Machine Learning for the Cluster Reconstruction in the CALIFA Calorimeter at R^{3}B

Author name[a]

[a]organization=University of the Moon,addressline=, city=Earth, postcode=, state=, country=

Abstract

The R^{3}B experiment at FAIR studies nuclear reactions induced by high-energy radioactive beams. A key detector, the CALIFA calorimeter, consists of 2544 CsI(Tl) scintillator crystals designed to detect gamma rays and light charged particles with high angular resolution and Doppler correction.

Precise cluster reconstruction from sparse hit patterns is essential for accurate energy determination. The current algorithm uses fixed cluster sizes and geometric thresholds. To enhance performance, it was compared with machine learning techniques such as Agglomerative Clustering, incorporating previously unused timing information from CALIFA. Additional improvements were explored using an Edge Detection Neural Network.

This study, based on Geant4 simulations, demonstrates notable gains in reconstruction accuracy, showcasing the potential of machine learning in high-energy nuclear physics experiments.

Keywords: keyword 1, keyword 2, keyword 3, keyword 4

1. Introduction

With the advancements in facilities dedicated to the production of radioactive beams at relativistic energies, such as the Facility for Antiproton and Ion Research (FAIR) at GSI, significant progress has been made in the study of exotic nuclei far from stability[1]. FAIR will provide high-intensity relativistic radioactive beams of rare isotopes with energies reaching up to 1 AGeV, enabling investigations in inverse kinematics with full kinematic reconstruction[2]. A key experimental setup designed for this purpose is the Reactions with Relativistic Radioactive Beams (R^{3}B) Setup, which allows for high-resolution particle spectroscopy. This setup serves as a unique tool for unveiling the structure of nuclei and their reaction dynamics with unprecedented precision.

At the core of the R^{3}B Setup is the CALIFA calorimeter (Calorimeter for the In-Flight Detection of Gamma Rays and Light Charged Particles), a highly segmented detection system composed of more than 2500 CsI(Tl) scintillator crystals that hermetically enclose the target area in the polar angular range of 7° to 140°. This design enables the simultaneous measurement of gamma rays down to 100 keV and light charged particles, such as protons and deuterons, up to several hundred A MeV[3]. To ensure optimal performance, extensive research has been conducted to refine the geometric design, minimize scattering and energy losses due to the holding structure[4], and develop a dead-time-free data acquisition system capable of handling high-rate experiments[5]. Furthermore, a seamless integration within the R3BRoot framework[6] has been achieved, enabling offline data analysis from the raw data level to the calibrated level and ultimately to the cluster level, where individual hits are recombined for the final energy reconstruction.

This study presents the results of applying machine learning-based graph networks to enhance the energy reconstruction of gamma rays in CALIFA. Using simulated Geant4 data, the performance of standard R3B clustering algorithms is compared to an agglomerative clustering model implemented with SciPy[7] and a generic neural network (NN) architecture, demonstrating the potential of machine learning techniques in improving reconstruction accuracy.

2. Methodology

2.1. Challenges in Relativistic Gamma Spectroscopy

While the detection of light charged particles such as protons typically yields well-localized energy deposits in segmented detector arrays, the detection of gamma rays emitted from reaction products moving at relativistic velocities ($\beta \approx 0.8$) presents significant challenges. These difficulties primarily arise from the inherently sparse and spatially distributed energy deposits resulting from the interaction mechanisms of photons with the scintillator material (see Fig. 1)[8].

At photon energies below approximately 300 keV, the photoelectric effect dominates the interaction cross-section. As the photon energy increases, Compton scattering becomes the predominant process. For photon energies exceeding the pair production threshold ($E_\gamma > 2m_ec^2 \approx 1.022 MeV$), electron-positron pair creation becomes possible and is the dominant interaction mechanism above $\backsim$ 6 MeV.

