BACHELOR'S THESIS

Gamma-hadron separation with the upcoming Cherenkov Telescope Array

Author:
Sofia Llàcer Caro

Supervisor:
dr. Manuela Vecchi

Second Examiner:
dr. Kristof de Bruyn

Bachelor's Thesis for the Programme:
BSc Physics - Particle Physics
Kapteyn Astronomical Institute
Faculty of Science & Engineering
University of Groningen
April 2023 - July 2023

university of groningen / faculty of science and engineering

Gamma-hadron separation with the upcoming Cherenkov Telescope Array

Sofia Llàcer Caro
S4111303

Abstract

Gamma-hadron separation techniques are required for ground detections in the area of gamma-ray astronomy. The aim of this work is to implement and compare different separation methods for the next generation of detectors: the upcoming Cherenkov Telescope Array. Cut-based, logistic regression and decision trees were implemented and tested on simulated CTA detections. Then, their performance was compared, finding that decision trees outperform the rest of methods implemented, reaching around 95% of rightly classified as gamma-events. An even higher performance of these algorithms could be attained through further adjustments of the hyperparameters involved in their implementation.

university of groningen / faculty of science and engineering

Gamma-hadron separation with the upcoming Cherenkov Telescope Array

Sofia Llàcer Caro
S4111303

Acknowledgements

— *Bonica, ¿dónde aterrizan esas estrellas que hacen "shiiium"?*

My great-grandmother, who is turning 100 this year, asked me - full of light in her eyes - where shooting stars land after I tried explaining my research. I would like to dedicate this work to her.

Of course, it would not have been possible without the guidance of my supervisor dr. Manuela Vecchi. She was both demanding and encouraging in equal parts, for which I am thankful. Remembered to tell me "Sofia, life is beautiful", but also "This makes no sense" when required.

I would also like to thank my family for the support they have provided during the months leading to this research. For all those afternoons of knitting and re-potting plants while thinking about Cherenkov radiation.

To all the friends I made during this BSc: Lisa, Michela, Agus, Abhi, Anmol, Jitae and countless more. Finally, to Adrian, for listening to my ramblings on SSTs while having a fever. For being a wonderful support, for your curiosity for the smallest things. All in all, for being the best partner I could have asked for.

Contents

1 Introduction

The exploration of fundamental physics in recent decades has been experimentally possible due to the development of particle accelerators. These operate at high energies to study elementary particles, as is the case of the remarkable Large Hadron Collider, where collisions can reach energies up to 13 TeV [1]. Particle accelerators on Earth are a very useful tool to study high energy processes in a controlled environment. However, they have energy limitations bound to factors including geographical considerations or restricted spatial extension [2]. This has set the focus on the detection of extraterrestrial particles, as some of them have been detected at energies in the order of magnitude of PeV and beyond [3] [4]. Thus, the study of high-energy particles upon their arrival on Earth is a powerful tool for physics to probe extreme environments beyond terrestrial laboratories [5].

Gamma rays are of particular interest among the particles that arrive at Earth. They are photons with energies of GeV and above. Their lack of electromagnetic charge means that they do not get deflected by magnetic fields in their trajectory. Consequently, gamma rays point back to their source and their detection can provide a faithful representation of cosmic sources [6]. The detection of gamma rays has been possible through ground-based telescopes since 1989, with the first gamma-ray image of the Crab Nebula [7]. Since then, major developments have been made in the area of ground-based gamma-ray astronomy, leading to a network of observatories all around the world [8]. In this context, the Cherenkov Telescope Array (CTA) is the future gamma-ray observatory which will be in the forefront of this kind of detections for the next decade and beyond. It will not only extend the energy range of gamma ray detections between 20 GeV and 300 GeV but also significantly improve the sensitivity of current measurements by a factor of 5-10 [9].

There are, however, other types of particles besides gamma rays that arrive on Earth at the energies at which CTA will operate. The major contribution of other types of events is due to cosmic rays, particularly protons. The occurence of gamma-ray events (signal) with respect to other detections (background) is around 0.1% [10]. The task of separating both types of particles is an occuring problem in the area of ground-based gamma-ray astronomy and is commonly referred to as gamma/hadron separation [1][11]. There are currently numerous lines of research open on the different methods available for the gamma/hadron separation applied to the Cherenkov Telescope Array. Improvements in the signal-background separation performance of CTA would allow for higher sensitivity. This would lead to higher significance in the detection of sources and, additionally, the detection of very faint sources e.g. dark matter [9].

The aim of this work is to provide a comparison between the traditional method used for this separation and recently introduced ways to perform this task through machine learning. All this will be applied to simulations of events detected by the Cherenkov Telescope Array.

[1]The term 'hadron' arises from the terminology for protons and other particles of the same kind in general. However, the most relevant species in this problem are protons in particular. Other particles which are not hadrons (i.e. electrons or positrons) are 100 times less abundant and 10000 less abundant, respectively, than protons at TeV energies, so were not considered in the present analysis.

2 Theoretical Overview

The study of high energy extraterrestrial particles through ground-based observatories is possible due to the physical phenomena that happen upon their arrival at the Earth's atmosphere. These physical phenomena are namely Extended Air Showers (EAS) and the Cherenkov effect. After these effects have happened, the detection of its products on the Earth's surface can take place. Upon detection, the information obtained can be used to reconstruct the energy, direction and type of the primary particle. The information of interest in the present work is whether the particle detected is a signal gamma ray or a background proton.

2.1 Physical Phenomena

2.1.1 Extended Air Showers

Extraterrestrial high-energy particles interact with the atoms composing the Earth's atmosphere. This gives rise to the so-called Extended Air Shower (EAS), which can be divided into two types depending on the types of interactions involved in their production. Gamma rays are associated with electromagnetic showers, whereas protons give rise to hadronic showers [2] [12].

- **Electromagnetic showers**
  This kind of EAS can be initiated by photons. They develop due to electromagnetic interactions through two combined processes: pair production and Bremsstrahlung. Both of them happen under the influence of the Coulomb fields of air atoms' nuclei. If its energy is sufficient[3], a photon can undergo pair production (into an electron and a positron). The resulting electron and positron can undergo the Bremsstrahlung emission of a high energy photon again (see Fig. 2.1). The succession of these phenomena in the atmosphere gives rise to the development of the electromagnetic shower [13].

- **Hadronic showers**
  These have more complex descriptions as they involve electromagnetic but also strong interactions, giving rise to different components. As can be seen in Fig. 2.1, hadronic showers generally start though the interaction of a high-energy hadron (for instance a proton) with the nucleus of an atom composing the air and causing its disintegration. The hadronic component includes events where hadrons interact via the strong force with atmospheric nuclei, as long as the energy suffices, to produce more hadrons and vice versa. Apart from hadrons, a nuclear disintegration can produce pions [4], which are highly unstable mesons and quickly decay. On one hand, charged pions, $\pi^{\pm}$, with an approximate half-life of $2.6 \cdot 10^{-8}$ s, will travel less than 8 meters through the atmosphere [15]. Their main decay channel is into $\mu^{\pm}$, which accounts for the

[2] Other particles can also produce EAS. However, only gamma rays and protons are presented in this work.

[3] The minimum energy requirement for pair production is around 1 MeV (equivalent to two electron rest masses).

[4] In addition to pions, kaons can also be produced in these interactions. However, they decay into muons and pions, which is why the discussion above is mainly centered around these last particles [14].

