






Article

Application of Machine Learning Ensemble Methods to ASTRI Mini-Array Cherenkov Event Reconstruction

Antonio Pagliaro ^{1,2,3,*} , Giancarlo Cusumano ^{1,3} , Antonino La Barbera ^{1,3} , Valentina La Parola ^{1,3} 
and Saverio Lombardi ^{3,4,5} 

¹ INAF IASF Palermo, Via Ugo La Malfa 153, 90146 Palermo, Italy; giancarlo.cusumano@inaf.it (G.C.); antonino.labarbera@inaf.it (A.L.B.); valentina.laparola@inaf.it (V.L.P.)

² Istituto Nazionale di Fisica Nucleare Sezione di Catania, Via Santa Sofia 64, 95123 Catania, Italy

³ ICSC—Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing, Italy; saverio.lombardi@inaf.it

⁴ INAF—Osservatorio Astronomico di Roma, Via Frascati 33, Monte Porzio Catone, 00040 Rome, Italy

⁵ ASI—Space Science Data Center, Via del Politecnico Frascati 33, 00133 Rome, Italy

* Correspondence: antonio.pagliaro@inaf.it

Abstract: The Imaging Atmospheric Cherenkov technique has opened up previously unexplored windows for the study of astrophysical radiation sources in the very high-energy (VHE) regime and is playing an important role in the discovery and characterization of VHE gamma-ray emitters. However, even for the most powerful sources, the data collected by Imaging Atmospheric Cherenkov Telescopes (IACTs) are heavily dominated by the overwhelming background due to cosmic-ray nuclei and cosmic-ray electrons. As a result, the analysis of IACT data necessitates the use of a highly efficient background rejection technique capable of distinguishing a gamma-ray induced signal through identification of shape features in its image. We present a detailed case study of gamma/hadron separation and energy reconstruction. Using a set of simulated data based on the ASTRI Mini-Array Cherenkov telescopes, we have assessed and compared a number of supervised Machine Learning methods, including the Random Forest method, Extra Trees method, and Extreme Gradient Boosting (XGB). To determine the optimal weighting for each method in the ensemble, we conducted extensive experiments involving multiple trials and cross-validation tests. As a result of this thorough investigation, we found that the most sensitive Machine Learning technique applied to our data sample for gamma/hadron segregation is a Stacking Ensemble Method composed of 42% Extra Trees, 28% Random Forest, and 30% XGB. In addition, the best-performing technique for energy estimation is a different Stacking Ensemble Method composed of 45% XGB, 27.5% Extra Trees, and 27.5% Random Forest. These optimal weightings were derived from extensive testing and fine-tuning, ensuring maximum performance for both gamma/hadron separation and energy estimation.

Keywords: machine learning; ensemble learning; imaging atmospheric Cherenkov telescopes; gamma/hadron separation; image analysis; pattern recognition



Citation: Pagliaro, A.; Cusumano, G.; La Barbera, A.; La Parola, V.; Lombardi, S. Application of Machine Learning Ensemble Methods to ASTRI Mini-Array Cherenkov Event Reconstruction. *Appl. Sci.* **2023**, *13*, 8172. <https://doi.org/10.3390/app13148172>

Academic Editor: Seokwon Yeom

Received: 14 June 2023

Revised: 10 July 2023

Accepted: 11 July 2023

Published: 13 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. General Overview

Gamma-ray astronomy has been one of the last branches of astrophysics to arise (the first telescope for observing gamma rays was sent into orbit aboard the Explorer 11 satellite in 1961). Gamma photons cannot reach the Earth's surface. Up to energies of a few hundred GeV, we can collect them by sending instruments aboard satellites in orbit above the atmosphere (see, for example, the Fermi Gamma-ray Space Telescope). However, for higher energies, these technologies are unsuccessful due to the low photon flux, which would require an unthinkably large collection area.

Imaging Atmospheric Cherenkov Telescopes (IACTs), with their specific characteristics [1], permit us to perform very high-energy (VHE) gamma astronomy (above a few tenths GeVs) from the ground with a very large effective collection area (above the order of 10^5 m²). The key

idea is to collect the Cherenkov light induced by the atmospheric cascade ignited by the VHE photon that interacts with the atoms of the atmosphere.

An atmospheric cascade is started when relativistic high energy particles (in this document, we will sometimes refer to them as primaries), such as cosmic rays and photons, reach the Earth's atmosphere. Cherenkov light is produced when secondary charged relativistic particles in the cascade, travelling through the atmosphere, disturb the equilibrium of the atoms in the atmosphere, which subsequently emit a faint blue-ultraviolet radiation as they regain their equilibrium.

As a result, on the focal plane of IACTs we do not have a direct image of the gamma rays coming from the outer space (and therefore of the astrophysical source) but a mere collection of footprints of the Cherenkov radiation triggered by each relativistic high-energy particle. The morphological and temporal features of these footprints can then be used to reconstruct the image of the source for scientific analysis. Since the showers are intrinsically blurred due to their development in the atmosphere, the footprints are approximate ellipsoids extended on a few degrees of arc, so that IACTs usually do not need high-performance optics. On the other hand, since the Cherenkov flash lasts only a few nanoseconds, very fast electronics are required [1].

One of the factors limiting IACTs sensitivity is how challenging it is to suppress the enormous number of cosmic ray background events. In fact, gamma rays make up a very minor fraction of the flux of cosmic rays, and IACTs detect and image the Cherenkov radiation regardless of its origin, either from a gamma photon or from a high-energy charged particle such as a proton. In order to detect gamma-ray sources, any IACT analysis method must therefore be capable of performing an effective background rejection. This involves separating the gamma-ray signal from the far more abundant background of hadron driven showers by exploiting the recognition of shape features in the image.

Usually, the image of a shower initiated by a gamma photon has a quite regular elliptical shape. On the other hand, a hadronic shower, which forms if a cosmic ray is the primary, even though it may contain electromagnetic sub-shower components, typically has a more irregular shape (see Figure 1).

Fortunately, most images have enough information to distinguish the gamma-ray signal from the predominant cosmic-ray background and to reconstruct the arrival direction and energy of the primary.

The extraction of pertinent features, such as the Hillas parameters [2] and more features that will be described later, from the camera pixel data is a prerequisite for IACT image analysis.

1.2. Context

The research presented in this paper stems from the ASTRI Mini-Array project. The ASTRI (*Astrofisica con Specchi a Tecnologia Replicante Italiana*) [3] Mini-Array is an international project led by the Italian National Institute for Astrophysics (INAF) to build and operate an array of nine 4 m class IACTs at the Observatorio del Teide (Tenerife, Spain). The project's scientific agenda focuses on investigating the origin of cosmic rays, the extragalactic background light (EBL) and other fundamental physics topics, and time-domain and multi-messenger astrophysics at the TeV and multi-TeV energy scales. An equally significant aspect of the research encompasses the examination of gamma-ray bursts and multi-messenger transients in the very high-energy (VHE) domain. Additionally, the ASTRI Mini-Array will conduct stellar intensity interferometry studies.

The ASTRI Mini-Array is designed to be sensitive to energies in the 1–200 TeV energy range, and to achieve an angular resolution of $\sim 3'$, and an energy resolution of $\sim 10\%$ above about 10 TeV.

Astrophysical observations in this energy range can offer valuable insights into high-energy particle sources, such as the so-called PeVatrons recently seen by LHAASO up to 1.4 PeV [4].

This performance would enable the extraction of crucial morphological and spectral data, which is essential for distinguishing between various proposed acceleration mechanisms. For further information on the ASTRI Mini-Array Core Science program, please refer to the dedicated paper by the ASTRI collaboration [5].

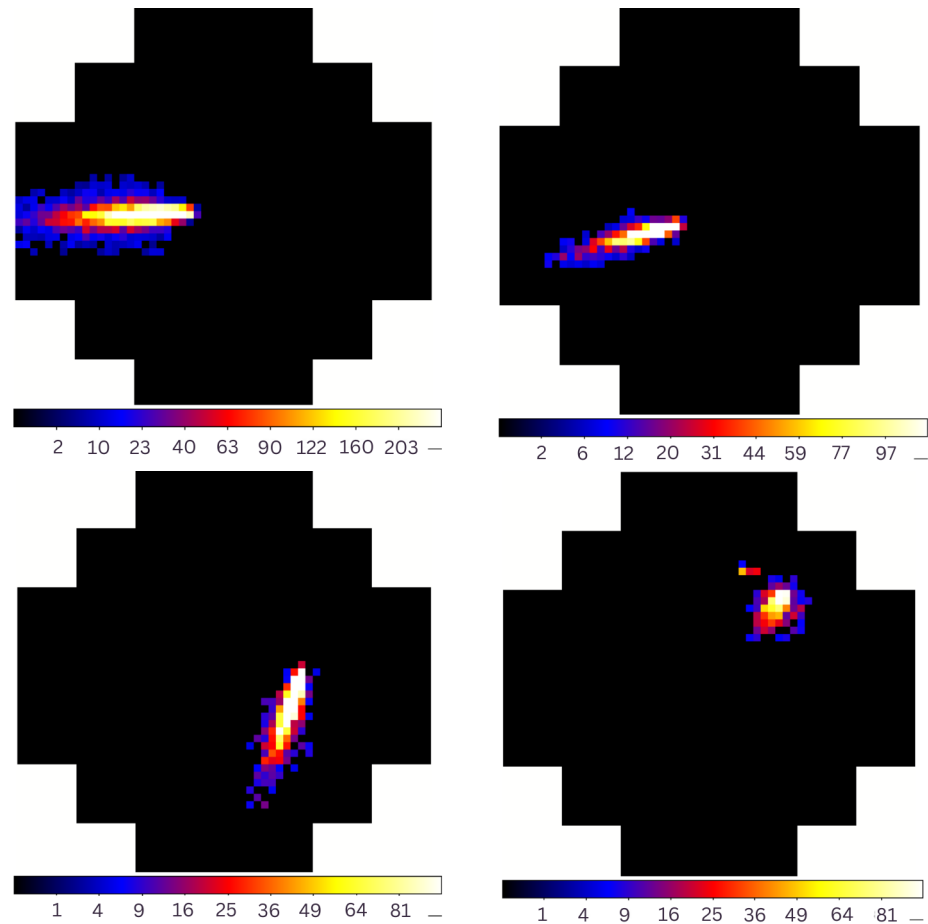


Figure 1. Cherenkov shower images simulated as observed by the ASTRI MiniArray telescopes, illustrating the morphological differences between different types of events (the contour matches that of the ASTRI telescope camera). The images include two gamma ray events (**Top**), a hadronic event with similarities to gammas (**Bottom Right**), and a distinct hadronic event (**Bottom Left**). These images showcase the often faint but discernible variations in the Cherenkov shower morphology.

1.3. Event Reconstruction

Cherenkov event reconstruction is a procedure used in IACTs to reconstruct the properties of VHE cosmic rays and gamma rays that have interacted with the Earth's atmosphere and refers to the process of analyzing the signals collected by the detectors in order to infer the properties of the incident particles. This typically involves several steps, including image cleaning, image parameterization, and stereo reconstruction.

Image cleaning is the process of removing unwanted or spurious signals from the detector image. This includes removing noise (both instrumental noise and night sky diffused background light), correcting for detector inefficiencies, and identifying and removing signals that are likely not related to Cherenkov radiation. Image parameterization consists of describing the cleaned image using a set of morphological indicators and allows us to quantify the properties of the particle interaction, such as the position, size, and orientation of the Cherenkov cone produced by the particle. Stereo reconstruction refers to the process of combining the information from multiple detectors in order to reconstruct the full 3D trajectory of the primary incident. This may involve triangulating the image of the shower as recorded by multiple detectors, or using other techniques to reconstruct the primary

trajectory. Finally, this whole set of information is used to reconstruct the energy, type, and direction of the primary particle (particle reconstruction). This can be achieved using Machine Learning methods, and the final output is the most probable characterization of the primary as either a photon or hadron (in the form of a global parameter, usually named *gammaness*), and an estimation of its energy.

The objective of this study is to identify the best reconstruction method for gamma/hadron discrimination and energy estimation using Machine Learning techniques, primarily those based on trees. We are not focusing on the reconstruction of the incoming particle direction, since the analytic stereo reconstruction method employed in ASTRI Mini-Array already produces optimal results in this regard.

While Machine Learning methods based on trees have been shown to be effective for gamma/hadron discrimination and energy estimation, it is possible that other Machine Learning techniques not considered in this study could perform better in certain scenarios. Examples of such techniques include Deep Learning algorithms (see [6] and references therein).

2. Data

2.1. Data Samples and Image Reconstruction

Monte Carlo (MC) simulations are a widely used computational technique for modeling the behavior of IACTs. In this context, MC simulations are used to model the interactions between very high-energy cosmic rays and gamma rays with the Earth's atmosphere.

This technique consists of generating a large number of random samples, or "trials", based on a set of probability distributions that describe the properties of the incoming particles, such as their energy, direction, and type. These samples are then used to simulate the shower of secondary particles that is produced when the incoming particles interact with the atmosphere. The resulting shower of particles is used to estimate the properties of the Cherenkov radiation that would be detected by the IACTs.

