Analyzing The Cosmic Ray Spectrum Based On Imaging Atmospheric Cherenkov Telescope Data Using Deep Learning Methods

Master's Thesis in Physics

Presented by

Christina Hillig

March 18, 2019

Erlangen Centre for Astroparticle Physics

Friedrich-Alexander-Universität Erlangen-Nürnberg

1. Supervisor: Prof. Dr. Stefan Funk
2. Referee: Prof. Dr. Gisela Anton

Zusammenfassung

Der Ansatz von Deep Learning Methoden zur Unterscheidung von Luftschauern induziert durch γ Strahlung und Hadronen aufgenommen mit H.E.S.S., erzielte Ergebnisse, welche vergleichbar sind mit den standard Analyse Methoden. Aus diesem Grund wird der Deep Learning Ansatz getestet, um Bilder von kosmischer Strahlung zu klassifizieren. Die Ziele dieser Arbeit sind die Unterscheidung der mit H.E.S.S. detektierten ausgedehnenten Luftschauerbilder von kosmischer Strahlung, deren Energierekonstruktion sowie die Erzeugung eines Eisenspektrums. Für die Zuordnung der Luftschauerbilder zu dem initialen kosmischen Teilchen wurde ein Deep Learning Ansatz gewählt. Es wurden für das Training des Deep Learning Algorithmuses Monte-Carlo (MC) Simulationen verwendet. Die Energie des primären Teilchens wurde ebenfalls über einen Deep Learning Ansatz rekonstruiert. Um aus diesen Daten ein Eisenspektrum zu erstellen, wurde der Foward Folding Algorithmus implementiert und erweitert, sodass er auf hadronischen Teilchen angewandt werden kann. Es wurden unterschiedliche Deep Learning Architekturen gestest. Die Architektur mit der höchsten Genauigkeit bei der Bestimmung des initialen Teilchens wurde für die weiterführende Analyse verwendet. Durch die zusätzliche Energierekonstruktion der hadronischen Teilchen konnte der Forward Folding Algorithmus angewendet werden. In dem Eisenspektrum sind systematische sowie statistische Fehler, die durch den Detektor, den Klassifizierungs und Energie Rekonstruktions Alorithmus entstanden sind, vernachlässigt worden. Der größte Fehler ist gegeben, durch die stark fehlerbehaftete Klassifizierung von Eisen Kernen mit dem Deep Learning Ansatz, welche sich bei der Anwendung auf reale Daten von PKS 2155-304 stärker ausgewirkt hat, aufgrund u.a. der Diskrepanz zwischen MC und realen Daten. Es war möglich ein Eisenspektrum aus MC Simulationen zu rekonstruieren.

Abstract

Deep Learning methods are shown to be applicable on the distinction task of γ-ray and hadrons induced extensive air showers (EASs) observed with H.E.S.S.. The results are comparable to the common analytical analysis methods. Therefore the Deep Learning approach is tested to distinguish cosmic radiation images. The goals of this thesis are to distinguish images of cosmic radiation EAS observed with H.E.S.S., to reconstruct their energy and to create an iron energy spectrum with these results. The Deep Learning approach was chosen for the classification of EAS images of an initial particle. Monte-Carlo (MC) simulations were used for the training of the Deep Learning algorithm. The energy of a primary particle was reconstructed using this approach, too. To create an iron energy spectrum out of these data, the Forward Folding algorithm was extended such that it can be applied on hadronic signal. Several algorithms were tested. The one having the highest accuracy was used for further analysis steps. Due to the energy reconstruction of the hadronic particles the Forward Folding algorithm can be used. The iron energy spectrum takes not into account the systematic and statistic errors caused by the response of the detector, the classification and energy reconstruction algorithm. The greatest error is given by the erroneous classification of iron nuclei using Deep Learning which has the most impact on the real data classification of PKS 2155-304 due i.a. to the discrepancy between MC and real data. It was possible to create a MC iron spectrum.

Contents

List of Abbreviations

γ-ray	gamma-ray
EAS	extensive air shower
DR-light	Direct-light
DL	Deep Learning
ML	Machine Learning
MC	Monte-Carlo
CR	cosmic-ray
SNR	supernova remnant
IACT	Imaging Atmospheric Cherenkov Telescope
VHE	very high energy
EM	electromagnetic
MAGIC	Major Atmospheric Gamma Imaging Cherenkov Telescopes
VERITAS	Very Energetic Radiation Imaging Telescope Array System
H.E.S.S.	High Energy Stereoscopic System
CTA	Cherenkov Telescope Array
PMT	photomultiplier tube
CT	Cherenkov telescope
FoV	field of view
photoelectrons	p.e.
PCA	Principal Component Analysis
SVD	singularity value decomposition
ImPACT	Image Pixel-wise fit for Atmospheric Cherenkov Telescopes
NN	Neural Network
ReLU	rectified linear unit
DNN	deep neural network
CNN	convolutional neural network
ResNet	Residual Network
LSTM	Long Short Term Memory
ROC	receiver operating characteristic
AUC	area under the curve
TP	true positive
FN	false negative
FP	false positive
TN	true negative
PDF	probability density function
AP	actual positives
-ln L	negative log-likelihood

List of Figures

List of Tables

Introduction

So far the key science question for the gamma-ray (γ-ray) observations in the energy range of MeV up to PeV is the origin of cosmic radiation. Therefore knowledge about sources which provide the energy for cosmic radiation, the acceleration process and their energy spectrum [28] is of interest. The observation of γ-ray and hadronic particles extensive air showers (EAS) is done by e.g. ground-based methods such as H.E.S.S.. For the reconstruction of γ-ray EAS analytical approaches are well established due to their elliptical shape in the camera image of the telescopes [55]. In the case of a detected hadronic EAS there are already some approaches developed e.g. focusing on the Direct Cherenkov light (DC-light) pixel in the camera image for reconstruction [6] or the use of a Boosted Decision Tree [10] for classification.
In this thesis a *Deep Learning* (DL) approach is tested to analyze images of hadronic EAS. DL is related to *Machine Learning* (ML). Arthur Samuel defined in 1959 ML as the

> *"[...] field of study that gives computers the ability to learn without being explicitly programmed".*

He was a pioneer in this field and is known for his Checkers-playing program which was the first self-learning program. On the basis of it he tried to teach computers how to learn from their previous tasks [46].
Another pioneer in the methods and applications of ML is Tom Mitchell. He described in 1998 the learning problem as follows:

> *"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E".*

Today ML is used in many scientific fields such as automotive driving [56], health care [11] and γ-ray astronomy [58].
In the γ-ray astronomy DL methods are used to distinguish heavier nuclei and reconstruct their energy. In addition, the reconstruction of a γ-ray gives information about the observed source. Analytical solution do already exist for this particle type and up to now it is not possible to achieve better results using a ML approach. Therefore it is tried to reconstruct marginalized groups such as the hadronic particles because the conventional methods are not able to analyze them.
The goals of this thesis are to use a ML approach to distinguish cosmic radiation represented by the six classes Gamma, Proton, Helium, Carbon, Silicon and Iron where Iron

represents all heavier elements. In addition, a reconstruction of their energy is done to create an iron nuclei spectrum using real data observations from PKS 2155-304.
In principle it is possible to create a spectrum of the other 5 classes using a ML approach but they are not part of this thesis. The ML tasks are trained on Monte-Carlo (MC) simulated shower events simulated under a zenith angle of 20 °, an azimuth angle of 180 °, with a maximum offset of 2.2 ° and 2.5 ° having an energy of $E > 5\,\mathrm{TeV}$ and $E < 100\,\mathrm{TeV}$.
To find the best architecture for the distinction of cosmic radiation several buildings are tested. An architecture is the formal expression of a mathematical approach used to define the given task. It is tried to improve the classification performance by investigating different architectures. The one having the highest accuracy is chosen for the multi-classification task distinguishing cosmic radiation. An other architecture is used for the energy reconstruction task. The results of both task are used for the application of the Forward Folding algorithm in order to create an iron nuclei spectrum.

The contents of this thesis contain among others an introduction to the theory of EAS and their observation, analysis methods and Fe^{+} spectrum reconstruction. In detail, the first chapter gives a brief overview to the origin of cosmic radiation and the related energy spectrum. Followed by the different types of air shower and their detection in the second chapter. In the third chapter the Hillas analysis is briefly explained and the ML approach is introduced in detail. The fourth chapter contains the methods to accomplish the results. The fifth chapter represents the results of the classification and energy reconstruction of the cosmic radiation and in addition, the created iron nuclei spectrum. The last to chapters contain the discussion and conclusion of the results.

1. Cosmic Radiation

Since Viktor Hess proved the existence of cosmic radiation by measuring increasing ionization radiation in higher altitudes a lot has been explored about cosmic radiation [35]. This chapter will give a brief overview about the origin of cosmic radiation and their energy spectrum in association to their differential fluxes.

1.1. Sources of ultra-high Cosmic Radiation

Cosmic radiation is a high energy radiation divided into γ-rays and cosmic-rays (CRs) nd comes from outside of the solar system [30]. γ-rays are very high energy photons in ranges higher than 100 keV. A γ-ray is created if a cosmic particle is accelerated in interstellar medium and a decay process or an inverse Compton scattering occurs [27].
Today, the origin of CRs is still not fully comprehended. CRs are heavier nuclei synthesized in stars [30]. So far, there are two general approaches for the emergence of CRs. One of them are top-down scenarios which assume that a CR is created as decay product of the decay of heavy particles. The others are bottom-up scenarios where particles are accelerated by fast cosmic fields also called (bow) shocks near supernova remnants (SNR) [18].
The bottom-up scenarios are the most plausible because their energy distribution is comparable to the observed one of the CRs. The energy distribution of CRs is introduced in sec. 1.2.

Fermi Acceleration of Cosmic Rays

Enrico Fermi originally tried to explain the acceleration mechanism of cosmic particles. He made the assumption that a relativistic particle is reflected and accelerated by cosmic fields. There it either gains or looses energy depending on its charge whereby the energy gain is more likely than the losses [26].
He showed that the average energy gain can be described by a power-law and that it is proportional to the square of the speed of the cosmic field v_{cf} divided by the square of the speed of the relativistic particle c_p: $\propto \left(\frac{v_{cf}}{c_p}\right)^2$ [18]. This is defined as second-order Fermi acceleration. Despite this approach some mechanisms can not be explained [18].
In comparison to the second-order acceleration, in the first-order Fermi acceleration the energy gain is linear: $\propto \left(\frac{v_{cf}}{c_p}\right)$. This leads to a more effective acceleration of relativistic particles in shocks near SNR. If the particle moves with isotropic velocity into a shock it will be diffusely reflected and also its velocity will be increased. The orientation of

the particle's direction is changed from either down- to upstream or up- to downstream [50]. After multiple repetitions of the processes by many particles the resulting energy gain can be approximated by a power-law:

$$N(E)\,dE \sim E^{-2}\,dE$$

with dE the fractional energy gain and $N(E)$ the number of particles. The power-law index is -2 for the first-order Fermi acceleration [18].

In summary, γ-rays are created from the interaction of CRs with nuclei in interstellar medium and decays of neutral pions π^0. CRs mostly originate outside of our galaxy due to their high energy and deflection in cosmic clouds and magnetic fields. Because, by now it is assumed that particles are diffuse accelerated in shocks which is described by first-order Fermi mechanism. It expects a linear energy gain which can be approximated for many particles by a power-law with an index of -2. This is close to the observed spectral index Γ for CRs and is therefore a good approximation. One has to consider that not every particle is equally accelerated in cosmic mirrors. Thus the energy distribution of cosmic radiation has a wide range which will be discussed in the next part.

1.2. Energy Spectrum

The energy spectrum of cosmic radiation reaches from 1 eV up to 100 EeV due to its different acceleration in cosmic fields and composition. To measure the energy spectrum of cosmic radiation it has to be detected directly or indirectly via detection methods introduced in section 2.
Subsequently, the measured cosmic radiation has to be separated into its components to describe the primary spectrum of cosmic radiation which is described by the differential flux measured in $\mathrm{m^{-2}\,s^{-1}\,sr^{-1}\,TeV^{-1}}$ per nucleus [67]:

$$\frac{d\Phi}{dE} = \Phi_0 \cdot \left(\frac{E}{E_0}\right)^{-\Gamma} \tag{1.1}$$

with Φ_0 the absolute flux normalization, Γ the spectral index and E_0 the normalization energy. The energy E describes the kinetic energy of the particle. The differential flux follows a power-law and is similar to the calculated power-law of first-order Fermi acceleration.
High energy CRs can be divided into ionized nuclei, electrons, and antimatter. The dominant components of the ionized nuclei are hydrogen (89 %) and helium (10 %) [47]. Whereas in comparison heavier nuclei and electrons are very rare [67]. Figure 1.1 depicts the differential flux of CRs.

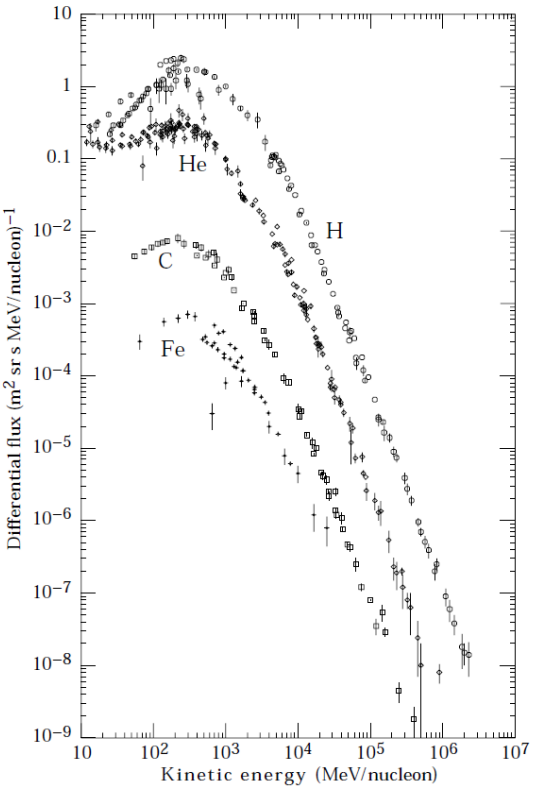


Figure 1.1. – *Differential flux of major components of primary CRs. Figure adapted from Ref. [30].*

Since the energy spectrum follows a power law at higher energies the low flux demands a greater effective area to ensure enough data. This can be achieved via the Imaging Atmospheric Cherenkov technique which is described in detail in the following chapter.

2. Detection of Cosmic Radiation using Imaging Atmospheric Cherenkov Telescopes

Already in 1911, Viktor Hess discovered during two balloon rides that in addition to the radiation of radioactive substances at the ground there must be other sources of gamma radiation. He came to this conclusion because, unlike expected, the ionization energy did not change with increasing altitude. In 1912, he tried again to measure the radioactive radiation in Earth's atmosphere at even higher altitudes independently of time and other external influences. He found out that from about 2000 m upwards the ionization of the Earth's atmosphere increases. This can only be explained by the fact that radiation is of cosmic origin [35].
Based on his research further investigations were made. For example, Robert Millikan, known for the oil drop experiment, developed automated, self-registering gauges which were used to measure ionization regardless of the altitude [65].
Cosmic radiation consists of high-energy particles ($E > 100\,\mathrm{MeV}$ [69] up to $E < 100\,\mathrm{EeV}$ [19]) and is entirely absorbed in the Earth's atmosphere. The reason is the opacity of the atmosphere which is highly energy dependent. Therefore, high energy γ-rays have to be detected directly via space-based satellite telescopes like the Large Area Telescope on the Fermi satellite or indirectly via ground-based techniques for example Imaging Atmospheric Cherenkov Telescopes (IACT).
However, for the detection of very high energy (VHE) particles, e.g. larger detection areas are needed because the differential flux is described by a power law see eq. 1.1. Due to the limited size of a satellite, the area of a detector is constrained and therefore ground-based techniques introduced in section 2.2 are necessary for the detection of air showers[28]. In the following the generation of air showers and Cherenkov light will be described in more detail.

2.1. Extensive Air Shower

Each time a VHE particle enters the Earth's atmosphere an EAS is induced. If primary cosmic radiation interacts with nuclei in Earth's atmosphere at a height of around $10 - 20\,\mathrm{km}$ interaction products are created, so called secondary particles. Each VHE secondary particle interacts again with atmospheric nuclei. This produces a cascade of particles.
A distinction is made between electromagnetic (EM) and hadronic showers depending on the interactions in the Earth's atmosphere and the primary cosmic radiation.

2.1.1. Cherenkov Radiation

During particle cascades many charged VHE particles are produced. Due to their high energy they move faster than the speed of light in the dielectric medium air. This leads to radiation of Cherenkov light in visible and ultraviolet wavelengths lasting for a few nanoseconds [69] [27]. Cherenkov light is also known as Cherenkov effect. It is emitted under a characteristic angle θ_C:

$$\cos(\theta_C) = \frac{c}{n_{air} \cdot v} \tag{2.1}$$

Where θ_C is determined by the speed of light c, the refractive index of air n_{air} and the velocity of the particle v. Since $\cos(\theta_C)$ must be smaller than 1, the maximum angle θ_C is constrained by the velocity of ultra-relativistic particles ($v \approx c$):

$$\theta_{C,max} = \arccos\left(\frac{1}{n_{air}}\right)$$

In the air the maximum angle $\theta_{C,max}$ is approximately $1\,^\circ$. Due to the dependency of refractive index n_{air} to the height in the atmosphere, Cherenkov photons are emitted in a light cone around the shower axis [27].

In conclusion, Cherenkov light is emitted by charged particles moving faster than the speed of light in air ($\frac{v}{c} > \frac{1}{n_{air}}$) around the shower axis in a light cone.
Secondary particles of EAS fulfill these criteria thus emitting Cherenkov photons. Those are created during interactions of a primary particle with air nuclei. The evolution and behavior of different air showers is explained in more detail below.

2.1.2. Electromagnetic Shower

Electromagnetic showers are induced by interactions of electrons, positrons or photons of high energy with electric fields of atomic nuclei in the air. The interaction process between a relativistic particle and a nucleus in the Earth's atmosphere depends on the type and energy of the particle. The dominant interactions are ionization, bremsstrahlung and pair production.
VHE electrons/positrons lose their energy through bremsstrahlung by emitting a (VHE) γ-ray. This VHE γ-ray interacts with an atom via pair production creating an electron-positron pair which in turn emits new (VHE) γ-rays through bremsstrahlung, see Fig. 2.1. Particle production processes increase the number of particles in the shower while decreasing the average energy per particle $\bar{E}$ exponentially:

$$\bar{E} = E_0\, e^{-\frac{X}{X_r}}$$

where E_0 is the initial energy of the primary particle and X is the atmospheric depth which corresponds to the interaction length. The characteristic length X_r is defined as the distance where the energy is reduced by $1/e$ due to interaction processes [32].

These processes are repeated until the energy $\bar{E}$ drops under a critical point ($E_{c,air} \approx 81\,\text{MeV}$ [30]). Below this threshold the loss of energy by ionization of air molecules dominates the loss of particle production processes. Particles with $\bar{E} < E_{c,air}$ do not radiate and lose their rest energy by e.g. Compton scattering, ionization and photoelectric effect. During interaction processes, the energy and momentum must be conserved. A simple model of an EM shower development was created by Heitler [45]. The model is divided into a longitudinal and lateral shower development. The axial shower profile is characterized by the maximum size which is determined by the penetration depth X_{max}. It depends logarithmically on the factor E_0/E_c and is given by:

$$X_{max} = X_0 \ln\left(\frac{E_0}{E_c}\right)$$

with X_0 as radiation length.
The lateral expanse of the EM shower depends also on the initial energy. This means for higher energies of the primary particle the shower becomes narrower. This is comparable to the Molière Radius R_M which describes the transverse expansion of an EM shower in a medium [20]:

$$R_M = X\,\frac{E_s}{E_c}$$

with E_s ($\sim 21\,\text{MeV}$ [20]) as scale energy. Along the shower axis within a radius of $3.5\,R_M$, 99 % of the energy is deposited[20].

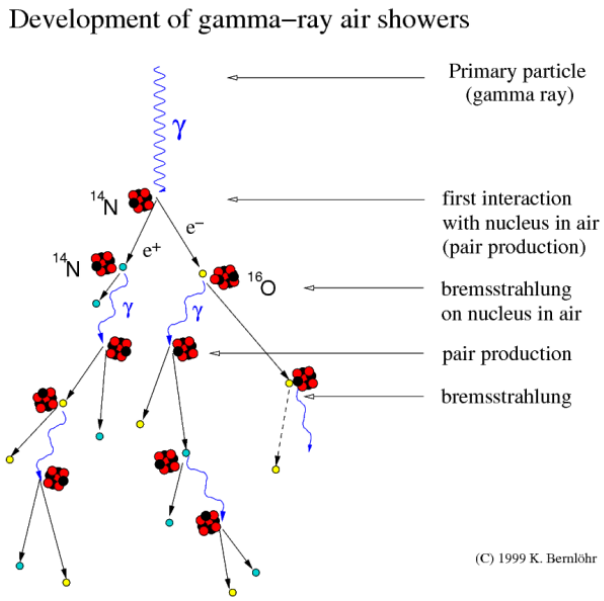


Figure 2.1. – *Example of a γ-ray induced cascade. The incoming primary particle interacts with the atmospheric nuclei producing an electron-positron pair. These created particles interact again with nuclei in the Earth's atmosphere in turn creating (VHE) γ-rays. Figure extracted from Ref. [14].*

Recapitulating, the number of particles in a shower reaches its maximum if $\bar{E} < E_c$. At this point no further particles are produced and the rest energy is deposited in air via ionization. This is synonymous to the shower maximum in penetration depth X_{max}. Where X_{max} provides information about the longitudinal shower development which depends also on the initial energy of the primary particle.

As mentioned in section 1.2, the radiation reaching Earth's atmosphere does not only consist of VHE γ-rays but also of CRs. In the following section the development of showers induced by CRs is described.

2.1.3. Hadronic Shower

A hadronic shower is induced via inelastic interactions between a CR and an atmospheric nucleus. If the resulting secondary particle is hadronic with an energy above a certain threshold then in turn it interacts inelastic with surrounding medium creating new secondary particles.
A hadron with an energy E_0 interacting in the atmosphere creates a_{tot} new particles with an energy E_0/a_{tot}. Those are separated into hadron (proportion $^2/_3$) and EM components (proportion $^1/_3$) [23]. Their energy is determined by the initial energy of the CR E_0 and the number of interactions t:

$$E_{hadron} = \left(\frac{2}{3}\right)^t E_0$$
$$E_{EM} = \left(\frac{1}{3}\right)^t E_0$$

The EM components in a hadronic shower result from the decay of charged muons and neutral pions. A muon decays in an electron/positron depending on its charge if its energy is below a certain decay energy E_{decay}.

$$\begin{aligned} \mu^- &\rightarrow e^- + \bar{\nu}_e + \nu_\mu \\ \mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu \end{aligned} \tag{2.2}$$

If the energy of a muon is higher than a minimal emission energy it emits radial Cherenkov photons forming a so called muon ring. At sea level the decay of a muon into a low-energy electron is dominant [30].
The main decay (98.82 % [54]) of neutral pions is:

$$\pi^0 \rightarrow 2\gamma$$

This process happens immediately and mainly takes place at high altitudes [23] [30].

The hadronic components result from charged π-mesons or kaons. If their energy is above the decay energy E_{decay}, they interact again inelastically with atmospheric nuclei.

If their energy is below E_{decay}, they tend to decay in a charged muon and a neutrino (99.98 % for pions and 63.56 % for kaons [54]) [23]:

$$\pi^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}(\bar{\nu}_{\mu})$$
$$K^{\pm} \rightarrow \mu^{\pm} + \nu_{\mu}(\bar{\nu}_{\mu})$$

The charged muon in turn contributes to the EM component of the hadronic shower, see eq. 2.2. The decay energy E_{decay} is similar to the critical energy E_c, introduced in section 2.1.2, which describes an energy dependent threshold for interaction processes. E_{decay} depends on the actual density gradient of the Earth's atmosphere and decreases with increasing initial energy of the primary particle. In consequents, heavier incoming CRs produce smaller EAS because of their energy dependency on the mass number. At the shower maximum the energy of the particle is equal to E_{decay} and the particles deposit their energy by ionization of the air molecules.

The spatial development in longitudinal direction can be described by the interaction length λ_{int} [7]:

$$\lambda_{int} = \frac{\mathcal{A}}{N_A \rho \sigma_{inel}}$$

with σ_{inel} as hadron effective collision cross-section of one nucleon and $\frac{N_A \rho}{\mathcal{A}}$ as the density of the atoms. The longitudinal development of hadronic showers is slightly broader than those of EM showers due to $\lambda_{int} > X$ [69] and inelastic interaction processes leading to a wider spreading of the cascade.
The lateral shower profile is characterized by the transverse narrow expansion of the EM components measured in R_M. The hadronic components consist mainly of non-relativistic particles and their interactions add a transverse momentum which broadens the lateral shower profile.

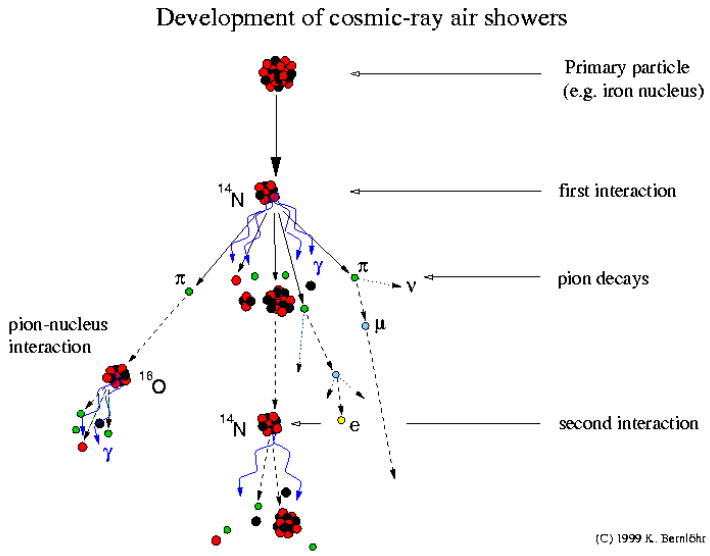


Figure 2.2. – *Example of an iron nucleus induced hadronic shower. The ionized iron nucleus interacts inelastically with atomic nuclei in the Earth's atmosphere inducing an EAS. Each interaction creates secondary particles e.g. pions emitting Cherenkov light. Charged pions decay in charged muons and neutrinos. Figure from Ref. [14].*

Another important contribution to the hadronic shower profile is the primary CRs ability of emitting Cherenkov light because in comparison to EM showers the primary particle is charged. This phenomenon is also called DC-light and creates an additional light cone with an radius of $\sim 100\,\mathrm{m}$ which is much narrower as the light cone produced by secondary particles due to the higher emission height, the resulting smaller Cherenkov angle as well as it is produced by the primary particle itself, see eq. 2.1. The intensity of the DC-light is described by the charge Z of the primary particle [6]:

$$I \sim Z^2$$

To fully understand the development of the hadronic shower profile a more detailed modeling of the interaction processes is needed. This can be done via MC simulations which will be briefly explained in section 2.3.2.

As mentioned above, EM and hadronic showers have different properties. They are compared in table 2.1.

Table 2.1. – *Comparison of attributes of the two types of EAS*

	EM shower	**Hadronic shower**
Interaction length	small	high
Secondary particles	γ, $e^{\pm}$	$\mu^{\pm}$, $\pi^{\pm}$, π^{0}, ...
Scattering	narrow	wide
DC-light	no	yes

The described EAS can be used to draw conclusions about the initial particles. Hence, detecting EAS is an indirect method to observe cosmic radiation. The next section presents a ground-based cosmic radiation detection technique.

2.2. Imaging Atmospheric Cherenkov Telescopes

The IACT technique is the most used technique for ground-based astronomy in the VHE range. It is a tool to detect visible Cherenkov light emitted by EAS generated by cosmic radiation during clear nights.
It was firstly implemented in the Whipple 10 m gamma-ray telescope which was located on Mount Hopkins at the Fred Lawrence Whipple Observatory in Southern Arizona. Due to the improvement in sensitivity (minimal detectable flux) by IACT technique it was possible to detect the first TeV γ-ray source, the Crab Nebula, in 1989 [66]. The Crab Nebula is a part of the Milky Way galaxy and was created by a supernova during an explosion of a massive star and is a supernova remnant.
Another improvement towards higher sensitivity in IACTs was achieved through the combination of a very large mirror and fast timing in Major Atmospheric Gamma Imaging Cherenkov Telescopes (MAGIC) and also through stereoscopic imaging arrays like Very Energetic Radiation Imaging Telescope Array System (VERITAS) and High Energy Stereoscopic System (H.E.S.S.) [27]. Another IACT experiment is planned, the Cherenkov Telescope Array (CTA). It will consist of three different telescope sizes: small, medium and large and it will be located at two different locations to be able to detect sources on the hole sky map. The IACT arrays will be located in Atacama Desert in Chile which belongs to the Southern Hemisphere and La Palma which is in Northern Hemisphere.

Cherenkov Light Detection

Cherenkov light is emitted by EAS in an altitude of about 20 km at the Cherenkov angle Θ_C of $\sim 1^\circ$ in air. The emitted light cone results in a circle on the ground with a diameter of 100 to 150 m depending on the local altitude of the IACT [37].
The technique uses very large focusing mirrors to catch and reflect Cherenkov light onto

the focal plane of the camera, see Fig. 2.3. Since the Cherenkov photons act like light pulses which expose the camera plane in a very short time fast photomultiplier tubes (PMTs) are needed. Due to the specular reflection of the incident Cherenkov light pulse on the telescope reflector, the angle is preserved. This can be used to reconstruct the direction of the initial γ-ray. IACTs have an angular resolution of $0.1\,^\circ$ [28]. With increasing reconstruction performance it becomes easier to [44]:

- resolve images of galactic sources
- measure flux variations with a minimal limit of 200 s
- measure energy spectra up to 70 TeV

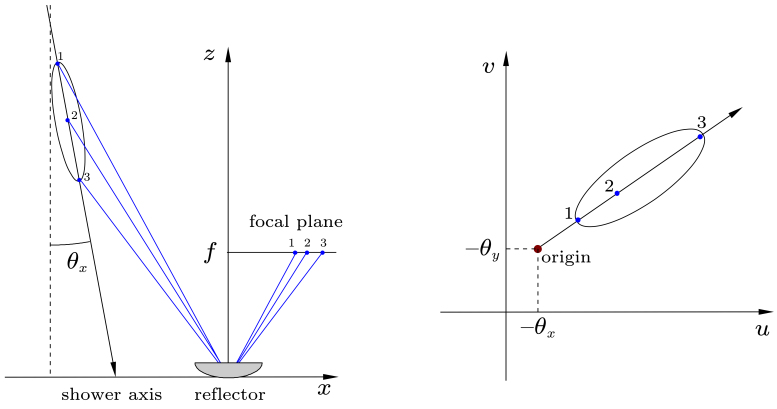


Figure 2.3. – *The left figure illustrates an EAS which emits Cherenkov light at the Cherenkov angel θ_x around the shower axis. The light is divided into three paths pointing onto the reflector which focuses the light beams to the focal plane in the drawn order. The figure on the right shows the imaged EAS in the camera plane (u, v). The red point at $(-\theta_x, -\theta_y)$ defines the origin of the primary cosmic particle. Figure extracted from Ref. [41].*

As mentioned above, one of the IACTs is H.E.S.S.. Because this thesis follows up with real observation data and Monte Carlo simulations of H.E.S.S. it will be described in further detail below.

2.3. H.E.S.S.

H.E.S.S. is based on the IACT technique and studies γ-ray sources in an energy range of 100 GeV up to an order of TeV. It is located in the Southern Hemisphere at the Khomas highland at 1800 m in Namibia [28]. Due to the sensitive technique Namibia was chosen because of its clear night sky and good weather conditions.
The stereoscopic imaging array is divided into two phases. H.E.S.S. I consists of four 12 m Cherenkov telescopes (CT1-4) with a mirror area of $108\,\mathrm{m}^2$ per telescope and

H.E.S.S. II adds a fifth 28 m telescope (CT5) with a mirror area of 614 m^2 to the array [38], see figure 2.4. It shows the stereoscopic imaging array H.E.S.S. with the position of the telescopes.