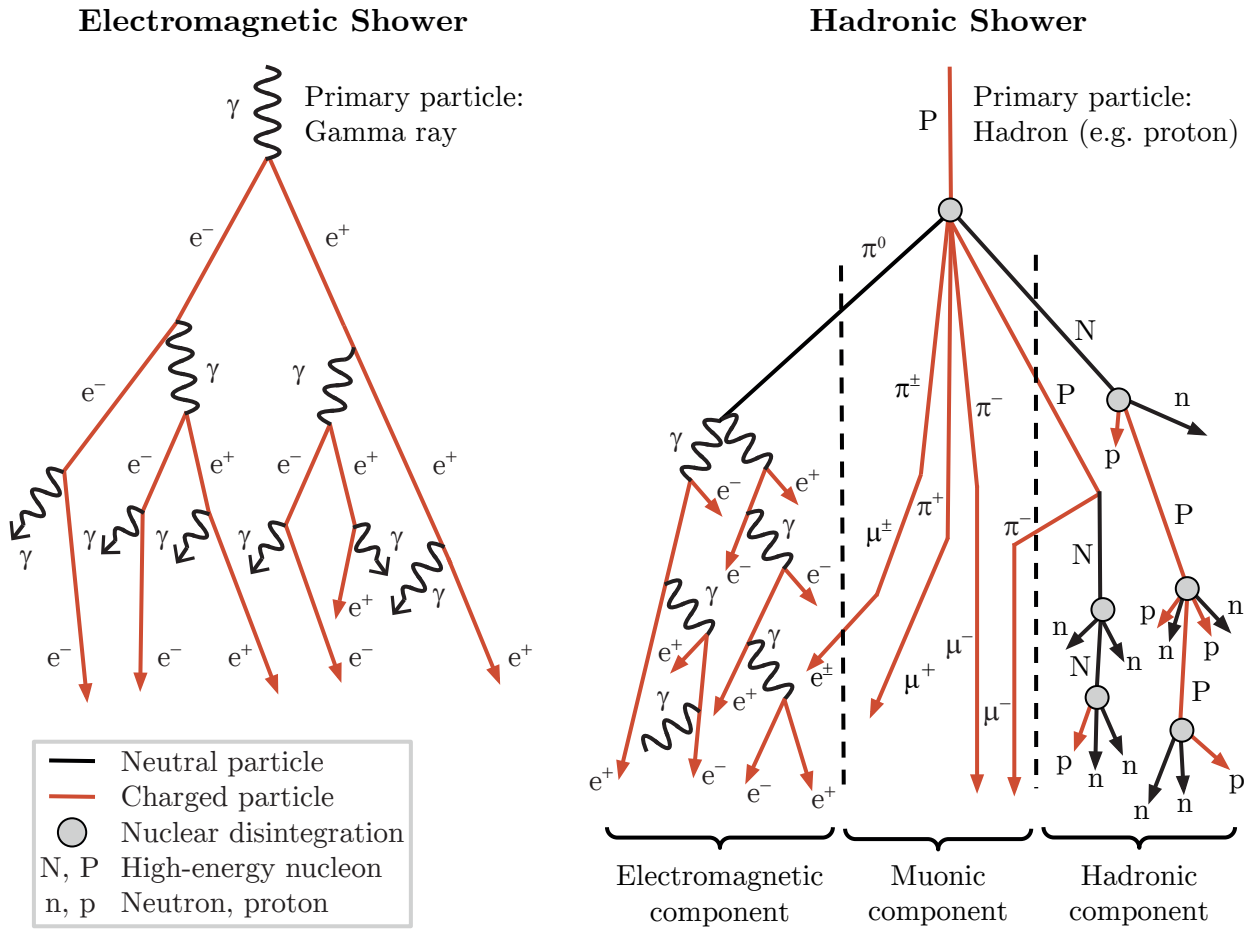

Figure 2.1: Schematic depiction of the composition of electromagnetic and hadronic showers.

muonic component of the shower[5]. On the other hand, neutral pions, π^0 have an even shorter half-life than $\pi^{\pm}$ and decay into gamma rays. Then these gamma rays undergo pair production and develop into an electromagnetic shower themselves, thus giving rise to the electromagnetic component of the hadronic shower [13].

The growth in the number of particles as an electromagnetic shower develops is exponential at the beginning. Nevertheless, the energy of the initial particle gets redistributed to other particles as the EAS develops. Consequently, the resulting particles will have less energy per particle. When the collision loss rate equals that of Bremsstrahlung, the rate of ionisation will reach the critical energy (around 83 MeV in the air). Then, the electromagnetic shower development desists [14]. Whereas the electromagnetic component of the hadronic shower follows the description above, the development of the muonic and hadronic component significantly differs from the electromagnetic component. The muons of the shower can pass through the material relatively unaffected, losing only a small fraction of their energy through ionization. The particles in the hadronic component undergo additional interactions, including elastic scattering, inelastic scattering and particle production, leading to the

[5]Other decay channels are not considered because the main channels for all types of pions account for 99% of the decay fraction in both cases [15].

production of further hadrons. The development of the hadronic component is complex due to the nature of the strong interaction and therefore leads to a different shape compared to the purely electromagnetic EAS [16]. The main variation lies in the fact that hadronic showers are in general more scattered, while electromagnetic ones remain narrower in their development [17]. This is especially noticeable on the higher energy ranges of detection of CTA [18].

2.1.2 Cherenkov Effect

Ultra-relativistic charged particles in air can induce Cherenkov light. This is due to two factors: first, the air is a dielectric medium and charged particles passing through it can polarise it. Second, air has a refractive index above one, so ultra-relativistic particles can travel at speeds faster than the speed of light in the medium. The result is the Cherenkov effect: the formation of a wavefront of coherent light named Cherenkov radiation [12]. Its wavelengths span from the ultraviolet region to visible light, and it is emitted in a cone-like shape as can be seen in Fig. (2.2).

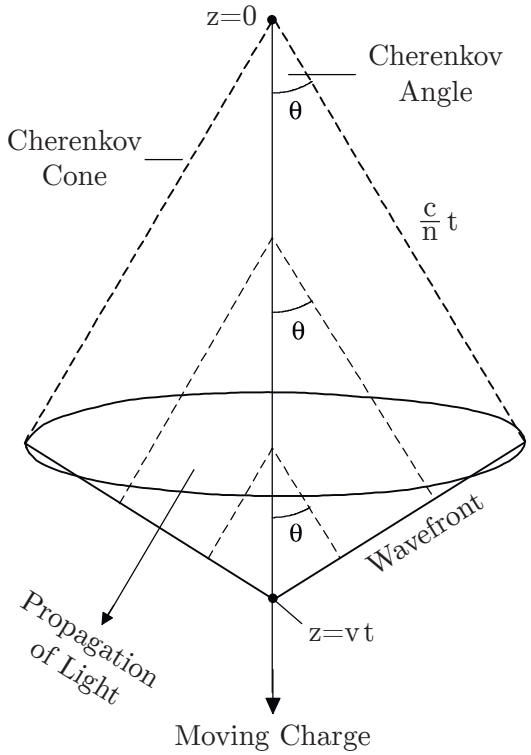

Figure 2.2: Basic geometry of Cherenkov radiation phenomenon. Shown are the Cherenkov emission angle, θ (see equation 2.1), of a charged relativistic particle moving along the z-axis, the instantaneous position of the wavefront and the direction of propagation of the Cherenkov photons. Adapted from [13].

