Journal of Artificial Intelligence and Metaheuristic

♦ ISSN: 2833-5597 Vol: 11, No. 01, 2026 ♦ PP. 50–55

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

FULL LENGTH ARTICLE

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

Received: August 20, 2025 Revised: October 20, 2025 Accepted: December 22, 2025 ⋆ Corresponding author

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 0.9704, 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

1. INTRODUCTION

Very-high-energy (VHE) gamma-ray astronomy is 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. Unlike charged cosmic rays, gamma rays propagate essentially undeflected by interstellar and intergalactic magnetic fields, thereby preserving directional information about their production sites. VHE observations therefore provide a uniquely informative probe of compact objects and energetic transients, including pulsars and their nebulae, active galactic nuclei, and supernova remnants.

Ground-based atmospheric Cherenkov gamma telescopes provide highly sensitive measurements in the VHE regime by using the Earth's atmosphere as a calorimetric detection medium. When a primary gamma ray enters the upper atmosphere, it initiates an electromagnetic cascade whose relativistic charged secondaries emit Cherenkov radiation. The resulting nanosecond-scale optical signal is recorded by fast photodetector cameras mounted on large reflectors. Imaging atmospheric Cherenkov telescopes (IACTs) reconstruct event properties by analyzing the spatial and intensity distribution of Cherenkov light in the camera plane.

A persistent bottleneck in IACT analysis is the gamma–hadron discrimination problem. Hadronic cosmic rays vastly outnumber astrophysical gamma rays and produce air-shower images that can partially resemble electromagnetic cascades. This overlap is especially problematic at low energies, where images are faint, fluctuations are larger, and image-parameter distributions become less separable. Classical cut-based methods based on Hillas parameters are interpretable and computationally efficient, but their linear or manually tuned boundaries are often insufficient for exploiting nonlinear correlations among image descriptors.

Machine learning and deep learning have improved event classification by learning nonlinear decision functions from Cherenkov image parameters or camera images. However, non-optimized architectures can still suffer from redundant features, suboptimal hyperparameters, and limited sensitivity to the most discriminative attributes. This work addresses these limitations by integrating an attention-guided recurrent classifier with adaptive metaheuristic optimization.

The main contributions are: (i) an enhanced attention-based LSTM model for gamma–hadron classification, (ii) an Adaptive Balanced Greylag Goose Optimization (ABGGO) strategy for feature selection and hyperparameter tuning, (iii) a systematic comparison with ANN, CNN, LSTM, and several optimized EALSTM variants, and (iv) a multi-metric evaluation covering accuracy, sensitivity, specificity, PPV, NPV, and F-score.

2. LITERATURE REVIEW

Gamma-ray astronomy has benefited from both physical modeling and data-driven classification. Early IACT pipelines relied heavily on image moments, directional reconstruction, and manually optimized selection cuts. These approaches remain useful because Hillas parameters encode physically meaningful information such as image length, width, size, orientation, asymmetry, and concentration. Nevertheless, the overlap between gamma and hadron distributions limits the discriminative power of fixed thresholds.

Machine learning methods such as decision trees, random forests, support vector machines, and ensemble classifiers have been widely adopted to improve background rejection. More recently, deep learning models have been explored for gamma-ray event classification, including CNNs trained on camera images, LSTMs and recurrent structures applied to ordered features, and hybrid approaches that combine physical descriptors with learned representations. Attention mechanisms are particularly attractive because they can weight the relative importance of features and focus the model on discriminative components of the event representation.

Metaheuristic optimization has also become important for model design. Genetic algorithms, particle swarm optimization, grey wolf optimization, and related population-based methods have been used to search high-dimensional spaces where gradient-based optimization is inconvenient or non-differentiable. In gamma–hadron separation, such methods can tune hyperparameters and select informative subsets of Cherenkov descriptors, reducing redundancy while improving generalization.

3. MATERIALS AND METHODS

3.1 Dataset Description

The study uses a Monte Carlo–generated MAGIC gamma telescope dataset containing continuous-valued event descriptors. Each event is represented by image parameters extracted from Cherenkov shower images and assigned to either the gamma class or the hadron class. The features include image size, length, width, concentration measures, asymmetry, higher-order moments, and source-position descriptors. These variables encode the geometry, intensity distribution, and orientation of shower images in the camera plane. Figure 1 illustrates the class-wise distributions of the ten commonly used MAGIC features, while Figure 2 summarizes their central tendency and tail behavior. Figure 3 focuses on the distributions of fSize, fLength, and fWidth, which show substantial overlap but still contain useful discriminative structure.

