data set.

Data set	σ (V)	Δ (ns)	f	τ (s)
Measured	N.A.	6	0.1	0.005
	0.05	6	0.175	0.0
Synthetic	0.01	6	0.045	0.0
	0.001	6	0.0175	0.0

Algorithm 1: Determine if a pulse is pile up.

Require: $\boldsymbol{x} \in \mathbb{R}^L, \Delta, f, \tau$
  $x_{\max} \leftarrow \max(\boldsymbol{x})$, $i^* \leftarrow \mathrm{argmax}(\boldsymbol{x})$
  $i \leftarrow 0$
  while $i \neq i^* - 2\Delta$ **do**
    $dx \leftarrow x[i + \Delta] - x[i]$
    if $dx \geq f\ x_{\max}$ **and** $dx \geq \tau$ **then**
      return True
    end if
    $i \leftarrow i + 1$
  end while
  $i \leftarrow i^*$
  while $i \neq L - \Delta$ **do**
    $dx \leftarrow x[i + \Delta] - x[i]$
    if $dx \geq f\ x_{\max}$ **and** $dx \geq \tau$ **then**
      return True
    end if
    $i \leftarrow i + 1$
  end while
  return False

4.2. SVM

Support vector machine (SVM) (Cortes and Vapnik, 1995; Schölkopf and Smola, 2002) is a powerful and flexible algorithm for both classification and regression problems. It identifies the optimal boundary or "hyperplane" that separates the classes in the data, while maximizing the margin, or the distance between the boundary and the closest data points from each class, known as support vectors. For non-linearly separable data, as in PSD with pile up, SVM employs the "kernel trick" by mapping the data to a higher-dimensional space where it becomes linearly separable. SVM has been widely used and has proven to be effective, but the choice of kernel function and the values of the tuning parameters are crucial for good performance.

In our PSD classification using SVM, we utilized a Gaussian (Radial Basis Function) kernel. The regularization strength was set to 1.0, and the spread of the kernel was specified with a value of 0.7. This configuration demonstrated effective performance on both synthetic and measured data sets.

4.3. Random forest

Random forest (Breiman, 2001) is a versatile ensemble ML technique that can be applied to both classification and regression problems. It is composed of multiple decision trees, each trained on a random subset of the data. One of the key advantages of random forest is its ability to reduce overfitting, a common issue in decision trees, by averaging or voting the predictions of all the decision trees, which reduces the variance of the final output. Furthermore, random forest also improves the accuracy of predictions by reducing the correlation among the decision trees through a technique known as "bagging", or bootstrap aggregating, which trains each tree on a different subset of the data. The application of this method results in an ensemble that is more robust and less prone to overfitting. The algorithm also includes a built-in feature selection mechanism, using a technique called "random subspace method", where at each split of the tree, a random subset

of features is considered. This approach helps to avoid overfitting and improves the generalization of the model. Additionally, it can handle large numbers of features, even if they are highly correlated or irrelevant to the target variable, and can handle missing values and outliers well. Overall, the random forest algorithm is user-friendly and easy to understand, requiring minimal data preparation and robust to noise.

For the PSD classification task, our random forest implementation had a tree depth limit set to ten and employed 20 trees in the forest. This setup was effective on both synthetic and measured data sets.

4.4. K-NN

The k-nearest neighbors (k-NN) algorithm (Cover and Hart, 1967; Altman, 1992; Hastie et al., 2009b) is a supervised learning technique that can be used for both classification and regression problems. It is based on the idea that similar data points tend to have similar labels. The algorithm works by taking a new data point and finding the k-nearest data points in the training set using a distance metric, such as the Euclidean distance. The label of the new data point is then determined by the majority label among the k-nearest data points. One of the main advantages of k-NN is its simplicity. The algorithm is easy to understand and implement, and it requires no training, making it memory efficient. Additionally, k-NN can handle non-linear decision boundaries and can work well with small data sets. The performance of k-NN can be affected by the choice of the value of k and the distance metric used. A smaller value of k can lead to a more complex decision boundary and a higher risk of overfitting, while a larger value of k can lead to a smoother decision boundary and a lower risk of overfitting. A common method to select k is cross-validation, in which different values of k are tested and the best k is chosen. The choice of distance metric can also affect the performance, with Euclidean distance being a commonly used metric.

In the PSD classification task, we utilized the k-NN algorithm, considering only the closest neighbor for classification. It yielded effective results on both synthetic and measured data sets. k-NN effectively compares pulses through the Euclidean distance in an L-dimensional space, where $L = 296$ in our experiments.

4.5. GMM

A Gaussian mixture model (GMM) (Dempster et al., 1977; Xu, 1993; Celeux and Govaert, 1995) is a probabilistic model that can be used for data clustering by assuming that all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Each Gaussian component has its own parameters, such as the mean and the covariance matrix. Formally, a GMM with c components is defined as:

$$\mathrm{p}(\boldsymbol{x}|\boldsymbol{\Theta}) = \sum_{i=1}^{c} \omega_i \, \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{4}$$

where $\boldsymbol{\Theta} = \{\omega_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1:c}$ are the parameters of the model, $\boldsymbol{x}$ is the data point, ω_i is the mixing coefficient for component i, $\boldsymbol{\mu}_i$ is the mean of component i, and $\boldsymbol{\Sigma}_i$ is the covariance matrix of component i. The goal of GMM is to find the parameters $\boldsymbol{\Theta}$ that maximize the likelihood of the data. This result can be achieved using the expectation–maximization (EM) algorithm. The EM algorithm is an iterative method that alternates between two steps:

- E-step: Compute the responsibility of each component for each data point. These responsibilities are the conditional probabilities that a data point belongs to a component given the current parameter estimates.
- M-step: Update the parameters of the model based on the responsibilities.