Figure 2.4. – *H.E.S.S. with all five telescopes. CT1-4 are arranged in a square of* 120 m *side length [38]. CT5 is located in the middle of the square. Figure from Ref. [2].*

2.3.1. Optical System of CT1-4

The optical system of the H.E.S.S. telescopes is mounted on an alt-azimuth motorized dish. It operates in the Southern Hemisphere. The reflector is partitioned in spherical mirror facets. The mirrors of the telescopes do focus to a distance of 10 km corresponding to the altitude at which EAS are induced. The incident light pulse is reflected by a mirror onto Winston cone light collectors focusing on photon detectors consisting of PMTs [16]. The photomultiplier camera is positioned perpendicular to the center of the dish held by four camera arms. It has a field of view (FoV) of 5 ° diameter and 960 hexagonal pixels [38]. The pixels are divided into 60 drawers whereby each drawer includes 16 pixels. The system has an angular resolution of 0.1 °, hence it is possible to observe not only point sources but also the extend of expanded sources.
The trigger conditions of H.E.S.S. I are summarized and ordered by priority in the following [38]:

- signal of 5 photoelectrons (p.e.) $\Rightarrow$ single pixel triggered
- 3 to 5 pixels trigger by coincidence within a 8×8 pixel sector in a time window of 1.5 ns $\Rightarrow$ telescope triggered
- two or more telescopes trigger $\Rightarrow$ multiplicity fulfilled

If these conditions are fulfilled data is taken. One data take lasts $\sim 28.2\,\text{min}$ and is called run. A run consists of events containing the intensity, triggered pixels and additional data. These include e.g. dead-time corrected run duration, number of participated telescopes, telescope alignment and Hillas parameters which will be introduced in section 3.1.

2.3.2. Data Representation

As mentioned above the IACT technique is used to detect Cherenkov photons. The data acquisition system of H.E.S.S. collects and performs a real time analysis on data from photo detectors. The observed data can then be illustrated in pixelated images as shown in Fig. 2.5. The elliptical shape of the imaged shower results from the elliptically formed EM showers. Hadronic showers are rejected by the offline analysis methods because of their wide scattering [37].

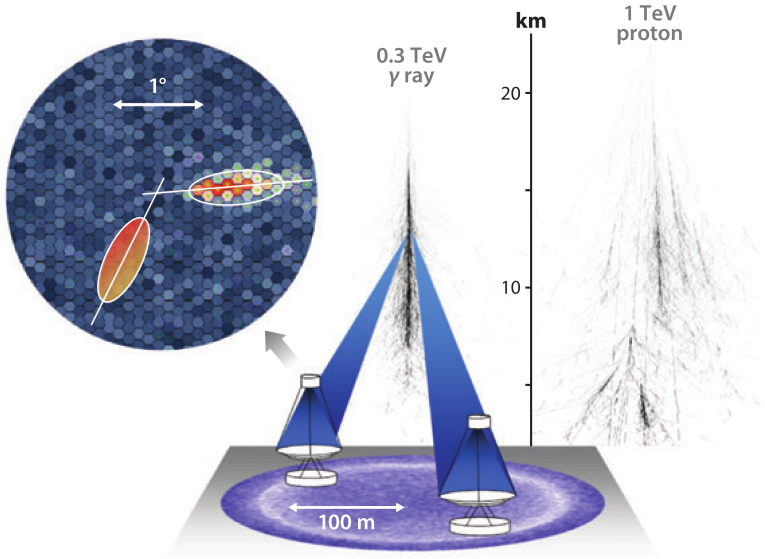


Figure 2.5. – *In an altitude of* $20\,\text{km}$ *a cosmic radiation induced EAS emits Cherenkov light in a light cone resulting in a circle on the ground. H.E.S.S. detects Cherenkov photons by reflecting those via mirrors and image them on the focal plane of the camera. Figure from Ref. [37].*

Real Data Observation

Real data observations were taken from PKS 2155-304. BL Lac PKS 2155-304 is a VHE γ-ray source with a redshift of $z = 0.116$ [5]. As the name implies it belongs to the group of blazars which are a class of Active Galactic Nuclei. These have active supermassive black holes in their centers [49] like our Milky way. Since 2002 PKS 2155-304 is observed by H.E.S.S. and is the brightest blazar in the Southern Hemisphere at the energy of 200 GeV [5]. It is monitored continuously to detect bright flares, which are needed to be able to study the VHE flux variability within the expanse of tens of seconds [5].
For this thesis the following selection criteria for PKS 2155-304 runs are required:

- observed data must be of phase 1b
- multiplicity of participated telescopes is 2
- maximum offset $\in [0, 5\,^\circ]$
- zenith angle $\in [19.5\,^\circ, 20.5\,^\circ]$

Monte-Carlo Simulations

The H.E.S.S. MC simulations used in this thesis were created with CORSIKA IACT option and simtel_array. CORSIKA is a program simulating EAS evolutions in dependency of the air opacity of cosmic radiation as well as their detection [15].
For the analysis done in this thesis diffuse MC EAS were simulated of phase 1b under a zenith angle of $20\,^\circ$, an azimuth angle of $180\,^\circ$, with a maximum offset of $2.2\,^\circ$ in an energy range of 5 TeV to 100 TeV.

With IACTs it is possible to detect Cherenkov light emitted by EAS and image those with a telescope camera. The main task of the IACT technique is to differentiate between EM and hadronic shower cascades.

2.3.3. Image Preprocessing

Image preprocessing of H.E.S.S. data is used to prepare observed shower images for further analysis, including shower classification and reconstruction with computational analysis tools. Some image preprocessing steps are necessary to transform the data to be able to apply additional computational algorithms. Others can improve the calculation performance by extracting features in advance.

Image Cleaning

Image cleaning is a standard preprocessing step in the H.E.S.S. analysis. It is used to reduce the noise level of observed data. Noise in images can originate from for example night sky, light pollution of cities and electronic noise. In this thesis MC simulation images are cleaned with *5-10 Image Cleaning*.
This type of image cleaning requires a threshold of at least 5 p.e. for each pixel, in

addition at least one neighboring pixel intensity has to exceed 10 p.e.. In images with hexagonal pixels each pixel has 6 neighbors but the ones at the edge of the image. If in sum the neighboring pixels do not reach the limit of 10 p.e., then the pixel is set to 0 p.e., otherwise the pixel retains its current value.

For the usage of our framework which is an analysis tool to classify and reconstruct EAS shower another preprocessing step has to be applied.

Pixel Conversion

As already explained in section 2.3.1, the camera used in H.E.S.S. telescopes consists of hexagonal pixels and since most numerical algorithms are applicable to images consisting of quadratic pixels, it is necessary to transform the hexagonal pixel grid into an image with quadratic pixels. Two kinds of pixel grid conversions are used in this thesis.

Interpolation is a common method for defining new pixel values and positions for given pixel vertexes and their intensities. The hexagonal pixel grid of the shower images is subdivided into 2-simplexes, triangles, forming a new meshgrid. The angle between vertexes is maximized by Delaunay triangulation such that no vertex is inside a circumscribed circle of any triangle following a linear interpolation on each simplex [12]. Linear interpolation is mostly applied on real data observations for tasks of energy and direction reconstruction due to the fact that pixel positions are influenced by the way pointing corrections are handled. Discussion of further details are beyond the scope of this thesis.

Rebinning is a method to translate hexagonal simplexes into square ones. Considering the observed data as hexagonal histograms gives the possibility to conserve the total intensity after transforming them into square histograms [64]. Rebinning uses sparse matrix multiplications thus reduce calculation time. In this thesis it is used on MC simulations of shower images and real data observations in classification tasks.

The following, optional preprocessing step was implemented during the research phase of this work in order to see if this improves the analysis. It is applied after the pixel conversion.

Principal Component Analysis

The Principal Component Analysis (PCA) is a method to find correlations in data and represent it by new orthogonal eigenvectors so called principal components reducing the dimension of a given dataset [4]. One advantage of dimension reduction is the lower computation time afterwards.
PCA is a tool operating on a stack of N 2D matrices with dimension d denoted as $X_{N,d\times d}$. Before PCA can be used the data must be standardized. Then singularity value decomposition (SVD) is applied calculating the singular values and eigenvectors of a square symmetric matrix, covariance matrix or correlation matrix [4]:

$$X = USV^T \ ,$$

where the matrices U and V contain the left and right singular vectors respectively and the entries of the diagonal matrix S represent the singular values.
The singular values have to be sorted in descending order to choose the k ($k < d$) largest ones where k is the dimension. These can be determined either by a user dependent threshold or variance. The associated eigenvectors construct one projection matrix $P_{d\times k}$ which transforms the original matrices $X_{N,d\times d}$ into a new feature subspace $C_{N,d\times k}$ containing the principal component scores:

$$X \approx C \cdot P^T$$

Figure 2.6 shows an example application of PCA on a dataset containing EAS images of two different particle types, γ-rays and H^+. On the left side of the figure an exemplary, combined γ-ray EAS image $X_{1,d\times d}$ with 1764 principal components is illustrated. The dimension of the image can be reduced by multiplying the feature subspace $C_{1,d\times k}$ with the projection matrix $(P_{d\times k})^T$. In this example the dimensions are $d = \sqrt{1764}$ and $k = \sqrt{223}$. The projection matrix contains all principal components with singular values fulfilling the constraint of having a variance of smaller than 90% of the maximal variance. The resulting projected image is shown on the right.
In this thesis PCA is performed as a preprocessing step on shower images after rebinning was applied, see section 5.3.

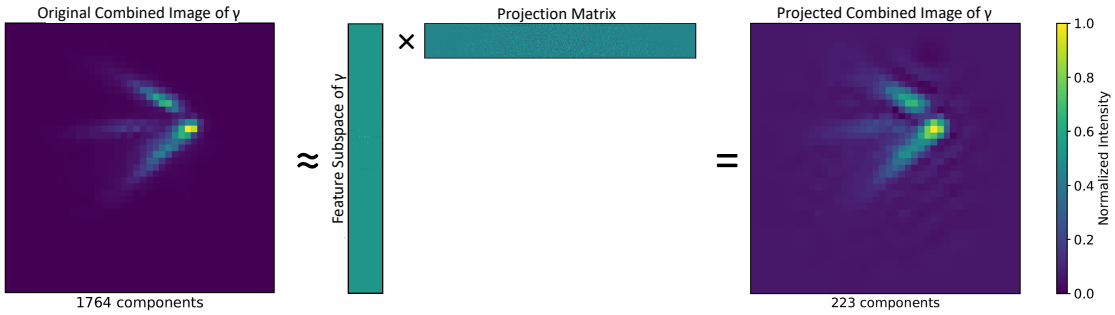


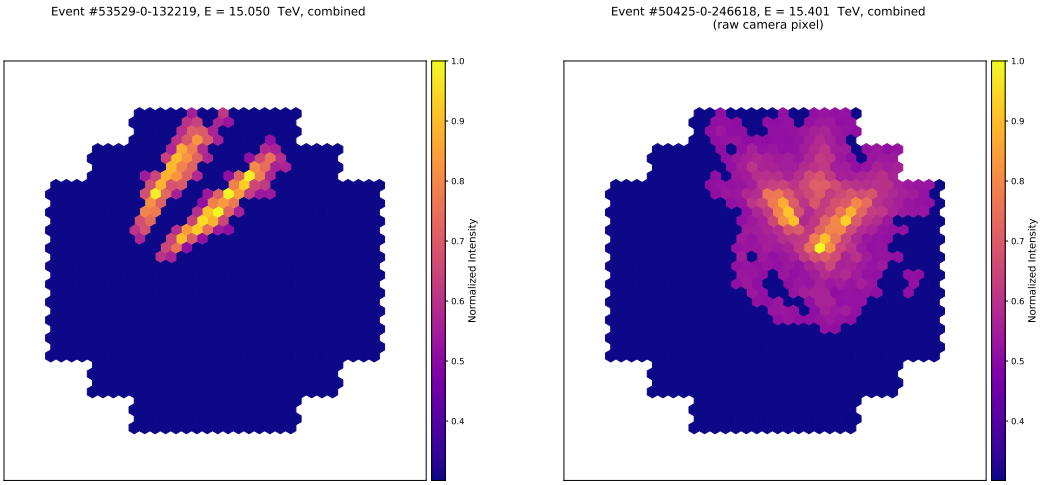
Figure 2.6. – *Illustration of the process of PCA illustrated on a combined image of a γ-ray EAS. Figure inspired by [8].*

3. Analysis Methods

This chapter introduces the used analysis methods which are applied after the preprocessing steps. γ-ray EAS observed by H.E.S.S. are reconstructed using the algorithm Image Pixel-wise fit for Atmospheric Cherenkov Telescopes (ImPACT) by part of the H.E.S.S. collaboration. ImPACT fits an assumed template to the triggered pixels in the camera using a likelihood fit [53].
Before the reconstruction of an EAS a selection of γ-ray and hadronic induced events needs to be done. The classical approach is to use Hillas analysis [55]. For a few years analysis methods have also been developed in the field of ML e.g. Boosted Decision Trees [10] and neural networks (NNs) [58]. Hillas analysis and NNs will be explained in more detail in the following. In addition, spectral analysis, a method for the reconstruction of the differential flux, is introduced.

3.1. Hillas Analysis

The Hillas analysis is the basic analysis method used for image observations with H.E.S.S. and is applied after the image cleaning [53]. The Hillas parameters are used for post-selection of γ-ray signal and hadron background in EAS images. The triggered PMTs in images of hadronic showers are randomly distributed over the hole camera plane due to their wide scattering in the Earth's atmosphere. Whereas for γ-ray induced EAS they are notably triggered more tightly near the image axis forming a two-dimensional ellipse see Fig. 3.1.



(a) *γ-ray EAS.* **(b)** *H^+ EAS.*

Figure 3.1. – *Example EAS shower of a γ-ray and a H^+ EAS observed by H.E.S.S.. The title above each combined telescope image indicates the event number including the corresponding run number and the energy of the observed EAS. The note raw means that no preprocessing step was applied expect Image cleaning. In a) a γ-ray image and in b) a H^+ EAS image is shown.*

This information can be used to reduce the image properties to a domain of boundaries which are described by an elliptical shape and orientation angle α. The ellipse is defined by the length l of the semi-major axis and the width w of the semi-minor axis. The geometrical parameters are pictured in Fig. 3.2.
The Hillas parameters are listed in the following [55]:

- Size (intensity of all pixels)
- Peak intensity (image center of gravity)
- Length l and width w of the ellipse
- Distance d to the image center
- Orientation angle α
- Azwidth (combined shape and orientation characteristics)
- Miss (distance between semi-major axis and image center)

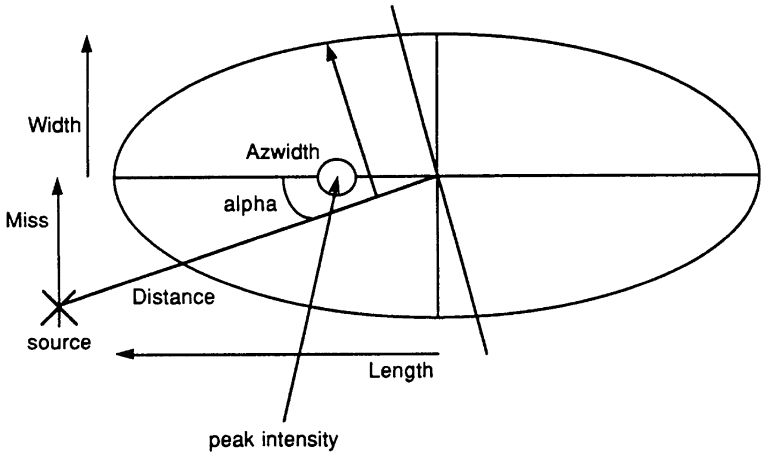


Figure 3.2. – *This figure illustrates the characteristic elliptical shape of a γ-ray shower parametrized by Hillas parameters. The ellipse is defined by the half width and length of semi-minor/major axis. Inside the ellipse is the center of gravity (peak intensity) marked as a circle. The distance and the orientation angle α define the range and orientation of the ellipse to the image center which is cross-marked as source. The azwidth is perpendicular to the distance and describes the effective extension of the ellipse. The parameter miss defines the distance from the image center to the semi-major axis. Figure extracted from Ref. [55]*

A detailed calculation of the parameters l, w, d and α can be found in *Survey of candidate gamma-ray sources at tev energies using a high-resolution cerenkov imaging system* from Reynolds et al. [55].

3.2. Machine Learning

In this thesis, the first attempt was using ML methods to distinguish not only γ-rays from hadrons [58], but CRs in general. The following subsections will explain the structure, learning types, training and evaluation of a ML task.

Neural Networks

To solve ML tasks artificial neurons are used. They receive inputs x which are mapped to an output using an activation function like the rectified linear unit (ReLU)

$$a(x) = max(0, x)\,.$$

The output depends not only on the activation function but also on the weighting $\vec{w}$ of the inputs. Higher weighted input data is more likely to be considered as 'important' and thus contributes more to the output.
The simplest ML architecture consists of one artificial neuron solving only linear problems. The more complex the problems become the more neurons are needed building a

NN. These networks are divided into multiple layers which consist of numerous neurons, so called nodes. The first layer is the input layer followed by some hidden layers. The result is given by the output layer. Several layers can be connected in series where each node is linked to the previous and following layer node except bias nodes. A bias node is an additional neuron with value one which provides more flexibility to fit the input data. It is used in hidden layers and in the input layer. It is also possible to fully connect each node with every node from the previous and following layer. Those layers are then called fully connected layers.
The aim of using NN is to describe a highly complex application by one function $h_W(\vec{X})$ which shall approximate the target function

$$f_{target}(\vec{X}) = \vec{y}_{sol}\,,$$

where $\vec{X}$ is the selected input data and $\vec{y}_{sol}$ denotes the right solution of the given application. An example of a simple architecture of a NN is illustrated in figure 3.3.

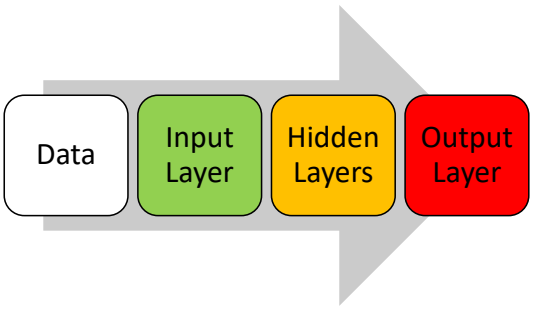


Figure 3.3. – *Schematic structure of a simple model. Input data is given to the first layer of the model. The data is further computed by the hidden layers leading to an output contributing to the specified task.*

There are several ways to train a NN for different tasks and given examples. Two are introduced in the following.

Supervised Learning

Scientists distinguish between supervised and unsupervised learning. In unsupervised learning there is no labeled training data provided and the NN has to find out how to solve the given problem by itself. In this thesis only supervised learning is discussed. Further details about unsupervised learning can be found in *Title* from author et al. [57].
In supervised learning the user provides a training data set $\vec{X}$ to the computer which contains in addition the right solution $\vec{y}_{sol}$ of the application. In general, supervised learning can be divided into two kinds of problems. The first one is the *regression problem*. It predicts a continuous quantity solution, a real-value output, for a given problem. A quantity could be something like the energy of incoming Cherenkov photons

of a shower where the energy value describes the real-value output and the name of the quantity would be the word "energy".
The second one is the *classification problem*. As the name implies, it maps the input data to their predicted classes. A classification problem can be defined by its characteristics, called features. For instance, a shower was detected by H.E.S.S.. The self-learning program has to decide whether it is a proton shower or a γ-ray shower based on the detected image. The features have to be extracted from the image (input data). A two class problem (e.g. proton- and γ-ray shower) is also called binary classification and a problem considering more than two classes is a multi-classification problem.

As already mentioned, to solve highly complex problems, deeper networks with more hidden layers are needed. In the following section these deep neural networks (DNNs) are explained in more detail.

3.2.1. Deep learning

Deep learning is a subset of ML using multiple layers. Often, a given problem is too complex to be described mathematically, so deep learning methods are used. They can extract highly abstract features [29] from training input data and use them to make predictions to different input data. Deep learning architectures are applied to e.g. speech recognition and image classification. All used DNN architectures in this thesis are introduced in this section.

Convolutional Neural Network

The convolutional neural network (CNN) is the most common one. CNNs are applied to data with grid-like topology e.g. images. They use the mathematical operation *convolution* in at least one of their layers [29].
In 2D a convolution is equal to a matrix multiplication of an image with a filter kernel. As an example one can imagine the original image as a 6×6 matrix and the filter as a 3×3 matrix applied on every pixel. The *kernel* operates on an area of 3×3 pixels in the original image and maps those to one pixel. This is repeated until the whole image is scanned by the filter kernel. The size of the output matrix is dependent on the step size of the convolution, *stride*. If in the given example the stride is equal to 2 the output would correspond to a 3×3 array which contains all features of the original image and is in ML commonly named *feature map*. If an input image is square with odd side lengths (e.g. 5×5 or 7×7) or has sides with unequal length (e.g. 6×7 or 6×8) the borders are filled, for instance, with zeros to get a square image. This modification of the input image is called *padding* and is usually used by convolutional (conv) layers and *pooling layers*. In CNNs it is common to insert pooling layers between conv layers which reduce the size of their output by focusing it onto its significant features.
The connection between conv layers differs from fully connected layers. Each neuron from the following layer is connected to a local field to the current layer. Such regions are called *receptive fields* and are of the same size as the kernel. Extending this definition

one feature contains information about a whole region of a specific size in the input image. These receptive fields are stable to simple transformations [43] and are used to verify the number of conv layers in CNNs.
In this thesis conv layers are a part of every build NN due to their important ability of extracting features on their own.

Residual Network

Residual Networks (ResNets) were introduced to rectify or to understand the degradation problem of training accuracy, see section 3.2.3, which occurs mostly for deeper networks [34]. In ResNets the convolutional layers are not only connected to their following layer. In addition a side path leads to an identity layer. These map the input back to themselves meaning the input equals the output. The result from an identity layer is concatenated to another following layer, thereby gaining additional degrees of freedom. This leads to an improved learning [34].
Residual layers were included in the architectures during the research phase of this work.

Recurrent Neural Network

Recurrent Neural Networks (RNNs) are characterized by the repeated application of the same operation on a sequential input. A sequence is interdependent data like a video. Its images are arranged in chronological order and are provided one by one to the RNN. RNNs are capable of finding correlations between the previous input image and the next. An extension of RNNs are Long Short Term Memory (LSTM) networks using LSTM cells. These support RNNs to extract and learn long-term dependencies. The core idea of LSTMs is to regulate an internal cell state, add or remove features, via gates [51].
In terms of shower images detected by H.E.S.S. telescopes, the use of RNNs is not immediately apparent. Since the telescopes do not trigger at the same time due to time shift of incoming Cherenkov photons, detected images can indirectly be seen as a sequence. In this work RNNs were used in the binary classification task.

Siamese Network

Siamese networks are used for similarity measure between two images. In comparison to the previously introduced models, the Siamese network consists of two identical networks which are concatenated after the output layer. These networks are not only identical, they also share their weights, hence the name Siamese network. The internal network's architecture differs from standard classifiers only in the last layer. Instead of a probability, the output vector contains the extracted weighted features of the input image. In this thesis the two images contain detected showers from different initial particles in diverse variations caused by:

- varying size of the shower due to energy fluctuations
- position of the shower in the image generated by different impact points

- expressions in the meaning of inner shower shape caused by diverse interactions of particles with air molecules
- varying background dependent on night sky, clouds and dust

In Siamese networks a distance metric learns to evaluate the similarity of provided training data $\vec{X}$. The distance metric utilized in this thesis is the Mahalanobis distance d_M, see equation 3.1, which searches for a linear transformation weighting matrix projecting the two outputs $h_W(X_a)$ and $h_W(X_b)$ of the Siamese network into a subspace. There the Mahalanobis distance equals the Euclidean distance [39].

$$d_M(X_a, X_b) = \|h_W(X_a) - h_W(X_b)\|_2 \tag{3.1}$$

where X_a and X_b are the inputs provided to the identical networks.
Since the detected showers are similar if induced by the same type of initial particle and different if created by another kind of initial particle, it was an approach chosen for this thesis that uses these two properties to classify the cause of the EAS. There the similarity measure and classification task are combined. Due to this, it is called multi-task network. It is based on [40] which unites for the first time classification and similarity measure.
The multi-task network was adapted for this thesis and implemented in the existing framework.

The model is composed of many statistical and numerical methods of physics and mathematics. However, the problem determines the architecture of the model and therefore how to proceed to get the right solution regarding to the task. The following sections explain the learning phase of a NN grouped into training and evaluation of a model.

3.2.2. Training

Before the training can be started the architecture has to be build. Between each convolution and fully connected block and between each fully connected layer a dropout is insert in the used architectures introduced in section 5.2.2.
Dropout is an approximation method used for regularization of a NN. It supports the NN during the learning phase by dropping out nodes which means setting a defined number of weights to zero. The probability that a neuron is set to zero is given by the dropout value defined by the user. After one learning epoch a model consisting of n neurons has 2^n possible variants of itself [59]. During the validation phase these are combined in one NN sharing their weights. There the NN is tested on unseen data without applying dropout. The nodes dropped out during one learning epoch are also considered as zeros. The resulting NN can be seen as the average over all possible variants. The dropout prevents the stagnation of learning accuracy and over-fitting. Both are explained in section 3.2.3.
In addition one can optimize the learning of the NN by tuning the hyperparameters which regularize the learning. The defined hyperparameters and their definitions are listed in

table 3.1. In the case of similarity measurement two additional hyperparameters, m and λ_s, were introduced.

Table 3.1. – *Hyperparameters for classification and multi-task NNs.*

Parameter	**Definition**
β	average over weights of the last training steps
No. epochs per decay	epochs after which the learning rate decays
α decay factor	learning rate decay value
α_{init}	initial learning rate value
staircase	exponential or gradual decay of learning rate
m	minimal distance between two dissimilar images
λ_s	scale value for contrastive loss

To find the function which describes the given problem best the NN has to be trained on an example dataset. During training phase input data $\vec{X}$ is gradually presented to the model, in so called batches. Batches are used due to computational limits. A higher batch size is better because this means more train examples are presented to the NN at once and therefore more variability of one class is presented to the NN. The weight W_{jk}^{l} of node j in layer l is adjusted to the associated input ${x_{in}}_k^{l-1}$ of node k in layer $l-1$. This process is repeated multiple times. In supervised learning the calculated outputs $a({x_{in}}_k^{l-1})_j^{l_{fin}}$ in the final layer l_{fin} are compared with the true solution Y after each iteration. The error of the final layer between predicted and desired output is calculated by a loss function L. The summation over each error of all neurons using back-propagation [17] is defined as the cost function $J(W)$. In back-propagation the cost of each neuron is calculated via the derivative of the loss function which is affected by weights and biases of the associated node. Ensuing the costs are summed up. An optimizer is used to minimize the cost of the model. This forces the system to adjust weights for the application and in the end to predict the correct output.

Loss Function

The loss is calculated during the training of each example. There are several loss function for each task. The ones used in this thesis are presented below.

Cross Entropy Loss works best in classification tasks because it avoids saturated states and prevents gradient vanishing [36] which are essential components of optimizers. The general form of cross entropy loss is

$$L_c(W, \vec{y}_{h_W(X)}, X) = -\sum_{class}^{N} y_{h_W(X),\,class} \log(P_{h_W(X),\,class})$$

where $\vec{y}_{h_W(X)}$ is a one-hot vector of dimension $1 \times N$, which has $N-1$ zero entries and a one at the correct label $y_{h_W(X)\,,class=sol}$ position. N is the number of classes and $P_{h_W(X)\,,class}$ is the predicted probability of an input example X to be of class $class$.

Contrastive Loss is used in Siamese networks to minimize the distance between similar image pairs and maximize it for dissimilar pairs. Using the Mahalanobis distance the contrastive loss has the following form:

$$L_s(W, y, X_a, X_b) = (1-y)\frac{1}{2}(d_M)^2 + y\frac{1}{2}max(0, m - d_M)^2$$

where the margin $m > 0$ defines a radius around the feature vector $h_W(X)$ [31]. The aim is to keep the distance d_M between classes larger than the margin. The labels y differ from those of the traditional classification task. They do no longer represent the actual label of the input image instead they symbolize the similarity $y = 0$ or dissimilarity $y = 1$ of input image pairs.
In the course of this thesis a multi-task network uniting classification and similarity measurement is implemented in the preexisting framework. The loss is a summation of cross entropy loss with contrastive loss [40]:

$$L = L_c + \lambda_s \cdot L_s$$

Here the parameter λ_s is used as a scale value to balance the losses.

Mean Absolute Error is used in regression tasks. As the name implies the absolute difference between predicted and true real-value is calculated by

$$L_r(W, \vec{y}_{h_W(X)}, X) = |\vec{v}_{h_W(X)} - \vec{y}_{h_W(X)}|$$

where $\vec{v}_{h_W(X)}$ is a vector containing the real-value outputs of the model.
The error of one train example can be calculated by the loss function whereas the cost of the whole network is determined by sum over the loss of all train examples.

Optimizer

As mentioned above optimizers in ML are used to minimize the cost function by updating the networks weights iteratively based on the loss function's gradient. In addition, the gradient is scaled by a learning rate α. This parameter α_{init} has to be chosen carefully. A too large learning rate causes the optimizer to jump back and forth between the optimal or local minimum not converging. Whereas using a too small learning rate the NN needs very long training time. The best case is to find the global minimum and not a local minimum or saddle point.

Adam is the most popular optimizer used for classification tasks due to its good result and low computation time. It is an algorithm based on first-order gradient stochastic optimization to obtain adaptive estimates of lower-order moments [42]. In comparison to higher-order optimizers Adam is able to handle noisy data with sparse gradients and with online, update of events after each iteration, and non-stationary cost functions.

Nesterov Adam (Nadam) is an extension of optimizer Adam adding Nesterov momentum [22]. This leads to faster decay along directions with small and nearly constant

gradients whereas the decay is slower along directions with strongly fluctuating gradients [61]. The optimizer is robust against applications having very complex features. Detailed information about Nadam can be found in *Incorporating nesterov momentum into adam* from Dozat et al. [22].

After each training epoch and after the learning phase a performance evaluations is done to test the NN on unseen data. Therefore validation and test set have to be created. The following section describes methods to evaluate the NN depending on the specific task.

3.2.3. Performance Evaluation

The training of the NN will continue until the criterion $h_{\vec{w}}(\vec{X}) \approx f_{target}(\vec{X})$ has been fulfilled. In practice the function of the application $f_{target}(\vec{X})$ is unknown, therefore other verifications have to be done.
In classification and regression the evaluation during training of a NN is done using a validation set. It contains data other than that used for learning and is only used to monitor the performance of a NN during training with the current weights and chosen hyperparameters. The loss of a validation set is an indication of how well the model behaves after an iteration. A decreasing validation loss is desired. If the loss remains nearly constant or even increases the model might be under-fitted or has run into an over-fit respectively.
Under-fitting occurs if the model can not capture all complex features of a problem. This usually comes along with low variance and a high bias. Whereas a model is over-fitting if it performs perfect on the train examples but guesses on unknown data [9]. In the case of high bias it is recommended to either train longer such that the loss is sufficiently minimized or to stop the training and redo it with other hyperparameters, more training examples or an extended architecture. Over-fitting can be prevented by choosing a smaller model or by stopping the training and reset the weights to an earlier state where the validation loss is minimal.
Additionally, in classification tasks the evaluation of the NN performance offers metrics such as accuracy, precision, recall, F_1-Score and receiver operating characteristic (ROC) curve [63].

- *Accuracy* is a measure of how good the classifier performs on the current data. It can only be used if the dataset contains equally distributed examples of all classes.
- *Precision* can be described as the proportion of correctly labeled events to the sum over all as positive predicted examples.
- *Recall* is the amount of all as positive predicted events and is also known as the sensitivity of a classifier.
- F_1*-Score* or F-measure is the harmonic mean of precision and recall. It is used if the example distribution of the classes are unequal or the focus is on the precision and recall of a classifier.

- *ROC* curve is determined with the area under the curve (AUC). The ROC curve is a measure of the models ability to distinguish between classes. A high AUC value, $> 90\,\%$, means the NN is able to separate good between the classes. However, a value $< 50\,\%$ means the classifier is guessing a class. In the ROC curve the recall, also called true positive rate, is plotted against false positive rate which is determined by (1-specificity). In figure 3.4 an example of a ROC curve is shown with descriptions. At point A a classifier is correctly predicting TNs, at point B the classifier correctly predicts all examples which would be the optimal classifier, at point C the classifier correctly predicts all TPs and at point D the classifier misclassified all events. The black curve is the expected result of this illustrated classifier and is calculated by the gradient of recall and the false positive rate.

