Enhancing Gamma–Hadron Separation in Imaging Atmospheric Cherenkov Telescopes Using Attention-Guided Deep Learning and Adaptive Balanced Greylag Goose Optimization

Ebrahim A. Mattar[1,*], S. K. Towfek[2,3]

[1]College of Engineering University of Bahrain, Bahrain
[2]Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA
[3]Jadara Research Center, Jadara University, Irbid 21110, Jordan

Emails: ebmattar@uob.edu.bh; sktowfek@jcsis.org

Abstract

Gamma–hadron discrimination remains a fundamental challenge in very-high-energy gamma-ray astronomy due to the strong overlap between gamma-ray–initiated and hadron-induced air showers recorded by imaging atmospheric Cherenkov telescopes, particularly at low energies where background contamination is severe. Traditional cut-based and non-optimized machine learning approaches often struggle to fully exploit the nonlinear and correlated nature of Cherenkov image parameters, leading to suboptimal background suppression and reduced telescope sensitivity. To address these limitations, this paper proposes a unified deep learning and metaheuristic optimization framework that combines an enhanced attention-based long short-term memory network (EALSTM) with advanced optimization strategies. In particular, a novel Adaptive Balanced Greylag Goose Optimization algorithm (ABGGO) is employed to jointly perform feature selection and hyperparameter optimization, enabling effective exploration–exploitation balancing while preserving physically meaningful feature representations. The proposed ABGGO+EALSTM framework is systematically evaluated against baseline deep learning models, including artificial neural networks (ANN), convolutional neural networks (CNN), and standard long short-term memory networks (LSTM), under identical experimental conditions. Experimental results on a Monte Carlo–generated Cherenkov telescope dataset demonstrate clear and consistent performance gains at every stage of the analysis. In the baseline evaluation stage, EALSTM achieves an accuracy of 0.9294 and an F-score of 0.9266, outperforming ANN, CNN, and LSTM. Following metaheuristic optimization, the proposed ABGGO+EALSTM model attains a peak accuracy of 0.9718, sensitivity of 0.9694, specificity of 0.9740, and F-score of 0.9705, representing absolute improvements exceeding 4% over the baseline EALSTM configuration and outperforming GA+EALSTM, GWO+EALSTM, and PSO+EALSTM variants. These results demonstrate that integrating attention-based deep learning with adaptive metaheuristic optimization significantly enhances gamma–hadron discrimination, leading to improved background suppression and signal retention. The proposed framework offers a scalable and robust solution for current and next-generation Cherenkov observatories, with strong potential for real-time event filtering, multi-telescope analysis, and future deployment on real observational data.

Keywords: Gamma–hadron discrimination; Imaging atmospheric Cherenkov telescopes; Attention-based deep learning; Metaheuristic optimization; Greylag Goose Optimization

DOI: https://doi.org/10.54216/JAIM.110105
Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

1 Introduction

Very-high-energy (VHE) gamma-ray astronomy has become a central pillar of contemporary high-energy astrophysics and astroparticle physics because it provides a direct observational channel to the most extreme particle-acceleration environments in the Universe [1], [2], [3]. In contrast to charged cosmic rays, gamma rays propagate essentially undeflected by interstellar and intergalactic magnetic fields, thereby preserving directional information about their production sites [4], [5], [6]. As a consequence, VHE observations provide a uniquely informative probe of compact objects and energetic transients, including pulsars and their nebulae, active galactic nuclei, and supernova remnants, among other source classes. These targets are closely tied to fundamental problems in non-thermal astrophysics, such as the origin of cosmic rays, the microphysics of relativistic shocks, and the role of magnetic reconnection in high-energy plasmas. Comprehensive overviews of the scientific context and instrumental landscape of ground-based gamma-ray astronomy can be found in established reviews of the atmospheric Cherenkov technique and modern imaging arrays [7].

From an instrumental standpoint, ground-based atmospheric Cherenkov gamma telescopes currently provide the most sensitive measurements in the VHE regime because they leverage the Earth's atmosphere as a calorimetric detection medium. When a primary gamma ray penetrates the upper atmosphere, it initiates an electromagnetic cascade whose relativistic charged secondaries emit Cherenkov radiation once their velocities exceed the local phase velocity of light. This nanosecond-scale emission, concentrated in the near-UV to optical band, reaches the ground as a faint light pool and can be recorded by fast photodetector cameras mounted on large reflectors. The imaging atmospheric Cherenkov technique (IACT) reconstructs the primary event properties by analyzing the spatial and intensity distribution of Cherenkov light in the camera plane, enabling inference of the arrival direction and energy, together with a probabilistic assessment of the primary particle type [8], [9], [10]. The methodological foundations and subsequent evolution of the IACT approach are discussed in classical and modern accounts.

A persistent methodological bottleneck in IACT data analysis is the gamma–hadron discrimination problem. Although gamma-ray–initiated electromagnetic showers produce comparatively compact and regular image morphologies, the recorded event stream is dominated by hadronic cosmic rays (protons and nuclei), which generate extensive air showers with Cherenkov signatures that can partially imitate gamma-like images. Because the astrophysical gamma-ray flux is typically orders of magnitude lower than the charged cosmic-ray background, even modest misclassification rates can substantially degrade the sensitivity of a telescope system [11], [12]. False acceptance of hadron-induced showers increases the residual background level and thus reduces source detection significance, while overly aggressive background rejection can discard true gamma events, diminishing effective area and biasing subsequent physical inference. Therefore, robust gamma–hadron separation is not a peripheral step but rather a core determinant of the end-to-end scientific performance of Cherenkov observations.

Historically, discrimination has relied on physically motivated image parametrizations that reduce the camera image to a small set of descriptors. A widely used approach is to compute characteristic parameters of an ellipse fitted to the light distribution, commonly known as Hillas parameters, which encode geometric properties such as elongation, orientation, and centroid displacement. While such engineered features are interpretable and computationally efficient, their discriminative power can be limited by nonlinear correlations among parameters and by energy-dependent morphological degeneracies between gamma and hadron classes. This motivates the use of machine learning (ML) models, which can learn complex decision boundaries directly from data and thereby exploit multivariate feature interactions that are difficult to capture using threshold-based selections [13], [14]. The transition from cut-based analysis to statistical learning and, more recently, deep learning has been accelerated by advances in computational resources and by the increasing scale of Cherenkov datasets, as documented across the IACT literature.

In this context, deep learning architectures have become particularly attractive because they can represent highly nonlinear mappings with strong generalization potential when properly regularized and optimized [15]. ANN models provide flexible nonlinear classifiers for tabular parameter spaces, CNN models can extract structured interactions when data are organized to preserve local dependencies, and recurrent models such as LSTM can capture ordered relationships among inputs when a meaningful

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

feature ordering or temporal structure is defined. Furthermore, EALSTM-type mechanisms can improve representational efficiency by adaptively weighting informative feature components, which is conceptually aligned with the physical reality that not all image parameters contribute equally across energy ranges and observation conditions. These considerations establish the motivation for investigating ANN, CNN, LSTM, and EALSTM models for Cherenkov gamma–hadron classification within a unified and reproducible framework [16].

A first challenge arises from the nature of the feature space derived from Cherenkov images. Hillas-type descriptors and higher-order moment parameters are constructed from the same underlying photon distribution, and are therefore frequently correlated. Such correlations can inflate model variance, obscure the marginal utility of individual predictors, and reduce robustness to dataset shifts when models overemphasize redundant signals. This is especially relevant because Cherenkov image morphology is simultaneously influenced by primary energy, impact parameter, and shower development fluctuations, meaning that multiple parameters co-vary even when the underlying class label remains unchanged. Consequently, classification performance is not solely a matter of selecting powerful model families; it also depends on disciplined feature handling and appropriate regularization to mitigate redundancy-driven instability.

A second difficulty is the intrinsic overlap between the gamma and hadron classes in the parameter space, particularly at low energies. At the lower end of the IACT threshold, shower images become fainter, more truncated, and more susceptible to noise and reconstruction uncertainties. Under these conditions, hadronic showers may appear more electromagnetic-like, while gamma showers may lose distinctive morphology due to limited photon statistics, which compresses the separability between the classes and necessitates highly nonlinear decision boundaries. Linear or shallow models can therefore struggle to deliver stable discrimination across the full operating range, motivating deeper architectures and robust optimization strategies.

A third challenge concerns hyperparameter sensitivity in deep learning. Models such as ANN, CNN, LSTM, and EALSTM typically introduce a nontrivial set of design and training hyperparameters, including learning rate schedules, layer widths and depths, regularization coefficients, and optimization settings. Inadequate hyperparameter selection can induce underfitting (insufficient capacity to capture class structure) or overfitting (memorization of training idiosyncrasies rather than learning physically meaningful patterns). This issue can be amplified in feature-based IACT analyses because the input dimension may be modest while model capacity can be large, making the bias–variance trade-off particularly delicate. Accordingly, systematic hyperparameter optimization is not optional but central to obtaining reproducible generalization in Cherenkov classification pipelines.

Finally, robust generalization to unseen events is essential because Cherenkov observations are influenced by atmospheric variability, detector response changes, and shower-to-shower fluctuations. Even in simulation-driven studies, models must be designed and validated to maintain stable behavior across realistic distributions of shower geometries and noise conditions. This requirement motivates careful dataset splitting, disciplined validation procedures, and optimization strategies that explicitly prioritize out-of-sample performance rather than training-set fit.

The overarching objective of this study is to develop a rigorous machine learning framework for gamma–hadron discrimination in atmospheric Cherenkov telescope data, with emphasis on methodological clarity and reproducibility. The work evaluates multiple deep learning models—ANN, CNN, LSTM, and EALSTM—under a consistent preprocessing and validation protocol, thereby enabling a controlled comparison of how different architectural biases represent the multivariate structure of Cherenkov-derived features. This comparative design is intended to clarify the extent to which model expressivity, inductive bias, and training stability jointly determine discrimination capability in the IACT setting.

A second objective is to integrate a metaheuristic-based optimization framework to address two tightly coupled challenges: feature selection and hyperparameter tuning. Feature selection is motivated by the correlated nature of Hillas-type parameters and the need to reduce redundancy, while hyperparameter tuning is motivated by the sensitivity of deep models to training configurations. Metaheuristic search provides a flexible strategy for navigating large, non-convex optimization spaces that arise when selecting feature subsets and tuning model hyperparameters simultaneously. The study therefore

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

adopts metaheuristic optimization as a principled mechanism for enhancing model efficiency and robustness while maintaining physical interpretability through disciplined feature usage.

This paper contributes an integrated deep-learning and metaheuristic optimization framework for Cherenkov gamma–hadron discrimination, with a particular focus on EALSTM-based modeling. The methodological contribution lies in the systematic coupling of metaheuristic optimization with deep learning so that feature relevance and hyperparameter settings are optimized in a coordinated manner, rather than in isolation. This integrated view is scientifically meaningful because redundant or weakly informative features can interact with hyperparameter choices to amplify overfitting or suppress useful nonlinear structure. By explicitly addressing both components within one framework, the study advances a reproducible strategy for improving classifier stability and interpretability in IACT analyses.

In addition, the paper provides a structured experimental design that enables transparent comparison across ANN, CNN, LSTM, and EALSTM models within the same dataset context, and evaluates multiple metaheuristic strategies (ABGGO, GA, GWO, and PSO) as optimization engines. The broader contribution is to illustrate, within a Cherenkov telescope setting, how intelligent optimization can be used to strengthen deep learning pipelines for astrophysical signal detection, complementing existing approaches in the IACT literature that emphasize likelihood reconstruction or classical multivariate techniques.

The remainder of this paper is organized to reflect a conventional and reproducible methodological narrative. Section 3 describes the dataset origin and structure, including the simulation context and the extracted image parameters, and details the preprocessing and experimental protocol used to ensure fair model comparison. The subsequent methodology section presents the learning models (ANN, CNN, LSTM, and EALSTM) together with the metaheuristic optimization strategy employed for feature selection and hyperparameter tuning, grounding each choice in the specific characteristics of Cherenkov-derived inputs. The results section then reports the experimental outcomes and comparative analyses using standardized evaluation metrics. The paper concludes by summarizing key methodological insights and outlining future directions for extending optimized ML-based discrimination frameworks to broader IACT analysis pipelines.

2 Literature Review

Gamma-ray astronomy increasingly depends on analysis pipelines capable of extracting rare and physically meaningful signals from data dominated by instrumental effects and astrophysical or particle backgrounds. Across ground-based Imaging Atmospheric Cherenkov Telescopes (IACTs), hybrid wide-field facilities, and space-based instruments, recent studies demonstrate a clear methodological shift toward machine learning (ML) and deep learning (DL) techniques for event classification, energy and direction reconstruction, time-domain inference, and computational acceleration. The reviewed literature shows that improvements in performance do not arise solely from model complexity; rather, they result from the interaction between data realism, representation strategies, and operational constraints such as latency, scalability, and robustness.

A persistent challenge in IACT analysis is gamma–hadron discrimination, where electromagnetic showers initiated by gamma rays must be separated from the dominant background of hadron-induced cosmic-ray showers. One study addresses this problem using the MAGIC gamma telescope dataset from the UCI Machine Learning Repository and explores recurrent neural architectures, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM models [17]. A key methodological contribution of this work is the application of a square-root transformation to mitigate skewness and kurtosis in the feature distributions, improving training stability and classification accuracy. The reported results suggest that careful preprocessing of Cherenkov-derived tabular features is as critical as model selection, highlighting a data-centric perspective in gamma–hadron classification [17].