Compton scattering broadens the clustering by the deflection of the incident gamma ray. According to the Klein–Nishina formula, the scattering is predominantly forward-focused for moderate to high photon energies[10], leading to clusters in neighboring crystals.
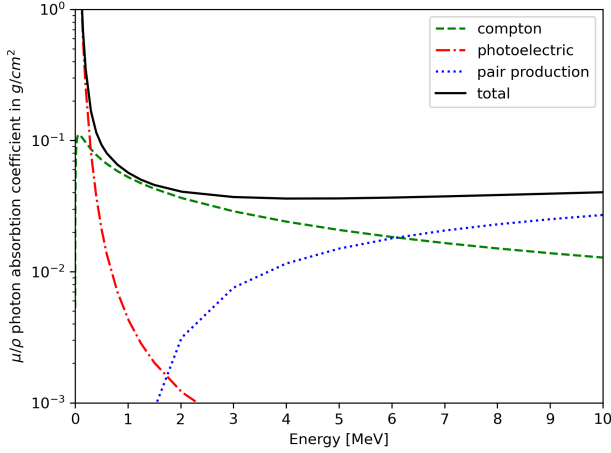
Figure 1: Mass attenuation coefficients for photons in CsI in the range from 100 keV to 20 MeV according to XCOM database[9].

In the case of pair production, which occurs above the $2m_e c^2$ threshold, the resulting annihilation of the positron yields two 511 keV gamma photons. These secondary photons often escape the initial interaction site, leading to a significant fraction of the incident photon's energy being deposited in multiple detector elements.

For gamma rays emitted by nuclei at rest, this behavior gives rise to well-defined single- and double-escape peaks in the recorded energy spectra – corresponding to the escape of one or both 511 keV photons, respectively – if these photons exit the cluster volume without interaction (see Fig. X).

In experiments involving relativistic ions, such as those conducted at R^3B, Doppler broadening significantly distorts spectral features, including single- and double-escape peaks. This effect hinders accurate reconstruction of the photon energy and complicates the extraction of absolute gamma-ray yields and reaction cross sections.

2.2. Data Structure and Standard R3B Clustering Algorithm

In the standard data acquisition (DAQ) configuration, all CALIFA detector hits occurring within a $\pm 4\,\mu$s time window are grouped into a single event. Each individual hit i in CALIFA is represented by a data structure containing the following calibrated information:

- Energy deposit E_i

- Polar angle θ_i

- Azimuthal angle ϕ_i

- Time stamp t_i (via White Rabbit Precision Time Protocol[11])

In the standard R3B clustering approach, the time information t_i is not utilized during the spatial reconstruction of clusters. The initial stage of the clustering algorithm begins by sorting all hits in descending order of energy. A user-defined geometric condition, typically a conical cluster shape with a default aperture of 0.5 rad, is applied. This value has been found to provide an optimal compromise between compact high-energy clusters and more diffuse gamma-ray showers.

The hit with the highest energy defines the seed or center of the first cluster. The algorithm then iterates through the remaining hits and includes each hit in the current cluster if it lies within the specified cone aperture relative to the seed direction. Once the list is fully processed for the current cluster, the next highest-energy unassigned hit becomes the seed of a new cluster. This procedure repeats until no unassigned hits remain.

2.3. Simulation Setup

To evaluate and compare the clustering algorithms presented in this work, simulated data are required, as supervised machine learning approaches rely on access to ground truth labels. For this purpose, the R3BROOT framework with a Geant4-based Monte Carlo [12] backend was employed.

The CALIFA detector geometry used in the simulation corresponds to the configuration implemented in early 2024. At that time, the iPhos region (polar angles $19° - 43°$) was fully instrumented, while only the forward half of the Barrel region ($43° - 87°$) was active. The forward-most CEPA region ($7° - 19°$) was not yet equipped.

Gamma-ray energies were sampled from a uniform distribution between 0.3 MeV and 10 MeV. The interaction of the primary gamma rays with the CsI(Tl) scintillation material was modeled using Geant4 transport physics.

To emulate realistic event topologies, three gamma rays were generated per event, resulting in multiple detector hits. Timing information was coarsely simulated by assigning each primary gamma a random emission time within the $\pm 4\,\mu$s event window. The corresponding hit times were then Gaussian-smeared with a standard deviation of 200 ns to reflect typical electronic channel timing variations.

The resulting dataset was split into training and test subsets, comprising 13,000 and 7,000 events, respectively.

2.4. Preformance Metrics

To quantitatively assess the performance of the clustering algorithms presented in this work, a set of four custom metrics was defined. Three of these are event-based, while an optional fourth metric evaluates clustering quality on a per-cluster basis:

- **True Positive (TP)**: All hits in an event are correctly assigned to their respective clusters.

- **False Positive (FP)**: At least one hit in an event is incorrectly merged into a cluster it does not belong to.

- **False Negative (FN)**: At least one hit is not merged into its true cluster and instead forms a spurious cluster.

- **False Mixed (FM)**: An event is classified as false mixed if it contains both FP and FN characteristics—i.e., at least one hit is wrongly assigned, and at least one true cluster is partially reconstructed.