The Cherenkov emission angle θ follows the relationship below due to coherence requirements,

$$\cos\theta = \frac{1}{\beta n} \tag{2.1}$$

Where β is the Lorentz factor and n is the refractive index of the medium, in this case air [13]. The particles in EAS from primary gamma rays and protons that fulfil the requirements undergo the

Cherenkov effect [6]. The large-scale effect for EAS that is a cone-shaped radiation that is generally called 'light pool'. EAS take place at about 10 km above sea level with an emission angle θ of approximately 1.0 to 1.3° [7] and it can be detected through dedicated observatories [13].

2.2 Detection Technique

2.2.1 Cherenkov Telescope Array

Ground-based gamma-ray astronomy has been benefiting from the aforementioned phenomena for decades through the use of Imaging Air-Cherenkov Telescopes (IACTS). The detection principle consists in the reflection of the light pool from EAS through mirrors onto a photon sensor. Then, the information is converted into digital data by using photomultipliers [9]. Placing numerous detection systems as such on the ground allows for the later reconstruction of details of the EAS and subsequently the primary particle (see Fig. 2.3)[8].

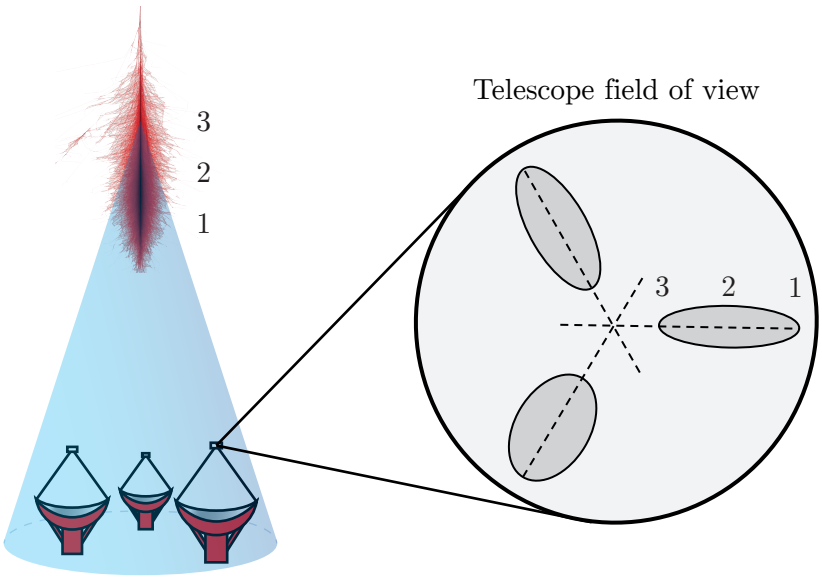

Figure 2.3: Schematic depiction of the detection technique used by IACTs. In red is a simulated image of an EAS. The grey shaded cones represent the emmitted Cherenkov radiation. The numbers correspond to different regions of the light pool as seen in the telescope field of view. Based on [19].

The Cherenkov Telescope Array (CTA) will be a IACT with a telescope arrangement covering both the northern and the southern skies, with a location in La Palma (Spain) and in Paranal (Chile). It will consist of three different types of telescopes: Large Size Telescopes (LST), Medium Size Telescopes (MST) and Small Size Telescopes (SST). The different sizes of telescopes will allow for an energy spectrum coverage ranging from 20 GeV to about 300 TeV [9]. All this together translates to

[6] For detail on which particles produced in EAS are charged, see Fig. 2.1

[7] For the energy range considered and the medium being air.

[8] Note that the EAS image in Fig. 2.3 is based on the CORSIKA simulation (see [17]) of an electromagnetic shower caused by a primary 50 GeV photon, taken from [17].

[9] Higher sizes corresponds to lower energies

an improvement in the sensitivity of detections above 1 TeV by a factor of 5-10 with respect to the previous generation of IACTs (H.E.S.S, VERITAS or MAGIC). In addition to this, it will also be able to measure shorter timescale phenomena than the current observatories. The angular resolution will be improved too, leading to higher capability to resolve extended sources [9].

2.2.2 Hillas Parameters

The images resulting from the detections through IACTs have an ellipse-like shape, which can be parametrised for quantitative analysis [20]. The most widely used technique for this are the Hillas parameters (see Fig. 2.4). They were first introduced in 1985 by A. M. Hillas in search for criteria on gamma/hadron separation and are still being used in CTA simulations for the same purpose [20].

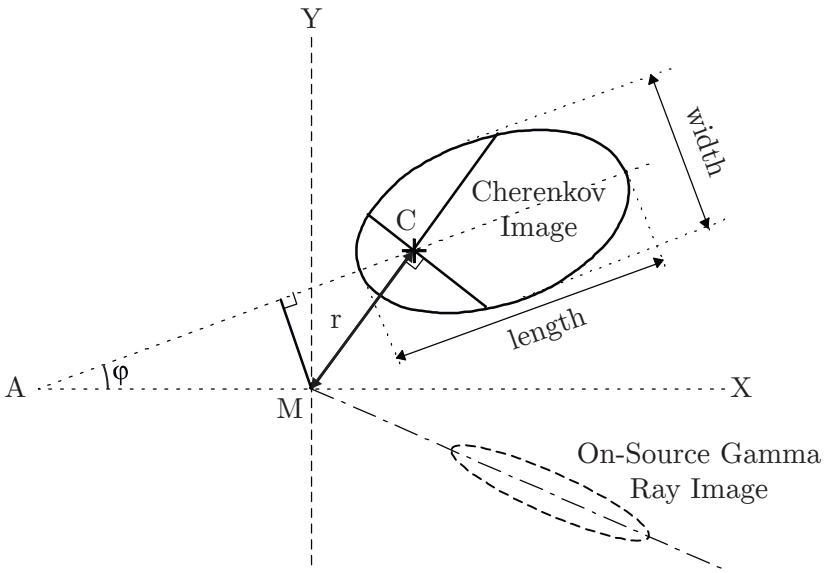

Figure 2.4: Principle and basic parameters of air Cherenkov image analysis. The solid ellipse indicates the pixel image contour, C is the centroid of the image (location of highest brightness) and M the center of the field of view. Based on [19].

The major and minor axis of the ellipse, labeled length and width in Fig 2.4, are two examples of the Hillas parameters. Another parameter of interest is the Hillas radius, representative of the distance between the centre of the telescope field of view and the centroid of the Cherenkov image. Except for the clean regular elliptic shape this image is also representative for hadronic showers. The dashed ellipse at the lower right with the extension of the major axis intercepting the center M of the mirror, labelled 'On-Source Gamma Ray Image', shows the typical narrow elliptic contour of a gamma ray shower when the mirror axis is pointing at the source [19].

3 Experimental Methods & Approach

The aim of this work is to compare the conventional method used for gamma/hadron separation to recently-introduced machine learning methods in the context of CTA. In short, this was done by first applying the different classification techniques to a data set consisting of simulated gamma-ray and proton events detected by CTA and then assessing the performance of each of the methods in this task.

The simulated data used in this work was obtained through a number of steps. First, the open source software CORSIKA was used to simulated the development of extensive air showers in the atmosphere, as well as their Cherenkov emission [17]. Then, the simulation of the arrays of Imaging Atmospheric Cherenkov Telescopes and their response to the shower-induced emission was performed using sim_telarray [21]. Finally, the output of sim_telarray was read and processed directly with ctapipe, a low-level data processing pipeline software for CTA (see [22]). The output of ctapipe in H5 format contained information on the particle type and its energy, as well as on the signals detected by the telescopes and their Hillas parametrisation[10]. The result was a data set of simulated events as detected by CTA and their associated primary particle. The data set used was class-balanced, with 9291 gamma events and the same number of proton events.