While tabular features such as Hillas parameters provide compact summaries of shower images, they inevitably discard fine-grained morphological information. Several studies therefore emphasize learning

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

directly from image data. A notable contribution trains a convolutional neural network based on the Inception ResNet V2 architecture using observational data from a giant flare of the Markarian 421 blazar observed by the MAGIC telescope in 2013 [18]. Unlike many prior approaches that rely exclusively on Monte Carlo simulations, this work directly confronts the simulation–observation mismatch that can introduce systematic biases in deep learning models. The competitive performance of the CNN relative to the standard MAGIC Random Forest analysis demonstrates the feasibility of training deep models on real observational data, while also underscoring the importance of addressing selection effects and generalization across different source states [18].

At the level of telescope arrays, next-generation IACT systems record multiple images per event across different telescopes, producing structured and variable-length inputs. Using simulated Cherenkov Telescope Array (CTA) data and the CTLearn framework, one study implements a hybrid CNN–RNN architecture to perform gamma–hadron separation and investigates whether sorting telescope images by total amplitude improves background rejection [19]. The absence of observed performance gains from such sorting suggests that amplitude-based ordering does not provide additional discriminative information beyond what the network can already infer. This result has practical implications for pipeline design, indicating that additional preprocessing heuristics may be unnecessary when models are capable of learning permutation-robust representations [19].

Hybrid detection concepts further expand the scope of deep learning applications in gamma-ray astronomy. The High Altitude Detection of Astronomical Radiation (HADAR) experiment combines the wide field of view of extensive air-shower arrays with the sensitivity of Cherenkov detectors. A study on HADAR demonstrates that incorporating physical background knowledge into dataset construction and model design enables accurate background identification, energy reconstruction, and directional estimation [20]. The reported improvements in sensitivity, including competitive performance relative to established Cherenkov telescopes in certain energy ranges, highlight the effectiveness of physics-guided deep learning approaches in handling strongly correlated and scattered detector data [20].

In contrast to event-level classification in ground-based instruments, space-based gamma-ray observatories often focus on population-level source classification. Using the 4FGL-DR4 catalog from the Fermi Large Area Telescope, one study applies Random Forest and Extreme Gradient Boosting methods to classify unassociated gamma-ray sources and identify new pulsar candidates [21]. By explicitly addressing class imbalance through data balancing strategies, the study demonstrates how supervised ML can support population studies and prioritize sources for multiwavelength follow-up, thereby linking catalog analysis to broader astrophysical interpretation [21].

Time-domain analysis represents another area where ML has introduced significant methodological advances. One study proposes a deep-learning-based approach to estimate adaptive binning windows for Fermi-LAT gamma-ray light curves directly from raw photon data, overcoming the computational cost and inaccuracies of traditional adaptive binning methods in highly populated regions [22]. Complementarily, another framework focuses on anomaly detection in multivariate time series by learning expected background behavior using feed-forward and recurrent neural networks and applying a triggering algorithm to identify statistically significant deviations in data from the Fermi Anti-Coincidence Detector [23]. Together, these works demonstrate how sequence learning enables efficient and responsive detection of transient phenomena in gamma-ray time series [22], [23].

Simulation remains a major computational bottleneck in IACT science, motivating the use of generative models as surrogates for expensive Monte Carlo pipelines. One study proposes using Generative Adversarial Networks (GANs) to generate source-dependent gamma-ray image parameters for the MACE telescope, addressing biases that arise when analysis cuts calibrated for a specific spectral index are applied to sources with different spectra [24]. A related work focuses on the TACTIC telescope and demonstrates that GAN-generated synthetic image parameters closely resemble Monte Carlo simulations, as validated through statistical comparisons and Random Forest-based classification tests. These studies position generative modeling as a viable strategy for reducing simulation costs while preserving the fidelity required for downstream scientific analyses [24].

Beyond classification, deep learning has also been applied to regression tasks such as gamma-ray energy reconstruction. Using TAIGA-IACT Monte Carlo data, one study compares traditional

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

Hillas-parameter-based energy reconstruction with deep learning approaches based on convolutional neural networks and autoencoder-derived latent features [25]. The results indicate that learned representations achieve comparable accuracy at low energies and significantly improved performance at high energies, suggesting that autoencoders and CNNs can capture subtle image features that are not fully described by classical moment-based parameters [25].

Emerging research directions include the exploration of quantum machine learning for astrophysical signal detection. A study evaluating Quantum Convolutional Neural Networks (QCNNs) for identifying gamma-ray-burst-like signals in simulated CTA Observatory light curves reports accuracies comparable to classical CNNs while using fewer parameters [26]. Although currently exploratory and limited to quantum simulators, this work establishes methodological benchmarks and clarifies the trade-offs between model complexity, encoding strategies, and performance in time-series classification tasks [26].

Operational constraints increasingly shape the deployment of ML in gamma-ray observatories. The Advanced Particle-astrophysics Telescope (APT) concept integrates neural networks into its on-board processing pipeline to perform real-time gamma-ray burst localization under strict size, weight, and power limitations [27]. Similarly, convolutional neural networks have been applied to optimize electron recoil path reconstruction in the GammaTPC concept, improving pointing capability in a liquid-argon time projection chamber detector [28]. A broader survey of autonomous telescopes further contextualizes these efforts, emphasizing how AI-driven systems enable continuous operation, real-time data analysis, and rapid transient detection, while also highlighting challenges related to reliability, data management, and ethical considerations [29].

Finally, comparative studies from adjacent areas of astronomy provide transferable insights. Work on automated galaxy morphology classification comparing Support Vector Machines, Random Forests, and Naive Bayes classifiers demonstrates the strong dependence of performance on feature representation, with engineered morphological features often outperforming PCA-compressed pixel data [30]. This observation reinforces a recurring theme across gamma-ray applications: the alignment between data representation and underlying physical structure is a dominant factor in determining the success and scalability of machine learning models [30].

In summary, the reviewed literature reveals a coherent evolution in gamma-ray data analysis toward deep, representation-driven, and operationally integrated machine learning approaches. Progress is most robust when models are developed in close connection with detector physics, data characteristics, and real-world constraints, ensuring that algorithmic advances translate into improved sensitivity, reliability, and scientific discovery potential [20], [27], [29].

3 Materials and Methods

3.1 Dataset description

The dataset employed in this study consists of simulated events designed to reproduce the physical and observational characteristics of very-high-energy gamma-ray detection with a ground-based atmospheric Cherenkov gamma telescope. The data were generated using a Monte Carlo (MC) approach, which is the standard methodology in Cherenkov telescope analysis for modeling extensive air showers and their associated Cherenkov light emission under controlled and well-defined physical conditions. Monte Carlo simulations play a crucial role in this domain because they allow the generation of large, labeled datasets in which the nature of the primary particle initiating each air shower is known with certainty, thereby enabling supervised learning and objective performance evaluation of classification algorithms.

The simulation framework used to generate the dataset is based on the CORSIKA (COsmic Ray SImulations for KAscade) code, a well-established and extensively validated Monte Carlo program for modeling the development of extensive air showers in the Earth's atmosphere. CORSIKA simulates the interactions of primary gamma rays and hadronic cosmic rays with atmospheric nuclei, tracking the production and propagation of secondary particles and the associated Cherenkov radiation.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

The version and configuration employed in this dataset were selected to enable the simulation of events down to energies below 50 GeV, which corresponds to a particularly challenging regime for atmospheric Cherenkov telescopes due to reduced photon statistics and increased background contamination. Detailed descriptions of the physical models, interaction processes, and numerical techniques implemented in CORSIKA can be found in the foundational documentation.

Each simulated event represents the detection of Cherenkov photons produced by an air shower and recorded by a camera composed of photomultiplier tubes arranged in a planar geometry. As a primary particle traverses the atmosphere and initiates a cascade, relativistic charged particles emit Cherenkov radiation that reaches the detector as a two-dimensional light distribution. Depending on the energy of the primary gamma ray or hadron, the number of detected Cherenkov photons can range from a few hundred to several tens of thousands. After standard preprocessing steps, which include image cleaning and noise suppression, the resulting light distribution forms a compact cluster in the camera plane, commonly referred to as the shower image. The morphological properties of this image encode essential information about the nature of the primary particle.

To convert the raw camera images into a form suitable for machine learning analysis, a set of descriptive features is extracted from each event. These features are primarily based on the classical Hillas parametrization, which characterizes the shower image by fitting an ellipse to the distribution of detected photons in the camera plane. The orientation, elongation, and centroid of this ellipse are strongly correlated with the shower geometry and primary particle type. In addition to second-order geometric moments, higher-order moments and intensity-based descriptors are also included to capture asymmetries and longitudinal development characteristics of the shower. This feature-based representation provides a compact yet physically interpretable summary of each event, making it well suited for supervised classification tasks.

The complete dataset contains a total of 19,020 instances, each corresponding to an individual simulated air-shower event. For each instance, ten continuous-valued attributes are provided, along with a binary class label indicating whether the event was initiated by a gamma ray (signal) or by a hadronic cosmic ray (background). The continuous attributes include geometric parameters such as the major and minor axes of the fitted ellipse, intensity-related measures describing the total and concentrated light content, asymmetry indicators along the principal axes, higher-order moment descriptors, and angular and spatial parameters relating the shower image to the camera center. Together, these attributes form a structured numerical feature space that reflects both the physical development of the air shower and the response characteristics of the Cherenkov detector.

The class label is defined as a categorical variable with two possible values, corresponding to gamma-induced and hadron-induced showers. This explicit labeling, made possible by the Monte Carlo simulation process, enables the formulation of the gamma–hadron separation problem as a binary classification task. From a machine learning perspective, this setup allows for direct optimization of discriminative models using supervised learning techniques, while from a physical standpoint it provides a controlled environment for studying the limits and capabilities of advanced classification strategies.

The dataset does not incorporate external auxiliary data sources, such as atmospheric monitoring information or telescope configuration parameters, in order to focus the analysis exclusively on image-derived features. This design choice ensures that all classification performance can be attributed to the intrinsic discriminative power of the extracted shower parameters and the modeling capacity of the applied machine learning and optimization techniques. Prior to model training and evaluation, the dataset is partitioned into training, validation, and testing subsets according to a predefined splitting strategy, which is described in detail in the subsequent methodology sections. This partitioning is performed in a manner that preserves the statistical properties of the original data and supports unbiased assessment of model generalization to unseen events.

A detailed exploratory analysis of the MAGIC gamma-ray telescope dataset is essential for understanding the statistical behavior and discriminative power of its extracted features prior to model development. The MAGIC dataset consists of parameters derived from the Hillas analysis of Cherenkov images, which characterize the shape, orientation, and intensity of particle-induced air showers recorded by Imaging Atmospheric Cherenkov Telescopes (IACTs). These parameters are

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

widely used to distinguish gamma-ray–induced showers from hadronic background events, a task that remains central to very-high-energy gamma-ray astronomy. Figure 1 illustrates the class-wise distributions of the ten most commonly used MAGIC features, including geometrical descriptors (e.g., *fLength*, *fWidth*), intensity-related attributes (e.g., *fSize*), concentration measures (e.g., *fConc*, *fConc1*), and orientation parameters (e.g., *fAlpha*, *fM3Long*, *fM3Trans*). The distributions are shown separately for gamma events (g) and hadron events (h), enabling a direct visual comparison of their statistical properties. As observed in Figure 1, several features exhibit noticeable differences in shape, spread, and central tendency between the two classes, indicating their relevance for classification. At the same time, significant overlap persists for certain parameters, highlighting the intrinsic complexity of the gamma–hadron separation problem and motivating the use of advanced machine learning and deep learning techniques.

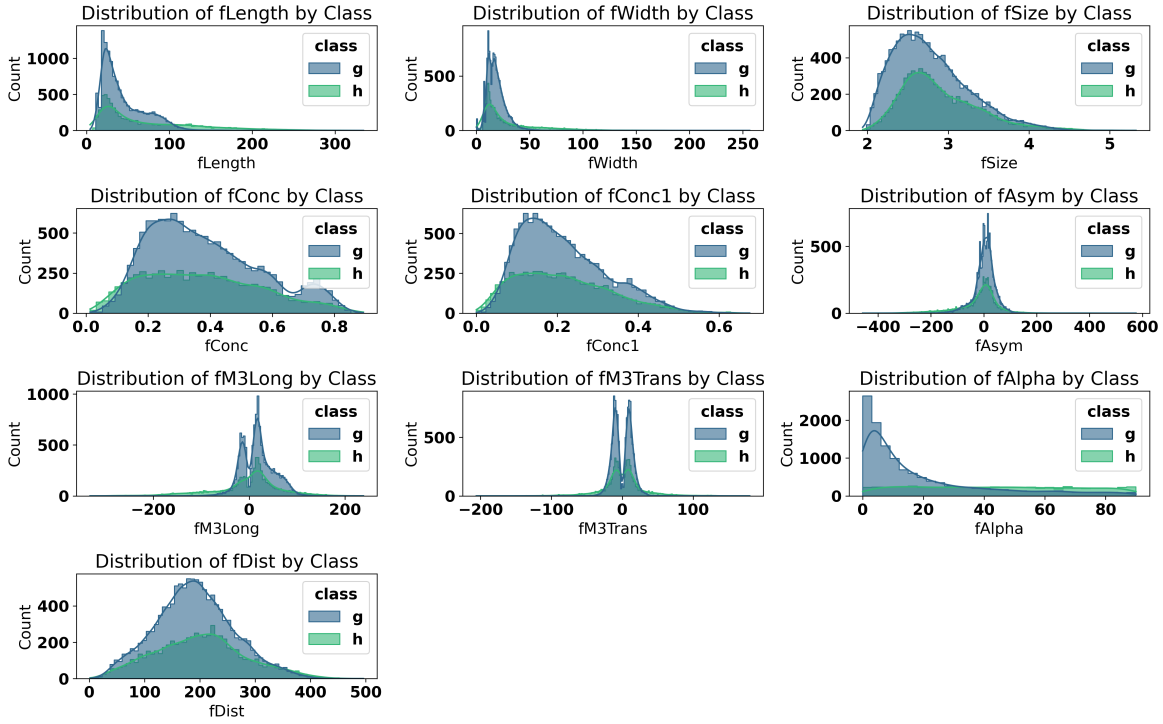

Figure 1: Class-wise distributions of the MAGIC gamma telescope features for gamma (g) and hadron (h) events.