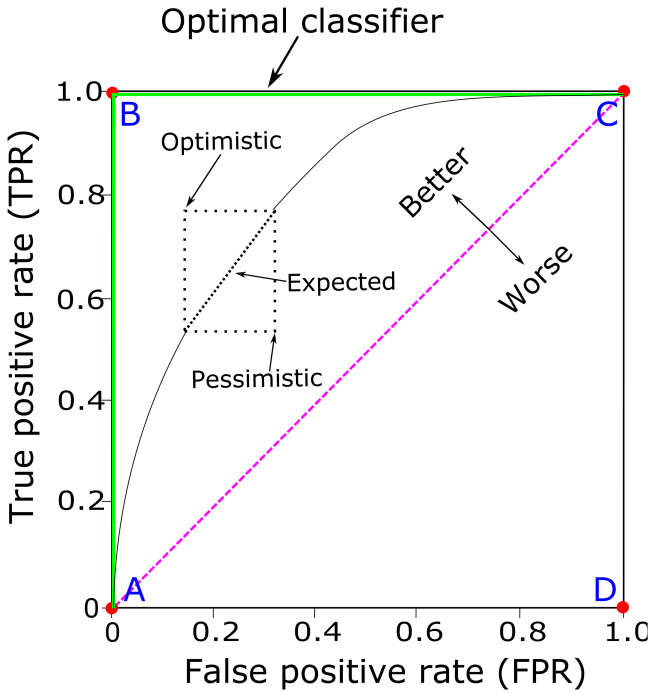


Figure 3.4. – *Illustration of a ROC curve. The green line represents the case of an optimal classifier and connects the points A, B and C. The violet, diagonal, dashed line marks the border between better or worse classifiers. Figure extracted from [63].*

The metric is further divided into binary and multi-classification evaluations. In binary classification an additional metrics, specificity and ζ-distribution, can be calculated.

- *Specificity* is the opposite of recall. It is the amount of all as negative predicted events [63].

- *ζ-distribution* is the distribution of events predicted with a ζ-value $\in [0, 1]$ [58]. The ζ-value is the output of the classifier. In this thesis an output of > 0.5 means

the classifier predicts this events as signal whereas an output of < 0.5 is denoted as background.

The general form of a confusion matrix in binary classification is shown in table 3.2. There the predicted events are subdivided into four categories. Based on this table the metric can be easily defined mathematically. The formulas are written down in appendix A.

Table 3.2. – *Symbolic confusion matrix in binary classification.*

Actual: / Predicted:	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

In multi-classification the confusion matrix slightly differs from the binary classification matrix. An example is shown in figure 3.5. The predicted labels, applied in rows, are compared with the true labels, applied in columns, of the single classes. Each column represents the actual true label as well as the predictions of a classifier for the corresponding label per row. Whereas each row represents the predicted classes and the occurring contamination by other classes. The aim is to achieve only entries on the diagonal of the matrix. Then one has found the perfect classifier for the application under the constraints that sufficient sample data has been provided to the NN and the classifier did not run into an over-fit.
Due to non-existing metrics for multi classification in the preexisting framework they were implemented during the research phase of this thesis.

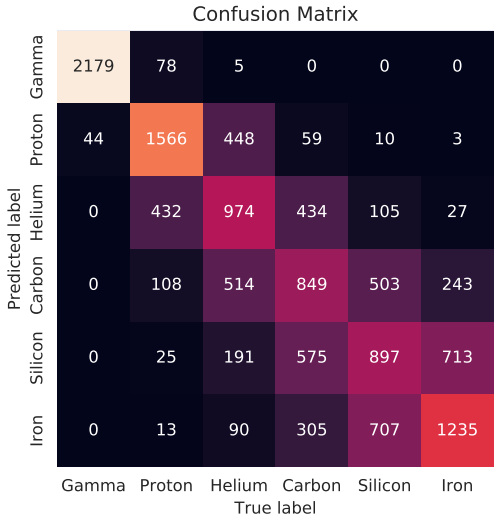


Figure 3.5. – *Example confusion matrix in multi-classification. The higher the values on the diagonal entries, the better this particles were classified. The best determined class was the one with label Gamma followed by Proton and Iron. All other classes were poorly distinguished. From the as Silicon classified events only* $\approx 37\,\%$ *are TP. In the case of a Multi-classifier with six different labels an accuracy of 16 % is equal to an estimate.*

The metric of regression tasks depends on the task itself. The performance of the energy estimator can be verified by the energy migration matrix and the energy bias E_{bias}.

- *Energy migration matrix* is a 2D energy resolved histogram containing the number of events reconstructed with a specific energy E_{reco} and true energy E_{true}. This matrix is directly related to the energy resolution and dispersion of the trained NN.

- *Relative energy bias* is defined as the mean value of reconstructed energy to true MC energy:

$$E_{bias} = \frac{E_{reco} - E_{true}}{E_{true}}$$

E_{bias} is a measure of the energy estimator resolution. The corresponding standard deviation can be derived by:

$$\sigma_{std} = \sqrt{\frac{\sum_{i=0}^{N}(E_{bias,\,i} - \bar{E}_{bias})^2}{N-1}}$$

where N is the number of bins within the energy bias range.

In conclusion NNs are tools to solve highly complex problems, e.g. classification of EAS or their energy reconstruction. The architecture and chosen optimizer depends on the task which determine the metric to evaluate the performance.
The general steps to solve a ML problem are:

1. Build NN model $h_{\vec{w}}(\vec{X})$
2. Determine training set $\vec{X}$ and associated label $\vec{Y}$
3. Define cost function $J(\vec{w})$
4. Minimize $J(\vec{w})$ via an optimizer
5. Adjust weights $\vec{w}$ of $h_{\vec{w}}(\vec{X})$
6. Repeat steps 4. and 5. until $h_{\vec{w}}(\vec{X}) \approx f_{target}(\vec{X})$

This thesis only uses TensorFlow for the creation and validation of neural networks. TensorFlow is an open source cross-platform software for numerical computation. It is implemented in C++ and Python and was originally developed by the Google Brain team to conduct machine learning algorithms and other scientific tasks [1]. All architectures presented below and evaluations were implemented in Python.

3.3. Spectral Analysis

The spectral analysis is implemented in the course of this work to the initial framework. It is done after the classification of the CR particles and reconstruction of their energy. The goal is to determine the source spectrum measured by H.E.S.S. by fitting the spectral index Γ and the absolute flux normalization Φ_0 of the differential energy spectrum $\frac{d\Phi}{dE}$, see eq. 1.1.
The differential particle rate is measured by H.E.S.S. and is described in terms of the reconstructed energy E_{reco}, the effective area $A_{eff}(E_{true}, \theta, \psi, \nu)$ and the probability density function (PDF) $P(E_{true}, E_{reco}, \theta, \psi, \nu)$ for each CR [60]:

$$\frac{dN}{dE_{reco}dt} = \int_{E_{min}}^{E_{max}} P(E_{true}, E_{reco}, \theta, \psi, \nu) A_{eff}(E_{true}, \theta, \psi, \nu) \frac{d\Phi}{dE_{true}} dE_{true} \ .$$

The PDF and the effective area are determined using MC simulations created at a zenith angle θ, azimuth angle ν and an offset ψ with an energy E_{true}.
First the on- and off-region has to be defined. The on-region is equated to the FoV of H.E.S.S. under the assumption that the detected signal is hadronic and diffusely distributed over the sky map. Consequently an off-region does not exist. In the case of a γ-ray induced EAS a direction reconstruction of the observed shower can be done after the determination of the on- and off-region.
The classified MC events in the on-region are grouped into their true labels. Then they are filled into two energy resolved E_{reco} histograms, one containing all true events of a particle type and the other containing events which were correctly classified by the trained NN. From these histogram pairs the effective area can be determined.

Effective Area

The effective area $A_{eff}(E, \theta, \psi, \nu)$ is an energy resolved and angular dependent correction factor to compensate the misclassified events of a particle type *type*. The geometric area $A_{geo} = \pi r^2$ of an EAS on the ground is multiplied with the ratio of TP and actual positives (AP):

$$A_{eff}(E_{true}, \theta, \psi, \nu) = A_{geo} \cdot \frac{N_{type}^{TP}}{N_{type}^{AP}} \ , \tag{3.2}$$

where N_{type}^{TP} is the number of TP of *type* MC events and N_{type}^{AP} is the number of all simulated *type* events.

Probability Density Function

The PDF $P(E_{reco}, E_{true}, \theta, \psi, \nu)$ indicates the probability of an event with a specific energy E_{true} to be reconstructed with an energy E_{reco}.

- The same MC events used for the effective area calculation are used to determine the PDF.
- The simulated events are filled in a 2D histogram E_{reco} vs. E_{true}.
- The 2D histogram is normalized such that each column contains the relative number of events:

$$\sum_{k=0,\ E_{true}=const.}^{N} E_{reco,k}\ dE_{reco} = 1$$

The created PDFs are used as look-up tables for specific θ, ψ and ν angles for comparison to real data. Further details can be found in *Emitters of VHE gamma-radiation as revealed by the H.E.S.S. Galactic plane survey* Hoppe [60].
In this thesis only PDFs with a zenith angle θ of 20 °, a maximum offset angle ψ of 2.2 ° and an azimuth angle ν of 180 ° are determined because the simulations were only carried out with these parameter values. An example PDF is shown in fig. 3.6.

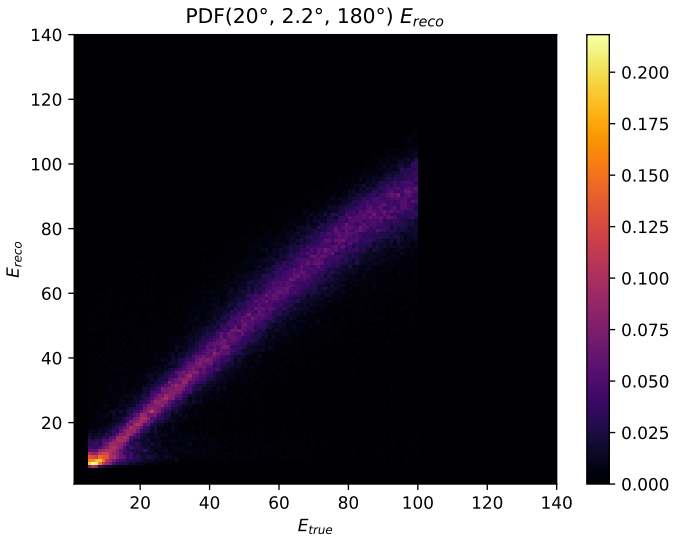


Figure 3.6. – *Illustration of a PDF with a zenith angle θ of $20\,^\circ$, a maximum offset angle ψ of $2.2\,^\circ$ and an azimuth angle ν of $180\,^\circ$. The diagonal stripe is an indication of the dispersion of the energy estimator. In the optimal case only the diagonal entries of the PDF would have values greater than zero resulting in a diagonal line.*

Spectrum Energy Threshold

The spectrum energy threshold is used to cut out energy ranges poorly reconstructed by the energy estimator. It only takes energies for which the energy bias E_{bias} is within the range of $\pm 10\,\%$ into account. The upper and lower energy boundary is derived by the deviation of the mean of correct predicted energies. The energies for which the $E_{bias} > 10\,\%$ are affected by large reconstruction errors. These would lead to systematic errors by the reconstruction of the source spectrum. In contrast, energies for which $E_{bias} < -10\,\%$ are not sufficient to determine a suggestive effective area [60].
Applying the spectrum energy threshold means to keep only events measured with energies $E_{reco} \in \mathcal{M} := \{x \in bias | x > 10\,\% \wedge x < -10\,\%\}$. The events are filled into energy resolved histograms. The number of those events in the off-region and on-region is denoted as N_{off} and N_{on}, respectively.

Forward Folding

Forward folding is the method used to determine the measured energy spectrum by fitting the parameters Γ and Φ_0 to obtain Φ_{fit}. The predefined values of $A_{eff}(E_{true}, \theta, \psi)$, $P(E_{reco}, E_{true}, \theta, \psi)$, N_{on} and N_{off} are needed to get the number of expected excess events given by the following relation:

$$n_\gamma \,|_{E_{reco,l}}^{E_{reco,l+1}} = \sum_{runs,\theta,\psi} t_{\theta,\psi} \int_{E_{reco,l}}^{E_{reco,l+1}} dE_{reco} \int_{E_{min}}^{E_{max}} \phi_{fit}(E_{true}) A_{eff}(E_{true}, \theta, \psi) P(E_{reco}, E_{true}, \theta, \psi)\ dE_{true}\ , \tag{3.3}$$

where $t_{\theta,\psi}$ is the observation time of a run obtained at a certain zenith angle θ and an offset ψ. $\phi_{fit}(E_{true})$ is the fitted spectrum.
The analytical solution of the number of expected background events is given by [60]:

$$n_b \,|_{E_{reco,l}}^{E_{reco,l+1}} = (\alpha(N_{on} + N_{off}) - (1+\alpha)n_\gamma)^2 + 4\alpha(\alpha+1)N_{off}n_\gamma\ ,$$

α is the background normalization factor derived from the number of determined off-regions, see reflected region method [13].
The probability of observing N_{on} and N_{off} events can be determined by the Poisson statistic fulfilling the following applicable assumptions

- events occur independently
- N_{on} events occur in an energy range ΔE
- constant particle rate, valid due to constant observations by H.E.S.S. in a time interval of 15 ns
- two events cannot occur at exactly the same time

The Poisson statistic can then be applied on the H.E.S.S. data given by the equation [60]:

$$P_p(N_{on}, N_{off} | n_\gamma, n_b) = \frac{(n_\gamma + \alpha n_b)^{N_{on}}}{N_{on}!} \exp(-n_\gamma - \alpha n_b) \cdot \frac{n_b^{N_{off}}}{N_{off}!} \exp(-n_b)$$

The first term describes the probability of observing a specific number of events in the on-region N_{on}. Those measured events are expected to consist of excess events n_γ and normalized background events n_b. The second term describes the probability of observing background events N_{off} over the whole off-regions if n_b are expected.
The measured flux is obtained by varying the assumed source spectrum and thereby minimizing the negative log-likelihood function ($-\ln L$):

$$\begin{aligned} -\ln L = -\ln P_p = N_{on} \cdot \ln(n_\gamma + \alpha n_b) + N_{off} \cdot \ln n_b \\ - ((1+\alpha)n_b + n_\gamma) - \ln N_{on}! - \ln N_{off}! \end{aligned}$$

During the optimization process the derivative of $-\ln L$ is determined. The constant terms $-\ln N_{on}!$ and $-\ln N_{off}!$ are then zero and can be neglected in $-\ln L$.
The optimal fit parameters inserted in equation 1.1 yield the measured energy spectrum multiplied by the ratio of measured events to expected excess events. The data points of the flux for broader bins are determined using the obtained spectrum Φ_{fit} and the term described by the observation time, effective area and PDF [60]:

$$\begin{aligned}\bar{E}_{reco,l} = \sum_{runs,\theta,\psi} t_{\theta,\psi} \int_{E_{reco,l}}^{E_{reco,l+1}} dE_{reco} \int_{E_{min}}^{E_{max}} dE_{true} \\ \phi_{fit}(E_{true}) \cdot E_{true} \cdot A_{eff}(E_{true}, \theta, \psi) \cdot P(E_{reco}, E_{true}, \theta, \psi) \cdot \frac{1}{n_\gamma}\end{aligned} \tag{3.4}$$

The observed flux of a specific source is [60]:

$$F(\bar{E}_{reco}) = \frac{(N_{on} - \alpha N_{off})}{n_\gamma} \cdot \phi_{fit}(\bar{E}_{reco}) \tag{3.5}$$

The negative maximum log-likelihood function can be approximated by an inverted Gaussian in 2D around its minimum. There an error estimation on the fit parameters can be done under the condition of a very fine scanning of the fit parameter space.
The probability that the found fit parameters fall within the true parameters value is given by R_{68} assuming the standard deviation σ, the expected number of events in the on-regions, are known [62].
The standard error 1σ on the estimators ϕ_0 and Γ form an ellipse centered around the global optimal fit parameters [62]. This behavior is given by [62]:

$$-\ln L = \ln L_{max} - \frac{1}{2}\ , \tag{3.6}$$

$\ln L_{max}$ is the maximum value of the log-likelihood function.
From this geometric shape, the standard deviations of the parameters and the orientation of the ellipse can be derived [62].

4. Methods

This chapter contains the methods used to create the datasets, apply preprocessing steps and determine the architectures. Further, an additional cut parameter was investigated to bias the NN decision respectively to heavier nuclei. Also, the Forward Folding algorithm adapted on the adjusted formula of the signal background description is explained. In addition, the application of the Forward Folding algorithm with the investigated formula on real data is described.

4.1. Dataset Creation

The dataset is very important for the specific task and performance of a NN. If not chosen carefully the results of the NN may be insufficient for further analysis steps. In the case of training a classifier the classes should occur in an equally ratio to prevent asymmetric learning. The classifier would then have a bias towards the class with more train examples which wants to be avoided in most cases.
All datasets contain MC simulations which were created in the southern hemisphere under a zenith angle of 20 ° and azimuth angle of 180 °.

4.1.1. Conditions

For the creation of the dataset standard analysis cuts, the same as used in standard analysis, were applied on the MC simulations. They are listed in table 4.1. The energy and offset parameters were defined by comparing the values in the MC simulations of all classes. Therefore only events with a maximum offset of 2.2 ° and within an energy range of 5 TeV to 100 TeV are allowed.
Only those who fulfill the cut conditions will be saved in a predetermined list. The list is compared to a histogram limiting the maximum number of events N_{max} for each energy bin. It is created by an energy spectrum with a user provided spectral index Γ:

$$k = \frac{1}{N_{tot}} \int_{E_{min}}^{E_{max}} \left(\frac{E}{E_0}\right)^{-\Gamma} dE$$
$$N_{max}|_{E_i}^{E_{i+1}} = \frac{1}{k} \int_{E_i}^{E_{i+1}} \left(\frac{E}{E_0}\right)^{-\Gamma} dE \ ,$$

where N_{tot} is the number of train examples and k the normalization factor. The normalization energy E_0 is set to 1 TeV and the bins are constructed such that they increase logarithmic in size due to lower statistics in regions of higher energies.
The conditions of equal energy distribution of the classes was introduced to prevent the

classifier from learning energy dependencies between the classes for example, iron events are more likely to have high energies compared to γ-ray events which mostly appear in the low energy sector.

Table 4.1. – *Cuts used to create the dataset containing EAS shower events to ensure more homogeneous data.*

CT1-4 parameter cuts	**Values**	**Units**
multiplicity	2	
size	$[60, \infty[$	p.e.
local distance	$[0, 0.525]$	m
offset	$[0, 2.2]$	deg
energy	$[5, 100]$	TeV

4.1.2. Preprocessing

Before the preprocessing steps the images are unnormalized and consist of $960\,\mathrm{px}^2$ hexagonal pixels. In this thesis only *rebinning* was applied on MC simulated images to train NN. For a better comparison real data images were rebinned too. Usually, interpolation is used on real data to assign pixel intensities with a pixel map to compensate foreign influences e.g. gravity. This can be corrected using the pointing correction.
After applying rebinning the images have a resolution of $42 \times 42\,\mathrm{px}^2$. This value was chosen due to technical limitations. Another reason was the following consideration. How many pixels are needed to convert the original image, see figure 4.1, into a square one? The needed number of pixels to fill up the edges is derived by using the following equation:

$$\mathbf{4} \cdot [3 \cdot (4\,\mathrm{px} \cdot 4\,\mathrm{px})] = 192\,\mathrm{px}^2 \ ,$$

the bold **4** is the number of edges and the term $[3 \cdot (4\,\mathrm{px} \cdot 4\,\mathrm{px})]$ subdivides the four edges into three partitions with $4 \cdot 4$ pixels each. The result is a square image with 1152 hexagonal pixels in total. The transformation of one hexagonal pixel with a circumradius R into a square one can be done by taking the ratio of their areas:

$$\frac{4R^2}{R^2 \frac{3}{2}\sqrt{3}} = \frac{8}{3\sqrt{3}} \ .$$

Multiplying this ratio with the total number of hexagonal pixels results in a resolution of $\sim 42 \times 42\,\mathrm{px}^2$.

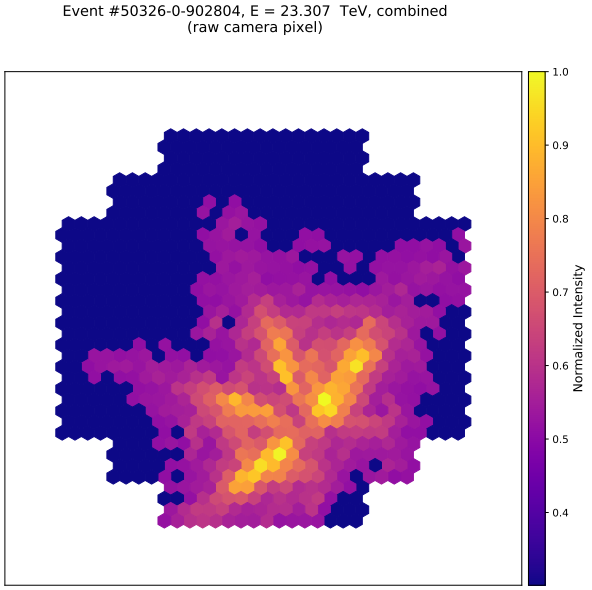


Figure 4.1. – *Combined, raw image of a MC He^+ nuclei EAS. The EAS was simulated with an energy 23.307 TeV. The inner and outer EAS gradient shape is similar to a star. He^+ nuclei EAS can not be described by an ellipse but this image contains features which look like an ellipse e.g. the pixels having higher intensities.*

The next preprocessing step is the application of *normalization*. The images are scaled such that the mean is zero and the standard deviation is one. In regression tasks an additional scaling factor is used to normalize the target value and the response of the NN to values between $[0, 1]$. In the case of the energy estimator the scaling factor was determined by the chosen maximum energy. A logarithmic and a linear scaling were tested to find the best normalization method.
An optional preprocessing step *PCA* was tested to find correlations between the classes and therefore reduce the number of dimensions. PCA was applied using variances 90 %, 95 %, 99 % on two different datasets.
PCA was first applied on a dataset contained γ-rays and protons. It was applied despite the very good separation of these two classes by the Impact analysis and ML approach to test if it can improve the classifier performance.
The other dataset contained silicon and iron events because the separation of silicon and iron by the classifier works moderately good. PCA is used to support the distinction. Subsequently, the model was trained using the dimensionally reduced shower events and comparing the result with the performance of the original dataset without this preprocessing step.

4.2. NN Definition and Evaluation

The determination of the initial particle was done using NNs. They were trained on six different classes denoted as Gamma, Proton, Helium, Carbon, Silicon and Iron. To get the best performance different architectures with diverse approaches were tested. In addition, the energy reconstruction of all six initial particles was done.

4.2.1. Architecture Selection

The minimum number of convolutional layers in the trained NN is determined by the resolution of the receptive field r_{out} which can be calculated by the following formula [21]:

$$r_{out} = r_{in} + (k+1) \cdot d_{in} \ .$$

The receptive field size depends on the input feature map size r_{in} and the kernel size k of the parameters and distance between two adjacent features d_{in} in regard to the stride size s. The number of layers in relation to the resolution of the receptive field is displayed in fig. 4.2. The graphic shows the resolution of the receptive field over the number of layers. The total number of layers includes the number of convolutional layers and pooling layers. The combination can be done due to the structure of the NN architecture. Between two convolutional layers one pooling layer was inserted.
Defining a resolution of 42 px a minimum number of six layers in total are needed to cover the original image including 3 convolutional and 3 pooling layers. In the figure 4.2 the resolution 31 px is the square root of 960 px^2 which was the total number of pixels in the originally hexagonal image. Whereas the resolution of 64 px was the best approximation after applying rebinning on the MC images investigated by [64].
Four different approaches and 6 architectures were tested to find the number of convolutional and fully connected layers for which the classifier performs best. The architectures altered in their number of convolutional layers from 3 to 4 and also the number of fully connected layers was varied from 3 to 5 which results in a total number of 6 combinations. The four approaches include the combined image provision, channel approach, the single image approach and the usage of a LSTM cell. The first approach was to gradually present combined CT1-4 images to the model. The CT1-4 images of one event were summed up resulting in a so called combined image. Those architectures are mark with a star *. The second one presented the input as a stack of all CT1-4 images treating them as RGBα channels but only in the first convolutional layer. These networks can be recognized by the shortcut α. The third approach provided all CT1-4 images one after another to the convolutional layers. The weights were shared which is comparable to a Siamese NN. The four outputs of the convolutional layers of each CT image were concatenated treating the feature maps as one feature vector. The vector was further provided to the fully connected layers. Last a LSTM cell was added to the NN. It was inserted between the convolution part and the fully connected layers.The architectures are then denoted as CNN_*, CNN_α, CNN and RNN in the corresponding order.
To decide which classifier has the best performance of the ζ-distribution in over the whole defined energy region of 5 TeV to 100 TeV was compared. This was done due to

the known bias of ±10 % within energy range 100 GeV to 100 TeV of H.E.S.S. [53]. The result of the architecture study was used to create the base frame of the architecture for the multi-classification task.

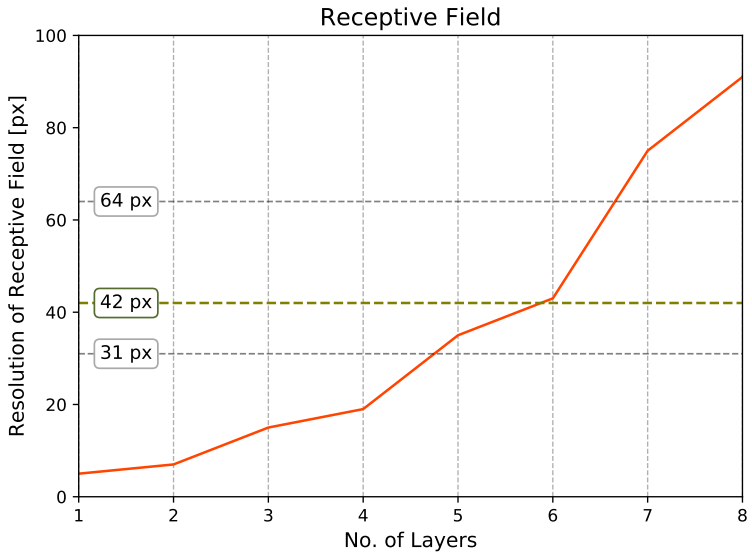


Figure 4.2. – *The growth of the size of the receptive field equals an exponential increase. The graph includes all layers, alternating convolution layer and pooling layer. Each step from an even number to an odd one shows the increase of the receptive field by applying alternately a convolution layer and the pooling layer. The resolution 31 × 31 px^2 needs at least 5 layers to cover the hole image and the resolution 64 × 64 px^2 needs at least 7 layers. For the chosen resolution of 42 × 42 px^2 in this thesis 6 layers are necessary.*

The NN architectures for the regression tasks were chosen by modifying the existing ones already fulfilling the condition of having enough convolutional layers. A recurrent layer was added between the 4th and 5th layer concatenating its output to the penultimate layer. The provided input for the task of energy reconstruction included the CT1-4 images and the event size. The size was added to the feature vector in the fully connected layers to support the learning. The size is known for MC simulations and real data. In addition, the dropout for this task was set to 50 %.
In the multi-classification task two additional network architectures were investigated. Due to the moderately good results of the energy reconstruction of the MC γ-ray EAS using the additional parameter size, the EAS image abilities size, skewness and kurtosis were also added to the classifiers in order to see if it improves the performance of multi-classification. This network is denoted with CNN_{SK}. The second tested DNNs were a combination of Siamese and classification tasks. The architecture differed only in the last layer which was divided into a classification and a feature vector output. For

the multi-task network the channel approach and single image convolution were tested. They got the shortcut SCNN_α or SCNN. The combined CT1-4 approach was not tested in the multi-classification task because no reasonable argument exists to try this approach and also the RNN architecture was not tested because it was computationally to expensive.
The output of the multi- and multi-task classifiers is a probability vector containing the six ζ-values for each class. The index of the highest ζ-value was chosen to determine the predicted class by the NN.

The best found base frame architecture was used for all classification tasks. In addition, the same datasets for all multi- and multi-task classifications were used. In the binary classification the same dataset containing γ-rays and H^+ events was used for all tasks. Further the architectures performance were optimized by choosing suitable hyperparameter.

4.2.2. Hyperparameter Optimization

The hyperparameter tuning helps to generalize the data and to find the optimal NN. The best hyperparameter set was determined by taking the same architecture and change the hyperparameters. The resulting performances were compared at the learning point where the loss of the validation set was minimal. Therefore the following conditions must be fulfilled:

- sufficient diversity of events in dataset
- training of the NN until it has converged or an overfit occurred
- the train set loss has no recognizable overfit during the learning phase

4.2.3. Model Evaluation

The goodness of the performance depends on the NN task. In the case of classification, the classifier having the highest accuracy was used to distinguish MC simulations and real data observed by H.E.S.S.. This is valid for binary classification and multi(-task)-classification. Whereas the regression task compares the energy yielding a $bias < 10\,\%$ and $bias > -10\,\%$. In addition, the corresponding resolution of the NN was derived by the standard deviation of the $bias$.
For further analysis of regression tasks there must be energies fulfilling the condition of having a $bias$ $\pm 10\,\%$. If not then the hyperparameters of the model have to be changed. If this has not led to an improvement in performance, it can be assumed that the architecture must be adjusted or an other analysis method must be applied to reconstruct the energy of the EAS.

4.3. Investigation of the DC-light Cut Parameter

DC-light is mostly found in heavier nuclei EAS images [6]. The idea was to get an additional cut parameter in order to support the distinction of the classifier. In [6] they investigated such a cut parameter for H.E.S.S. to get the real data initial particle and create an iron energy spectrum. In this thesis a similar approach was done to find the highest pixel in the camera image. Therefore the mean over all pixels in one image for 100000 events per class was taken. The pixel having the highest value was divided by the mean of the image. The outcome is a ratio of the highest pixel to the mean of per image. The results of all classes were compared to find a further cut parameter.

4.4. Forward Folding Application

The forward folding algorithm can be applied after the classification and energy reconstruction using the best NNs of each task. The effective area and the PDF were calculated using datasets consisting of MC simulations. From now on all as signal determined events can be either of hadronic type or γ-rays and all remaining classes are consequently background events. The following steps were done to make the forward folding algorithm applicable for the determination of heavier nuclei spectra.
First the implemented forward folding was tested on a MC dataset containing 128000 events in total. This was done to check if the input spectrum can be reproduced. The dataset was created with a spectral index Γ of -2 containing only iron nuclei events. To be able to compare the fit result with the original parameters the energy distribution of the MC events was calculated which has been defined during the dataset creation. From which the original flux normalization $\phi_{0,MC}$ can be derived. It is given for $\forall E_{true} \in \mathcal{M} : \{bias > -10\,\% \ \wedge \ bias < 10\,\%\}$ by:

$$\phi_{0,MC} = \frac{N_{MC}}{A_{MC} \cdot t_{MC}} \int_{E_{min}}^{E_{max}} \left(\frac{E_{true}}{E_0} \right)^{-\Gamma} dE_{true} \tag{4.1}$$

under the assumption that for MC simulations the area $A_{MC} = 1\,\mathrm{m}^2$ and also the observation time $t_{MC} = 1\,\mathrm{s}$. The normalization energy E_0 was set to the mean energy in the bias interval $\pm$ 10 %.
Further the whole FoV of 5 ° was defined as on-region consequently the background normalization factor α, the measured off-region events N_{off} and the expected background events n_b are set to 0. This means the measured events in total N_{tot} are the same as the events from the on-region N_{on}. The negative log-likelihood then simplifies to:

$$-\ln L = N_{tot} \cdot \ln n_{iron} - n_{iron} - \ln N_{tot}!$$

The differential flux in each mean energy $\bar{E}_{reco}$ bin is then given by:

$$F(\bar{E}_{reco}) = \frac{N_{tot}}{n_{iron}} \cdot \Phi_{fit}(\bar{E}_{reco}) \tag{4.2}$$

Then background was added to the dataset to derive the function which defines the expected background. The dataset contained MC Fe^+ (signal) and Si^+ (background) nuclei events and was used to investigate the formula of expected events in total containing TP and FP of one particle type:

$$n_{tot} = n_{signal} + \alpha \sum_{class} n_{b,class}$$

where the background normalization factor α is 1 and $n_{b,class}$ defines the expected number of a background particle wrongly classified as signal.
The formula was motivated by the assumption that the on-region is the whole FoV and the measured signal is a mixture of events from all classes. The valid formula of the determination of expected excess events is written down in section 3.3 in equation 3.3. However, the number of expected background events of one class per energy bin depends on the efficiency $\epsilon_{signal,class}$ and PDF $P_{signal,class}(E_{reco}, E_{true}, \theta, \psi)$ of the specific background class. The efficiency and PDF were both determined using MC simulations of the corresponding class. Considering the confusion matrix of the best classifier, the efficiency can be determined by

$$\epsilon_{signal,class} = A_{geo} \cdot \frac{N_{signal,class}}{N_{tot,class}}$$

where $N_{signal,class}$ is the number of events in the row of *signal* and the column of *class* and $N_{tot,class}$ the total number of events of the considered background class. In figure 4.3 the total number of the background class, e.g. $N_{tot,silicon}$ marked by the yellow color and the number of the Si^+ nuclei events classified as class Iron marked with green, is shown.

