VIP Experiment Data Analysis Summary
June - July 2025

Jason Yip
July 28, 2025

1 Abstract

This report presents a data analysis focusing on the years 2021 and 2025. Multiple methodologies are employed and compared to optimize the signal-to-background ratio, with particular emphasis on feature selection and machine learning (ML) approaches. The application of ML to the 2021 dataset yields an improvement in energy resolution (FWHM), achieving 1.84 keV compared to the previous result of 2.2 keV [1]. The ML workflow comprises two stages: initially, a denoising autoencoder (DAE) is utilized to reconstruct pulses with significant noise, followed by a convolutional neural network (CNN) that selects only physically meaningful pulses.

Goal. The objective is to lower the energy threshold from 25 keV while effectively rejecting background pulses and enhancing sensitivity to lower energies. In order to probe potential Pauli Exclusion Principle (PEP) violations, it is essential to acquire data down to at least 7 keV. However, as the energy decreases, noise becomes increasingly significant, necessitating the development of advanced methods to extract pulses underlying noise. This challenge is explicitly addressed and focused in this analysis.

2 Background

This section provides a brief overview of previous results and introduces the ML-based approaches utilized in this work.

Previous Results. Piscicchia et al. previously performed pulse shape discrimination on the 2021 dataset using three feature selection techniques [1]:

1. Comparison of mean values at the beginning and end of the pulse, as well as the pulse height and the mean value at the end of the pulse.
2. χ^2 discrimination at the end of the pulse.
3. Pulse derivative selection by fitting a Gaussian function (see Figure 1).

The resulting energy spectrum is shown in Figure 2. The energy resolution was determined to be 2.2 keV (FWHM) at the ^{210}Pb peak, and the intrinsic low-energy threshold was found to be approximately 25 keV.

Denoising Autoencoder (DAE). The denoising autoencoder (DAE) is designed to recover underlying signals from noisy data by learning an efficient representation through sequential encoding and decoding processes. The application of DAEs has demonstrated significant improvements in energy resolution and the extraction of key features for pulse shape discrimination in data collected from Germanium detectors [2].
This approach is particularly valuable in our context, as noise becomes increasingly significant at low energies, the region of interest for the VIP experiment. A method capable of reconstructing and extracting features from low-energy pulses would therefore be highly advantageous. During training, the model receives input data X (pulses with artificially added noise) and targets Y (the corresponding clean pulses), as visualized in Figure 3.

Convolutional Neural Network (CNN). A convolutional neural network (CNN) is designed to automatically learn and extract important features from input data through layers of convolution and pooling operations (see Figure 4). This makes CNNs particularly effective for analyzing

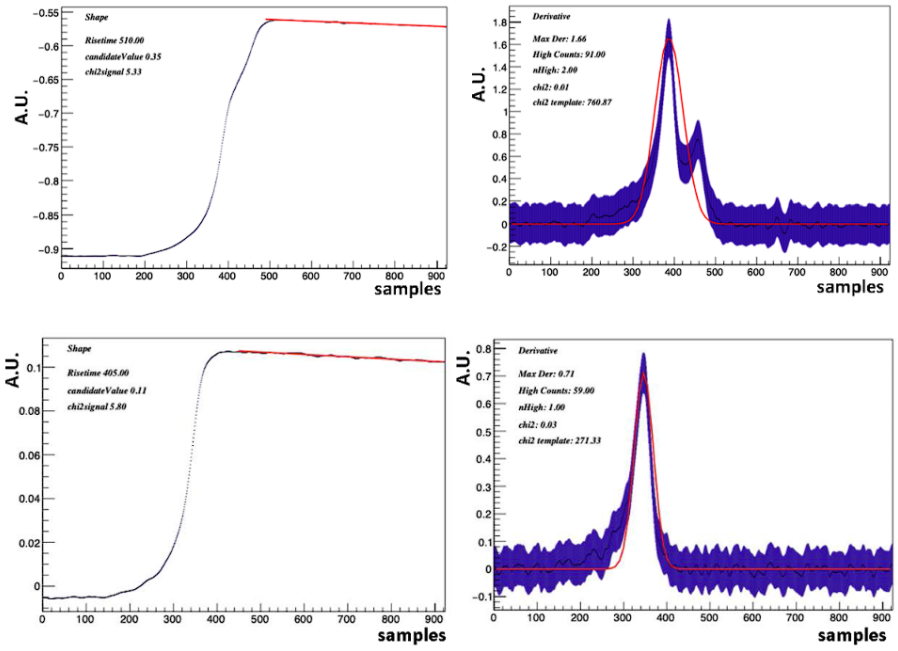


Figure 1: Top panel: a typical multi-site event that passes selection cut 2 but not selection cut 3. Bottom panel: a signal event selected by the algorithm. Figures on the right show the derivatives of the corresponding curves on the left. The full acquisition time window is 2560 ns; each acquisition contains 1024 samples, with a sampling rate of 400 MHz (one sample corresponds to 2.5 ns).

structured signals, such as pulse shapes from detectors. In our case, the CNN can be trained to distinguish between signal and background pulses, using X (pulse data) as input and Y (corresponding labels) as targets.

3 Methods

3.1 Features definitions

Various features are defined for pulses from each event:

1. **Rise Time.** Rise time is defined as 10% to 90% from the maximum value of the pulse after baseline is subtracted.

2. **Decay Time.** Decay time is defined as 90% to 10% from the peak. In both data 2021 and data 2025, the decay time is much longer than the time window and cannot be calculated. On the other hand, in data 2022 and data 2023, due to the preamplifier and wideband amplifier, the tail of the pulses is shaped such that decay time can be calculated.

3. **Peak Time.** Peak time is defined as the interval between rise time and decay time, which is 90% from each side of the peak. This is particularly useful for data 2022 and 2023 as it tends to be longer for saturated events. However this is not as discriminative for data 2021 and 2025 as the shape of pulses is significantly different.

4. **Npeak.** This is the number of peaks detected in the **normalized 1st derivative** space to capture pile-up events. The normalization is essential due to the fact that otherwise the parameters for the peak-finding algorithm need to be finetuned for every individual pulse. The algorithm used to find the number of peaks is through SciPy find_peaks [3].

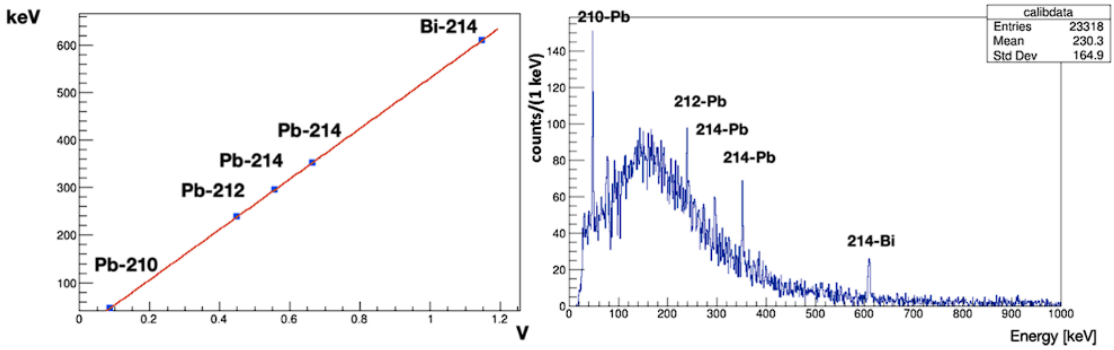


Figure 2: Left: Energy calibration line. **Right**: Energy-calibrated spectrum.