To complement the distributional analysis and provide a robust summary of central tendency, dispersion, and outlier behavior, boxen plots are employed for a detailed comparison of the MAGIC gamma telescope features across event classes. Boxen plots extend traditional box plots by visualizing the distribution of data at multiple quantile levels, making them particularly effective for large datasets with skewed distributions and heavy tails. This representation is well suited to the MAGIC dataset, where several features exhibit pronounced asymmetry and extreme values due to the stochastic nature of atmospheric air showers. Figure 2 presents the boxen plots of the ten extracted features for gamma-ray events (g) and hadronic background events (h). As shown in Figure 2, clear differences are observed in the median values, interquartile ranges, and tail behaviors between the two classes for several parameters, such as *fLength*, *fWidth*, *fAlpha*, and *fDist*. In contrast, features like *fConc* and *fConc1* demonstrate overlapping central distributions while still differing in their upper quantiles. These observations highlight both the discriminative potential and the intrinsic overlap of the feature space, underscoring the need for advanced classification models capable of capturing nonlinear relationships and subtle statistical variations.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

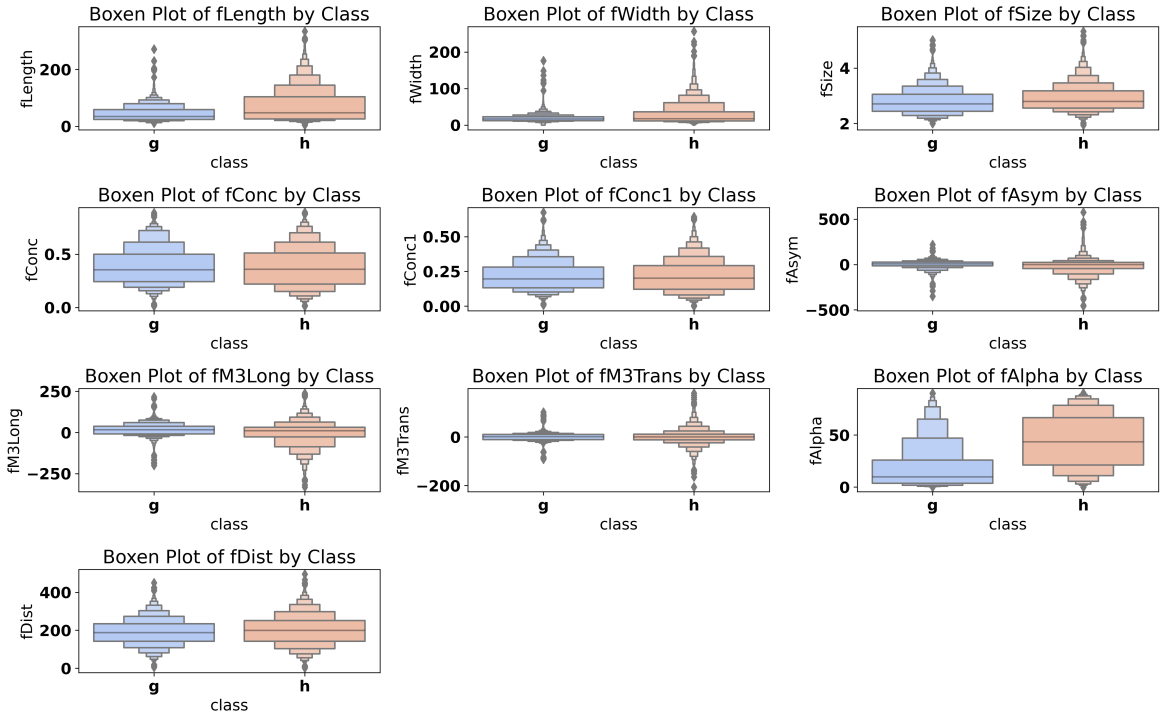

Figure 2: Boxen plots of the MAGIC gamma telescope features for gamma (g) and hadron (h) events.

Figure 3 presents overlaid histograms of *fSize*, *fLength*, and *fWidth* for gamma (g) and hadron (h) classes. As illustrated in Figure 3, all three features exhibit pronounced right-skewed distributions, with hadronic events generally extending toward larger values, particularly in *fLength* and *fWidth*. While substantial overlap between the two classes is observed, differences in peak location, spread, and tail behavior suggest that these features contain meaningful discriminatory information. These distributional characteristics underscore the importance of appropriate scaling and the potential benefit of nonlinear classifiers capable of capturing subtle class-dependent patterns.

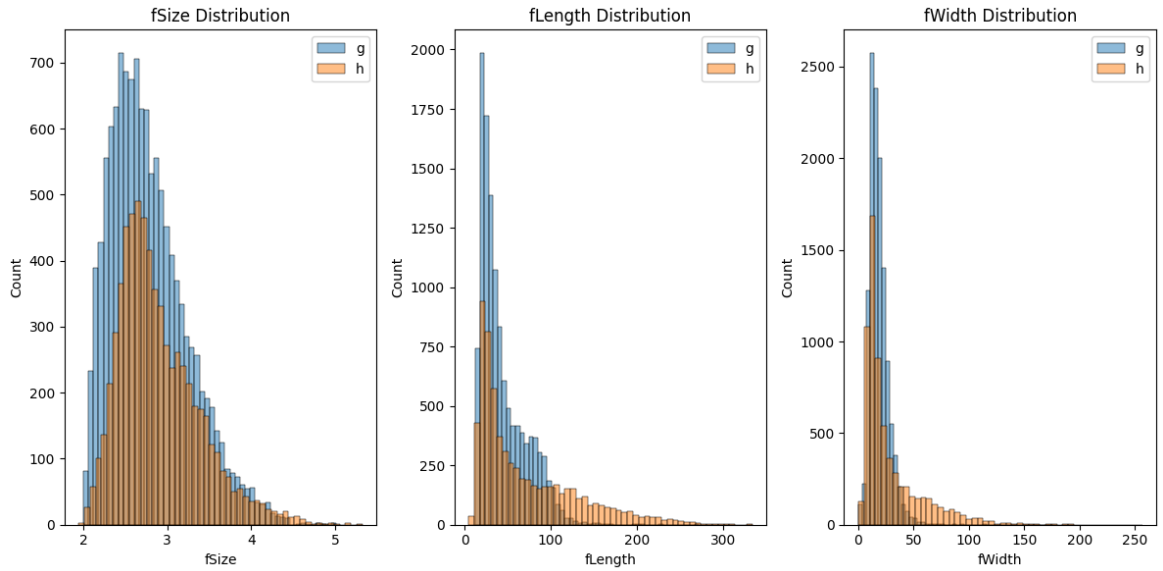

Figure 3: Distributions of *fSize*, *fLength*, and *fWidth* for gamma (g) and hadron (h) events.

DOI: https://doi.org/10.54216/JAIM.110105
Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

3.2 Data preprocessing

Data preprocessing constitutes a critical stage in the development of reliable machine learning models for atmospheric Cherenkov telescope analysis, as the quality and statistical properties of the input features directly influence both learning stability and generalization capability. In the context of gamma–hadron discrimination, preprocessing must be carried out with particular care because the extracted image parameters are physically correlated, span different numerical ranges, and are affected by intrinsic shower fluctuations and detector-related noise. The preprocessing strategy adopted in this study is therefore designed to standardize the feature space, mitigate undesirable numerical effects, and preserve the physically meaningful information encoded in the simulated Cherenkov images. All input variables in the dataset are continuous-valued descriptors derived from the parameterization of shower images after standard image cleaning procedures. As the dataset originates from Monte Carlo simulations, it does not contain missing values in the conventional sense. Nevertheless, preprocessing is still required to address differences in feature scale, distribution shape, and variance, which can otherwise bias the learning dynamics of gradient-based deep learning models. In particular, parameters such as image size and higher-order moments can exhibit substantially different magnitudes compared to angular or geometric descriptors, leading to disproportionate influence during model training if left unnormalized.

To ensure numerical stability and facilitate efficient optimization, feature scaling is applied uniformly across all continuous attributes. Scaling transforms the original feature values into a common numerical range, preventing features with large absolute magnitudes from dominating gradient updates and accelerating convergence during training. This step is especially important for deep learning architectures such as ANN, CNN, LSTM, and EALSTM, whose training dynamics are sensitive to the relative scale of input variables. The scaling procedure is fitted exclusively on the training subset and subsequently applied to the validation and test subsets using the same transformation parameters, thereby avoiding information leakage and preserving the integrity of the evaluation process. In addition to scaling, the statistical structure of the feature space is examined through correlation analysis. The Hillas-based parameters and related descriptors are derived from the same underlying light distribution and are therefore expected to exhibit non-negligible correlations. Strongly correlated features can introduce redundancy, increase model variance, and obscure the contribution of truly informative variables. While deep learning models are, in principle, capable of handling correlated inputs, excessive redundancy can still degrade performance and increase the risk of overfitting, particularly when model capacity is high relative to feature dimensionality. Consequently, correlation analysis serves as a diagnostic tool to inform subsequent feature selection and optimization stages rather than as a hard filter at the preprocessing stage. The class labels associated with each event are categorical, indicating whether the corresponding air shower was initiated by a gamma ray or by a hadronic cosmic ray. For the purposes of supervised learning, these labels are encoded in a numerical form compatible with the loss functions employed by the deep learning models. This encoding is performed in a consistent and unambiguous manner across all experiments to ensure comparability between different model architectures and optimization strategies. Following feature scaling and label encoding, the dataset is partitioned into training, validation, and testing subsets. The splitting strategy is designed to preserve the overall statistical characteristics of the original dataset, including the relative proportions of gamma and hadron events, in each subset. Such stratified partitioning is essential for obtaining unbiased estimates of model performance and for ensuring that hyperparameter tuning decisions made on the validation set generalize to unseen data. The training subset is used exclusively for parameter learning, the validation subset supports model selection and hyperparameter optimization, and the testing subset is reserved for final performance assessment. Importantly, all preprocessing steps are incorporated into a unified and reproducible pipeline. This pipeline ensures that identical transformations are applied consistently across all models and experimental configurations, thereby eliminating confounding factors that could otherwise arise from inconsistent data handling. By standardizing the input representation and enforcing strict separation between training, validation, and testing data, the preprocessing framework establishes a robust foundation for subsequent machine learning modeling and metaheuristic optimization. This disciplined approach is essential for drawing reliable conclusions about the effectiveness of different classifiers and optimization strategies in the gamma–hadron discrimination task.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

3.3 Deep learning models

The classification framework developed in this work relies on deep learning models that are capable of approximating complex nonlinear decision functions defined over structured numerical feature spaces. Let $\mathbf{x} \in \mathbb{R}^d$ denote the input feature vector corresponding to a single Cherenkov event, where d represents the number of extracted image parameters, and let $y \in \{0, 1\}$ denote the associated class label, corresponding to hadron-induced and gamma-induced showers, respectively. The objective of all considered models is to learn a mapping

$$f_{\boldsymbol{\theta}} : \mathbb{R}^d \to [0, 1], \tag{1}$$

parameterized by $\boldsymbol{\theta}$, such that $f_{\boldsymbol{\theta}}(\mathbf{x})$ estimates the posterior probability of an event belonging to the gamma-ray class. Model training is formulated as the minimization of a supervised loss function, typically expressed in terms of binary cross-entropy,

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \right], \tag{2}$$

where N denotes the number of training samples and $\hat{y}_i = f_{\boldsymbol{\theta}}(\mathbf{x}_i)$ is the model output for the i-th event.

Although the dataset does not constitute a classical time series, the ordered and interdependent nature of the extracted Cherenkov image parameters motivates the exploration of architectures commonly employed in time-series and sequential learning tasks. In particular, parameters describing shower geometry, intensity distribution, and asymmetry are not statistically independent but encode correlated aspects of the same physical cascade. Modeling these dependencies explicitly can improve the discriminative representation learned by the classifier. Based on these considerations, four deep learning architectures are selected as baseline models: ANN, CNN, LSTM, and EALSTM.

The artificial neural network (ANN) serves as a foundational baseline model for tabular data. An ANN with L layers implements a sequence of affine transformations followed by nonlinear activation functions,

$$\mathbf{h}^{(l)} = \varphi\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right), \quad l = 1, \ldots, L, \tag{3}$$

where $\mathbf{h}^{(0)} = \mathbf{x}$, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias vector of the l-th layer, respectively, and $\varphi(\cdot)$ denotes a nonlinear activation function. The final layer maps the learned representation to a scalar output through a sigmoid function to produce a class probability. ANNs are universal function approximators and provide a flexible baseline for assessing the added value of more specialized architectures. However, because ANNs treat all input features symmetrically, they do not explicitly encode structured dependencies among features derived from common physical origins.