Note that the GMM can also be trained in a semi-supervised manner when partially labeled data are available. In this case, the responsibilities of the labeled data are fixed by vectors of the known labels of the ground truth in the E-step.

In the PSD classification, we employed a GMM approach on the data after projecting it into a 3D space using PCA. This technique yielded enhanced results compared to applying GMM on the raw data directly. It is worth noting that our GMVAE shares features with the GMM. The main difference between our GMVAE and the GMM is that the GMVAE encoder acts as a data-driven nonlinear dimensionality reduction while the GMM simply applies a linear dimensionality reduction step.

5. Classification results of synthetic and measured data sets

We tested the performance of the proposed GMVAE utilizing synthetic and measured data sets with three classes: (0) gamma-ray pulses, (1) neutron pulses and (2) pile-up pulses introduced in Section 3.

5.1. Synthetic data sets

We evaluated the performance of the GMVAE in terms of classification accuracy by training it with different numbers of labeled data points (from 0 to 100% of the size of the training data set) and varying noise levels ($\sigma = 0.001, 0.01, 0.05$). The noise added to the pulses, which are normalized to their peak values to ensure the maximum amplitude of each pulse is normalized to one, follows a Gaussian distribution with a mean of zero and standard deviation of σ (σ is dimensionless because pulses are normalized). It is worth noting that supervised classifiers (SVM, random forest and k-NN) are trained only on the subset of labeled data while the GMVAE and GMM are effectively trained on both labeled and unlabeled pulses.

Fig. 7 compares the classification accuracy, defined as the ratio of the number of correct predictions to the total number of input pulses, obtained by PSD-CI, GMVAE, GMM and the three supervised classifiers: k-NN, SVM and random forest on the synthetic data set with different noise levels. Our analysis reveals that the GMVAE demonstrates remarkable robustness and efficiency. When assessing the impact of varying noise levels, GMVAE demonstrated significant resilience. Even with a minimal set of 100 labels, the model achieved 99.5% accuracy at a noise level of $\sigma = 0.001$ and a remarkable 96.7% at a high noise level of $\sigma = 0.05$. This result lies within 2% of its peak performance at $\sigma = 0.05$, *i.e.*, 98.3% accuracy, making GMVAE exceptionally efficient in achieving high classification accuracy with minimal supervision. Contrast this with supervised methods like k-NN, SVM, and random forest, which demonstrated a more gradual increment in performance with increasing data set size. For instance, SVM needed at least 1000 pulses to reach within 2% of its maximum accuracy at $\sigma = 0.05$. k-NN and SVM displayed similar trends, requiring larger data sets to converge to optimal performance. Further, when compared to PSD-CI, GMVAE demonstrated superior performance, especially in high noise scenarios. Notably, GMVAE attained an accuracy of 96.7% at $\sigma = 0.05$ with minimal supervision (100 labels), whereas PSD-CI declined to 82.8%, indicating that GMVAE is better suited for environments with increased uncertainty and noise. Finally, when compared to the semi-supervised approach GMM, GMVAE illustrated clear benefits in both unsupervised and semi-supervised scenarios. Notably, GMVAE benefited from a data-driven nonlinear dimensionality reduction in its encoder, which distinctly outperformed GMM's linear dimensionality reduction approach, particularly at high noise levels. For instance, at $\sigma = 0.05$, GMM accuracy saturates at 84% for all tests with different numbers of labeled data. It is worth noting that our GMVAE can be trained on unlabeled data alone when labeled data are unavailable. Notably, the fully unsupervised GMVAE achieves classification accuracy of 90.5%, 91.1%, and 62.3% for the considered noise levels. These results demonstrate that GMVAE can perform well with no supervision