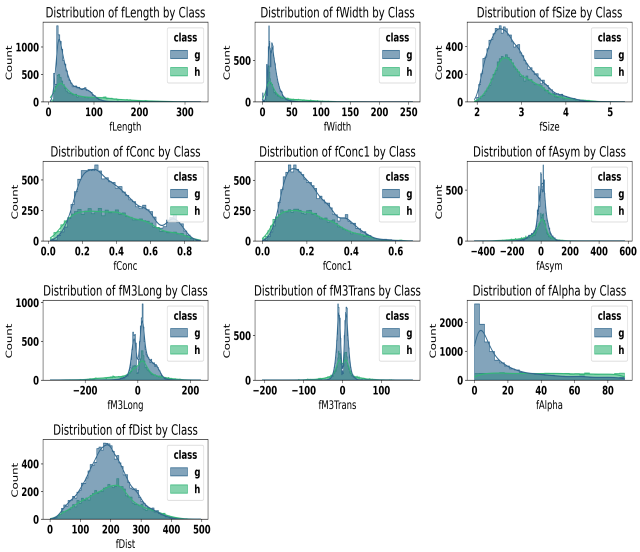

Figure 1. Class-wise distributions of MAGIC gamma telescope features for gamma and hadron events.

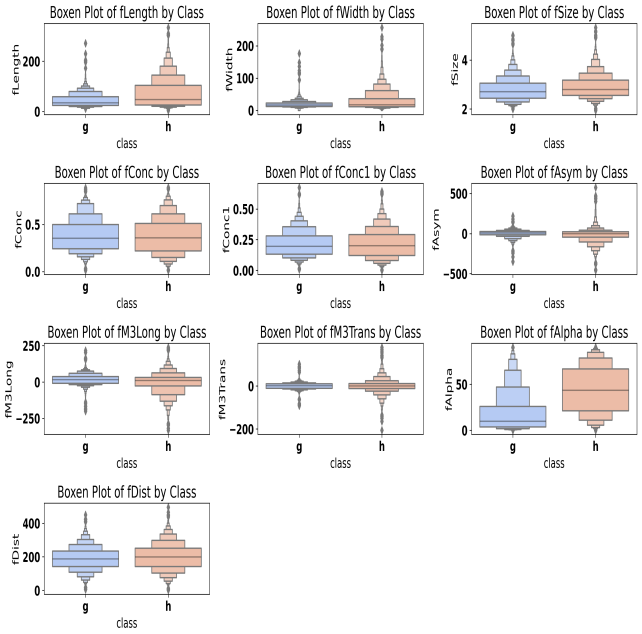

Figure 2. Boxen plots of MAGIC gamma telescope features for gamma and hadron events.

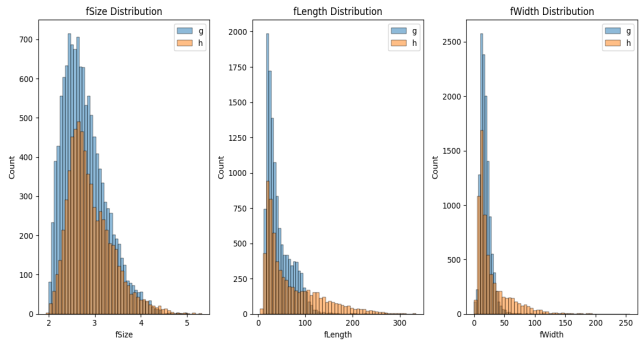

Figure 3. Distributions of fSize, fLength, and fWidth for gamma and hadron events.

3.2 Data Preprocessing

Data preprocessing is critical because the extracted image parameters are physically correlated, span different numerical ranges, and are affected by shower fluctuations and detector-related noise. All input variables are continuous descriptors derived after standard image cleaning procedures. Since the data are generated by Monte Carlo simulation, conventional missing values are not expected; however, preprocessing is still required to address differences in scale, variance, and distribution shape.

Feature scaling is applied uniformly across continuous attributes so that variables with large magnitudes do not dominate gradient updates. Scaling parameters are fitted exclusively on the training subset and then applied to validation and testing subsets to avoid information leakage. Class labels are encoded numerically as $y \in \{0,1\}$, and the dataset is partitioned into stratified training, validation, and testing subsets. This pipeline ensures that all models and optimizers receive consistent input representations.