The convolutional neural network (CNN) is included to investigate whether local feature interaction modeling can enhance discrimination capability. In this setting, the input feature vector is reshaped or ordered in a manner that preserves meaningful adjacency relations among parameters. Convolutional layers apply learnable filters that perform local aggregations,

$$\mathbf{h}_k^{(l)} = \varphi\left(\sum_j \mathbf{w}_{k,j}^{(l)} * \mathbf{h}_j^{(l-1)} + b_k^{(l)}\right), \tag{4}$$

where $*$ denotes the convolution operation and $\mathbf{w}_{k,j}^{(l)}$ represents the filter weights connecting input channel j to output channel k. By stacking multiple convolutional layers, CNNs can learn hierarchical feature representations that capture increasingly abstract interactions among groups of parameters. This inductive bias may be beneficial for Cherenkov image descriptors that naturally cluster into geometric, intensity-based, and higher-order moment features.

The long short-term memory (LSTM) network is selected to explicitly model ordered dependencies among input parameters. LSTM architectures extend recurrent neural networks by introducing gating

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

mechanisms that regulate information flow through an internal memory cell. For a given input sequence $\{\mathbf{x}_t\}_{t=1}^{T}$, the LSTM updates are governed by

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i\right), \tag{5}$$
$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f\right), \tag{6}$$
$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o\right), \tag{7}$$
$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_c\mathbf{x}_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c\right), \tag{8}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \tag{9}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \tag{10}$$

where $\mathbf{i}_t$, $\mathbf{f}_t$, and $\mathbf{o}_t$ denote the input, forget, and output gates, respectively, $\mathbf{c}_t$ is the cell state, $\mathbf{h}_t$ is the hidden state, and $\sigma(\cdot)$ denotes the sigmoid activation function. In the present context, the input sequence corresponds to an ordered representation of Cherenkov image parameters, allowing the LSTM to learn dependencies that reflect the physical relationships among features.

The enhanced attention-based long short-term memory (EALSTM) model extends the standard LSTM by incorporating an attention mechanism that adaptively weights input features or hidden states according to their relevance for classification. Let $\{\mathbf{h}_t\}_{t=1}^{T}$ denote the sequence of hidden states produced by the LSTM. The attention mechanism computes a weighted context vector

$$\mathbf{c} = \sum_{t=1}^{T} \alpha_t \mathbf{h}_t, \tag{11}$$

where the attention weights α_t satisfy

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T}\exp(e_k)}, \quad e_t = \mathbf{v}^\top \tanh(\mathbf{W}_a\mathbf{h}_t + \mathbf{b}_a). \tag{12}$$

This formulation allows the model to emphasize informative components of the feature sequence while attenuating less relevant or redundant information. From a physical standpoint, this adaptive weighting aligns naturally with the Cherenkov discrimination problem, as certain image parameters contribute more strongly to gamma–hadron separation depending on shower morphology and observational conditions.

All four architectures—ANN, CNN, LSTM, and EALSTM—are implemented within a unified experimental framework and trained using identical preprocessing pipelines and data partitions. This ensures that observed differences in behavior can be attributed to architectural characteristics rather than inconsistencies in data handling. Among these baseline models, EALSTM is selected as the primary candidate for subsequent metaheuristic optimization. This selection is motivated by its combined ability to model ordered dependencies and to dynamically reweight feature contributions, making it particularly well suited for handling correlated and heterogeneous Cherenkov image parameters. By focusing optimization efforts on EALSTM, the study aims to enhance a model architecture that is already closely aligned with the statistical and physical structure of the gamma–hadron discrimination task.

3.4 Metaheuristic optimization algorithms

3.5 Greylag Goose Optimization (GGO) Algorithm

This study employs the Greylag Goose Optimization (GGO) algorithm as a population-based metaheuristic designed to achieve an adaptive balance between exploration and exploitation. GGO is inspired by the cooperative and dynamic behavior observed in greylag goose flocks, where individuals alternate between wide-area exploration and focused exploitation to locate optimal feeding regions. The overall workflow of the GGO algorithm is formally described in Algorithm 1.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

At initialization, GGO randomly generates a population of candidate solutions, where each individual Y_i, for $i = 1, 2, \ldots, n$, represents a feasible solution within the search space and n denotes the population size. The quality of each candidate is evaluated using an objective function F_n, and the best-performing individual is denoted by Z, which serves as the current leader of the population. This leader is updated iteratively as better solutions are discovered during the optimization process.

A defining feature of GGO is its dynamic population structure. At each iteration, the population is divided into two groups: an exploration group of size n_1 and an exploitation group of size n_2, such that $n_1 + n_2 = n$. Initially, the population is evenly split between exploration and exploitation to ensure broad coverage of the search space. As iterations progress, the algorithm gradually shifts more individuals toward exploitation in order to refine promising solutions. However, if the objective function value remains unchanged for three consecutive iterations, the algorithm increases the number of exploration agents to introduce additional diversity and reduce the risk of premature convergence to local optima. This adaptive group resizing mechanism enables GGO to respond dynamically to stagnation in the optimization process.

During the exploration phase, agents search for new promising regions in the vicinity of their current positions. To prevent the population from being overly influenced by a single leader, three randomly selected individuals, Y_{Paddle1}, Y_{Paddle2}, and Y_{Paddle3}, are used to guide the update process. This strategy enhances diversity and mitigates directional bias. When the condition $|B| \geq 1$ is satisfied, the position of an exploration agent is updated. The coefficients w_1, w_2, and w_3 are dynamically adjusted within the interval $[0, 2]$ to maintain stochastic variability. The parameter p, decreases exponentially with the iteration index, allowing extensive exploration in early stages and progressively favoring convergence in later stages.

The exploitation phase focuses on intensifying the search around high-quality solutions. In this phase, agents refine their positions by following multiple sentry solutions rather than a single leader. Specifically, non-sentry agents ($Y_{NonSentry}$) update their positions based on three sentry solutions, $Y_{\text{S entry1}}$, $Y_{\text{S entry2}}$, and $Y_{\text{S entry3}}$, which guide the population toward the expected location of the optimal solution. This mechanism reduces sensitivity to noise and supports stable convergence. The exploitation process is governed by conditional update rules based on iteration parity and stochastic parameters, as detailed in Algorithm 1.

From a theoretical perspective, the exploitation stage leverages the Triangle Inequality, a fundamental property of metric spaces, to justify bounded and stable position updates. In addition, the exploration process is conceptually supported by the Law of Large Numbers, which explains how repeated stochastic sampling across a sufficiently large population increases the likelihood of convergence toward the global optimum. Together, these principles provide a mathematical rationale for the convergence behavior of GGO.

Algorithm 1 summarizes the complete GGO procedure, including population initialization, fitness evaluation, dynamic group partitioning, position updates for exploration and exploitation agents, stagnation detection, and adaptive adjustment of group sizes. Through the coordinated interaction of these components, GGO achieves a robust balance between diversification and intensification across iterations.

3.6 Dynamic Optimization

Dynamic optimization addresses problems in which the objective function or the feasible region evolves over time, requiring algorithms to continuously adapt their search behavior. Such problems arise in numerous scientific and engineering domains where environmental conditions, system parameters, or optimization goals are not static. In this context, adaptive population-based algorithms are particularly effective due to their inherent ability to maintain diversity while tracking moving optima.

A general dynamic optimization problem can be formulated as:

$$\min / \max \; F(L, t), \quad \text{subject to } L \in \mathcal{S}(t),$$

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

Algorithm 1 GGO Algorithm

Require: Population (Y_i), size (n), maximum iterations ($t_{\max}$), objective function (F_n), parameters ($b, B, D, \dots$)
Ensure: Best agent position (Z)
Initialize the population (Y_i), size (n), maximum iterations ($t_{\max}$), and parameters ($b, B, D, \dots$)
Calculate the objective function (F_n) for each agent and determine the best position (Z)
Divide agents into exploration (n_1) and exploitation (n_2) groups
for each iteration t from 1 to $t_{\max}$ **do**
    Update agent positions in the exploration group (n_1) using specific equations
    Update agent positions in the exploitation group (n_2) based on conditions (e.g., parity of t and random parameters)
    Adjust parameters and evaluate the objective function (F_n)
    if no improvement in the best objective function (F_n) over two iterations **then**
        Adjust the sizes of n_1 and n_2
    end if
end for
return The best agent position (Z)

where $F(L,t)$ denotes a time-dependent objective function, $L \in \mathbb{R}^d$ represents a candidate solution, and $\mathcal{S}(t)$ denotes a feasible region that may change with time. One of the principal challenges in such settings is maintaining an effective balance between exploration and exploitation. Exploration enables the discovery of new promising regions, while exploitation refines existing high-quality solutions. Excessive emphasis on either aspect can lead to suboptimal performance, motivating adaptive strategies that regulate their interaction.

In population-based methods, this balance is commonly implemented by partitioning the population P_t into exploration and exploitation subgroups:

$$P_t = P_t^{\text{explore}} \cup P_t^{\text{exploit}}, \quad |P_t^{\text{explore}}| + |P_t^{\text{exploit}}| = |P_t|.$$

The relative sizes of these groups are adjusted dynamically using a feedback-driven control parameter $\alpha(t)$:

$$\frac{|P_t^{\text{explore}}|}{|P_t|} = \alpha(t), \quad \frac{|P_t^{\text{exploit}}|}{|P_t|} = 1 - \alpha(t),$$

where

$$\alpha(t) = \begin{cases} \alpha(t-1) + \Delta\alpha & \text{if stagnation occurs,} \\ \alpha(t-1) - \Delta\alpha & \text{if progress is significant.} \end{cases}$$

This adaptive scheme increases exploration when improvement stagnates and strengthens exploitation when consistent progress is observed, thereby supporting robust convergence behavior in non-stationary or highly multimodal landscapes.

Balanced Chain Greylag Goose Optimization Algorithm (BCGGO)

Building upon the dynamic optimization principles underlying GGO, this paper adopts the Balanced Chain Greylag Goose Optimization Algorithm (BCGGO) as an enhanced multi-leader variant. BCGGO extends the original GGO framework by increasing the number of guiding leaders from three to five, which enables a more flexible and robust balance between exploration and exploitation. The presence of multiple leaders allows the population to follow several promising regions simultaneously, thereby improving diversity preservation and reducing sensitivity to premature convergence.

The BCGGO algorithm begins by initializing the population and evaluating the fitness of each candidate solution. The population is then dynamically divided into exploration and exploitation groups, whose sizes are adjusted based on optimization feedback. In early iterations, BCGGO emphasizes exploration to ensure wide coverage of the search space. As the optimization progresses, exploitation is gradually

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

strengthened to refine solutions near the most promising leaders. When stagnation is detected, the algorithm increases exploration pressure to reintroduce diversity.

The complete operational procedure of BCGGO, including multi-leader guidance, conditional update rules, stagnation handling, and adaptive group resizing, is detailed in Algorithm 2. This algorithm preserves the mathematical structure and update equations of the original design while embedding them within a dynamic population management framework. Through this design, BCGGO effectively integrates exploration diversity, exploitation intensity, and adaptive feedback, making it well suited for complex and dynamic optimization problems.

Algorithm 2 BCGGO Algorithm

Require: Population (L_i), size (n), maximum iterations ($t_{\max}$), objective function (F_n). Parameters ($a, A, C, b, l, c, \dots$).
Ensure: Best agent position (P) as the optimal solution.
Initialize population and parameters.
Evaluate initial fitness (F_n) and determine best position (P).
Divide population into exploration (n_1) and exploitation (n_2) groups.
for each iteration t from 1 to $t_{\max}$ **do**
  for each agent in exploration group (n_1) **do**
    if $t \mod 2 = 0$ **then**
      if $r_3 < 0.5$ **then**
        if $|A| < 1$ **then**
          $L(t+1) = L^*(t) - A \cdot |C \cdot L^*(t) - L(t)|$
        else
Select 5 random agents ($L_1, \dots, L_5$)
          $z = 1 - \left(\frac{t}{t_{\max}}\right)^2$
          $L(t+1) = w_1 L_1 + z w_2 (L_2 - L_3) + (1-z) w_3 (L_4 - L_5)$
        end if
      else
        $L(t+1) = w_4 |L^*(t) - L(t)| e^{bl} \cos(2\pi l) + [2w_1(r_4 + r_5)] L^*(t)$
      end if
    else
      $L(t+1) = L(t) + D(1+z) w (L - L_{\text{Flock1}})$
    end if
  end for
  for each agent in exploitation group (n_2) **do**
    if $t \mod 2 = 0$ **then**
      $L_1 = L_{\text{Sentry1}} - A_1 |C_1 L_{\text{Sentry1}} - L|$
      $L_2 = L_{\text{Sentry2}} - A_2 |C_2 L_{\text{Sentry2}} - L|$
      $L_3 = L_{\text{Sentry3}} - A_3 |C_3 L_{\text{Sentry3}} - L|$
      $L(t+1) = \frac{1}{3} \sum_{j=1}^{3} L_j$
    else
      $L(t+1) = L(t) + D(1+z) w (L - L_{\text{Flock1}})$
    end if
  end for
Recalculate F_n, update parameters.
  if no improvement in 2 iterations **then**
Increase (n_1), decrease (n_2)
  end if
end for
return Best agent position (P).