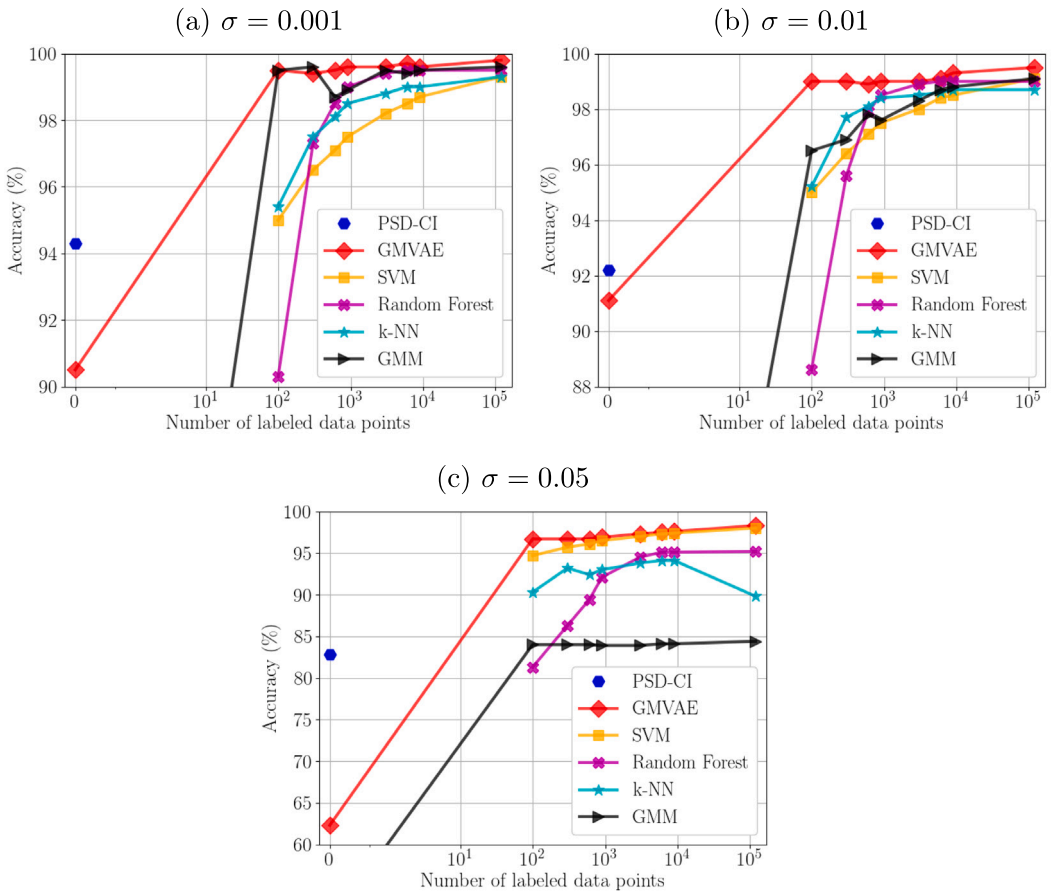


Fig. 7. Synthetic data set: Classification accuracy comparison of PSD-CI, GMVAE, GMM, k-NN, random forest, and SVM against the number of labeled data points under noise levels ($\sigma = 0.001, 0.01, 0.05$). The x-axis's initial point represents an unsupervised scenario (zero labeled data points), showcasing GMVAE and GMM's capability to utilize unlabeled data, unlike the supervised-only SVM, random forest, and k-NN. PSD-CI is shown as a single point, indicating its independence from labeled data quantity and its reliance on the full unlabeled data set. Visualization is optimized by adjusting the y-axis range in each panel.

in environments with low noise levels, while minimal supervision is recommended for noisy environments.

Fig. 8 displays the GMVAE membership probabilities of the observed classes (gamma, neutron, and pile-up) for various noise levels. Note that the class membership probabilities sum up to one across all classes for a given data point. It is evident that the mean probability of the true class, as determined by GMVAE, remains around 47% for all tests with varying numbers of labeled data points, while the mean probability of the false classes is approximately 27%. Notably, the probability uncertainty does not decrease significantly with a decrease in the number of labeled data points, indicating the effectiveness of the minimally-supervised GMVAE. However, an increase in the probability uncertainty with an increase in the noise level is observed, as could be expected intuitively (especially for the pile-up class).

The confusion matrices show the GMVAE classification fraction for each category. The confusion matrices produced by GMVAE using 100 labels (minimal supervision) for different noise levels are shown in Fig. 9 (a, b, c), and that produced using 120,000 labels (full supervision) for the highest $\sigma = 0.05$ is shown in Fig. 9 (d). It can be seen that the minimally-supervised GMVAE achieves a classification accuracy of 99.5%, 99%, and 96.7% for the noise levels $0.001, 0.01$ and 0.05, respectively. The decrease in accuracy is due to the fact that pile-up pulses can be challenging at high noise levels as the second peak can be obscured by the noise (see Fig. 4). Notably, the classification accuracy of the minimally-supervised GMVAE is only 1.6% lower than that of the fully-supervised GMVAE for the highest noise level ($\sigma = 0.05$). These results demonstrate the robustness of the minimally-supervised approach. Fig. 10 displays the raw data and GMVAE latent vectors of the synthetic data set with a noise level of $\sigma = 0.05$, projected onto a 3D space using PCA. Unlike the raw data in Fig. 10 (a), the GMVAE latent space in Fig. 10 (b) demonstrates a distinct separation between the different classes in the testing data points.

5.2. Measured data sets

Given the high noise levels in the real data, we validate the performance of all methods on the measured data set after training on a synthetic data set of size 120,000 pulses (40,000 pulses per class) with a noise level of $\sigma = 0.05$.

The parameters of PSD-CI need to be re-tuned to perform PSD of the measured data set effectively, as using parameters found with synthetic pulses leads to poor results. This result highlights the lack of robustness in charge integration methods, which require significant user intervention and parameter tuning, making them unsuitable for automated systems. Fig. 11 displays gamma and neutron counts obtained by different classifiers, plotted against the distance from the ^{137}Cs source to the detector. The pile-up counts are not shown as they are usually considered outliers in real-world applications. Moreover, the pile-up rate is low, maximum 1.3%, with the absolute number of pile-up events is correspondingly small, which precludes a detailed assessment of the fit with the theoretical decay of the pile-up rate as the distance from the detector increases. As discussed in Section 3, the gamma counts are expected to decrease proportionally to the square of the distance. This is confirmed in Fig. 11 (a), where all classifiers show a decline following the $1/distance^2$ trend. The neutron counts, on the other hand, are expected to remain constant, which is demonstrated in Fig. 11 (b)