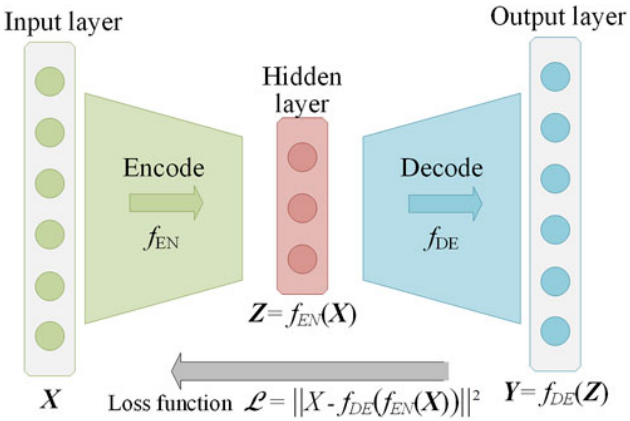


Figure 3: Visual representation of a Denoising Autoencoder (DAE).

5. **FWHM Time.** The full-width-half-maximum time (FWHM) time is defined as the time interval of the width at half maximum on both side of the peak. For data 2021 and data 2025, this should be defined in the **normalized derivative** space of each pulse as it has a higher probability to capture pile-up events that pass through the Npeak selection than peak time. This is very similar to peak time. Figure 5 shows a visual example of some feature definitions.

6. **L1 norm**. In pulse shape analysis, comparing the shape of an observed pulse to a reference pulse is a common method for identifying or classifying signals. One way to quantify the difference between two pulse shapes is by using the L1 norm, which sums the absolute differences between corresponding points of the two signals [4]. This is calculated after subtracting baseline. It is particularly robust to detect outliers as this value is larger for outliers. For data 2021, the reference pulse is calculated by taking the average of pulses within 1σ of rise time and FWHM time, with only 1 peak detected. Figure 6 show the reference pulse calculated.

After defining features, a histogram of each feature is constructed for feature selections. The following criteria need to be satisfied for a pulse to be considered as signal in the non-ML based approach, otherwise background:

1. Rise time within 1σ
2. Decay time within 1σ (applicable to data 2022 and 2023 only)
3. FHWM time within 1σ

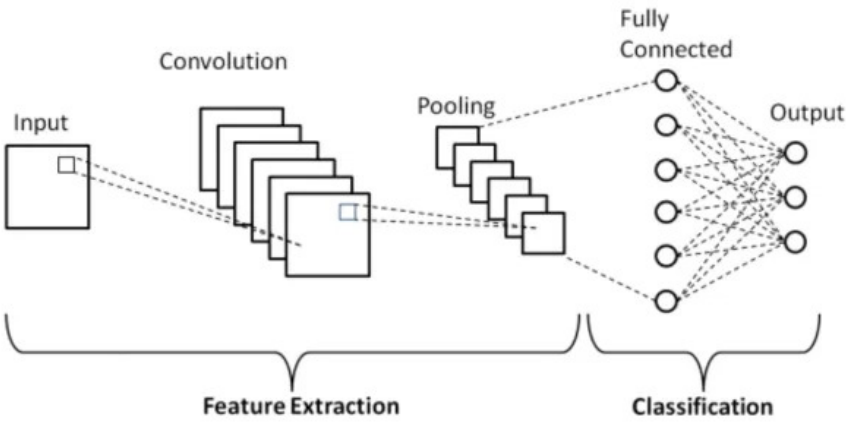
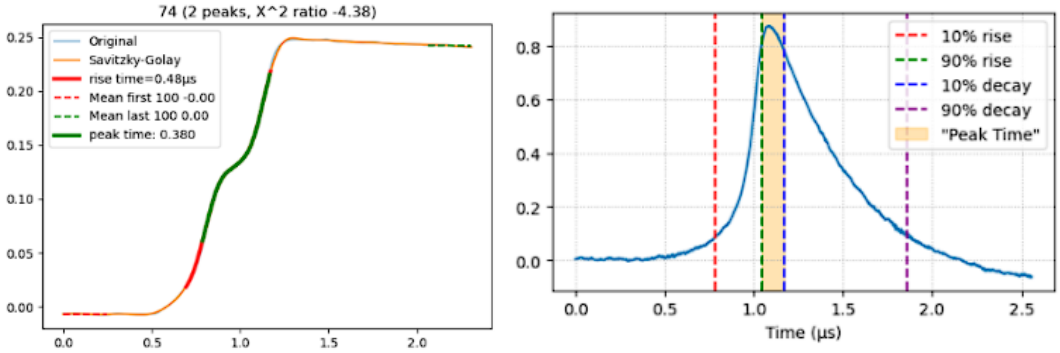


Figure 4: A representation of the structure of a CNN.



(a) pulse from data 2021 **(b)** pulse from data 2022

Figure 5: Visual representation of feature definition from **(a)** 2021 where orange is after smoothing with Savgol filter [3] and **(b)** 2022 data.

4. L1 norm within 1σ
5. A pulse must have exactly 1 peak.

Peak time is not used as FHWM has been proven to be more discriminative than peak time. Direct extraction of features on pulses is enough in the absence of noise. Unfortunately, in our case the region of interest lies at very low energy, where noise is very significant (see Figure 7 for an example). Extracting features directly from raw pulses can be very inaccurate and lead to rejections of signal events. If one can accurately reconstruct pulses from noise, feature selections and/or ML-based approaches can be much more discriminative.

3.2 DAE Reconstruction and CNN Classification

In the ML-based approach, a DAE reconstruction is followed by a CNN classification. This reduces the computational power needed and the complexity of the CNN as the reconstructed pulses are much more distinguishable between what a signal and background event is, therefore lowering the complexity of the CNN.

3.2.1 DAE

To accurately reconstruct noisy pulses, synthetic pulses are generated and used as training dataset. A direct use of pulses from data as training dataset was explored but did not show improvement

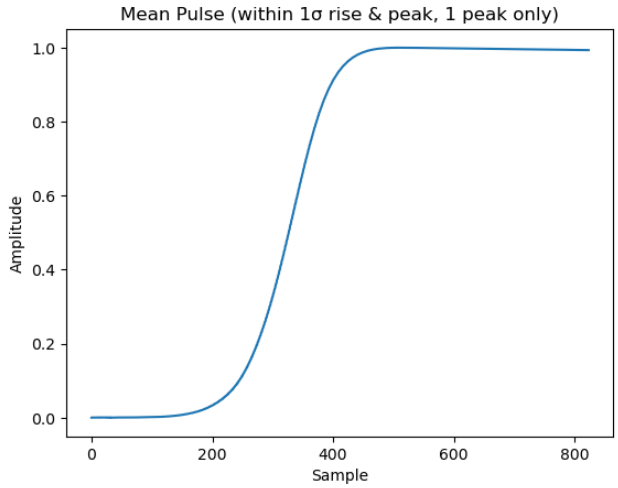


Figure 6: The 2021 normalized reference pulse for L1 norm calculation.