3.3 Deep Learning Models

Let $\mathbf{x} \in \mathbb{R}^d$ denote the input feature vector for one Cherenkov event, and let $y \in \{0,1\}$ denote its class label. A generic neural classifier estimates

$$\hat{y} = f_\theta(\mathbf{x}), \tag{1}$$

where θ contains trainable parameters. For binary classification, the output probability is produced with a sigmoid activation,

$$\hat{p} = \sigma(z) = \frac{1}{1+e^{-z}}. \tag{2}$$

Model parameters are trained by minimizing binary cross-entropy,

$$\mathscr{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{p}_i) + (1-y_i)\log(1-\hat{p}_i)\right]. \tag{3}$$

Four models are compared. ANN provides a fully connected baseline; CNN captures local feature interactions; LSTM models ordered dependencies across the feature sequence; and EALSTM extends LSTM with attention to emphasize informative components. In an LSTM cell, the gates are defined as

$$\mathbf{i}_t = \sigma(\mathbf{W}_i\mathbf{x}_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i), \tag{4}$$
$$\mathbf{f}_t = \sigma(\mathbf{W}_f\mathbf{x}_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f), \tag{5}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o\mathbf{x}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o), \tag{6}$$
$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c\mathbf{x}_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c), \tag{7}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \tag{8}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \tag{9}$$

The attention layer computes normalized weights

$$\alpha_t = \frac{\exp(e_t)}{\sum_j \exp(e_j)}, \qquad e_t = \mathbf{v}^T \tanh(\mathbf{W}_a\mathbf{h}_t + \mathbf{b}_a), \tag{10}$$

and forms the context vector

$$\mathbf{c} = \sum_t \alpha_t \mathbf{h}_t. \tag{11}$$

3.4 Metaheuristic Optimization

The optimization stage searches for feature subsets and hyperparameters that maximize validation performance. A candidate solution encodes selected features and model-control parameters. The objective function rewards strong classification performance while discouraging unnecessarily complex solutions:

$$F(\mathbf{s}) = \lambda_1 \,\text{Accuracy}(\mathbf{s}) + \lambda_2 F_1(\mathbf{s}) - \lambda_3 \frac{|\mathscr{S}|}{d}, \tag{12}$$

where $\mathscr{S}$ is the selected feature subset and d is the total feature count.

3.5 Greylag Goose Optimization

Greylag Goose Optimization (GGO) is a population-based search method inspired by cooperative movement and role allocation in greylag goose groups. A population Y_i, $i = 1,2,\ldots,n$, represents feasible solutions. The best individual

Z guides the population while exploration and exploitation groups are dynamically balanced. At iteration t, the population is divided as

$$P_t = P_t^{\text{explore}} \cup P_t^{\text{exploit}}, \qquad |P_t^{\text{explore}}| + |P_t^{\text{exploit}}| = n. \quad (13)$$

The algorithm increases exploration when improvement stagnates and shifts toward exploitation as promising regions emerge.

Table 1. GGO algorithmic procedure.

Algorithm 1: GGO Algorithm
Require population Y_i, size n, maximum iterations $t_{\max}$, objective function F_n, and parameters $b, B, D, \ldots$
Ensure best agent position Z
Initialize the population and calculate the objective function for each agent.
Determine the best position Z and divide agents into exploration and exploitation groups.
For each iteration $t = 1$ to $t_{\max}$, update exploration agents using stochastic search equations.
Update exploitation agents based on iteration conditions and random parameters.
Evaluate the objective function and update Z if a better solution is found.
If no improvement occurs over successive iterations, adapt the sizes of the exploration and exploitation groups.
Return the best agent position Z.

3.6 Dynamic Optimization and ABGGO

Dynamic optimization considers objectives or feasible regions that vary with time:

$$\min / \max\ F(\mathbf{L}, t), \qquad \mathbf{L} \in S(t), \quad (14)$$

where $F(\mathbf{L}, t)$ is time-dependent and $S(t)$ is a feasible region. The proposed ABGGO extends GGO with an adaptive balance between exploration and exploitation, preserving di-versity when the search stagnates and intensifying the search near strong candidate solutions. This dynamic mechanism is well suited to feature selection and hyperparameter optimiza-tion because the effective search landscape changes as model validation performance evolves.