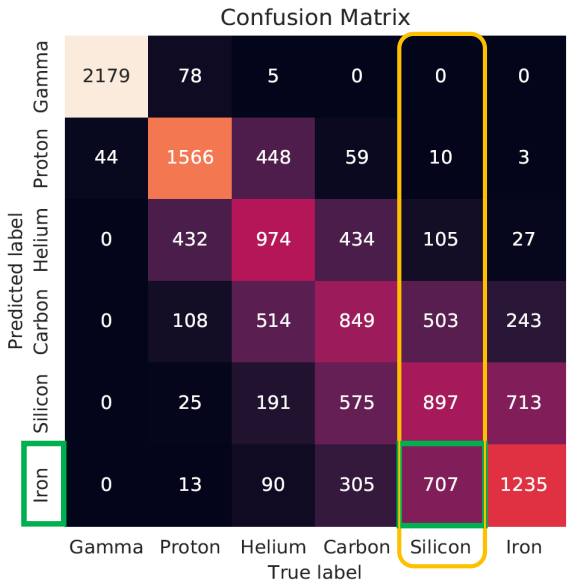


Figure 4.3. – *Example of the efficiency drawn in a confusion matrix. The yellow border includes the true events of class Silicon and the green square includes the number of events of Si^+ nuclei classified as Iron events. The ratio of the events in the green square to the sum of events within the yellow border is given by the efficiency.*

The PDF $P_{signal,class}(E_{reco}, E_{true}, \theta, \psi)$ is determined by the probability of a background event with a specific true energy which is reconstructed with an energy E_{reco} using the estimator trained on the signal class. The mathematically procedure is the same as described in 3.3.
Therefore, the number of expected background events of one class is given by:

$$n_{b,class}\,|_{E_{reco,l}}^{E_{reco,l+1}} = \sum_{runs,\theta,\psi} t_{\theta,\psi} \int_{E_{reco,l}}^{E_{reco,l+1}} dE_{reco} \int_{E_{min}}^{E_{max}} dE_{true} \\ \phi_{fit,class}(E_{true}) \epsilon_{signal,class}(E_{true}, \theta, \psi) P_{signal,class}(E_{reco}, E_{true}, \theta, \psi) \tag{4.3}$$

where $\phi_{fit,class}(E_{true})$ is the assumed energy spectrum of the background class. This formula is only valid for a dataset containing sufficient MC examples of each particle type and for a classifier trained on an adequate variety of events.
Applying these cognizances on the dataset containing Fe^+ and Si^+ events the resulting negative log-likelihood function is:

$$-\ln L = N_{tot} \cdot \ln(n_{iron} + \alpha n_{b,silicon}) - n_{iron} - \alpha n_{b,silicon} - \ln N_{tot}!$$

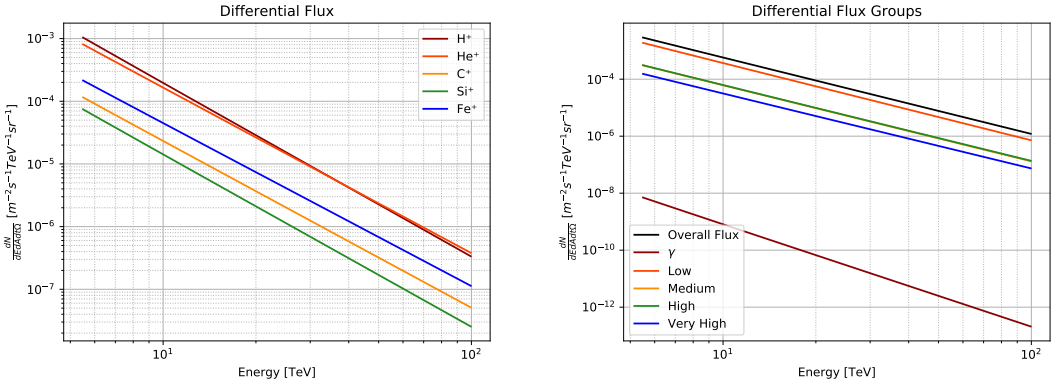
Extending these on a dataset containing multiple classes the formula can be generalized to:

$$-\ln L = N_{tot} \cdot \ln(n_{signal} + \alpha \sum_{class} n_{b,class}) - n_{signal} - \alpha \sum_{class} n_{b,class} - \ln N_{tot}! \quad (4.4)$$

The last application of the forward folding algorithm was on real data including the source PKS 2155-304. The events were classified by the best found classifier and their energy was reconstructed using the estimator trained on Fe^+ events because the spectral index and normalization flux of an Fe^+ spectrum shall be determined.
The classifier can distinguish between six classes γ, H^+, He^+, C^+, Si^+ and Fe^+. Real data contains events beyond those particles and therefore the total number of expected events in the case of real data is given by:

$$n_{tot} = n_{low} + n_{medium} + n_{high} + n_{VH} \quad (4.5)$$

where n_{low} is the number of expected hydrogen and helium nuclei, n_{medium} is the number of all particles with charges $Z \in [6, 8]$, n_{high} represents the particles $Z \in [10, 16]$ and n_{VH} contains all particles with very high charges $Z \in [17, 26]$.
These groupings were taken from *Cosmic Rays VII. Individual element spectra: prediction and data* Wiebel-Sooth [68] which were used for the creation of a MC dataset. In this paper a grouping of the observed data was done according to their charge Z. The spectrum parameter values of those groups serve as reference. Their differential fluxes as well as the individually fluxes are visualized in figure 4.4. In [10] similar groupings adapted from [68] were expanded to higher energies. These are later used for the background estimation on real data. Therefore the flux normalization ϕ_0 was recalculated at the normalization energy E_0.



(a) *Single particle differential flux.* **(b)** *Grouped differential flux.*

Figure 4.4. – *The differential fluxes were determined amongst other by satellite detectors and balloon flights [68]. In a) the differential flux of all particle types used for the training of the classifier is shown. The right graphic b) shows the differential flux over all grouped particle types of [68] sorted by the charge Z. In addition, the differential flux of PKS 2155-304 is shown. Its data was taken from [24].*

After the determination of the fit parameters ϕ_0 and Γ by the optimizer the result can be compared to the dataset. The chosen optimizer to minimize the negative maximum log-likelihood $-\log L$ was basin hopping. This optimizer is designed to find the global minimum of inverted Gaussian-like functions [52] which is true for the negative log-likelihood in the large sample limit [62]. The boundaries for the optimizer was set to $\phi_0 \in [0, \infty[$ and $\Gamma \in]-\infty, \infty[$. The error on both fit parameters was determined by the formula 3.6. In addition, the nonlinear least square problem was used to fit the expected distribution of events on the measured one. Therefore the curve-fit function was applied using the Levenberg-Marquardt algorithm under the constraint that no boundaries were set[3]. It is an numerical approach for solving non linear least square problems χ^2. Further details can be found in [48]. Reasons for the choice of this function are it considers the error on the measured data and return the fit parameters as well as the covariance matrix. The diagonal entries of the covariance matrix contain the fit errors. In addition, the confidence interval containing 99.999943 % of the standard error hold by 5σ [62] can easily determined using the curve-fit function. Whereas for the log-likelihood this would be to expensive in computational time. The measured events can be described by the Poisson statistic therefore the absolute uncertainty is given by:

$$\Delta N_{signal} = \sqrt{\left(\frac{\partial N_{signal}}{\partial N_{tot}} \Delta N_{tot}\right)^2} = \sqrt{N_{tot}} \tag{4.6}$$

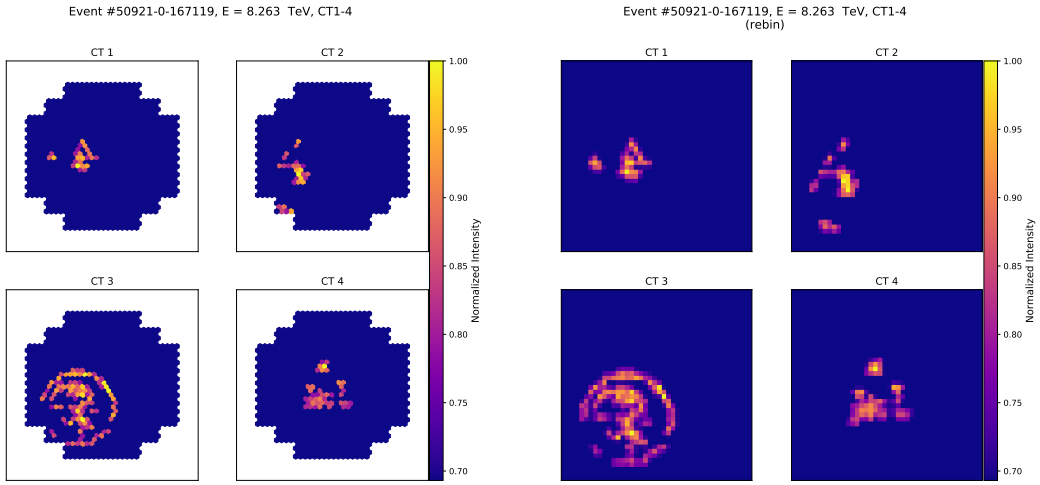
where the N_{signal} is given by $N_{tot} - N_{class}$ and is the number of measured signal events. Further errors like the statistic error on A_{eff} were neglected which could have been taken into account due to the small dataset the classifier was trained on.

5. Results

This chapter represents the results of this work. First the result of the preprocessing step rebinning is shown. Then the classification of cosmic EAS shower images is investigated focusing at first on the binary classification task. In the case of binary classification finding a reasonable good architecture, study the learning dependency of the NN on the energy distribution of the train examples and investigate the charge difference ΔZ is reviewed. Next, the application of the preliminary findings on the multi-classification task which is one of the main goals of this thesis is discussed. Followed by the investigation of the optional preprocessing step PCA. Last, the results of the energy reconstruction and the iron energy spectrum are shown which represent the second and third main part of this work.

5.1. Rebinning

Rebinning was applied to use the standard numerical algorithms of TensorFlow. The hexagonal pixel grid was transformed into an image consisting of square pixels. The chosen resolution is $42 \times 42\,\mathrm{px}^2$. An event of an iron nuclei from the MC simulated run 50046 is shown in figure 5.1. The MC raw images have the typical hexagonal shape due to the camera pixels applied on the Winston cones. The orientation and size of the images is preserved after the application of rebinning on the raw images. Even the feature of single pixels having higher intensity are retained in the rebinned CT1-4 images. The iron nuclei MC showers' morphology cannot be described by a geometrical form, thus it could not be analyzed with the standard analysis tools established for the γ hadron separation. In the CT3 image a muon is detected having a higher energy than its minimal emission energy which can be seen in the form of a muon ring. All CT1-4 contain a DC-light pixel which occurs if the primary particle is of hadronic type and which becomes also more likely with increasing charge Z.
In this work rebinning was also applied on real data. The usual preprocessing step is to interpolate the real data images to compensate the outer influences on the telescope which shifts the pixels. Despite the error it was neglected for this work due to the trained NN on rebinned MC event images.



(a) *Raw EAS* **(b)** *Rebinned EAS*

Figure 5.1. – *Example MC simulation images of an iron nuclei event. The run and event number can be taken form the title. Here, the first value* 50046 *indicates the run number and the last one* 550305 *is the event number. The energy of the simulated EAS is 11.83 TeV. Each CT shows an irregular shape which is typical for hadronic EAS. a) On the raw image of a MC event Image cleaning was applied. The image contains therefore only pixel groups* > 2*. In CT3 a part of a muon-ring was detected. b) The shape and orientation in the rebinned MC image is preserved. The image itself is blurred.*

5.2. Cosmic Radiation Classification

The distinction of the cosmic radiation via NNs is investigated by testing different approaches of providing CT1-4 images to the model. In addition, to achieve the best possible result the performance on different architectures and the dependency on the chosen energy distribution were studied. Further the binary task was expanded to a multi-classification task in order to try to distinguish all classes at once. Also the cut-parameter DC-light was investigated which could support the classifier in its distinction of heavier nuclei.

5.2.1. Binary Classification

First step towards the "best" possible architecture for binary classification was done by testing one dataset containing γ-ray MC simulations. The evaluation was done at the epoch having the lowest validation loss. The results are shown in table 5.1. The table contains the accuracy of each classifier. It can be seen there is no "best" architecture independent of the provided input to the NN.

The determination of the architecture using combined images did not reveal an explicit result therefore other approaches treating CT1-4 images as a RGBα channel image in the first convolutional layer, providing single images and using a LSTM cell were tested in order to find an indication.

Table 5.1. – *Tested number of layers in binary classification models. The results correspond to the models accuracy which were trained on datasets containing γ-ray and H^+ MC events with an energy distribution having a spectral index of $\Gamma = -2$.*

Layer		**Fully connected**			Classes
	No.	3	4	5	
Convolutional	3	97.33 %	97.22 %	97.30 %	CNN_*
	4	97.93 %	97.84 %	97.79 %	
	3	98.20 %	98.11 %	98.15 %	CNN_α
	4	98.04 %	98.13 %	98.14 %	
	3	97.03 %	97.66 %	97.66 %	CNN
	4	97.54 %	97.60 %	97.62 %	
	3	97.19 %	97.22 %	97.22 %	RNN
	4	97.39 %	97.53 %	97.52 %	

The single performances of the NN are similar to each other in the context of numerical uncertainties during the learning process. Thus, the ζ-distribution in the energy region between 5 TeV and 100 TeV is considered to try to evaluate which architecture is the best. The comparison of the ζ-distribution results in no definite difference for each ζ-cut. Taken into account the accuracy in that energy range the best classifier is the CNN_α having 3 convolutional and 3 fully connected layer with an accuracy of 98.26 %. The ζ-distribution of the classifier can be seen in figure 5.2. The ζ-cut is set to 0.5 meaning all events labeled with a ζ-cut greater than 0.5 are by definition signal events in this case γ-ray induced MC EAS. The background events are labeled with a ζ less than 0.5 , here the H^+ events. Due to the highest peaks at the ζ values 0.0 and 1.0, for background and signal respectively and low percentage of the events in between these ζ values it can be inferred that the classifier is evidently good in the distinction of γ-ray and H^+ MC events. In the direct comparison to the classifiers having an accuracy of 98.14 % in the energy region between $5 - 100$ TeV and are build of 3 or 4 convolutional and 5 fully connected layers no visible difference can be identified.
In addition, the average accuracy per architecture was used to try to determine the "best" architecture for the application of EAS classification. It is calculated by taking

the average of the accuracies per architecture in the table 5.1. The highest average accuracy 97.78 % do have the classifiers build of 4 convolutional and 5 fully connected layers. It must be mentioned that the difference in the average accuracy to the second best architecture is 0.2 %. Hence, it is not expressive.
Due to no coincidence of the two approaches to determine the best architecture and the fact that all architectures are nearly equal in their performance the architecture consisting of 4 convolutional and 5 fully connected layers was chosen for further analysis.

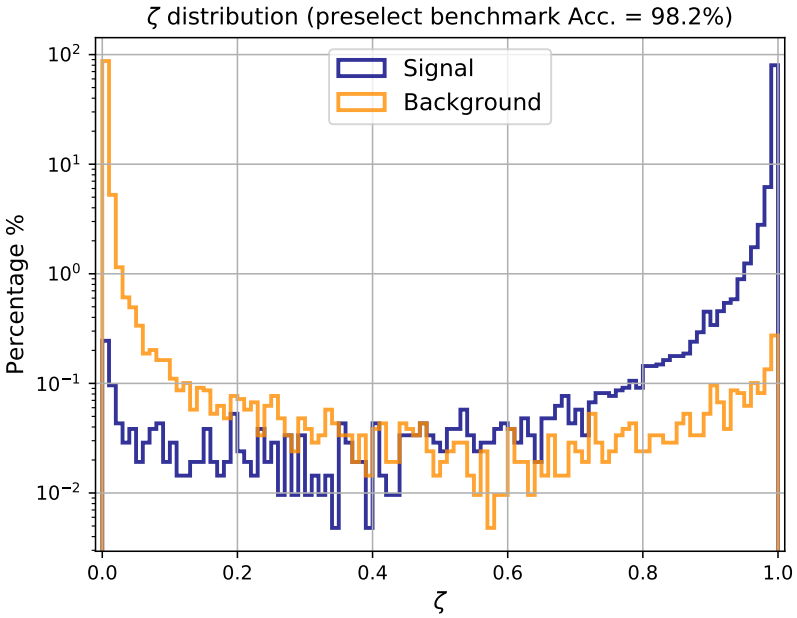


Figure 5.2. – *ζ-distribution in the energy range between 5 TeV and 100 TeV of the classifier build of* 3 *convolutional layers and* 3 *fully connected layers is plotted against the percentage of events per ζ value. The total number of events in this energy region is* 41666 *achieving an accuracy of 98.2 %. The signal and background events in these context are the γ-ray and H^+ events. The distribution has its peaks at the ζ values* 0.0 *and* 1.0 *indicating the classifier is ∼99 % and ∼97 % sure in its decision for the background events and signal events, respectively. In between the ζ values* 0.0 *and* 1.0 *the percentage of the number of events is low which supports the statement that the classifier can evidently good distinguish between γ-ray and H^+ events.*

The loss and the metric of the determined "best" architecture are shown in the figures 5.3, 5.4 and 5.5. The hyperparameter used for the training of the CNN_α are $\beta = 0.995$, No. of epochs per decay = 2.0, α decay factor = 0.5, α_{init} = 0.0005 and staircase = False.
The normalized loss per batch of the classifier shows a decrease in the first epochs for both sets. The loss of the training set steadily decreases. At epoch 6 the classifier has

its minimum loss value on the validation set. The increase after epoch 6 is due to an over-fit of the CNN_α which can be also seen in figure 5.4. There the accuracy of the validation set reaches its maximum value. After epoch 6 the specificity decreases further meaning the classifier learns in the over-fit to classify special shapes of H^+ events but unlearns features which describe in general the class Proton. Considering the recall of the validation set which describes how good the classifier can identify γ-ray events the course of this graph stays nearly constant. This can be an indication that γ-ray events have a typical elliptical shape in the most cases.
An other reference for the over-fit is the increasing metric of the training set. The NN would be nearly perfect on the training dataset at an epoch > 23 by proceeding the training which would also lead to a loss with a value close to zero including numerical errors.
The ROC curve shown in figure 5.5 is near to the optimal form which is similar to the geometrical shape of a triangle.

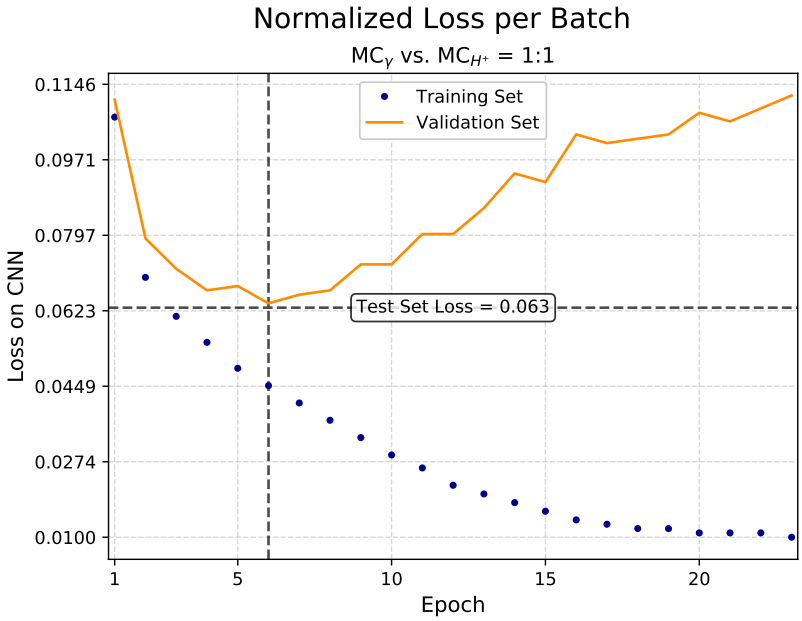


Figure 5.3. – *The normalized loss per training batch gives an overview of the current learning state of the NN. The loss of the training set is steadily decreasing. Whereas the validation set loss first decreases until epoch 4. There the α decays the second time due to the hyperparameter No. of epochs per decay which was set to 2. At epoch 5 the validation loss seems to over-fit but reaches its minimal value at epoch 6. After reaching its minimal validation loss value the NN is over-fitting which can be seen by the increasing validation loss curve and the further decreasing training set loss. The model was evaluated at epoch 6 having a test set loss of 0.063. It is marked by the dashed lines.*

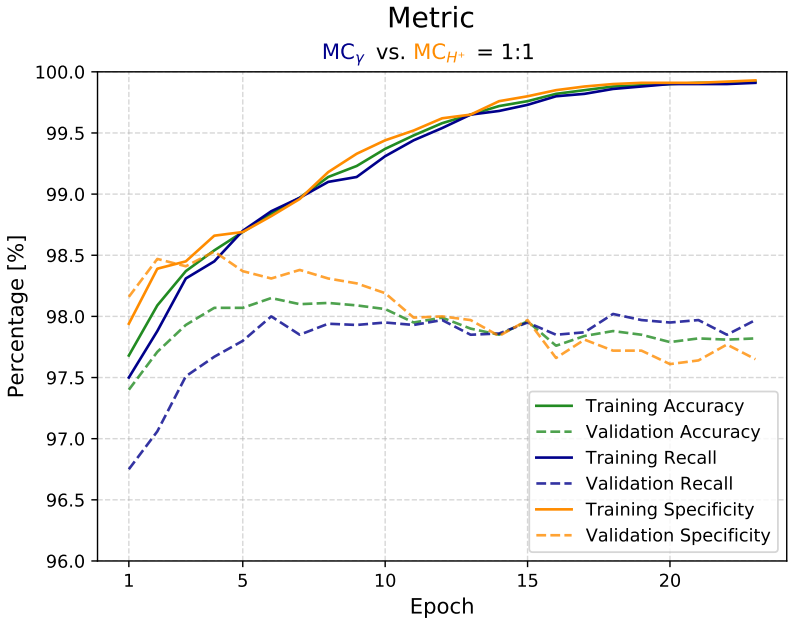


Figure 5.4. – *The evaluation metric of γ-rays vs. H^+ nuclei events. The accuracy, recall and specificity of the training set is steadily increasing. They are marked by the solid lines. The validation set metric is marked by the dashed lines. Here the color in the title indicates the class corresponding to the recall and specificity, respectively. The dataset contains the same number of events of each class which can be seen in the title "MC_γ vs MC_{H^+} = 1:1". In this case, the accuracy is for each epoch the mean value of recall and specificity. The course of the specificity recall has its peak in the beginning of the training. Its value is higher as those of the training metric. After some epochs the value decreases until epoch* 16 *where it stays nearly constant. The recall of the validation set represents the class containing γ-ray events. In the first epochs the NN learns the features of this class which can be seen by the increasing percentage and remains nearly constant after epoch* 6 *with small fluctuations. The validation accuracy has its maximum value at epoch* 6*. The NN is over-fitting after epoch* 6 *due to the decreasing accuracy.*

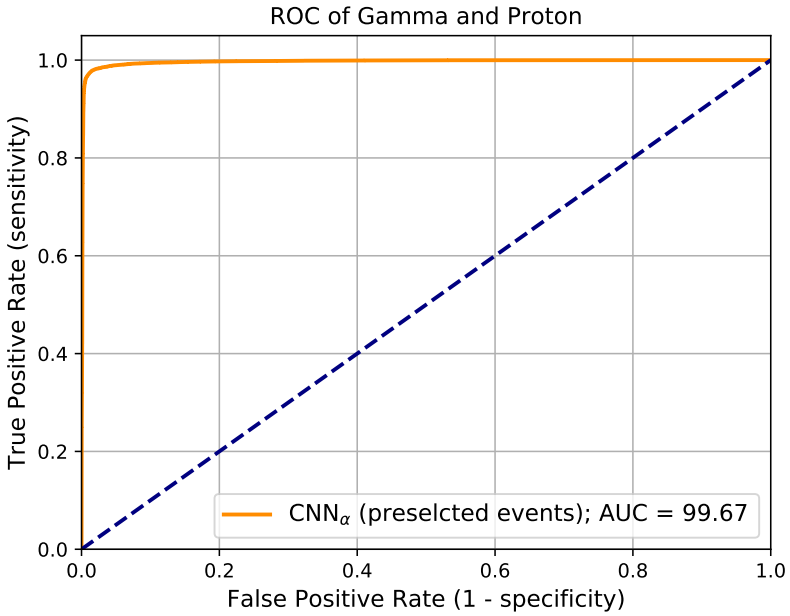


Figure 5.5. – *ROC curve of γ-rays vs. H^+ nuclei events. The dashed, blue diagonal line marks the minimal possible gradient of the orange curve. If the orange curve is the same as the diagonal line it can be concluded that the classifier is guessing both classes. Whereas the best case would be an appearing triangle with the edge points (0,0) $\xrightarrow{orange}$ (0,1) $\xrightarrow{orange}$ (1,1). The ROC of the γ-rays vs. H^+ nuclei events is close to the optimal shape. Here, the AUC is 99.67 %.*

Energy Distribution Study

The CNN$_\alpha$ is further used for the learning dependency study on the energy distribution of the events. An overview is listed in table 5.2. The dataset containing γ-rays and H^+ events are created using the standard cuts, see table 4.1. In the table the performances of the best and worst classifiers on a certain test set with spectral index Γ are confronted. The best classifier per test set is marked bold. The classifier trained on an energy distribution with a spectral index $\Gamma = -1.5$ vs. $\Gamma = -2.0$ is the best for datasets containing the same energy distribution for both classes. In comparison, the classifier trained on energy distributions with the spectral indexes of $\Gamma = -2.0$ vs. $\Gamma = -2.0$ is the best classifier on the mixed spectral index test sets. Interesting is the fact that this classifier is worse on its own test set where it has an accuracy of 98.14 %. One possible reason could be that this test set has more complex features which the training set does not contain and therefore are not shown to the NN during training phase. Using more train examples would resolve this assumption. The classifier trained on an arbitrary energy distribution is worse in its performance compared to the predefined energy distributions on all used test sets. To be fair it needs to be mentioned that these arbitrary distributions are similar to those with a spectral index of $\Gamma = -1$ and $\Gamma = 0$.

The reason is the energy distribution of the simulated MC events of these two classes. The distributions are shown in figure 5.6.
In conclusion they have all similar results but the classifier having an energy distribution with a spectral index of $\Gamma = -2$ for both classes is the best. It can be seen that the classifier having an nearly arbitrary energy distribution does have the energy as feature due to its worse result for all used test sets.

Table 5.2. – *Results of arbitrary and predefined energy distributions.*

Test Set Γ	**Γ vs. Γ**	**Accuracy**
arbit. vs. arbit.	arbit. vs. arbit.	97.81 %
	−2.0 vs. −2.0	97.96 %
−1.5 vs. −2.0	arbit. vs. arbit	98.14 %
	−2.0 vs. −2.0	98.33 %
−2.0 vs. −2.0	arbit. vs. arbit	97.98 %
	−1.5 vs. −2.0	98.26 %
−1.5 vs. −1.5	arbit. vs. arbit	98.01 %
	−1.5 vs. **−2.0**	98.07 %
−2.0 vs. −1.5	arbit. vs. arbit	98.01 %
	−2.0 vs. −2.0	98.18 %

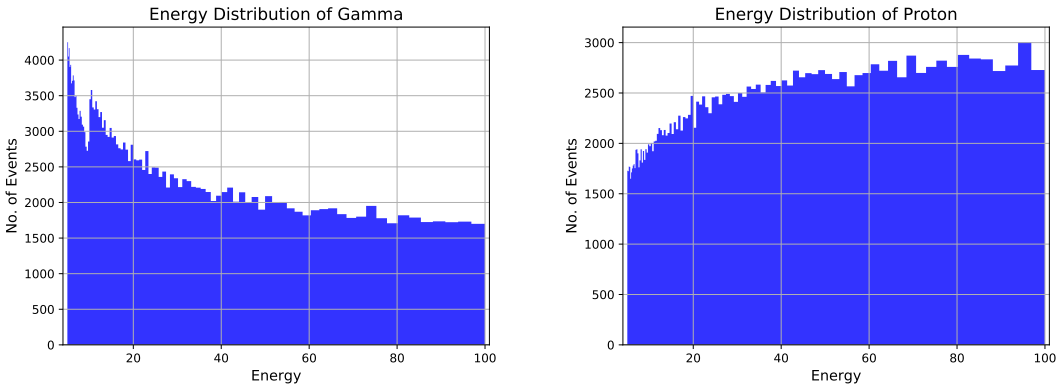


Figure 5.6. – *Distributions of the classes Gamma and Proton for a non predefined spectrum. The MC event were created with different minimum energies and spectral indexes and where saved in one folder per class. During the dataset creation these events are collected randomly.*

The CNN_α and the best performance of the classifier trained on an energy distribution with a spectral index $\Gamma = -2$ were used to investigate the binary classification task of heavier nuclei.

Metric Results of all Binary CNN_α Classifiers

The "best" architecture as well as the predefined energy distribution with a spectral index $\Gamma = -2$ were used to build or train the classifier. The datasets contain in total 300000 events and 150000 of each class. In table 5.3 the metric results are shown. The classifiers are sorted by the charge Z of the primary particle of the signal class. It can be seen that the classifiers performance increases if the mass difference of the signal and background classes also increases. The binary distinction task of H^+ against heavier nuclei and γ-ray events achieves the highest accuracy compared to the binary classification task of the heavier nuclei. The lowest accuracy is reached by the classification of C^+ vs. Si^+ and Si^+ vs. Fe^+ events. Conspicuous is the accuracy of the classifier C^+ vs. Fe^+. It has an accuracy of 76.60 % which is higher than the ones of the C^+ vs. Si^+ and Si^+ vs. Fe^+ events.
In order to investigate the phenomenon of the increasing mass difference to the improved performance of the classifier a plot was created in order to try to visualize the correlation. The figure 5.7 shows the recall dependency of the binary classification to the mass difference. The points color and label mark the signal vs. background class. An increase in the recall can be seen for the signal classes Proton, Helium and Iron for increasing mass difference. The overall best distinction is achieved by the classification of H^+ and γ-ray events. An unique course show the classifiers with the signal classes Carbon and Silicon. The recall of Carbon vs. background class decreases at first with increasing mass difference and at a mass difference of 44 u the recall has a value of 79.34 %. This means that the shower images of C^+ vs. He^+ and C^+ vs. Si^+ have similar features especially at lower energies, see the a), b), d), e) figures of 5.8. There the classifier predicts more FP compared to higher energies where the classifiers TP increase. The classification task of Si^+ and Fe^+ events scored the worst results compared to all other ones. The mass difference of Fe^+ and Si^+ regarding to the signal class Iron is the lowest one. All other background classes have a higher mass difference. Considering the TP energy distribution of this classifier for Fe^+ and Si^+ events in the figures 5.8c and 5.8f it can be seen that the TP of the class Silicon is diminished in the low energy range and for higher energies the TP are near to the true distribution whereas the TP distribution of class Iron is smaller as the true one especially in the energy range of 10 TeV to 45 TeV. It can be concluded, that the background classes which have a higher charge Z compared to the signal class look different in the low energy region.