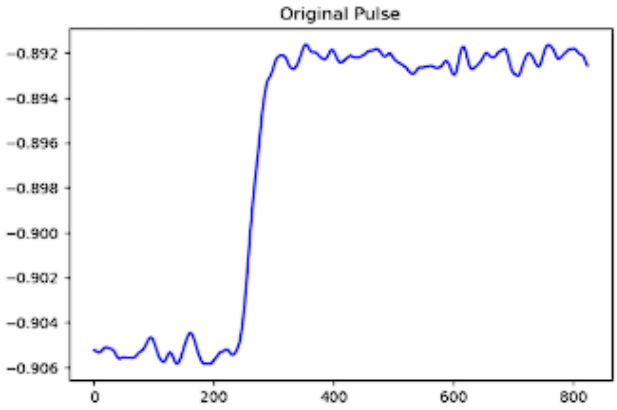
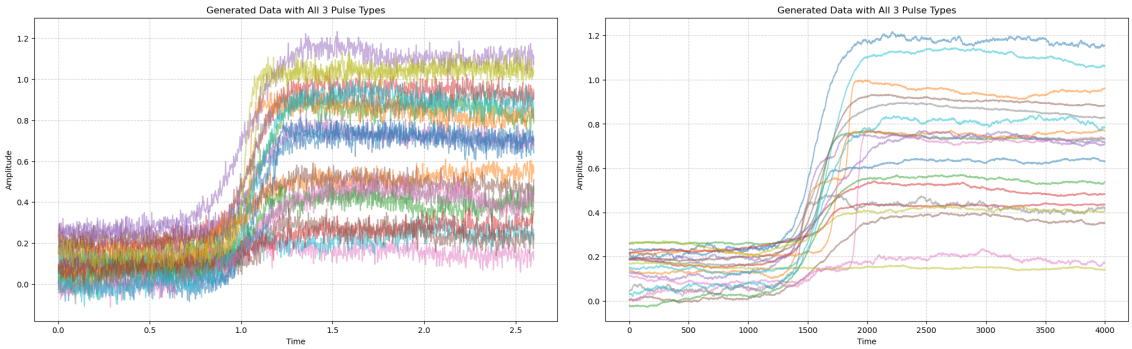


Figure 7: An example of a low-energy signal event with significant noise.

compared to the smoothing algorithm *savgol_filter* from SciPy. Three types of synthetic pulses are generated: normal, pile up, and slow rise. Two set of training data are generated, clean pulses and noisy pulses (one can simply add noise to the clean pulses). The noise level is adjusted to match the noise level in the data. See Figure 8 for an example of different synthetic pulses. Pulses with two noise level are generated, and the final training dataset consists of 50% each, with a total of 5000 pulses.

Each pulse is then normalized by amplitude found using a trapezoidal filter. Same normalization method is applied prior to every ML model. A 1D convolutional autoencoder is employed to reconstruct input signals. The encoder consists of three Conv1D layers with decreasing filter sizes (64, 32, and 16 filters) and kernel sizes of 16, 8, and 4, each followed by ReLU activation and max pooling, which reduces the sequence length at each stage. This compresses the input into a compact encoded representation. The decoder then mirrors this process with Conv1D layers (16, 32, and 64 filters) and upsampling layers to gradually restore the original sequence length. After upsampling, a Cropping1D layer is used to precisely match the output length to the input. The final output layer is a Conv1D layer with a single filter and linear activation, producing the reconstructed signal. The model is trained using the Adam optimizer [5] and mean absolute error (MAE) loss function. After reconstruction, features can be extracted accurately for non-ML approach, and the reconstructed pulses can be used for ML-based approach.



(a) Synthetically generated noisy pulses (x axis in unit of microsecond)

(b) Synthetically generated noisy pulses with reduced noise level (x axis in unit of sample)

Figure 8: Examples of some synthetically generated noisy pulses. In **(b)**, the length of each pulse is matched to data 2025.

3.2.2 CNN

For supervised CNN classification, it is essential to construct high-quality training datasets representing both good and bad pulses. Considerable effort was devoted to this step, employing various feature selection techniques in combination with low reconstruction error criteria. Despite these efforts, some outliers persisted in both the good and bad datasets. Subsequently, thanks to Simone, a ground-truth dataset was established through manual labeling, along with strict feature selection (requiring all features within 1σ, very low reconstruction error from the DAE, and very low L1 norm). For CNN classification, the last 100 samples of each reconstructed pulse were omitted to avoid artifacts caused by signal padding. The resulting .pkl file contains features, manual labels, original pulses, and reconstructed pulses, and is intended for use in future analyses. This file can be accessed using *Pandas* [6].

A 1D CNN for pulse classification is used, with the input being both the reconstructed pulses and the corresponding normalized derivative. The model consists of several Conv1D layers with increasing filter sizes, each followed by batch normalization and max pooling to progressively extract and condense features from the input signal. After flattening the feature maps, a dense layer with dropout is used for further processing, and the final output layer uses a sigmoid activation for binary classification. The model is trained with the Adam optimizer [5] and binary cross-entropy loss. Each pulse is assigned a number between 0 and 1, with 0 being the absolute background and 1 being the absolute signal. The optimal threshold is determined by using the Receiver-operating characteristic (ROC) curve to systematically evaluate all possible thresholds and selecting the one that provides the best trade-off between true positive and false positive rates.

4 Results

The original spectrum is shown in Figure 9. It is clear that after applying a trapezoidal filter, the background is already reduced, and the energy resolution (FWHM) determined at the ^{210}Pb is 2.1 keV. The peaks used for calibrations are ^{210}Pb at 46 keV, ^{214}Pb at 295 keV, ^{214}Pb at 352 keV and ^{214}Bi at 609 keV. All spectrum are fitted by likelihood due to the low count of events.

4.1 Non-ML Approach

The feature selections for non-ML approach are those described in section 3.1, along with manual label selection. The $S/\sqrt{B}$ and the energy resolution are determined to be 23.04 and 2.04 keV. See Figure 10 for calibration result and Figure 11 for spectrum. Feature extraction with DAE is detailed in Section 4.2.1.

For a more detailed spectrum, see Figure 12. It is evident that many lower-energy events are excluded by the selection criteria, resulting in reduced sensitivity in low energy.