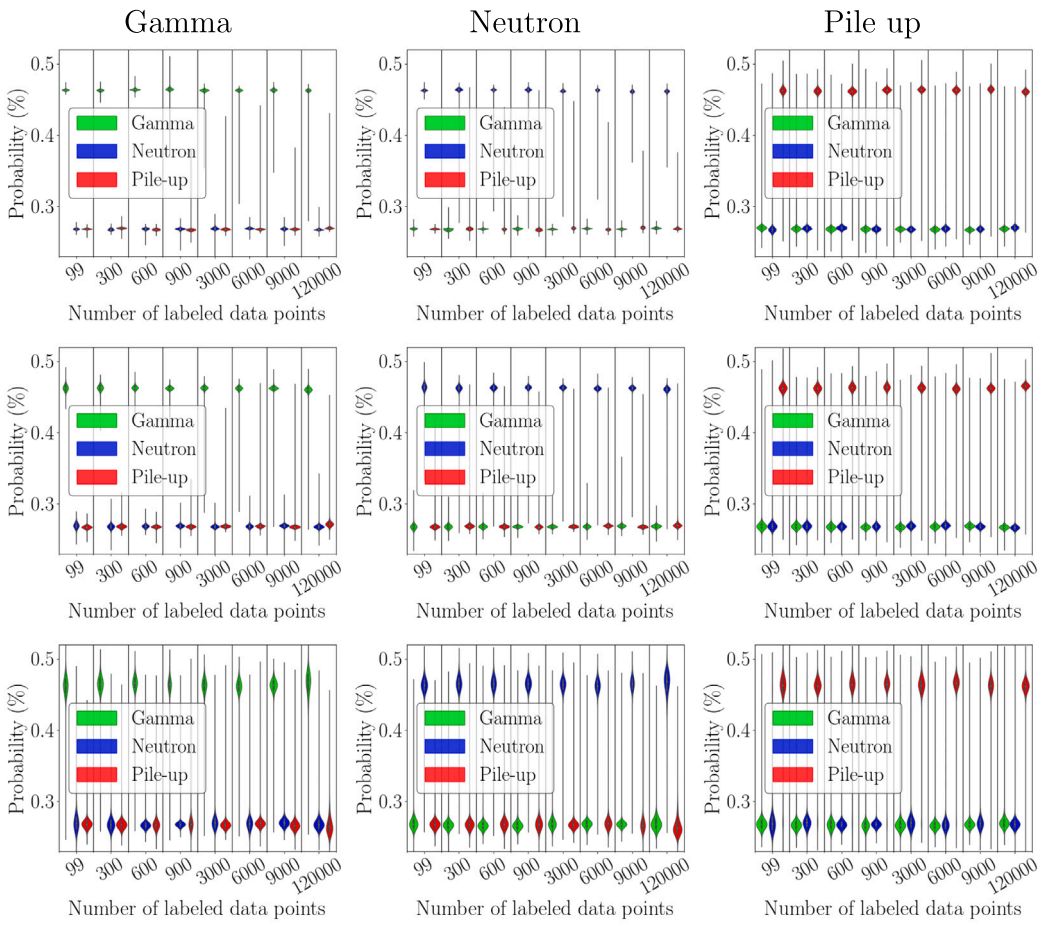


Fig. 8. Synthetic data set: GMVAE class probability distributions. Columns correspond to true classes (Gamma, Neutron, Pile-up), and rows correspond to noise levels with $\sigma = 0.001$ (top), 0.01 (middle), and 0.05 (bottom). Violin plots depict the distribution of predicted probabilities for each class, illustrating the GMVAE performance across varying number of labeled data points (represented by divisions along the x-axis) and noise conditions (across rows).

for 100 labeled pulses and Fig. 11 (c) for 120,000 labeled pulses. The proposed GMVAE exhibits more stable and consistent neutron counts across the nine measured scenarios with increasing pile-up rate, as indicated by lower normalized uncertainties of 0.008 and 0.009 for cases (b) and (c), respectively. These values are compared to 0.010 and 0.015 for SVM, 0.020 and 0.037 for random forest, 0.012 and 0.009 for k-NN, and 0.050 for PSD-CI.

We examined pulses that GMVAE identified as neutrons, comparing these classifications with those from SVM and k-NN. Results showed that SVM mainly misidentifies these as pile-ups, while k-NN often sees them as gammas. These results implies that both SVM and k-NN may underestimate neutron counts.

Fig. 12 shows the raw data and GMVAE latent vectors of the measured data set for the case where the ^{137}Cs source was positioned 30 cm from the detector surface, projected onto a 3D space using PCA. It is evident that the different classes, especially the pile ups, are more distinctly separated in the GMVAE latent space in Fig. 12 (b) compared to the raw data in Fig. 12 (a). These results clearly highlight the ability of the GMVAE model to better distinguish between different pulse classes, even in the presence of pile-up noise.

Finally, our GMVAE model demonstrates high computational efficiency, enabling real-time inference capabilities. Specifically, it can process 10^5 pulses within 0.34 s when utilizing an NVIDIA Quadro RTX 6000 GPU. This performance metric underscores the model's suitability for real-time applications, highlighting its practicality in scenarios requiring swift data processing.

6. Conclusion

In this paper, we presented a novel semi-supervised classification approach for solving the challenging problem of PSD in the presence of pile-up pulses, when performing measurements with organic scintillation detectors, stilbene-d_{12}. Accurately identifying and separating pile-up events is critical for obtaining accurate energy spectra, which is essential for a wide range of applications from nuclear safeguards to physics and astrophysics nuclear reaction studies.

Our approach leveraged a GMVAE, which can be trained in an unsupervised fashion, and introduced a semi-supervised version with auxiliary loss functions to take advantage of scarce labeled data. We thoroughly evaluated the performance of the GMVAE utilizing synthesized and measured data sets, and the results confirmed its superior performance compared to other traditional and state-of-the-art ML classifiers. The synthesized data set was based on pulse templates obtained with a 14.1 MeV monenergetic neutron source, while the measured data were obtained from a ^{239}PuBe radionuclide source. In both cases, the GMVAE outperformed the traditional deterministic CI approach and other state-of-the art ML algorithms, owing to its semi-unsupervised ability of learning not only from labeled pulses but also from the non-labeled ones, hence its nature. Notably, while the GMVAE model was developed using training data set from a monoenergetic neutron source, *i.e.*, produced by a DT neutron generator, the model exhibited excellent performances also when used to perform PSD classification of a polyenergetic mixed field produced by a ^{239}PuBe neutron and gamma-ray source and a ^{137}Cs gamma-ray source. These results demonstrated