By simulating a statistically significant number of events, we can accurately assess and compare the performance of different Machine Learning methods for gamma/hadron segregation and energy estimation on the ASTRI Mini-Array Cherenkov telescopes. This allows us to optimize the performance of our Machine Learning techniques for the specific properties of the data generated by the IACTs.

The data used in this work (dubbed *ASTRI MA Prod2-Teide*) [7] are a large set consisting of 4×10^7 on-axis photon events and 2×10^9 proton simulations, assumed to fall in an area that includes the nine telescopes of the ASTRI Mini-Array on the Teide site. Heavy hadrons constitute a relatively small fraction of the overall hadronic events and have been excluded from the analysis. The simulation parameters have been defined according to the site settings (coordinates, height, atmosphere features, etc.) and to the telescopes' positions on the ground. The simulation consists of two subsets, according to the direction of the primary particles: 20° (zenith angle) South and 20° North. The energy range of the simulated primary particles is 0.1–330 TeV for photons and 0.1–600 TeV for protons, with a spectral slope of -1.5 . These spectral slopes differ from the actual measured spectra (which are steeper [7]) in order to allow us to reach sufficient statistics at the highest energies. This choice does not impact the analysis, which is performed with a differential approach.

The first step of the simulation tracks the development of each simulated shower along the atmosphere to the ground. To this end, we use the Corsika software package [8], whose output is the distribution of the Cherenkov light pool (photon position and direction) on the ground. In the second step, we perform the ray-tracing of each Cherenkov photon through the telescope optics to the focal plane camera and the camera response to the signal, producing an image of the shower as seen by the telescope. This is carried out using the *sim_telarray* package, a highly configurable software for the simulation of arrays of imaging atmospheric Cherenkov telescopes [9].

2.2. Data Format

The data levels for ASTRI project [10] were originally defined in compliance with those of the Cherenkov Telescope Array. ASTRI data levels of interest for the work presented here are:

- Level 2a (DL2a): array-wise merge data. For each event, they contain the single-telescope image parameters plus the stereoscopic shower event parameters;
- Level 2b (DL2b): array-wise fully reconstructed data. These also include the estimated event energy and its gammaness, i.e., the likelihood that the primary is a gamma event instead of a hadronic event.

The FITS data format [11] has been adopted for all ASTRI data levels. For our tests, we use FITS files in the official ASTRI format (level DL2a) as input data. In the output, we produce FITS files in the official ASTRI format (level DL2b).

In the ASTRI data reduction and scientific analysis software (A-SciSoft) package [10], which we use as a benchmark, data at level DL2a are reduced to level DL2b. A-SciSoft uses Random Forest methods both for gamma/hadron discrimination and for energy estimation.

3. Machine Learning Techniques

For a long time, scientists involved in ground-based gamma-ray astronomy have been investigating Machine Learning techniques. Bock et al. [12] started the initial efforts. Later, two operational ground-based observatories, H.E.S.S. [13–15] and MAGIC [16], showed the efficacy of tree-based multivariate classifiers for gamma hadron discrimination.

More recently, Sharma et al. [17] evaluated and compared five different Machine Learning methods to decide which of these methods is most suitable for gamma/hadron discrimination: the Random Forest method outperforms the Artificial Neural Network method in terms of signal strength and misclassification rate by almost 20%. Moreover, the Random Forest method has an advantage over perceptron-based methods in terms of computational time.

Next, we outline the models that we examined, providing a succinct summary of each.

3.1. Random Forest

Random Forest is a Machine Learning algorithm that is used for classification and regression tasks [18]. It uses multiple decision trees to make predictions.

A decision tree is a tree-like model in which an internal node represents a feature (or attribute) and the branches represent the decisions based on that feature. Each leaf node represents a class label. The decision tree algorithm works by recursively partitioning the data into smaller and smaller subsets, based on the value of the features.

In a Random Forest, multiple decision trees are trained on different subsets of the data and with different subsets of the features and the predictions made by each tree are then combined to make a final prediction. This is one of the main advantages of this method, since it reduces the chance of overfitting to any particular pattern in the data. Additionally, Random Forest can handle missing values and outliers in the data more effectively than a single decision tree.

3.2. Gradient Boosting

Gradient Boosting is an ensemble method that combines several weak learners, typically decision trees, to create a strong learner. The method works by iteratively adding new trees to the model, where each new tree is trained to correct the errors made by the previous tree. During the training process, the algorithm minimizes a loss function through gradient descent, using the negative gradient of the loss function to adjust the weights of the trees.

3.3. LightGBM

LightGBM (LGBM) is a gradient boosting framework that is designed to be efficient, scalable, and highly accurate. It is based on decision tree algorithms and uses a unique

approach called “Gradient-based One-Side Sampling” (GOSS) to reduce the number of data points needed for training, while still achieving high accuracy. LGBM uses a leaf-wise approach for growing trees, which enables it to handle large datasets with many features and avoid overfitting. In addition, it provides several advanced features such as early stopping, cross-validation, and regularization to further improve model performance.

3.4. Histogram-Based Gradient Boosting

Histogram-based Gradient Boosting (HGB) is a gradient boosting framework that is specifically designed to handle continuous numerical features with high cardinality. It is based on decision tree algorithms, but uses histograms to discretize numerical features, which greatly reduces the number of splitting points needed during training. This enables HGB to handle datasets with millions of features and samples while still achieving high accuracy. HGB also uses a novel algorithm for gradient computation, which reduces the computational cost of gradient boosting and improves its scalability. In addition, HGB provides several advanced features such as early stopping, cross-validation, and regularization to further improve model performance.

3.5. Extra Trees

Extra Trees are a fast and efficient Machine Learning algorithm, and they are well-suited to applications where the data are noisy or have a large number of features. They are similar to Random Forest in that they use multiple decision trees to make predictions. However, there are a few key differences between the two algorithms.

In Extra Trees, the decision trees are trained on random subsets of the data and with random subsets of the features, rather than using the entire dataset and all the features. This makes Extra Trees more random than Random Forest and reduces the chances of overfitting to any particular pattern in the data.

3.6. Extreme Gradient Boosting

Extreme gradient boosting (XGBoost or XGB) is a powerful Machine Learning algorithm and represents an improvement over traditional gradient boosting algorithms, which train decision trees one at a time, in a sequential manner. In contrast, XGBoost trains decision trees in parallel, using the boosting technique.

One of the main advantages of XGBoost is that it is very fast and efficient. It is implemented using parallel processing and cache-aware algorithms, which makes it much faster than traditional gradient boosting algorithms. Additionally, XGBoost has a number of advanced features that allow it to handle a variety of data types and complexity. For example, it has support for missing values, handling of imbalanced data, and handling of high-dimensional data.

4. Parameters

The reconstruction of Cherenkov events detected by ASTRI Mini-Array telescopes requires the use of several parameters (see Figure 2) to fully characterize their properties. In our study, we have utilized multiple sets of parameters to analyze and interpret the data. Each set of parameters includes a specific subset of the parameters listed below, commonly employed in Cherenkov analysis [2].

1. $\log_{10}(SIZE)$ is the decimal logarithm of the total content in photo-electrons of the cleaned image;
2. *WIDTH* is a Hillas parameter [2]: the minor half axis of the ellipse that best represents cleaned image;
3. *LENGTH* is a Hillas parameter: the major half axis of the ellipse that best represents the cleaned image;

4. *DENS* is defined as:

$$DENS = \log_{10} \frac{SIZE}{WIDTH \times LENGTH} \quad (1)$$

5. *CONC* is the concentration of the image defined as the ratio of the sum of the intensity in the two brightest pixels over *SIZE*;
6. *NUSEDTEL* is the number of telescopes used for the stereoscopic reconstruction;
7. *M3LONG* is the third momentum descriptor of the image elongation;
8. *TELIP* is the telescope impact defined as the distance between the stereo reconstructed core position and the given telescope;
9. *STMAXH* is the stereo reconstructed shower maximum height;
10. *LEAKAGE* is defined as the ratio of the sum of the pixel signals at the edge of the camera over *SIZE*;
11. *NUMCORE* is the number of core pixels in the image (see [10] for more details on the image cleaning method);
12. *NUMBOUNDARY* is the number of boundary pixels in the image (see [10] for more details on the image cleaning method);
13. *DELTA* is the angle between the Cherenkov photon emission direction and the projection of the photon arrival direction onto the plane perpendicular to the Cherenkov radiation cone;
14. *ASYM* is the distance from the highest pixel to the center of the image, projected onto the major axis;
15. *ECCE* is the eccentricity of the image defined as:

$$ECCE = \frac{\sqrt{(LENGTH^2 - WIDTH^2)}}{LENGTH} \quad (2)$$

16. *ELONG* is the elongation of the image defined as:

$$ELONG = 1 - \frac{WIDTH}{LENGTH} \quad (3)$$

17. *NUMISLAND* is the number of islands of the image. This parameter is used only for filtering.

	1	2	3	4	5	6	7	8
	LOG10 SIZE	WIDTH	LENGTH	DENS	CONC	NUSED TEL	M3 LONG	TELIP
GH SET 0								
GH SET 1								
GH SET 2								
GH SET 3								
EN SET 0								
EN SET 1								
	9	10	11	12	13	14	15	16
	ST MAXH	LEAK	NUM CORE	NUM BOUND	DELTA	ASYM	ECCE	ELONG
GH SET 0								
GH SET 1								
GH SET 2								
GH SET 3								
EN SET 0								
EN SET 1								

Figure 2. Sets of parameters used in our analysis. “GH SET” refers to sets used in gamma/hadron discrimination while “EN SET” refers to sets used in energy estimation. A full box indicates that the corresponding parameter is part of the set. “GH SET 3” and “EN SET 1” are highlighted since they are our top-performing sets, as we will discuss in Sections 5.5 and 6.3.

Figure 2 reports the different sets of parameters used in our analysis. Each set was tested against the benchmark combination (GH SET 0).

5. Tests on New Methods for Gamma/Hadron Separation

5.1. Tests on Gamma/Hadron Separation

Our primary objective is to achieve accurate discrimination between gamma-ray and hadron events. To accomplish this, we employ a discriminant parameter known as “gammaness”, which serves as a measure of the likelihood that an event originates from a gamma-ray source. The gamma/hadron separation code is based on Machine Learning methods. In input, it reads FITS files in the official ASTRI format (level DL2a). In output, it writes FITS files in the official ASTRI format (level DL2b). In our test, input data are ASTRI MA Prod2-Teide level DL2a and output data are ASTRI MA Prod2-Teide level DL2b. Our benchmark is the method based on Random Forest implemented in A-SciSoft and the following set of twelve discriminating parameters (from now on, these will be referred to as “GH SET 0”)

1. $\log_{10}(SIZE)$
2. *WIDTH*
3. *LENGTH*
4. *DENS*
5. *CONC*
6. *NUSEDTEL*
7. *M3LONG*
8. *TELIP*
9. *STMAXH*
10. *LEAKAGE*
11. *NUMCORE*
12. *NUMBOUNDARY*

Random Forest gamma/hadron filtering is performed as follows:

- *NUSEDTEL* > 1—specifies that the event must involve at least two telescopes in order to be considered for reconstruction;
- *STMAXH* > 0—the maximum height of the reconstructed shower must be greater than zero, indicating that the stereoscopic reconstruction is meaningful;
- *SIZE* > 50 phe—the size of the event must exceed a threshold of 50 to exclude events with low signal-to-noise ratios;
- *NUMISLAND* < 2—the number of isolated clusters of pixels belonging to a single image must be less than two;
- *NUMCORE* > 2—this is a lower limit on the number of pixels composing the image;
- *LEAKAGE* < 0.1—the fraction of light detected in the outermost region of the cameras must be less than 10%;
- *LENGTH* > 0, *WIDTH* > 0—null values in these two parameters indicate a degeneration of the image that makes it useless for reconstruction purposes.

By satisfying these conditions, reconstructed Cherenkov events are more likely to be of high quality and free from noise and background signals.

Several tests were performed on different sets, using different Machine Learning methods and different training hyperparameters tunings.

We employ the method of a mean decrease in impurity to calculate the importance of features. This method measures the reduction in the impurity criterion achieved through the splits made on a specific feature across all the trees in the ensemble. Gini impurity (see [19]) is a measure that indicates the likelihood of misclassification of new, randomly generated data if they were assigned a random class label based on the class distribution present in the dataset. By evaluating the mean decrease in impurity, we gain insights into the relative relevance of different features as they contribute to improving the overall purity of the decision trees. This analysis provides valuable information about the significance and

contribution of each feature in the learning process, shedding light on their importance for accurate predictions and classification. The Random Forest algorithm has a built-in feature importance evaluation algorithm. However, tree-based models have a strong tendency to overestimate the importance of continuous numerical features (the continuous feature provides more opportunity for the tree-based models to split the data in half). So, we also used the method of permutation feature importance.

The method of permutation feature importance is used in Machine Learning to evaluate the importance of different features in a dataset for predicting a particular outcome. It consists of training a model on a dataset and then randomly permuting the values of one feature at a time and re-evaluating the model performance on a validation set. The decrease in model performance when a particular feature is permuted is used as a measure of that feature's importance.