3.7 Evaluation Metrics

Performance is assessed with six complementary metrics. Accuracy measures overall correctness; sensitivity or TPR measures gamma-event detection; specificity or TNR mea-sures hadron rejection; PPV and NPV quantify predictive reliability; and F-score summarizes precision–recall balance.

Table 2. Balanced and adaptive GGO procedure.

Algorithm 2: BCGGO/ABGGO Algorithm
Initialize population, solution encodings, optimizer parameters, and the EALSTM search space.
Evaluate all candidates using validation metrics and identify the current best solution.
Partition the population into balanced exploration and exploitation subgroups.
Perform exploration using multiple guide solutions to maintain population diversity.
Perform exploitation around sentry or elite solutions to refine high-quality candidates.
Adapt the subgroup sizes when objective values stagnate.
Continue until the maximum number of iterations is reached, then return the selected features and optimized hyperparameters.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (15)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \quad (16)$$

$$\text{Specificity} = \frac{TN}{TN+FP}, \quad (17)$$

$$\text{PPV} = \frac{TP}{TP+FP}, \quad (18)$$

$$\text{NPV} = \frac{TN}{TN+FN}, \quad (19)$$

$$F_1 = \frac{2\,\text{PPV} \times \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}}. \quad (20)$$

Table 3. Evaluation metrics used for gamma–hadron classification.

Metric	Interpretation
Accuracy	Overall proportion of correctly classified events
Sensitivity (TPR)	Fraction of gamma events correctly retained
Specificity (TNR)	Fraction of hadron events correctly rejected
PPV	Reliability of positive gamma predictions
NPV	Reliability of negative hadron predictions
F-score	Harmonic mean of precision and recall

4. EXPERIMENTAL RESULTS

4.1 Baseline Model Performance Before Optimization

Four deep learning architectures are evaluated in baseline form: ANN, CNN, LSTM, and EALSTM. The results show a clear hierarchy. EALSTM achieves the highest accuracy of 0.9294 and the highest F-score of 0.9266, outperforming LSTM, CNN, and ANN. The attention mechanism helps the model emphasize physically informative Cherenkov descriptors and improves both signal retention and background rejection. Table 4 reports the numerical comparison, and

Figures 4–6 provide the corresponding visual comparisons across the same metrics.

Table 4. Baseline performance of deep learning models before optimization.

Model	Accuracy	TPR	TNR	PPV	NPV	F-score
EALSTM	0.9294	0.9231	0.9353	0.9302	0.9286	0.9266
LSTM	0.9091	0.9009	0.9167	0.9091	0.9091	0.9050
CNN	0.8980	0.8950	0.9009	0.8991	0.8969	0.8970
ANN	0.8915	0.8889	0.8940	0.8930	0.8899	0.8910

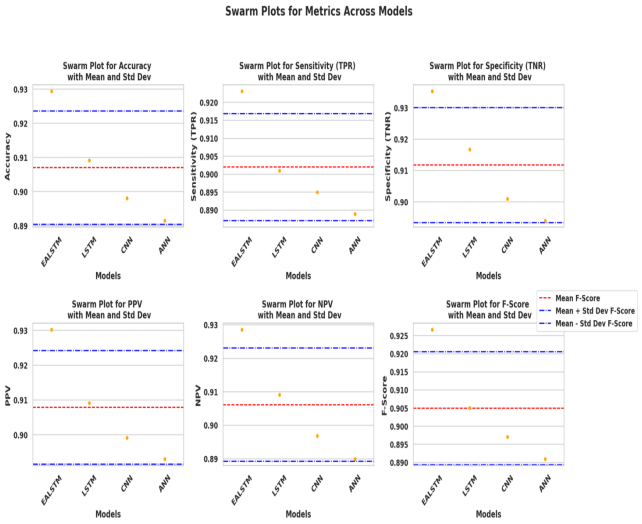

Figure 4. Swarm plots of performance metrics across ANN, CNN, LSTM, and EALSTM.

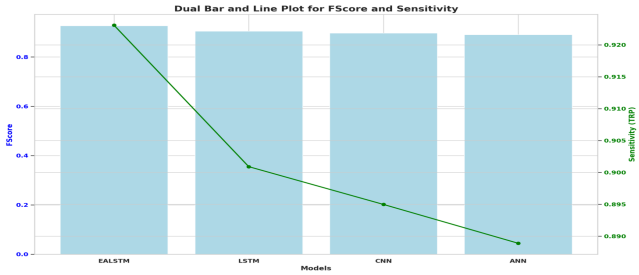

Figure 5. Dual bar and line plot illustrating F-score and sensitivity across baseline models.