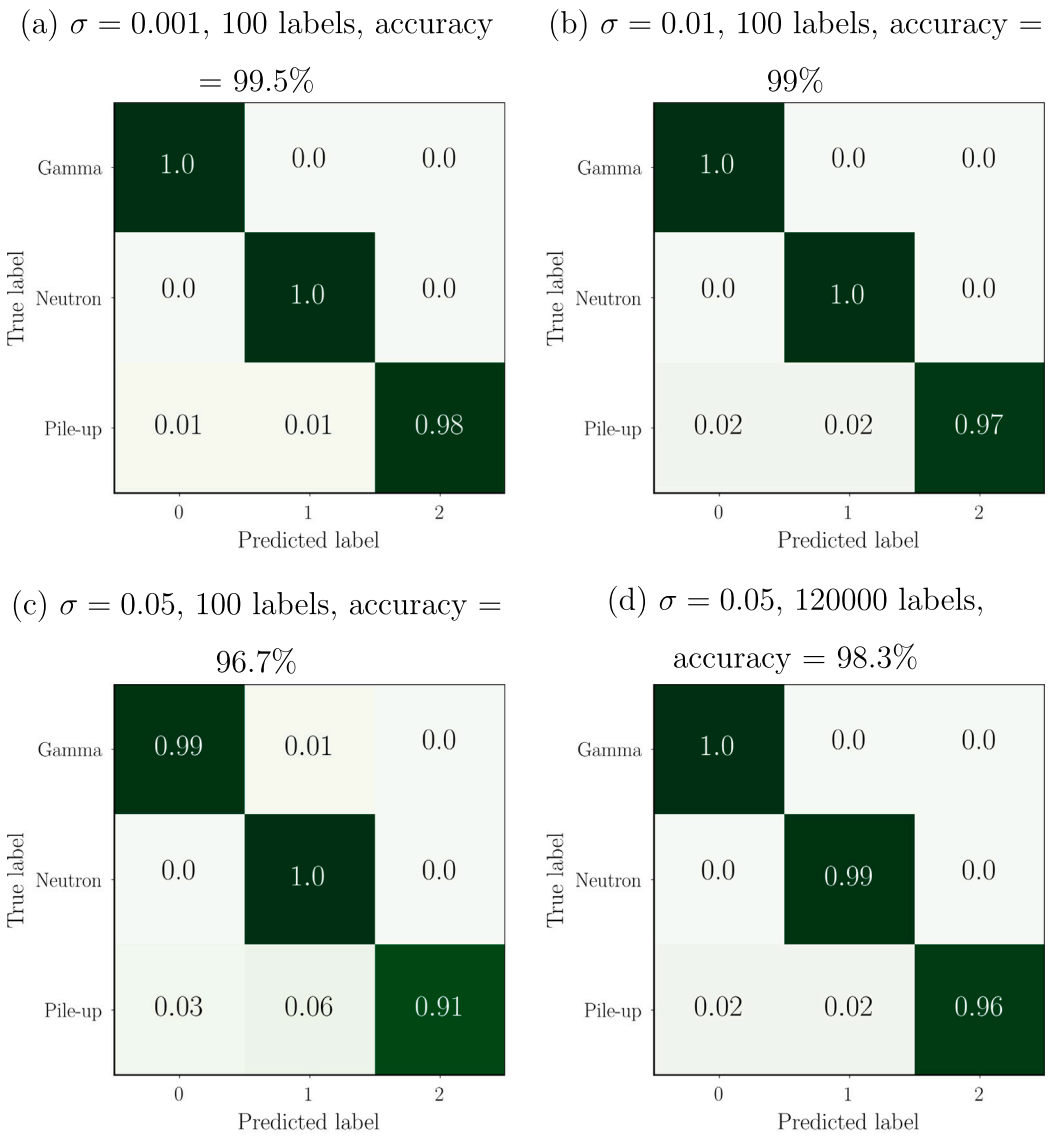


Fig. 9. Synthetic data set: Confusion matrices of the proposed GMVAE obtained after training with 120,000 pulses, out of which 100 are labeled (minimal supervision) and with a noise level of (a) $\sigma = 0.001$, (b) $\sigma = 0.01$, and (c) $\sigma = 0.05$. Panel (d) corresponds to training the GMVAE with 120000 labeled pulses (full supervision). Each panel shows testing on 480,000 pulses with the same noise level as used in training.

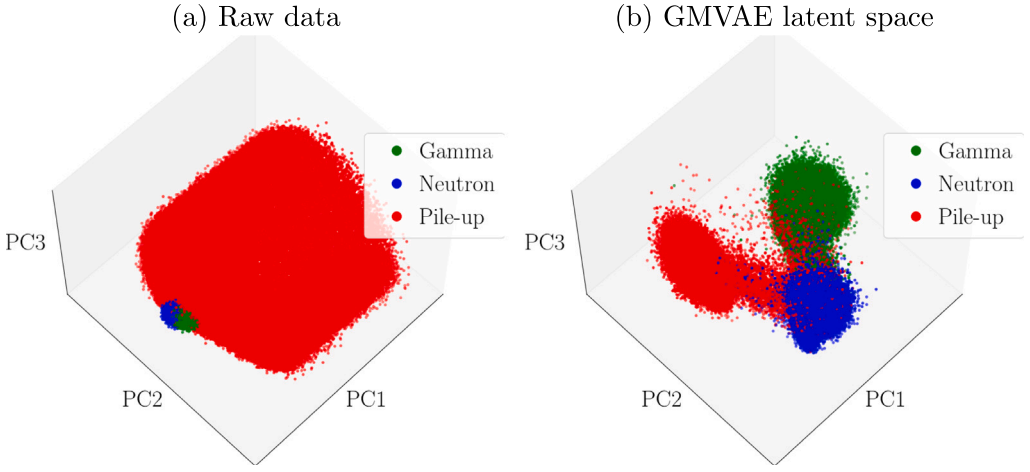


Fig. 10. Synthetic data set: 3D PCA projection of the raw data (a) and the GMVAE latent vectors (b) of 480000 pulses with a noise level of $\sigma = 0.05$. The GMVAE was trained with 120000 pulses, out of which only 100 were labeled (minimal supervision).