We train the model using 80% of the training sample as the training set and evaluating its performance on the remaining 20% of the data, which serves as the test set. This score will be our baseline. Then, we shuffle one feature at a time on the test set and feed the data to the new model to obtain a new score. If the shuffled feature is important, the model should suffer a drastic drop in its score. On the other hand, if the feature is not important, the model should not be impacted.

By means of these two feature importance methods, we evaluate more parameters set. We add a random feature to test our feature importances evaluation (see Figure 3). The random feature is expected to have importance equal to zero.

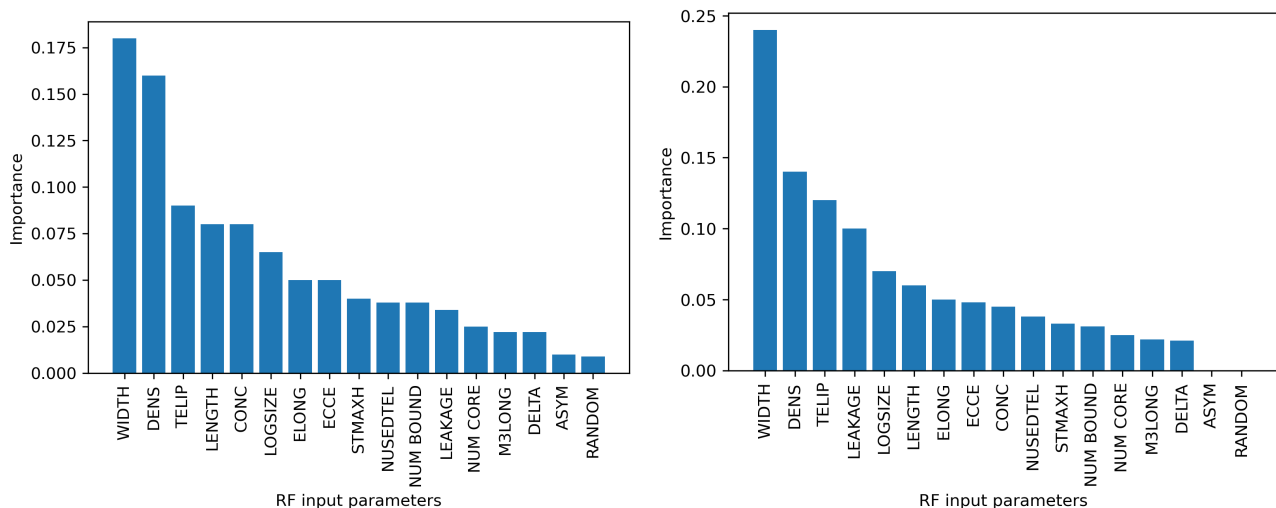


Figure 3. Feature importances for gamma hadron discrimination computed with two methods: (Left) Mean decrease in impurity method. (Right) Permutation method.

Different sets of features have been considered for testing, and all sets share the following common set of ten parameters. The benchmark parameters M3LONG and NUMCORE have been removed from the common set because they were found to have lower performance in terms of feature importance estimation (see Figure 3).

1. $\log_{10}(SIZE)$
2. *WIDTH*
3. *LENGTH*
4. *DENS*
5. *CONC*
6. *NUSEDTEL*
7. *TELIP*
8. *STMAXH*
9. *LEAKAGE*
10. *NUMBOUNDARY*

“GH SET 1” adds two parameters for a total of twelve:

1. *DELTA*
2. *ASYM*

“GH SET 2” adds five parameters for a total of fifteen:

1. *NUMCORE*
2. *DELTA*
3. *ASYM*
4. *ECCE*
5. *ELONG*

“GH SET 3” adds four parameters for a total of fourteen:

1. *M3LONG*
2. *NUMCORE*
3. *DELTA*
4. *ASYM*

All the parameters are level DL2a input parameters or a combination of them (e.g., *ECCE* and *ELONG*, defined previously). We have used the same training samples for our tests and for the benchmark.

Pruning is performed on the samples used for training with the aim of balancing the number of hadron and gamma images in each bin of $\log_{10}(SIZE)$. Specifically, the simulated images were divided into 100 logarithmic bins based on their sizes, and within each bin, the number of hadron images and the number of gamma images were made equal by pruning images in excess of the lower count. For our experiments, we had a total of 136,498 gamma images and 126,645 hadron images. After pruning, we removed 14,204 gamma images and 4351 hadron images. The remaining hadron and gamma images were balanced, with 122,294 of each remaining. By performing this pruning step, we ensured that the training data were less likely to be skewed towards one type of particle or the other, thereby reducing the risk of poor performance on the test data.

Hyperparameter tuning was performed by means of a randomized search that implements a “fit” and a “score” method on the following hyperparameters:

- *n estimators* is the number of decision trees in the Random Forest;
- *max features* is the maximum number of features that are considered at each split in the decision tree;
- *max depth* is the maximum depth of the decision trees in the Random Forest;
- *min samples split* is the minimum number of samples required to split an internal node in the decision tree;
- *minimum samples leaf* is the minimum number of samples required to be at a leaf node in the decision tree;
- *bootstrap*: in Machine Learning, the process of bootstrapping consists of replacing random samples of the training data with replacement from the original dataset to create multiple different subsets of the data. Each subset is used to train a separate model, and the results are combined to produce a final model. In the context of Random Forests, bootstrapping is used to create multiple decision trees, each trained on a different subset of the training data. The idea is that by training on different subsets of the data, the decision trees will differ from each other in ways that help to reduce overfitting and improve generalization performance. The bootstrap hyperparameter is a Boolean value that indicates whether or not to use bootstrapping when building the decision trees. When bootstrap is set to True, bootstrapping is used to create the subsets of training data used to train each decision tree. When bootstrap is set to False, each decision tree is trained on the entire training set.

Our best results, achieved for the Random Forest model and for all the sets of features considered, are as follows:

- n estimators = 100. In our case, 100 decision trees were used, which was not the best choice for achieving the highest performance. However, we chose to keep the network simple to balance performance and computational complexity. We noted that there was little improvement in performance when using hundreds of trees, but using thousands of trees would result in a very heavy and slow network;
- max features = square root of the number of features used during fit;
- max depth = 20
- min samples split = 2
- min samples leaf = 2
- $bootstrap$ = True

5.2. Evaluating More Machine Learning Models

We have implemented and tested several Machine Learning methods using a fast test run with Lazypredict [20] on 42 different models in order to compare them to the benchmark method based on Random Forest.

Lazypredict is a Python library used for quickly evaluating and comparing the performance of multiple Machine Learning models. It allows users to easily train and test a large number of models with minimal coding, providing a convenient way to quickly assess the suitability of different models for a particular task.

Based on these preliminary test results, it appears that Gradient Boosting, LGBM Regressor, and Histogram Gradient Boosting show promising performance for gamma/hadron separation. However, it is important to note that these results are obtained from a limited sample and should be considered as approximate (refer to Table 1). Our primary objective is to improve the Q factor (refer to Section 5.4 for its definition), which heavily relies on impostor rejection. In more comprehensive tests, we have found that methods based on trees and XGB exhibit better performance in achieving this objective.

Table 1. Evaluation metrics for different Machine Learning models for gamma/hadron separation. Only the first 10 best-performing models out of 42 are shown. It is worth noting that this is a fast test run on 10,000 samples on the benchmark set of parameters (“GH SET 0”) and the results are only approximate. R-squared measures the proportion of variance in the dependent variable explained by the independent variables, while adjusted R-squared takes into account the number of independent variables in the model. Root Mean Squared Error (RMSE) measures the average deviation of predicted values from actual values.

Model	Adj R-Squared	R-Squared	RMSE	Training Time
Grad Boosting	0.30	0.31	0.41	2.09
LGBM	0.29	0.30	0.41	0.68
HGB	0.29	0.29	0.42	0.97
Random Forest	0.28	0.28	0.42	7.16
Extra Trees	0.27	0.28	0.42	0.84
MLP	0.25	0.26	0.43	2.39
XGB	0.23	0.23	0.43	0.38
SVR	0.22	0.22	0.44	3.55
Ada Boost	0.21	0.22	0.44	0.14
Bagging	0.20	0.20	0.44	0.47

5.3. Ensemble Learning Methods

Ensemble Learning is a general meta approach to Machine Learning that involves training multiple models and combining their predictions to make a final prediction. The goal of Ensemble Learning is to improve the accuracy, stability, and robustness of the final prediction by using the strengths of multiple models.

There are several types of Ensemble Learning methods, including Stacking. Stacking is a type of Ensemble Learning in which multiple models are trained on the same data

and their predictions are combined using a meta-model. The goal of stacking is to use the strengths of different models to make more accurate predictions.

We tested Stacking Ensemble models composed as follows (see Figure 4 and Table 2):

- GH ENSEMBLE 0: 1.00 Random Forest (same model as the benchmark).
- GH ENSEMBLE 1: 0.65 XGB, 0.21 Extra Trees, 0.14 Random Forest.
- GH ENSEMBLE 2: 0.60 Extra Trees, 0.40 Random Forest.
- GH ENSEMBLE 3: 0.42 Extra Trees, 0.30 XGB, 0.28 Random Forest.

Table 2. The table displays the percentage of each Machine Learning model in four different Stacking Ensembles consisting of Random Forest (RF), XGB, and Extra Trees (ET) models.

GH Ensemble	RF	XGB	ET
0	1.00	0.00	0.00
1	0.14	0.65	0.21
2	0.40	0.00	0.60
3	0.28	0.30	0.42

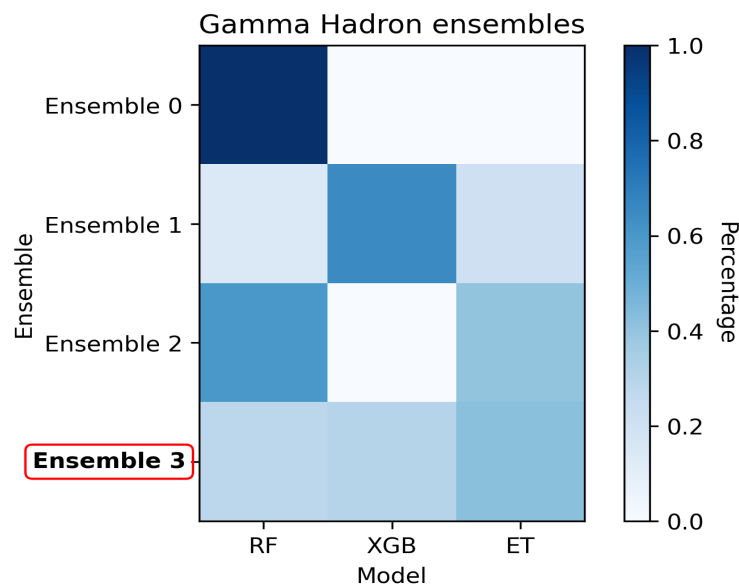


Figure 4. Heatmap of gamma/hadron separation Stacking Ensembles. The figure displays the percentage of each Machine Learning model in four different Stacking Ensembles consisting of Random Forest (RF), XGB, and Extra Trees (ET) models. The ensembles were chosen through a series of tests aimed at optimizing the Quality factor of the models. Our best model is Ensemble 3. The heatmap visualization provides an easy-to-read overview of the model composition of each ensemble, with darker shades of blue indicating a higher percentage of the model.

The ensembles mentioned above were chosen through a series of tests aimed at optimizing the quality factor of the models (see next paragraph). In order to identify the best combination of models, we conducted several experiments using different ensembles and evaluated their performance using the quality factor metric. Based on the results of these experiments, we selected the four ensembles mentioned above. ENSEMBLE 0 comprises a single model, while the other ensembles are composed of a combination of XGB, Extra Trees, and Random Forest models, with varying weights assigned to each model. These ensembles were found to produce the best results in terms of the quality factor and were thus chosen for further analysis.

5.4. Quality Factor

The quality (Q) factor is a metric used to evaluate the performance of a classification method in distinguishing between gamma (signal) and hadron (background) events. It is defined as

$$Q = \frac{\epsilon_{\gamma}}{\sqrt{\epsilon_{bkg}}}, \quad (4)$$

where ϵ_{γ} is the γ acceptance rate while ϵ_{bkg} represents the hadron acceptance rate. The γ acceptance rate is defined as the correctly classified γ events out of the total number of γ events. The hadron acceptance rate is defined as the ratio of proton events which behave like γ events after the classification (background contamination). We also define the hadron rejection rate as the number of hadron events which have been correctly classified out of the total number of hadron events:

$$1 - \epsilon_h = \epsilon_{bkg} \quad (5)$$

The higher the Q factor, the better the discrimination between gamma and hadron events, and therefore, the better the performance of the method.

5.5. Results for Gamma/Hadron Separation

To identify the most effective approach for gamma/hadron segregation, we assessed several supervised Machine Learning methods, including the Random Forest method, Extra Trees method, and Extreme Gradient Boosting (XGB), and compared them to the A-SciSoft benchmark.

To optimize the performance of each method in the ensemble, we conducted extensive experiments involving multiple trials and cross-validation tests. Based on these tests, we identified the most effective models for our purposes.