The data set was composed of a combination of LST, MST and SST detections. However, further analysis was performed on each telescope type individually and on different camera types in the case of the MSTs. Unless explicitly stated, the results provided were obtained based on a combination of all telescope sizes together.

3.1 Classification Methods

3.1.1 Cut-Based Method

Historically, the method used for the gamma/hadron separation was to establish a cut-off value or threshold for the features that were more image-sensitive [23]. This procedure was applied by plotting the distributions of gamma (signal) and proton (background) events for a given feature in the data set. Then, a threshold was chosen for the features which showed a separation between the two types of events (see Fig 3.1).

3.1.2 Machine Learning

The advent of machine learning has been a breakthrough for the processing of large volumes of data in the area of astroparticle physics [24]. Along this line, the majority of methods chosen for the gamma/hadron separation in CTA simulated events are based on machine learning. The main focus is particularly on supervised learning classification algorithms. These algorithms must be trained on a number of features such as the Hillas parameters in order to determine whether an event is gamma-

[10]The term "feature" may be used for the Hillas parameters as it aligns with machine learning terminology.

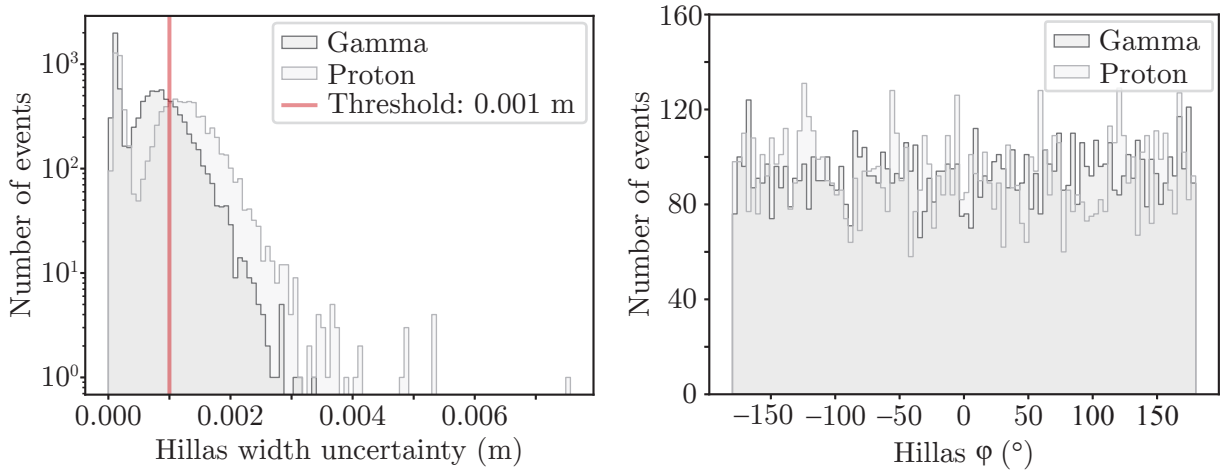

Figure 3.1: Example of the gamma and proton distributions of the number of events for two different features. On the left is the Hillas width uncertainty, which was used for the cut-based method. On the right is a feature where the separation was not discernible and therefore unusable for such purpose.

or proton- induced. Thus, the type of the initial particle must be known, which is provided in the data set used in this analysis. In general, they work as follows: first, they are trained on a training set which contains both feature values and the target class or actual particle type. Then, they are given a test set which contains only the feature values but not the target class. They are therefore made to predict an event's class through its feature values [25]. Thus, in the test set, the algorithm is given the Hillas parameters of an event and not the type of particle the associated with it. For the present analysis, all the features present in the data set (40) were used in the machine learning algorithms [11].

The proportion used for the training and test set was 80% to 20% respectively, which is a common practice in the area of machine learning. The variable hyperparameters were chosen to be the standard ones provided by the widely used machine learning python library "Scikit - learn" (see [26]). The particular algorithms tested for the gamma/hadron separation were the following:

- **Logistic Regression**
  This is one of the simplest methods used for classification, and it works by obtaining a probability between 0 and 1 of the event belonging to a given class. This is calculated through linearly combining the features and assigning a weight or an importance to them based on the training set [24]. Then, this linear combination is put into a sigmoid function. This is the general function of choice because it is a smooth S-like function bounded between 0 and 1, making it a very useful tool for expressing probabilities [27]. Once the importance for each feature is given through the training set, the function is applied to the test set in order to determine the probability of an event belonging to a given class, say a gamma-ray event. The probability threshold for an event to be considered gamma-ray event is 0.5, the standard value [28] [12].

[11] Using fewer features yielded very similar results, so we decided to use all the information available in the data set.
[12] For a further study on the choice of threshold, see appendix A

- **Decision trees**

  Decision trees are sequential models, which logically combine a sequence of simple tests; each test compares a numeric attribute against a threshold value of a feature. When a data point falls in a partitioned region, a decision tree classifies it as belonging to the most frequent class in that region [29]. Two variants of decision trees were used: random forest and boosted decision tree (BDT). The first kind incorporates the construction of many of these decision trees and then choosing the most popular class among them (gamma or proton) to be the predicted one [30]. BDTs have been used before in other IACTs like H.E.S.S [31] and they incorporate continuous values of features into quantiles instead of a unique threshold and binary choice like basic decision trees [32].

3.2 Performance Assessment

After applying all the methods chosen for the classification of events into gamma- or proton- events (cut-based, logistic regression, random forest and boosted decision tree), their performance was assessed. This was done through the use of confusion matrices, from which the F-score was calculated as a metric for evaluating the performance of each of the classification models. Note that, for the cut-based method, the performance assessment was done for the highest-separation feature.

3.2.1 Confusion Matrix

A confusion matrix is a widely used tool used in the assessment of classification algorithms [33]. It consists on a grouping of the correctly classified events and the wrongly classified events. In the case of a binary class problem, this results in a 2x2 matrix which contains percentages where the entries are generally referred to as true positive (TP), false positive (FP), false negative (FN) and true negative (TN), (see Fig. 3.2).

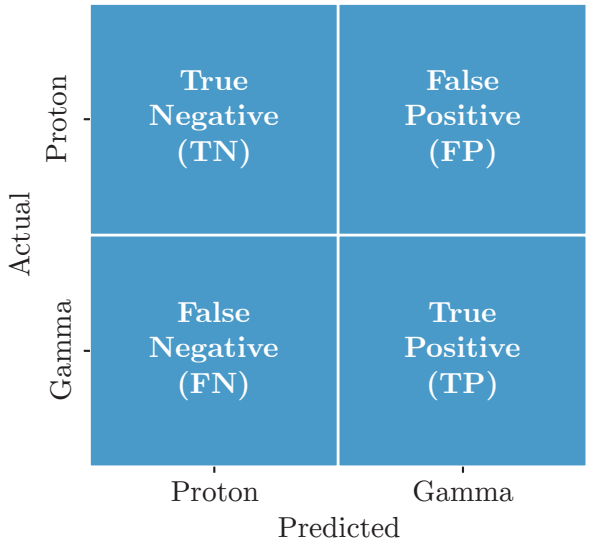

Figure 3.2: Annotated diagram of a confusion matrix in a binary classification problem. Here, the positive class is considered to be gamma (signal) and vice versa.