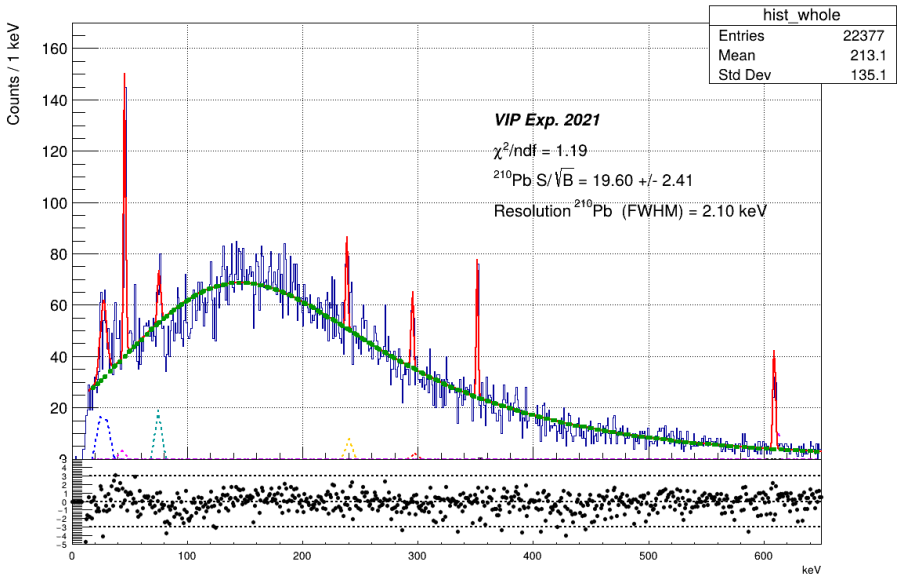


Figure 9: The 2021 total spectrum before any data cleaning, along with pull plot at the bottom.

4.2 ML Approach

4.2.1 Feature extractions

The result of the DAE reconstruction is shown in Figure 13. This is a successful denoise model.

As one can see, if features are extracted directly from the original pulse, it would be very inaccurate and could lead to false background rejection. For example, Npeak in the derivative of the original pulse (bottom right of Figure 13) would be almost impossible to have 1 peak only, but finding Npeak after denoising accurately would yield only 1 peak (bottom left). This pulse corresponds to roughly 24.7 keV. See Figure 14 for the reconstruction error of the DAE on the validation dataset, which characterizes how accurate the model reconstruct pulses. The larger the error the more likely the pulse is anomalous. However this is not always the case, as for noisy signal the noise would cause larger reconstruction error than less noisy signal, as one can see in Figure 15.

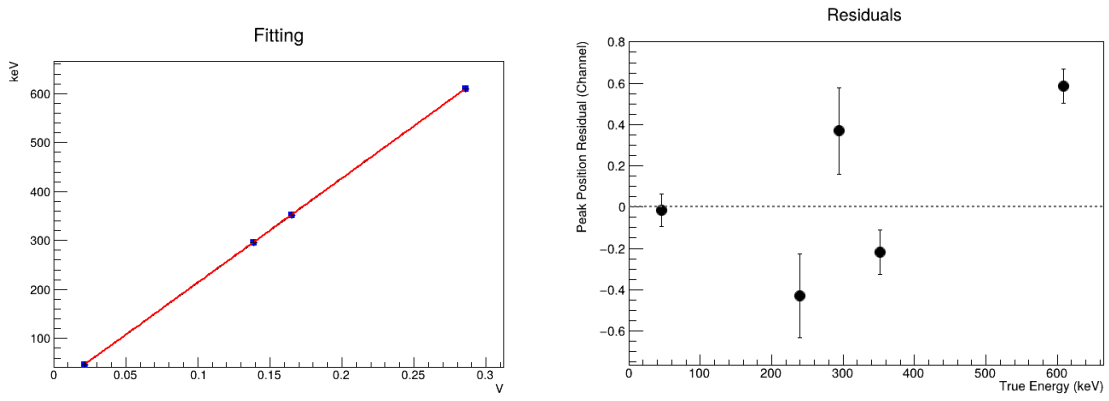
After DAE reconstruction, features can be extracted accurately even in the presence of noise. See Figure 15 for an example. The distributions of features also receive corrections. See Figure 16 for the distributions of features.

The resulting energy spectrum through the chain of DAE and CNN has also shown an increase in signal-to-background ratio. The $S/\sqrt{B}$ and the energy resolution are determined to be 24.08 and 1.84 keV. The performance is very comparable and is easily reproducible, and this method should be used if one desires reproducible results. After strict selections discussed in Section 3.2.2, a ground-truth good and bad dataset is obtained, and a simple CNN is trained on the reconstructed ground truth. The training performance is characterized by a high AUC of 0.999, as shown in Figure 17.

The resulting energy spectrum is presented in Figure 18, with a more detailed view shown in Figure 19. As illustrated, the use of the CNN is able to recover lower-energy events, resulting in higher sensitivity at lower energies.

5 Discussion

This method, which combines DAE and CNN, not only reduces computational requirements but also improves the accuracy of feature extraction. The application of the trapaziodal filter along with feature selections or ML has led to a substantial increase in signal-to-background ratio compared



(a) **(b)**

Figure 10: **(a)** Calibration function and **(b)** residuals

to the previous analysis method, with the latest method yielding 23.04 ± 2.22 and 24.08 ± 2.43 compared to the previous 14.9 assuming similar uncertainty. The improvement corresponds to about a 2.5-2.7σ effect. The two methods perform similarly within uncertainties but the ML-based approach exceeds with its reproducibility. In my view, DAE is especially powerful and well-suited for the VIP experiment, as our primary interest lies in the low-energy spectrum where noise is significant. Without the benefit of automatic background rejection as in the 2025 data acquisition (discussed later), a DAE may be necessary to enhance sensitivity at lower energies. This approach is also likely to be effective for the 2022–2023 data, and its application should be considered if appropriate. While other ML methods for background rejection may exist, it is important to ensure that any chosen method is motivated by the physics of our specific low-energy region.

For the 2022–2023 data, a wideband amplifier was implemented, resulting in the observed Gaussian-like pulses (see Figure 5b). Although amplitude can still be measured to construct the energy spectrum, a challenge arises in handling saturated events (roughly those with amplitude > 1.25). Ideally, such events should be rejected; however, doing so would lead to a lack of events in the higher-energy region of the spectrum, which should require further discussions on spectrum construction.

Based on discussions with Alberto, the new digitizer is capable of rejecting background events with zero energy channels. Nevertheless, further consideration is required regarding the data analysis workflow for the new acquisition system. Figure 20 presents the spectrum obtained as of July 15, 2025, comprising 1,619,986 total events.

With appropriate normalization and processing, the current DAE model should be able to reconstruct all pulses in the 2025 dataset, and the CNN model is expected to reject most background events. However, additional discussion is needed to determine whether further data cleaning is necessary. A preliminary result of this ML-based approach on the 2025 data is shown in Figure 21.

Notably, in the 2021 data, the "pumping" noise from the power supply was absent, whereas it is present in the 2025 data. The initial DAE model was not trained on this type of noise, which led to poor reconstruction performance (as illustrated in Figure 22), but it was subsequently retrained to enable accurate signal reconstruction under this type of noise. It is worth noting that after the implementation of a linear power supply on July 17, 2025, the pumping noise should not be present anymore.