Our best-performing models are:

- GH SET 1-GH ENS 0: a pure Random Forest model on set of parameters “GH SET 1”;
- GH SET 3-GH ENS 1: a Stacking Ensemble model on set of parameters “GH SET 3” made of 0.65 XGB Regressor, 0.21 Extra Trees Regressor and 0.14 Random Forest;
- GH SET 3-GH ENS 2: a Stacking Ensemble model on set of parameters “GH SET 3” made of 0.40 Random Forest and 0.60 Extra Trees;
- GH SET 3-GH ENS 3: a Stacking Ensemble model on set of parameters “GH SET 3” made of 0.28 Random Forest, 0.42 Extra Trees and 0.30 XGB

Results for the true identification of γ , hadron rejection, and quality factors are shown in the following Table 3 and in Figures 5–8. Our best threshold for gammaness is 0.8. This threshold was determined through extensive testing and optimization to provide the optimal balance between gamma identification and hadron rejection.

Table 3. True identification of gammas, hadron rejection and quality factors. Our best threshold for gammaness is 0.8, which was used to classify events as gamma-like or hadron-like. This threshold was determined through extensive testing and optimization to provide the optimal balance between gamma identification and hadron rejection. AUC is the Area Under the Curve, a commonly used metric in classification tasks. AUC represents the overall performance of a classification model by measuring the area under the receiver operating characteristic (ROC) curve (see Section 5.6).

GH Model (thres = 0.8)	% True Id γ	% h Rejection	Quality Factor	AUC
A-SciSoft	67.63	98.70	5.943	0.83
SET 1-ENS 0	65.01	98.98	6.431	0.82
SET 3-ENS 1	40.51	99.54	5.998	0.75
SET 3-ENS 2	65.38	99.03	6.639	0.75
SET 3-ENS 3	64.46	99.08	6.736	0.83

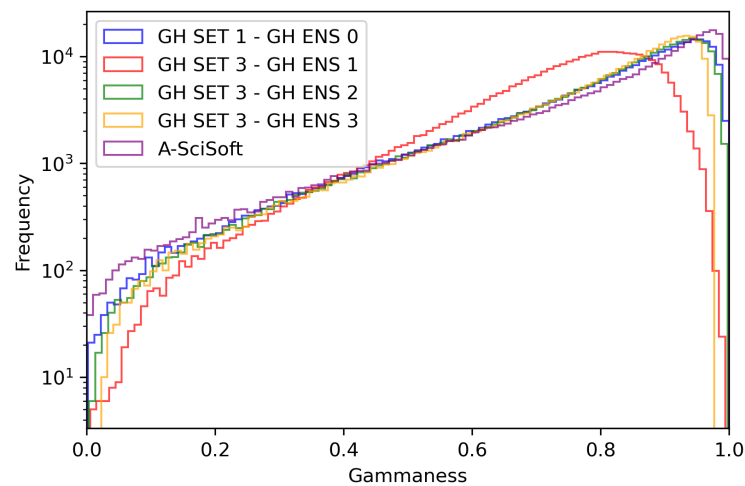


Figure 5. Classified γ (based on the gammaness classification parameter) out of the total number of γ events for different models and sets of parameters.

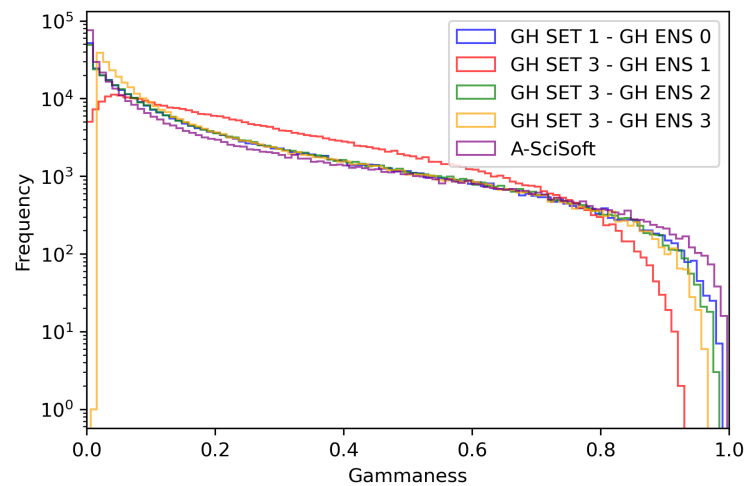


Figure 6. Classified hadron events (based on the gammaness classification parameter) out of the total number of hadron events for different models and sets of parameters.

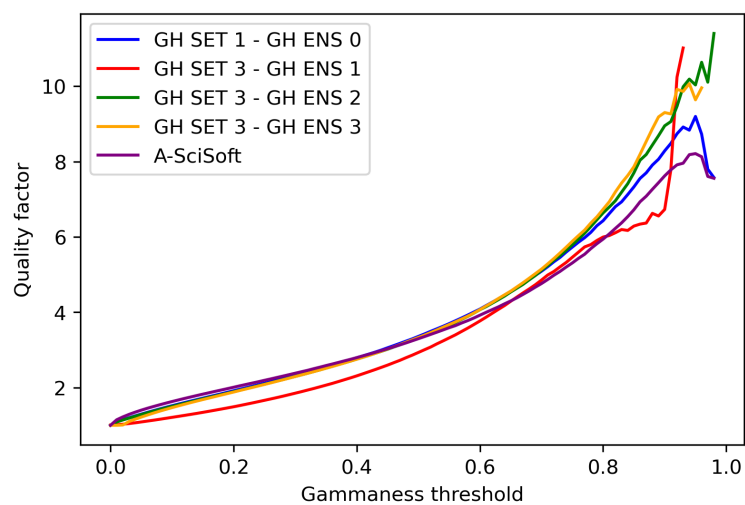


Figure 7. Quality factor for different gammaness threshold and for different models and sets of parameters.

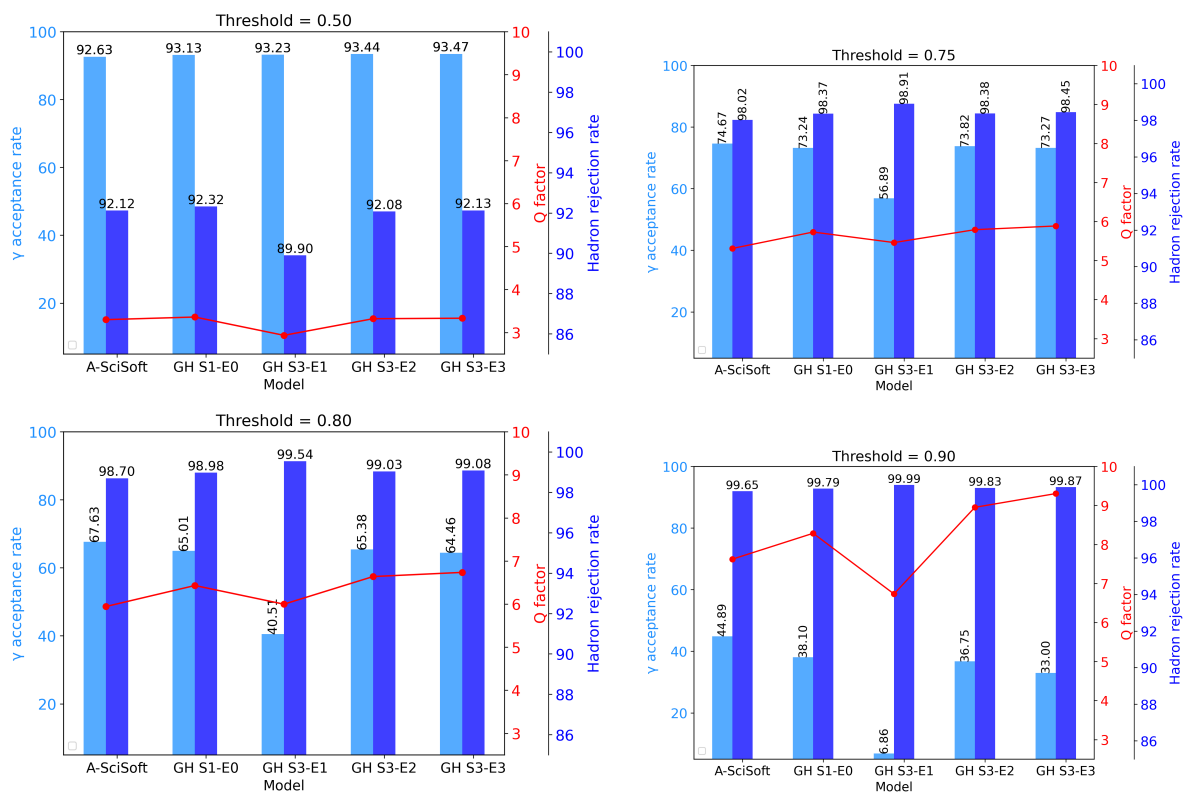


Figure 8. Bar plots showing the performance of five models in terms of true identification rate (γ acceptance rate, in dodgerblue) and rejection rate of impostor matches (Hadron rejection rate, in blue), along with the quality factor (in red) for each model for four different gammaness thresholds (0.5, 0.75, 0.8 and 0.9).

5.6. Receiver Operating Characteristic (ROC) Curve

While the Q factor is considered a more important parameter, the receiver operating characteristic (ROC) curve is also commonly used to evaluate the performance of gamma/hadron discrimination methods. The ROC curve (see Figure 9) shows the trade-off between correctly identifying gamma rays (true positives) and misidentifying hadrons as gamma rays (false positives), and is a useful tool for comparing different methods. However, the Q factor is preferred because it is a proxy for the significance of the detection of gamma-ray signal over the overwhelming hadronic background in Cherenkov observations. Therefore, in gamma/hadron discrimination for IACTs, the Q factor is a crucial parameter that must be carefully considered in the development and evaluation of Machine Learning methods. AUC values are reported in Table 3.

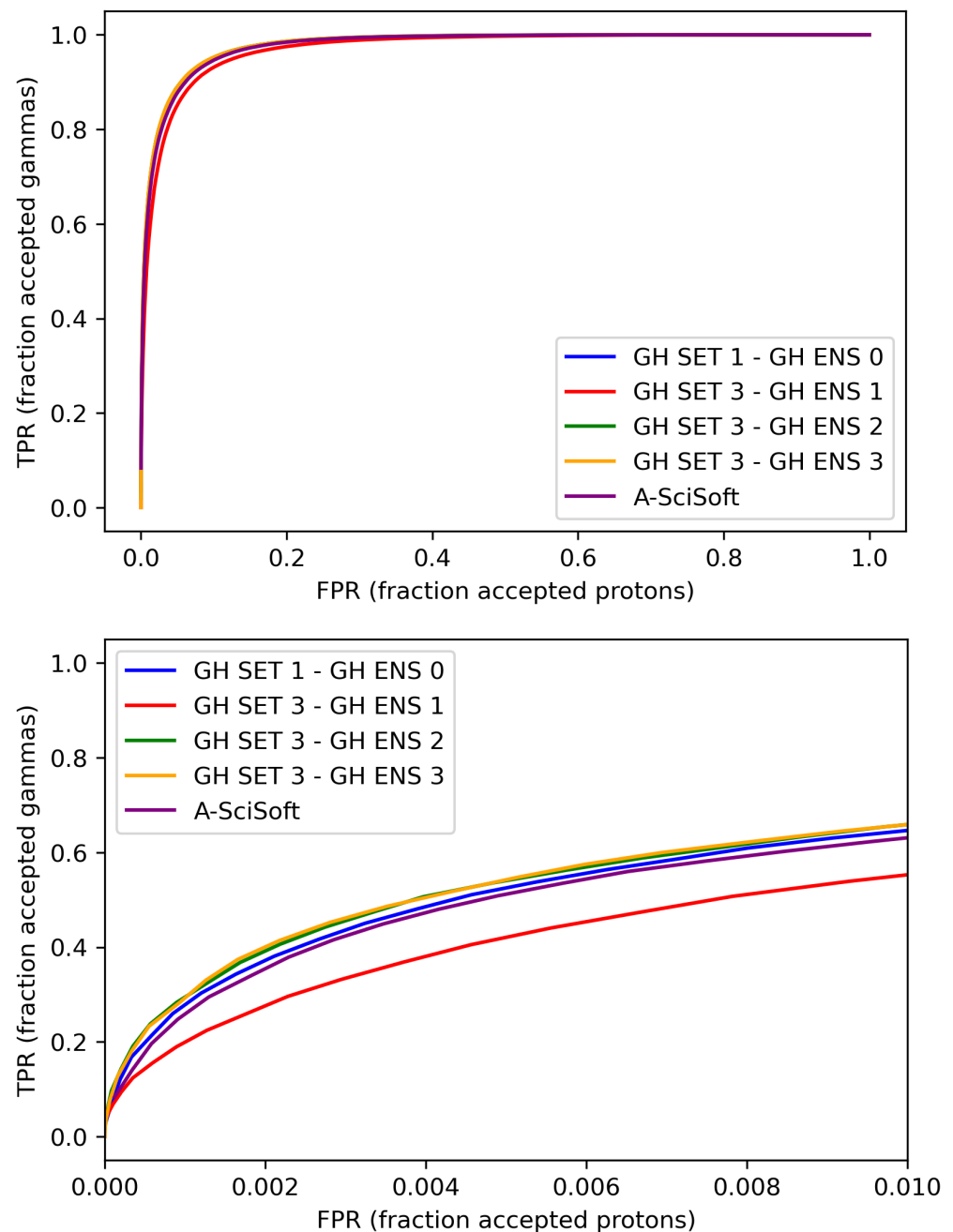


Figure 9. Comparison of receiver operating characteristic (ROC) curves for gamma/hadron discrimination methods. The first plot shows the full ROC curve, with the x-axis ranging from 0 to 1, while the second plot is a zoomed-in view of the false positive rate (x-axis ranging from 0 to 0.06) to better visualize the performance of the classifiers at low false-positive rates.

