PROCEEDINGS OF SCIENCE

Neural Networks for Photon Searches with AugerPrime

Ezequiel Rodriguez[a,b,*] **for the Pierre Auger Collaboration**[c]

[a] *Instituto de Tecnologías en Detección y Astropartículas (CNEA, CONICET, UNSAM), Av. General Paz 1555 (B1630KNA) San Martín, Buenos Aires, Argentina*

[b] *Institute for Astroparticle Physics (IAP), Karlsruhe Institute of Technology, P.O. Box 3640, 76021 Karlsruhe, Germany*

[c] *Observatorio Pierre Auger, Av. San Martín Norte 304, 5613 Malargüe, Argentina*
Full author list: http://www.auger.org/archive/authors_2024_11.html

E-mail: spokespersons@auger.org

Ultra-high-energy photons ($E \geq 10^{17}$ eV) are expected as by-products of interactions between ultra-high-energy cosmic rays (UHECRs) and background radiation fields or galactic matter, as well as from decay of super-heavy dark matter. Despite these various production mechanisms, the diffuse photon flux is too low for direct detection. Consequently, photon searches at UHE must rely on large ground-based detector arrays. In this contribution, we present a method for photon-hadron discrimination based on deep learning algorithms applied to detector simulations within the context of the Pierre Auger Observatory. Our method correlates information from the Surface Detector (SD), sensitive to air-shower particles arriving to the ground, and the Underground Muon Detector (UMD), sensitive to muons with energies above ~ 1 GeV. We chose graph neural networks (GNNs) for their effectiveness in handling the discrimination task, allowing for an easy and flexible correlation of information from the SD and UMD. This approach is particularly suitable for handling the irregular structures found in SD and UMD configurations, where stations may be missing due to technical issues. Using simulations, the performance indicates that the method has strong potential for identifying photons, suffering at most 10^{-4} background contamination at 0.5 signal efficiency. Future studies will delve into how much that background contamination can be diminished.

7th International Symposium on Ultra High Energy Cosmic Rays
(UHECR) 2024
17 November – 21 November, 2024
Malargüe, Argentina

*Speaker

PoS(UHECR2024)111

1. Introduction

The study of the diffuse photon flux produced by ultra-high-energy (UHE) cosmic rays is an ingredient in understanding the origin and acceleration mechanisms of these particles. Measuring the diffuse photon flux, particularly in the energy range of tens to hundreds of PeV, helps probe the interactions of cosmic rays with matter and radiation fields near their sources, providing insights into Galactic and nearby extragalactic phenomena. Diffuse photons, generated through interactions with the cosmic microwave background, serve as indirect tracers of cosmic-ray propagation and composition [1, 3]. An important contribution to the flux is expected from UHE cosmic rays interacting with Galactic disk matter [2]. Furthermore, more exotic scenarios suggest a flux component from the decay of super-heavy dark matter [4]. Overall, the diffuse photon flux at UHE remains a complex subject with many open questions.

At tens to hundred PeVs, the expected photon flux is at least four orders of magnitude below the cosmic-ray flux; therefore, direct detection is unfeasible, making indirect detection by surface arrays a suitable alternative. The Pierre Auger Observatory is sensitive to extensive air showers induced by both hadronic particles and gamma-rays [5]. As a hybrid air-shower detector, the Observatory is comprised of a Surface Detector (SD) and a Fluorescence Detector. Within the SD, we can differentiate between three different triangular grids of water-Cherenkov detectors (WCDs) with different spacing; the SD-1500, SD-750 and SD-433. The SD-1500 having stations spaced at 1500 meters apart, contains the SD-750, which at the same time contains the SD-433 in a nested manner. Every SD station posseses three photomultiplier tubes (PMTs) that measure Cherenkov light emitted by the particles in the shower front crossing the water volume.

The Pierre Auger Observatory is currently undergoing an upgrade known as AugerPrime. As part of this upgrade, the stations of the two nested arrays are paired with buried scintillation modules, composing the Underground Muon Detector (UMD). Each module is composed of 64 scintillation bars and shielded by 2.3 meters of soil. When traversed by charged particles, the bars emit photons, which are transported to silicon photomultipliers (SiPM) using green-shifting optical fibers. The SiPM electronics operate in two modes: a binary mode and an integrator mode, both designed to estimate the number of muons. In this work, we utilize the binary mode. The UMD will enable precise measurements of the muonic component of air showers [6]. Since photons produce air showers with distinct characteristics, such as deeper atmospheric development and lower muon content compared to hadron-induced showers, the UMD is particularly effective in differentiating primary photons from the dominant background of charged cosmic rays.