Table 5.3. – *Results of all binary classifiers.*

Class vs. Class	Accuracy	Recall	Specificity
γ vs. H^+	98.14 %	97.71 %	98.58 %
H^+ vs. Fe^+	95.29 %	96.87 %	93.81 %
H^+ vs. Si^+	92.52 %	94.57 %	90.66 %
He^+ vs. Fe^+	88.96 %	92.14 %	86.22 %
H^+ vs. C^+	86.67 %	89.48 %	84.23 %
He^+ vs. Si^+	83.49 %	87.17 %	80.47 %
C^+ vs. Fe^+	76.60 %	79.34 %	74.32 %
He^+ vs. C^+	71.58 %	73.81 %	69.74 %
H^+ vs. He^+	71.11 %	74.49 %	68.55 %
C^+ vs. Si^+	66.87 %	68.53 %	65.47 %
Si^+ vs. Fe^+	62.88 %	63.78 %	62.09 %

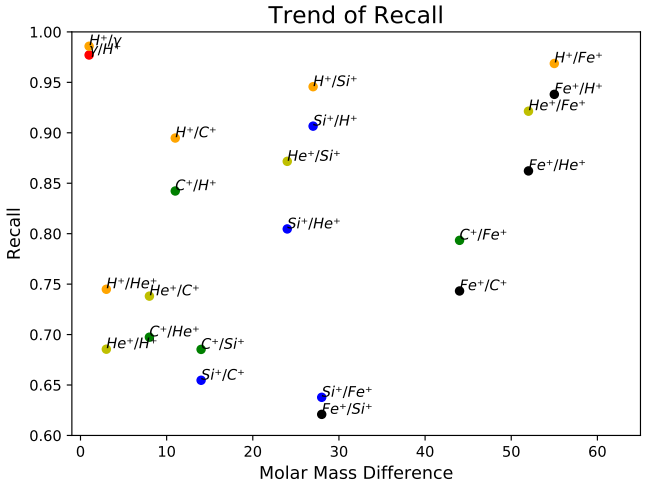
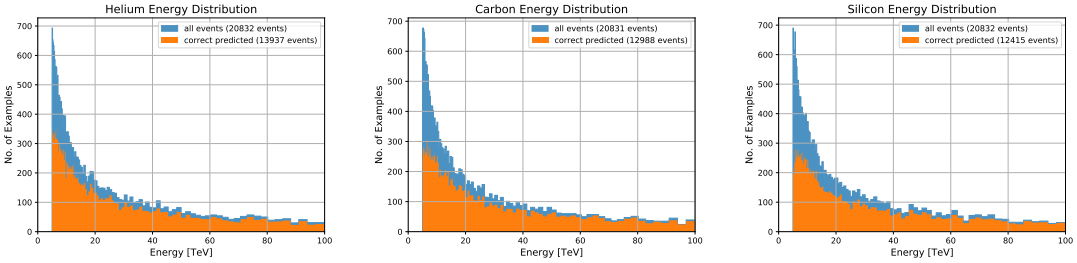
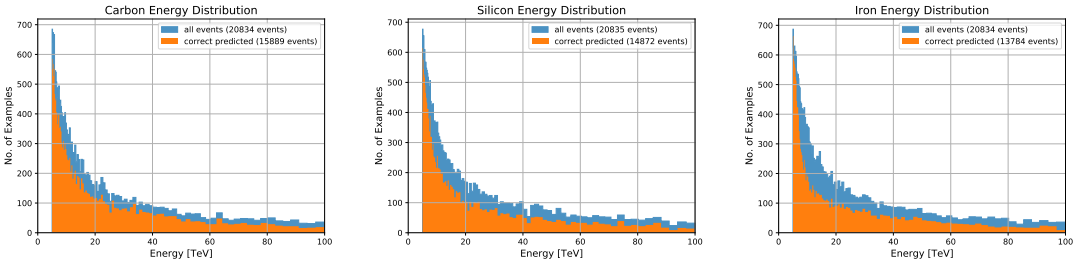


Figure 5.7. – *The graphic shows the dependency of the mass difference to the recall of the signal class. The points label mark the signal versus background class recall result for the corresponding mass difference. The orange points (signal class Proton), yellow points (signal class Helium) and black points (signal class Iron) show an increase in the recall value. Whereas the green points (signal class Carbon) have their highest value for the classifier trained on Carbon and Proton and their lowest recall for the classifier C^+/Si^+. The course of the green points shows an decrease for increasing mass difference but the one C^+ vs. Fe^+. The classifiers marked by the blue points (signal class Silicon) shows a similar behavior like the classifiers represented by the green points.*



(a) He^+ *vs.* C^+ **(b)** C^+ *vs.* Si^+ **(c)** Si^+ *vs.* Fe^+



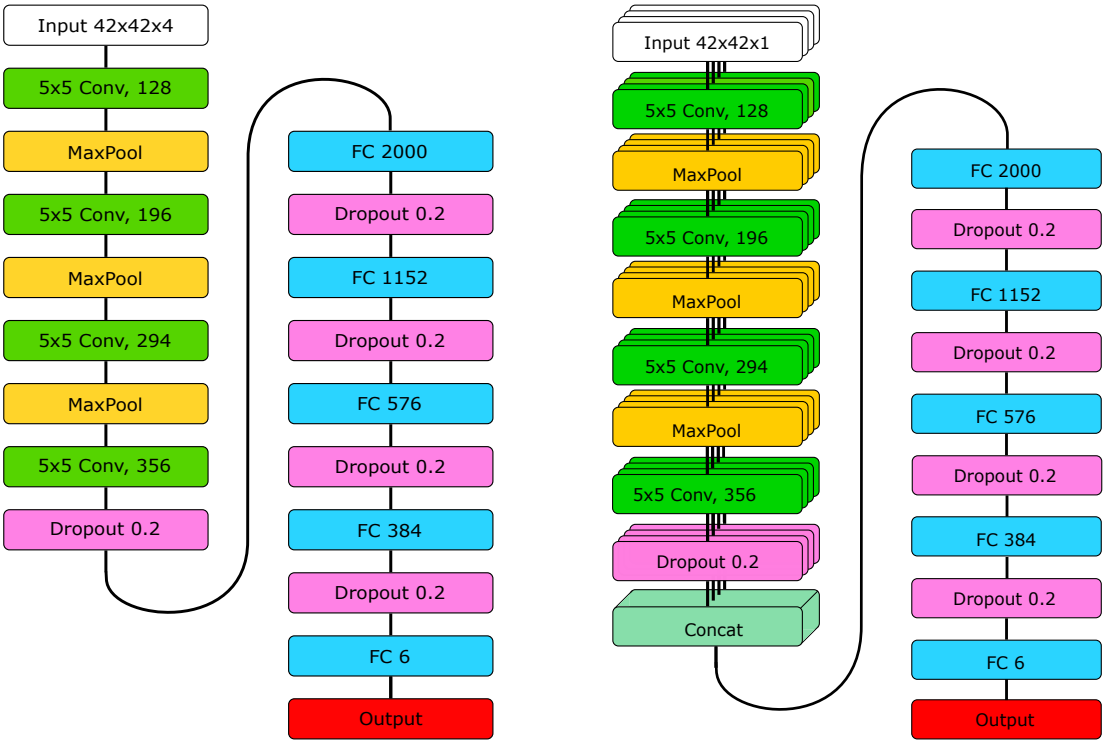
(d) He^+ *vs.* Ck **(e)** C^+ *vs.* Si^+ **(f)** Si^+ *vs.* Fe^+

Figure 5.8. – *Illustration of the energy distribution of the classifiers* C^+ *vs.* He^+, C^+ *vs.* Si^+ *and* Si^+ *vs.* Fe^+. *The upper distributions belong to the ones below. Conspicuous is the energy distribution of the signal and background classes. The distributions of a), b) and c) look similar and also d), e) and f). The classifiers show a diminished TP rate at lower energies and an increased at higher energies for each classes.*

In conclusion the best performance was achieved using the architecture with 4 convolutional and 5 fully connected layers but the performances of the other tested architectures were similar. Further the energy distribution study ensued that it is reasonable to train the classifier on a examples having an energy distribution with a spectral index of $\Gamma = -2$. The binary classification task of the classes Proton, Helium and Iron depends on the mass difference due to the increasing recall value with increasing mass difference. The classifiers C^+ vs. He^+, C^+ vs. Si^+ and Si^+ vs. Fe^+ show an unique behavior. These results are used to investigate the multi- and multi-task classification of all six classes at once.

5.2.2. Multi- and Multi-Task Classification

For the task of distinguishing all classes by one classifier the architecture investigated for the binary classification was used. In addition, a multi-task network which focuses on both differences and similarities between and within the classes was tested to improve the performance of the multi-classification task. The figure 5.9 shows the CNN_α and CNN of the 5 used architectures. Their base architecture was investigated in section 5.2.1. The others can be found in the appendix B.1.
For the training of these classifier in total 96000 train examples were used. They are equally divided which means 32000 events of each class. The energy distribution have a spectral index of $\Gamma = -2$. The optimizers Adam and Nadam were used for the training of the multi- and multi-task classifier, respectively. The NN architectures are similar in their structure. All have 4 convolutional layers and 5 fully connected layers. Between each convolutional layer a maximum pooling layer is inserted with stride 2 and padding same. This means the feature maps shrink in their size by half after each pooling layer and keeps only the features with the highest values within a certain kernel size. Stride 2 means that the scanning of the feature map in the pooling layer considers every second pixel. A dropout of 0.2 is chosen to prevent over-fitting. In subfigure 5.9a the architecture using the channel approach is visualized. The input consists of all CT1-4 images having a resolution of $42 \times 42\,\text{px}^2$ after the rebinning preprocessing step. The subfigure 5.9b shows the single image approach. The 4 towers consist of 4 convolutional layers and in between maximum pooling layers. In addition the towers share theirs weights. After the towers the feature maps are concatenated and provided to the fully connected layers consisting of 2000 nodes. The output contains the predicted class label.
The evaluation of the chosen "best" classifier CNN_α is done at the epoch where the validation loss is minimal, see figure 5.10. The determined minimum of the validation loss is at epoch 6. At this epoch the classifier was proven on the test set. The resulting loss is 0.998. Interesting is the fact that the validation loss is higher than the test set loss at this point. This phenomenon could already be seen in figure 5.3 showing the binary classification loss. Reasons can be that classifiers learned features which are more similar to extracted one in the test sets than to the ones in the validation set. The differences in the loss value are relatively small therefore numerical uncertainties can also lead to a slightly smaller test set loss. After this epoch the validation loss increases steadily. In contrast, the training set loss decreases continuously. It seems that the classifier is over-fitting after epoch 6. The loss values are relatively high in comparison to the normalized loss in the binary classification 5.3. In other words the loss of CNN_α is about 16 times higher than the loss of CNN_α used in the binary classification.



(a) CNN_{α} **(b)** CNN

Figure 5.9. – *Illustration of multi-classification networks having the same structure as the ones used for the binary task expect the last fully connected layer which has* 2 *instead of* 6 *nodes. The convolutional layers are marked with the color green. They are initialized with a kernel size* 5×5 *having* 128 *nodes. The maximum pooling layers have stride* 2 *and use padding same. After the convolutional layers a dropout of 0.2 is inserted and after each fully connected layer to prevent over-fitting. The dropout contains probabilities for each class. a) The channel approach convolves over the stack of CT1-4 images treated as one RBG*α *image in the first layer. b) Single CT images are provided one after another to the network sharing their weights. They get concatenated after the convolution layers and first dropout.*

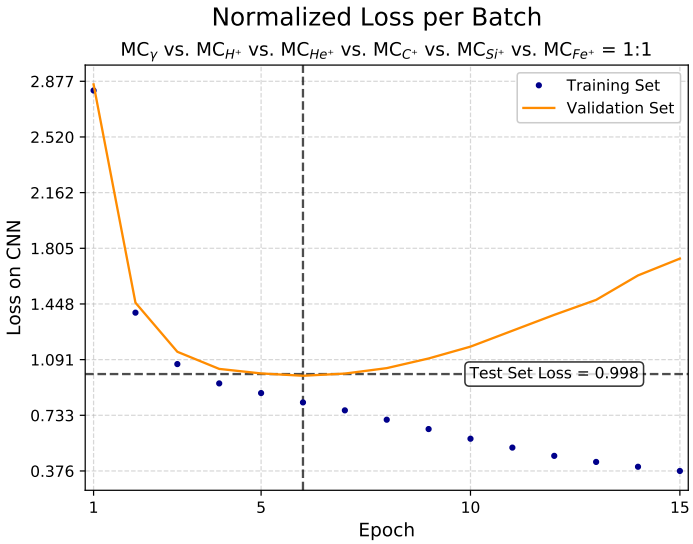


Figure 5.10. – *Figure of the normalized loss of the multi classifier CNN$_\alpha$. The classifier was trained on MC simulations, see title. The trainings set loss decreases steadily. In comparison, the validation set loss decreases until epoch* 6 *where it has its minimum. After epoch* 6 *the validation loss increases. The CNN$_\alpha$ is over-fitting after this epoch which can be seen by the increasing loss of the validation set and further decreasing of the training set loss. At epoch* 6 *the classifier was tested on the test dataset. The resulting loss is* 0.998*. However, the loss of the validation set is higher and the one of the training set is smaller.*

The performance result of the best classifier on class Iron is shown in table 5.4 and the other classifier results can be found in the appendix B.3. One result they all have in common is the good differentiation of γ-ray EAS events against all other classes the classifiers were trained on. The classifier having the highest accuracy is the CNN$_{SK}$ using the addition parameters skewness and kurtosis. Due to time limitations the CNN$_{SK}$ was not chosen for further analysis steps but CNN$_\alpha$. Both performance results are comparable therefore the reconstructed iron energy spectrum would be similar, too.
The focus in the performance of the CNN$_\alpha$ was to suppress the background events in each class and increase the TP number which is shown by the F$_1$-Score. The recall of class Iron is 47.97 % which means that 52.03 % of true class Iron were predicted as FN. The class which lost the most events after the classification is Carbon. On Carbon the CNN$_\alpha$ has a TP rate of 39.11 %. Overall the classifier does not guess any of the classes it was trained on. In the case of guessing the recall would have a value smaller than 16.67 %. The classifier has on its worst recognized class more than the double of guessing value.

Table 5.4. – *Performance results of CNN_α reaching a total accuracy of 56.68 %.*

Class	**Precision**	**Recall**	**F_1-Score**
γ	96.04 %	97.25 %	96.65 %
H^+	73.71 %	66.25 %	69.78 %
He^+	48.21 %	42.44 %	45.14 %
C^+	36.21 %	39.11 %	37.60 %
Si^+	37.03 %	47.07 %	41.45 %
Fe^+	55.99 %	47.97 %	51.67 %

Further the recall of the validation set was considered to understand the learning of the CNN_α. In figure 5.11 the recall of the validation set is shown. It can bee seen that the recall of class Gamma stays nearly constant over the hole learning phase. The course of the accuracy (green solid line) is similar. The recall of the classes Helium and Carbon seem to be correlated due to their altering recall values. The same phenomenon can be seen by the recall values of Silicon and Iron.
In comparison to the validation set recall of the binary classifier the multi-classifier seems to forget features corresponding to the classes Silicon and Iron in order to learn the other classes. The binary classifier learns both classes until epoch 6 where it has its minimal validation loss value. In addition, the binary classifier has at this epoch its maximal accuracy value. Considering the accuracy of the multi-classifier it seems that at epoch 5 and 7 the classifier has a higher accuracy than at epoch 6 where the validation loss is minimal.

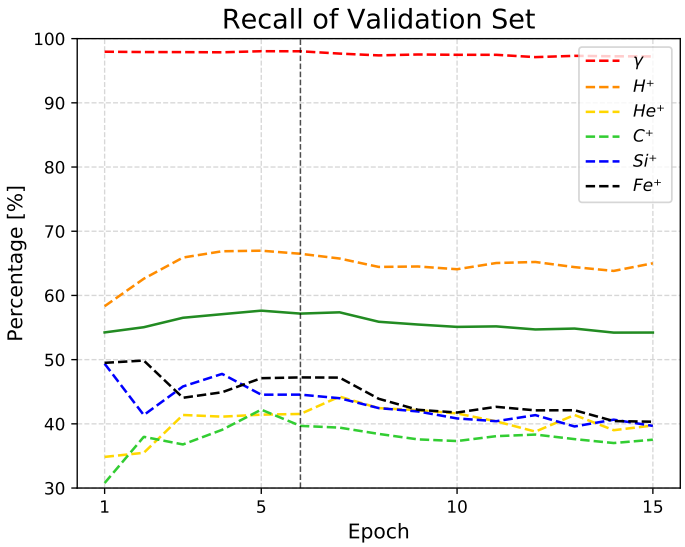


Figure 5.11. – *Illustration of the validation set recall of CNN$_\alpha$. The overall accuracy is drawn in as green solid line and is nearly constant. The classifier was evaluated at epoch 6 marked with the thin, black, dashed line. The recall of each class is drawn as colored, dashed line. The recall value of class Gamma is stays almost unchanged. The classifier seems to learn some features of the classes Proton, Helium and Carbon but at the same time forgetting some features of the classes Silicon and Iron in the first* 6 *epochs. After this epoch the recall of all classes decreases with some fluctuations.*

The result of the CNN$_\alpha$ can be illustrated in one confusion matrix. The confusion matrix shown in figure 5.12 gives an overview of the number of events assigned to their true and predicted labels. The column contains all events to the corresponding true label and the rows the events predicted as a specific class. The classifier predicts the most true label Gamma events as class Gamma which agrees with the result of table 5.4. The H^+ induces EAS events have the second highest true positive ratio. The most FN of class Proton were classified as Helium events and only 0.63 % were misclassified as Iron. The number of FN of the classes Helium, Carbon, Silicon and Iron is higher than the number of FN of class Gamma and Proton. In comparison the number of FP decreases with increasing charge difference ΔZ. The FN events of class Iron were mostly predicted as class Silicon. This result is in agreement with the one in the table 5.3 in the binary classification. The binary and multi-classifier can hardly distinguish between classes which have lower charge difference ΔZ relative to the considered signal class.

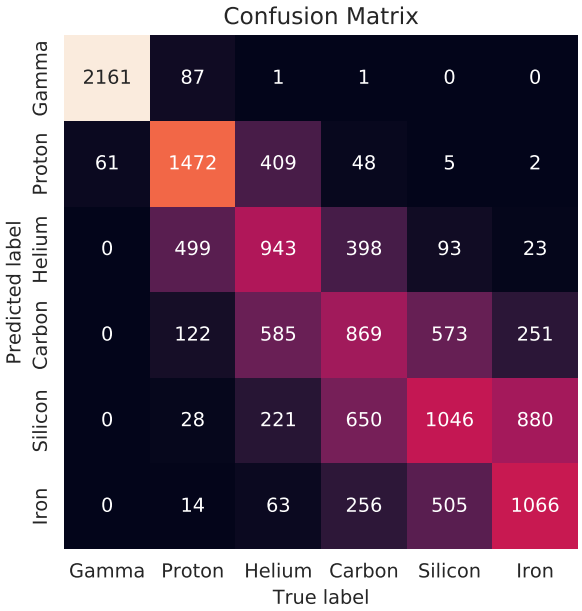


Figure 5.12. – *Confusion matrix of the CNN$_\alpha$. The columns contain the numbers of the events corresponding to the true class. Whereas the rows symbolize all events predicted as a certain class. The classifier dedicates the most γ-ray events correctly to class Gamma. The uncertainty or the ratio of FP of CNN$_\alpha$ decreases with increasing charge Z considering only the heavier nuclei (H^+ and all heavier elements). In the rows, the neighbor classes having the smallest charge difference ΔZ (above and below) to the corresponding true class got the most FP labels. In comparison, the FP ratio of class Proton is much smaller. For the classes Helium, Carbon and Silicon the number of FP exceeds the number of TPs. Despite the increasing number of FP per row the diagonal entries containing the TPs of each class are the highest numbers.*

The row of the confusion matrix containing events predicted as class Iron is further considered. It is also directly related to the precision of the classifier. In figure 5.13 the number of events per true class are plotted against the true energy of the events. The energy distribution of the TP of class Iron does not have the original spectral index of -2. The classifier correctly predicts Iron events in the lower energy bins but the number of misclassified events of other classes predicted as Iron increases, too. In the higher energy region some bins do not contain any events. Events from the classes Carbon and Silicon are classified as Iron until a true energy of 83 TeV. No events of class Gamma were predicted as Iron.

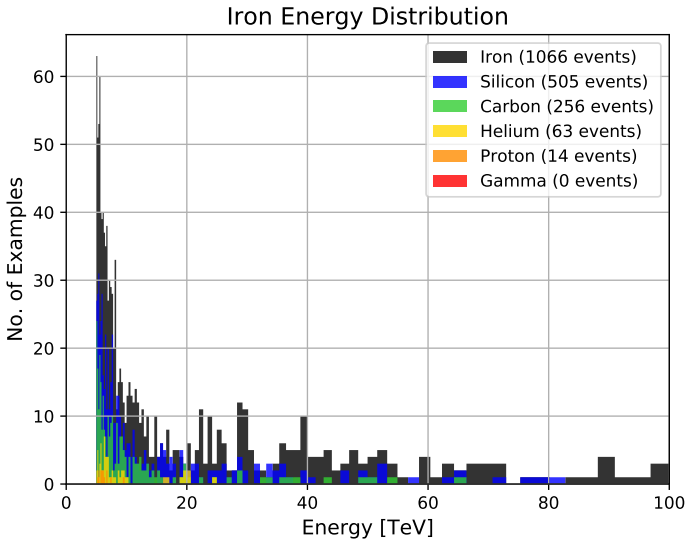


Figure 5.13. – *Energy distribution of all as class Iron labeled events. The TP of Iron dominate the histogram. In the distribution of TP some bins do not contain events. The distribution of events misclassified as Iron predominate in the low energy region. In the higher energy bins only Silicon and Carbon events were predicted wrongly.*

The classification of the six classes Gamma, Proton, Helium, Carbon, Silicon and Iron did work but can be improved. None of the tested classifier did reach an accuracy higher than 60 %. To further improve the performance of the classifier an additional cut parameter was investigated.

5.2.3. DC Cut Parameter

Since DC-light only occurs for hadronic particles the idea was to use this feature as an additional cut parameter to improve the distinction of hadrons especially iron nuclei. Therefore the DC-light ratio of 100000 events of each class was calculated. The result can be seen in 5.14. The bins in the shown histogram have a width of 4 p.e.. The iron nuclei events are present in each bin but the last one. Also the gradient of the number of events for this class and class silicon is relatively small in contrast to the gradient of the classes containing carbon or helium nuclei events.
One possible cut could be set at a ratio of 39 p.e.. At this ratio only carbon and iron nuclei events pass the cut. The dataset would then contain in total 125 events whereby 25 are carbon nuclei events and 100 are iron nuclei events. Hence, no reasonable cut can be set without loosing too much of statistic. This can also be seen by considering the distributions in which no obvious feature can be seen to be used as cut parameter. In the aspect of creating an iron nuclei spectrum the probable DC-light cut of 39 on real data would lead to a 80 % recall additionally no classifier would be needed.

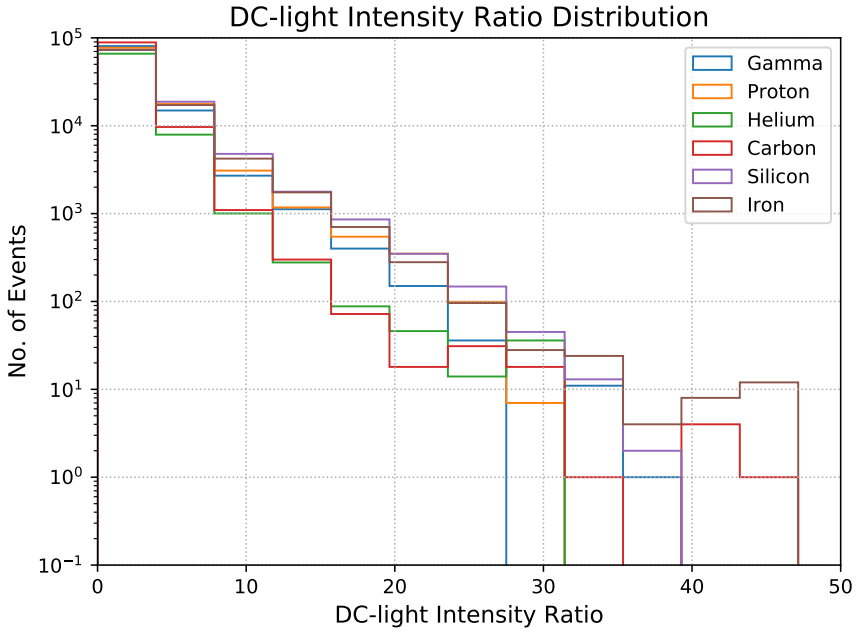


Figure 5.14. – *DC-light distribution of* 100000 *events per class. The number of events is given by the DC-light ratio. The number of events decreases linearly with increasing intensity ratio. Between the ratios 39 p.e. and 47 p.e. only events of the classes Carbon and Iron are left. The number of events of carbon nuclei events decreases in the DC-light ratio faster compared to the other classes. However the heaviest nuclei iron and the silicon nuclei have a smaller gradient in contrast to the other classes.*

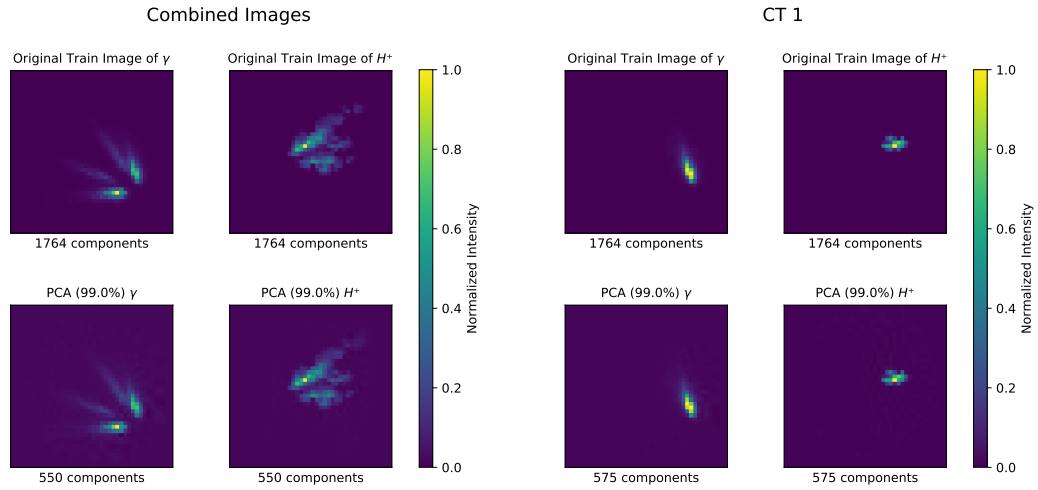
5.3. Application of PCA

An optional preprocessing step PCA was implemented and tested to find correlation between the classes and therefore reduce the number of dimensions. PCA was applied using variances 90 %, 95 %, 99 %. Already rebinned CT1-4 images of the training dataset with a resolution of $42 \times 42\,\mathrm{px}^2$ are given to the algorithm. PCA must not be applied on the validation and test dataset.

5.3.1. γ vs. Proton

PCA was firstly applied on a dataset containing γ-rays and H^+ MC events. The dimension reduction method was tested despite the evidently good separation of these two classes by the Hillas analysis and ML approach in order to improve the distinction by the classifier.
The training dataset contains 128000 MC events in total, 64000 of each class. The energy distribution of each has a spectral index of $\Gamma = -2$. The outcome of the dimension

reduction algorithm is shown in figure 5.15. The original images are in the top row, whereas the results can be seen in the bottom row. An example of a γ-ray induced EAS is shown in both images on the left side. It has the typical elliptical shape. From such images the impact point can be determined and therefore the direction of the EAS. The H^+ nuclei events are shown in the right images in both figures. Their geometrical form cannot be approximated with an ellipse. Comparing the original with the dimension reduced images the pixels not counted among the EAS ones have a value greater than 0. Whereas in the original images those pixels were set to 0 during the Image cleaning preprocessing step.



(a) *PCA(99 %) on combined images.* **(b)** *PCA(99 %) on CT1 images.*

Figure 5.15. – *Comparison of the original combined and CT1 images to the dimension reduced images of both classes. The variance for this example was set to 99%. In the direct comparison the dimension reduced images are smoother and contain more noise. The number of principal components in the original images is* 1764 *which is equal to the number of pixels of the resolution* $42 \times 42\,\mathrm{px}^2$*. a) The images in the second row show the dimension reduced images. The EAS images have after the PCA* 550 *principal components. b) In contrast the dimension reduced EAS of one CT have* 575 *principal components.*

The number of principal components depends on the chosen variance. In figure 5.16 the behavior of the correlation between these and the variance is shown. The number of principal components depends on the chosen explained variance and indirectly on the resolution of the CT1-4 images. The more pixels an image consists of the more features can be extracted.
The course of the graph in figure 5.16 shows an exponential decrease in the total number of principal components for decreasing explained variance. The original number princi-

pal components 1764 correspond to the total number of pixels in the rebinned image. Using an explained variance of 99 % reduces this number by 1215 principal components. Choosing variances below 90% is not useful since the image then contains only few visible features and more noise, see 5.17. The figure shows an example from a dataset which was reduced to 80 % of the origin principal components. The dimension reduced CT1-4 images are blurry and contain additional features. These look like Moiré patterns or alising effects. The cause of these artifacts was not further investigated.

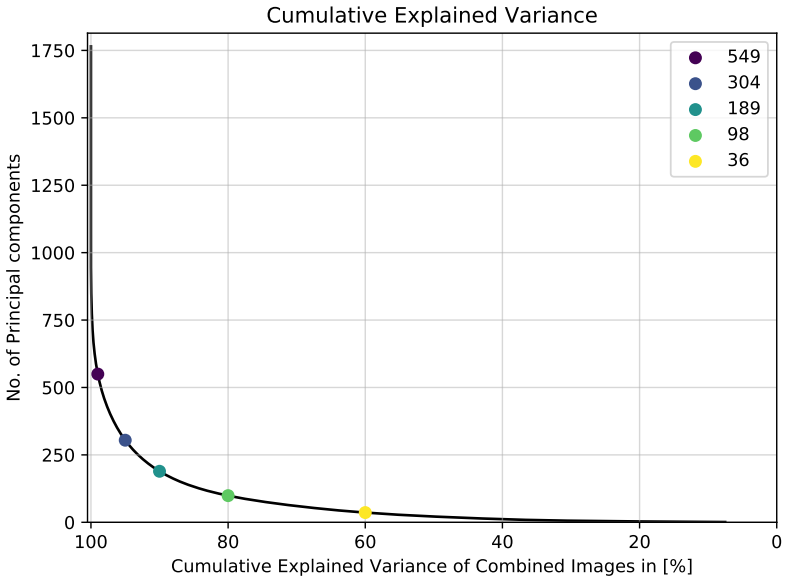


Figure 5.16. – *The number of principal components decreases exponentially with decreasing cumulative explained variance. The variances of 60 % up to 99 % are marked as color points. A cumulative explained variance of e.g. 80 % has in total 98 principal components.*

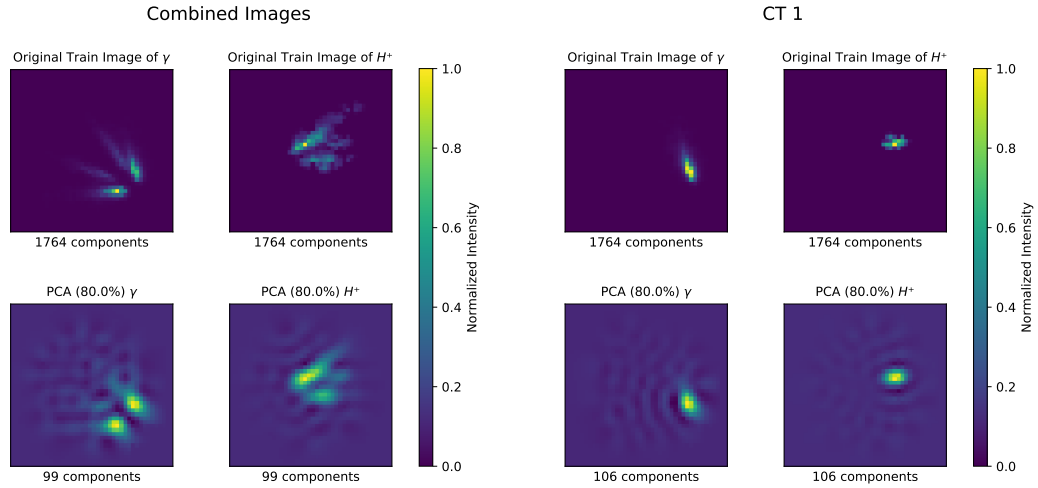


Figure 5.17. – *The images show an example of PCA with 80 % explained variance. The resulting dataset contains additional noise and the shape of the original shower is modified but the orientation of the shower is preserved.*

The training results on the dimension reduced datasets for variances of 90 %, 95 %, 99 % are shown in table 5.5. In comparison to the dataset with 100 % variance the accuracy is decreased. Therefore it is not useful to apply PCA on a dataset containing γ-ray and H^+ events. The recall represent the ratio of TP events classified as Gamma. The decrease in the recall after PCA was applied means that the most principal components with an explained variance of $> 99\,\%$ correspond to the features describing the class Gamma. This can also be seen in the figure 5.15. There the reconstructed image of class Gamma is more blurry compared to the reconstructed image of class Proton. In addition, in the dimension reduced, combined image of class Proton in figure 5.15 the shower part in the upper right corner is also blurred. Considering the corresponding image in CT1 that part is not blurred. Hence, the blurred part in the combined image containing 99 % variance was a part of the 1 % describing a γ-ray EAS feature. Conspicuous is the increase in the TP rate of the class Proton for the explained variances 99 % and 95 %. Using an explained variance of 90 % or 80 % reduces significantly the specificity from 99.09 % to 92.76 % and 78.01 %, respectively. Therefore it can be concluded that several features of class Proton have an explained variance greater than 80 % due to a lower decrease in the recall for the same explained variance.

