Rijksuniversiteit Groningen

Bachelor Thesis

A machine learning approach to gamma-ray identification with the Cherenkov Telescope Array

rijksuniversiteit groningen

Author:
Wiegert I. Kaal

Supervisor:
dr. M. Vecchi

Date:
Summer 2023

Abstract

The Cherenkov Telescope Array (CTA) is expected to be far more accurate and have a five to ten times higher sensitivity than the current generation of ground-based gamma-ray detectors. In this work, we tested the effectiveness of Boosted Decision Trees (BDTs), a widely used Machine Learning algorithm that is easy to implement and has shown promising results in similar classification problems, for the separation of gamma-ray-initiated air showers and cosmic-ray-initiated air showers, which constitute the main background for gamma-ray astronomy, being about 1000 times more abundant. We looked at what properties of the air shower image were the most important for the BDT, and how accurately it could identify gamma-ray showers for CTA. "Out of a total sample of 39 features which were provided in the simulated events provided by the CTA consortium, we identified 4 features that provided a good separation power. Using all 39 features, our BDT reaches a precision of over 98 percent. This makes the BDT a very suitable method for the separation of gamma-ray-initiated and cosmic-ray-initiated air showers.

Contents

1 Introduction

Across the universe, there are objects that accelerate particles to very high energies, and when these particles enter our atmosphere they initiate air showers: cascades of interactions between the atmospheric nuclei and highly energetic charged particles [1]. Currently (at the time of writing), a new telescope array, called the Cherenkov Telescope Array Observatory (CTAO), is being built to investigate these objects through air showers. In this work, we aim to investigate what methods are most suitable to be used by the CTAO for the identification of γ-ray initiated air showers, which can be used to study these objects. The main problem in identifying these showers is that there is a lot of background from cosmic rays, which initiate very similar air showers. Therefore, we need to separate the two shower types in some way.

There have already been investigations on similar topics. For H.E.S.S, a telescope array with a similar objective as CTAO, multiple methods of separation have already been investigated [2], and for VERITAS, another similar observatory, extensive research has been done to investigate the use of Boosted Decision Trees (BDTs) for the separation of γ and hadron air showers [3]. BDTs are a commonly used machine learning algorithm, that has been shown to be an effective classification method in similar investigations. We hope to contribute in this area by testing the effectiveness of a machine-learning method for the CTAO.
Therefore, our thesis was formulated as:

- Is the Boosted Decision Tree suitable for separating air showers initiated by γ-ray from those initiated by cosmic rays?

Before answering this question, we will explain everything necessary to understand the problem and the methods used to solve them in the first chapter. Will explain why the objects that produce γ-rays and cosmic rays are interesting to us, and then how they are generally investigated. Then, we will explain how CTA is built to access a new energy range, its importance in its field, and the simulated data that was made to investigate its performance.
In the methods, we will first explain how the method will be evaluated, and then look further into what parts of the simulated data we will use. When those are clear, we can finally explain the BDT method we used for discriminating the two shower types. We will explain why this method was chosen, the choices in its setup, and how it will process the data. Finally, we will look further into what parts of the simulated data were most important for the BDT, and how we found them.
In the results, we will look at the relationship between the precision of the method and its parameters, to finally obtain its optimal efficiency. These results are then compared and evaluated.
In the conclusion, we will summarise our findings, look at their importance from a broader perspective, and discuss the limitations of this investigation.

2 γ-ray astronomy and CTA

2.1 γ-ray astronomy

γ-rays are photons with a very short wavelength, which means they have very high energies: typically they are classified as photons with an energy higher than 10^5 eV [4] and were measured for the first time in 1989 [1]. Since then, the arrays developed for observing these particles have become a hundred times more sensitive, and have proven to be an effective method to study cosmic ray sources and the non-thermal universe [5].

2.1.1 γ-ray sources

To understand the importance of γ-rays, it is important to understand how they are created.
γ-rays sources are objects in the universe capable of accelerating hadronic particles to speeds close to the speed of light, with either particle interactions or magnetic or gravitational potentials inside or around the object. Although the production of these particles is still a partially unsolved problem, we know there are only a few objects that are capable of accelerating particles to these energies. Examples of objects that are suitable are supernova explosions, accreting black holes, the center of an active galactic nuclei, and pulsars. We call these particles primary cosmic rays [4].
When these cosmic rays, which are charged, interact with the interstellar medium, they can create photons with energies similar to the energy of the particle through various processes. For instance, they can be deflected by magnetic or coulomb fields, and radiate a photon through bremsstrahlung or synchrotron radiation. Most of the time, this happens close to the source that created the cosmic ray, which makes it a γ-ray source [1].

Besides acceleration in a γ-ray source, cosmic rays can also be created far away from the source of their particle, by secondary acceleration, for instance by interaction with extensive gas clouds when the particle propagates through the interstellar or intergalactic medium [4]. However, the concentration of γ-rays from these cosmic rays will be much lower compared to the γ-ray concentration near the γ-ray sources [1].

These objects are interesting for us to study for several reasons. As mentioned before, they are capable of accelerating particles to energies that we currently cannot recreate here on Earth, which is interesting for various fields of science. Moreover, they would help us to understand the origin of cosmic rays and gives us insight into the extreme conditions of these objects, where the currently accepted physics can be tested, or new physics can be found. Furthermore, it gives access to a completely new energy range, which historically always enabled the discovery of new phenomena [6].

2.1.2 Cosmic rays and γ-rays

To study these objects, we need to investigate the particles that we receive from them, which are the γ-rays and cosmic rays.

More than 90 percent of all cosmic rays are protons, but there are also electron, positron, and antiproton cosmic rays, which are, as mentioned before, all charged [6]. This is the cause of a major problem with studying the γ-ray sources with the detection of cosmic rays. During its propagation from its source to us, the cosmic ray is subject to a great number of interactions,

like the creation of secondary particles by interaction with interstellar gas or the secondary acceleration we mentioned before. The decay of radioactive components, and energy loss through ionization are also of importance because they change the energy of the particle. The alteration of energy is not only a problem because it influences how external factors can change the direction of the particle, but the initial energy is also one of the properties we are most interested in. Additionally, the cosmic rays might be subject to solar modulation. These interactions are of greater influence for cosmic rays with low energies but are strong enough to influence cosmic rays with higher energies as well. Because there is still known to little about these parameters, the propagation of cosmic rays is not well-understood. This makes it impossible to trace the origin of a cosmic ray when we observe it on earth [7].
Even though we cannot trace back their origin, the research on cosmic rays has taught us more about the composition of the universe and the origin of certain elements, like Lithium [8]. Moreover, they are a promising candidate to teach us more about dark matter, as several accesses in the cosmic ray spectrum might be explained by the production of cosmic rays by dark matter [9].

But while cosmic rays still have their use, we cannot rely on them to study specific γ-ray sources. Therefore we use the γ-rays instead. Unlike cosmic rays, γ-rays do not change their direction or energy while traveling through the interstellar or intergalactic medium, which means that they can be linked to a specific source with only the small piece of their trajectory that we observe. As the energy of the γ-ray depends on the energy of the cosmic ray that created it, it is possible to study the energy output of these sources directly [10]. The downside of studying γ-rays is that there are very few of them compared to cosmic rays. Of all cosmic ray and γ-ray detections, only 0.01 percent or less is a γ-ray. Because both are detected in very similar ways, it is necessary to separate the γ-ray events from the CR events with great precision [1].

2.1.3 Air shower development

When a Cosmic ray or γ-ray enters our atmosphere, it will decay into charged particles with high energies and thus initiate a cascade of interactions, called an extensive air shower.
γ-rays induce electromagnetic showers, made from electrons, positrons, and photons. Electromagnetic air showers can be described with the Heitler model. In this model, there are two dominant production processes: The first is the creation of an electron and a positron from a photon by pair production, where both particles get the same energy, and the second is the production of a photon by an electron or positron through bremsstrahlung, where the electron or positron loses half of its energy. Additionally, the electrons can lose energy through ionization [11]. The first and second processes will repeat after the particle has traveled a specific distance λ_e called the depth, which is dependent on the radiation length λ_r, by the formula $\lambda_e = \lambda_r * ln(2)$ [12]. A simple illustration of this process is shown in Figure 1. This repeats itself until the particles reach a critical energy E_c, then the creation of more particles stops.

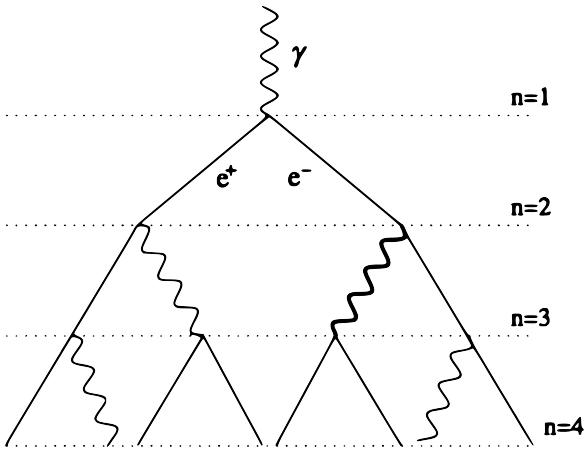

Figure 1: Heitler's model of electromagnetic shower development [13]

While this model is the most simplified version, it does incorporate the most important properties of the electromagnetic air shower: the particles that are produced, the dependence of the number of particles on the traveled distance x, $N = exp[x/\lambda_r]$, and when the shower approximately stops. We have to use this approximate model because real showers can contain more than 10^{10} particles, and our knowledge of reaction cross-sections and particle production is still insufficient to simulate the actual showers. Therefore these approximations are used to get a general idea of the electromagnetic shower development [12].