In addition, a cluster-based metric is defined:

- **Well Reconstructed (WR)**: The ratio of correctly reconstructed clusters to the total number of true clusters in the dataset.

These metrics allow a comprehensive evaluation of clustering accuracy, robustness, and failure modes.

2.5. Agglomerative Clustering

To incorporate temporal information into the clustering process—unlike the standard R3B algorithm, which omits it—a generic, well-established method was adopted: agglomerative clustering as implemented in the `fclusterdata` function from the `SciPy` library. This approach enables flat clustering based on hierarchical linkage with a user-defined threshold.

Each hit was mapped into spherical coordinates (θ, ϕ, r), where the radial component r encodes time information. To ensure non-negative radii, the acquisition time window of $\pm 4\,\mu s$ was shifted by $+4.5\,\mu s$. The Ward linkage criterion [13], which minimizes intra-cluster variance, was employed as the distance metric.

The threshold parameter was optimized to yield the best performance according to the custom-defined metrics. A comparison between the standard R3B clustering and the agglomerative method is presented in Fig.2. As shown in Fig. 2, the agglomer-

Figure 2: Comparison of clustering performance on the test dataset using standard R3B clustering (blue) and agglomerative clustering (orange) incorporating hit-time information. The dataset comprises simulated γ-rays with energies uniformly distributed between 0.3 and 10 MeV. The metrics shown include event-based and cluster-based measures: *well_reco*, *true_positive*, *false_positive*, *false_negative*, and *false_mixed*.

ative clustering algorithm demonstrates improved performance both on an event level (true positive rate) and on a cluster level (correctly reconstructed clusters). However, this improvement is accompanied by an increased false negative rate, indicating that the algorithm tends to under-merge hits near the edges of clusters. This limitation motivated the development and application of an edge detection neural network, which is introduced in the following subsection.

2.6. Implementation of the Edge Detection Neural Network

To enhance the clustering performance, particularly at the boundaries of hit distributions, a generic neural network (NN) was developed to perform pairwise classification of detector hits. This model is applied either to individual raw hits or to

hits pre-clustered via agglomerative clustering, on an event-by-event basis.

The model takes 12 input features for each hit pair (i, j): absolute values of energy (E_i, E_j), polar angle (θ_i, θ_j), azimuthal angle (ϕ_i, ϕ_j), and time (t_i, t_j). Additionally, four differential features are computed: $\Delta E = |E_i - E_j|$, $\Delta\theta = |\theta_i - \theta_j|$, $\Delta\phi = |\phi_i - \phi_j|$, and $\Delta t = |t_i - t_j|$. These differential inputs are essential for training stability and convergence. In particular, $\Delta\phi$ resolves the discontinuities caused by the periodicity of the azimuthal angle (e.g., distinguishing between $\phi = 355°$ and $\phi = 5°$), which would otherwise introduce large erroneous differences in angular comparisons.

Of the 12 features, only the hit time is normalized to the $[0, 1]$ interval; all other values are used in their native physical units.

The architecture of the neural network is illustrated in Fig. 3. The 12-dimensional input vector is passed through a fully connected feed-forward network with one hidden layer of 10^3 nodes, followed by a ReLU activation. Two additional hidden layers, each with 10^2 nodes, are applied sequentially. The output layer consists of a single node with a sigmoid activation, yielding a score in the interval $[0, 1]$, where values close to 1 indicate that the hits (or clusters) are likely to originate from the same event cluster.
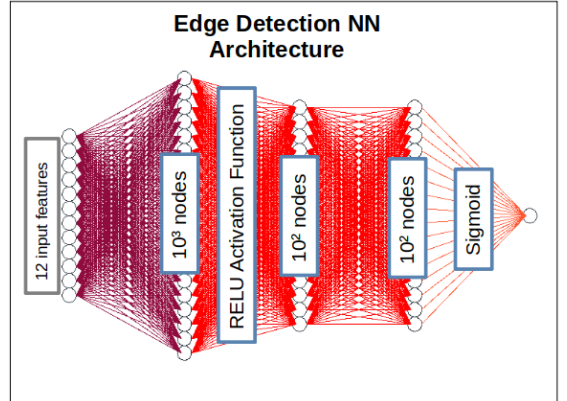

Figure 3: Schematic architecture of the edge detection neural network. The model consists of three fully connected hidden layers. The final output layer comprises a single node with a sigmoid activation function, yielding a prediction score in the range $[0, 1]$. Values closer to 1 indicate that the input hit pair originates from the same physical cluster and should be merged, while values near 0 suggest they should remain separate.