Table 5.5. – *Results of PCA applied on a γ-ray and H^+ dataset.*

Variance	Accuracy	Precision	Recall	Specificity
100 %	97.72 %	98.31 %	97.15 %	98.29 %
99 %	97.02 %	99.46 %	94.83 %	99.43 %
95 %	96.33 %	99.14 %	93.85 %	99.09 %
90 %	93.09 %	92.70 %	93.43 %	92.76 %
80 %	83.45 %	73.73 %	91.51 %	78.01 %

The application of PCA on the γ-ray and H^+ EAS events dataset did not improve the performance. PCA is investigated to test if an improvement of the accuracy using the same NN but an other dataset containing Si^+ and Fe^+ events is possible. Considering the corresponding result in 5.3 the classifier has not a high accuracy compared to dataset containing γ-ray and H^+ EAS events.

5.3.2. Silicon vs. Iron

The second dataset contains silicon and iron events. This dataset is chosen to investigate if PCA can support the distinction of this classifier.
Subsequently, the model was trained using the dimensionally reduced shower events and compare it with the result of the original dataset. The results are shown in table 5.6. Example of the original images compared to the dimension reduced ones can be seen in figure 5.18. The figure shows EAS simulated shower of Si^+ and Fe^+. The irregular shape and the morphology are typical for these initial particle types. The difference of the original and the dimension reduced images are optically relatively small due to the chosen explained variance of 99 %. Considering the performance results the NN trained on CT1-4 images with 99 % and 95 % of the original variance are better than the classifier trained on the only rebinned images. Therefore it can be concluded that the principal components having a high variance are feature of both classes which strongly correlate. This assumption is supported by the increasing recall of class Silicon for both expalined variances and a increasing specificity in the case of keeping the principal components which have a variance smaller than 95 %. A variance of 90 % does not improve the performance of the classifier.
Summarizing the result of the application of PCA on the dataset containing MC Si^+ and Fe^+ nuclei events using a variance of 95 % does improve the accuracy and also the precision of the classifier. Choosing a variance smaller than 95 % is not useful.

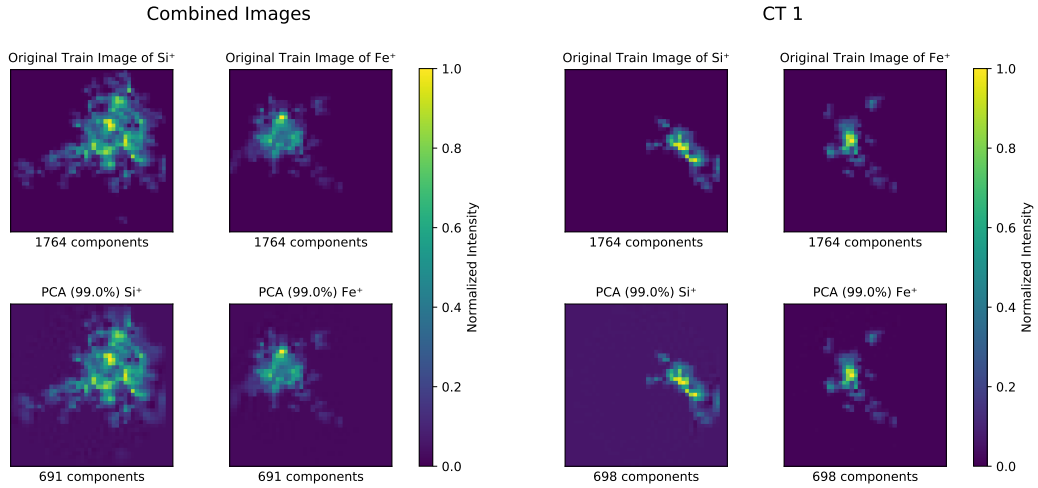


Figure 5.18. – *The images show the origin images and the rebuilt images after PCA was applied with a cumulative explained variance of 99 %. The left image shows combined images of silicon and iron. The dimension reduction applied on combined images results in fewer principal components compared to applying PCA separated on CT1-4. It can be seen in the right subfigure with in total 698 principal components.*

Table 5.6. – *Results of PCA applied on a Si^+ and Fe^+ dataset.*

Variance	Accuracy	Precision	Recall	Specificity
100 %	61.29 %	63.18 %	60.88 %	61.74 %
99 %	61.87 %	59.73 %	62.40 %	61.38 %
95 %	61.74 %	64.27 %	61.17 %	62.38 %
90 %	60.98 %	61.36 %	60.90 %	61.06 %

The optional preprocessing step, PCA, did not achieve any improvement in the accuracy of classifier trained on γ-ray and H^+ events. In the case of Si^+ and Fe^+ events the accuracy was higher using a variance of 99 % or 95 % compared to the classifier trained on the original images. Since PCA did not advance the classifier for the dataset containing γ-rays and H^+ and only insufficient for Si^+ and Fe^+, it was not further applied on any dataset for any task.

5.4. Energy Reconstruction

The energy reconstruction of CR is an common part of the analysis. To achieve the best result a study of the choice of the spectral index and the scaling factor of the response and the true energy was done. In addition, the energy reconstruction for all six considered particle types is done by training the energy estimators on each class separately. The architecture for this task consists of 9 convolutional, 2 pooling 1 residual and 8 fully connected layers. It is illustrated in figure 5.19. For this architecture the channel approach was chosen. The first convolutional layer having 96 nodes with a kernel size of 5×5 convolves over the stack of CT1-4 images. Followed by three more convolutional layers. The outcome of the first maximum pooling layer is transfered to the fifth convolutional layer and copied for later usage. The feature maps are further convolved until the residual layer. There the copy and the feature map of the eighth convolutional layer are combined. This means further information from the result of the first maximum pooling is applied to the current feature map. Afterwards the outcome is provided to the last maximum pooling layer followed by the last convolutional layer. Then the feature map is transfered to a one dimensional feature vector and a dropout of 0.5 is applied. The tower consisting of fully connected layers with dropouts of 0.5 in between. It is divided into two parts. The first one contains layers where the number of nodes decreases until 16. At this so called bottle neck the size of the CT1-4 is concatenated to the feature vector. The second part consists only of fully connected layers with firstly increasing and the last one has a decreased number of nodes. The output is the predicted energy for the corresponding event.

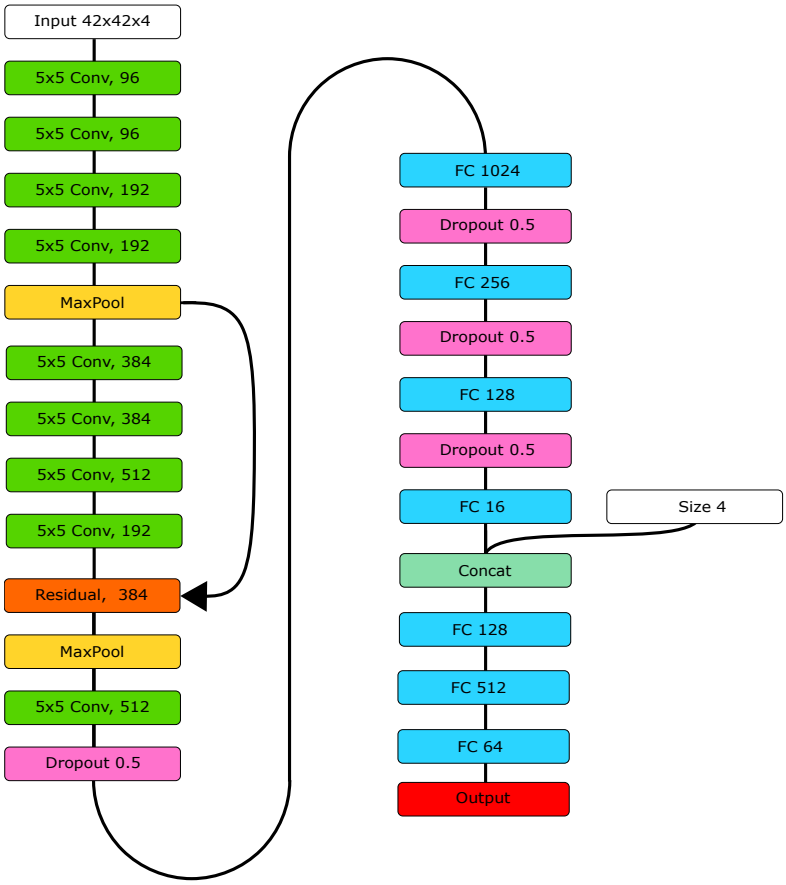
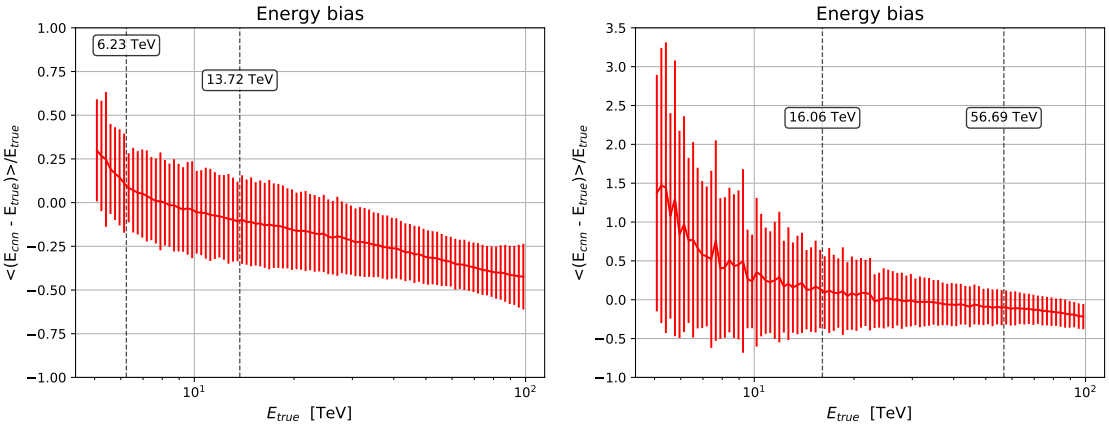


Figure 5.19. – *Illustration of energy estimator architecture consisting of* 9 *convolutional layers and* 8 *fully connected layers. The convolutional layers are marked with the color green. The channel approach convolves over the stack of CT1-4 images treated as one RBGα image in the first layer. They are initialized with a kernel size* 5×5 *having* 96 *nodes. The maximum pooling layers have stride* 2 *and use padding same. After the convolutional layers a dropout of 0.5 is inserted and after each fully connected layer to prevent over-fitting. A connection from the first maximum pooling to the residual layer to demonstrate the residual functionality was created. In addition, at the bottle neck the size was added to the feature vectors.*

5.4.1. Choice of Spectral Index

The energy estimator was trained and tested on a dataset containing only MC Fe^{+} events. During the training and testing the energy was logarithmically scaled.
The first training dataset was created with a spectral index Γ of -1 with 320000 training examples and the second dataset was created with $\Gamma = 0$ and the same number of events. The resulting energy bias of the test set for both spectral indexes is shown in figure 5.20. Both subfigures show the relative bias which was stored in bins increasing logarithmically in size, same as explained in 4.1.1.

The energy range within ±10 % *bias* for the dataset with an energy distribution created with $\Gamma = -1$ is $\Delta 7.49\,\mathrm{TeV}$. Whereas the dataset with $\Gamma = 0$ achieves an energy range of $\Delta 40.63\,\mathrm{TeV}$. A further feature of training the energy estimator on a $\Gamma = -1$ energy spectrum is the equally spreading of the relative dispersion. Thus the energy estimator can reconstruct the energy in each bin with the same uncertainty. The dataset depending on the flux normalization Φ_0 broadens the energy range in which the *bias* is ±10 % but the dispersion is increased.
The result for the task of the energy reconstruction is to use an equally distributed energy because it improves noticeably the energy estimator with the compromise of increased dispersion.



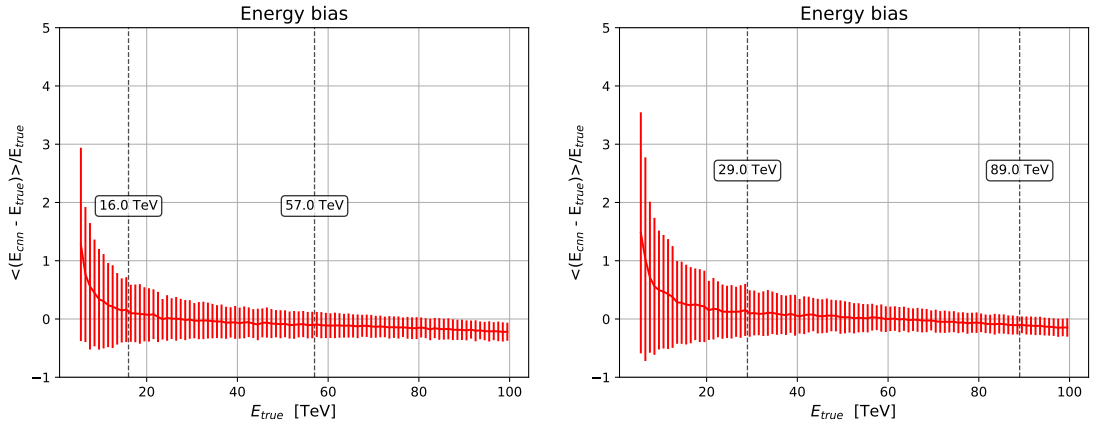
(a) $\Gamma = -1$ **(b)** $\Gamma = 0$

Figure 5.20. – *The figures show the relative bias for an energy distribution with a spectral index of* $\Gamma = -1$ *and* $\Gamma = 0$*. The dashed lines mark the energy range where the bias is* ±10 %*. a) The resolution of the relative energy bias is nearly constant over the whole energy range. Between the energies* $E = 50\,\mathrm{TeV}$ *and* $E = 70\,\mathrm{TeV}$ *the resolution of the bias is slightly better compared to the other energies beyond this range. In total the bias is* ±10 % *within the energy range 6.23 TeV and 13.72 TeV. b) The resolution of the relative energy bias and the energy bias itself is higher compared to a) over the whole considered energy region. In total the range within the relative energy bias is* ±10 % *begins at 16.06 TeV and ends at 56.69 TeV. The energy range is in b) broader then in a).*

5.4.2. Comparison Log- vs. Linear-scaling

The different variants of the input and response scaling in regression task was done to see if the results differ and if they do which scaling variant is better for this application. The training dataset created for the choice of the spectral index, containing 320000 iron nuclei events having an energy distribution with a spectral index $\Gamma = 0$, was reused for

the comparison of log- and linear-scaling.
The energy *bias* limit within ±10 % of the test set, to see in figure 5.21, is wider for the linear scaling. Using a logarithmically scaling (log-scaling) the *bias* interval results in a total energy range of Δ41.0 TeV. Whereas the energy range applying linear-scaling scores Δ60.0 TeV but is worse considering the dispersion at lower energies in comparison to the log-scaling.
In the analysis methods only the energies within the *bias* limit are used for calculations and the dispersion outside the *bias* region is not considered therefore a reasonable result is to use a linear-scaling.



(a) *Log-scale* **(b)** *Linear-scale*

Figure 5.21. – *The figures show the relative bias for the case of log-scaling and linear scaling of the true energy and the predicted energy. The dashed lines mark the energy range where the bias is* ±10 %. *a) The standard deviation of the relative energy bias in the lower energy regime is greater than in the higher energy region. b) The resolution of the energy estimator is between* $E = 5\,\text{TeV}$ *and* $E = 10\,\text{TeV}$ *in comparison to the mean in the range where the bias is* ±10 % *relatively high. The energy range is in b) greater then in a).*

5.4.3. All Initial Particle Energy Reconstruction

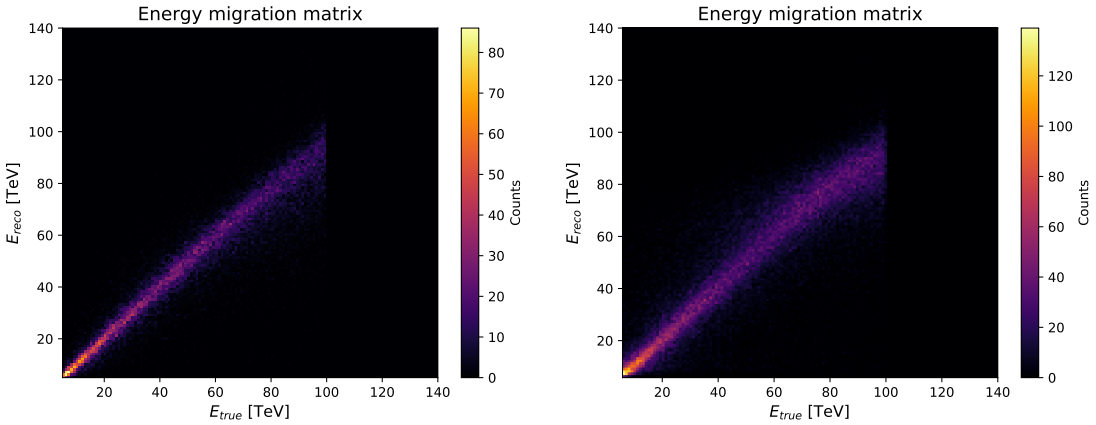
The energy reconstruction was done for all six classes. The result of the test sets is shown in table 5.7. The energy distribution of each class has a spectral index of $\Gamma = 0$. Due to the limitation in the total number of events passing the cut parameters, see table 4.1, the total training examples of the classes Gamma and Proton is 128000 and 243200, respectively. The energy estimators of the classes Helium, Carbon, Silicon and Iron were trained on 320000 events each. The widest energy range for a *bias* ±10 % is achieved for the class Gamma despite of only 128000 train examples. The total energy range is 87 TeV. In contrast, the energy reconstruction of the nuclei have a smaller *bias* ± 10 %

range. The smallest range of 52 TeV has the class Proton. Due to the goal of creating an iron energy spectrum the energy estimator of the class Iron is considered in more detail. The *bias* interval for the class Iron is 60 TeV. Therefore it is possible to apply further analysis steps in this energy region.

Table 5.7. – *Energy range in which bias is* $\pm 10\,\%$ *of the test set.*

Particle type	$E_{reco,min}$ [TeV]	$E_{reco,max}$ [TeV]	ΔE_{reco} [TeV]
γ	7.0	94.0	87.0
H^+	19.0	71.0	52.0
He^+	17.0	75.0	58.0
C^+	32.0	89.0	57.0
Si^+	24.0	85.0	61.0
Fe^+	29.0	89.0	60.0

The energy reconstruction of the class Gamma has the widest energy range in which the *bias* is $\pm 10\,\%$. The resolution and dispersion is relatively low compared to the energy reconstruction of Fe^+ events which can be seen in the figure 5.22. It shows the energy migration matrix of the classes Gamma and Iron. One noticeable feature of the energy migration matrix of Iron is that it is wider scattering at higher energies due to the choice of the logarithmically increasing bin width which results in lower statistics, see section 4.1.1. This phenomenon can also be seen in the migration matrix of class Gamma. In addition, the energy migration matrix of Gamma has less uncertainties in this region but more compared to the low energy region. There the energy estimator predicts the correct energy for the most of the events. The number of total counts differs by 192000 therefore the colorbar of class Gamma shows less maximal events. The range for the true and reconstructed energies were set to the lowest and highest true or predicted energy value of class Iron. Hence, it has to be mentioned that one event of class Gamma was predicted to have an energy of ~ 408.962 TeV but having a true energy of 94.04066 TeV. This behavior of the energy estimator is suspicious because it had only seen events having energies of 5 TeV to 100 TeV. It was only one event having a higher predicted energy than 140 TeV therefore it was neglected for the creation of this plot 5.22a. The migration matrix of class Iron has a broader diagonal line compared to the energy migration matrix of class Gamma. The dispersion increases for increasing energy and at an energy of $\sim$ 80 TeV the diagonal line has a salient point. Further the true and predicted energies of class Iron was considered further in energy resolved histograms.



(a) *Class Gamma* **(b)** *Class Iron*

Figure 5.22. – *The resulting energy migration matrices of the classes Gamma and Iron after the energy reconstruction using a DNN. The colorbar indicates the number of events per energy bin. The limits of the ordinate- and abscissa-axis were set to the minimum and maximum of either E_{reco} or E_{true} depending on which has the lower or higher energy in the boundaries, respectively. a) One event in the energy migration matrix of Gamma was reconstructed outside the shown energy range. The axes are adjusted to the ones in figure b). The dispersion is relatively small in comparison the the energy migration matrix of class Iron. b) In the low energy regime the dispersion is relatively low compared to the high energy region. The course of the distribution in the higher energy regions $E_{true} > 80\,\mathrm{TeV}$ and $E_{true} < 100\,\mathrm{TeV}$ is more scattered and has a salient point at $E_{true} = 80\,\mathrm{TeV}$.*

Comparing the energy migration of iron with the energy resolved histogram of the reconstructed to the true energy in figure 5.23, one can see that the energy estimator is less accurate in its prediction due to overshooting in the region $60 - 80\,\mathrm{TeV}$ and under-estimation of the energy in $80 - 100\,\mathrm{TeV}$. In addition, some events are predicted having an energy higher than 100 TeV although it was trained on events having energies within the range of 5-100 TeV. The histogram containing the predicted energies has a non-linear increase up to 80 TeV. In contrast the histogram showing the distribution of the true energy has a linear increase of the number of events due to the chosen binning and spectral index of $\Gamma = 0$.

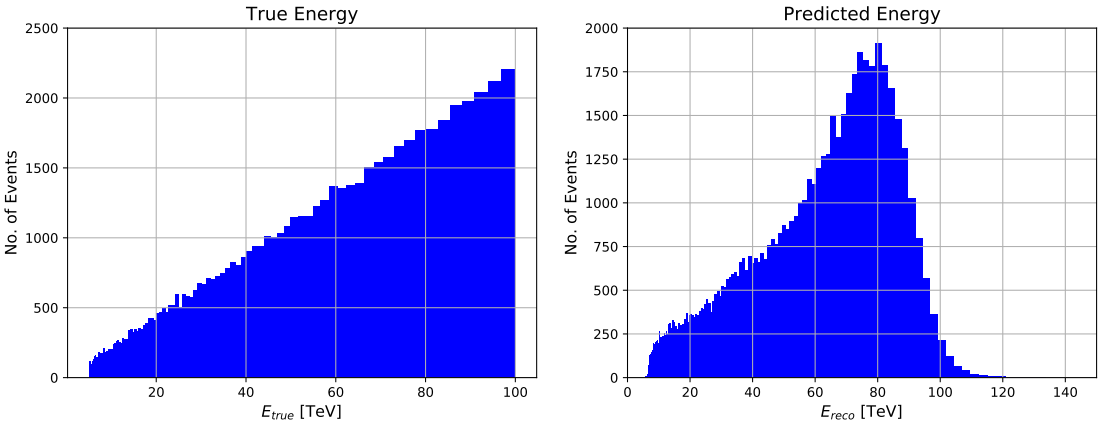


Figure 5.23. – *Comparison of original energy E_{true} and the predicted energy E_{reco} of the NN. a) shows the energy distribution of the MCs. Due to logarithmically increasing bin width the distribution for a spectral index of $\Gamma = 0$ has not its typical constant shape. b) The distribution of the predicted energies has a non-linear increase up to 80 TeV. After the peak it decreases rapidly. The energy estimator predicts higher energies than 100 TeV for a few events.*

The energy reconstruction results of the classes Gamma, Proton, Helium, Carbon and Silicon can be found in the appendix C. One feature all energy estimators have in common is the decreasing accuracy in the high energy region. The reason could be the lower statistic in those regions due to the chosen logarithmically increasing bin size. In conclusion, it is possible to do further calculations using NN for the task of the energy reconstruction of iron nuclei.

5.5. Differential Flux Determination of MC Fe^+

The Forward Folding algorithm was first calibrated on class Iron on order to reconstruct an iron energy spectrum without background. After the successful reconstruction of an iron energy spectrum with an original spectral index of $\Gamma = 0$ and $\Gamma = -2$ of MC simulations containing only signal events background was added in order to investigate the formula 4.3. Last, the classification and energy reconstruction results of a MC dataset created with real data spectral indexes are provided to the implemented and tested Forward Folding algorithm to test the reconstruction the MC real data-like iron energy spectrum.

5.5.1. Initialization of the Forward Folding Algorithm

The calibration of the forward folding for one specific class was done using the classifier and energy estimator response of the corresponding class. The dataset contains 320000 iron nuclei MC events created with a spectral index $\Gamma = 0$. The true energy is set from 1 TeV to 140 TeV which is maximal value of the predicted energy of class Iron.
The effective area is derived by the quotient of the TP to the total number of iron events and is multiplied by $1\,m^2$ only for MC events. It is shown in figure 5.24. In comparison to the effective area of γ-ray events reconstructed with the standard analysis tools the effective area of iron decreases in the energy region of 5 TeV to 13 TeV from 75 % to 28 %. The expected outcome is to have an increased ratio of TP to the total number of events in the regions of higher energies remaining at some point constant. However, the course of the iron nuclei effective area fluctuates having a minimum and two maxima in the energy range of 5-100 TeV.
The effective area of the classes Gamma, Proton, Helium, Carbon and Silicon are shown in the appendix D.1. In the direct comparison of those effective areas to the effective area of the class Iron its increased TP rate in the low energy regime is quiet unique. To get an understanding of the behavior of the classifier MC event images in the energy range of 5 TeV to 13 TeV are considered, see figure 5.25. Both figures show disjointed areas with undefinable geometrical shapes except the CT3 in 5.25a. There a muon with an energy higher than its minimal emission energy was detected forming a muon-ring. Possible features could be the DC-light or the disjointed areas of triggered pixels.
Further the energy migration matrix was normalized resulting in the PDF which was calculated using the predicted energies of the energy estimator trained on iron nuclei events. The PDF of the events of the TP of class Iron is shown in figure 5.26. It was created for a zenith angle θ of 20 ° and a maximum offset angle ψ of 2.2 °. The tightly packed, higher probability bins in the low energy regime is caused by the low scattering of the uncertainties. Whereas in higher energies the dispersion is increased due to lower statistics in the dataset which can be seen in the 5.23, too. Therefore the behavior coincides with the energy migration matrix of the class Iron.

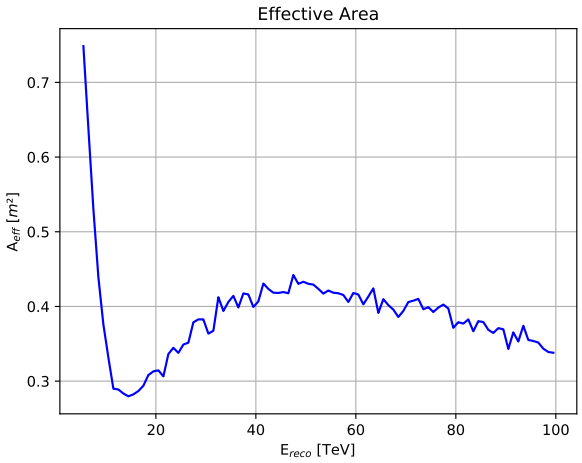
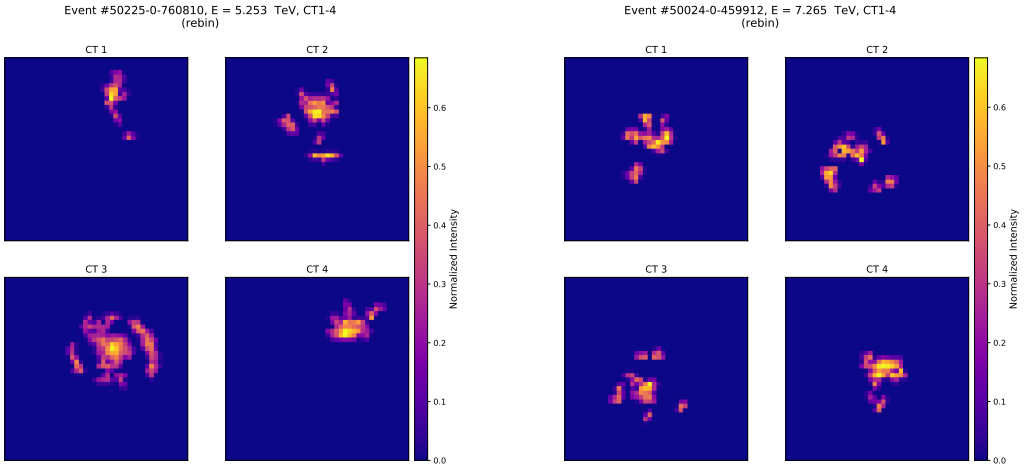


Figure 5.24. – *The figure shows the effective area of MC Fe^+ EAS in the range where it has real value entries . The correction factor of the classifier on class Iron has an enhanced ratio at lower energies as well as an increasing ratio between 30 TeV and 50 TeV. Whereas in the regions of higher energy the effective area is decreasing. The minimum value of the effective area is at 16 TeV.*



(a) C^+ *nuclei EAS* **(b)** Fe^+ *nuclei EAS*

Figure 5.25. – *Example MC EAS events of the classes Carbon and Iron. The CT1-4 images are rebinned with a resolution of* $42 \times 42\,\mathrm{px}^2$. *Both MC EAS show irregular and disjointed shapes. In addition, features which indicate DC-light can be seen, too. a) Event* 760810 *of run* 50225 *is simulated with an energy of 5.253 TeV. In CT3 a part of a muon ring is pictured. b) Event* 459912 *of run* 50024 *has an energy of* $E = 7.265\,\mathrm{TeV}$. *The telescope images show disjointed, triggered pixel areas.*

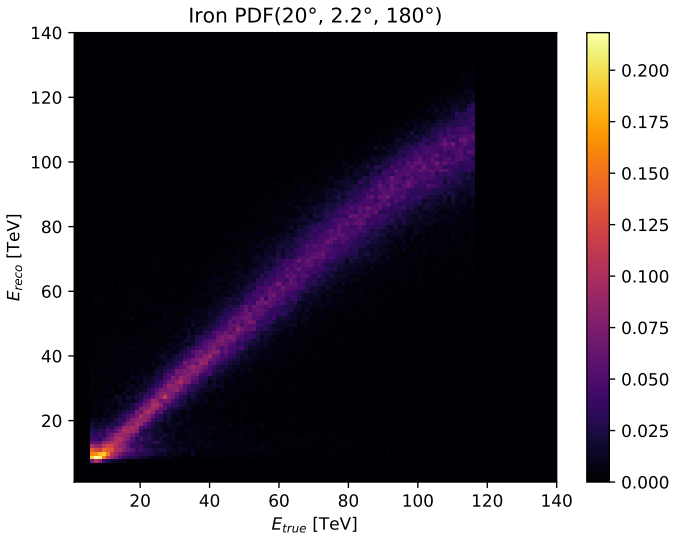
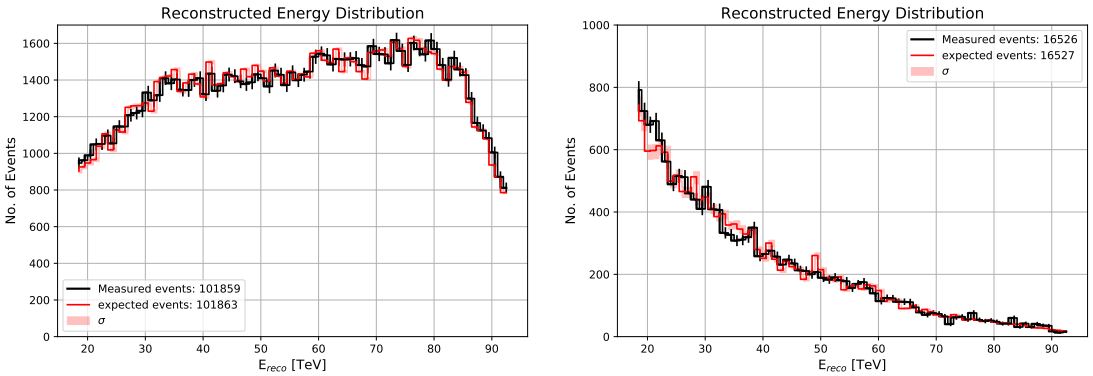


Figure 5.26. – *The PDF contains information about the probability of an Iron event having the energy E_{true} to be reconstructed with an energy E_{reco}. The probability per bin can be extracted from the colorbar. In the low energy regions the probabilities are higher compared to the energies above 10 TeV.*

5.5.2. Iron spectrum without Background

First application of the Forward Folding algorithm was on the calibration dataset containing only MC iron nuclei events with an energy distribution of $\Gamma = 0$. The spectrum energy threshold was applied on the measured energy distribution in order to suppress high errors of the energy estimator in the analysis. For the calculation of the assumed spectrum in the Forward Folding process the normalization energy E_0 value was set to the mean of the energy range fulfilling the condition of bias $\pm$ 10 %. The respone of the classifier and energy estimator as well as the fit result of the basin hopping minimizer is shown in the subfigure 5.27a.
In comparison to the original energy distribution of the events considering this energy range the most statistic is missing at the boundaries $E = 19\,\text{TeV}$ and $E = 93\,\text{TeV}$. In between the measured distribution fluctuates. Not knowing the original distribution it is not possible to guess the original parameters of ϕ_0 and Γ. The expected energy distribution of the iron nuclei events is similar to the measured distribution. Therefore it can be concluded that the implemented fit works. The fluctuations in the energy bins are due to the erroneous measured distribution but the total number of events differs only by 1 event.
In addition, the calculated PDF and effective area were used to reconstruct the energy distribution of MC Fe^+ EAS simulated with a spectral index of $\Gamma = -2$ and in total 128000 events. This was done to test if the implemented forward folding algorithm can reconstruct unseen data. In figure 5.27b the fitted event distribution compared to the

measured one is shown. The error of the measured data is $\sqrt{N_{on}}$. The expected energy distribution of the Fe^{+} events is similar to the measured one. In some bins the fit over estimates the expected events but these are compensated by the under estimation in other energy bins. In sum the expected number of events exceeds the total number of measured events by 1. The fit error of the negative log-likelihood is higher in the energy bins containing more statistic. The error was determined using the formula 3.6 which corresponds to σ and therefore the error falls in the 68 % containment radius.