While feature extraction on the 2025 data is feasible, it poses significant demands on computational and memory resources; for instance, processing the entire dataset acquired as of July 17, 2025, using a NVIDIA A100-SXM4-40GB GPU took less than two minutes to apply the ML model, but required more than 300 GB of memory.

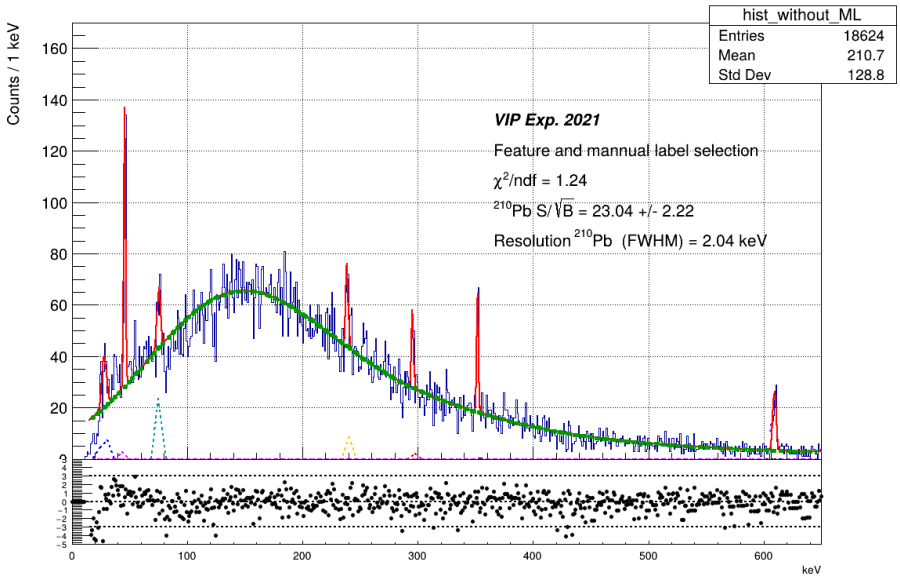


Figure 11: Energy Spectrum with feature selections and manual labels, with pull plot at the bottom.

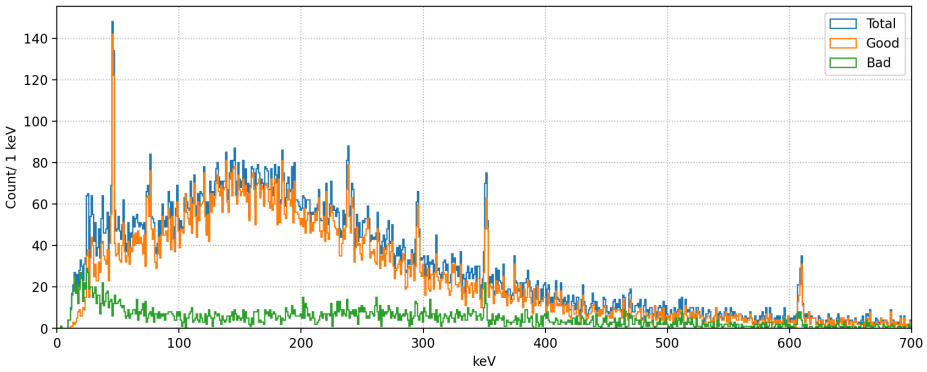


Figure 12: More detailed energy spectrum with feature selections and manual labels.

6 Conclusion

Although the 2021 dataset does not provide sufficient statistics to observe potential PEP violations, the results demonstrate that ML-based methods are highly effective in discriminating background events and enhancing sensitivity to low-energy signals. In comparison to ground-truth background rejection (Figure 11), the ML-based approach achieves comparable performance and exhibits a notable capacity to recover signals amid significant noise, particularly for low-energy events. Furthermore, the integration of DAE and CNN architectures reduces both the computational demands and the complexity associated with training the CNN. This method can be easily extended to future analyses.

7 Future directions

The current machine learning model (CNN) is a supervised approach, requiring labeled signal and background datasets for event classification. However, it is also possible to develop unsupervised models that can independently learn event features without the need for ground-truth labels. Such models are currently under development, for example, using variations of autoencoders (AEs) or

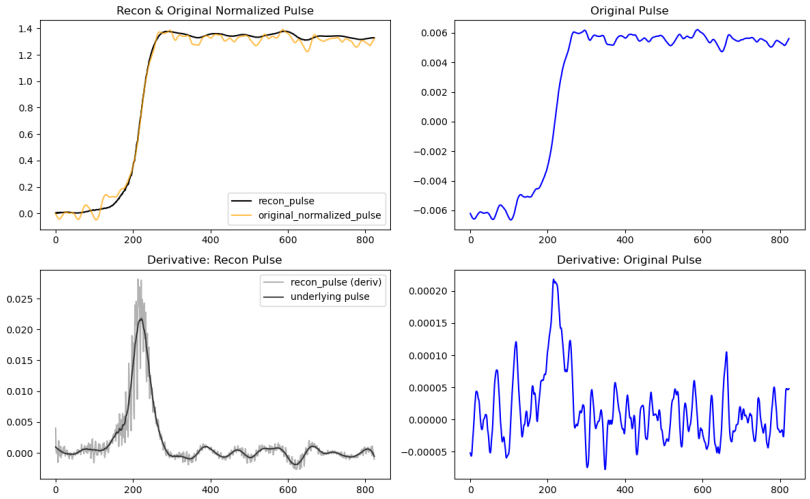


Figure 13: Top left: normalized original (orange) and reconstructed (black) pulse. **Top right**: original pulse before normalization. **Bottom left**: derivative of reconstructed pulse. **Bottom right**: derivative of original pulse.

other architectures that focus on pulse characteristics. While it should be feasible to extract features from the 2022 and 2023 datasets, it might not be as feasible from the 2025 dataset.

A popular alternative is the semi-supervised approach, in which a supervised method is combined with an unsupervised method [4, 7, 8]. For example, in early June 2025, I performed a semi-supervised ML-based discrimination on the raw 2022–2023 data by first training an AE to reconstruct pulses in an unsupervised fashion and then applying a simple deep neural network (DNN) in a supervised manner to the latent space, where features are compressed. Although initial results were promising, the lower energy spectrum was not selected due to noise (see Figure 23 for reference). It is worth noting that, at that time, the energy spectrum was constructed using the pulse area rather than the amplitude, as discussed in Section 5.

Similar strategy of the chain DAE to CNN can be applied to data 2022-2023 and compare performance, with the focus on lower energy region. While some methods might demonstrate excellent performance, one should always check whether the low-energy events are being selected, as those events often behave like background and can be easily rejected. In the new data acquisition, a multichannel analyzer (MCA) has been adapted to be able to produce two channels for each event, with channel 0 having four times higher resolution than channel 2 but a lower energy range up to 1 MeV than channel 2 up to 4 MeV. Calibration can be done with channel 2. Energy spectrum from channel 0 can then be used to extract physics and put new bounds on θ-Poincaré model (in particular the non-commutativity scale Λ_{QG}) and the measurement problem.