The data acquired by the Observatory is highly complex, particularly in the spatio-temporal dimension. Deep learning algorithms offer distinct advantages for processing and extracting information from such intricate datasets. In this work, we explore the use of Graph Neural Networks (GNNs) for the task of hadron-photon discrimination. We detail how events are represented as graphs, describe the design of the neural network architecture, and assess the high-level performance of this method.

2. Simulation sets

The training, validation, and test datasets consist of the simulated SD-433 and UMD detector response [7] to air showers generated with CORSIKA [8] using the hadronic interaction model

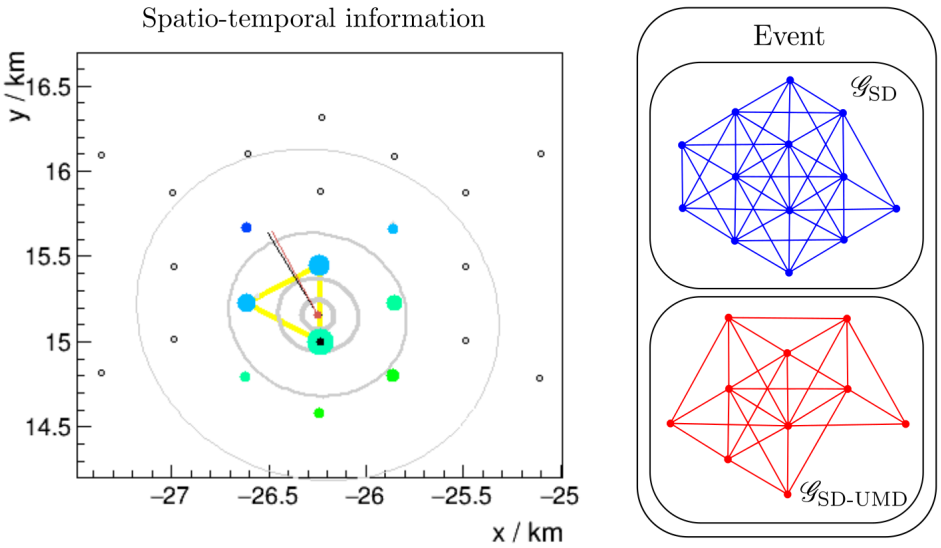

Figure 1: The spatio-temporal information from a combined SD-MD simulated event (left) can be represented by two graphs where nodes represent triggered SD stations and their respective features, including spatial offsets ($\Delta x, \Delta y, \Delta z$) and a time offset ($\Delta t$) relative to the station with the highest signal, the number of working PMTs, the muon density (ρ_μ), and the effective area of the UMD station. In this work, the graph-based representation of an event relies on a graph $\mathcal{G}_{\mathrm{SD}}$ representing the SD-only information, and $\mathcal{G}_{\mathrm{SD-UMD}}$ that besides the SD, integrates additional information from the UMD (right).

EPOS-LHC [9]. The simulations cover primary energies ranging from $\log(E_{\mathrm{MC}}/\mathrm{eV}) = 16.5$ to 17.5, with an isotropic distribution uniform in $\sin^2\theta$ and uniform in azimuthal angle. The simulation process includes modeling the arrival of secondary particles at the SD and propagating those that penetrate the soil to the UMD. The response of each detector is then computed to reconstruct the event, providing a comprehensive representation of the air shower. The resulting outputs include a variety of quantities at both the station and event levels, encompassing temporal and spatial information.

The training dataset comprises 118,000 simulated and reconstructed events, while the validation dataset includes 40,000 events, both containing an equal mix of photon- and proton-induced air showers. To evaluate the network's performance in the presence of a higher muon content, helium and iron nuclei are incorporated into the test dataset alongside photons and protons. The test dataset consists of approximately 40,000 reconstructed photon- and proton-induced events, as well as around 50,000 helium- and iron-induced events each. All events, regardless of the primary particle involved, were reconstructed using the photon energy scale.

2.1 Graph representation of events

In order to feed the neural network the information in the simulations, a graph representation of an event needs to be defined. A simple homogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ can be constructed using two matrices: the adjacency matrix $A \in \mathbb{R}^{N\times N}$ and the feature matrix $X \in \mathbb{R}^{N\times F}$ [10]. The set of nodes (or vertices) is denoted by $\mathcal{V}$, where $|\mathcal{V}| = N$, and the set of edges is denoted by $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$, representing the connections between nodes. The adjacency matrix A encodes the structure of the

graph, where $A_{ij} = 1$ if there is an edge connecting nodes i and j, and $A_{ij} = 0$ otherwise. The feature matrix X contains F features for each of the N nodes, where each row corresponds to the feature vector of a specific node. This representation allows graph neural networks to efficiently model the relationships between entities and their attributes.