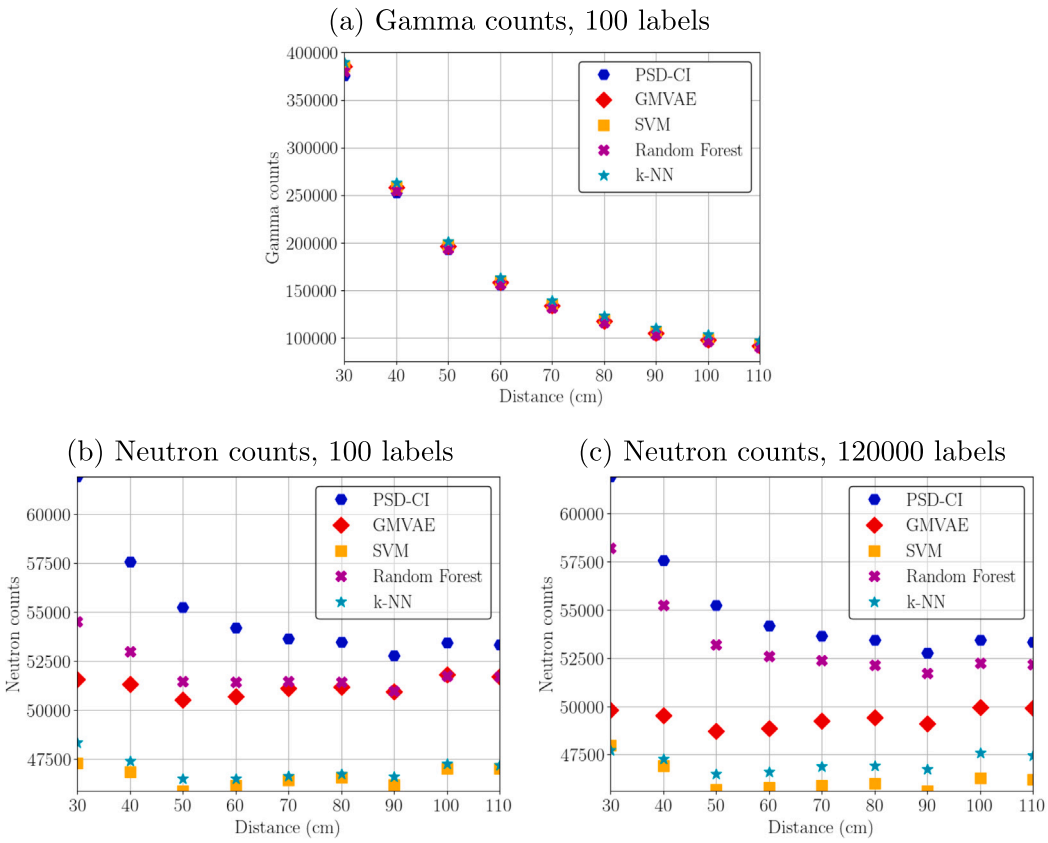


Fig. 11. Measured data set: Gamma and neutron counts obtained by various classifiers are plotted against the distance from the ^{137}Cs source to the detector. (a) Gamma counts and (b) neutron counts were obtained by training the GMVAE with 120000 synthetic pulses with a noise level of $\sigma = 0.05$, out of which only 100 were labeled (minimal supervision), while the supervised classifiers SVM, k-NN, and random forest were trained only with the 100 labeled pulses, as they cannot use unlabeled pulses. (c) Neutron counts were obtained by training all classifiers with the full synthetic data set of 120000 labeled pulses with a noise level of $\sigma = 0.05$. Notably, gamma counts for 120000 labels are omitted as all algorithms consistently perform similarly to the 100 label scenario, rendering further illustration unnecessary.

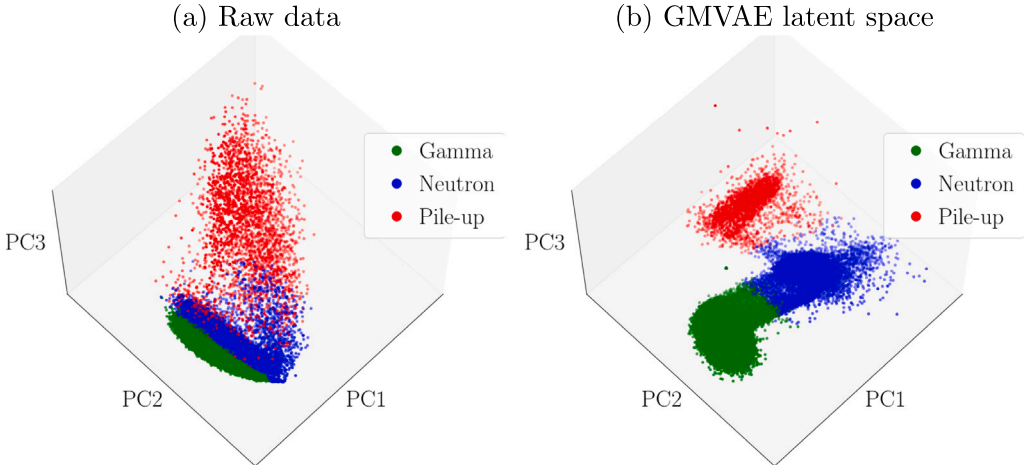


Fig. 12. Measured data set: 3D PCA projection of the raw data (a) and the GMVAE latent vectors (b) for the case where the ^{137}Cs source is positioned 30 cm from the detector surface. Both (a) and (b) contain 439777 pulses, and the GMVAE was trained with 120000 synthetic pulses with a noise level of $\sigma = 0.05$, out of which only 100 were labeled (minimal supervision).

the effectiveness of the training, the robustness of the classification, and the general applicability of the model, even when the neutron energy range of the data used for training differs substantially from the tested one.