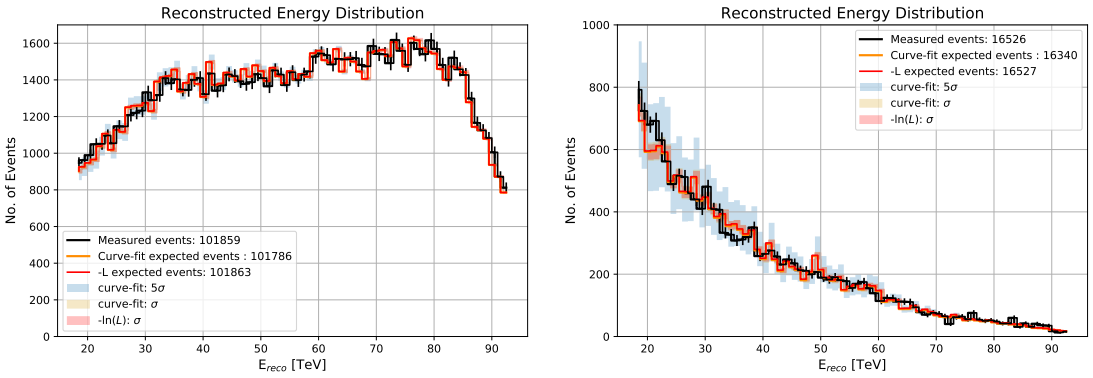


(a) $\Gamma = 0$ **(b)** $\Gamma = -2$

Figure 5.27. – *Fitting result of the calibration dataset and test dataset containing only iron nuclei events. The events not passing the spectrum energy threshold were set to zero. The expected energy distribution was fitted on the measured distribution. a) The sum over the events of both distribution differs by* 4*. Whereas considering the difference of the event number in each bin the difference fluctuates more. The error band of* σ *is hardly visible in the lower energy bins. At higher energies the error is near to zero. b) The total number of expected events is exceeds the measured number of events by* 1*. The energy distributions fluctuate more in the lower energy bins. At these energies the statistic is increased compared to the higher energies where the bins have only a few events for both distributions. Here the standard error of* σ *is nearly zero whereas at energies of 19 TeV to 55 TeV the measured and expected distribution differ more and therefore the error is higher.*

The negative log-likelihood fit is compared to the python function curve-fit. The advantage of curve-fit is that the error on the measured data can be taken into account during the fit. Both fit results are directly compared in the figure 5.28 for the datasets created with a power-law spectral index of $\Gamma = 0$ and $\Gamma = -2$. The 5σ error of the curve-fit in the subfigure 5.28a is smaller in comparison to the 5σ error for the distribution created with a spectral index of $\Gamma = -2$. One possible reason could be that the PDF and effective area were defined using the dataset with a spectral index equals zero. The σ error of the curve-fit and log-likelihood are very similar but the total number of expected events differ by 73 which results in $7.2 \cdot 10^{-4}$ %. In addition, the error is in energy region of 19 TeV to 50 TeV higher in comparison to the energies above

50 TeV. There the distribution of the expected and measured counts are similar to each other. In the subfigure 5.28b the total number of expected events of the curve-fit and the negative log-likelihood differ by 186 which corresponds to $6.05 \cdot 10^{-5}$ %. Hence, it can be concluded that the expected events distribution determined by the negative log-likelihood fit preserves the total number of events but the difference in the result to the curve-fit is small therefore both algorithms can be used.

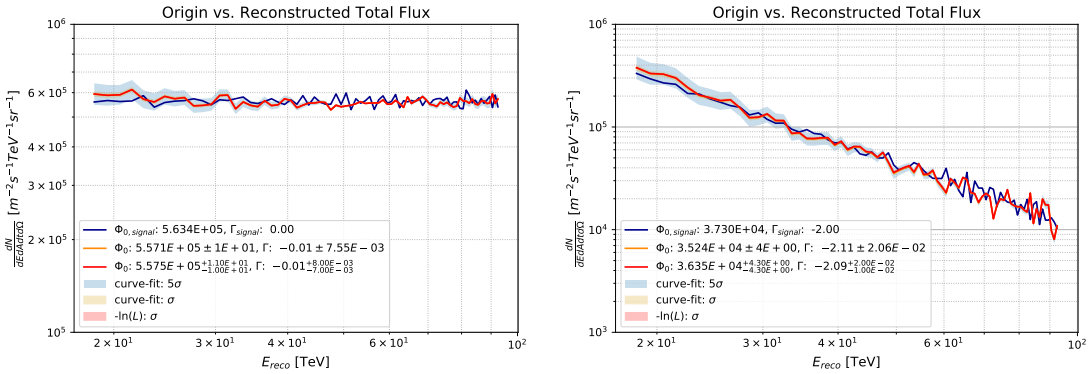


(a) $\Gamma = 0$ **(b)** $\Gamma = -2$

Figure 5.28. – *Comparison of the curve-fit and negative log-likelihood results. The graphics contain the measured and expected counts and the standard deviations* 5σ *and* σ*. a) The results of the curve-fit and negative log-likelihood are similar in their distribution. In the energy region of 18 TeV to 50 TeV the difference of the measured to the expected events per energy bin is higher in comparison to the energies above 50 TeV. This phenomenon expresses also in the error band of* 5σ *which is broader for energies of 19 TeV to 50 TeV. b) In the energy region of 19 TeV to 70 TeV the error band defined by* 5σ *is broader and also a difference of the error by* σ *of the curve-fit result in contrast to the fit error of* $-\ln L$ *is visible. For energies above 70 TeV the errors are smaller.*

The obtained fit parameters of $-\ln L$ were used to reconstruct the energy distribution of the input dataset containing only MC Fe^+ iron events. The ratio of the measured distribution to the expected is multiplied by the fitted power-law spectrum. The results for both datasets are shown in figure 5.29. There the original and fitted spectra are compared. In addition the error of the negative log-likelihood is delineated. It can be seen that the reconstructed spectrum is similar to the original spectrum. The fitted parameters $\phi_0 = 5.575 \cdot 10^5\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$ and $\Gamma = -0.01$ are similar to the original parameter values. The fit result of the tested dataset having an original spectral index of $\Gamma = -2$ is $\phi_0 = 3.635 \cdot 10^4\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$ and $\Gamma = -2.09$. The deviation of the fit parameters to the original is greater compared to the dataset having an original spectral index of $\Gamma = 0$. The reason is that the PDF and effective area were determined for the dataset with $\Gamma = 0$. For datasets where the Forward Folding algorithm has no prior knowledge the fit result can be worse. So far, the implemented Forward Folding

algorithm considering only pure signal datasets works correctly and the errors of the fit are relatively small.



(a) $\Gamma = 0$ **(b)** $\Gamma = -2$

Figure 5.29. – *Reconstructed iron energy spectrum of MC events with error impact of the Forward Folding process. Whereas the error of the effective area and PDF are neglected. The original spectrum (blue) is shown in comparison to the reconstructed iron spectrum suing curve-fit (orange) and* -ln L *(red). The red and orange fluxes have similar values therefore they cannot be distinguished in the plot. a) The spectrum reconstruction of the calibration dataset is similar to the original iron spectrum. The calculated errors on the fit parameters are not within the true value range of the original parameters. b) The iron energy spectrum reconstruction of the dataset containing iron events with a distribution of* $\Gamma = -2$ *fluctuates around the original iron spectrum. The error band of the fit parameter determined by the negative log-likelihood is hardly visible.*

The formula describing the pure power-law spectrum of the signal class in the whole FoV approximates the measured events with an uncertainty of the square root of the measured distribution per energy bin. Since this error for σ can not be provided to the used minimizer on -ln L and it is similar to the error of the spectrum fit it is neglected in the -ln L fit. For more realistic results using a σ multiplicity of σ higher than 1 is appreciated. The determination of the -ln L error has a higher computation time in comparison to the calculation of the curve-fit error. The latter can be determined from the returned covariance matrix. Furthermore, the results of the curve-fit and -ln L are similar. Summarizing, the Forward Folding algorithm for pure signal events can reconstruct the original MC parameters of $\phi_{0,MC}$ and γ_{MC} for the calibration dataset with a spectral index of $\Gamma = 0$ and the test dataset with a spectral index of $\Gamma = -2$. The next step was to use a mixed dataset containing signal and background events.

5.5.3. Iron spectrum with Silicon Background

The second application of the Forward Folding algorithm was to investigate the formula of the background estimation. Therefore a dataset containing in total 128000 MC Fe^+ and MC Si^+ events, 64000 per class was created having an energy distribution with a power-law spectral index of -2 for both classes.
To determine the number of background events for each energy bin the efficiency containing the ratio of FP predicted as Iron of class Silicon and the PDF of these predicted events were calculated. They are shown in figure 5.30. In the energy region 1 TeV to 5 TeV and 100 TeV to 120 TeV the efficiency is not defined due to the cut parameters of the used dataset. It can be seen that the efficiency is increased in the first 9 energy bins. This result coincides with the effective area of class Iron which is also increased in the first energy bins. Therefore it can be concluded that at lower energies these Si^+ EAS images have similar features as Fe^+ EAS images. After the peak the efficiency decreases steadily which is requested. The optimal result would be an efficiency of zero in each energy bin meaning the classifier could perfectly distinguish Iron from all other classes. The PDF was created by reconstructing the energy of the Si^+ events using the energy estimator trained on Fe^+ events. In figure 5.30 the PDF containing the probability of a Si^+ event to be reconstructed with a specific energy under the constraint that it was classified as Iron is shown. A diagonal shape in the PDF is visible but it is wider and scatters much more especially for higher energies and between 40 TeV and 70 TeV then the desired one. In addition, some high energetic events were predicted to have an energy of about 10 TeV. The dispersion in the higher energy bins is attributed to the lower statistic on which the energy estimator for Fe^+ events was trained on. Conspicuous is the fact that it is much more probable that an event is reconstructed with a lower energy.

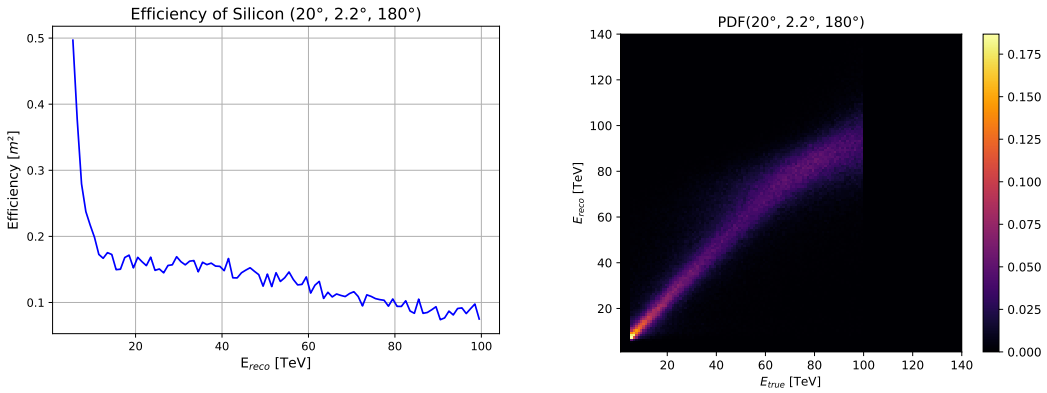
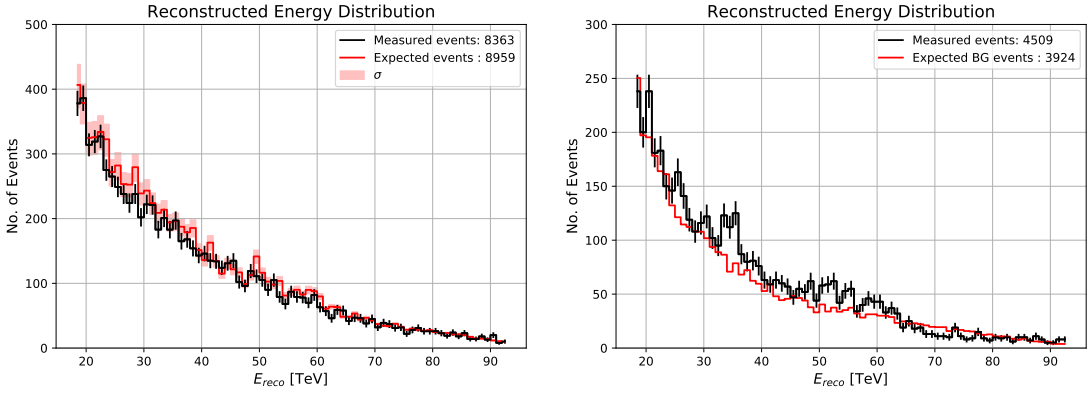


Figure 5.30. – *The figure shows the efficiency of class Silicon of Si^+ events predicted as Iron and their PDF. a) The efficiency is increased in the first 9 energy bins. From the energy of 11 TeV to 14 TeV the efficiency rapidly decreases from 0.5 m^2 to 0.16 m^2. Afterwards it decreases steadily. b) In the PDF an approach of a diagonal line can be seen. Higher uncertainties occur in the middle and part and high energy regime. Overall the reconstructed energies are overestimated compared to their true value.*

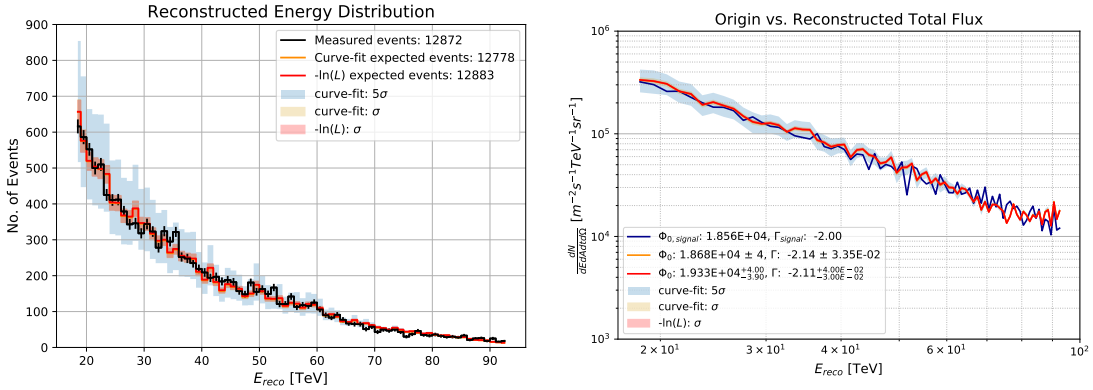
The efficiency and PDF of true Silicon events classified as Iron were multiplied with the predefined power-law spectrum of the MC Si^+ events. The result was added to the expected signal events of class Iron for an assumed spectrum. The comparison of the measured to the expected signal and also the measured to estimated background events are shown in the figure 5.31. In the subfigure 5.31a the error on the measured signal and of the fit using the negative log-likelihood were considered. The result of the expected distribution has a higher uncertainty in the lower energy bins. There the distribution is above the distribution of the measured signal events. For increasing energy the difference gets smaller. In total, the number of expected events has 596 events more than the measured one which corresponds to 7.13 %. In the subfigure 5.31b the measured and expected background events are compared. In the optimal case they would have the same number of evens and their distributions would also be similar. Due to measurement errors and errors on the efficiency and PDF of Si^+ events predicted as Iron the expected events deviate from the measured ones. The defined formula for the background estimation works but can be improved using more train examples for the multi classifier. In addition, a deviation is to be expected because this dataset was not the one the efficiency and PDF were created with. Fitting the signal and background expected events is not useful since it is only possible to fit the total spectrum using the formula 4.4. Therefore only the parameters ϕ_0 and Γ of the signal class were fitted.
The total number of measured and expected events and also the reconstructed total flux are shown in the figure 5.32. In the subfigure 5.32a it can be seen that the total number of expected and measured events differ by 94 for the curve-fit and by 11 for the -ln L approach. Overall the expected distribution of two fit approaches are sim-

ilar in their course and also the standard deviation errors for σ coincide. The error on the measured data is given by the square root the the measured count per energy bin. In the higher energy bins the errors are smaller compared to the lower energy bins where the uncertainty is increased. In the subfigure 5.32b the origin and reconstructed total flux is shown. The fluxes have a similar signal normalization flux $\phi_{0,signal}$. The fit results of the spectral index for the curve-fit is $\Gamma = 2.14$ and for the -ln L approach $\Gamma = 2.11$. Their signal normalization ϕ_0 is $1.868 \cdot 10^4\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$ and $1.933 \cdot 10^4\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$, respectively. Both signal flux normalization results exceed the original one by $120\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$ or $770\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$. The standard deviation for 5σ contains both reconstructed and the origin flux normalization. The origin normalization of the Si^+ background is $\phi_{0,BG} = 1.857 \cdot 10^4\,\mathrm{m^{-2}s^{-1}TeV^{-1}sr^{-1}}$.



(a) *Signal Fe^+ events* **(b)** *Background Si^+ events*

Figure 5.31. – *The figures show the measured and expected events of the signal and background class. Interesting are the parts where the distribution of the expected signal events are over estimated and the background events are under rated. a) The expected distribution overshoots the measured one in nearly each bin. This can be also seen in the total number of events. The expected distribution has 596 events more than measured one. b) The expected background distribution is is similar in their shape to the measured one. The distribution of the expected number of events has 585 less events compared to the sum over the measured background events.*



(a) *Total measured and expected distributions*

(b) *Total Flux*

Figure 5.32. – *Illustration of the measured and expected energy distribution and the origin and reconstructed total flux. The error band of* 5σ *and* σ *are shown. The fitted and corresponding distributions look similar. a) The error of the total number of measured events is considered. The* σ *error of the curve-fit and* $-\ln L$ *fit are nearly the same in each energy bin. Whereas the* 5σ *error determined by the curve-fit expresses range of possible results in the ordinate. The distribution of the total expected events of the curve-fit and* $-\ln L$ *are can not be distinguished in this plot. b) The origin and reconstructed total flux are compared in this figure. It can be seen that the fluxes have a similar total flux normalization. Only the error of* 5σ *is visible and contains the original flux in the energy region of 19 TeV to* $\sim$ *50 TeV.*

In conclusion the investigated formula for the background estimation can be used to approximate the expected background events. The total number of events is not preserved but possible reasons are the error on the PDF and efficiency of as Iron predicted Si^+ events and that the tested dataset is not the calibration dataset. The 5σ error contains the true distribution and flux and should be used for a more realistic result. The results of the formulas for the expected signal and background events tested on MC datasets containing only signal or both types of events revealed that it is possible to use the response of the classifier and energy estimator to reconstruct a spectrum. Therefore this approach was tested on real data of PKS 2155-304. The results are shown in the following section.

5.5.4. Iron spectrum with Unknown Background

To verify the investigated formula for the signal and background estimation one further test was done on a MC dataset where all labels are the same. This was done to simulate a real dataset but with known composition of the of all six classes. In total the dataset contains 498729 events. The spectral index of each class are the same as for real data. The result of the classification task is compared to the original division in table 5.8. The predicted total number of events of each class contains TP and FP. The percentage of the events of the classes Proton and Helium is more distributed over all classes compared to the others. The flux normalization $\phi_{0,55}$ was defined at a normalization energy $E_0 = 55\,\text{TeV}$. The values of the effective area and time are equal to one for MC datasets and the value of the solid angle Ω is set to $0.00598\,sr$. The spectral indexes are taken from [68].

Table 5.8. – *Comparison of the total number of predicted and origin events. The predicted events contain FPs and TPs.*

Particle type	$\phi_{0,55}\ [TeV^{-1}m^{-2}s^{-1}sr^{-1}]$	Γ	No. AP	No. predicted events
γ	872.88	2.0	3000	7105
H^+	15073.07	2.69	167729	120564
He^+	15564.69	2.69	173200	130928
C^+	5330.93	2.67	57200	99280
Si^+	4815.89	2.62	47200	82217
Fe^+	5330.93	2.6	50400	58635

The outcome of the energy reconstruction using the energy estimator trained on MC Fe^+ events in comparison to the true energy is shown in the figure 5.33. In the left figure the true energy distribution of all events from all classes is shown. The high number of examples in the low energy bins is due to the used spectral indexes of the classes. Therefore low energetic events occur more often compared to the high energetic ones. Considering the predicted energy distribution the number of events is increases in the first energy bins. After the peak at $\sim 13\,\text{TeV}$ the counts decrease steadily until 55 TeV. At this energy the distribution slightly increases and decreases further right after the saddle point at the energy 59 TeV. Taken the fluctuations into account this energy region can also be seen as plateau. The interval of 60 TeV to 100 TeV has a much higher gradient compared to the region of 13 TeV to 55 TeV. No definite flux can be determined from this reconstructed energy distribution.
In the figure 5.34 the expected events in comparison to the measured events is shown. The measurement decreases in the energy range 19 TeV to $\sim 60\,\text{TeV}$. At this energy a saddle point can be seen which was already visible in the reconstructed energy distribution of all events. Afterward it decreases further. The total number of expected events

are under- and overestimated of the curve-fit and negative log-likelihood, respectively. The course of both expected distributions fluctuates around the truth. The plateau can not be fitted with the assumed formula of the signal and background events. In comparison to the right figure in 5.33 the classifier sorts the most events which contribute to the plateau out. This means Fe^+ events do not cause this shape containing two different gradients. The confidence interval of 5σ contains the measured distribution in the interval from 19 TeV to 68 TeV. Afterwards the error and the difference between the expected and measured examples is nearly zero.
The figure 5.33 shows the origin and reconstructed signal fluxes with the true and fit parameters of the class Iron. Also the total fitted fluxes are illustrated. In the subfigure 5.35a it can be seen that the fit parameter of $\phi_0 = 1.025 \cdot 10^4 \pm 4.22\,\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ is nearly twice as much compared to the true value but the shape is preserved. This is true for using the -ln L due to the well fitted spectral index of $-2.66^{+6\cdot10^{-2}}_{-5\cdot10^{-2}}$. For the curve-fit both power-law parameters were overestimated. The signal normalization flux is $\phi_0 = 8.762 \cdot 10^3 \pm 3\,\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ and $\Gamma = -2.88 \pm 4.39 \cdot 10^{-2}$. In the subfigure **??** the reconstructed total fluxes are shown. Both graphs look similar. In the high energy region the difference gets greater due to dissimilar spectral indexes. The containment radius of 5σ contains both fitted, total fluxes in the energy region 19 TeV to 75 TeV.
The overall result of the Forward Folding algorithm on a mixed dataset containing all classes reproduces the spectral index with a given uncertainty which contains the true value at least if $-\ln L$ is used. The flux normalization is overestimated by both fit functions. Possible reasons could be the neglected error on the PDFs, efficiencies and effective area. The error on the efficiencies which should compensate the classifiers decision could be the greater than the error of the PDFs.

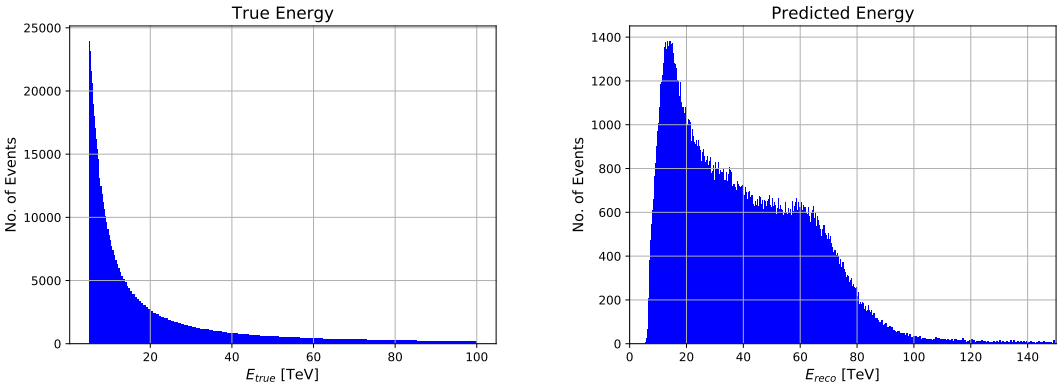


Figure 5.33. – *Comparison of the true and predicted energy distribution of all events. The original distribution shows a smooth definable decrease of counts. Whereas the predicted energy distribution shows two different gradients. One from 13 TeV to 55 TeV and the second from 60 TeV to 100 TeV. In between a plateau containing a saddle point can be seen.*

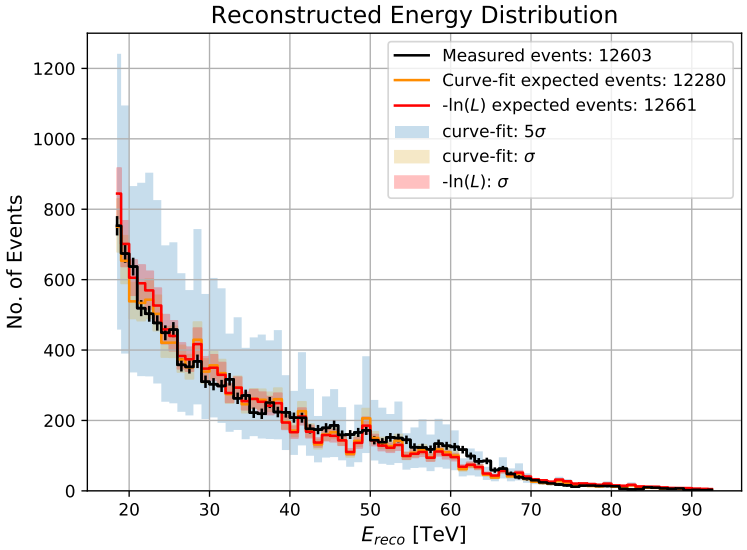
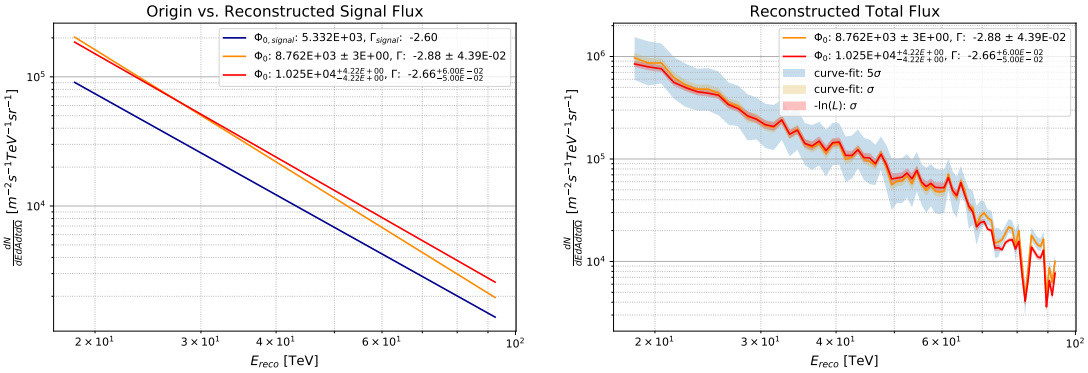


Figure 5.34. – *The figure shows the measurement and expected reconstructed energy distribution. The total number of events is underrated considering the result of the curve-fit and overestimated by the* -ln L *fit. The* 5σ *standard error of the curve-fit contains the truth. Whereas the containment radius* R_{68} *of* σ *does not contain the measured event distribution.*



(a) *Measured vs. Expected events* **(b)** *Total flux*

Figure 5.35. – *The figures show the comparison of the original vs. the reconstructed fluxes and total fluxes. The origin flux is marked by the blue solid line, the flux reconstructed using the* -ln L *is shown in red and the flux determined using the curve-fit function is drawn in orange. a) The fitted fluxes are above the original flux. The course of all three signal fluxes is similar. b) Both total fluxes have a decrease. The* 5σ *error contains both fit results in the energy bins 19 TeV to 75 TeV.*

To summarize, the defined formula to estimate the number of background events in the measurement works on MC EAS events. The flux normalization is not preserved and is overestimated in the case of a mixed dataset containing signal and background events. Especially, if all 6 classes are considered. The γ-ray events have the lowest percentage to be classified as Iron compared to the heavier nuclei. However, the PDF of γ-ray events classified as Iron shows that the predicted energies of these events are nearly guessed and therefore contribute unpredictable to the measured event distribution. The efficiency and PDF of all background classes can be found in the appendix D.2. The results can be improved using a classifier having a higher recall on Fe^+ events and overall a higher accuracy. In the following section, the verified background formula, the CNN_α and the energy estimator of class Iron are applied on real data PKS 2155-304 events.