Cosmic rays induce air showers that have a much more complex development and are called hadronic showers. These hadron showers are too complex to describe analytically, but we can apply a generalized Heitler model to gain a better understanding of them. When a hadron interacts with our atmosphere, it produces a number of pions. Of these pions, two-thirds is charged and one-third is neutral. Each time the energy is distributed evenly, so two-thirds of the energy goes into the charged pions and one-third into the neutral pions. The charged pions will split up again after 1 mean interaction length into other pions, of which two-thirds are charged and one-third is neutral, as with the initial hadron. The neutral pions on the other hand quickly decay into 2 γ-rays: $\pi^0 \longrightarrow 2\gamma$. This means that hadronic showers have both a hadronic component and an electromagnetic component from the decaying neutral pions. These two processes repeat until the particles reach the energy E_{dec}, where the remaining charged pions decay into a muon [11].

Because these showers are so complex, it is not yet possible to create simulations that reproduce the shower we observe, and thus reverse-engineer the properties of the initial particle. Therefore, using what we know about the simulation to distinguish γ-ray-initiated air showers from cosmic-ray-initiated air showers is not a viable option. Still, there are a few important differences, that could be used to distinguish between the two air shower types. For instance, the secondary hadrons that are produced in hadronic showers get a transverse momentum, momentum perpendicular to the shower axis, of approximately 350 to 400 MeV, almost independent of the energy of the original particle. This causes low-energy hadrons to have a large angle with respect to the shower axis, which makes the hadronic showers have a larger lateral distribution than electromagnetic showers [11].
More details on the differences between γ-ray- and cosmic-ray-induced air showers are discussed in section 4.2.

2.2 γ-ray detection

2.2.1 Space-based detection

Instead of trying to find the initiating particle of an air shower, it is also possible to measure the γ-rays and cosmic rays directly in space. This way, it is also possible to detect the γ-rays that would otherwise be blocked by our atmosphere. Another advantage is that the calibration of a satellite can be performed in a laboratory because the atmosphere is not part of the detection.
The most important space-based γ-ray detector at the moment is Fermi-LAT. To detect γ-rays it depends on the pair-production of an electron and positron by the γ-ray within the detector. By measuring their energy, the energy of the initial photon can be constructed. These detectors still need to separate γ-ray events from CR events, which is done both by the satellite itself and afterward on the ground [14].
However, there are limitations to space-based detectors. Because the particle flux decreases logarithmically with energy, a larger detection area is required to realistically detect such a particle within a reasonable time limit. For instance, with a detection area of 1 m^2, you would measure 1 particle with an energy above 10 PeV in a year. Because of technical limitations, it is currently not possible to make a detector in space with a detection area suitable to observe γ-rays with energies higher than 1 PeV within a reasonable time frame. Therefore, ground-based observatories, which can cover a much larger area, have an essential role in detecting these high-energy particles [15].

2.2.2 Ground-based detection

Cherenkov light
As explained in section 2.1.3, γ-rays and CR interact with the atmosphere upon entry, creating air showers. One way to observe these air showers is through Cherenkov light. When a charged particle travels with a speed close to the speed of light in a vacuum through a medium with a refractive index, it is possible the particle travels faster than the photons in the medium. The charged particle can then momentarily polarise the medium [1], as a moving charge emits electromagnetic waves. When the medium returns to its ground state after the particle has passed, it re-emits the energy in the form of a photon. Because the photons created through this process travel slower than the particle, they form a cone behind the particle. the shape of this cone depends on the relative speed of the particle with respect to the photons [16].

Cherenkov arrays
While the charged particles in the shower decay, the Cherenkov light created by them does not disappear and can be observed on the ground. Because these showers only last for a very short time, they appear as a faint blue flash, and because they are so faint they are only visible on clear moonless nights. This means that the telescopes that observe Cherenkov light have a low-duty cycle, but they also have some strong advantages. Because the air showers occur high in the air, the Cherenkov light spreads around to a large area. A telescope in this area can catch this light, which functions as the fingerprint of the air shower. This way, the telescope makes a 'picture' of the part of the air shower that was visible from the telescope, as is shown in Figure 2. This fingerprint is very suitable for discrimination between γ- and cosmic rays [1] and can be used to derive other important properties of the particle, like the energy [17].

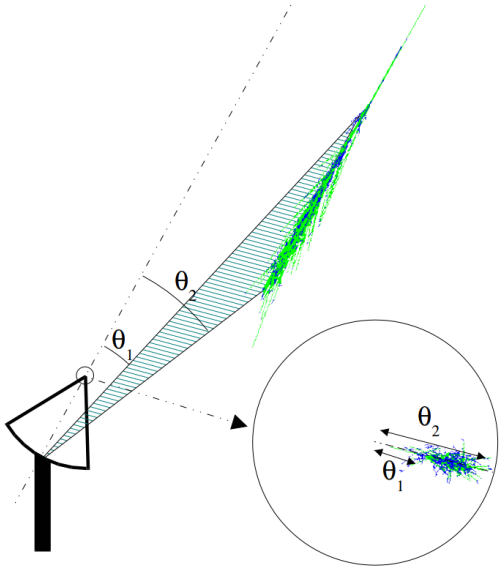

Figure 2: The Cherenkov light forms an image that is projected in the focal plane of the telescope [1]

2.3 The Cherenkov Telescope Array

2.3.1 The goals of CTA

The Cherenkov Telescope Array Observatory, or the CTAO, is a ground-based γ-ray observatory that is currently still under construction. The CTAO builds on the technology developed by previously made Cherenkov array telescopes, like H.E.S.S., MAGIC, and VERITAS, but is expected to be between 5 and 10 times more sensitive and have much higher accuracy. While the current arrays consist of five or fewer telescopes, the CTAO will consist of 60 telescopes divided between two locations, on the northern and southern hemispheres.
The goal of CTA is to investigate gamma rays with energies between 20 GeV and 300 TeV. This energy range has not been covered by any observatories so far and is expected to uncover more than a thousand new objects. Moreover, it will be possible to make these observations in greater detail than before. Because it consists of more telescopes, it can cover a much broader area, thus enabling a much quicker sky survey. With the help of these improvements, they hope to advance our knowledge in several fields.
Mainly, the objective of CTA is to gain more insight into the workings of cosmic accelerators, the sources of cosmic rays mentioned in 2.1.1. Specifically, they are interested in them because those objects are much better accelerators than those we can build on Earth, and played a role in shaping the universe. Other extreme regions, like voids, black holes, and neutron stars are also of interest to CTA. Additionally, they hope to learn more about dark matter, and through dark matter about the nature of space-time. In short, CTA hopes to advance the field of γ-ray astronomy on many frontiers, by providing a new energy range with higher accuracy than currently available [18].
A final important feat of the CTAO is that its data will be open to the scientific community. By using proposals from outside the ground-based γ-ray astronomy community, they wish to involve a broad research community and increase the scientific output [19].

2.3.2 Design CTAO

Now that we understand the goals of CTA, we can investigate the design of CTA in further detail.
The first objective of CTA is the energy range, from 20 GeV to 300 TeV. To achieve this, three different telescope types will be built: Large-Sized Telescopes (LSTs), Medium-Sized Telescopes (MSTs), and Small-Sized Telescopes (SSTs). [20]

The Large-Sized Telescopes have a mirror with a diameter of 23 meters and can detect air showers with energies between 20 GeV and 150 GeV. They are also able to accurately observe air showers with energies up to 3 TeV but are outperformed by the MSTs in this energy spectrum. Because of their large area, they can collect more light than the other telescope types and thus detect fainter flashes of Cherenkov light. The light collected by the mirror is reflected at the LST camera, which has a field of view of 4.3 degrees. Four LST telescopes will be built on the northern CTA location, and the foundation for another four in the southern hemisphere will be placed in preparation for a future expansion of the telescope array [21]. What the telescope will look like is shown in Figure 3.

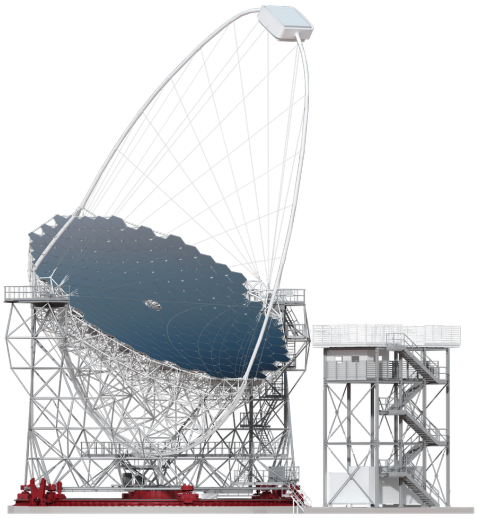

Figure 3: The design of the Large-Sized Telescope [21]

The Medium-Sized Telescopes will have a mirror diameter of 12 meters, and detect air showers with energies ranging from 150 GeV to 5 TeV, the core energy range of CTA. Two different types of cameras were designed for the MSTs, with similar properties. The first camera, the NectarCam, is similar to the LST camera and has a field of view of 7.5 degrees, and a slightly lower readout event rate than the second camera type, FlashCam, which has a field of view of 7.8 degrees. 14 MSTs will be built at the southern CTA location, and 9 will be built in the northern location [22]. What the telescope will look like is shown in Figure 4.