Figure 14: MSE reconstruction error of DAE on validation dataset.

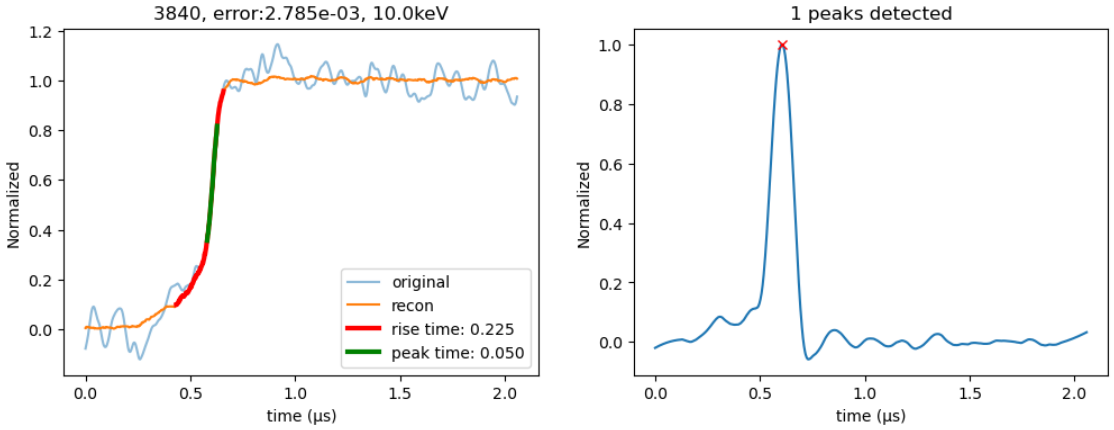
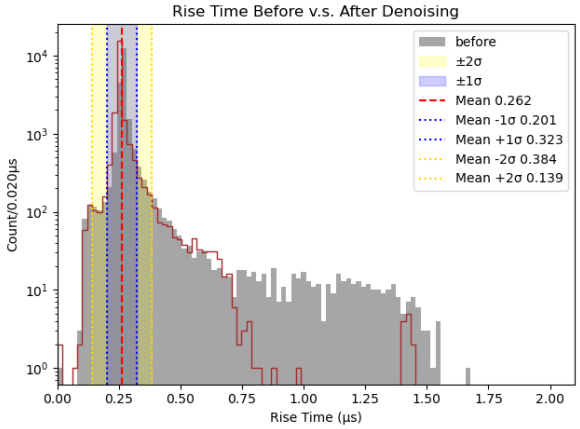
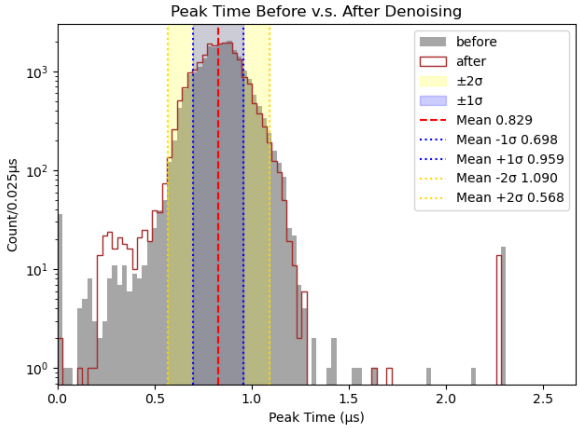


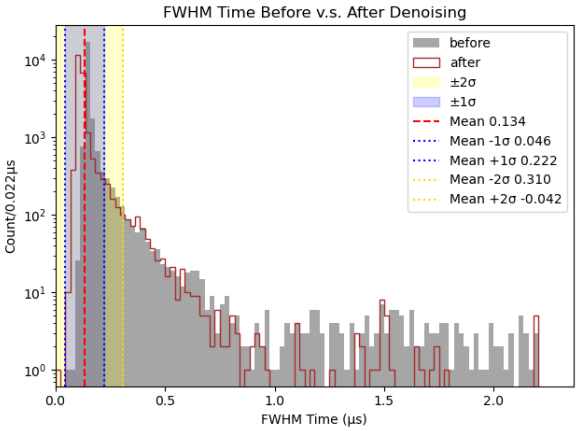
Figure 15: Left: normalized pulse before and after reconstruction, with rise time and peak time shown explicitly. **Right**: normalized derivative of reconstructed pulse.



(a) Rise time distribution



(b) Peak time distribution



(c) FWHM time distribution

Figure 16: Each distribution shows the mean before and after reconstruction.

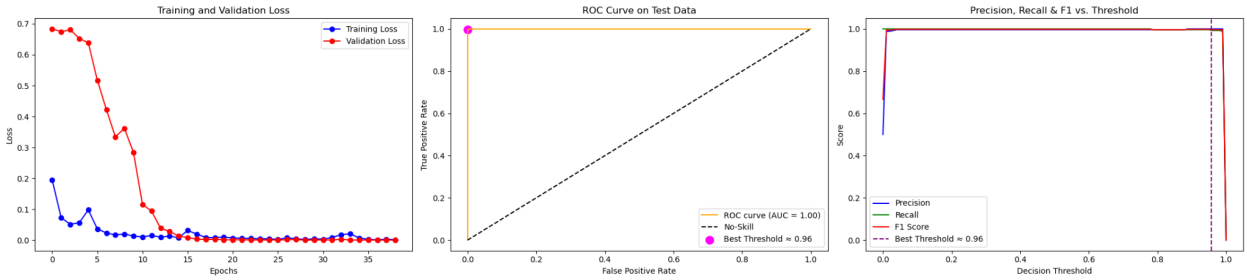


Figure 17: **(a)** loss vs epochs, **(b)** ROC curve, and **(c)** precision, recall and F1 curve.

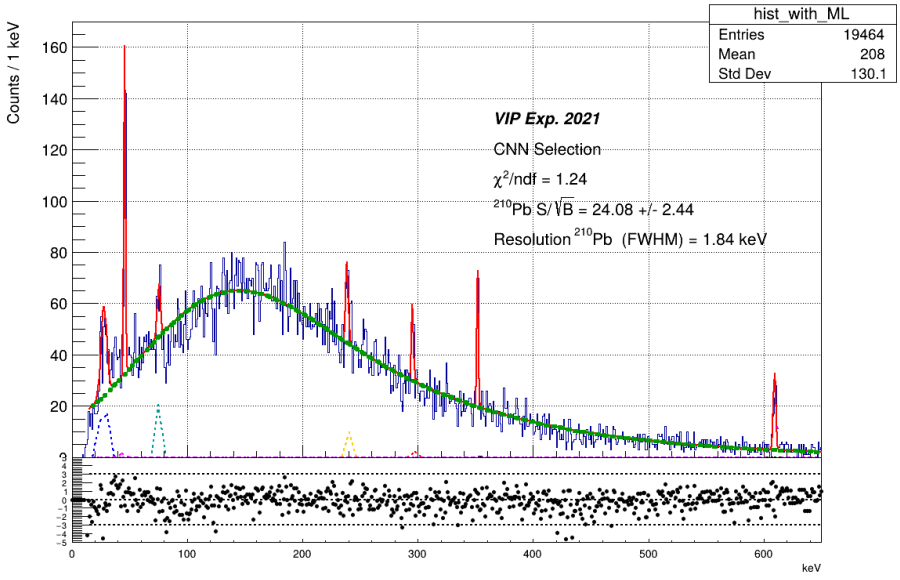


Figure 18: Energy Spectrum with ML approach.