3.2.2 F-Score

Confusion matrices are a very useful tool for the evaluation of classification algorithms. They provides a great starting point and a visual representation of the performance of a classifier. However, further metrics for model performance assessment can be defined in order to summarise this information compactly in a single value. The choice of performance metric generally depends on the type of problem at hands and considerations dependent on the priorities and goals of the analysis.

Accuracy is the overall rightly classified events over the total of events. Even though this is a valuable measure for the performance of the model, there are two inconveniences in its use for this case. Firstly, assuming the classifier will most likely not be perfect, there are two possible types of errors: FP and FN. However, the measure of accuracy gives no information about the distribution onto these two categories. Secondly, we have to take into account the fact that we are dealing with a classification problem where the signal (gamma) has much more importance over the background (proton). That is why we are not so interested in knowing about the TN events (rightly classified as background), but rather about our two types of error. These errors can be assessed though the metrics of sensitivity and precision, which are widely used metrics in the area of astroparticle physics for the gamma-hadron separation problem in Cherenkov detectors. The precision and sensitivity are defined as follows [34].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.2}$$

Equation 3.1 and 3.1 can be combined in the concept of the F-score [35]. This widely used metric in machine learning is a harmonic mean between precision and sensitivity, and provides information on the performance of a classifier while taking into account the importance of one of the classes (signal or gamma) over the other (background or proton). In addition, this metric has already been used before in the CTA collaboration for the performance assessment of classifiers in the context of gamma/hadron separation [36].

$$\text{F-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{3.3}$$

Where the first part of equation 3.3 offers a definition with respect to precision and sensitivity, and the second part is expressed with respect to the confusion matrix terms[13].

To summarise, the widely used confusion matrix provides a non-case-sensitive way to visually assess the performance of a classification model. Additionally, the defined F-score offers a case-dependent yet compact quantitative measure of the performance of each of the classifiers in this particular problem, aiming at easing the comparison between models.

[13] As a clarification, the absolute TN, TP, FN, FP were used instead of the rate because the data set considered consisted of 50% gamma and 50% proton events

4 Results & Discussion

Fig. 4.1 shows the confusion matrices obtained after applying the different classification methods for the data including all telescopes (SST, MST, LST). It can be seen that 73.42% of the gamma events were correctly classified in the case of the cut-based method. This is for the feature which displayed the greatest distribution separation, which turned out to be the Hillas width uncertainty. For this particular feature, proton events displayed higher values than gamma ones (see Fig. 3.1 for reference). The threshold in the cut-based method, which was chosen to be 0.001 m, indicated that although the separation of the two distributions was not very high, it was noticeable enough. This may be due to the fact that gamma showers are generally more narrow and defined than proton showers [19]. Hadronic showers are known to be more dispersed, which may explain a greater uncertainty when identifying the width of the ellipse arising from the light pool of Cherenkov radiation. This may have been reflected in an overall greater uncertainty in the simulated data for the Hillas width.

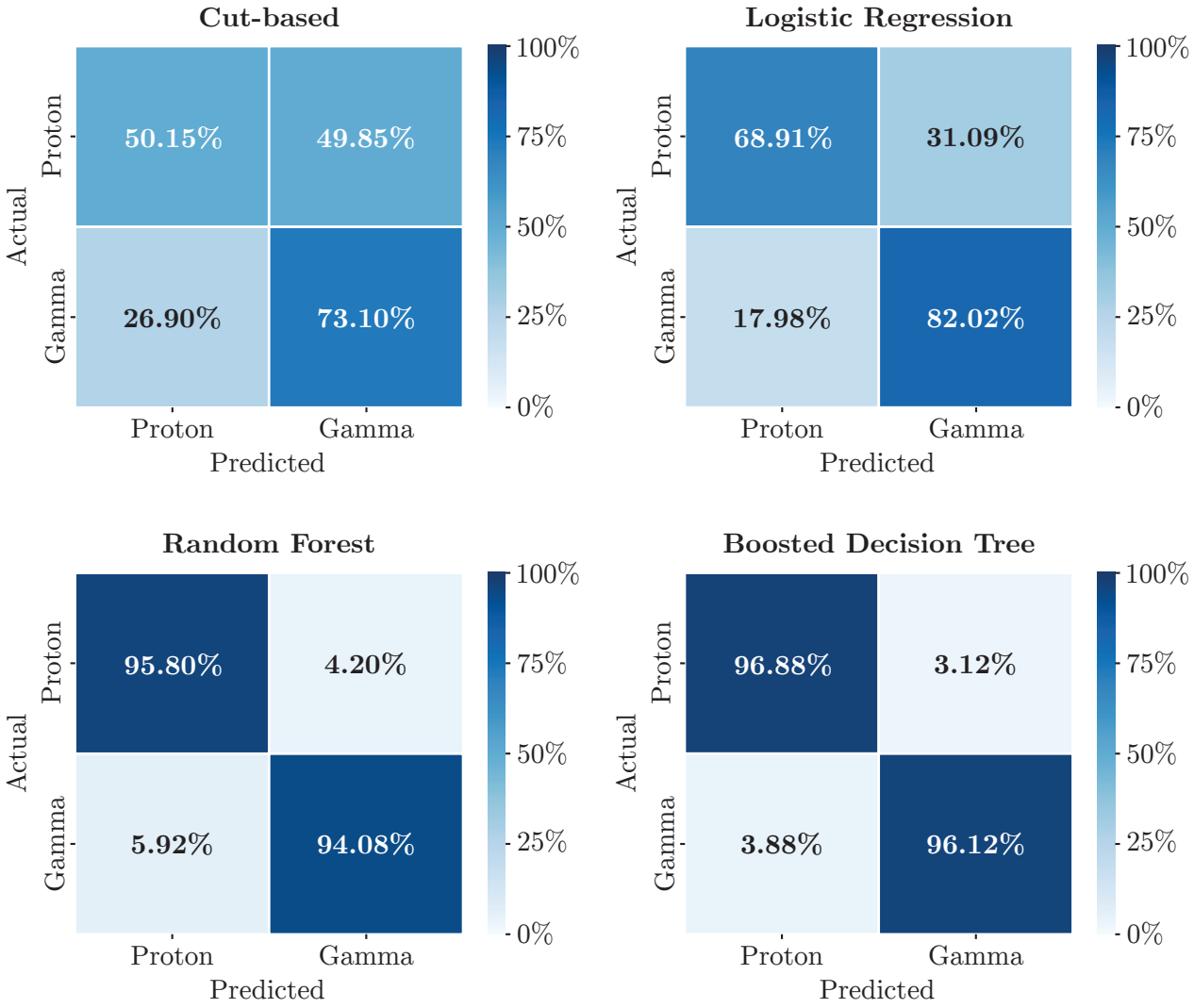

Figure 4.1: Confusion matrices obtained after applying the different classification methods (cut-based, logistic regression, random forest and boosted decision tree).

Logistic regression as a classification algorithm yielded an 83.10% correctly identified gamma events. Even though this supposes a 10% improvement with respect to the historically used method, this is only the case when the entire energy range of CTA is considered (20 GeV to 300 TeV). Upon further inspection, it was found that the cut-based method performance can be majorly improved by only considering the data from SSTs, operating in the higher end of this energy range (for more detail, see appendix C). In order to examine whether the differences had to do with the instruments used or with a real physical variation, different types of cameras were examined when possible. This was the case for MSTs, where FlashCam and NectarCam were analysed separately. It was found that the type of camera did not have an impact in the separation per feature, and the order of most relevant features for the cut-based method was the same for both types of cameras. This indicates that the relevance of features is not dependent on the instrument used, but rather on the physical aspect detected. Taking all the previous into account, there is strong indication that the development of gamma and proton EAS at higher energies (above 5 TeV, operating range of SSTs) is more different than at lower energies.