6. Tests on New Methods for Energy Reconstruction

6.1. Tests on Energy Reconstruction

Our main focus is to address the bias, as there is significant room for improvement in this aspect. Bias refers to the systematic deviation or error in the predictions made by our models. Instead of solely prioritizing the optimization of energy resolution, as even significant differences in the ensembles had negligible impact on it, our emphasis is on enhancing the ensembles' performance in terms of reducing bias. We achieve this by carefully selecting the ensembles and their corresponding weights to specifically optimize and minimize bias, as it plays a crucial role in achieving accurate results.

The energy reconstruction code is based on Machine Learning methods for energy regression. In our tests, input data are DL2a files from ASTRI MA Prod2-Teide MC production and output data are DL2b.

Our benchmark is the method implemented in A-SciSoft based on Random Forest, which makes use as default of the following eight discriminating parameters (“EN SET 0” in Figure 2):

1. $\log_{10}(SIZE)$
2. *WIDTH*
3. *LENGTH*
4. *DENS*
5. *CONC*
6. *TELIP*
7. *STMAXH*
8. *LEAKAGE*

Random Forest reconstruction filtering for training quality is performed by applying the same conditions as for the gamma/hadron separation.

By means of the two feature importance methods already described in Section 5.1, we evaluate more parameters set. In order to test our feature importance evaluation (see Figure 10) we add a random control feature whose importance is expected to be equal to zero.

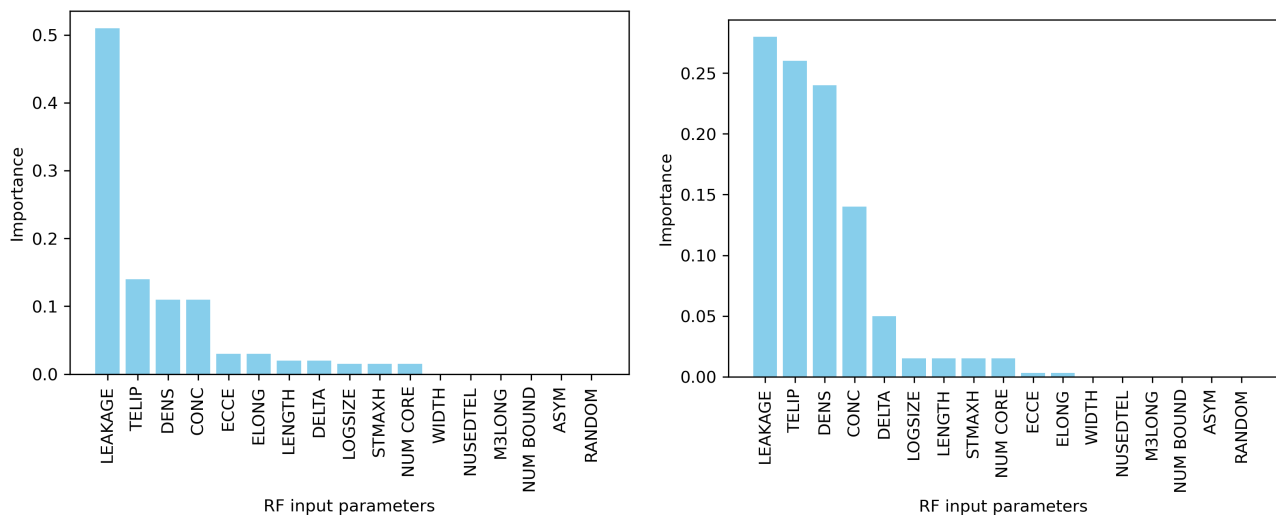


Figure 10. Feature importances for energy reconstruction computed with two methods: **(Right)** Mean decrease in impurity method. **(Left)** Permutation method.

A second set of eleven parameters (from now dubbed as “EN SET 1”) has been considered for tests:

1. $\log_{10}(SIZE)$
2. *WIDTH*
3. *LENGTH*
4. *DENS*
5. *CONC*
6. *TELIP*
7. *STMAXH*
8. *LEAKAGE*
9. *DELTA*
10. *ECCE*
11. *ELONG*

For our tests and for the benchmark, the same training samples have been used. All sets of parameters both for energy estimation and for gamma hadron discrimination are summarized in Figure 2.

6.2. More Machine Learning Methods Evaluation

Several Machine Learning methods were implemented and tested using a fast test run with Lazypredict [20] on 42 different models, as already made for the gamma/hadron separation.

According to the results of this test, the best performance for energy reconstruction should be achieved using Extreme Gradient Boosting (XGB) [21], Extra Trees and Random Forest (see Table 4).

Table 4. Evaluation metrics for different Machine Learning models for energy estimation. Only the first 10 models out of 42 are shown. It is worth noticing that this is a fast test run on 10,000 samples on the benchmark set of parameters (“EN SET 0”) and the results are only approximate. R-squared measures the proportion of variance in the dependent variable explained by the independent variables, while adjusted R-squared takes into account the number of independent variables in the model. Root Mean Squared Error (RMSE) measures the average deviation of predicted values from actual values.

Model	Adj R-Squared	R-Squared	RMSE	Training Time
XGB	0.33	0.33	0.02	0.83
Extra Trees	0.30	0.31	0.02	0.74
Random Forest	0.30	0.30	0.02	7.07
NuSVR	0.29	0.30	0.02	21.90
Grad Boost	0.28	0.28	0.02	1.67
HGB	0.27	0.28	0.02	1.76
Bagging	0.24	0.24	0.02	0.46
KNeigh	0.15	0.15	0.02	0.05
Tran Target	0.11	0.12	0.02	0.03
Linear	0.10	0.12	0.02	0.01

6.3. Ensemble Learning Methods

As already completed for gamma/hadron separation, we applied Stacking Ensemble Learning on different combinations of models. Stacking Ensemble Learning is a technique in which multiple models are trained on the same dataset and their predictions are combined using a meta-model. The objective of stacking is to leverage the individual strengths of different models to produce more accurate predictions. As such, stacking is an effective way to improve the performance of Machine Learning models.

We tested Stacking Ensemble models composed as follows (see Figure 11 and Table 5):

- EN ENSEMBLE 0: 0.65 XGB, 0.21 Extra Trees, 0.14 Random Forest.
- EN ENSEMBLE 1: 0.80 XGB, 0.16 HGB, 0.04 Extra Trees.
- EN ENSEMBLE 2: 0.34 XGB, 0.33 Extra Trees, 0.33 HGB.
- EN ENSEMBLE 3: 0.55 XGB, 0.45 Extra Trees.
- EN ENSEMBLE 4: 0.45 XGB, 0.275 Random Forest, 0.275 Extra Trees.

Table 5. The table displays the percentage of each Machine Learning model in five different Stacking Ensembles consisting of XGB, Extra Trees (ET), Random Forest (RF), and HGB models.

EN Ensemble	XGB	ET	RF	HGB
0	0.65	0.21	0.14	0.00
1	0.80	0.04	0.00	0.16
2	0.34	0.33	0.00	0.33
3	0.55	0.45	0.00	0.00
4	0.45	0.275	0.275	0.00

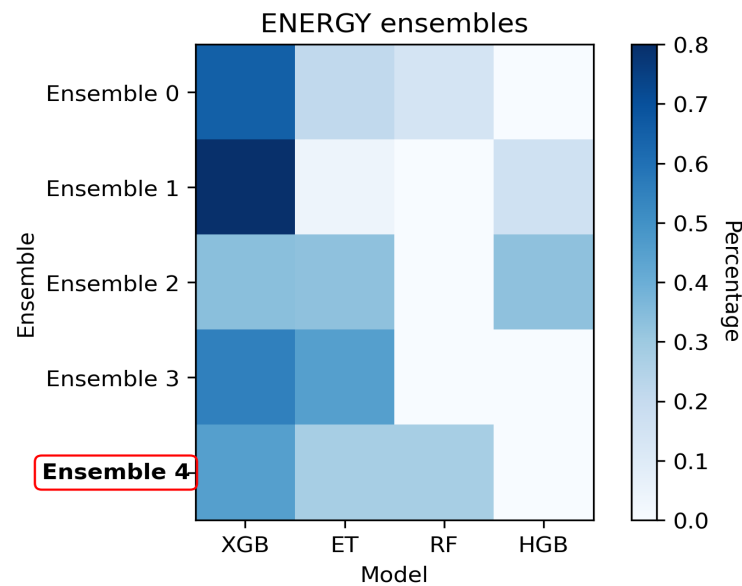


Figure 11. Heatmap of energy Stacking Ensembles. The figure presents the percentage of each Machine Learning model in five different Stacking Ensembles consisting of XGB, Extra Trees (ET), Random Forest (RF), and HGB models. The ensembles and their weights were selected to optimize the bias. Our best model is Ensemble 4. The heatmap visualization provides a clear representation of the relative contribution of each model to each ensemble, with darker shades of blue indicating a higher percentage of the model.

These combinations were derived using a general method for creating Stacking Ensemble models, which involves training multiple base models on the same dataset and then combining their predictions using a meta-model. The specific weights assigned to each base model in the ensembles were determined through a process of hyperparameter tuning and cross-validation, which seeks to identify the optimal combination of models and weights for the specific task at hand.

Our primary objective was to improve the bias of the ensembles, as there was a substantial scope for enhancement in this area. Therefore, we prioritized the optimization of the ensembles based on bias rather than energy resolution. Consequently, the ensembles and their weights were selected to optimize the bias and not energy resolution.

In our tests, all these methods computed on “EN SET 1” show better performance than any method made using a single bagging or boosting ensemble.

6.4. Energy Resolution

Energy resolution is computed as follows. We divided our data set in 16 log bins between $E_{rec} = 10^{-0.7}$ and $E_{rec} = 10^{2.5}$ TeV, according to the reconstructed energy. In each bin, we built the $(E_{rec} - E_{true})/E_{true}$ distribution, where E_{rec} is the reconstructed energy and E_{true} is the true (simulated) energy of the events. The distribution has an almost Gaussian trend within the interval $[MEAN - k_1 \times RMS, MEAN + k_2 \times RMS]$ where $MEAN$ is the average value of the distribution and RMS is the root mean square of the distribution. However, to take into account non-Gaussianity, we chose different values for k_1 and k_2 : $k_1 = 1.5$ and $k_2 = 0.75$. With a Gaussian fit in the range considered, we extracted the values:

- Energy resolution = sigma of the Gaussian fit;
- Energy bias = mean of the Gaussian fit.

The energy resolution curve is therefore given by the sigma obtained with the Gaussian fits of the distributions for each reconstructed energy bin.

Both for the default method implemented in A-SciSoft and for the new tested methods, energy resolution is computed after applying a set of cuts that includes multiplicity cut, gammaness cut, and theta2 cut. A different set of cuts is performed in each energy bin.

It is worth noting that the set of cuts applied to the new methods is the same as the one optimized for the sample fully reconstructed by A-SciSoft default methods [7]. The reason behind this approach is that the distribution of gammaness is similar for both the default and new methods. As a result, the comparison between the default method and the new methods can be considered conservative, since the cuts were not optimized for the new methods.

In Figure 12, we present the results of our analysis, comparing the performance of different Machine Learning methods and two different sets of parameters against the A-SciSoft benchmark. When considering energy resolution alone, our top-performing method is XGB. However, it is crucial to note that this method exhibits a significant negative bias, as we will discuss in detail in the following section (see Figure 13).

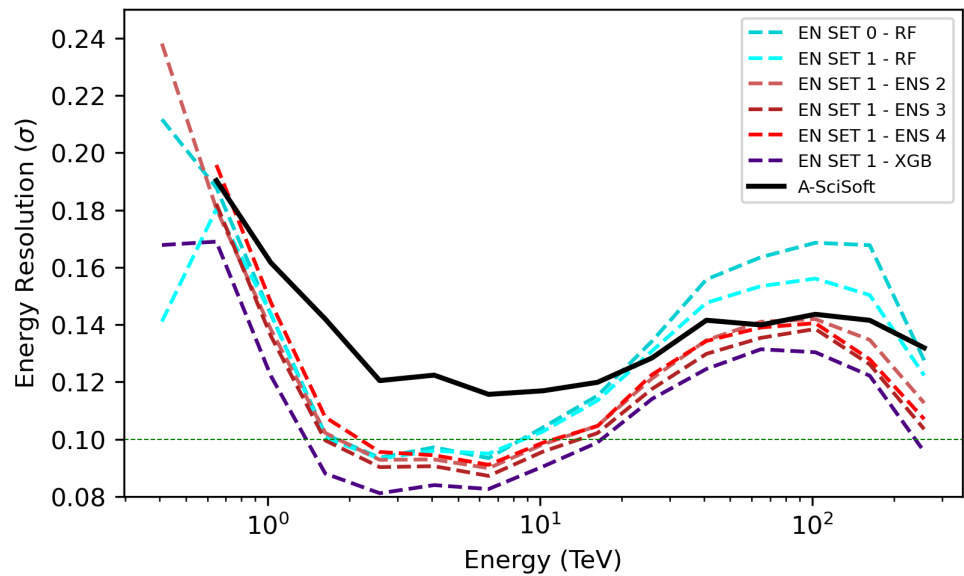


Figure 12. Energy resolution for different Machine Learning methods and two different sets of parameters against the benchmark (A-SciSoft).