Figure 4: The design of the Medium-Sized Telescope [22]

The Small-Sized Telescopes will have the smallest mirror, and therefore the most sensitive to highly energetic air showers that produce the most amount of light. With a mirror diameter of 4.3 meters, it can detect air showers between 1 TeV and 300 TeV but performs best between 5 TeV and 300 TeV. Its camera has a field of view of 10.5 degrees, and all 37 telescopes of this type will be placed at the southern location [23]. What the telescope will look like is shown in Figure 5.

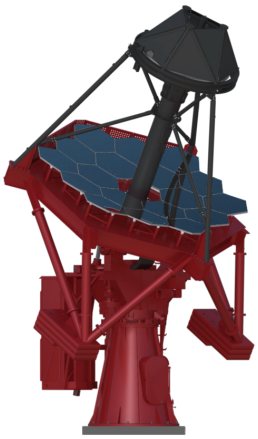

Figure 5: The design of the Large-Sized Telescope [23]

The second objective is to improve the accuracy of the observations. This is why the number of telescopes at both locations is going to be much higher than the current γ-ray telescope arrays have. This was achieved by increasing the collection area, increasing the number of telescopes, increasing the angular resolution with improved technology, and improving background rejection. The latter was achieved by having a high image multiplicity and is especially important for deep exposures used to study extended, low-surface brightness objects and low-flux objects [19].

The third objective is to advance the γ-ray astronomy field in a great variety of areas. This

is why the locations were chosen the way they were: to investigate all objects CTA is interested in, they needed to be able to survey both hemispheres. The search for dark matter and investigations of cosmic accelerators are focused on our galaxy and require high-energy shower detectors, which is why the SSTs are placed in the southern hemisphere. In the north, the galactic center is not visible, and therefore this location has its focus on extra-galactic targets. As the extra-galactic background light can suffer from gamma-gamma absorption, the reverse process of annihilation [24], it is the most efficient to use LSTs for these observations. Another important property of the location is its altitude.

Because the opaqueness of the atmosphere decreases with height, higher altitudes result in better images. On the other hand, lower altitudes give larger effective areas, an improvement of calorimetric capabilities, and better shower containment because the shower has developed further. This makes the optimal altitude around 1500 meters [1]. Therefore the CTAO will be built at an altitude of approximately 2100 meters [25].
The last important feat of the CTAO is its response time. All telescopes are capable of quickly changing their position and thus adjusting their target to any other target in the night sky in 20 seconds. This way, interesting events can be investigated within a short period of external alerts [19].

2.3.3 CTA simulation steps

To test whether the design of the CTAO satisfies all its requirements and works as expected, prototypes of the MST and LST have been built in 2012 and 2018 respectively [22] [21]. However, it is also interesting to test, given that the design satisfies the requirements, how the data that the CTAO will produce can be used, or which methods, like different machine learning types, can be used on the CTA data. For this reason, simulations were used and processed by CTA to create artificial data for the CTAO.

The development of extensive air showers in the atmosphere (including Cherenkov emission) is simulated using a general-purpose open-source software called CORSIKA [26].
The simulation of the arrays of imaging atmospheric Cherenkov telescopes and their response to the shower-induced emission is performed using Sim-telarray [27]. The output of Sim-telarray is then read and processed directly with ctapipe [28] a low-level data processing pipeline software for CTA (the Cherenkov Telescope Array).
In this work, we used the output of ctapipe, which consists of files in h5 format containing the following set of information, provided by our colleague Andres Delgado of the CTA Consortium:

- the simulated truth on the particle properties, such as the type of particle (gamma, proton, electron, etc.), and its simulated energy,
- the information on the events detected by the telescopes, such as the number of pixels which recorded a signal, the number of clusters of pixels in every image, the position of every telescope which recorded a signal, and more.
- The output of the reconstruction of the images on the cameras, providing the Hillas parameters, which will be discussed in section 2.4.

2.4 Hillas parameters

To distinguish between the two air shower types, it is common to use the hillas parameters. This is a set of features that were found to be suitable to distinguish γ-ray showers from cosmic ray

showers by A.M. Hillas in 1985 [29]. These parameters measure the position and general shape of the air shower image on the camera. Most of the parameters are measured with respect to the center of the image or a main axis that indicates the core of the shower. This main axis is determined by drawing a line from the start of the shower to the center of the image, as shown in Figure 6.

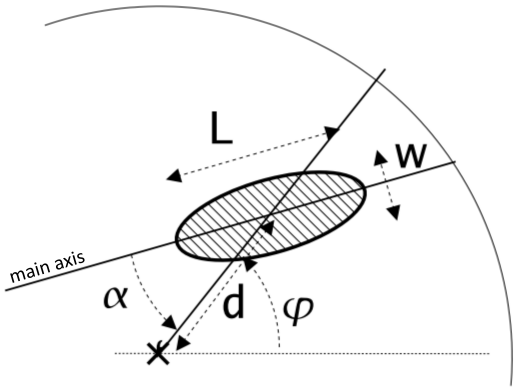

Figure 6: The geometrical definitions of hillas parameters, where w is the width, L the length, d the radius, ϕ the azimuthal angle of the image main axis, and α the orientation angle. In this report we will use ψ instead of α [1]

With hillas parameters, it is possible to give these images a simple geometrical description of their arrival direction and impact parameters [1].
They are useful because gamma-ray showers generally have different values for these parameters than cosmic-ray showers, making their distributions different. By exploiting this property, and sometimes by combining parameters, they can be used to accept or reject an observed air shower as a gamma-ray event [30].
The **hillas width** measures the spread of the light in the direction perpendicular to the shower main axis [29]. This mostly indicates the lateral development of the shower [31], which is, as explained in section 2.1.3, an important factor in distinguishing the two shower types.

The **hillas length**, on the other hand, measures the spread of light along the shower's main axis and is therefore related to the longitudinal development of the shower [31], which is how long the shower keeps developing after it is initiated.

Hillas phi is the orientation angle, the angle between the main axis and the line from the center of the image to the center of the ellipse. For showers initiated by gamma rays from sources on the image axis, it is expected that the hillas ϕ is smaller than 15 degrees. The showers initiated by cosmic rays, on the other hand, are expected to have a more even distribution, because all arrival directions are equally probable [31].

Hillas psi is the azimuthal angle of the image axis, the angle between the horizontal camera plane and the line from the center of the image to the center of the ellipse. This is mostly influenced by the positioning of the telescope array [32].

Hillas radius is the nominal distance between the image center and the center of the ellipse, which is also largely dependent on the telescope position. In combination with the size of the

shower, it can be used to estimate the shower energy [33].

While not mentioned in Figure 6, **Miss** is also one of the hillas parameters and is the minimal distance between the horizontal camera plane and the center of the ellipse [29]. In the simulated data, this parameter is called hillas X.

Besides the features mentioned in the image, more parameters were added later or gotten out of use over time.
Frac (2) is the amount of intensity in the 2 pixels with the highest measured intensity, relative to the amount of intensity that was measured in all pixels. The **Azimuthal width** measures the spread of the light in the direction perpendicular to the line from the telescope image center to the center of the ellipse [29]. Because they are not available in the simulated data set, it is assumed that they have gotten out of use.
When the shower is partially out of view, it cannot simply be treated as an ellipse as this would give degenerate results. To solve this, **hillas skewness** and **hillas kurtosis** were added to hillas parameters [33]. Hillas skewness is the distance between the ellipse center and the pixel with the highest intensity [30] and indicates the asymmetry of the shower. Hillas kurtosis, on the other hand, measures the tailedness of the ellipse, which means it is a measure of the number of outliers [34].

3 Methods

As mentioned in 2.3.3, CTA made simulated events for the future CTAO. In this research, we will use these simulated events to test the effectiveness of BDTs. First, we will explain on what criteria the method will be judged, then explore the simulated data in further detail, and then how the BDT works. Finally, the features of the air shower that are used are evaluated to see if they make sense from a theoretical perspective.

3.1 Criteria

Before we explain the methods we used to separate γ-rays from cosmic rays, it is important to explain some relevant terms and how we will evaluate the methods.
Gamma rays constitute our signal, while cosmic rays, mainly made of protons, constitute our background. Electrons and positrons also constitute a relevant source of background for gamma-ray astronomy, but electrons are about a hundred times less abundant than protons at the TeV energies and positrons are about 10000 times less abundant than protons at TeV energies. For this reason, the scope of our work is limited to gamma-proton separation, and we did not take electrons and positrons into account.

Our methods will take the data sets of both background and signal events, and predict which of the two they think each event is based on the camera image. To describe the performance of our classification method, we use a confusion matrix, as it is usually done to assess the outcome of classification problems. A confusion matrix contains two rows and two columns that report the number of true positives, false negatives, false positives, and true negatives.
True Positives are actual signal events that are correctly identified as such, True negatives are actual background events that are correctly identified as such, False positives are actual background events that are wrongly identified as signal events, and False negatives are actual signal events that are wrongly identified as background events [35].