3.6.1 State-of-the-art metaheuristic models

To realize the hyperparameter optimization strategy, several state-of-the-art metaheuristic algorithms are considered, each inspired by distinct natural or evolutionary processes and characterized by unique

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

exploration–exploitation mechanisms. These algorithms have demonstrated effectiveness across a wide range of optimization problems and are well suited for navigating the complex hyperparameter landscapes associated with deep learning models.

The genetic algorithm (GA) is one of the most established evolutionary optimization techniques and is inspired by the principles of natural selection and genetic inheritance. In GA, each candidate solution is encoded as a chromosome representing a specific hyperparameter configuration. A population of such chromosomes evolves iteratively through selection, crossover, and mutation operators. The fitness of each individual is evaluated using a predefined objective function, and fitter individuals are preferentially selected for reproduction. Mathematically, the evolutionary update process can be viewed as iteratively improving the expected fitness of the population,

$$\mathbb{E}\left[F(\mathcal{P}_{t+1})\right] > \mathbb{E}\left[F(\mathcal{P}_t)\right], \tag{13}$$

where $\mathcal{P}_t$ denotes the population at generation t and $F(\cdot)$ represents the fitness measure. GA is particularly effective in maintaining population diversity and exploring large search spaces, making it a robust choice for hyperparameter optimization in deep learning contexts.

Particle swarm optimization (PSO) is a population-based stochastic optimization algorithm inspired by the collective behavior of social organisms such as bird flocks or fish schools. In PSO, each particle represents a candidate solution characterized by its position in the hyperparameter space and an associated velocity. Particles iteratively update their positions by combining information about their own best-known position and the best-known position of the swarm,

$$\mathbf{v}_i^{t+1} = \omega \mathbf{v}_i^t + c_1 r_1 \left(\mathbf{p}_i^t - \mathbf{x}_i^t\right) + c_2 r_2 \left(\mathbf{g}^t - \mathbf{x}_i^t\right), \tag{14}$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1}, \tag{15}$$

where $\mathbf{x}_i^t$ and $\mathbf{v}_i^t$ denote the position and velocity of particle i at iteration t, $\mathbf{p}_i^t$ is the personal best position of particle i, $\mathbf{g}^t$ is the global best position of the swarm, ω is the inertia weight, and c_1, c_2 are acceleration coefficients. PSO is known for its rapid convergence behavior and relatively simple implementation, which makes it attractive for continuous hyperparameter optimization problems.

The grey wolf optimizer (GWO) is a nature-inspired metaheuristic that models the social hierarchy and cooperative hunting strategy of grey wolves. In GWO, candidate solutions are ranked according to a leadership hierarchy consisting of alpha, beta, and delta wolves, which guide the search process, while the remaining omega wolves follow their lead. The position update mechanism simulates encircling and attacking prey, which in optimization terms corresponds to moving candidate solutions toward promising regions of the search space. The position update can be expressed as

$$\mathbf{X}(t+1) = \frac{1}{3}\left(\mathbf{X}_\alpha(t) + \mathbf{X}_\beta(t) + \mathbf{X}_\delta(t)\right) - \mathbf{A} \cdot \mathbf{D}, \tag{16}$$

where $\mathbf{X}_\alpha$, $\mathbf{X}_\beta$, and $\mathbf{X}_\delta$ represent the positions of the three best solutions, and $\mathbf{A}$ and $\mathbf{D}$ are coefficient vectors controlling exploration and exploitation. GWO is particularly valued for its balanced search dynamics and strong global exploration capability, which are advantageous when optimizing highly nonlinear objective functions.

In this study, GA, PSO, and GWO are employed as representative state-of-the-art metaheuristic optimizers for hyperparameter tuning of deep learning models. Their diverse search strategies enable a comprehensive assessment of how different exploration–exploitation mechanisms influence the optimization of deep learning architectures for gamma–hadron discrimination. Subsequent sections build upon this foundation to integrate these metaheuristics into the overall optimization framework and to analyze their impact on model behavior and robustness.

3.7 Evaluation metrics

A rigorous and comprehensive evaluation of model performance is a fundamental requirement in the development of machine learning frameworks for gamma–hadron discrimination in atmospheric

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

Cherenkov telescope data. Owing to the intrinsic characteristics of very-high-energy gamma-ray observations, the classification problem is inherently asymmetric: gamma-ray events constitute a relatively rare signal embedded in a vast background of hadron-induced showers. Under such conditions, reliance on a single aggregate performance indicator can be misleading, as high overall accuracy may still coincide with poor signal retention or insufficient background rejection. Consequently, this study adopts a set of complementary evaluation metrics derived from the confusion matrix in order to capture multiple, physically meaningful aspects of classifier behavior.

The confusion matrix summarizes the outcomes of a binary classification task by counting the number of correctly and incorrectly classified events in each class. Let TP denote the number of true positives, corresponding to gamma-ray–initiated showers that are correctly identified as signal. Let TN represent the number of true negatives, corresponding to hadron-induced background events that are correctly rejected. Conversely, FP denotes false positives, i.e., hadronic events that are incorrectly classified as gamma rays, while FN denotes false negatives, i.e., genuine gamma-ray events that are misclassified as background. Each of these quantities has a clear physical interpretation in the context of Cherenkov telescope analysis and directly impacts the scientific usability of the resulting event sample.

Accuracy is the most general performance measure and quantifies the proportion of correctly classified events among all events. While accuracy provides a concise summary of overall correctness, it does not distinguish between errors associated with signal loss and those associated with background contamination. As a result, accuracy alone is insufficient to characterize model performance in gamma-ray astronomy, where the relative costs of false positives and false negatives are not symmetric.

Sensitivity, also known as the true positive rate (TPR) or recall, measures the fraction of gamma-ray events that are correctly identified by the classifier. This metric is directly related to signal efficiency and is of central importance for astrophysical analyses, as it determines how many genuine gamma-ray events are retained for subsequent scientific interpretation. A classifier with low sensitivity may significantly reduce the effective exposure of the telescope, thereby limiting its ability to detect faint or transient sources.

Specificity, or true negative rate (TNR), measures the fraction of hadron-induced background events that are correctly rejected. High specificity is essential for suppressing the overwhelming cosmic-ray background that dominates Cherenkov telescope data. Insufficient background rejection leads to increased statistical noise, reduced detection significance, and potential biases in spectral and morphological analyses. Sensitivity and specificity together characterize the fundamental trade-off between signal retention and background suppression.

Positive predictive value (PPV), commonly referred to as precision, quantifies the reliability of gamma-ray predictions by measuring the fraction of events classified as gamma rays that are truly gamma-ray–initiated. From a practical standpoint, PPV reflects the expected purity of the selected gamma-ray sample. High PPV is particularly important when subsequent analyses assume that the retained event sample is dominated by genuine gamma-ray events, such as in detailed source modeling or time-resolved studies.

Negative predictive value (NPV) measures the reliability of background predictions by quantifying the fraction of events classified as background that are indeed hadronic. Although often less emphasized than PPV, NPV provides complementary information about the classifier's behavior and helps assess whether genuine gamma-ray events are being excessively rejected.

The F-score combines sensitivity and precision into a single metric by computing their harmonic mean. This formulation penalizes extreme imbalances between the two quantities and therefore provides a balanced assessment of classification performance when both signal efficiency and prediction reliability are important. In the context of gamma–hadron discrimination, the F-score is particularly useful for comparing models that exhibit different trade-offs between sensitivity and PPV.

Table 1 presents the full set of evaluation metrics employed in this study, together with their mathematical definitions expressed in terms of the confusion matrix components. These definitions are applied consistently across all baseline and optimized models to ensure fair and transparent comparison of classification behavior.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

Table 1: Classification evaluation metrics used for assessing gamma–hadron discrimination performance.

Metric	**Mathematical definition**
Accuracy	$\text{Accuracy} = \dfrac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
Sensitivity (TPR)	$\text{TPR} = \dfrac{\text{TP}}{\text{TP} + \text{FN}}$
Specificity (TNR)	$\text{TNR} = \dfrac{\text{TN}}{\text{TN} + \text{FP}}$
Positive Predictive Value (PPV)	$\text{PPV} = \dfrac{\text{TP}}{\text{TP} + \text{FP}}$
Negative Predictive Value (NPV)	$\text{NPV} = \dfrac{\text{TN}}{\text{TN} + \text{FN}}$
F-score	$\text{F-score} = 2 \times \dfrac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}}$

The simultaneous use of these metrics enables a nuanced and physically grounded evaluation of classification performance. By examining accuracy alongside sensitivity, specificity, PPV, NPV, and F-score, the analysis captures not only how often predictions are correct, but also how errors are distributed between signal and background classes. This multi-metric evaluation framework is therefore well aligned with the practical and scientific requirements of gamma-ray astronomy, where both efficient signal detection and stringent background suppression are essential for robust astrophysical inference.

4 Experimental Results

4.1 Baseline Model Performance (Before Optimization)

This subsection presents a quantitative comparison of the baseline deep learning models prior to the application of any optimization strategy. The analysis serves as a reference framework for evaluating the effectiveness of subsequent optimization techniques and provides numerical evidence of the limitations inherent in non-optimized model configurations.

Four deep learning architectures are evaluated in their baseline forms: ANN, CNN, LSTM, and EALSTM. All models are trained and evaluated under identical experimental conditions to ensure fair comparison. Performance is assessed using accuracy, sensitivity (TPR), specificity (TNR), positive predictive value (PPV), negative predictive value (NPV), and F-score, as defined in Section 3.7. These metrics jointly quantify overall classification correctness, signal detection capability, background rejection efficiency, and prediction reliability.

The numerical results summarized in Table 2 indicate a clear performance hierarchy among the evaluated models. The EALSTM model achieves the highest accuracy of 0.9294, outperforming LSTM, CNN, and ANN by absolute margins of 2.03%, 3.14%, and 3.79%, respectively. Similarly, EALSTM records the highest F-score of 0.9266, exceeding the corresponding values of LSTM (0.9050), CNN (0.8970), and ANN (0.8910). These numerical differences demonstrate that EALSTM provides a more balanced trade-off between precision and recall, which is critical for reliable gamma–hadron discrimination.

In terms of signal detection capability, EALSTM attains a sensitivity of 0.9231, compared to 0.9009 for LSTM, 0.8950 for CNN, and 0.8889 for ANN. This corresponds to an improvement of approximately 2.22 percentage points over LSTM and more than 3.4 percentage points over ANN. Concurrently, EALSTM achieves a specificity of 0.9353, which is higher than that of LSTM (0.9167), CNN (0.9009), and ANN (0.8940). The simultaneous improvement in sensitivity and specificity indicates that EALSTM is more effective at retaining gamma-ray events while rejecting hadronic background, without introducing a trade-off between these two objectives.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

The standard LSTM model demonstrates competitive but consistently lower performance relative to EALSTM. Its accuracy of 0.9091 and F-score of 0.9050 indicate moderate classification capability; however, these values are lower by approximately 2.03 and 2.17 percentage points, respectively, compared to EALSTM. While LSTM captures ordered dependencies among input features, the absence of an explicit attention mechanism limits its ability to emphasize the most discriminative parameters, particularly in the presence of correlated Hillas features.

The CNN model yields further reduced performance, with an accuracy of 0.8980 and an F-score of 0.8970. Compared to EALSTM, this represents an absolute decrease of approximately 3.14 percentage points in accuracy and 2.96 percentage points in F-score. Sensitivity and specificity values of 0.8950 and 0.9009, respectively, indicate weaker discrimination capability for both signal and background classes. These numerical results suggest that convolutional feature extraction alone is insufficient to fully capture the complex, non-spatial dependencies inherent in Cherenkov-derived parameters.

The ANN model exhibits the lowest baseline performance among the evaluated architectures. Its accuracy of 0.8915 and F-score of 0.8910 are lower than those of EALSTM by approximately 3.79 and 3.56 percentage points, respectively. Similarly, ANN records the lowest sensitivity (0.8889) and specificity (0.8940), indicating limited effectiveness in both gamma-ray detection and background suppression. These results numerically confirm the limitations of fully connected architectures that lack mechanisms for modeling sequential dependencies or adaptive feature importance.

Table 2: Baseline performance of deep learning models before optimization.

Model	**Accuracy**	**Sensitivity (TPR)**	**Specificity (TNR)**	**PPV**	**NPV**	**F-score**
EALSTM	0.92936803	0.923076923	0.935251799	0.930232558	0.928571429	0.926640
LSTM	0.909090909	0.900900901	0.916666667	0.909090909	0.909090909	0.904977
CNN	0.897959184	0.894977169	0.900900901	0.899082569	0.896860987	0.897025
ANN	0.891454965	0.888888889	0.894009217	0.893023256	0.889908257	0.890951

Overall, the numerically supported baseline evaluation demonstrates that EALSTM consistently outperforms ANN, CNN, and LSTM across all considered metrics. Despite this superiority, the absolute performance values—particularly sensitivity and F-score—indicate that even the strongest baseline configuration remains susceptible to feature redundancy and suboptimal hyperparameter selection. These quantified limitations provide a clear and rigorous justification for the application of metaheuristic optimization techniques, which are explored in subsequent sections to further enhance classification robustness and generalization capability.