Taking the previous considerations into account, when only the SSTs are considered, the cut-based method can yield an F-score as high as 0.77 ± 0.01 as can be seen in Fig. 3.3. Logistic regression would offer a comparable performance when applied to all telescopes (0.78 ± 0.01). As clearly seen in both Fig. 4.1[14] and 3.3, decision trees are the best option when it comes to separation of gamma- and proton- events in CTA. Both random forest and boosted decision tree show over 95% rightly identified gammas. The F-score of the random forest was 0.96 ± 0.01 [15], while the boosted decision tree's was 0.97 ± 0.01. Even though random forest has a lower F-score, the difference with respect to the boosted decision tree lies within the error margin, which makes these two methods' performance comparable and effectively considered the same. These two machine learning methods are encompassed in the same family of classification algorithms (decision trees), which can be explain the fact that their F-scores are in agreement with each other.

At first, the F-scores associated with decision trees seemed excessively high, so some additional search on the possible correlation between the features provided to the algorithms was done. However, the correlation matrix did not yield any particular high values (the highest value found for the correlation coefficient was 0.24). Once the we tested that there was no leakage of information or correlation leading to such a high performance, we can propose that these algorithms could be even improved. This could be done through adjusting the variables or hyperparameters chosen during their implementation. The standard values of the built-in functions were used, and further analysis on the effect of hyperparameter tuning in the functions should be explored, possibly yielding different performances when used on the test data. Along these lines, in Fig. A.1, the different classification methods are given for comparison. However, it must be noted that this information cannot be used

[14] Further performance metrics are presented in appendix D

[15] Detail on the calculation of the errors for the metrics is given in appendix B

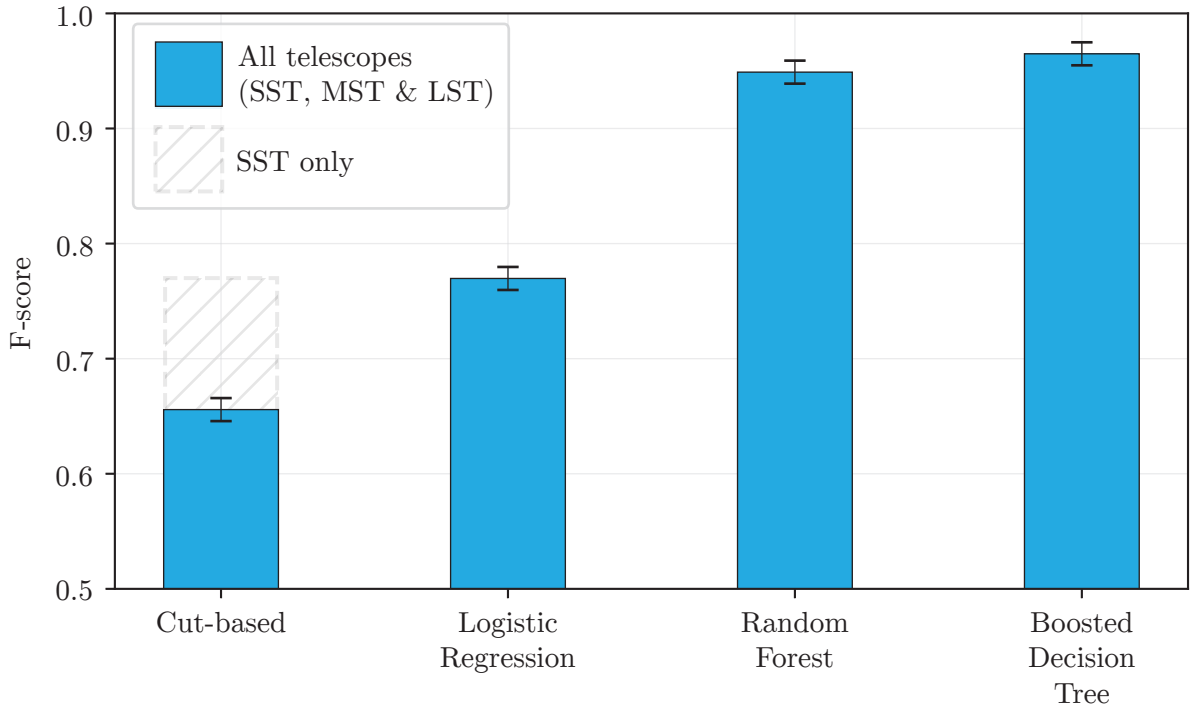

Figure 4.2: F-score for the different classification methods implemented. Results are given for a combination of all telescopes, and the cut-based F-score when only SSTs are considered is also displayed for reference.

as an overall statement on the performance of classification methods. Rather, it must be taken with the caution of taking into account that this is only a comparison for the used hyperparameters and the provided dataset. Namely, the information displayed is rather a statement on the performance of these methods under the particular conditions chosen. As a remark, decision trees are prone to overfitting [37], which is an aspect to be considered when further investigating the hyperparameters of this family of algorithms.

In spite of the fact that decision trees yielded very accurate classifications, a few additional factors should be considered when assessing their performance when it comes to CTA. First, the analysis performed in the present work evaluates the performance of the classification methods on class-balanced datasets. Namely, the amount of signal and background was considered to be the same. This can be chosen when working with a simulation. However, this is not expected to be the case with real data, as there is a high class-imbalance where the majority of instances are background and the signal represents a very small fraction 0.1% of the detections [10]. A possibility would be to investigate data sets where the proportion of signal to background mimics the real case. In this way, the evaluation of the performance of the algorithms based on the simulation would be more applicable to real detections. There have been numerous studies on the class imbalance problems applied to the context of decision trees [38], and even algorithms developed, like the Class Confidence Proportion Decision Tree, which take this class imbalance into account [39]. These special case of decision trees could bring the analysis on gamma-hadron separation to a more realistic and applicable case in the area of gamma-ray detection.

Finally, analysis on real detections should be performed in order to assess the capabilities of decision trees as classification algorithms. Furthermore, other classification methods besides those considered in the present work should be investigated for gamma-hadron separation. Neural networks, for instance, are being expored as a very promissing method in image reconstruction in gamma-ray astronomy [40]. This type of algorithm has been applied to other IACTs before [41] and is being applied particularly to one of the first real detections of CTA [42]. The efforts for obtaining highly accurate results for gamma-hadron separation in the data of CTA range from implementation of decision trees to convolutional neural networks to numerous others [43]. This is a topic not only relevant for a small portion of research groups, but rather a problem that has been investigated for decades and should remain evolving. Finding better strategies to separate ground-based gamma-ray detections will also allow us to increase the confidence when answering questions currently still open in the area of physics: What are the most energetic events in our Universe? Is there new physics beyond the Standard Model?

5 Conclusion

In summary, the best performing classification algorithms for gamma-hadron separation tested on this work for all types of telescopes of CTA are decision trees. On the other hand, more rudimentary machine learning techniques like logistic regression yielded similar results to the cut-off method. The latter's performance highly benefited from considering each size of telescope separately. This is expected because the energy range analysed is narrower and therefore the features have greater differences between gamma-ray and proton-initiated EAS. Furthermore, the differences between these EAS are strongly believed to be greater for higher energies (above 5 TeV). Limitations of this work include the number of events considered for each class (gamma-ray and proton) when analysing the performance of the classification methods. A balanced (50% gamma, 50% proton) data set was considered. However, this is not the case in the occurence of EAS, which have a high class imbalance, where only 0.1% of the are gamma-ray initiated [10]. A suggested improvement for this aspect is to make a randomly chosen data set where the proportion of gamma to proton detections mimics that which is naturally occurring. Furthermore, other types of classification methods should be examined, such as convolutional neural networks, in order to improve gamma-hadron separation techniques. This would improve the reliability of overall results that ground-based gamma-ray astronomy can offer, particularly with the upcoming Cherenkov Telescope Array.