Training is performed using the binary cross-entropy loss function[14][15] and stochastic gradient descent (SGD)[16] with a fixed learning rate of 5×10^{-3}. Given the moderate size of the training dataset, full-batch training is employed without mini-batching. The model is trained for 8×10^5 epochs. After training, a tunable threshold is applied to the prediction scores to classify hit pairs. Final clusters are then formed by grouping all connected hit pairs based on the predicted associations.

The edge detection NN was implemented and tested in three configurations:

- **Plain Edge NN:** The model is applied directly to individual hits without any pre-clustering. All clustering is performed based solely on the NN predictions.

- **Agglomerative + Edge NN:** The data is first clustered using the agglomerative algorithm described in the previous subsection. For each resulting cluster, an energy-weighted center of mass is calculated, replacing individual hits. The NN is then trained on false negative cases (i.e., under-merged hits) from this initial clustering. In application, agglomerative clustering is first applied to the test data, followed by the NN to refine the boundaries and reduce the false negative rate (see Fig. 2).

- **R3B Clustering + Edge NN:** This configuration mirrors the Agglomerative + Edge NN approach, but omits the use of time information. In line with the standard R3B clustering methodology, the time inputs are fixed to a constant value of 1 to reflect their exclusion from the clustering process.

3. Results

Clustering Model	TP	FP	FN	FM	WR
R3B Standard Clustering	60.6	5.3	25.2	8.9	80.4
Agglomerative Clustering	62.8	3.3	32.0	1.9	84.1
Edge Clustering (no time)	63.4	7.2	24.8	4.6	82.4
Edge Clustering (with time)	74.7	3.4	20.5	1.4	89.2
R3B + Edge (no time)	67.4	8.5	16.0	8.0	82.2
Agglo + Edge (with time)	81.3	5.1	12.2	1.5	91.0

Table 1: Summary of performance metrics as defined in Subsection 2.4, evaluated for the different clustering algorithms. The models *R3B Standard Clustering*, *Edge Clustering (no time)*, and *R3B + Edge (no time)* utilize angular and energy information on a per-hit basis for cluster reconstruction. In contrast, *Agglomerative Clustering*, *Edge Clustering (with time)*, and *Agglo + Edge (with time)* additionally incorporate time-of-hit information into the clustering process.

The results of this study are summarized in Table 1, where the ordering of the clustering models reflects the methodological workflow followed during the investigation. The initial comparison was conducted between the baseline *R3B Standard Clustering* algorithm and a generic *Agglomerative Clustering* model that incorporates time-of-hit information.

The agglomerative model shows improved performance over the R3B baseline in terms of both event-level true positives (TP) and cluster-level (WR) values. However, it exhibits inferior performance with respect to the false negative (FN) rate, indicating a tendency to miss relevant hits during reconstruction. This limitation motivated the development of an Edge Detection Neural Network, initially evaluated as a standalone clustering algorithm and subsequently integrated into the agglomerative framework, yielding the combined model denoted as *Agglo + Edge*.

The *Agglo + Edge* model demonstrates superior performance across all evaluated metrics, achieving an overall correct reconstruction rate of 81.3%, significantly outperforming the *R3B Standard Clustering* algorithm, which reaches 60.6%.

To further explore the capabilities of neural network-based clustering approaches, two additional models were evaluated: a standalone *Edge Detection Neural Network* and a hybrid approach combining *R3B Standard Clustering* with edge-based postprocessing (*R3B + Edge*). Notably, both of these models operate without incorporating time-of-hit information, similar to the R3B baseline. Nonetheless, both outperform the *R3B Standard Clustering* in all reported metrics, underscoring the potential of edge-based neural network models for improving cluster reconstruction in high-granularity detector systems.

4. Discussion and Outlook

The results presented in the previous section clearly demonstrate that high-level machine learning approaches, such as the Edge Detection NN, can significantly enhance the accuracy of cluster reconstruction. These models not only reduce distortions in the measurement process but also exhibit increased sensitivity to low-statistics reactions—an important feature for experiments targeting rare processes.

It is noteworthy that even the models, which do not utilize time-of-hit information (similarly to the *R3B Standard Clustering*), outperform the baseline method. This underscores the general effectiveness of neural network-based methods in extracting structural features from detector data.