Our analysis allowed us to evaluate the absolute performance of the GMVAE in the presence of random noise overlapped to the synthetic signals and its relative performance compared to existing state-of-the-art algorithms. GMVAE maintained a classification accuracy exceeding 96%, even under noise conditions severe enough to induce false pile-up identifications in conventional methods. When compared to supervised methods under severe noise conditions, GMVAE required the fewest training pulses (100) to achieve a classification accuracy within 2% of

its optimum performance, *i.e.*, 98.3% accuracy. Conversely, supervised methods, such as SVM, exhibited a steady increase of the classification accuracy with the size of the training set, which had to encompass at least 1000 pulses to maximize the classification accuracy. Among semi-supervised approaches, GMVAE also benefits from data-driven non-linear dimensionality reduction in the encoder, which yields superior results in terms of minimum size of the training set and classification accuracy compared to the GMM, whose encoder is based on linear dimensionality reduction. GMVAE-based algorithms achieve excellent PSD classification performance while requiring minimum supervision.

Our GMVAE model demonstrated exceptional computational efficiency and was capable of real-time inference, processing 10^5 pulses in just 0.34 s. This efficiency underscored the model's aptness for real-time applications, emphasizing its practicality for fast data processing demands.

Future work involves exploring the performance of the GMVAE on other types of detectors and sources and implementing the GMVAE on field programmable gate array (FPGA)-based platforms and evaluate its effectiveness when performing real-time PSD classification in the field. Additionally, the proposed method could be extended to other applications beyond PSD where labeled data is scarce or costly to obtain, making it a promising approach for a wide range of classification problems.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yoann Altmann reports financial support was provided by Royal Academy of Engineering. Stephen McLaughlin reports financial support was provided by Engineering and Physical Sciences Research Council, and UK MOD University Defence Research Collaboration (UDRC) in Signal Processing. Angela Di Fulvio reports financial support was provided by Department of Energy National Nuclear Security Administration.

Data availability

The data sets supporting this work are available at the following DOI: 10.17861/73b6ec57-a691-4337-8781-2087459dd48a.

Acknowledgments

This work was supported by the Royal Academy of Engineering, UK under the Research Fellowship scheme RF201617/16/31, by the UK Engineering and Physical Sciences Research Council project EP/T00097X/1, EP/S000631/1, by the UK MOD University Defence Research Collaboration (UDRC) in Signal Processing and in part by the Department of Energy National Nuclear Security Administration, USA through the Nuclear Science and Security Consortium under Award Number DE-NA0003996.

References

Abdulaziz, A., Zhou, J., Di Fulvio, A., Altmann, Y., McLaughlin, S., 2022. Semi-supervised Gaussian mixture variational autoencoder for pulse shape discrimination. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 3538–3542.

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. In: American Statistical Association, vol. 87, pp. 10–14.

Bourne, M.M., 2016. Organic Scintillators in Nonproliferation Applications With a Hybridized Double-Pulse Rejection Technique (Ph.D. thesis). University of Michigan.

Bourne, M., Clarke, S., Paff, M., Di Fulvio, A., Norsworthy, M., Pozzi, S., 2017. Digital pile-up rejection for plutonium experiments with solution-grown stilbene. Nucl. Instrum. Methods Phys. Res., Sect. A 842, 1–6.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brooks, F.D., 1959. A scintillation counter with neutron and gamma-ray discriminators. Nucl. Instrum. Methods 4 (3), 151–163.

Brooks, F., 1979. Development of organic scintillators. Nucl. Instrum. Methods 162 (1–3), 477–505.

Caldwell, A., Cossavella, F., Majorovits, B., Palioselitis, D., Volynets, O., 2015. Signal recognition efficiencies of artificial neural-network pulse-shape discrimination in HPGe $0\nu\beta\beta$-decay searches. Eur. Phys. J. C 75 (7), 1–9.

Celeux, G., Govaert, G., 1995. A tutorial on expectation-maximization and its application to parameter estimation for gaussian mixture and hidden markov models. Comput. Stat. Data Anal. 2 (4), 441–458.

Charakorn, R., Thawornwattana, Y., Itthipuripat, S., Pawlowski, N., Manoonpong, P., Dilokthanakul, N., 2020. An explicit local and global representation disentanglement framework with applications in deep clustering and unsupervised object detection. arXiv preprint arXiv:2001.08957.

Collier, M., Urdiales, H., 2019. Scalable deep unsupervised clustering with concrete gmvaes. arXiv preprint arXiv:1909.08994.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inform. Theory 13 (1), 21–27.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 1–38.

Di Fulvio, A., Shin, T., Jordan, T., Sosa, C., Ruch, M., Clarke, S., Chichester, D., Pozzi, S., 2017. Passive assay of plutonium metal plates using a fast-neutron multiplicity counter. Nucl. Instrum. Methods Phys. Res., Sect. A 855, 92–101.

Di Fulvio, A., et al., 2018. Fast-neutron multiplicity counter for active measurements of uranium oxide certified material. Nucl. Instrum. Methods Phys. Res., Sect. A 907, 248–257.

Esposito, B., Fortuna, L., Rizzo, A., 2004. Neural neutron/gamma discrimination in organic scintillators for fusion applications. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), vol. 4, pp. 2931–2936 vol.4.