In this study, each event is represented by two distinct graphs, $\mathcal{G}_{\text{SD}}$ for the SD and $\mathcal{G}_{\text{SD-UMD}}$ for the combined SD and UMD (Fig. 1). The nodes in these graphs represent triggered stations, connected up to their second neighbors in a 433-meter triangular grid. Each graph is described by three matrices: the adjacency matrix A, the feature matrix X, and an additional matrix X_{traces}, which contains the average trace of the available PMTs.

For $\mathcal{G}_{\text{SD}}$, the feature matrix includes spatial and temporal information such as $\Delta x, \Delta y, \Delta z, \Delta t$ relative to the station with the highest signal, as well as the number of working PMTs (n_{PMTs}). In addition to these features, $\mathcal{G}_{\text{SD-UMD}}$ also incorporates the muon density ρ_μ and the effective area of the associated UMD station. The effective area represents the projected area of the UMD modules in the shower front, accounting for the orientation and geometric alignment of the modules relative to the incoming air shower. On the other hand, the traces matrix X_{traces} includes the first 60 time bins (25 ns each) of the averaged trace from all available PMTs of the SD station. The selected features were specifically chosen to minimize reliance on high-level reconstructed quantities.

The dual representation is necessary due to the inability of neural networks to handle missing UMD stations in an event effectively and the decision to avoid imputation. In some cases, a station with a valid SD measurement may lack a corresponding UMD measurement, either due to technical issues or the absence of a deployed UMD station. Imputation refers to the process of filling in missing data with estimated or substitute values to make the dataset complete. However, this approach can introduce biases or inaccuracies, especially when the imputed values do not accurately reflect the missing measurements. Using two graphs allows the retention of SD nodes even when UMD measurements are missing. In terms of structure, $\mathcal{V}_{\text{SD}-\text{UMD}} \subseteq \mathcal{V}_{\text{SD}}$ and $\mathcal{E}_{\text{SD}-\text{UMD}} \subseteq \mathcal{E}_{\text{SD}}$. Additionally, if we denote the feature set as $\mathcal{F}$, then $\mathcal{F}_{\text{SD}} \subseteq \mathcal{F}_{\text{SD}-\text{UMD}}$. While this approach introduces some redundancy in the information contained in the graphs, it ensures the flexibility required to accommodate diverse SD-UMD configurations and retains all the information. Alternative representations, such as using binary features to indicate missing UMD stations or assigning unphysical values to represent the absence of valid measurements, are found to be less effective.

Due to the size limitations of the dataset, data augmentation is performed to introduce more variability and to prevent overfitting to the training set. The augmentation process was carefully designed to ensure that the reconstructed quantities at the event level remained unchanged for the most part. For example, during the masking of SD stations, the stations that drive the zenith angle estimation were kept intact. Masking refers to the process of intentionally dropping certain stations from an event. By omitting a portion of the information, the goal is to encourage the model to learn more robust features. Augmentation techniques accounted for factors such as PMT/module masking, calibration, and saturation effects, as well as SD/UMD station masking, ensuring robustness in the network's training and evaluation processes.

3. Network architecture and training

The proposed neural network architecture is exhibited in Fig. 2. The design is based on three key sub-components: the Trace Analyzer (TA), the Graph Analyzer (GA), and a set of dense layers