The inclusion of time-of-hit information proves to be a critical factor for enhancing clustering performance. As this observable is typically available for CALIFA at R^3B, the results of this study support the recommendation to incorporate it into the reconstruction pipeline wherever possible.

Furthermore, these findings are intended to encourage broader adoption of advanced machine learning techniques by experimental groups, particularly in setups involving highly granular detectors. Such tools offer substantial performance benefits and can support more precise event reconstruction.

However, a consideration of computational complexity is essential. Both the *R3B Standard Clustering* and *Agglomerative Clustering* methods exhibit a time complexity of $O(N^2)$[17], where N is the number of hits. The proposed *Edge Detection Neural Network*, whether used in standalone mode or as an auxiliary module (*R3B + Edge*, *Agglo + Edge*), requires additional computational resources. In particular, the neural network involves large matrix operations—for instance, the second fully connected layer with 10^3 input and 10^2 output features—whose computational cost dominates for events with $N \sim O(10^1)$. Consequently, the current implementation is not suitable for online processing and is better suited for detailed offline analyses, where accuracy is prioritized over execution speed.

Future work should focus on optimizing the Edge Detection Neural Network for speed and scalability. Additionally, transformer-based models[18]—capable of analyzing full event topologies—may offer further improvements in clustering accuracy by capturing complex, global features.

Appendix A. Appendix title 1

Appendix B. Appendix title 2

References

[1] Experiments at the Interface of Nuclear, Atomic, and Hadron Physics with FRS at GSI and Super-FRS at FAIR.

[2] Y. Leifels, Status and physics perspectives of FAIR, IL NUOVO CIMENTO 100 (59) (2025) 48.

[3] D. Cortina-Gil, H. Alvarez-Pol, T. Aumann, V. Avdeichikov, M. Bendel, J. Benlliure, D. Bertini, A. Bezbakh, T. Bloch, M. Böhmer, et al., CALIFA, a Dedicated Calorimeter for the R3B/FAIR, Nuclear Data Sheets 120 (2014) 99–101.

[4] H. Alvarez-Pol, N. Ashwood, T. Aumann, D. Bertini, P. Cabanelas, E. Casarejos, J. Cederkall, D. Cortina-Gil, P. D. Fernández, I. Durán, et al., Performance analysis for the CALIFA Barrel calorimeter of the R3B experiment, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 767 (2014) 453–466.

[5] T. Le Bleis, M. Bendel, R. Gernhäuser, W. Henning, P. Klenze, R. Krücken, M. Winkel, A Digital Readout for CALIFA.

[6] D. Bertini, R3BRoot, simulation and analysis framework for the R3B experiment at FAIR, in: Journal of Physics: Conference Series, Vol. 331, IOP Publishing, 2011, p. 032036.

[7] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature methods 17 (3) (2020) 261–272.

[8] H. Kolanoski, N. Wermes, Teilchendetektoren, Springer, 2016, Ch. 13.5.2.

[9] S. Seltzer, XCOM-Photon Cross Sections Database, NIST Standard Reference Database 8, http://www.nist.gov/pml/data/xcom/index.cfm, accessed: 2025-04-10 (2010).

[10] O. Klein, Y. Nishina, Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac, Zeitschrift für Physik 52 (11) (1929) 853–868.

[11] M. Lipiński, T. Włostowski, J. Serrano, P. Alvarez, White rabbit: A ptp application for robust sub-nanosecond synchronization, in: 2011 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, IEEE, 2011, pp. 25–30.

[12] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al., Geant4—a simulation toolkit, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303.

[13] F. Nielsen, F. Nielsen, Hierarchical clustering, Introduction to HPC with MPI for Data Science (2016) 195–211.

[14] S. Mannor, D. Peleg, R. Rubinstein, The cross entropy method for classification, in: Proceedings of the 22nd international conference on Machine learning, 2005, pp. 561–568.

[15] P.-T. De Boer, D. P. Kroese, S. Mannor, R. Y. Rubinstein, A tutorial on the cross-entropy method, Annals of operations research 134 (2005) 19–67.

[16] D. Newton, R. Pasupathy, F. Yousefian, Recent trends in stochastic gradient descent for machine learning and big data, in: 2018 Winter Simulation Conference (WSC), IEEE, 2018, pp. 366–380.

[17] S. Sieranoja, P. Fränti, Fast agglomerative clustering using approximate traveling salesman solutions, Journal of Big Data 12 (1) (2025) 21.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).