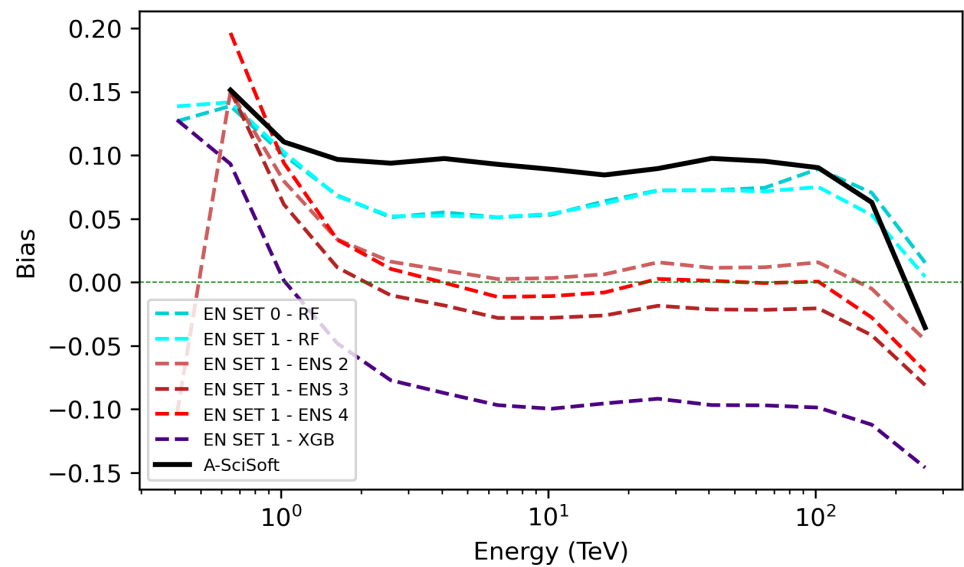


Figure 13. Energy bias for different Machine Learning methods and two different sets of parameters against the benchmark (A-SciSoft).

6.5. Energy Bias

Performing the tests on ensemble methods, we noticed that the energy bias was always positive when using Random Forest or Extra Trees methods, while it turned negative when the XGB method was used. As a result, an ensemble composed of methods based on trees combined with XGB can be tuned to obtain a value of the bias very close to zero in the approximate range 2–100 TeV. While the specific combination of methods used in this study may have been optimized for the particular dataset and range of energies analyzed, the general approach of using a combination of tree-based methods with XGB is a widely-used technique in Ensemble Learning. Therefore, it is likely that similar combinations of methods could be effective for other datasets and ranges of energies. Among the Stacking Ensemble considered, the best performing method is ENSEMBLE 4, which is made of 0.45 XGB, 0.275 Random Forest and 0.275 Extra Trees. In Figure 13, we show our results for different Machine Learning methods and two different sets of parameters against the benchmark (A-SciSoft).

7. Results

Through the Monte Carlo simulation of Cherenkov events as seen by the ASTRI Mini-Array, we assessed and compared the performance of several supervised Machine Learning methods for gamma/hadron segregation and energy estimation. The methods investigated included the Random Forest method, Extra Trees method, and Extreme Gradient Boosting (XGB).

For gamma/hadron segregation, our findings indicate that the most sensitive technique is “GH SET 3-GH ENS 3”, a Stacking Ensemble Method composed of 42% Extra Trees, 28% Random Forest, and 30% XGB on the set of parameters “GH SET 3” (see Figure 2). Comparatively, this ensemble method delivered better performance than each of the individual methods. The optimal weightings for this ensemble were determined through a series of trials and cross-validation tests, ensuring maximum performance for gamma/hadron separation.

Regarding energy estimation, our results show that the most effective technique is “EN SET 1-EN ENS 4”, a Stacking Ensemble Method composed of 45% XGB, 27.5% Extra Trees, and 27.5% Random Forest on set of parameters “EN SET 1” (see Figure 2). Similar to the gamma/hadron segregation ensemble, the optimal weightings for the energy estimation ensemble were derived from extensive testing and fine-tuning, leading to improved energy estimation performance compared to any single method alone.

To further evaluate the performance of the ensemble methods, we compared them to the more classical Random Forest method as implemented in A-SciSoft. Our results demonstrate that the Stacking Ensemble methods proposed in this study outperform the Random Forest method for the ASTRI Mini-Array event reconstruction both for the energy resolution and bias and for the gamma/hadron discrimination.

In summary, our results indicate that the Stacking Ensemble methods identified in this study deliver superior performance for gamma/hadron segregation and energy estimation compared to individual and classical methods. These findings suggest the potential for these ensemble methods to enhance the analysis of data collected by the ASTRI Mini-Array, although further validation with real observational data is needed to confirm their effectiveness in practice.

8. Computational Costs

We used a Intel® Xeon® CPU @ 2.20 GHz and a AMD EPYC 7B12 processor both with 13 GB dedicated Ram on Google Colaboratory to train all the models on our datasets and export the best parameters with the best scores obtained during training. In the gamma-hadron discrimination case, training our models on approximately 10^5 events takes approximately 120 s. In the energy reconstruction case, training on approximately 10^5 events takes approximately 750 s. The notable difference in computational time can be attributed to the fact that, in the ASTRI file system, the simulated values of energy are stored in a separate file, requiring an additional search operation to retrieve and associate

the energy values with the corresponding events. The libraries used for our models in Google Colaboratory are capable of deploying code to both CPUs and GPUs. Tests on GPUs NVIDIA© Tesla© T4 and NVIDIA© Tesla© K80, both with dedicated 12 GB GDDR5 VRAM, showed no noticeable improvement in computing time.

9. High-Level Performance Comparison

In order to provide a final high-level comparison between the new methods considered in this work and those implemented in A-SciSoft, we derived the main performance metrics by means of the standard routine implemented in the ASTRI pipeline [7]. For this high-level comparison, we considered the DL2b samples obtained with the best new methods investigated in this work (namely “GH SET 3-GH ENS 3” for the gamma/hadron separation and “EN SET 1-EN ENS 4” for the energy estimation) and those produced with A-SciSoft.

The main performance metrics that we considered were the differential flux sensitivity (for an exposure time of 50 h), energy resolution, and angular resolution. All of these quantities were comprehensively evaluated from the DL2b data achieved with the best new Machine Learning methods and with A-SciSoft for on-axis source observations in the (reconstructed) energy range between $10^{-0.5} \simeq 0.3$ TeV and $10^{2.5} \simeq 300$ TeV.

The background (proton) and gamma-ray events of the different Level 2b (DL2b) samples were reweighed according to experimental measurements of their spectra, following the same procedure adopted in [22]. In particular, the gamma-ray events were reweighed considering the Crab Nebula spectrum, as measured by the HEGRA Collaboration [23]. No electron background component was considered in this high-level comparison because for the typical energies for which the ASTRI Mini-Array is sensitive ($E > \sim 1$ TeV), it does not contribute significantly to the irreducible gamma-like background (which was by far dominated by proton events).

For each DL2b sample, separately, the final analysis cuts were based on the background rejection, shower arrival direction, and event multiplicity parameters. They were defined, in each considered energy bin, by optimizing the sensitivity for a 50 h exposure time. Then, five standard deviations (5σ , with σ defined as shown in Equation (17) of [24]) were required for a detection in each energy bin and off-axis bin, considering the same exposure time (as in the cut optimization procedure) and a ratio of the off-source to on-source exposure equal to 5. In the analysis of Cherenkov data, the gamma-ray signal within the “on-source” region of interest must always be compared with suitable “off-source” background control regions so that the fraction of irreducible background events surviving all analysis cuts in the signal region can be properly estimated and subtracted to determine the final excesses from the gamma-ray source. Typically, for observations of point-like gamma sources, Cherenkov data are taken in such a way that it is always possible to define multiple background control regions (five by default), which all have the same acceptance as the signal region. In addition, the signal excess was required to be larger than ten and at least five times the expected systematic uncertainty in the background estimation (assumed to be $\sim 1\%$). All of these assumptions are commonly adopted in the IACT community (see e.g., [22]) and allow us to derive performance results under coherent analysis conditions.

Figures 14–16 show the on-axis point-like source differential sensitivity (in 50 h), angular resolution, and energy resolution, respectively, achieved with A-SciSoft (blue points) and the best Machine Learning methods investigated in this work (orange points). The ratios between the results of the two methods are also shown.

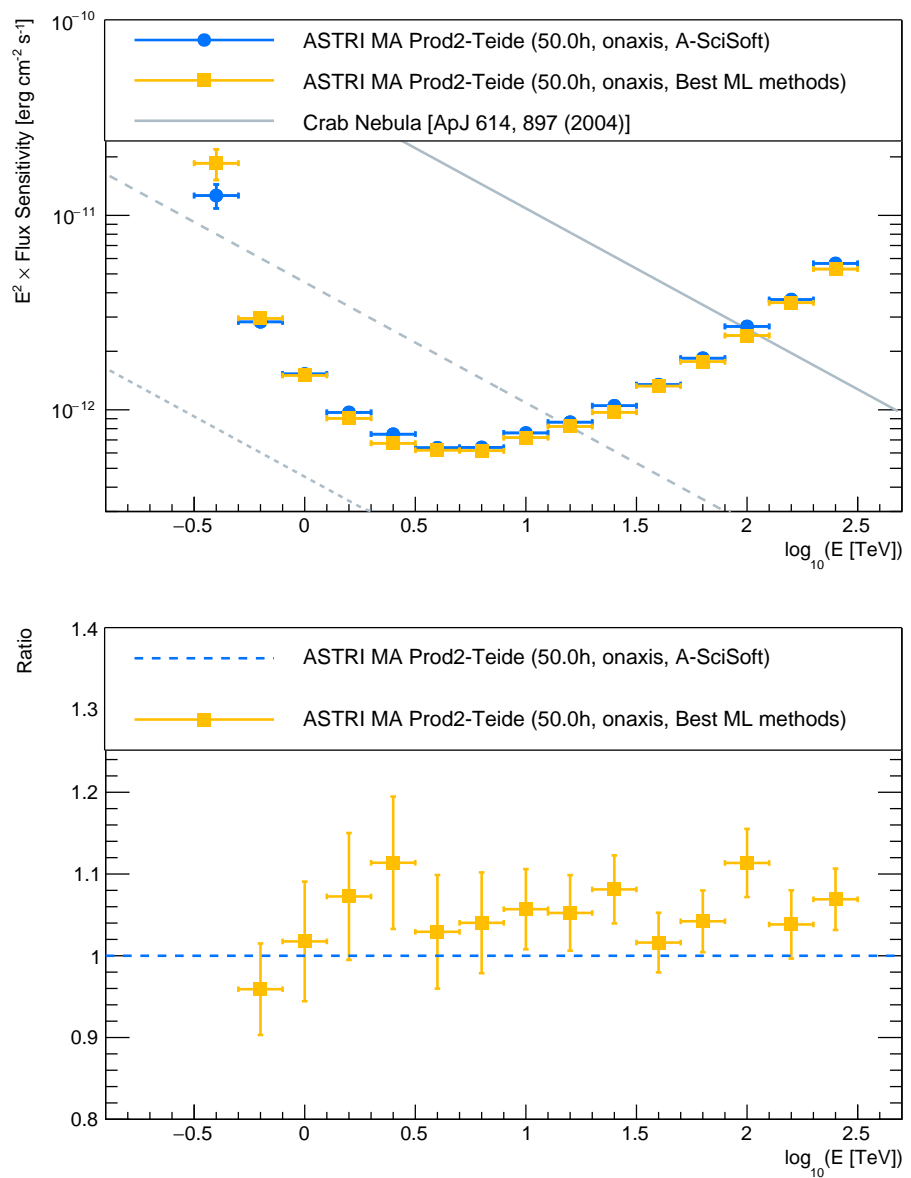


Figure 14. **Up:** On-axis differential sensitivity (in 50 h) achieved with the best Machine Learning methods (orange points) against the benchmark (A-SciSoft, blue points) and Crab Nebula [25] **Bottom:** Comparison between the on-axis differential sensitivity achieved with the best Machine Learning methods and the one achieved with A-SciSoft. The ratio is calculated so that higher values correspond to better performance.