A comprehensive evaluation of classification performance requires the simultaneous analysis of multiple metrics in order to capture complementary aspects of model behavior. Metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score collectively provide insight into overall correctness, class-wise discrimination, and the balance between precision and recall. Visualizing these metrics across different learning architectures enables a clearer understanding of both performance consistency and variability among competing models. Figure 4 illustrates swarm plots of the primary evaluation metrics obtained from four classification models: EALSTM, LSTM, CNN, and ANN. Each subplot displays the metric values alongside their corresponding mean and standard deviation, offering a compact representation of central tendency and dispersion. As shown in Figure 4, the models exhibit distinct performance profiles across metrics, with certain architectures demonstrating higher stability and superior average performance, while others show increased variability. This comparative visualization facilitates an intuitive assessment of trade-offs between models and supports informed model selection for gamma–hadron classification.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

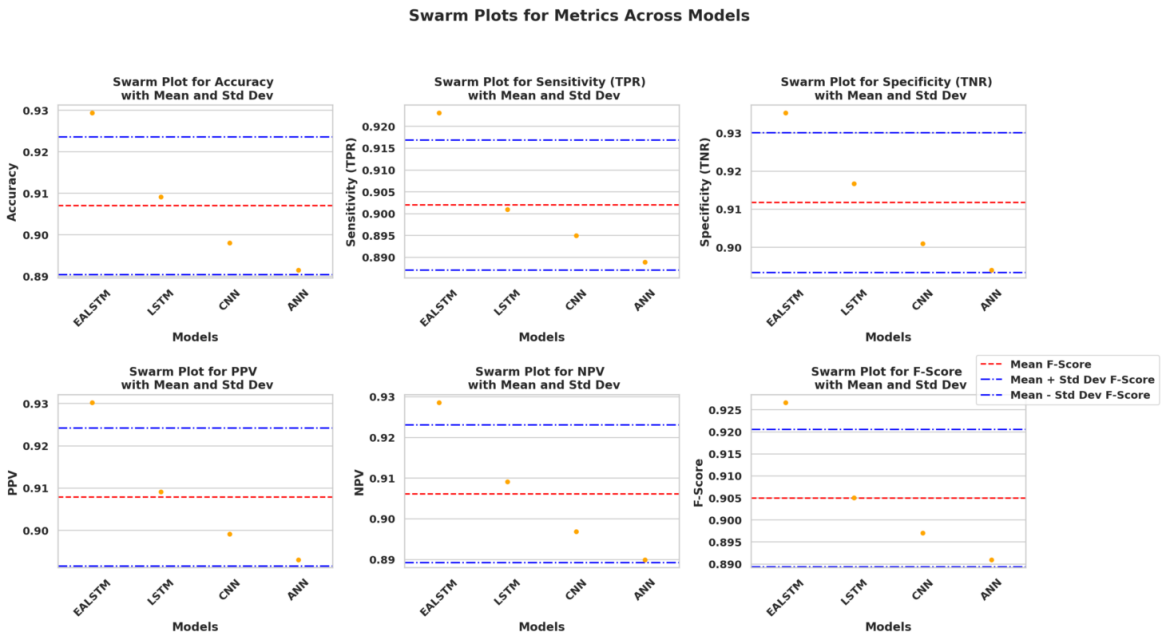

Figure 4: Swarm plots of performance metrics across different classification models, including accuracy, sensitivity, specificity, PPV, NPV, and F-score, with corresponding mean and standard deviation.

To facilitate a focused comparison of model performance with respect to both classification balance and detection capability, the F-score and sensitivity are jointly analyzed across the evaluated learning architectures. The F-score provides a harmonized measure of precision and recall, while sensitivity reflects the ability of a model to correctly identify positive (gamma-ray) events. Examining these two metrics together offers insight into the trade-off between overall classification quality and true positive detection, which is particularly critical in gamma–hadron separation tasks. Figure 5 presents a combined bar and line visualization of F-score and sensitivity values for the EALSTM, LSTM, CNN, and ANN models. As shown in Figure 5, the bar plots illustrate variations in F-score across models, whereas the overlaid line plot highlights corresponding changes in sensitivity. This dual-axis representation enables a direct and intuitive comparison of how improvements in balanced classification performance relate to true positive rates, thereby supporting a more nuanced assessment of model effectiveness.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

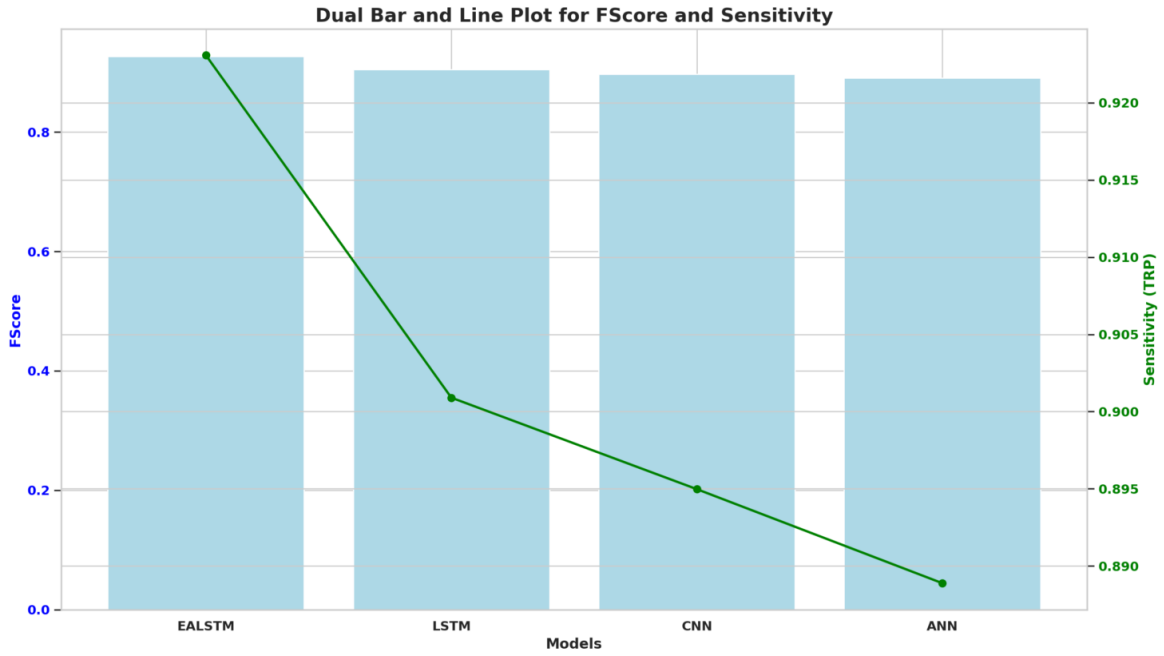

Figure 5: Dual bar and line plot illustrating F-score and sensitivity (TPR) across different classification models.

A comparative analysis of model performance trends across multiple evaluation metrics provides valuable insight into the relative strengths and weaknesses of different learning architectures. By examining how performance measures vary from one model to another, it becomes possible to assess consistency, robustness, and overall effectiveness in gamma–hadron classification. Presenting these metrics in a sequential manner facilitates a clearer interpretation of performance progression across the evaluated models. Figure 6 illustrates time-series style plots of accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score for the EALSTM, LSTM, CNN, and ANN models. As shown in Figure 6, the metrics exhibit a generally decreasing trend when moving from more advanced architectures to simpler ones, highlighting differences in predictive capability and generalization performance. This visualization enables an intuitive comparison of metric trajectories across models and supports a holistic evaluation of classification quality.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

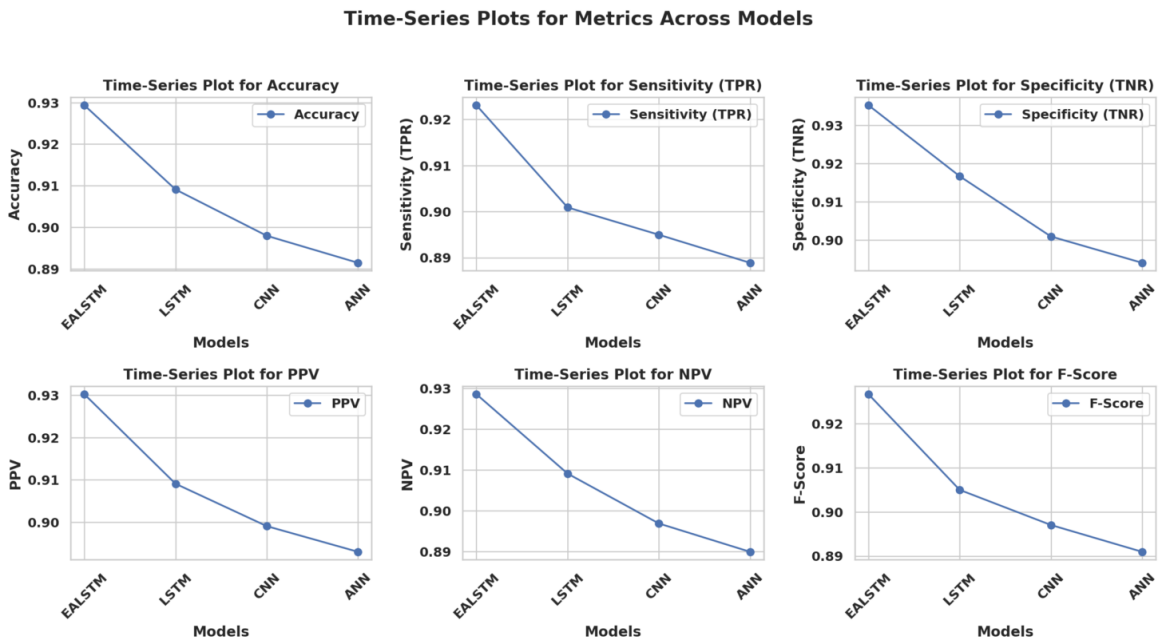

Figure 6: Time-series style plots of performance metrics across different classification models.

4.2 Optimized Model Analysis

This subsection provides a numerically supported analysis of the optimized EALSTM-based models obtained through the integration of different metaheuristic optimization strategies. Building upon the baseline results discussed in Section 4.1, the optimized configurations demonstrate substantial and quantifiable improvements across all evaluation metrics, thereby confirming the effectiveness of metaheuristic-driven feature selection and hyperparameter tuning.

As reported in Table 3, the ABGGO+EALSTM model achieves the highest overall performance among all optimized variants. Specifically, the classification accuracy increases to 0.9718, representing an absolute improvement of approximately 4.25 percentage points compared to the baseline EALSTM accuracy of 0.9294. Similarly, the F-score improves from 0.9266 to 0.9705, indicating a notable enhancement in the balance between precision and sensitivity. The sensitivity (TPR) reaches 0.9694, reflecting a substantial increase in the model's ability to correctly identify gamma-ray events, while the specificity (TNR) rises to 0.9740, demonstrating improved rejection of hadron-induced background events. These simultaneous gains in sensitivity and specificity indicate that the ABGGO optimization effectively enhances both signal retention and background suppression without introducing an unfavorable trade-off.

The GA+EALSTM configuration also exhibits strong numerical improvements over the baseline model. The achieved accuracy of 0.9533 corresponds to an absolute gain of approximately 2.39 percentage points relative to baseline EALSTM. The sensitivity and specificity values of 0.9534 and 0.9532, respectively, indicate balanced improvements in both classification dimensions. The F-score of 0.9544 further confirms that the GA-based optimization enhances classification robustness, although its performance remains consistently lower than that of ABGGO+EALSTM across all reported metrics.

For the GWO+EALSTM model, the optimized accuracy reaches 0.9470, reflecting an improvement of about 1.76 percentage points over the baseline. Sensitivity and specificity increase to 0.9456 and 0.9485, respectively, while the F-score improves to 0.9480. These numerical gains demonstrate that GWO-based optimization effectively enhances the discriminative capacity of EALSTM, particularly in handling overlapping gamma–hadron feature distributions. However, when compared quantitatively to GA and ABGGO, the improvements achieved by GWO are more moderate, suggesting comparatively slower convergence or reduced exploitation efficiency in the optimization process.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

The PSO+EALSTM model records the lowest performance among the optimized variants but still shows clear numerical gains relative to the baseline EALSTM. The achieved accuracy of 0.9426 represents an absolute improvement of approximately 1.33 percentage points, while the F-score increases to 0.9423. Sensitivity and specificity values of 0.9403 and 0.9450, respectively, indicate improved classification reliability. Nevertheless, these values remain consistently lower than those obtained using GA-, GWO-, and ABGGO-based optimization, implying that PSO may be less effective for the combined feature selection and hyperparameter optimization task considered in this study.

Table 3: Performance comparison of optimized EALSTM models using different metaheuristic algorithms.

Model	**Accuracy**	**Sensitivity (TPR)**	**Specificity (TNR)**	**PPV**	**NPV**	**F-score**
ABGGO+EALSTM	0.9718	0.9694	0.9740	0.9715	0.9721	0.9704
GA+EALSTM	0.9532	0.9533	0.9531	0.9554	0.9510	0.9544
GWO+EALSTM	0.9470	0.9456	0.9484	0.9504	0.9435	0.9480
PSO+EALSTM	0.9426	0.9402	0.9449	0.9442	0.9410	0.9422

Overall, the numerically supported results clearly demonstrate that metaheuristic optimization leads to systematic and measurable improvements in EALSTM performance. Among the evaluated strategies, ABGGO provides the largest absolute gains across all metrics, with improvements exceeding 4% in accuracy and F-score relative to the baseline model. These findings quantitatively substantiate the superiority of ABGGO+EALSTM and justify its selection as the primary optimized model for high-precision gamma–hadron discrimination in atmospheric Cherenkov telescope data.