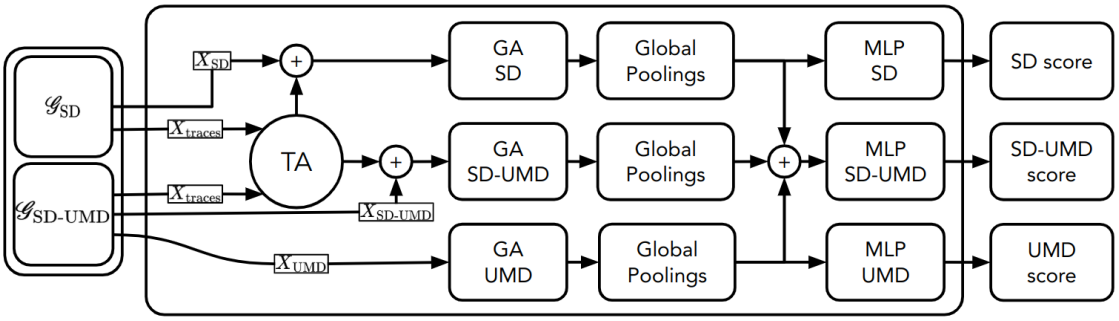

Figure 2: Simplified diagram of the architecture of the neural network used for photon-hadron classification. For each event, two graphs, $\mathcal{G}_{SD}$ and $\mathcal{G}_{SD-UMD}$, are provided as inputs. Traces from SD stations are first processed by the Trace Analyzer (TA). The features extracted by the TA are dynamically concatenated with the input feature matrices (X_{SD} and X_{SD-UMD}) before being processed by the Graph Analyzers (GA). A subgraph with isolated UMD information is also forwarded to a separate GA path. Global pooling operations are applied to the graph embeddings generated by the GAs, creating dense graph-level representations. These dense representations are then forwarded to dense layers, providing photon-hadron classification scores for each path.

(Multi-Layer Perceptrons, denoted as MLPs). The TA is the first stage of the forward pass and processes temporal information from the traces using three layers of one-dimensional convolutions, effectively extracting relevant features from time-series data. The output features from the TA are dynamically concatenated with the input feature matrices of the SD and SD-UMD paths before being passed to their respective graph analyzers. The output features from the TA are dynamically concatenated with the input feature matrices of the SD and SD-UMD paths before being passed to their respective graph analyzers. A subgraph with isolated UMD information is at the same time streamed to another path. The GA, implemented with three layers of Graph Attention Networks (GATv2) [11], processes the spatial and relational information embedded in the graph structures. To create dense representations of the graph embeddings, global readout operations, such as mean and max pooling, are applied to aggregate the node-level information into graph-level features. These dense representations are then forwarded to the dense layers, which consist of three fully connected layers for each information flow, transforming the extracted features into photon-like scores.

The network selects and processes information from the feature matrices through three distinct flows corresponding to the SD, the combined SD-UMD, and the UMD alone. Each flow processes its respective input independently, with the TA serving as a shared subcomponent for the SD and SD-UMD paths. This separation of information enables better isolation and understanding of each detector's contribution to the information processing. At the final stage, scores from the three flows are computed independently, providing introspective capabilities. Despite its intricate design, the network remains relatively small compared to state-of-the-art neural networks, summing $\sim 55,000$ parameters. This compact size ensures efficient training and evaluation.

The network training stage is driven by a global loss function, which is defined as the sum of the cross-binary entropy of the three independent scores corresponding to the SD, SD-UMD, and UMD paths. During training, all individual losses and the overall ROC-AUC metric are tracked to

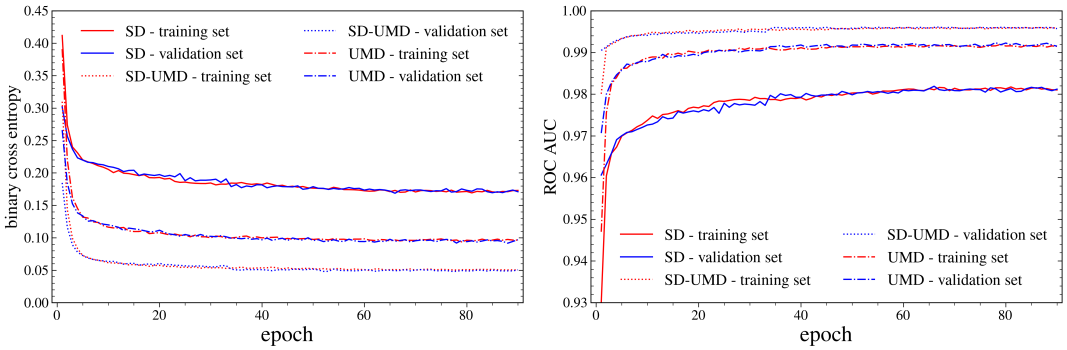

Figure 3: The figure on the left shows the binary cross-entropy loss during training and validation for the three network paths: SD, SD-UMD, and UMD. The close alignment of the training and validation curves indicates no significant overfitting throughout the training process. The figure on the right displays the ROC-AUC metric for training and validation, illustrating the model's capability to distinguish photon-like events.

monitor the network performance. Although the paths are independent during the forward pass, the backward propagation introduces interactions between the information flows.