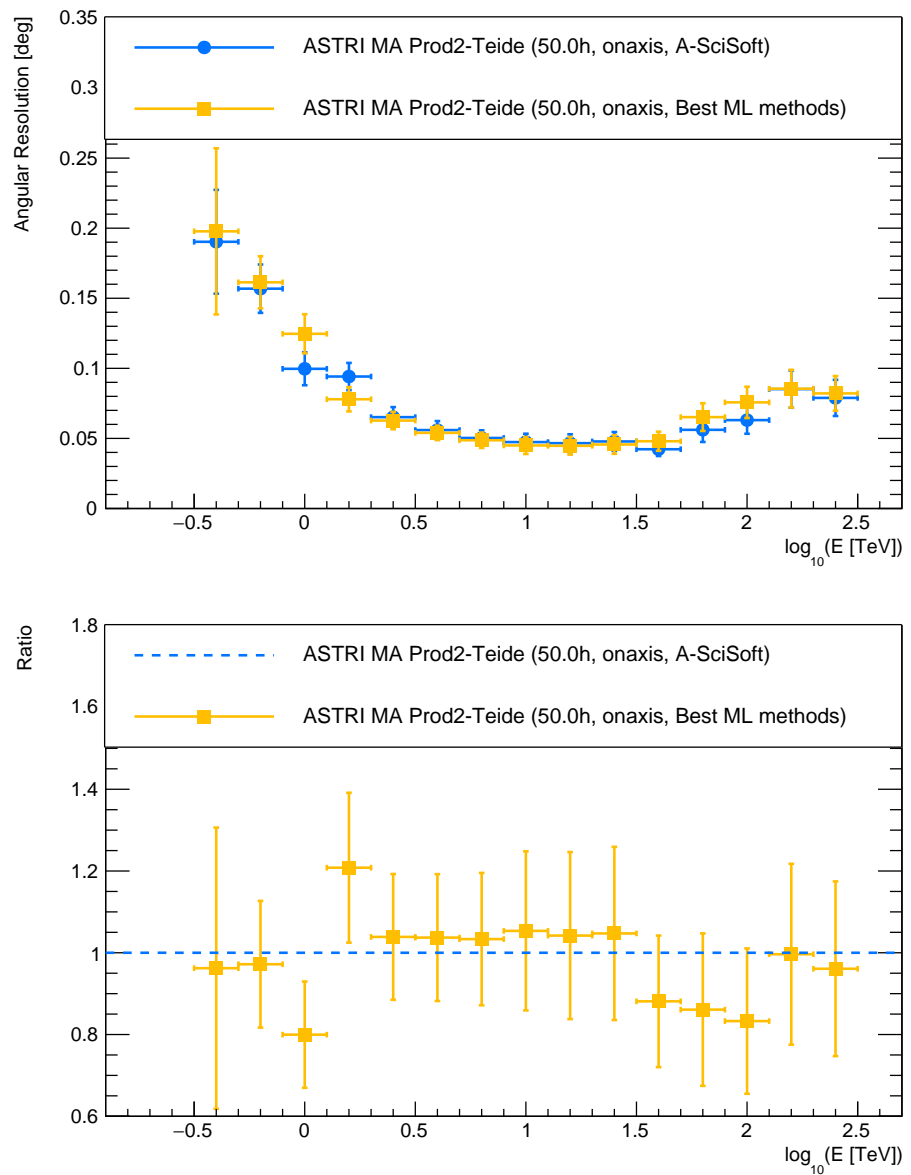


Figure 15. **Up:** On-axis angular resolution achieved with the best Machine Learning methods (orange points) against the benchmark (A-SciSoft, blue points). **Bottom:** Comparison between the on-axis angular resolution achieved with the best Machine Learning methods and the one achieved with A-SciSoft. The ratio is calculated so that higher values correspond to better performance.

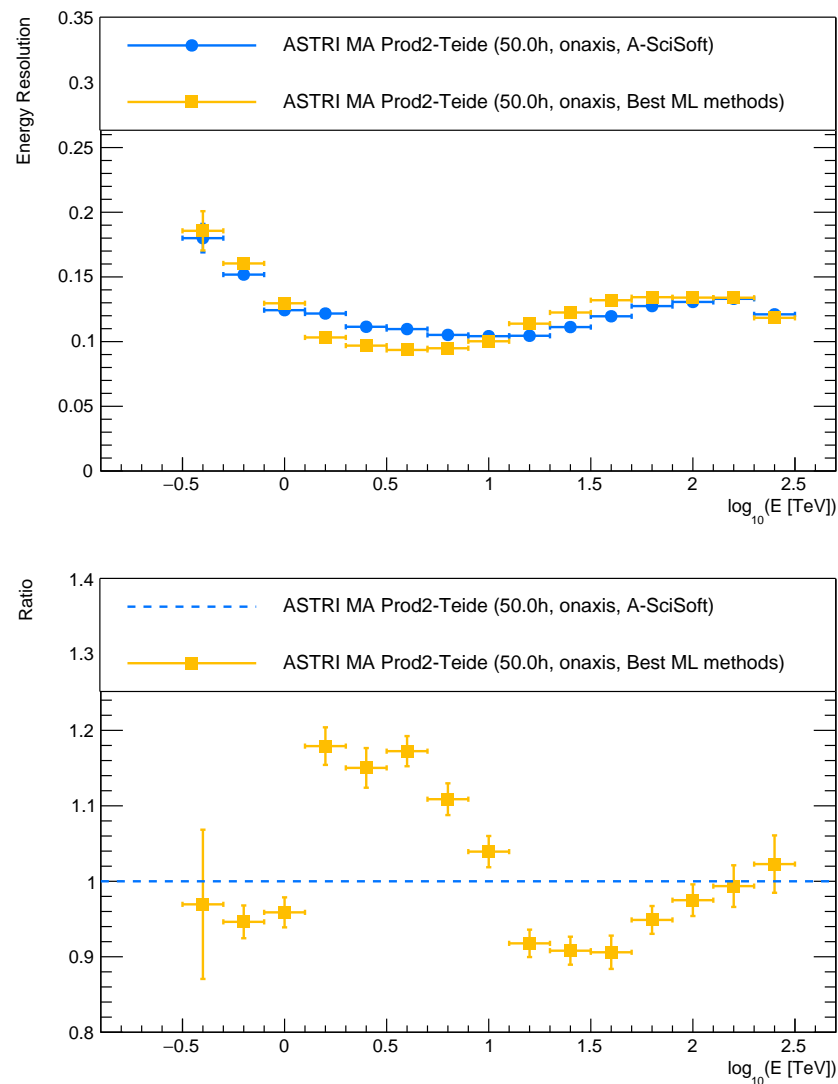


Figure 16. **Up:** On-axis energy resolution achieved with the best Machine Learning methods (orange points) against the benchmark (A-SciSoft, blue points). **Bottom:** Comparison between the on-axis energy resolution achieved with the best Machine Learning methods and the one achieved with A-SciSoft. The ratio is calculated so that higher values correspond to better performance.

As shown, the new Machine Learning methods provide slightly better differential sensitivity above ~ 1 TeV, a comparable angular resolution in the whole energy range, and a better energy resolution in the range ~ 1 – ~ 10 TeV. On the one hand, these results are interesting per se, as they go in the right direction of improving the performance achievable by the data reconstruction. On the other hand, they are also important for confirming that the standard reconstructed methods implemented in A-SciSoft already provide satisfactory, although improvable, results.

However, the relevant improvement of the new Machine Learning methods lies in the bias of the energy reconstruction. In this respect, it is worth mentioning that the standard tools used to compute the high-level performance include an energy bias correction routine. This routine uses a polynomial fit of the projection along the reconstructed energy of the energy migration matrix to correct the reconstructed energy of the events. This routine is applied by default in order to provide high-level performance as that is affected as little as possible by energy reconstruction biases. However, the desirability of having a method that performs as well as possible in terms of energy bias prior to the application of this correction is a very important point for the analysis, especially with a view to the analysis of

real data. Figure 17 show the on-axis point-like source energy bias achieved with A-SciSoft methods (blue points) and the best Machine Learning methods investigated in this work (orange points), before and after the application of the bias correction. From the plot on the left, it is evident that the new Machine Learning methods are much more effective, providing less bias in the energy reconstruction than the standard method currently implemented in A-SciSoft. It should be noted that the energy resolution curves shown in Figures 12 and 13 were obtained prior to any application of an energy bias (and without considering optimized cuts for each analysis) and therefore show different behavior from that which is shown in Figure 17

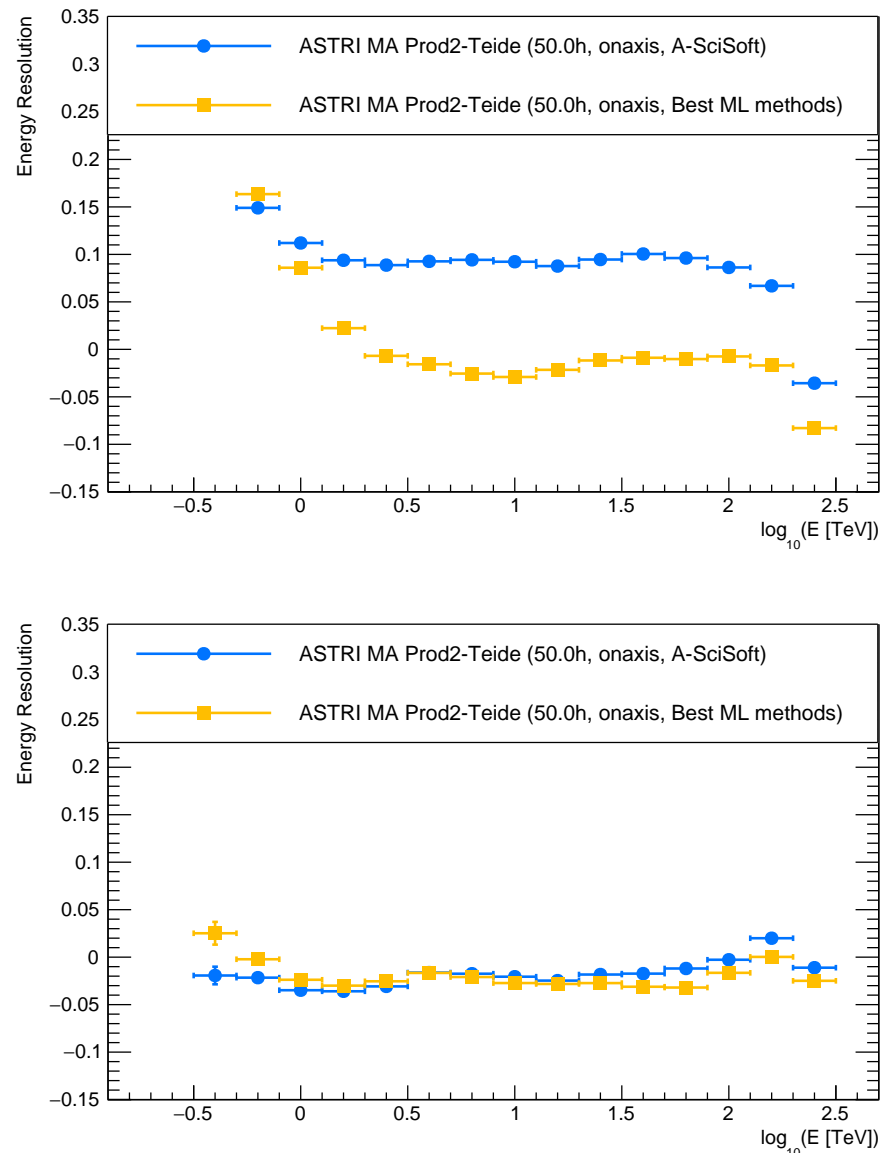


Figure 17. On-axis energy bias achieved with the best Machine Learning methods (orange points) against the benchmark (A-SciSoft, blue points), before (**Up**) and after (**Bottom**) the correction of the energy bias.

10. Discussion

In this work, we have investigated novel Machine Learning techniques distinct from the existing methods implemented in A-SciSoft, the official data reduction and analysis software of the ASTRI project, and showed the effectiveness of their application on simulated ASTRI Mini-Array data.

We have restricted our studies to the case of on-axis gamma-ray analysis. Generalization to off-axis gamma-ray samples may be carried out in future works. However, given the rather flat acceptance of the ASTRI Mini-Array performance as a function of off-axis [7], we expect that the methods investigated in this work can also be effective in the off-axis case and give similar results, in terms of gamma/hadron separation and energy reconstruction, to those found in the on-axis analysis case.

We have assessed and compared several supervised Machine Learning methods for gamma/hadron segregation and energy estimation, including the Random Forest method, Extra Trees method, and Extreme Gradient Boosting (XGB). Through extensive testing and optimization, we identified two Stacking Ensemble methods as the most effective for our purposes.

For gamma/hadron segregation, the optimal Stacking Ensemble method was composed of 42% Extra Trees, 28% Random Forest, and 30% XGB. This composition was determined based on the performance of each individual method in the ensemble, and the specific values were derived through a process of hyperparameter tuning and cross-validation.

For energy estimation, the optimal Stacking Ensemble method was composed of 45% XGB, 27.5% Extra Trees, and 27.5% Random Forest. Once again, these weights were determined through extensive testing and optimization to provide the best performance for our specific dataset and analysis.

By carefully selecting and fine-tuning the composition of the Stacking Ensemble methods, we were able to achieve high accuracy and sensitivity in both gamma/hadron segregation and energy estimation.

We observe from the results shown in the previous sections that the proposed Stacking Ensemble methods demonstrate superior performance compared to the more classical Random Forest method for the ASTRI Mini-Array event reconstruction both for the energy resolution and bias and for the gamma/hadron discrimination.