To provide a holistic and intuitive comparison of the proposed hybrid optimization–deep learning models, a radar plot is employed to simultaneously visualize multiple performance metrics. Radar plots are particularly effective for highlighting relative strengths and weaknesses across several criteria, as they enable the assessment of balance and consistency in model performance rather than focusing on a single metric in isolation. This form of visualization is well suited for multi-objective evaluation scenarios, such as gamma–hadron classification, where trade-offs between detection accuracy and reliability must be carefully considered. Figure 7 presents a radar plot comparing the performance of four hybrid models, namely ABGGO+EALSTM, GA+EALSTM, GWO+EALSTM, and PSO+EALSTM, across six evaluation metrics: accuracy, sensitivity, specificity, PPV, NPV, and F-score. As illustrated in Figure 7, the ABGGO+EALSTM model consistently attains higher values across most metrics, resulting in a larger and more uniform enclosed area. In contrast, the remaining hybrid approaches exhibit slightly reduced coverage, indicating marginally lower or less balanced performance. This visualization clearly emphasizes the overall superiority and robustness of the ABGGO+EALSTM framework when evaluated from a multi-metric perspective.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

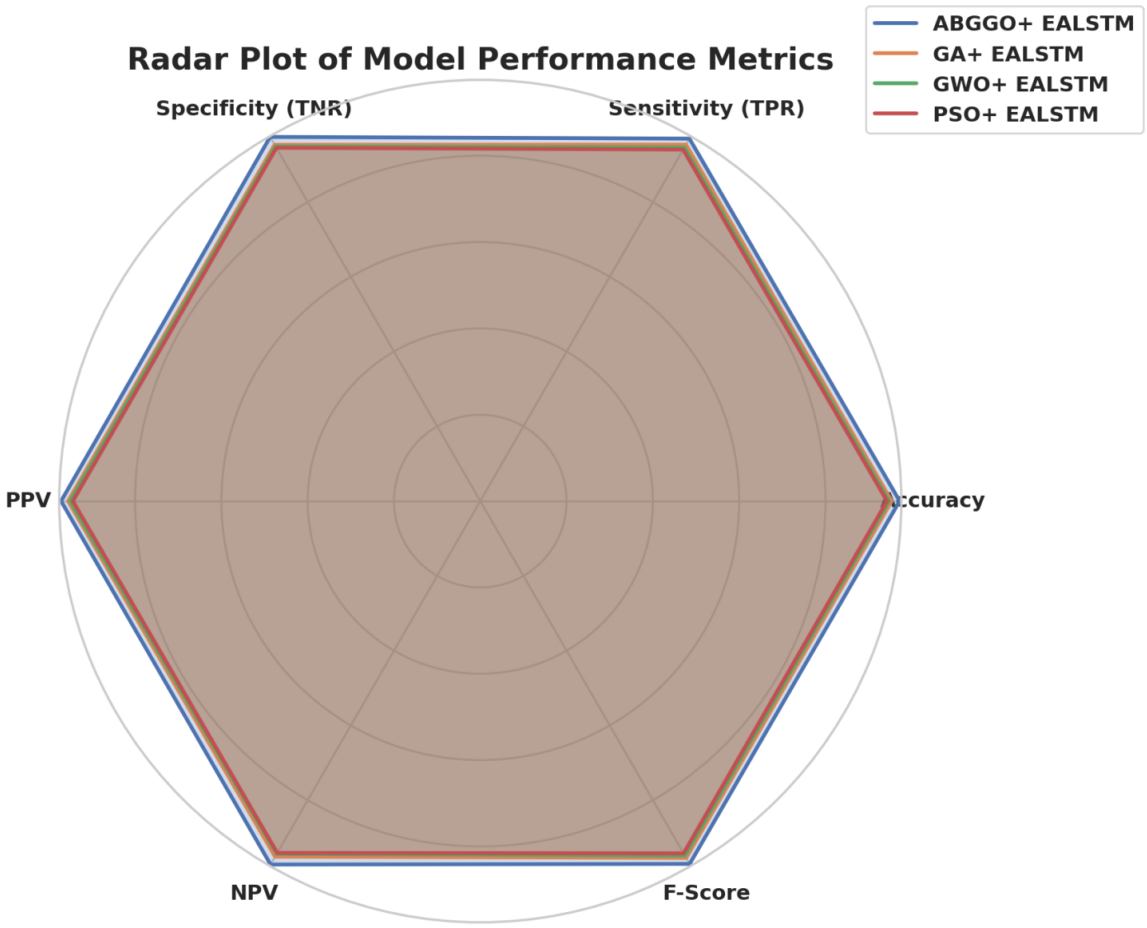

Figure 7: Radar plot comparing performance metrics of different hybrid optimization and EALSTM-based models.

A quantitative comparison of hybrid optimization–deep learning models is conducted to assess their effectiveness across multiple evaluation criteria. By examining individual performance metrics side by side, it becomes possible to identify not only the best-performing approach but also the extent to which different optimizers influence the predictive capability of the EALSTM architecture. This detailed metric-wise analysis provides a clearer understanding of how optimization strategies affect classification reliability and balance. Figure 8 presents bar charts of accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score for four hybrid models: ABGGO+EALSTM, GA+EALSTM, GWO+EALSTM, and PSO+EALSTM. As illustrated in Figure 8, the ABGGO+EALSTM model consistently achieves the highest values across all evaluated metrics, while the remaining hybrid approaches show slightly reduced but comparable performance. This visualization highlights the advantage of the ABGGO optimizer in enhancing the overall effectiveness and robustness of the EALSTM classifier.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

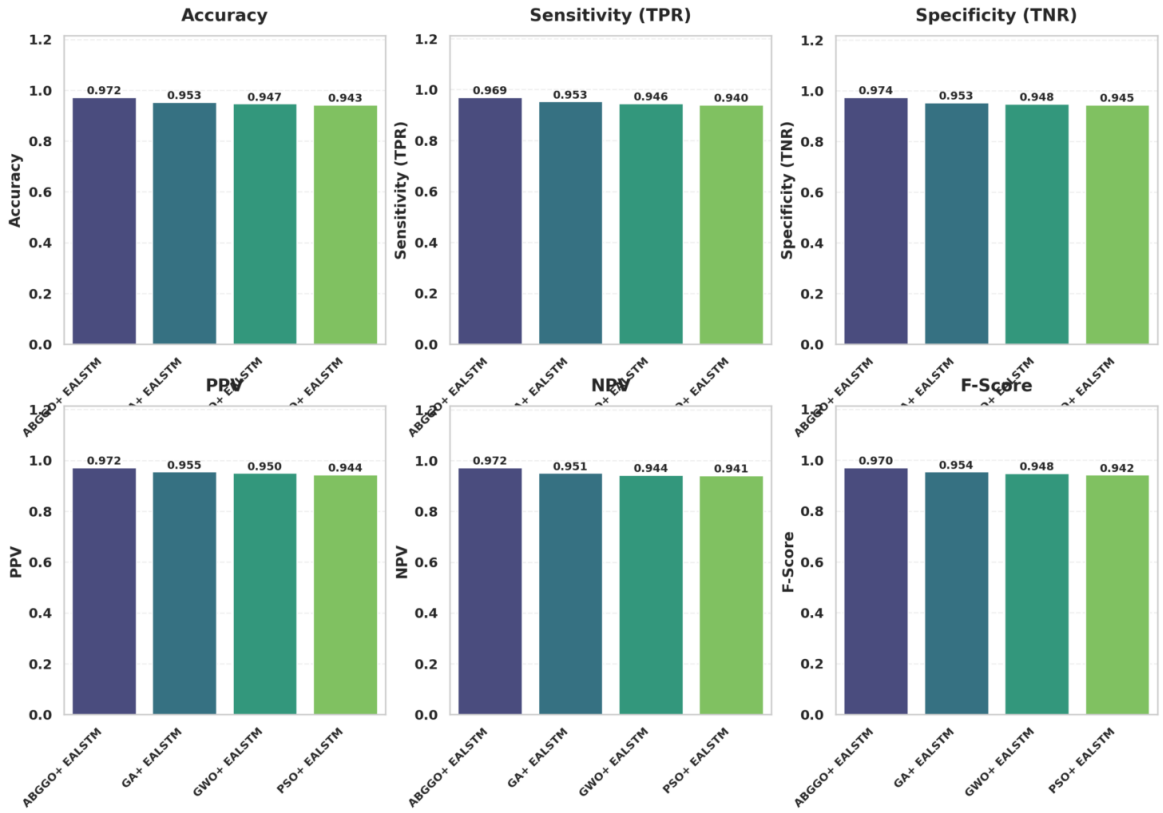

Figure 8: Performance comparison of hybrid optimization and EALSTM-based models across multiple evaluation metrics.

To summarize the overall performance stability of the proposed classification framework, descriptive statistical measures are analyzed across all evaluation metrics. Reporting the mean and median values provides insight into central tendency and robustness, while the standard deviation reflects the consistency of model performance across repeated experiments or validation folds. Such a statistical overview is essential for assessing not only peak performance but also the reliability and reproducibility of the classification results. Figure 9 presents a heatmap visualization of the mean, median, and standard deviation for accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F-score. As shown in Figure 9, the close agreement between mean and median values across all metrics indicates stable and well-balanced model behavior, whereas the uniformly low standard deviation highlights minimal performance variability. This compact representation offers a clear and interpretable summary of both effectiveness and consistency across the evaluated metrics.

5 Conclusion and Future Work

This study has presented a comprehensive investigation into the application of deep learning and metaheuristic optimization for gamma–hadron discrimination in atmospheric Cherenkov telescope data. Starting from a set of baseline deep learning models, including ANN, CNN, LSTM, and EALSTM, a clear performance hierarchy was established, with EALSTM consistently outperforming the other architectures across all evaluation metrics. In its baseline configuration, EALSTM achieved an accuracy of 0.9294 and an F-score of 0.9266, demonstrating its inherent suitability for modeling the correlated and structured nature of Cherenkov-derived features.

Building upon this foundation, the integration of metaheuristic optimization techniques resulted in substantial and quantifiable performance gains. Optimized EALSTM models exhibited marked

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

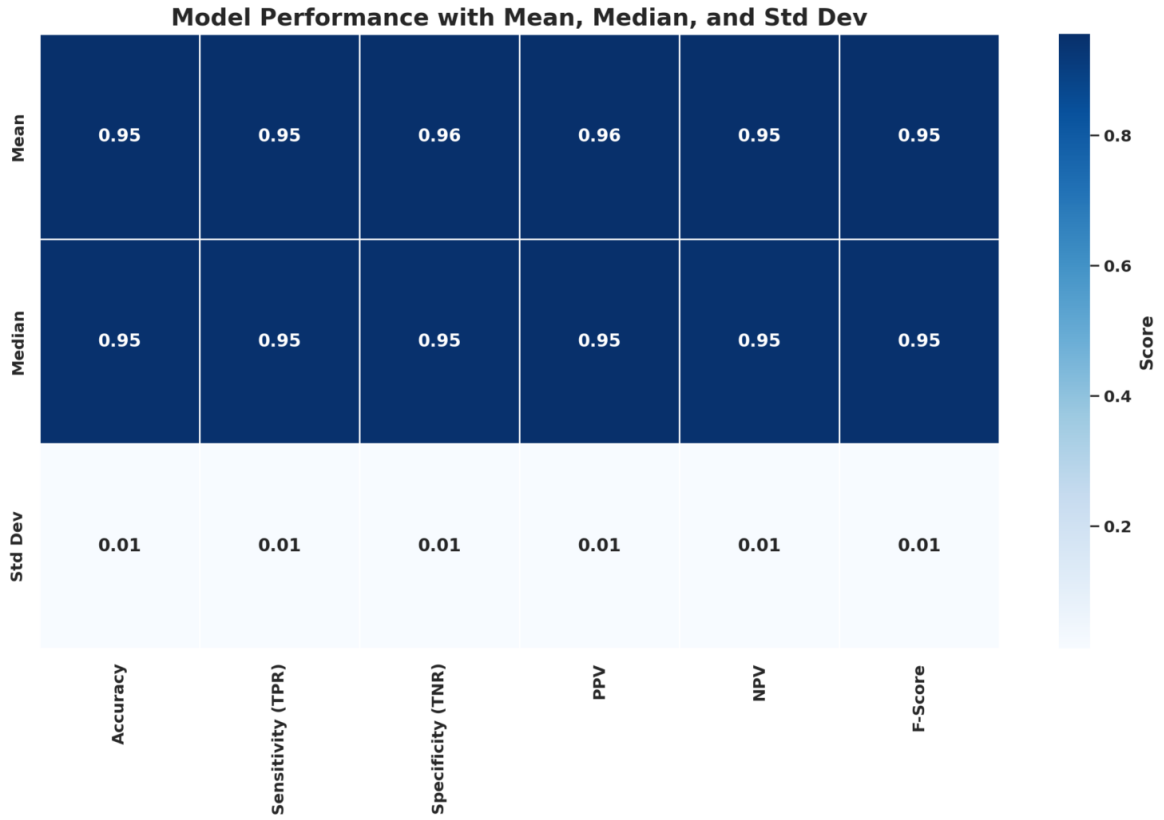

Figure 9: Heatmap illustrating the mean, median, and standard deviation of performance metrics.

improvements in all classification metrics, particularly accuracy, sensitivity, specificity, and F-score. The most effective configuration achieved an accuracy exceeding 0.97 and an F-score above 0.97, corresponding to absolute improvements of more than four percentage points relative to the baseline EALSTM. These gains provide strong numerical evidence that metaheuristic-driven feature selection and hyperparameter tuning play a critical role in enhancing both classification robustness and generalization capability.