In the confusion matrix, we show how much percent of the true event type is classified correctly, so the number of true positives and false negatives is divided by the total number of signal events, while the number of false positives and true negatives is divided by the total number of background events that were classified. Figure 7 shows the basic structure of a confusion matrix.

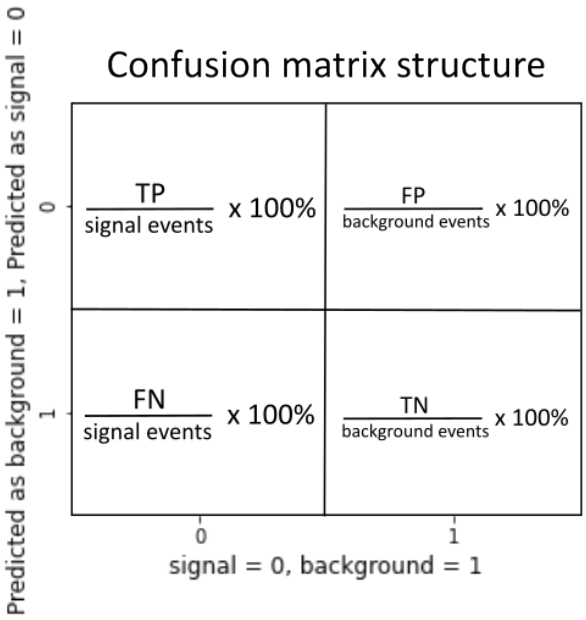

Figure 7: The basic structure of the confusion matrix as we will use it to display our results.

A perfect separator would classify every event correctly and would have 100 percent in the true positive and true negative sections. However, it is more important to have as few false positives as possible: because most research will mainly use the signal events, we want to be as sure as possible that the events we predict to be signal events are signal indeed signal events. We measure this through the precision: namely the ratio between the number of true positives and all positively identified events. So,

$$precision = \frac{TP}{TP + FP}$$

3.2 Used data

3.2.1 Feature distributions

The simulated data consists of a great number of γ-ray and cosmic ray events. Because most of the cosmic rays are protons, this investigation only used protons as air shower initiators for the background events. Each of these events has one or more images, depending on how many telescopes would have observed the simulated air shower. In total, 39 features collectively describe the position and size of the air shower image on the camera, its intensity and intensity morphology, and the time evolution of the shower.
Each of these features is the product of shower development, and therefore possibly influenced by the shower type, hadronic or electromagnetic. We cannot draw conclusions based on a single event, but if the feature is dependent on the shower type, this will be visible in a shift in distribution between proton- and photon-induced showers, as shown in Figure 8.
For instance, as the lateral distribution is often smaller for electromagnetic showers, it is expected that the width of a shower on the image is generally higher for the proton showers than for the gamma-ray showers [30].

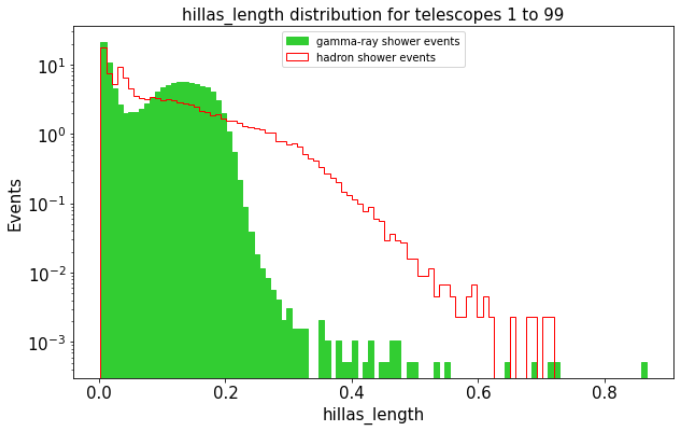

Figure 8: A histogram of the standard deviations of the intensity for both the proton and gamma-ray showers, with the hillas length on the x-axis and the number of events with that hillas on the y-axis (in log scale). This includes events observed by all telescopes

Generally, the preferred distribution would be one where the two distributions have their peaks at distinct locations, as this would make it easier to distinguish signal events from background events, although it can be enough to have a different distribution shape to be able to use the feature to make guesses that would result in better accuracy than random guessing. However, as we will see, even features that do not show this relation can be useful when combined with other parameters.

3.2.2 Event information

The data does not only contain the value of each feature for each event but also the telescope ID, event ID, and what the initiating particle was.
With the event ID, it is possible to 'track' events that were observed by more than one telescope. This way, more data is available for that air shower which ought to make the methods used to determine the initiating particle more accurate.
That we know the true initiating particles of each event is very important to test the effectiveness of our methods and in training the BDT. While the γ-ray and proton air shower simulations are stored in different datasets, during the testing they are mixed. Of course, they are not used for discrimination by the methods, as this information would not be available for non-simulated data.

To keep track of what telescope type recorded the event, we use the telescope ID. This is useful because each telescope type has slightly different camera properties, like the number of photo-multipliers, their field of view, and their energy range. These factors result in stark differences for several feature distributions, as is shown in figure 9.

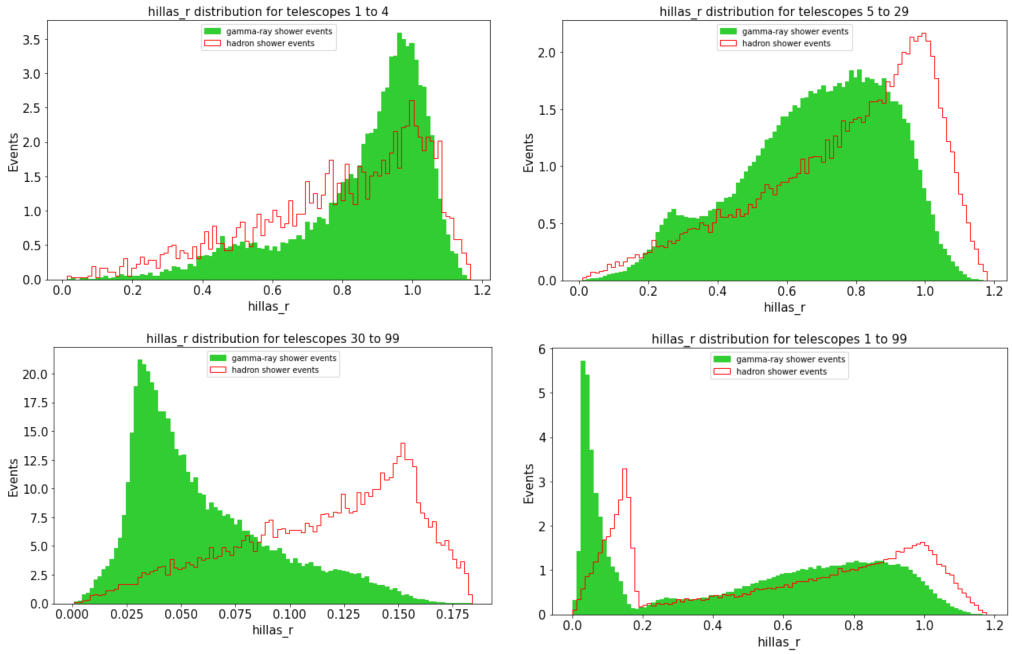

Figure 9: The distributions of hillas radius on different energy ranges. The figures contain histograms of the value of the hillas radius for the different telescope types.
Top left: figure 9A, LTS data, with energy range 20 GeV to 3 TeV.
Top right: figure 9B, MST data, with energy range 150 GeV to 5 TeV.
Bottom left: figure 9C, SST data, with energy range 5 TeV to 300 TeV.
Bottom right: figure 9D, all telescope types, with energy range 20 GeV to 300 TeV.
There is a very clear shift in the distributions between the telescopes. Not only the x-axis of the SST data is different, but also the separation between the peaks changes significantly. This can be caused by the differences in telescope design, or the different energies of the events they observe. Because the peak is in different locations for some telescope types, two peaks appear in the distribution that includes the data of all telescopes.

This means that while a feature might be useful in one energy range, with two clear distinct peaks in the distribution as in figure 9C, it can be almost useless in another, which almost happens in figure 9B. This is also the reason a feature can have more than one peak in a single distribution when using the data from all telescopes at once, see for example Figure 9D. Because the telescopes are hard to compare because of their differences besides their energy range, we decided to keep using all telescope data for the final results. However, we will also check the features that work well for specific telescope types, to better understand how the BDT works.

Test and train data

The methods that were used to classify the data, both required training by a similar data set of which the initial particle was known before they were tested with data of which they did not know the initial particle. Therefore, the events were randomly divided into a train and test data set. That this happened randomly was of great importance, because both data sets need to have similar distributions for the methods to be accurate. This was achieved by using the model-selection.train-test-split module from sklearn. With this method, the data was split into two sets, with 75 percent of the events in the train data set and 25 percent in the test data set. Both the γ-ray and cosmic ray events were distributed evenly over both data sets, and the

information of the event type was stored separately to be able to use it in training and checking classifications made by the methods. As shown in figure 10, this method is suitable for splitting the data as the distributions are similar enough.