As shown in Fig. 3, the learning curves indicate that there are no significant signs of overfitting, since the validation curves closely follow the training curves for both the binary cross-entropy loss and the area under the Receiver Operating Characteristic curve (ROC-AUC). From the results, it is evident that the SD-UMD path provides the most significant contribution to the overall performance, likely due to the combined information from both detectors. Furthermore, the UMD score appears to be more efficient at separating classes, as it achieves higher ROC-AUC values and lower losses compared to the SD score. These observations highlight the importance of combining complementary information from different detector components for optimal event classification. They also emphasize the advantage of having a direct muon measurement for the task of photon-hadron discrimination.

4. Photon-hadron discrimination

Only test events are considered to evaluate the high-level network's performance. The general separation between the photon and hadron primaries is notable, as displayed in the normalized score distributions in Fig. 4. The events have been weighted as $E_{\mathrm{MC}}^{-\alpha}$, where $\alpha = 2$ for photons and $\alpha = 3$ for hadrons, to mimic the cosmic-ray background and the expected photon flux. The separation improves as the mass of the primary increases, as seen by the steeper decline in the tail of the iron score distribution compared to the proton distribution. No background events are observed beyond the median of the photon score distribution, highlighting the network's capability to distinguish photon events. Taking into account the number of protons in the test set, the classifier should suffer at most 10^{-4} background contamination at 0.5 signal efficiency. However, a proper estimation of lower background contamination levels requires a detailed investigation of the tails of the score distributions, which will be addressed in future studies. Estimations of the irreducible error set by the physics of the problem will also help to benchmark contamination levels.

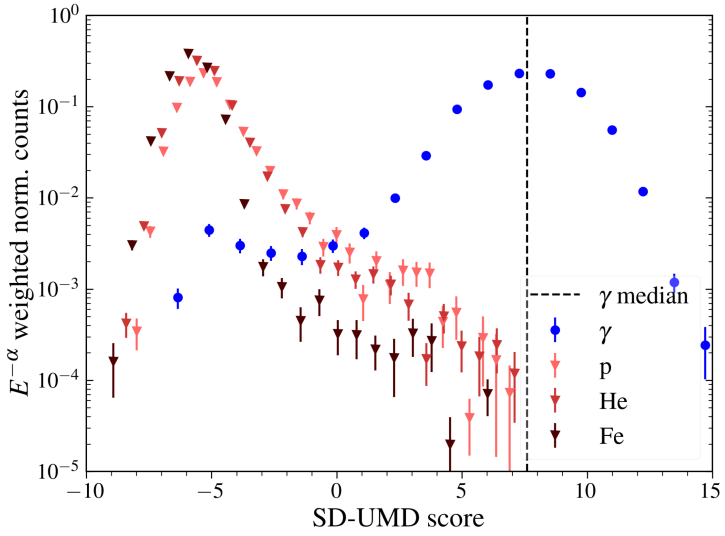

Figure 4: Normalized score distributions for test events, weighted proportionally to $E^{-\alpha}$, where $\alpha = 2$ for photons and $\alpha = 3$ for hadrons. The dashed line indicates the median of the photon score distribution. The photon distribution is represented in blue, while the hadronic distributions are shown in a red palette, with darker shades corresponding to higher primary masses.

The scores of the neural network are largely independent of the reconstructed photon energy and zenith angle, as illustrated in Fig. 5. Due to delayed shower development and the lack of a significant muon component, the energy reconstruction method optimized for hadronic showers tends to underestimate the energy in the case of photon primaries. For this reason, the photon energy is estimated with a different method to the one employed with hadronic showers [5].

The separation between photon and hadron primaries remains robust across reconstructed observables, with the SD-UMD scores showing the best discrimination, followed by the UMD and SD scores. However, a slight broadening of the photon point cloud is observed at higher zenith angles and lower reconstructed energies. This broadening suggests a minor degradation in the network's performance under these conditions, but does not significantly affect its overall discrimination capability. As noted before, the separation improves with the primary mass, as seen by the thinner hadronic point clouds for heavier primaries.

5. Summary

This work presents a novel method for discriminating photons from hadrons at ultra-high-energies using the Pierre Auger Observatory's Surface Detector and Underground Muon Detector. The method leverages Graph Neural Networks to process complex spatio-temporal data from detector simulations. The events were modeled as graphs to capture the irregular geometry and diverse configurations of the detectors. Separate graphs for SD-only and combined SD-UMD data allow for better handling of missing values. The network architecture consists of separate processing paths for SD, SD-UMD, and UMD information and allows a certain level of introspection.