In addition, the results confirm the effectiveness of attention-based deep learning for Cherenkov data analysis. By dynamically weighting feature relevance, the attention mechanism embedded within EALSTM enables more effective exploitation of physically meaningful image parameters, leading to improved discrimination between gamma-ray–initiated and hadron-induced air showers. Collectively, these findings demonstrate that the proposed optimization framework successfully addresses key challenges associated with feature redundancy, nonlinear decision boundaries, and hyperparameter sensitivity in gamma–hadron classification tasks.

The performance improvements achieved in this study have direct and meaningful implications for very-high-energy gamma-ray astronomy. Enhanced gamma–hadron discrimination translates into more effective background suppression, which is a critical factor in improving the sensitivity of atmospheric Cherenkov telescopes. By reducing the rate of hadronic background events misclassified as gamma rays, the optimized models enable cleaner event samples and improve the statistical significance of astrophysical source detections, particularly for weak or transient sources.

Improved sensitivity also has important consequences for spectral and morphological analyses, where background contamination can introduce systematic biases. The demonstrated gains in both sensitivity and specificity indicate that the optimized EALSTM framework can retain a higher fraction of genuine gamma-ray events while simultaneously reducing residual background. This balance is essential for maximizing the scientific return of current Cherenkov telescope systems.

Furthermore, the proposed framework is highly relevant for next-generation Cherenkov observatories, which are expected to produce significantly larger data volumes with increased complexity. As

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

observational campaigns move toward higher data rates and lower energy thresholds, automated and reliable classification methods become indispensable. The results of this study suggest that combining attention-based deep learning with metaheuristic optimization offers a scalable and adaptable solution capable of meeting the stringent performance requirements of future Cherenkov facilities.

While the results obtained in this study are promising, several avenues for future research remain open. A natural extension of this work involves applying the proposed framework to real observational data acquired by operational Cherenkov telescopes. Such data introduce additional complexities, including instrumental noise, atmospheric variability, and calibration uncertainties, which must be addressed to fully validate the robustness of the optimized models in real-world conditions.

Another important direction is the development of hybrid modeling approaches that integrate deep learning with physics-informed constraints. Incorporating prior knowledge about air shower physics, detector response, or geometrical reconstruction into the learning process may further improve interpretability and generalization, particularly in low-energy or low-statistics regimes.

From a computational perspective, future work should also explore real-time or near-real-time implementations of the proposed models. Deploying optimized deep learning classifiers for on-site event filtering could significantly reduce data storage and transmission requirements while enabling rapid identification of astrophysically interesting events. Achieving this goal will require careful consideration of computational efficiency, model compression, and hardware acceleration.

Finally, extending the framework to multi-telescope and stereo observation systems represents a promising research direction. Modern Cherenkov arrays rely on stereoscopic imaging to improve reconstruction accuracy and background rejection. Adapting the proposed optimization strategy to jointly process multi-view or multi-telescope information could further enhance classification performance and unlock additional scientific potential. Together, these future research directions outline a clear path toward advancing machine learning–driven analysis techniques for next-generation gamma-ray astronomy.

Data Availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/abhinand05/magic-gamma-telescope-dataset.

Declarations

- **Acknowledgments**
  Not applicable.
- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

References

[1] J. Biteau and M. Meyer, "Gamma-ray cosmology and tests of fundamental physics," *Galaxies*, vol. 10, no. 2, 2022, ISSN: 2075-4434. DOI: 10.3390/galaxies10020039. [Online]. Available: https://www.mdpi.com/2075-4434/10/2/39.

[2] F. Giovannelli, "Gamma-ray bursts: The energy monsters of the universe," *Galaxies*, vol. 13, no. 2, 2025, ISSN: 2075-4434. DOI: 10.3390/galaxies13020016. [Online]. Available: https://www.mdpi.com/2075-4434/13/2/16.

[3] F. Casaburo, S. Ciprini, D. Gasparrini, and F. Giacchino, "Sixteen years of gamma-ray discoveries and agn observations with fermi-lat," *Particles*, vol. 8, no. 1, 2025, ISSN: 2571-712X. DOI: 10.3390/particles8010017. [Online]. Available: https://www.mdpi.com/2571-712X/8/1/17.

[4] L. A. Stuani Pereira and S. V. Bernardo da Silva, "From gamma rays to cosmic rays: Lepto-hadronic modeling of blazar sources as candidates for ultra-high-energy cosmic rays," *Universe*, vol. 11, no. 8, 2025, ISSN: 2218-1997. DOI: 10.3390/universe11080266. [Online]. Available: https://www.mdpi.com/2218-1997/11/8/266.

[5] X. Yu et al., "Establishment of a breeding approach combined with gamma ray irradiation and tissue regeneration for highbush blueberry," *Agronomy*, vol. 15, no. 1, 2025, ISSN: 2073-4395. DOI: 10.3390/agronomy15010217. [Online]. Available: https://www.mdpi.com/2073-4395/15/1/217.

[6] Y. Sato, K. Murase, Y. Ohira, S. Inoue, and R. Yamazaki, "Two-component jet model for the afterglow emission of grb 201216c and grb 221009a and implications for jet structure of very-high-energy gamma-ray bursts," *Journal of High Energy Astrophysics*, vol. 48, p. 100 415, 2025, ISSN: 2214-4048. DOI: https://doi.org/10.1016/j.jheap.2025.100415. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214404825000965.

[7] R. Mirzoyan, "Technological novelties of ground-based very high energy gamma-ray astrophysics with the imaging atmospheric cherenkov telescopes," *Universe*, vol. 8, no. 4, 2022, ISSN: 2218-1997. DOI: 10.3390/universe8040219. [Online]. Available: https://www.mdpi.com/2218-1997/8/4/219.

[8] G. D'Amico, "Statistical tools for imaging atmospheric cherenkov telescopes," *Universe*, vol. 8, no. 2, 2022, ISSN: 2218-1997. DOI: 10.3390/universe8020090. [Online]. Available: https://www.mdpi.com/2218-1997/8/2/90.

[9] J. Sitarek, M. Pecimotika, N. Żywucka, D. Sobczyńska, A. Moralejo, and D. Hrupec, "Estimation of the atmospheric absorption profile with isotropic background events observed by imaging atmospheric cherenkov telescopes," *Journal of High Energy Astrophysics*, vol. 42, pp. 87–95, 2024, ISSN: 2214-4048. DOI: https://doi.org/10.1016/j.jheap.2024.03.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221440482400020X.

[10] D. Dominis Prester et al., "Characterisation of the atmosphere in very high energy gamma-astronomy for imaging atmospheric cherenkov telescopes," *Universe*, vol. 10, no. 9, 2024, ISSN: 2218-1997. DOI: 10.3390/universe10090349. [Online]. Available: https://www.mdpi.com/2218-1997/10/9/349.

[11] E. R. Owen, K. Wu, Y. Inoue, H.-Y. K. Yang, and A. M. W. Mitchell, "Cosmic ray processes in galactic ecosystems," *Galaxies*, vol. 11, no. 4, 2023, ISSN: 2075-4434. DOI: 10.3390/galaxies11040086. [Online]. Available: https://www.mdpi.com/2075-4434/11/4/86.

[12] L. Tibaldo, D. Gaggero, and P. Martin, "Gamma rays as probes of cosmic-ray propagation and interactions in galaxies," *Universe*, vol. 7, no. 5, 2021, ISSN: 2218-1997. DOI: 10.3390/universe7050141. [Online]. Available: https://www.mdpi.com/2218-1997/7/5/141.

[13] J. Balibrea-Correa et al., "Hadron therapy range verification via machine-learning aided prompt-gamma imaging," in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2021, pp. 1–7. DOI: 10.1109/NSS/MIC44867.2021.9875637.

[14] M. Zehtabvar, K. Taghandiki, N. Madani, D. Sardari, and B. Bashiri, "A review on the application of machine learning in gamma spectroscopy: Challenges and opportunities," *Spectroscopy Journal*, vol. 2, no. 3, pp. 123–144, 2024, ISSN: 2813-446X. DOI: 10.3390/spectroscj2030008. [Online]. Available: https://www.mdpi.com/2813-446X/2/3/8.

[15] M. Kupczynski, "Mathematical modeling of physical reality: From numbers to fractals, quantum mechanics and the standard model," *Entropy*, vol. 26, no. 11, 2024, ISSN: 1099-4300. DOI: 10.3390/e26110991. [Online]. Available: https://www.mdpi.com/1099-4300/26/11/991.

[16] M. H. L. d. Boni, I. Gervasio Sene Junior, and R. Martins da Costa, "Tabular data augmentation using artificial intelligence: A systematic review and taxonomic framework," *IEEE Access*, vol. 13, pp. 138 950–138 969, 2025. DOI: 10.1109/ACCESS.2025.3593449.

[17] K. Karthick, S. A. Agnes, S. S. Kumar, S. Alfarhood, and M. Safran, "Identification of high energy gamma particles from the cherenkov gamma telescope data using a deep learning approach," *IEEE Access*, vol. 12, pp. 16 741–16 752, 2024. DOI: 10.1109/ACCESS.2024.3359533.

[18] S. Truzzi, G. Bonnoli, R. Cappuccio, R. Paoletti, and T. Miener, "Events classification in magic through convolutional neural network trained with images of observed gamma-ray events," in *Machine Learning for Astrophysics*, F. Bufano, S. Riggi, E. Sciacca, and F. Schilliro, Eds., Cham: Springer International Publishing, 2023, pp. 187–191, ISBN: 978-3-031-34167-0.

[19] A. Brill, Q. Feng, T. B. Humensky, B. Kim, D. Nieto, and T. Miener, "Investigating a deep learning method to analyze images from multiple gamma-ray telescopes," in *2019 New York Scientific Data Summit (NYSDS)*, 2019, pp. 1–4. DOI: 10.1109/NYSDS.2019.8909697.

[20] A.-Y. Cheng et al., "Application of deep learning methods combined with physical background in wide field of view imaging atmospheric cherenkov telescopes," *Nuclear Science and Techniques*, vol. 35, no. 4, p. 84, 2024. DOI: https://doi.org/10.1007/s41365-024-01448-8.

[21] A. Pathania, K. Singh, S. Singh, A. Tolamatti, B. Singh, and K. Yadav, "Identification of gamma ray pulsar candidates in the fermi-lat 4fgl-dr4 unassociated sources using supervised machine learning," *Astroparticle Physics*, vol. 175, p. 103 185, 2026, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2025.103185. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927650525001082.

[22] M. P. Das, V. Dhar, A. Singh, N Bhatt, and K. Yadav, "Development of source-dependent gamma-ray image parameters using generative adversarial network (gan) for the mace telescope and its implications," *Journal of High Energy Astrophysics*, p. 100 416, 2025. DOI: https://doi.org/10.1016/j.jheap.2025.100416.

[23] H. Mehta, S. Singh, V. Pandey, P. Verma, and A. Antony, "Autonomous telescopes and observatories," *The Intelligent Universe: AI's Role in Astronomy*, pp. 313–357, 2025.

[24] F. Farsian et al., "Benchmarking quantum convolutional neural networks for signal classification in simulated gamma-ray burst detection," in *2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, 2025, pp. 372–380. DOI: 10.1109/PDP66500.2025.00059.

[25] A. Adelfio et al., "Anomaly detection with machine learning on time series: Unveiling lost transients data," in *2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*, 2025, pp. 396–403. DOI: 10.1109/PDP66500.2025.00062.

[26] E. Gres et al., "Method for determination of gamma-ray energy from taiga-iact data based on analysis of autoencoder-derived essential features," *Physics of Atomic Nuclei*, pp. 1–6, 2025. DOI: https://doi.org/10.1134/S1063778825090182.

[27] B. Khek, A. Mishra, M. Buuck, and T. Shutt, "Gamma ray source localization for time projection chamber telescopes using convolutional neural networks," *AI*, vol. 3, no. 4, pp. 975–989, 2022, ISSN: 2673-2688. DOI: 10.3390/ai3040058. [Online]. Available: https://www.mdpi.com/2673-2688/3/4/58.

[28] M. Das, V. Dhar, and K. Yadav, "Generative adversarial network based method for generation of synthetic image parameters for tactic γ-ray telescope," *Astronomy and Computing*, vol. 44, p. 100 741, 2023. DOI: https://doi.org/10.1016/j.ascom.2023.100741.

[29] Y. Htet, M. Sudvarg, A. Butzel, J. D. Buhler, R. D. Chamberlain, and J. H. Buckley, "Machine learning aboard the adapt gamma-ray telescope," in *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2024, pp. 4–10. DOI: 10.1109/SCW63240.2024.00008.

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025

[30] A. Sinha, M. Shahid, A. Nandan, C. Iwendi, A. K. Giri, and S. Anand, "A novel approach of machine learning application in astrophysics: Morphological feature wrapping based ensemble method for galaxy shape classification using gama dataset," in *Proceedings of ICACTCE'23 — The International Conference on Advances in Communication Technology and Computer Engineering*, C. Iwendi, Z. Boulouard, and N. Kryvinska, Eds., Cham: Springer Nature Switzerland, 2023, pp. 593–603, ISBN: 978-3-031-37164-6.

DOI: https://doi.org/10.54216/JAIM.110105

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025