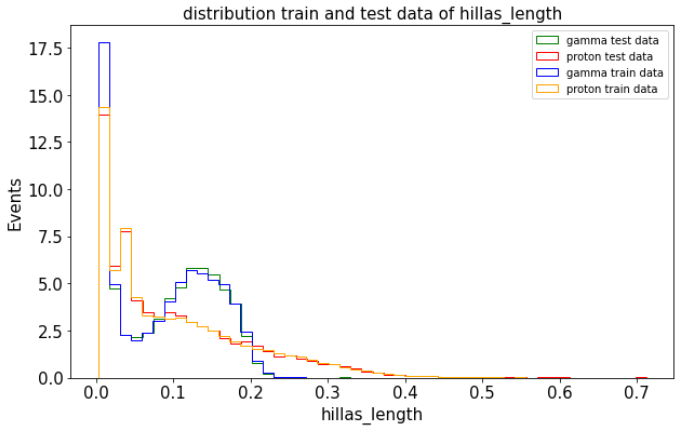

Figure 10: The distributions of both the γ-ray and cosmic ray air shower event data of the hillas length, for the train and test dataset. As expected, the distributions are very similar for the train and test data sets. A similar relation is seen in the other features.

3.3 The Boosted decision tree

One of the methods used to analyze the data was the Boosted Decision Tree (BDT), a machine-learning method that uses characteristic information of an object or event to classify it into a pre-determined object or event type, using the information of other objects or events it does know the type of. In the following chapters, we will explain how they work, why they are suitable for this project, and what the specific settings were of the BDT that we used.

3.3.1 How BTDs work

BDTs work with a method very similar to flow charts, a common way to give something a classification based on its collection of distinctive properties. At its basis, flow charts are built from a series of decision nodes and leaf nodes, which are small individual steps in the categorization process, depicted as circles in Figure 11. Each decision node tests a property of the data and splits the events based on that property into two categories. A category might for instance be having a specific feature with a value below a certain limit. The events then each go to a new decision node or leaf node based on their categorization. The series of decision nodes are called the branches of the decision tree, and the very first decision node is called the root. Each branch eventually ends in a leaf node: this is the final categorization of the decision tree. A simple example of a decision tree is shown in figure 11. In the case of our decision tree, this final categorization will either be '0' for signal events or '1' for background events [36].

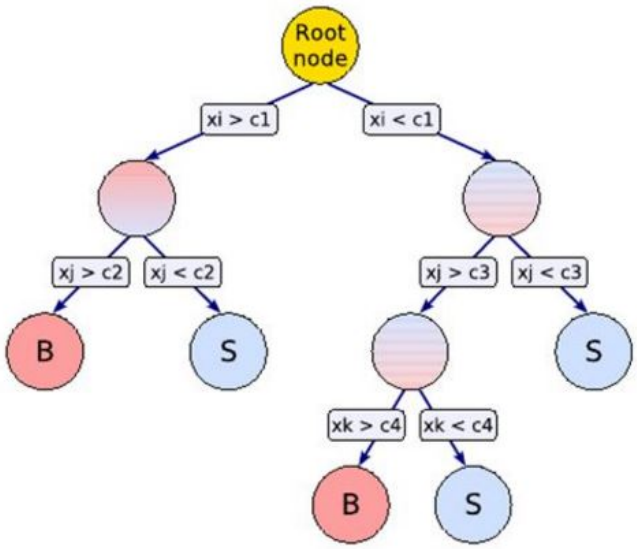

Figure 11: A very basic example of a decision tree where a classification is given based on a series of properties. Here, these properties are called xi, xj, and xk, and are categorized on whether the value of these parameters is above or below the limits c1, c2, and c3. The BDT uses comparable structures for its weak learners [37]. This decision tree is not related to the BDTs used in this investigation, apart from its basic structure.

It is possible to build a tree that perfectly fits the training data, but this will result in poor performance on the test data because the training data is slightly different from the test data. Making the algorithm fit the training data too well, as shown in figure 12 is called overfitting the data. To build a tree capable of categorizing events with high accuracy, machine learning methods employ ensemble learning, a method that uses many different models and combines them. This results in a decision tree that is flexible enough to be accurate for both the train- and test data set.
The BDT uses an ensemble learning method called 'boosting'. The tree starts at its root with a weak learner, which means the separation is only slightly more accurate as a random guess. These trees usually only consist of a few decision nodes at most. Then, another weak learner is made and added to the tree to improve it. The tree is placed on a branch that is the worst at categorizing or has residential events, thus improving the parts of the tree that needed the most improvement. This process is repeated many times, thus creating a strong learner from pieces of weak learners [36].

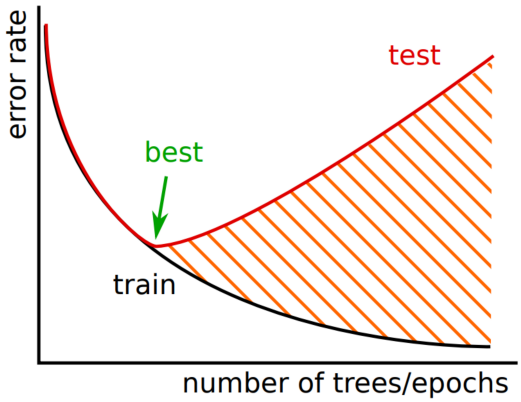

Figure 12: Graph of error rate against the number of improvement rounds, for both the training data and test data. While the classification of the training data keeps improving with the number of trees, the classification of the test data starts rising again due to overfitting [38].

Not only the type and amount of data are of importance for the performance of the BDT. There are some parameters and restrictions that can be applied to the BDT, called hyperparameters [36]. These will be explained in chapter 3.3.3.

3.3.2 Why BDTs

While we used the BDT as a machine learning method, there are other machine learning methods available that are capable of making categorizations based on data. This chapter will discuss these alternatives and explain why the BDT method was best for this investigation.

The machine learning method that is most like the BDT is the Random Forest method, but Neutral Networks, likelihood estimators, Fisher discriminants, and linear discriminant analysis are also suitable for this task. There are also other types of machine learning algorithms like Support Vector Machines and wavelet-based denoising, but those are more suitable for images with more distinctive properties [39].
Linear discriminant analysis is a method that tries to find linear relations that maximize data separation but is notably slower than the BDT method. Moreover, linear methods, which include Likelihood estimators and Fisher discriminants, assume a linear relation between the input parameters and the category, while this is not necessarily true for the features of this project. They, therefore, lack the flexibility that BDTs and Neutral Networks do have [40].
Neural Networks also work with decision nodes, but instead of forming a tree, all nodes are connected in series with an assigned weight. The sum of outcomes then determines the classification [41]. The advantage of BDTs compared to Neural Networks is that Neural networks cannot ignore features even when they have no separation value while BDTs can, so they don't suffer from that extra liability.
Finally, the Random Forest is a method very similar to BDTs, but instead of boosting their trees, they use a method called bagging. This means that instead of building a large tree in series, a lot of small trees are used in parallel. The classification most trees agree with when applying all trees is then used as the final classification. The downside of this is that this method is much slower than the BDT method. Moreover, because the random forest is less susceptible to overfitting, BDTs tend to be more accurate as random forests [42].

In conclusion, BDTs are a method with a short run time, which is flexible enough to handle the data we analyze and is still very accurate.

3.3.3 Implementation of the BDT

To implement the BDT, we used the module xgboost. This module takes the data and parameters you want to use and makes a fitting BDT. It was chosen because it was easy to implement, suitable for Python but also faster and more accurate than the other methods [43]. Moreover, it has proven to be useful by Patrick Luijks for the separation of iron- and proton cosmic rays for the CTAO [44].

For xgboost to be able to use the train data to make a BDT, it needs to know of each event whether it belongs to a γ event or proton event. As mentioned in 3.2.2, this information could be stored separately by using the model-selection.train-test-split module.
When the BDT is processing the training data, we often restrict the BDT in some ways. This way we can limit the processing time and/or increase its accuracy. For this research, we restricted the number of estimators, learning rate, base score, the maximum number of leaves, maximum depth, minimum child weight, and minimum loss reduction.

When the process of improving the previous tree is repeated too often the BDT will start overfitting the data, while if it is repeated not often enough the tree is not as efficient as it could be. This can be regulated by changing the number of trees, or the **number of estimators,** we allow the algorithm to make [36].
Another important factor is how heavy the alterations to the tree at each boost are: heavier improvements make for a steeper learning rate, so fewer trees are required to achieve higher accuracy at the start. However, by using smaller improvements the tree becomes more robust and can eventually reach a higher peak efficiency, at the cost of needing more improvement steps. We manage this with the **learning rate**, which determines a weight shrinkage at each new improvement. This way, overfitting is also prevented.
The BDT estimates the likelihood an event is a signal or background event. It is possible the chance an event is a background event can be different from the chance an event is a signal event, without even considering the data, because the number of signal events and background events can be different. This can be taken into account with the **base score**, which is somewhere between 0 (signal) and 1 (background) [45]. In this investigation, an equal number of signal and background events were used, but in reality, there will be 10000 times more background events than signal events [1]. The optimal base score we find will therefore be very different from the one that CTA should use when it is finished.
The tree can grow very large if not kept within certain limits. Therefore, the number of final nodes can be limited by restricting the **maximum number of leaves**.
The complexity of the tree depends partially on the number of decision nodes in the branches of the decision tree. As the number of nodes increases exponentially with the length of the branches, increasing this length also takes a heavy toll on the processing time of the BDT. Therefore, a **maximum depth** is applied, which restricts the complexity of the tree.
When a leaf has a sum of instance with a weight lower than the set **minimum child weight**, the BDT stops trying to improve that leave by adding other decision trees.
Another way of stopping the BDT from improving a branch further is by setting a minimum on the amount the new branch improves the tree. This stops the BDT to become more complex unnecessarily and is called the **minimum loss reduction** [45].