References

[1] CERN (official website), 2023. [Online]. Available: https://home.cern/science/accelerators Accessed: 19/6/2023.

[2] B. Holzer, "Introduction to particle accelerators and their limitations," *CERN Yellow Reports*, vol. 1, pp. 29–29, 2016.

[3] D. J. Bird *et al.*, "Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high energies," *Phys. Rev. Lett.*, vol. 71, pp. 3401–3404, 21 Nov. 1993. DOI: 10.1103/PhysRevLett.71.3401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.71.3401.

[4] Z. Cao *et al.*, "Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray galactic sources," *Nature*, vol. 594, no. 7861, pp. 33–36, 2021.

[5] J. Stacy and W. Vestrand, "Gamma-ray astronomy," in *Encyclopedia of Physical Science and Technology (Third Edition)*, R. A. Meyers, Ed., Third Edition, New York: Academic Press, 2003, pp. 397–432, ISBN: 978-0-12-227410-7. DOI: https://doi.org/10.1016/B0-12-227410-5/00274-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B012227410500274X.

[6] B. Degrange and G. Fontaine, "Introduction to high-energy gamma-ray astronomy," *Comptes Rendus Physique*, vol. 16, no. 6, pp. 587–599, 2015, Gamma-ray astronomy / Astronomie des rayons gamma, ISSN: 1631-0705. DOI: https://doi.org/10.1016/j.crhy.2015.07.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1631070515001292.

[7] M. Catanese and T. C. Weekes, "Very high energy gamma-ray astronomy," *Publications of the Astronomical Society of the Pacific*, vol. 111, no. 764, p. 1193, Oct. 1999. DOI: 10.1086/316435. [Online]. Available: https://dx.doi.org/10.1086/316435.

[8] A. Mitchell, "Status of Ground-based and Galactic Gamma-ray Astronomy," in *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*, vol. 395, 2021, p. 046. DOI: 10.22323/1.395.0046.

[9] B. S. Acharya *et al.*, *Science with the Cherenkov Telescope Array*. World Scientific, Nov. 2018. DOI: 10.1142/10986. [Online]. Available: https://doi.org/10.1142%5C%2F10986.

[10] J. Hinton, "Ground-based gamma-ray astronomy with cherenkov telescopes," *New Journal of Physics*, vol. 11, no. 5, p. 055 005, May 2009. DOI: 10.1088/1367-2630/11/5/055005. [Online]. Available: https://dx.doi.org/10.1088/1367-2630/11/5/055005.

[11] M. Vecchi, Personal communication via email, Jun. 2023.

[12] C. Grupen, G. Cowan, S. Eidelman, and T. Stroh, *Astroparticle physics*. Springer, 2005, vol. 50.

[13] P. K. Grieder, "Atmospheric cherenkov radiation," in *Extensive Air Showers: High Energy Phenomena and Astrophysical Aspects A Tutorial, Reference Manual and Data Book*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 835–878, ISBN: 978-3-540-76941-5. DOI: 10.1007/978-3-540-76941-5_16. [Online]. Available: https://doi.org/10.1007/978-3-540-76941-5_16.

[14] M. de Naurois and D. Mazin, "Ground-based detectors in very-high-energy gamma-ray astronomy," *Comptes Rendus Physique*, vol. 16, no. 6-7, pp. 610–627, 2015.

[15] R. L. Workman *et al.*, "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, 2022. DOI: 10.1093/ptep/ptac097.

[16] S. K. Makelvani and M. Bahmanabadi, "Discrimination between gamma and proton-induced showers using the parabolic shape of the air shower front," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1042, p. 167 468, 2022.

[17] T. Pierog, D. Heck, and R. Ulrich, *Corsika*, version 7.7410, May 2021. DOI: 10.5281/zenodo.5246070. [Online]. Available: https://doi.org/10.5281/zenodo.5246070.

[18] F. Schmidt and J. Knapp, 2005. [Online]. Available: https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html.

[19] J. Holder, "Atmospheric cherenkov gamma-ray telescopes," in *The WSPC Handbook of Astronomical Instrumentation: Volume 5: Gamma-Ray and Multimessenger Astronomical Instrumentation*, World Scientific, 2021, pp. 117–136.

[20] A. M. Hillas, "Cerenkov light images of eas produced by primary gamma," in *19th Intern. Cosmic Ray Conf-Vol. 3*, 1985.

[21] K. Bernlöhr, "Simulation of imaging atmospheric cherenkov telescopes with corsika and sim_telarray," *Astroparticle Physics*, vol. 30, no. 3, pp. 149–158, 2008.

[22] K. Kosack *et al.*, *Cta-observatory/ctapipe: V0.19.3*, version v0.19.3, Jun. 2023. DOI: 10.5281/zenodo.8063139. [Online]. Available: https://doi.org/10.5281/zenodo.8063139.

[23] M. Bunse, "Machine learning for acquiring knowledge in astro-particle physics," Ph.D. dissertation, Dissertation, Dortmund, Technische Universität, 2022, 2022.

[24] C. Weihs *et al.*, "A case study on the use of statistical classification methods in particle physics," in *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, W. A. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, and J. Kunze, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 69–77, ISBN: 978-3-642-24466-7.

[25] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons*, vol. 4, pp. 51–62, 2017.

[26] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[27] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.

[28] A. Subasi, "Chapter 3 - machine learning techniques," in *Practical Machine Learning for Data Analysis Using Python*, A. Subasi, Ed., Academic Press, 2020, pp. 91–202, ISBN: 978-0-12-821379-7. DOI: https://doi.org/10.1016/B978-0-12-821379-7.00003-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128213797000035.

[29] S. B. Kotsiantis, "Decision trees: A recent overview," *Artificial Intelligence Review*, vol. 39, pp. 261–283, 2013.

[30] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[31] S. Ohm, C. van Eldik, and K. Egberts, "γ/hadron separation in very-high-energy γ-ray astronomy using a multivariate analysis method," *Astroparticle Physics*, vol. 31, no. 5, pp. 383–391, 2009, ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2009.04.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927650509000589.

[32] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794, ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. [Online]. Available: https://doi.org/10.1145/2939672.2939785.

[33] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sensing of Environment*, vol. 62, no. 1, pp. 77–89, 1997, ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(97)00083-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425797000837.

[34] C. Bockermann *et al.*, "Online analysis of high-volume data streams in astroparticle physics," in *Machine Learning and Knowledge Discovery in Databases*, A. Bifet *et al.*, Eds., Cham: Springer International Publishing, 2015, pp. 100–115, ISBN: 978-3-319-23461-8.

[35] D. Olson and D. Delen, *Advanced Data Mining Techniques*. Springer Berlin Heidelberg, 2008, ISBN: 9783540769170. [Online]. Available: https://books.google.es/books?id=2vb-LZEn8uUC.