The most significant results we obtained from our study can be summarized as follows:

- The investigated new Machine Learning methods can provide a performance, in terms of sensitivity, angular resolution, and energy, that is in line if not better than the standard methods implemented in A-SciSoft (based on Random Forest);
- In particular, the new energy reconstruction methods provide a significant reduction in the energy bias with respect to that obtained by the standard methods; this is particularly important in view of the reconstruction of real data (soon to be taken with the ASTRI Mini-Array);
- The results obtained with the new reconstruction methods, which are completely independent of the standard ones, provide an important verification of the robustness of the current standard reconstruction chain implemented in A-SciSoft, which have already been used to obtain detection of the Crab Nebula source with the ASTRI-Horn prototype [26] and to evaluate the performance of the Mini-Array [7].

In future investigations, to ensure the reliability and robustness of our findings, we plan to include a sample of diffuse gamma ray events in our analysis. These events are characterized by having directions uniformly distributed throughout the field of view, which helps us eliminate any potential bias stemming from a preferential source direction present in the data used thus far.

Nevertheless, we anticipate that the performance of our method will remain consistent. This expectation is based on the fact that all the image parameters employed in the process of gamma/hadron separation and energy reconstruction are independent of the event direction.

Although the inferences that can be made from the results hold true for Monte Carlo simulated events and would be necessary to revalidate these conclusions when they are extrapolated to different data samples (such as real data from observations made with actual Cherenkov telescopes), this study may however allow for the elimination of some of the less effective methods.

The methods under investigation all make use of a space of image parameters that is well suited to Monte Carlo scenarios: real data are influenced in ways that cause this space to be distorted. When using Cherenkov telescopes, the night sky background changes during observation. Additionally, the atmospheric conditions can vary greatly, causing inevitable detector changes and malfunctions. None of these distortions have been the focus of our investigation.

In light of these results, the Stacking Ensemble methods identified in this study hold promise for improving the performance of gamma/hadron discrimination and energy estimation in Cherenkov telescope observations. However, further research is needed to validate these findings on real data and to explore the potential of both temporal parameters (e.g., arrival time gradient and RMS of the images [27]) and other Machine Learning methods (e.g., Deep Learning models [28] such as Convolutional Neural Networks, Recurrent Neural Networks, or Generative Adversarial Networks, as well as unsupervised learning techniques [29] like clustering or dimensionality reduction) that were not considered in this study. The ultimate goal is to integrate new high-performing Machine Learning techniques into the official data reduction and analysis pipeline of the ASTRI project. By doing so, we aim to offer the most advanced tools for reconstructing real events acquired by the ASTRI Mini-Array.

Finally, although the simulated data samples considered in this work are specific to the ASTRI Mini-Array, it is reasonable to think that some of the results that we have obtained can also be generalized, at least in qualitative terms, to other Cherenkov telescope arrays, in particular the CTAO small-sized telescopes (SSTs), which share many of the technical characteristics and scientific performance response with the ASTRI telescopes.

11. Conclusions

Our investigation demonstrates the effectiveness of Machine Learning techniques, particularly the Stacking Ensemble methods, for gamma/hadron segregation and energy estimation in the analysis of ASTRI Mini-Array data. The optimized compositions of these methods showed superior performance compared to the traditional Random Forest method. The proposed methods not only offer a performance on par with the standard methods implemented in A-SciSoft but also provide significant improvements in terms of energy bias reduction. These findings validate the robustness of the current reconstruction chain and open up possibilities for integrating high-performing Machine Learning techniques into the official data reduction and analysis pipeline of the ASTRI project.

Although our study focused on simulated data specific to the ASTRI Mini-Array, we anticipate that the results can be qualitatively generalized to other Cherenkov telescope arrays, particularly the CTAO small-sized telescopes (SSTs). Future research should validate these findings using real data and explore the potential of incorporating temporal parameters and other Machine Learning methods, such as Deep Learning models and unsupervised learning techniques, which were not considered in this study. The ultimate goal is to enhance the performance of gamma/hadron discrimination and energy estimation in Cherenkov telescope observations and provide advanced tools for reconstructing real events acquired by the ASTRI Mini-Array.

Author Contributions: Conceptualization, A.P., G.C., A.L.B., V.L.P. and S.L.; Methodology, A.P., G.C., A.L.B., V.L.P. and S.L.; Software, A.P., V.L.P. and S.L.; Validation, A.P., G.C., V.L.P. and S.L.; Formal analysis, A.P., G.C., A.L.B., V.L.P. and S.L.; Investigation, A.P., G.C., A.L.B., V.L.P. and S.L.; Data curation, A.P., V.L.P. and S.L.; Writing—original draft, A.P., A.L.B., V.L.P. and S.L.; Writing—review & editing, A.P., G.C., A.L.B., V.L.P. and S.L.; Supervision, A.P., G.C., A.L.B., V.L.P. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ASI INAF (2022-14-HH.0) and ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing (European Union—NextGenerationEU).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to ASTRI experiment simulation data policy.

Acknowledgments: This work was conducted in the context of the ASTRI Project. We gratefully acknowledge support from the people, agencies, and organisations listed here: <http://www.astrinaf.it/en/library/> (accessed on 10 July 2023). This work is partially supported by ICSC—Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by the European Union—NextGenerationEU. We acknowledge financial support from the ASI-INAF agreement n. 2022-14-HH.0. This paper went through the internal ASTRI review process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. de Naurois, M.; Mazin, D. Ground-based detectors in very-high-energy gamma-ray astronomy. *C. R. Phys.* **2015**, *16*, 610–627. [[CrossRef](#)]
2. Hillas, A.M. Cherenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei. In Proceedings of the 19th International Cosmic Ray Conference, San Diego, CA, USA, 11–23 August 1985; Volume 3, pp. 445–448.
3. Scuderi, S.; Giuliani, A.; Pareschi, G.; Tosti, G.; Catalano, O.; Amato, E.; Antonelli, L.A.; Gonzales, J.B.; Bellassai, G.; Bigongiari, C.; et al. The ASTRI Mini-Array of Cherenkov telescopes at the Observatorio del Teide. *J. High Energy Astrophys.* **2022**, *35*, 52–68. [[CrossRef](#)]
4. Cao, Z.; Aharonian, F.A.; An, Q.; Axikegu, Bai, L.X.; Bai, Y.X.; Bao, Y.W.; Batieri, D.; Bi, X.J.; Bi, Y.J.; et al. Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ -ray Galactic sources. *Nature* **2021**, *594*, 33–36. [[CrossRef](#)] [[PubMed](#)]
5. Vercellone, S.; Bigongiari, C.; Burtovoi, A.; Cardillo, M.; Catalano, O.; Franceschini, A.; Lombardi, S.; Nava, L.; Pintore, F.; Stamerra, A.; et al. ASTRI Mini-Array core science at the Observatorio del Teide. *J. High Energy Astrophys.* **2022**, *35*, 1. [[CrossRef](#)]
6. Bruno, A.; Pagliaro, A.; La Parola, V. Application of Machine and Deep Learning Methods to the Analysis of IACTs Data. In *Intelligent Astrophysics Emergence, Complexity and Computation*; Zelinka, I., Brescia, M., Baron, D., Eds.; Springer: Berlin, Germany, 2021; Volume 39, pp. 115–136.
7. Lombardi, S.; Antonelli, L.A.; Bigongiari, C.; Cardillo, M.; Gallozzi, S.; Green, J.G.; Lucarelli, F.; Saturni, F.G. Performance of the ASTRI Mini-Array at the Observatorio del Teide. In Proceedings of the 37th International Cosmic Ray Conference, Berlin, Germany, 15–22 July 2021; Volume 884.
8. Heck, D.; Knapp, J.; Capdevielle, J.N.; Schatz, G.; Thouw, T. *Report FZKA, 6019*; Forschungszentrum Karlsruhe: Karlsruhe, Germany, 1998.
9. Bernlöhr, K. Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_telarray. *Astropart. Phys.* **2008**, *30*, 149. [[CrossRef](#)]
10. Lombardi, S.; Antonelli, L.A.; Bigongiari, C.; Cardillo, M.; Lucarelli, F.; Perri, M.; Stamerra, M.; Visconti, F. ASTRI data reduction software in the framework of the Cherenkov Telescope Array. In Proceedings of the 10707 Software and Cyberinfrastructure for Astronomy V, Austin, TX, USA, 10–13 June 2018; Volume 107070R.
11. Pence, W.D.; Chiappetti, L.; Page, C.G.; Shaw, R.A.; Stobie, E. Definition of the Flexible Image Transport System (FITS), version 3.0. *Astron. Astrophys.* **2010**, *524*, A42. **2009**, *209*, id.01001 [[CrossRef](#)]
12. Bock, R.K.; Chilingarian, A.; Gaug, M.; Hakl, F.; Hengstebeck, T.; Jiřina, M.; Klaschka, J.; Kotrč, E.; Savický, P.; Towers, S. Methods for multidimensional event classification: A case study using images from a Cherenkov gamma-ray telescope. *Nucl. Instrum. Methods Phys. Res. Sect. A* **2004**, *516*, 511–528. [[CrossRef](#)]
13. Ohm, S.; van Eldik, C.; Egberts, K. Gamma/hadron separation in very-high-energy gamma-ray astronomy using a multivariate analysis method. *Astropart. Phys.* **2009**, *31*, 383–391. [[CrossRef](#)]
14. Fiasson, A.; Dubois, F.; Lamanna, G.; Masbou, J.; Rosier-Lees, S.; Optimization of multivariate analysis for IACT stereoscopic systems. *Astropart. Phys.* **2010**, *34*, 25–32. [[CrossRef](#)]
15. Dubois, F.; Lamanna, G.; Jacholkowska, A. A multivariate analysis approach for the imaging atmospheric Cherenkov telescopes system H.E.S.S. *Astropart. Phys.* **2009**, *33*, 73–88 [[CrossRef](#)]
16. Albert, J.; Aliu, E.; Anderhub, H.; Antoranz, P.; Armada, A.; Asensio, M.; Baixeras, C.; Barrio, J.A.; Bartko, H.; Bastieri, D. Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC. *Nucl. Instrum. Methods Phys. Res. Sect. A* **2008**, *588*, 424–432. [[CrossRef](#)]
17. Sharma, M.; Nayak, J.; Koul, M.K.; Bose, S.; Mitra, A. Gamma/hadron segregation for a ground based imaging atmospheric Cherenkov telescope using Machine Learning methods: Random Forest leads. *Res. Astron. Astrophys.* **2014**, *14*, 1491–1503. [[CrossRef](#)]
18. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
19. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman & Hall: Boca Raton, FL, USA, 1984.
20. Available online: <https://pypi.org/project/lazypredict/> (accessed on 10 July 2023).
21. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Volume 13–17, pp. 785–794

22. Acharyya, A.; Agudo, I.; Angüner, E.O.; Alfaro, R.; Alfaro, J.; Alispach, C.; Aloisio, R.; Batista, R.A.; Amans, J.-P.; Amati, L. Monte Carlo studies for the optimisation of the Cherenkov Telescope Array layout. *Astropart. Phys.* **2019**, *111*, 35. [[CrossRef](#)]
23. Aharonian, F.; Akhperjanian, A.G.; Barrio, J.A.; Bernlöhr, K.; Bojahr, H.; Calle, I.; Contreras, J.L.; Cortina, J.; Denninghoff, S.; Fonseca, V. The energy spectrum of TEV gamma rays from the Crab Nebula as measured by the HEGRA system of imaging air cerenkov telescopes. *Astrophys. J.* **2000**, *539*, 317–324. [[CrossRef](#)]
24. Li, T.-P.; Ma, Y.-Q. Analysis methods for results in gamma-ray astronomy. *Astrophys. J.* **1983**, *272*, 317. [[CrossRef](#)]
25. Aharonian, F.; Akhperjanian, A.; Beilicke, M.; Bernlöhr, K.; Börst, H.G.; Bojahr, H.; Bolz, O.; Coarasa, T.; Contreras, J.L.; Cortina, J.; et al. The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA Stereoscopic Air Cerenkov Telescopes. *Astrophys. J.* **2004**, *614*, 897–913. [[CrossRef](#)]
26. Lombardi, S.; Catalano, O.; Scuderi, S.; Antonelli, L.A.; Pareschi, G.; Antolini, E.; Arrabito, L.; Bellassai, G.; Bernlöhr, K.; Bigongiari, C.; et al. First detection of the Crab Nebula at TeV energies with a Cherenkov telescope in a dual-mirror Schwarzschild-Couder configuration: the ASTRI-Horn telescope. *Astron. Astron. Instrum.* **2020**, *634*, A22. [[CrossRef](#)]
27. Aliu, E.; Anderhub, H.; Antonelli, L.A.; Antoranz, P.; Backes, M.; Baixeras, C.; Barrio, J.A.; Bartko, H.; Bastieri, D.; Becker, J.K. Improving the performance of the single-dish Cherenkov telescope MAGIC through the use of signal timing. *Astropart. Phys.* **2009**, *30*, 293–305. [[CrossRef](#)]
28. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, UK, 2016.
29. Greene, D.; Cunningham, P.; Mayer, R. Unsupervised Learning and Clustering. In *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*; Cunningham, P., Cord, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2008.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.