By optimizing the BDT to fit the data sets we use, the following hyperparameters were found:

- Number of estimators = 160
- Learning rate = 0.23
- Base score = 0.74
- Maximum number of leaves = 30
- Maximum depth = 7
- Minimum child weight = 1
- Minimum loss reduction = 0.76

It should be noted that the Maximum depth is at the limit we set manually to limit processing time.

3.4 Used features

In this section we will first explain in further detail how the outcome of the methods will be judged, and why. Then we can dive further into the features we will use to get the best out of both methods and what these features mean.

3.4.1 Categories

To gain a further understanding of the BDT, we will find its highest precision, both in the optimal situation where it has access to all data and in the sub-optimal situation where some of the data is restricted. This way, we want to learn what parts of the data are most effective for the CTAO. Its performance of the BDT will be compared specifically in two situations: when it uses all 39 available features and when only 4 features can be used.
The first situation is the situation where the BDT functions best, while the second situation should give an idea of both the most effective features and how well the BDT performs with relatively little data. This way, we can compare how well the method works far away from its peak performance.

Ideally, we would find the performance of the BDT for the complete range of features, because this would show how the BDT improves as a function of feature number, but this would take too long for this research. We did, however, have enough time to find the best precisions of the BDT when it had access to 1, 2, 3, 4, 38, and 39 features, which will hopefully provide a clear enough image of the dependence of the precision on the number of features. While the optimal features are found for these numbers of permitted features, only the features found when using 4 and 39 features will be discussed in detail.

3.4.2 Feature selection

As mentioned before in Section 3.4.1, it was not possible to check every possible combination of four features, because the number of possible combinations increases significantly with the number of features in the pool you draw from. We would have liked to try these combinations because even when the features do not have a difference in distribution between signal and background events, they could have very useful relations with other features. To limit the number of

possible combinations, we performed a selection process. With this, we excluded features until the pool thereof was small enough to try all possible combinations in the pool.

We looked at the features that had the biggest difference between the distribution of signal and background events, both when the data of all telescopes were used and when the data of only a single telescope type was used. This way, we would be able to see if the BDT rather uses features per telescope type or rather features that work well for all telescopes. This resulted in a pool of 20 features that were suitable for separating signal and background events. Running all combinations of four features would still take too long because this would entail running the BDT $\frac{20!}{4!(20-4)!} = 4845$ times. Therefore, we made one even smaller pool of tree features, of which at least one needed to be included. These three features were chosen by running the BDT for all 40 features except 1: the three features that gave the worst result when left out were chosen. Now, we only had $3 * \frac{19!}{3!(19-3)!} = 2907$ combinations, which was low enough to run.

Following this procedure, we found these features for the individual telescopes:
For the SSTs, we found hillas ϕ, hillas X, hillas ψ, and the slope of the timing.
For the MSTs, we found hillas radius, hillas ϕ, hillas ψ, and the slope of the timing.
For the LSTs, we found the standard deviation of the peak time, the hillas radius, the standard deviation on the intensity, and the hillas width.

But, the features that are used when using the data from all telescopes matter the most: hillas ϕ, hillas ψ, the intensity leakage over the width, and the slope of the timing. The effectiveness of the BDT is also slightly dependent on the small difference between the training en test data sets, and therefore its final precision is partially dependent on chance. Because the top qualifiers for the best combinations were all very close, these four might not be the absolute best four features. However, the other feature combinations that came close to the precision of these feature combinations were very similar combinations, where only one or two features were swapped for another feature. Thus, while these features are possibly not the absolute best combination, they are close enough to be used in this investigation.

4 Results

Using these features and BDT settings, we estimate its performance. First, we will show how the precision of the BDT changes as a function of the number of parameters it can use. Then, we will look at the confusion matrix in the two situations and discuss if the result is satisfactory.

4.1 Precision as a function of feature number

As mentioned in the section above, it was not possible to find the maximum performance for all number of features. Therefore we only look at its maximum performance with a few features, with one feature missing, and with all features. The results of this are shown in Figure 13.

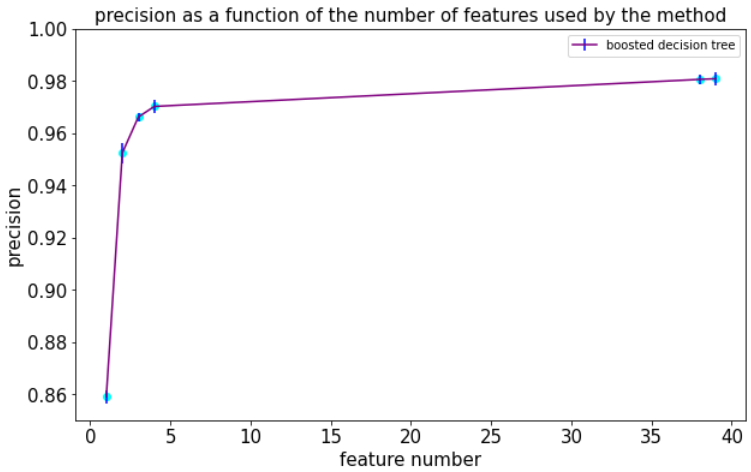

Figure 13: The maximum precision of the BDT method as a function of the number of features it uses. With each feature, the BDT becomes more effective

While the results are relatively consistent, the BDT does perform significantly less well when it only has access to one feature, although even then the precision is very high with a value of 85,9 percent. This makes sense: When there is only one feature, the BDT cannot make connections between features, which is one of its great strengths.
Not only the precision is high, but also the number of events it correctly identifies as a signal is satisfactory, both when it uses 4 features and when it can use all features, with 97,0 and 98,2 percent of the actual signal events identified as such, as shown in Figure 14 and Figure 15.

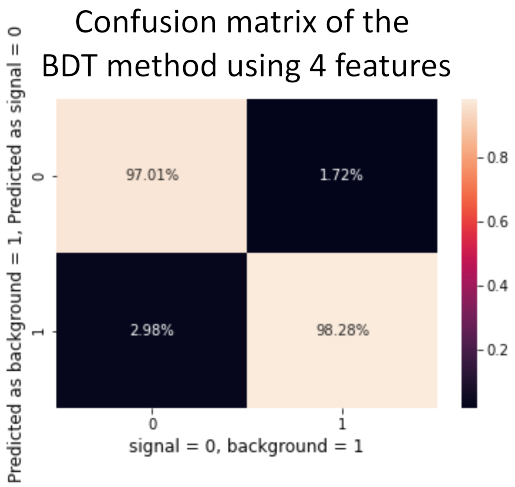

Figure 14: The confusion matrix of the BDT when it uses 4 features. The true positive and false negative rates (top and bottom left) are expressed as the number of events in that category, divided by the total number of signal events. The true negative and false positive rates (bottom and top left) are expressed as the number of events in that category, divided by the total number of background events.

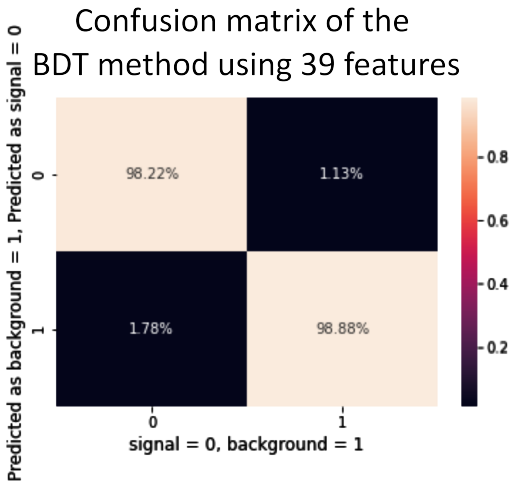

Figure 15: The confusion matrix of the BDT using all 39 features. The true positive and false negative rates (top and bottom left) are expressed as the number of events in that category, divided by the total number of signal events. The true negative and false positive rates (bottom and top left) are expressed as the number of events in that category, divided by the total number of background events.

4.2 Evaluation features

While we found 4 features, we also wanted to see what the physical importance of these features was.
The intensity leakage over the width is the fraction of the intensity in the border of one or two pixels wide on the edge of the shower image. We know the background events have less regular intensity distributions on the image [30]. It makes sense that the signal events, that have a clearer outline and a more well-defined core have less intensity on their edges. The irregularity of the hadronic showers could both put more intensity on the edge of the shower and increase the number of pixels on the outer side of the shower image. This is also recognized in the distribution, as shown in Figure 16. Therefore it is reasonable to believe the intensity leakage over the width can be used to distinguish between signal and background events.