Figueroa, J.A., 2019. Semi-supervised learning using deep generative models and auxiliary tasks. In: NeurIPS.

Figueroa, J.A., Rivera, A.R., 2017. Learning to cluster with auxiliary tasks: A semi-supervised approach. In: IEEE SIBGRAPI. pp. 141–148.

Fu, C., Di Fulvio, A., Clarke, S., Wentzloff, D., Pozzi, S., Kim, H., 2018. Artificial neural network algorithms for pulse shape discrimination and recovery of piled-up pulses in organic scintillators. Ann. Nucl. Energy 120, 410–421.

Gaughan, N., Zhou, J., Becchetti, F., Torres-Isea, R., Febbraro, M., Zaitseva, N., Altmann, Y., Di Fulvio, A., 2021. Characterization of stilbene/d12 for neutron spectroscopy without time of flight. Nucl. Instrum. Methods Phys. Res., Sect. A.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Hastie, T., Rosset, S., Zhu, J., Zou, H., 2009a. Multi-class adaboost. Stat. Interface 2 (3), 349–360.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning. Springer.

Hastie, T., Tibshirani, R., Friedman, J., 2009b. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

Jang, E., Gu, S., Poole, B., 2016. Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic gradient descent. In: ICLR. pp. 1–15.

Kingma, D., Welling, M., 2014. Auto-encoding variational bayes. In: ICLR.

Kingma, D.P., Welling, M., 2019. An introduction to variational autoencoders. Found. Trends® Mach. Learn. 12 (4), 307–392.

Knoll, G.F., 2010. Radiation Detection and Measurement. John Wiley & Sons, Inc., Hoboken, New Jersey.

Liu, G., Aspinall, M., Ma, X., Joyce, M., 2009. An investigation of the digital discrimination of neutrons and gamma rays with organic scintillation detectors using an artificial neural network. Nucl. Instrum. Methods Phys. Res. A 607 (3), 620–628.

Loh, W.-Y., 2011. Classification and regression trees. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 1 (1), 14–23.

Maddison, C., Mnih, A., Teh, Y., 2017. The concrete distribution: A continuous relaxation of discrete random variables. In: ICLR.

Marrone, S., Cano-Ott, D., Colonna, N., Domingo, C., Gramegna, F., Gonzalez, E., Gunsing, F., Heil, M., Käppeler, F., Mastinu, P., Milazzo, P., Papaevangelou, T., Pavlopoulos, P., Plag, R., Reifarth, R., Tagliente, G., Tain, J., Wisshak, K., 2002. Pulse shape analysis of liquid scintillators for neutron studies. Nucl. Instrum. Methods Phys. Res., Sect. A 490 (1–2), 299–307. http://dx.doi.org/10.1016/S0168-9002(02)01063-X.

Monterial, M., Marleau, P., Clarke, S., Pozzi, S., 2015. Application of Bayes' theorem for pulse shape discrimination. Nucl. Instrum. Methods Phys. Res., Sect. A 795, 318–324. http://dx.doi.org/10.1016/j.nima.2015.06.014.

Morishita, Y., Di Fulvio, A., Clarke, S., Kearfott, K., Pozzi, S., 2019. Organic scintillator-based alpha/beta detector for radiological decontamination. Nucl. Instrum. Methods Phys. Res., Sect. A 935, 207–213.

Nakhostin, M., 2019. A general-purpose digital pulse shape discrimination algorithm. IEEE Trans. Nucl. Sci. 66 (5), 838–845. http://dx.doi.org/10.1109/TNS.2019.2910153.

Schölkopf, B., Smola, A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In: IEEE CVPR. pp. 815–823.

Schütze, H., Manning, C.D., Raghavan, P., 2008. Introduction to Information Retrieval, vol. 39, Cambridge University Press Cambridge.

Shu, R., 0000. Gaussian mixture vae: Lessons in variational inference, generative models, and deep nets, URL http://ruishu.io/2016/12/25/gmvae/.

SpA, C., 2020. UM4380 - 725 and 730 DPP-PSD User Manual Rev. 6.

Svensén, M., Bishop, C.M., 2007. Pattern Recognition and Machine Learning. Springer Cham, Switzerland.

Tambouratzis, T., Chernikova, D., Pzsit, I., 2013. Pulse shape discrimination of neutrons and gamma rays using kohonen artificial neural networks. J. Artif. Intell. Soft Comput. Res. 3.

Varolgüneş, Y.B., Bereau, T., Rudzinski, J.F., 2020. Interpretable embeddings from molecular simulations using Gaussian mixture variational autoencoders. Mach. Learn.: Sci. Technol. 1 (1), 015012.

Xu, L., 1993. Unsupervised learning by EM algorithm based on finite mixture of Gaussians. In: World Congress on Neural Networks (Portland, OR), vol. 2, pp. 431–434.

Yousefi, S., Lucchese, L., Aspinall, M., 2009. Digital discrimination of neutrons and gamma-rays in liquid scintillators using wavelets. Nucl. Instrum. Methods Phys. Res. A 598 (2), 551–555. http://dx.doi.org/10.1016/j.nima.2008.09.028.

Zhou, J., Abdulaziz, A., Altmann, Y., Di Fulvio, A., 2023. Generalized method for the optimization of pulse shape discrimination parameters. Nucl. Instrum. Methods Phys. Res., Sect. A 1050, http://dx.doi.org/10.1016/j.nima.2023.168184.

Zhou, J., Gaughan, N., Becchetti, F., Torres-Isea, R., Febbraro, M., Zaitseva, N., Di Fulvio, A., 2022. Light output quenching in response to deuterium-ions and alpha particles and pulse shape discrimination in deuterated trans-stilbene. Nucl. Instrum. Methods Phys. Res. A 1027, 166287.