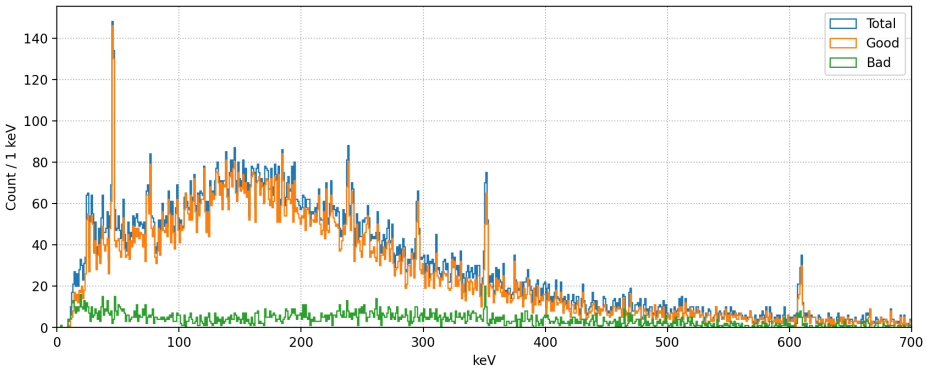


Figure 19: More detailed energy spectrum with ML selection.

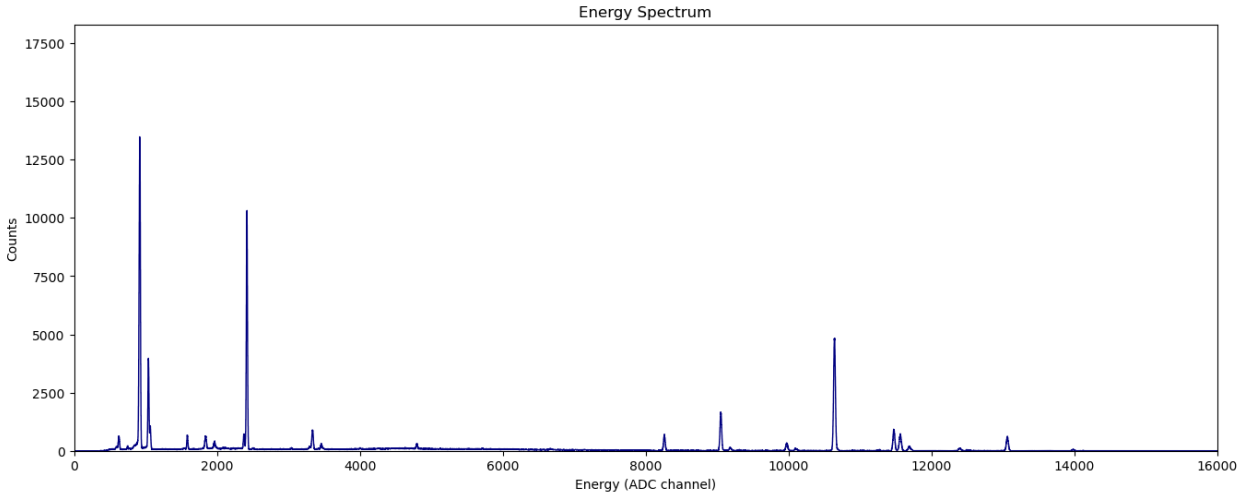
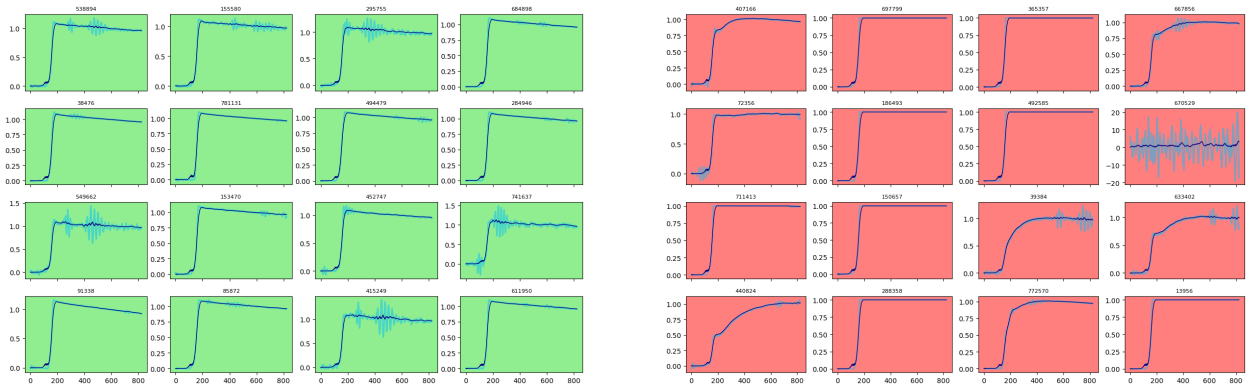
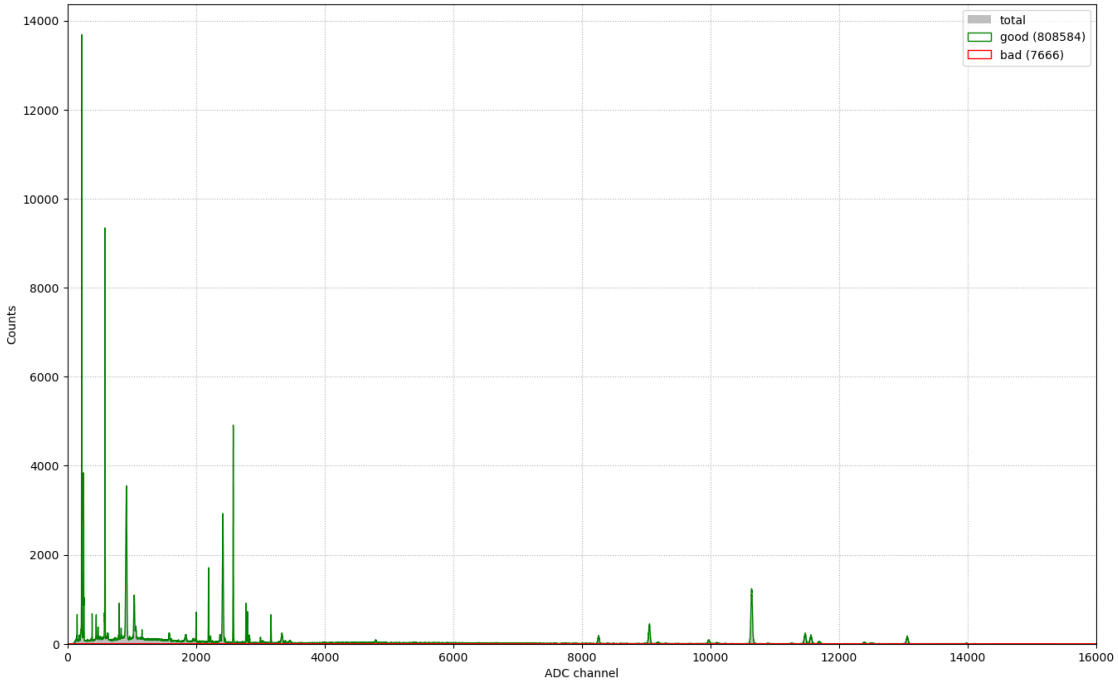


Figure 20: Energy spectrum acquired as of July 15, 2025.