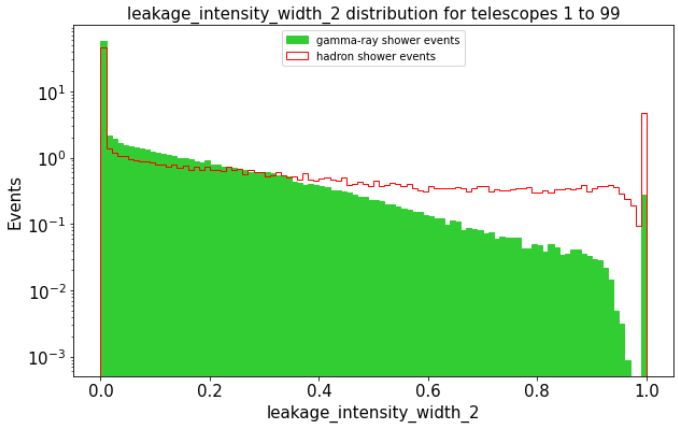

Figure 16: A histogram of the intensity leakage for both the signal and background events, with the intensity leakage on the x-axis and the number of events with that intensity leakage on the y-axis (in log scale). This includes events observed by all telescopes

The slope of the timing is found by fitting a line through the data points of the arrival times along the main axis of the shower [34]. A low slope, therefore, means that the shower has very short arrival times along the main axis, while a large slope means that the arrival times are relatively high at the end of the shower, so the shower would move forward slower. As the speed of the particles is dependent on the energy of the particle, and hadronic showers lose their energy quicker because it is distributed over more particles, it makes sense that the signal events have a higher peak for the low timing slopes, which we see in Figure 17. So we can conclude that this feature could indeed be useful in separating signals from background events.

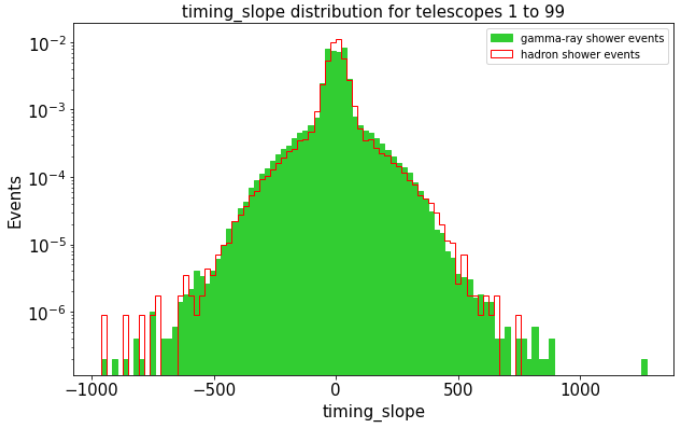

Figure 17: A histogram of the slope of the timing for both the signal and background events, with the slope on the x-axis and the number of events with that slope of the timing on the y-axis (in log scale). This includes events observed by all telescopes

The meaning of the last two features, hillas ϕ and hillas ψ, were already treated in section 2.4. We saw that the hillas ϕ was the direction of the shower, and could be used for gamma rays with their source on the telescope axis. The hillas ψ determines the position of the center of the air shower with respect to the telescope axis. Although the hillas ψ does not seem to be very useful on its own, it turned out to be very useful when it was combined with hillas ϕ, as these were also the two features that were found to be most effective when only 2 features could be found. Possibly, it enhanced the determination of the source of the shower, or the data of all telescopes could be leveraged better by combining these features. Further research would be required to determine why the combination of these features is so useful, but it is reasonable to believe their combination is useful because of the individual use of hillas ϕ, and the mutual relation to the telescope axis.

When looking at the 4 features leading to the highest precision for the BDT, we can compare them with the features that give the optimal result when only one telescope type is used.

Both for the SST and MST data, the hillas ϕ, the hillas ψ and the slope of the timing were good separators, although their fourth feature differed. The LST data however preferred very different features. Comparing them to the features used for all features, hillas ϕ, the hillas ψ and the slope of the timing return as important features, while the leakage of the intensity over the width is not the most important feature for any of the individual telescope types.

This indicates the relation between hillas ϕ, hillas ψ and the slope of the timing is a good separator for most telescope types, mostly for the MSTs and SSTs, but the relation is also effective enough for the LSTs, otherwise, these features would be so useful when using all telescope types.
We also see that the features that are used for the individual telescopes are not always the same. Possibly, the relation between the features is energy-dependent, or the difference might be caused by the difference in camera type, although this seems less likely because the camera used by LST is similar to one of the camera types of the MST.

5 Conclusion

Is the Boosted Decision Tree a suitable method for distinguishing between air showers initiated by γ-ray and those formed by the more abundant cosmic rays? To address this question, we built, trained, and tested the performance of a BDT using simulated events detected by the upcoming Cherenkov Telescope Array, an upcoming observatory that will be the world's largest ground-based gamma-ray project with one order of magnitude improvement in sensitivity, extended energy and sky coverage with respect to current generation experiments.

We used 50646 simulated showers induced by gamma rays and 50646 simulated showers simulated by cosmic-ray protons. These showers are initiated by primary particles with energies between 20 GeV and 300 TeV, which is the range explored by CTA. A colleague of our group provided the simulated events. They included 39 variables related to the position of the signal on the cameras, the corresponding shape, and the topology of the event. We used these features as input for our Machine Learning method we developed from scratch, using as a reference the work done by a former student of the CTA group at Kapteyn Institute and the xgboost tool.

We tested the BDT performance for different numbers of features used and found that even with one feature, the precision of our method is better than 85 percent. The highest precision of 98.2 percent was found when using 39 features.

This ability arises from the BDTs capability to combine features rather than treating them individually. This investigation primarily confirms the suitability of the BDT for the Cherenkov Telescope Array (CTA), aligning with previous successful applications on similar telescope arrays [2][3]. In addition, our contribution to the field includes identifying optimal settings for the BDT method's suitability for CTA. Additionally, we identified potentially valuable features for discriminating between signal and background events.

Research following up on this research could investigate how other machine learning algorithms, such as Deep Learning Methods or Convolutional Neural Networks could be used for classification, possibly resulting in a higher precision or further confirming the BDT use.

6 Acknowledgements

Writing this dissertation did not come without its obstacles, which I would not have been able to overcome without the help of others. First of all, I want to thank my supervisor, Manuela Vecchi, for guiding me in the right direction and warning me when I started to walk astray. Secondly, I want to thank Sofia llacer Caro and Patrick Luijks, who supported my thesis with theirs. I also want to thank my parents and family, for supporting me throughout the years so that I had the time and space to study. I want to give special thanks to Niall Foley and Giacomo Nijland, for always providing kindness and advice when I needed it. Then, I also want to thank Lieke Middag, for her presence in the Kapteynborg in the first weeks, which motivated me to start working from the very beginning, and everyone that came to the Kapteynborg in the following weeks, where they provided the atmosphere that was so suitable for writing this dissertation.
Besides those that helped me directly, there were a lot of people that helped me indirectly: family, friends, and good acquaintances that I cannot all name explicitly, but that always cheered me up and gave me the sense of fulfillment that I needed to try my very best the last few weeks. Finally, because I wish to remember this for years to come, I want to mention Aka Akasaka for writing the stories that made me look forward to the next Wednesday every week and provided some welcome distraction when I needed a break.

References

[1] M. de Naurois and D. Mazin. "Ground-based detectors in very-high-energy gamma-ray astronomy". en. In: *Comptes Rendus Physique*. Gamma-ray astronomy / Astronomie des rayons gamma 16.6 (Aug. 2015), pp. 610–627. ISSN: 1631-0705. DOI: 10.1016/j.crhy.2015.08.011. URL: https://www.sciencedirect.com/science/article/pii/S1631070515001462.

[2] Y. Becherini et al. "A new analysis strategy for detection of faint -ray sources with Imaging Atmospheric Cherenkov Telescopes". en. In: *Astroparticle Physics* 34.12 (July 2011), pp. 858–870. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2011.03.005. URL: https://www.sciencedirect.com/science/article/pii/S0927650511000673.

[3] M. Krause, E. Pueschel, and G. Maier. "Improved /hadron separation for the detection of faint -ray sources using boosted decision trees". en. In: *Astroparticle Physics* 89 (Mar. 2017), pp. 1–9. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2017.01.004. URL: https://www.sciencedirect.com/science/article/pii/S0927650517300166.

[4] C. Grupen. *Astroparticle Physics*. en. Undergraduate Texts in Physics. Cham: Springer International Publishing, 2020. ISBN: 978-3-030-27341-5 978-3-030-27339-2. DOI: 10.1007/978-3-030-27339-2. URL: http://link.springer.com/10.1007/978-3-030-27339-2.

[5] A. Mitchell. *Status of Ground-based and Galactic Gamma-ray Astronomy*. July 2021. URL: https://pos.sissa.it/395/046/pdf.

[6] R. Ong. "Very high-energy gamma-ray astronomy". en. In: *Physics Reports* 305.3 (Nov. 1998), pp. 93–202. ISSN: 0370-1573. DOI: 10.1016/S0370-1573(98)00026-X. URL: https://www.sciencedirect.com/science/article/pii/S037015739800026X.