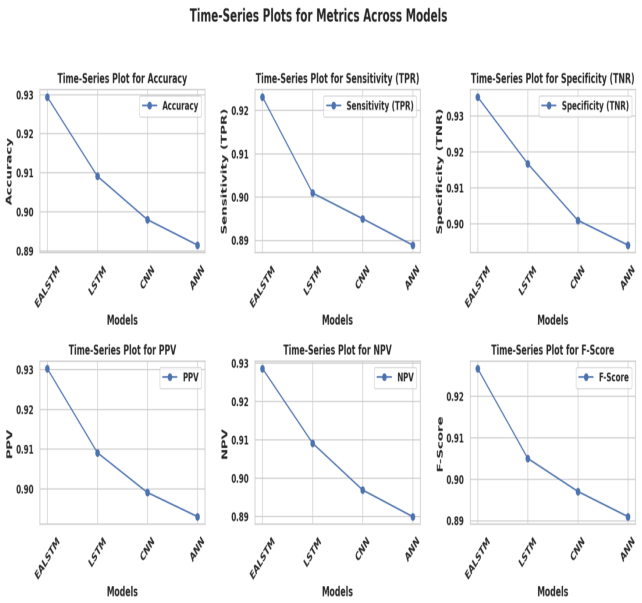

Figure 6. Time-series style plots of performance metrics across baseline classification models.

The ANN model exhibits the weakest baseline performance, while CNN improves local nonlinear modeling but remains less effective than recurrent methods. LSTM captures ordered dependencies among input features but lacks explicit attention. EALSTM provides the most balanced performance across accuracy, sensitivity, specificity, PPV, NPV, and F-score.

4.2 Optimized Model Analysis

Metaheuristic optimization is applied to the EALSTM framework using GA, GWO, PSO, and the proposed ABGGO. The optimized models improve over the baseline configuration, confirming that feature selection and hyperparameter tuning are important for reducing redundancy and improving generalization. ABGGO+EALSTM achieves the best overall results. Table 5 gives the optimized model comparison, while Figures 7–9 visualize the metric balance and stability of the optimized configurations.

Table 5. Performance comparison of optimized EALSTM models using different metaheuristic optimizers.

Model	Accuracy	TPR	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

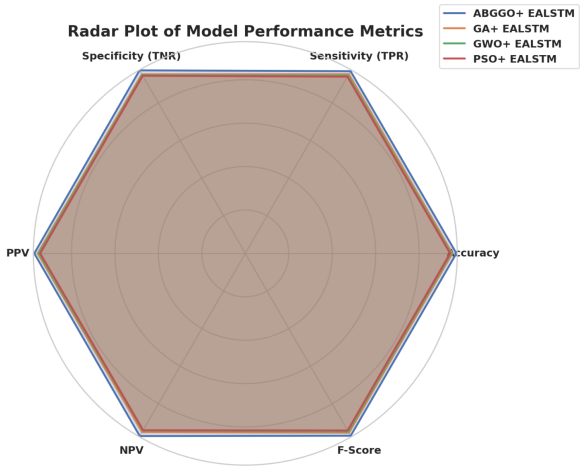

Figure 7. Radar plot comparing performance metrics of hybrid optimization and EALSTM-based models.

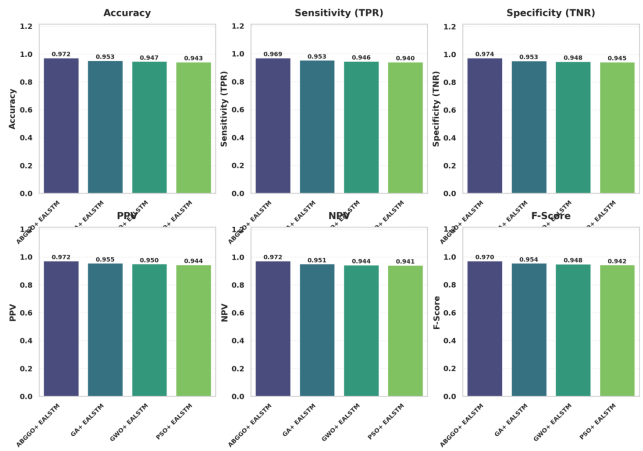

Figure 8. Performance comparison of hybrid optimization and EALSTM-based models across multiple evaluation metrics.