The trained model effectively separates photon and hadron events, with higher performance for the combined SD-UMD. The results are robust across reconstructed energy and zenith angles, with

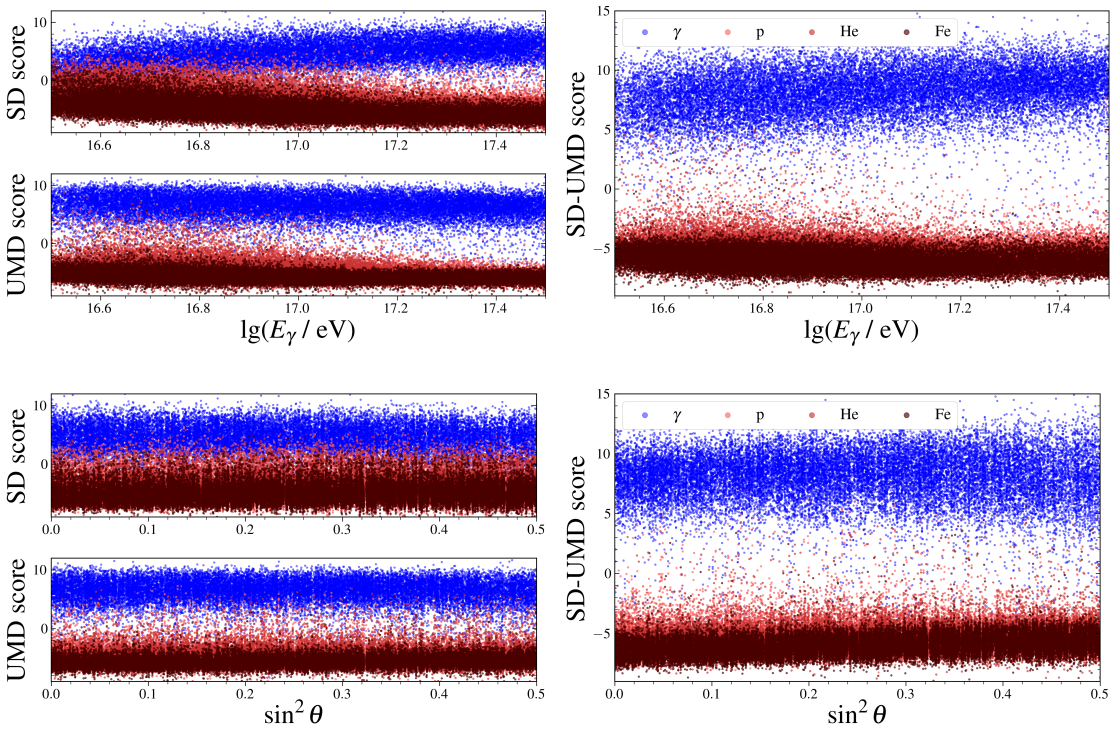

Figure 5: Network scores as a function of $\log(E_\gamma/\mathrm{eV})$ (top panel) and $\sin^2\theta$ (bottom panel). Photon events (blue) exhibit higher scores than hadronic events (red palette), with the SD-UMD scores providing the clearest separation. Photon events are represented in blue, while hadronic events are shown in a red palette, with darker shades corresponding to higher primary masses.

minimal degradation at higher zenith angles or lower energies. Overall, the method demonstrates the potential of GNNs in handling joint SD and UMD data for photon searches.

References

[1] A. Bobrikova et al., PoS **395** (2021) 449.
[2] C. Bérat et al., Astrophys. J. **929** (2022) 55.
[3] B. Sarkar et al., ICRC (Beijing) **2** (2011) 198.
[4] P. Abreu et al. [Pierre Auger Coll.], Phys. Rev. D **107** (2023) 042002.
[5] N. González et al. [Pierre Auger Collaboration], PoS **444** (2023) 238.
[6] A. Aab et al. [Pierre Auger Collaboration], arXiv:1604.03637.
[7] S. Argiro et al., Nucl. Instrum. Meth. A **580** (2007) 1485-1496.
[8] D. Heck et al., Forschungszentrum Karlsruhe Report FZKA **6019** (1998) .
[9] T. Pierog et al., Phys. Rev. C **92** (2015) 034906.
[10] L. Wu et al., Graph Neural Networks: Foundations, Frontiers, and Applications, Springer (2022).
[11] S. Brody et al., arXiv:2105.14491.