[7] A. Obermeier et al. "THE BORON-TO-CARBON ABUNDANCE RATIO AND GALACTIC PROPAGATION OF COSMIC RADIATION". en. In: *The Astrophysical Journal* 752.1 (May 2012). Publisher: The American Astronomical Society, p. 69. ISSN: 0004-637X. DOI: 10.1088/0004-637X/752/1/69. URL: https://dx.doi.org/10.1088/0004-637X/752/1/69.

[8] A. Ćiprijanović. "Galactic cosmic-ray induced production of lithium in the Small Magellanic Cloud". en. In: *Astroparticle Physics* 85 (Dec. 2016), pp. 24–28. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2016.09.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0927650516301311.

[9] R. Leane et al. *Snowmass2021 Cosmic Frontier White Paper: Puzzling Excesses in Dark Matter Searches and How to Resolve Them*. arXiv:2203.06859 [astro-ph, physics:hep-ex, physics:hep-ph]. Mar. 2022. URL: http://arxiv.org/abs/2203.06859.

[10] L. Tibaldo, D. Gaggero, and P. Martin. "Gamma Rays as Probes of Cosmic-Ray Propagation and Interactions in Galaxies". en. In: *Universe* 7.5 (May 2021). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 141. ISSN: 2218-1997. DOI: 10.3390/universe7050141. URL: https://www.mdpi.com/2218-1997/7/5/141.

[11] R. Engel, D. Heck, and T. Pierog. "Extensive Air Showers and Hadronic Interactions at High Energy". In: *Annual Review of Nuclear and Particle Science* 61.1 (2011). _eprint: https://doi.org/10.1146/annurev.nucl.012809.104544, pp. 467–489. DOI: 10.1146/annurev.nucl.012809.104544. URL: https://doi.org/10.1146/annurev.nucl.012809.104544.

[12] J. Matthews. "A Heitler model of extensive air showers". en. In: *Astroparticle Physics* 22.5 (Jan. 2005), pp. 387–397. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2004.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0927650504001598.

[13] A. De Angelis and M. Pimenta. *Introduction to Particle and Astroparticle Physics*. Undergraduate Lecture Notes in Physics. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-78180-8 978-3-319-78181-5. DOI: 10.1007/978-3-319-78181-5. URL: http://link.springer.com/10.1007/978-3-319-78181-5.

[14] S. Funk. "Ground- and Space-Based Gamma-Ray Astronomy". In: *Annual Review of Nuclear and Particle Science* 65.1 (2015), pp. 245–277. DOI: 10.1146/annurev-nucl-102014-022036. eprint: https://doi.org/10.1146/annurev-nucl-102014-022036. URL: https://doi.org/10.1146/annurev-nucl-102014-022036.

[15] L. Baldini. *Space-Based Cosmic-Ray and Gamma-Ray Detectors: a Review*. arXiv:1407.7631 [astro-ph]. Aug. 2014. DOI: 10.48550/arXiv.1407.7631. URL: http://arxiv.org/abs/1407.7631.

[16] *An Introduction to Cherenkov Radiation*. URL: http://large.stanford.edu/courses/2014/ph241/alaeian2/.

[17] K. Kawata et al. "Energy determination of gamma-ray induced air showers observed by an extensive air shower array". en. In: *Experimental Astronomy* 44.1 (Oct. 2017), pp. 1–9. ISSN: 1572-9508. DOI: 10.1007/s10686-017-9530-9. URL: https://doi.org/10.1007/s10686-017-9530-9.

[18] *Home*. en-US. URL: https://www.cta-observatory.org/.

[19] The CTA Consortium. *Science with the Cherenkov Telescope Array*. English. Accepted: 2021-02-12T03:02:43Z. World Scientific Publishing Co., 2019. ISBN: 978-981-327-009-1. DOI: 10.1142/10986. URL: https://directory.doabooks.org/handle/20.500.12854/58998.

[20] *Technology*. en-US. URL: https://www.cta-observatory.org/project/technology/.

[21] *LST - Cherenkov Telescope Array*. URL: https://www.cta-observatory.org/project/technology/lst/.

[22] *MST*. en-US. URL: https://www.cta-observatory.org/project/technology/mst/.

[23] *SST*. en-US. URL: https://www.cta-observatory.org/project/technology/sst/.

[24] L. Dreyer and M. Böttcher. *Gamma-Gamma Absorption in Gamma-Ray Burst Environments*. arXiv:1704.07232 [astro-ph]. Apr. 2017. DOI: 10.48550/arXiv.1704.07232. URL: http://arxiv.org/abs/1704.07232.

[25] information@eso.org. *Cherenkov Telescope Array Observatory*. en. URL: https://www.eso.org/public/teles-instr/paranal-observatory/ctao/.

[26] D. Heck. "Introduction to CORSIKA and Historical Review". en. In: (2014).

[27] K. Bernlohr. "Simulation of Imaging Atmospheric Cherenkov Telescopes with CORSIKA and sim_telarray". In: *Astroparticle Physics* 30.3 (Oct. 2008). arXiv:0808.2253 [astro-ph], pp. 149–158. ISSN: 09276505. DOI: 10.1016/j.astropartphys.2008.07.009. URL: http://arxiv.org/abs/0808.2253.

[28] *Prototype CTA Pipeline Framework (ctapipe) — ctapipe v0.19.3.dev47+g3f217808*. URL: https://ctapipe.readthedocs.io/en/latest/index.html.

[29] A. Hillas. *Cerenkov light images of EAS produced by primary gamma*. Aug. 1985.

[30] A. Hillas. "Differences between gamma-ray and hadronic showers". en. In: *Space Science Reviews* 75.1 (Jan. 1996), pp. 17–30. ISSN: 1572-9672. DOI: 10.1007/BF00195021. URL: https://doi.org/10.1007/BF00195021.

[31] A. Haungs et al. "First results on characterization of Cherenkov images through combined use of Hillas, fractal and wavelet parameters". en. In: *Astroparticle Physics* 12.3 (Nov. 1999), pp. 145–156. ISSN: 0927-6505. DOI: 10.1016/S0927-6505(99)00081-X. URL: https://www.sciencedirect.com/science/article/pii/S092765059900081X.

[32] A. Walker et al. "Measurement of the angular resolution of an extensive air shower array using a Cherenkov light detector". en. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 301.3 (Mar. 1991), pp. 574–578. ISSN: 01689002. DOI: 10.1016/0168-9002(91)90025-L. URL: https://linkinghub.elsevier.com/retrieve/pii/016890029190025L.

[33] M. de Naurois. "Analysis methods for Atmospheric Cerenkov Telescopes". en. In: (2005).

[34] *ctapipe.containers — ctapipe v0.19.3.dev19+g87591436*. URL: https://ctapipe.readthedocs.io/en/latest/_modules/ctapipe/containers.html#ImageParametersContainer.

[35] T. Fawcett. "An introduction to ROC analysis". en. In: *Pattern Recognition Letters* 27.8 (June 2006), pp. 861–874. ISSN: 01678655. DOI: 10.1016/j.patrec.2005.10.010. URL: https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X.

[36] Gaurav. *An Introduction to Gradient Boosting Decision Trees*. en-us. June 2021. URL: https://www.machinelearningplus.com/machine-learning/an-introduction-to-gradient-boosting-decision-trees/.

[37] *Decision Trees, Explained To Kids | Science 2.0*. en. Aug. 2014. URL: https://www.science20.com/tommaso_dorigo/decision_trees_explained_to_kids-224948.

[38] Y. Coadou. "Boosted decision trees". In: arXiv:2206.09645 [hep-ex, physics:physics]. Mar. 2022, pp. 9–58. DOI: 10.1142/9789811234033_0002. URL: http://arxiv.org/abs/2206.09645.

[39] S. Ohm, C. van Eldik, and K. Egberts. "/hadron separation in very-high-energy -ray astronomy using a multivariate analysis method". en. In: *Astroparticle Physics* 31.5 (June 2009), pp. 383–391. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2009.04.001. URL: https://www.sciencedirect.com/science/article/pii/S0927650509000589.

[40] S. Malakouti. "Discriminate primary gammas (signal) from the images of hadronic showers by cosmic rays in the upper atmosphere (background) with machine learning". en. In: *Physica Scripta* 98.4 (Mar. 2023). Publisher: IOP Publishing, p. 045506. ISSN: 1402-4896. DOI: 10.1088/1402-4896/acc1b2. URL: https://dx.doi.org/10.1088/1402-4896/acc1b2.

[41] *What are Neural Networks? | IBM*. en-us. URL: https://www.ibm.com/topics/neural-networks.

[42] M. Simic. *Gradient Boosting Trees vs. Random Forests | Baeldung on Computer Science*. en-US. Feb. 2022. URL: https://www.baeldung.com/cs/gradient-boosting-trees-vs-random-forests.

[43] J. Brownlee. *A Gentle Introduction to XGBoost for Applied Machine Learning*. en-US. Aug. 2016. URL: https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/.

[44] P. Luijks. *BSc Thesis - Cosmic ray identification with the Cherenkov Telescope Array*.

[45] *XGBoost Parameters — xgboost 1.7.5 documentation*. URL: https://xgboost.readthedocs.io/en/stable/parameter.html.