The ABGGO+EALSTM model consistently obtains the largest and most balanced metric profile. Its accuracy of

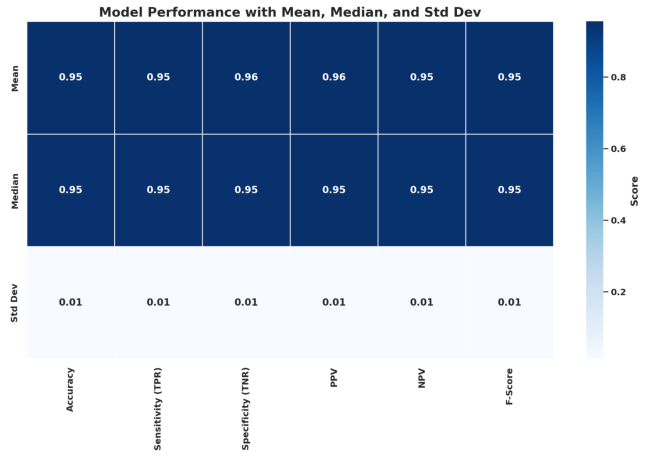

Figure 9. Heatmap illustrating mean, median, and standard deviation of performance metrics.

0.9718, sensitivity of 0.9694, specificity of 0.9740, and F-score of 0.9704 indicate strong gamma-event retention and hadronic background suppression. Compared with baseline EALSTM, these values represent absolute improvements exceeding four percentage points in the principal metrics.

The radar, bar, and heatmap visualizations show that ABGGO not only improves peak performance but also stabilizes the model across repeated evaluations. The close agreement between mean and median metric values and the low standard deviation indicate reliable and reproducible behavior.

5. CONCLUSION AND FUTURE WORK

This study presented a deep learning and metaheuristic optimization framework for gamma–hadron discrimination in atmospheric Cherenkov telescope data. Starting from baseline ANN, CNN, LSTM, and EALSTM models, EALSTM established the strongest non-optimized performance, achieving an accuracy of 0.9294 and an F-score of 0.9266. Its attention mechanism enabled more effective exploitation of physically meaningful image parameters.

The integration of metaheuristic optimization produced substantial gains. The proposed ABGGO+EALSTM configuration achieved the best results, with accuracy of 0.9718 and F-score of 0.9704, outperforming GA+EALSTM, GWO+EALSTM, and PSO+EALSTM. These gains show that adaptive feature selection and hyperparameter tuning can significantly improve background rejection and signal retention.

Future work will focus on validating the framework on real observational data, extending the approach to multi-telescope array analysis, improving real-time event filtering, and integrating camera-image representations with physically interpretable Hillas parameters. Further investigation may also explore uncertainty quantification and deployment on next-generation Cherenkov observatories.

REFERENCES

[1] A. M. Hillas, "Cerenkov light images of EAS produced by primary gamma rays and by nuclei," in *Proc. 19th International Cosmic Ray Conference*, 1985.

[2] T. C. Weekes, *Very High Energy Gamma-Ray Astronomy*. Institute of Physics Publishing, 2003.

[3] E. Lorenz and R. Wagner, "Very-high energy gamma-ray astronomy," *European Physical Journal H*, vol. 37, pp. 459–513, 2012.

[4] R. K. Bock et al., "Methods for multidimensional event classification: A case study using images from a Cherenkov gamma-ray telescope," *Nuclear Instruments and Methods in Physics Research A*, vol. 516, pp. 511–528, 2004.

[5] K. Karthick et al., "Identification of high energy gamma particles from the Cherenkov gamma telescope data using a deep learning approach," *IEEE Access*, vol. 12, pp. 16741–16752, 2024.

[6] S. Truzzi et al., "Events classification in MAGIC through convolutional neural network trained with images of observed gamma-ray events," in *Machine Learning for Astrophysics*, Springer, 2023, pp. 187–191.

[7] A. Brill et al., "Investigating a deep learning method to analyze images from multiple gamma-ray telescopes," in *2019 New York Scientific Data Summit*, 2019.

[8] 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, 2024.

[9] M. P. Das et al., "Development of source-dependent gamma-ray image parameters using generative adversarial network for the MACE telescope," *Journal of High Energy Astrophysics*, 2025.

[10] S. Mirjalili, "Genetic algorithm, particle swarm optimization and grey wolf optimizer: Theory and applications," in *Evolutionary and Swarm Intelligence Methods*, 2019.