[36] D. Riquelme, M. Araya, S. Borquez, B. Panes, and E. Carquin, "Deep learning semi-supervised strategy for gamma/hadron classification of imaging atmospheric cherenkov telescope events," in *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods*, – Science and Technology Publications, 2023, pp. 725–732, ISBN: 978-989-758-626-2. DOI: 10.5220/0011611500003411.

[37] A. Rogozhnikov, "Reweighting with boosted decision trees," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 762, 2016, p. 012 036.

[38] D. A. Cieslak and N. V. Chawla, "Learning decision trees for unbalanced data," in *Machine Learning and Knowledge Discovery in Databases*, W. Daelemans, B. Goethals, and K. Morik, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 241–256, ISBN: 978-3-540-87479-9.

[39] W. Liu, S. Chawla, D. A. Cieslak, and N. V. Chawla, "A robust decision tree algorithm for imbalanced data sets," in *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM)*, pp. 766–777. DOI: 10.1137/1.9781611972801.67. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611972801.67. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611972801.67.

[40] I. Shilon *et al.*, "Application of deep learning methods to analysis of imaging atmospheric cherenkov telescopes data," *Astroparticle Physics*, vol. 105, pp. 44–53, 2019.

[41] S. De, W. Maitra, V. Rentala, and A. M. Thalapillil, "Deep learning techniques for imaging air cherenkov telescopes," *Physical Review D*, vol. 107, no. 8, p. 083 026, 2023.

[42] M. Jacquemont, T. Vuillaume, A. Benoit, G. Maurin, P. Lambert, and G. Lamanna, "First full-event reconstruction from imaging atmospheric cherenkov telescope real data with deep learning," in *2021 International Conference on Content-Based Multimedia Indexing (CBMI)*, 2021, pp. 1–6. DOI: 10.1109/CBMI50038.2021.9461918.

[43] E. Lyard, R. Walter, V. Sliusar, N. Produit, and for the CTA Consortium, "Probing neural networks for the gamma/hadron separation of the cherenkov telescope array," *Journal of Physics: Conference Series*, vol. 1525, no. 1, p. 012 084, Jun. 2020. DOI: 10.1088/1742-6596/1525/1/012084. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1525/1/012084.

Appendix

Gamma-hadron separation with the upcoming Cherenkov Telescope Array

Appendix to the Bachelor's thesis, '*Gamma-hadron separation with the upcoming Cherenkov Telescope Array*' (2023) by Sofia Llàcer Caro. Written for the BSc Physics: Particle Physics programme at the University of Groningen.

Contents

A ROC curve for machine learning algorithms

Even though a probability of 0.5 was considered as the threshold in order to deem an event gamma or hadron induced, other values were also considered as an additional exploration. Below is the receiver operating characteristic (ROC) curve for the logistic regression, random forest and boosted decision tree algorithms implemented. It can be seen that, regardless of the value chosen as a discriminating threshold for the binary classification, decision trees remain the best classification method out of those implemented.

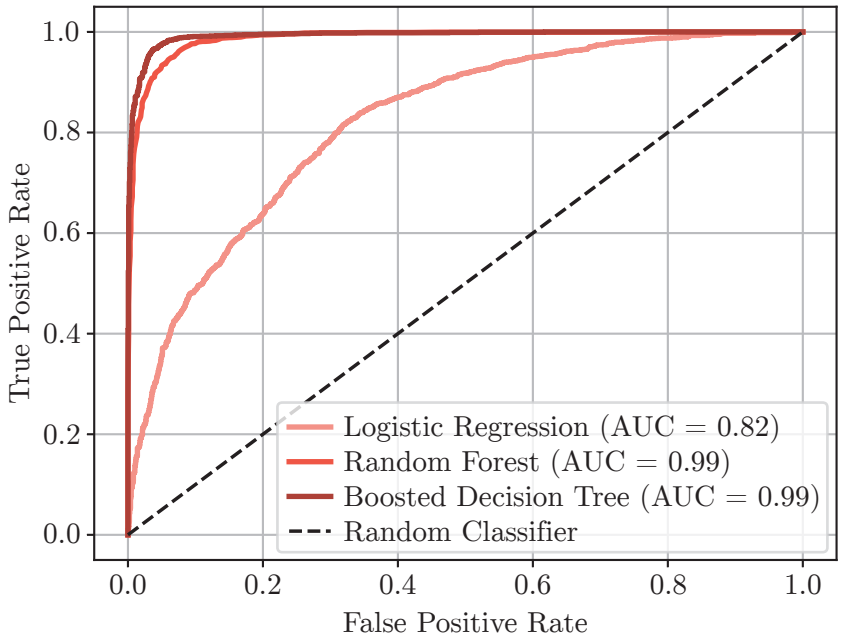

Figure A.1: ROC curve for the different machine learning algorithms implemented in the present work. The area under the curve (AUC) is also presented for each of them.

B Error analysis

The error associated to the performance metrics was calculated through the estimation of the effect of a number of factors in the final results. The different performance metrics were calculated several times while changing the ratio of training to test set. Then, the random seeds used along the code in order to choose random subsets of data were changed. The maximum deviation of the final result was as estimated though the uncertainties provided.

C Cut-based analysis on SSTs

The Hillas parameter which showed the greatest separation between gamma and proton distributions was the Hillas radius when only SSTs were considered. This encompasses an energy range from 5 TeV to 300 TeV. The threshold value used for the F-score calculation for SSTs only in Fig. A.1 can be seen in Fig. C.1 below.

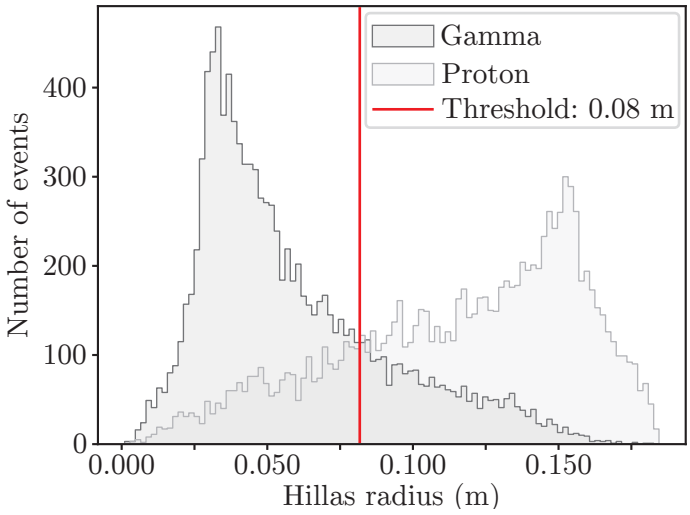

Figure C.1: Distribution of number of gamma- and proton- events for a given Hillas radius, including the threshold chosen for the cut-based method when only SSTs were considered.

D Further performance metrics

The results for various commonly-used performance metrics for the different classification methods implemented is presented in Table 1. The F-score is as presented in Fig. A.1 and the precision and sensitivity were calculated through 3.1 and 3.2 respectively. The accuracy was calculated as follows

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where the terms are those of the confusion matrix. The error was estimated through the method explained in appendix B. Note that the results presented are for a chosen discriminating threshold between gamma and proton of 0.5.

Table 1: Performance metrics for the different classification methods implemented.

Method	**Performance Metrics (± 0.01)**			
	Accuracy	**Precision**	**Sensitivity**	**F-score**
Cut-based	0.61	0.59	0.72	0.65
Logistic Regression	0.76	0.74	0.80	0.77
Random Forest	0.95	0.96	0.95	0.95
Boosted Decision Tree	0.97	0.97	0.98	0.97