(a) **(b)**



(c)

Figure 21: A preliminary result with ML-based data cleaning on data 2025, where **(a)** contains signal pulses, **(b)** contains background pulses, and **(c)** is the energy spectrum on partial data 2025.

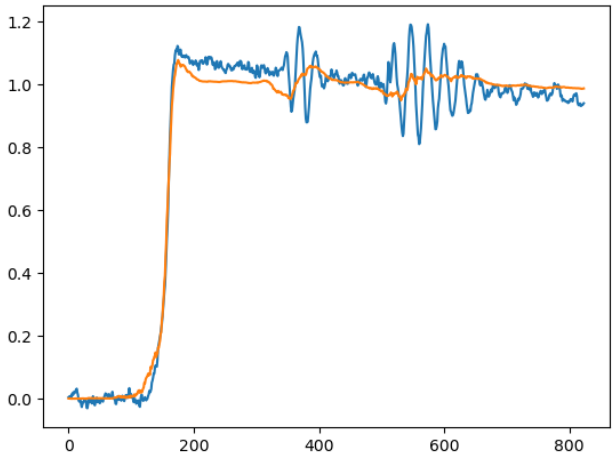
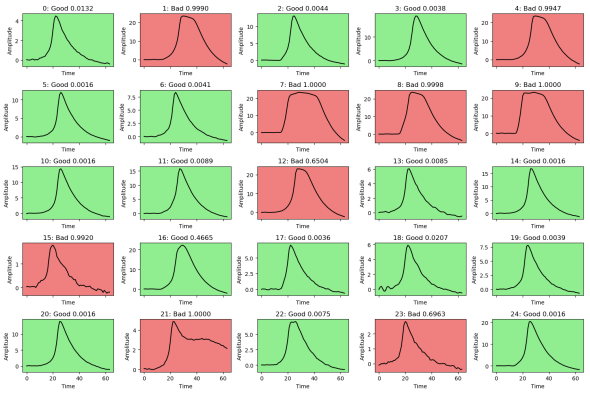
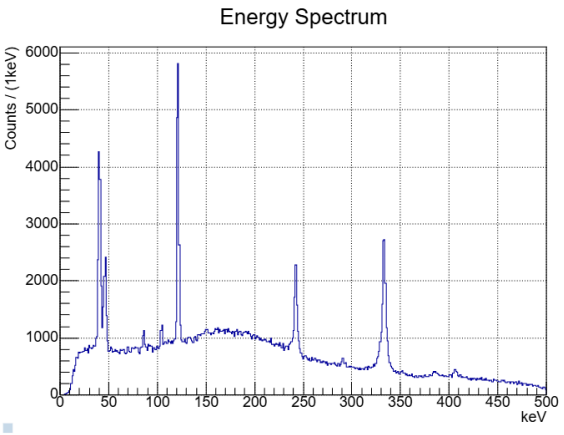


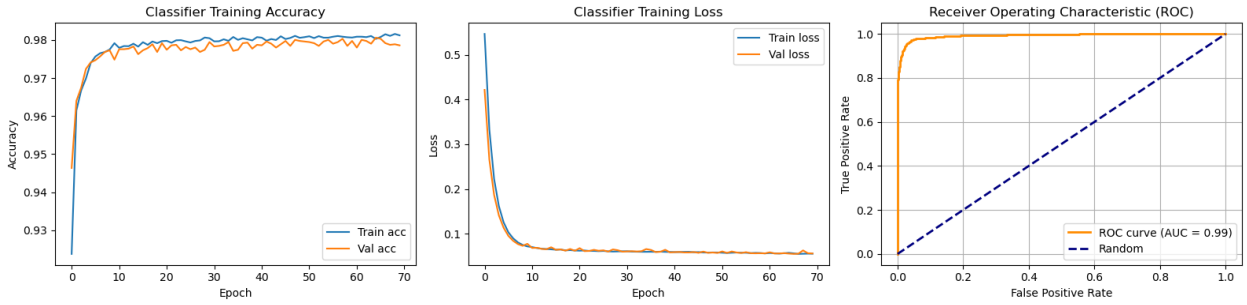
Figure 22: Blue is a pulse from data 2025 and **orange** is the reconstructed pulse using 2021 DAE model before training on the pumping noise. This pulse has a length of roughly 4000 sample but is downsampled to 925 to match input shape of the DAE model.



(a) Pulses 2022-2023



(b) Spectrum 2022-2023



(c) ROC 2022-2023

Figure 23: Overview of **(a)** pulses, **(b)** spectrum, and **(c)** ROC for 2022-2023 data.

References

[1] K. Piscicchia *et al.*, "Optimization of a BEGe Detector Setup for Testing Quantum Foundations in the Underground LNGS Laboratory," *Condens. Mat.*, vol. 9, no. 2, p. 22, 2024.

[2] M. R. Anderson, V. Basu, R. D. Martin, C. Z. Reed, N. J. Rowe, M. Shafiee, and T. Ye, "Performance of a convolutional autoencoder designed to remove electronic noise from p-type point contact germanium detector signals," *The European Physical Journal C*, vol. 82, Dec. 2022.

[3] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[4] G. Baccolo, A. Barresi, D. Chiesa, A. Giachero, D. Labranca, R. Moretti, M. Nastasi, A. Paonessa, M. Picione, E. Previtali, *et al.*, "Machine learning-assisted techniques for compton-background discrimination in broad energy germanium (bege) detector," *The European Physical Journal C*, vol. 85, no. 3, p. 332, 2025.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[6] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445, pp. 51–56, Austin, TX, 2010.

[7] E. León, A. Li, M. B. Schott, B. Bos, M. Busch, J. Chapman, G. Duran, J. Gruszko, R. Henning, E. Martin, *et al.*, "Machine learning-powered data cleaning for legend: a semi-supervised approach using affinity propagation and support vector machines," *Machine Learning: Science and Technology*, vol. 6, no. 1, p. 015064, 2025.

[8] P. Holl, L. Hauertmann, B. Majorovits, O. Schulz, M. Schuster, and A. Zsigmond, "Deep learning based pulse shape discrimination for germanium detectors," *The European Physical Journal C*, vol. 79, no. 6, p. 450, 2019.