5.6. Iron Energy Spectrum using PKS 2155-304 Data

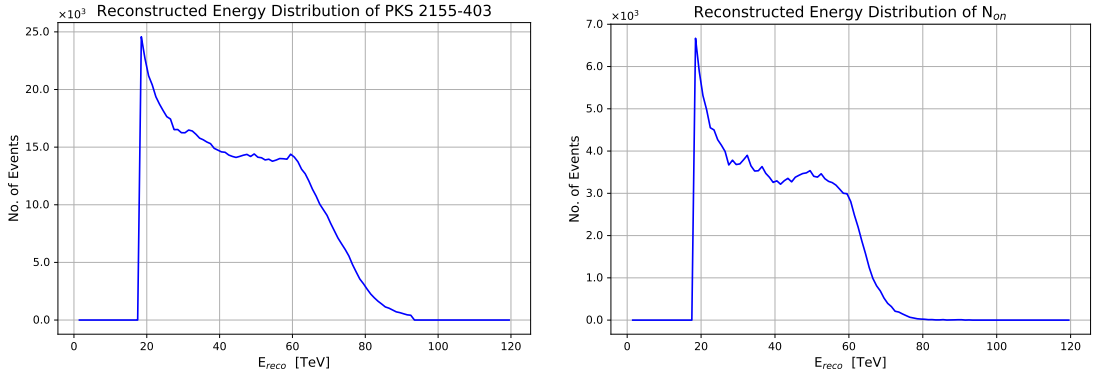
The best multi-classifier CNN_α and energy estimator of the class Iron were used to classify and reconstruct the real data of PKS 2155-304 events. The real runs were observed under a zenith of $\in [19.5\,^\circ, 20.5\,^\circ]$ and had a maximum offset of $2.5\,^\circ$. These observation parameters were chosen because the NNs were trained on MC events simulated under a zenith angle of $20\,^\circ$, an azimuth angle of $180\,^\circ$, with a maximum offset of $2.2\,^\circ$ having energies of 5 TeV up to 100 TeV. In addition, a multiplicity of 2 was specified for MC and real data. The effective area of real data is multiplied with $A_{geo} = \pi 1500^2\,\text{m}^2$ which is the maximum radius of the MC EASs on the ground. The total observation time of all considered PKS 2155-304 runs is 24733 s and the solid angle Ω is 0.00598 sr.
The PKS 2155-304 classification result of more than $1.8 \cdot 10^6$ events are shown in table 5.9. The flux normalization ϕ_0 is calculated at the normalization energy $E_0 = 55$ TeV. The spectral indexes are adapted from [10]. The predicted events per class of real data events can be interpreted as *class*-like events because these have predominantly features of a specific MC EAS described by one class. The most events were labeled as class Iron. In total, 642120 of the considered PKS 2155-304 events have been classified as Iron which corresponds to $\sim 1/3$ and about 2 % of all events were predicted to be of class Gamma.

Table 5.9. – *Information of the assumed background flux and total number of predicted events per class. The predicted events contain FPs and TPs.*

Particle-like type	ϕ_0 $[TeV^{-1}m^{-2}s^{-1}sr^{-1}]$	Γ	No. of predicted events
γ	$1.71 \cdot 10^{-12}$	3.59	34686
H^+	$1.70 \cdot 10^{-6}$	2.66	258092
He^+	$2.09 \cdot 10^{-6}$	2.58	272938
C^+	$2.89 \cdot 10^{-7}$	2.68	261763
Si^+	$4.73 \cdot 10^{-7}$	2.64	314620
Fe^+	$7.39 \cdot 10^{-7}$	2.59	642120

The result of the energy estimator is shown in the figure 5.36. The unclassified and classified events are directly compared. The distributions in both figures are similar e.g. combinations of several spectral indexes can be seen by different gradients. Each distribution has one conspicuous saddle point. These are at 59 TeV and 49 TeV for the unclassified case and the classified one, respectively. In 5.36a the total number of events is about four times higher than in 5.36b. Also the second visible gradient from 59 TeV to 93 TeV is steeper compared to the second one in 5.36b. In the energy interval 49 TeV to 62 TeV the distribution is nearly constant. Further, the distribution contains events in each energy bin in the energy region where the *bias* is $\pm$ 10 %. In comparison, the counts in the right figure are close to zero at an energy of 80 TeV. In 5.36b the course within

the energy interval 29 TeV to 49 TeV can be seen as almost constant. Such trends in the distribution can hardly be fitted by the Forward Folding algorithm using the formula 4.4.
Despite this unique course of the reconstructed measurement the Forward Folding algorithm was applied to fit the response of the classifier and reconstruct the flux. The outcome can be seen in figure 5.37. The measurement has an unique shape. It first decreases in the energy interval 28 TeV to 52 TeV. Afterwards it stays almost constant. This part can also be described as a plateau. Usually, a steadily decreasing of the course is expected which would be more fitting to the assumed power-laws of the signal and background classes. Whose expected course is shown by the blue solid line. The fitted expected distributions coincide more to the origin expected events than to the measurement. In the energy interval 37 TeV to 60 TeV the fitted distributions are close to the origin one. In the lower energy region the result of the -ln L fit matches better than the curve-fit whereas in the higher energy region the curve-fit is nearer to the origin and measured distribution. As already mentioned, it is not possible to fit the measured event distribution well. The expected ones overshoot in the low energy interval to compensate the plateau in the middle part because overall the total number of events differs by $\sim$ 1411 derived form the -ln L compared to the measured ones. Whereas the curve-fit underestimates the total number of events by about 30000 events. It can be concluded that the approach of the negative log-likelihood fit has a higher accuracy compared to the curve-fit function. Therefore on real data or on unique measured shapes the -ln L results in a more reliable outcome.



(a) *Unclassified events* **(b)** *Classified as Iron*

Figure 5.36. – *The figure shows the reconstructed energy distribution of unclassified and as Iron classified PKS 2155-304 events in the energy range where the bias is ± 10 %. Both distributions look similar to each other. The number of events are almost four times less in b). a) The reconstructed energy distribution of PKS 2155-304 events show a combination of different gradients due to the saddle point at 59 TeV. b) The events classified as Iron show an unique course. The distribution first decreases and increases again in the interval 41 TeV to 49 TeV. After the peak at 49 TeV it decreases further.*

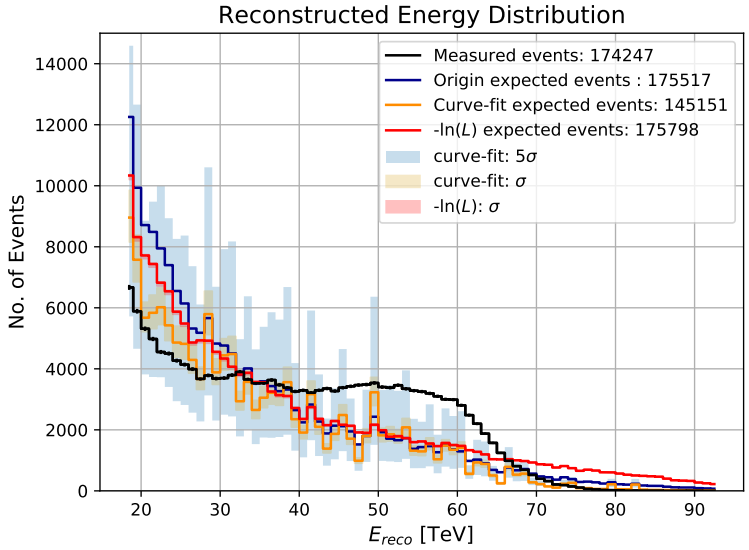
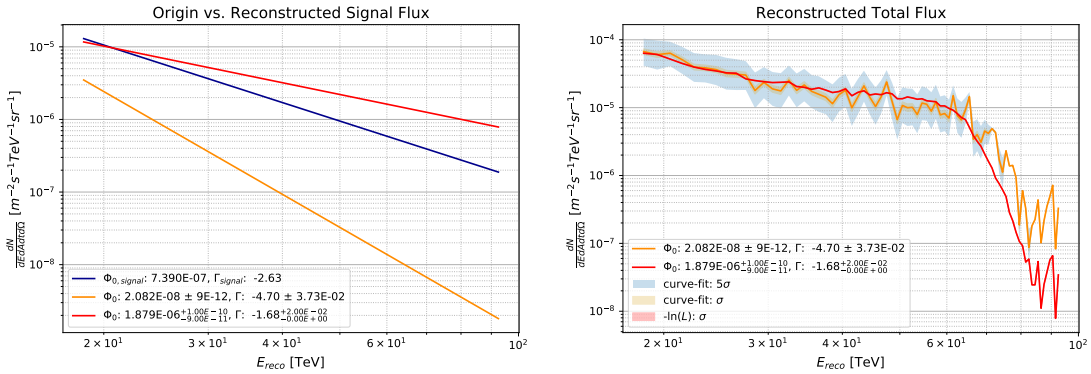


Figure 5.37. – *The figure shows the measured and the expected events. The measurement decreases in the first energy bins. In the energy range of 28 TeV to 52 TeV it stays almost constant with smaller fluctuations. Afterwards it decreases further. In total, 174247 events were predicted to be Fe^+-like. Both fit method results show a different course compared to the measured distribution especially in the energy range where the measurement has a plateau. The standard errors σ and 5σ on the curve-fit function do not enclose the measured distribution. Which can also be seen for the -ln L fit. The fit outcomes are similar to the distribution of the origin expected events. Especially in the energy interval 37 TeV to 60 TeV. In the lower energy region the -ln L fit is next to the origin expected distribution. Whereas the curve-fit fits more in the high energy region.*

Further the origin and reconstructed signal flux as well as the reconstructed total fluxes in the *bias* $\pm$ 10 % range were considered. They are shown in the figure 5.38. In the subfigure 5.38a it can be seen that the fitted fluxes do not match the assumed flux of Fe^+-like particles using the spectral index and flux normalization from the table 5.9. The curve-fit spectral index Γ is higher compared to the origin one which results in a much steeper decrease. In contrast, the spectral index determined by the -ln L has a smaller value and decreases therefore slower. The relatively high deviation to the assumed origin spectral index can hardly be seen in the expected distribution, see figure 5.37, due to the similar courses. Especially the resulting slope of the curve-fit matches the original expected events. Reasons could be the stronger fluctuating number of events within the energy range where the plateau occurs in the measurement. The magnitude of the fit errors of the curve-fit and -ln L of ϕ_0 are also curious. It is in the order of four magnitudes smaller than the value of the fitted flux normalization. This means the fit algorithms are sure in their predictions and in addition that the origin value is not within the error range. This statement is also valid for the spectral index Γ. In

the subfigure 5.38b the reconstructed total fluxes are shown. The unique shape is due to the multiplication of the fitted pure power-law with the ratio of the measured and expected events per energy bin. To conclude, the result of the classification task needs to be improved due to the strong impact on the reconstructed flux course.



(a) *Signal flux* **(b)** *Total flux*

Figure 5.38. – *The figures show the signal and total flux of class Iron in the bias range where the limits ± 10 % are hold. a) The origin flux (blue) is between the predicted flux using* -ln L *(red) and the reconstructed signal flux of the fit method curve-fit (orange). The gradient of the curve-fit is steeper compared to the origin and fitted gradient of* -ln L*. The fitted fluxes do not match the truth. b) The figure shows the reconstructed total fluxes. The orange line marks the curve-fit and the red line corresponds to the* -ln L *fit result. Within the energy region of 19 TeV and 60 TeV the total fluxes have a descent decrease compared to the rapidly decreases for higher energies. The standard error of* 5σ *contains the total flux of the* -ln L*, too.*

The application of the multi-classifier CNN_α and energy estimator trained on MC Fe^+ EAS events on real data PKS 2155-304 runs resulted in a unique distribution. The distribution has a plateau within the energy range 28 TeV to 52 TeV which was not possible to be fitted by neither fit methods. The fit algorithms overestimated the expected events in the lower energy regions to compensate the plateau. This resulted in a too high and too low Γ value of the curve-fit and -ln L, respectively. The standard error of 5σ did also not contain the measured distribution. Interesting was the result of providing the known origin power-law to the Forward Folding algorithm. It was similar to the fitted ones. Which might be an indication that the classifier is worse on real data because MC simulations contain slightly different features compared to real data. Further, the reconstructed total flux which was calculated using the formula 3.5 shows no definite slope. The measurement's impact is too high to be compensated by the fit results.
The results of the preprocessing step, the classification and energy reconstruction task as well as the fit results of the Forward Folding algorithm are recapitulated and further discussed in the following chapter.

6. Discussion

The rebinning result of the raw CT1-4 images using a resolution of $42 \times 42\,\mathrm{px}^2$ did preserve the orientation and intensity of the raw images. In addition, the EAS's shape remained almost unchanged. Thus, the chosen resolution of $42 \times 42\,\mathrm{px}^2$ can be used as a compromise of possibly loosing features which are possibly contained in the rebinned images of $64 \times 64\,\mathrm{px}^2$ and the opportunity to use the built architecture for the energy reconstruction task due to computational limitations.

The first main part is the classification of heavier nuclei, H^+ and γ-ray EAS images. Therefore an architecture study was done to investigate the accuracy dependency. The results were that all tested structures were similar. One possible reason could be the chosen energy range of 5 TeV to 100 TeV where the CT1-4 images of γ-ray and H^+ EAS have increased different features. This can be concluded due to the result of [58]. There the NN was trained on events in the energy range of 0.02 TeV to 100 TeV and achieved a total test set accuracy of 96.10 %. In comparison the total test set accuracy of CNN_α is 98.14 % in the energy range of 5 TeV to 100 TeV. An additional possible reason could be the diminished features due to the applied lower resolution of $42 \times 42\,\mathrm{px}^2$. A further study of the classification of low energy EAS images or a morphology study of the images could enhance the understanding of this behavior.
The energy distribution study did not reveal an energy dependency of the trained NN but that it is useful to choose a predefined spectrum which can indeed increase the accuracy. The images of γ-rays and H^+ events are too different and therefore a further study training classifier on heavier nuclei EAS images might reveal an energy dependency. The images of heavier nuclei look similar to the human eye and thus the NN model needs to learn features e.g. the energy to distinguish the classes.
The metric results of all binary classifiers did reveal a trend of increasing accuracy if the molar mass difference Δu of the signal and background class increases. The classifiers C^+ vs. He^+, C^+ vs. Si^+ and Si^+ vs. Fe^+ did not follow this phenomenon. They were examined in more detail. It could be seen that they must have similar features over the whole energy region and especially at lower energies due to more FP prediction in this interval. It can be concluded that signal EAS images having a higher molar mass difference to their competitive background class can be distinguished better.
The models created for the multi- and multi-task classification based on the architecture found by the architecture study have similar results. The result of the classifier was used later for the reconstruction of an iron spectrum. The chosen one was the CNN_α having an accuracy of 56.68 %. The results of this classification task are not representative due to the small total number of 96000 MC events the classifier was trained on. Usually, to get an expressive outcome the classifier should be trained on 10 times more data

examples. Internally the results are comparable because all classifiers were trained on the same dataset. It was shown that the images of class Gamma can be distinguished best from all other classes. Whereas for increasing charge Z the F_1-Score for the single classes decreases. This is not valid for the class Iron where the F_1-Score is higher compared to the ones of the other heavier nuclei initial particle classes Helium, Carbon and Silicon. This means that the class Iron has more features which are different from the heavier nuclei of the other 3 classes. Whereas the features of these classes are more correlated. The classification results were considered energy resolved to see at which energies the images have more uncorrelated features. The result is that higher energetic EAS images can be better distinguished as the ones having low energies. For the classes Helium, Carbon and Silicon these conclusion is also true but not as obvious as for the other 3 classes. The performance results of the multi-classification could be improved using more nodes in the layers or use residual layers to transport information from the upper levels to the deeper levels.
During the training of the different classifier architectures some difficulties occurred. Especially the hyperparameter determination for the multi-task networks was challenging. The learning dependency on the margin m as well as on λ_s is sensitive. A small change in these values led to the inability of learning one class or in the worst case of learning only one class. A parameter study to determine the m like it was done in [31] could help to set this hyperparameter only once. In addition, for these tasks the optimizer Nadam was used because the error decreased in the first five steps rapidly. The optimizer Adam could not handle this behavior and predicted too high gradients for the next steps which resulted in an invalid loss. This means the optimizer was further beyond the minimum. Using a smaller learning rate decay could lead to useful loss results of the optimizer Adam for the multi-task of classification and similarity measurement.
To support the classification task a further cut parameter was investigated using the DC-light occurring for EAS induced by heavier nuclei. This cut parameter study did not reveal any useful further cut for the creation of the dataset such that the classifier learns to focus on the DC-light feature. A more specific determination of the DC-light cut parameter could lead to a more distinct result. Therefore the approach of [6] searching for the next neighbors in a hexagonal grid and determine the ratio of the main pixel to its next neighbors would be more expressive.

The optionally preprocessing step PCA was investigated to enhance the distinction task. The algorithm was applied with the intension of removing redundant information in the EAS images considering a dataset of two classes. First application of PCA on a dataset containing γ-ray and H^+ EAS events using explained variances of 99 %, 95 % and 90 % did not improve the accuracy of the classifier. Reasons could be that these images have few correlated features. This hypothesis is supported by the result using explained variances 99 % and 95 % where the recall of the class Gamma is reduced and the specificity of the class Proton is increased. This is an indication that the missing 1 % and 5 % were features which describe parts of the γ-ray EAS. Interesting would be which principal components correspond to these missing features and what do they describe. In addition, an explained variance of 80 % was tested on this dataset. The

reconstructed images look blurred and distorted by additional features. The NNs accuracy was decreased by 14.27 % thus it is not useful to use explained variances $< 90\,\%$. Further, PCA was applied on an other dataset containing Si^+ and Fe^+ EAS events. For the variances 99 % and 95 % the accuracy was increased compared to the trained NN on the dataset with the original rebinned images. This means correlated information in the EAS of Si^+ and Fe^+ was removed and the classifier was able to focus on features which are less correlated. One possible feature could be the DC-light which occurs more often by heavier nuclei inital particles. Both particles have a higher charge Z compared to He^+ thus the DC-light feature is more likely. To proof this claim the investigation of DC-light as it was done on the raw data in [6] could be applied on these dimension reduced images. In addition, an explained variance of 90 % was tested but the results are worse compared to the classifier trained on the original dataset. In conclusion, the application of PCA on heavier nuclei could improve the classification results under the constraint that the classifier is trained on all classes except γ-ray EAS images.

The second main part of this thesis considered the energy reconstruction of the 6 classes. The architecture for the energy prediction task is similar to the one used for the classification task. Therefore it could also be applied on the distinction task of heavier nuclei to test if the residual layer improves the accuracy. The study of the choice of the spectral index was done to test if a dataset containing events dependent or independent of the energy distribution increases the relative energy bias range. The outcome was to use a spectrum with $\Gamma = 0$ which is reasonable because the events are equally distributed over the whole energy interval. Therefore the features of images from any energy region are not preferably learned. The study of the scale factor showed that applying a linear scaling increases the energy range within the $bias \pm 10\,\%$ interval from 41 TeV to 60 TeV. Log-scaling might be useful if the energy interval is large (> 1 PeV) or the requirement is to have a similar energy resolution for low and high energies compared to the linear scaling where the resolution is increased in the low energy interval but is smaller in the *bias* range $\pm$ 10 % and for high energies. During the dataset creation the energy bin width was defined to increase logarithmically in size with increasing energy due to lower statistics in those energy regions. Thus the distribution with a spectral index of $\Gamma = 0$ seemed to increase instead of being constant. The requirement was that the events should be equally distributed but this is only possible for finite binning. Therefore the questions are could this be improved by collecting all similar events in an energy interval assuming that the events are then equally distributed over the whole energy range and what is the meaning of an equal distribution. Further improvements could be achieved by using more events in the high energy region or a broader energy interval to increase the $bias \pm 10\,\%$ interval.
Comparing the energy reconstruction *bias* range $\pm$ 10 % of the class Gamma of [53] using the ImPACT analysis which reaches from 100 GeV to 100 TeV to the achieved range from 7 TeV to 94 TeV using DNN it can be seen that the *bias* value is smaller using ImPACT. Remembering that the energy estimator was only trained on MC event in the energy interval 5 TeV to 100 TeV it could compete if it is trained on a wider true energy range and on more event examples. In conclusion, it is possible to reconstructed the

energy of hadronic particle which is not possible with the standard analysis tools.

The third main part was to use the response from the CNN_α and energy estimator on PKS 2155-304 real data to reconstruct an iron energy spectrum. For this reason the Forward Folding algorithm was implemented and first tested on MC Fe^+ events to reconstruct the input dataset's spectral index and normalization flux. Therefore the effective area and the PDF of the MC Fe^+ events which are the correction factors of the classifier and energy estimator were derived. The increased effective area in the first energy bins was suspicious thus the MC images from class Iron to other classes were compared. One noticeable feature was the uncorrelated triggered pixel areas in the Fe^+ EAS images in the low energy regions. In the higher energy interval the heavier nuclei EAS have similar shapes therefore the effective area of Iron is decreased. The effective area of classified γ-rays is similar in their shape compared to effective area of [53] determined using ImPACT in the energy interval 5 TeV to 100 TeV. The effective area in [53] is multiplied by the geometric area whereas the effective area of the class Gamma was not, see figure D.1a. Considering the effective area determined by the NN the fluctuations are relatively small and the course increases in the first energy bins until $\sim$ 40 TeV. Afterwards it slightly decreases which can also be seen in the post cuts effective area of [53]. The PDF of reconstructed Fe^+ events contains the probabilities with which energy the event is reconstructed for a given true energy. These are decreased and more scattered in the higher energy bins. This means that events having a higher true energy are more likely to be reconstructed with a greater difference to the true energy than for events with lower true energy. The effective area and PDF were folded with an assumed spectrum. Two approaches were tested to get the best fit parameters of the spectrum. The first one was the minimization of the negative log-likelihood and the second was the curve-fit function. The results were compared to the measured data. The σ standard error calculation of the log-likelihood was expensive in calculation time due to a fine grid scanning of the fit parameter space. For a more realistic result the standard error using 5σ should be considered. Using the formula of [62] the error on the fit parameters can then be determined by $-\ln L = -\ln L_{min} + 25/2$. Another disadvantage was that the measurement error could not be considered during the fit. However, the curve-fit could take the error on the measured data into account and due to the returned covariance matrix the σ and 5σ error could be easily calculated. Both approaches were directly compared. In the sense of the large sample limit the non linear least square problem χ^2 and $-\ln(L)$ can be approximated by a Gaussian function and therefore the errors are similar in their impact.
The Forward Folding algorithm was first tested on datasets containing only MC Fe^+ events having a spectral index of $\Gamma = 0$ and $\Gamma = -2$. The parameter results of both fit approaches were near to the original dataset values of ϕ_0 and Γ. The next step was to investigate the formula for the background estimation. It contains the efficiency of background events of one class predicted as FP of the signal class. Whereas the PDF contains reconstructed Si^+ events using the energy estimator of the signal class. The efficiency and PDF were folded with the known spectrum of the background. It is not possible to fit the signal and background spectrum simultaneously because only the total

flux can be fitted using the defined formula 4.4.
The dataset which was used to investigate the background estimation formula contained signal Fe^+ and background Si^+ MC events. The course of the efficiency of the Si^+ events is increased in the first energy bins which matches with the effective area of the signal class Iron. This means the Si^+ and Fe^+ events have similar features in this energy interval. An efficiency close to zero would be optimal. The entries PDF of the background events are wider scattered especially for the higher true energies. Overall the PDF Si^+ events reconstructed using the energy estimator trained on Fe^+ images looks similar to the PDF created from Fe^+ nuclei which in turn explains the relatively low accuracy of the binary classifier trained on Fe^+ and Si^+ EAS images. The energy estimators of the classes Iron and Silicon learn similar features independent of the other class. Again both approaches were used to get the parameter values of the Fe^+ spectrum. The measured signal and background events were compared to the fitted and known distributions. The Forward Folding algorithm for the determination of the background events underestimates the total number of events by $\sim 1/8$. Possible reasons are the errors on the efficiency as well as PDF of the background events. More training examples for the classifier would lead to a more expressive result of the efficiency. Also this dataset was not used to calculate the efficiency and PDF therefore higher deviations are more likely. Due to the fact that the formula 4.4 fits the total spectrum parameters the energy bins where the background events were underestimated are compensated by the fit resulting in an overestimation of signal events in these bins. The result of the fit parameter were for the -ln L $\Gamma = -2.66^{+6\cdot10^{-2}}_{-5\cdot10^{-2}}$ and $\phi_0 = 1.025 \cdot 10^4 \pm 4.22\,\mathrm{TeV^{-1}m^{-2}s^{-1}sr^{-1}}$ and for the curve-fit function $\Gamma = -2.88 \pm 4.39 \cdot 10^{-2}$ and the $\phi_0 = 8.762 \cdot 10^3 \pm 3\,\mathrm{TeV^{-1}m^{-2}s^{-1}sr^{-1}}$. They were slightly increased compared to the original parameters $\Gamma = -2.60$ and $\phi_0 = 5.332 \cdot 10^3\,\mathrm{TeV^{-1}m^{-2}s^{-1}sr^{-1}}$. For the fit of the power-law variables the choice of the normalization energy $E_0 = 55\,\mathrm{TeV}$ was important. If E_0 was not within the *bias* $\pm$ 10 % the -ln L did contain several minima which led to a false determination of the fit values despite the usage of the basin hopping minimizer because diverse combinations of the fit parameter led to a supposed good fit result. In addition, the fit error calculation did not match to the fit results. It can be concluded that the choice of E_0 affects the fit result and should be determined within the considered energy range.
Further a dataset containing real data compositions, taken from [68], of the six classes was tested. This was done in order to see if the classifier and energy estimator can handle unequal distributions and if it is possible to reconstruct the Fe^+ spectrum. The energy reconstruction of these events had a unique shape containing a plateau. This is due to the unpredictable behavior of the energy estimator on the background classes. Events of the classes Gamma and Proton have the highest dispersion in their PDFs created by using the energy estimator trained on Fe^+ events. In addition, the total number of H^+ and He^+ events are the greatest in the dataset which means these events have the highest impact on the energy reconstruction and cause this unique shape. After the classification of the energy reconstructed events the plateau in the measurement was compensated by the classifier. This supports the statement that the unique shape of the energy reconstructed events were caused by background events. Both fit results overesti-

mated the flux normalization. Whereas the spectral index was well reconstructed by the -ln L. In contrast, the curve-fit function estimated a higher value for Γ. One possible reason could be that the negative log-likelihood function takes not only the absolute difference into account which is done by the least square method χ^2 used in the curve-fit function.
However, it can be concluded that the formula for the background estimation is a useful approximation and can be used in combination with the formula for the expected signal events to determine the spectral index of the signal events and normalization flux ϕ_0. Therefore this approach can be applied on real data events.

The application of the energy estimator trained on Fe^+ events on the real dataset containing PKS 2155-304 runs resulted in a similar distribution as the MC dataset created with spectral indexes of cosmic radiation. Both distribution had a plateau within the interval 49 TeV to 59 TeV. In addition, the saddle points are at the same energy of 59 TeV. Afterwards the multi-classifier CNN_α was used to group the events into six particle-like classes which where adapted from [10] [68]. The result showed a suspicious distribution within the energy region 29 TeV to 49 TeV where it stays nearly constant. In comparison to the classified events of the MC dataset the multi-classifier applied on real data did not result in a similar distribution of classified PKS 2155-304 events. Therefore it can be concluded that the CNN_α did eventually learn features which appear in MC images but not in real data observations. This supports the hypothesis of a MC real data discrepancy as mentioned in [58]. However, the energy estimator distributions looked similar in their course which weakens the claim of a MC real data discrepancy. Comparing the total number of predicted Fe^+ events which was 642120 to the overall predicted events for each class it can be seen that the percentage of the predicted Fe^+ events is not in agreement with the expected dominant components of the heavier nuclei. The dominant elements are hydrogen (89 %) and helium (10 %) [47]. In addition, the classifier was not trained on a sufficient number of events which also led to higher FP rate. More events, applying more nodes or layers to the architecture as well as using the CNN_{SK} could improve the accuracy. Further improvements could be achieved by a higher size cut on the real data images to sort out events having low intensities and eventually only a few triggered pixels. Dark images implicate a low initial particle energy and the DNN were only trained on MC events within the energy interval of 5 TeV to 100 TeV. An other approach which could be tried is to take only events for the analysis which got a ζ-value higher than e.g. 0.7. The classifier then is sure in its prediction that this event was induced by for example an iron nuclei.
The response of the DNNs was provided to the Forward Folding algorithm. The fit results were not in agreement with the assumed spectral index of $\Gamma = 2.59$ and flux normalization $\phi_0 = 7.39 \cdot 10^{-7}\,\text{TeV}^{-1}\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}$ taken from [10] and adapted to the normalization energy 55 TeV. The determined spectral index of [68] is $\Gamma = 2.62 \pm 0.03$ is similar. In contrast, the fitted response of the DNN approach resulted in $\Gamma = 1.68^{+2\cdot10^{-2}}_{-0}$ for the -ln L and $\Gamma = 4.70 \pm 3.73 \cdot 10^{-2}$ using the curve-fit function. These differ strongly to the previous predicted spectral indexes of Fe^+-like hadronic particles. Whereby the

cause is the worse classification of the real data events. The flux normalizations are $\phi_0 = 1.879 \cdot 10^{-6}{}^{+1 \cdot 10^{-10}}_{-9 \cdot 10^{-11}}\,\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ and $\phi_0 = 2.082 \cdot 10^{-8} \pm 9 \cdot 10^{-12}\,\mathrm{TeV}^{-1}\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1}$ for the -ln L and χ^2 fit at the normalization energy $E_0 = 55\,\mathrm{TeV}$. The statistic errors done by the CNN_α and the energy estimator were considered by the Forward Folding algorithm. Whereby, the systematic errors of e.g. the effective area, the PDF, the opacity of the atmosphere, the hadronic interaction processes and the PMT's quantum efficiency [10] were neglected for this work but they need to be taken into account. The error of the effective area using the CNN_α might have a bigger influence than in [10].

To summarize, the response of the DNN used to reconstruct the energy of PKS 2155-304 data was similar to the MC dataset containing real data distributions. Whereas the Deep Learning approach to distinguish heavier nuclei resulted in an improvable outcome. If this task can be improved by e.g. either using additional cut parameters or solving the MC real data discrepancy the Deep learning approach can be used as a further analysis tool and therefore it will be possible to reconstruct an iron spectrum of real data. In the next chapter a more detailed conclusion and an outlook is done.

7. Conclusion

The goals of this thesis were to classify CRs and reconstruct their energy as well as to create an iron energy spectrum. Therefore the DNNs were trained on MC H.E.S.S. I simulations due to the known labels. In addition, the Forward Folding algorithm was implemented and expanded to determine the pure power-law of hadronic particles with hadronic and γ-ray background.

In this thesis DNNs were used to distinguish heavier nuclei EAS images. The investigated NN architectures consist of 4 convolutional and 5 fully connected layer. These achieved good results in the binary classification task especially for the γ-ray vs. H^+ and H^+ vs. heavier nuclei events. For the training of the DNN events fulfilling the standard cuts and having an energy within the range 5 TeV to 100 TeV were used. The same architecture was applied on the multi- and multi-task classification. The outcome showed that it is possible to distinguish MC EAS images of heavier nuclei using DNN. The highest accuracy achieved by CNN_{SK} was 57.76 % due to time limitations only the CNN_{α} having an accuracy of 56.68 % was used for further analysis steps. A guessing of any class would correspond to a recall of 16 % which was not the case for any of the primary particle types. It needs to be mentioned that the DNN for the multi and multi-classification task were trained on only 96000 events. For more representative results more trainings examples are needed which could also lead to an improvement of the performance. To investigate the number of nodes in each layer iterative pruning [33] should be implemented. It can prevent over-fitting if too much nodes are used or it can improve the performance using more nodes. An additional step to increase the accuracy of the DNN could be to train the classifier momentarily on only one class. During this training phase it focuses on features which describe this class best. The same procedure can be repeated multiple times. In the meantime training on all classes could support the NN to focus on unique features of each class.
The approach of using DNN facilitate the classification and energy reconstruction of hadronic particles which is not possible with the standard analysis tools because these focus on the background rejection and on the geometrical information in a γ-ray EAS image. The energy estimators were trained on events in the energy range 5 TeV to 100 TeV and achieve a *bias* interval of about 2/3 of the total energy region on the test set. On γ-rays Hillas reconstruction and ImPACT have within 5 TeV to 100 TeV a *bias* smaller than $\pm$ 10 % [53] and in the energy interval 5 TeV to 60 TeV a *bias* near to zero [53] considering H.E.S.S. I. The energy estimators are tools for the energy reconstruction of hadronic particles. In conclusion, these can be used as an optional method for the energy reconstruction of γ-ray EAS because the results were promising and could be improved by using more train examples and study the events in specific energy ranges

to define the bin size which defines the energy distribution of the events.
Considering the direction reconstruction of cosmic radiation, in [58] a direction reconstruction of γ-ray EAS images using DNN was done. The direction reconstruction of hadronic particles is only useful on showers having a similar elliptical shape as γ-ray EAS and/or having energies in the order of $> 10\,\mathrm{EeV}$ [25]. Therefore the Hillas reconstruction, ImPACT or a DNN approach can be used because at these ultra high energies the hadronic particle was not distracted from any magnetic field in space [25].
For the creation of an iron spectrum the results of the CNN_α and energy estimator are needed. Due to erroneous outcomes of these two tools the Forward Folding algorithm was implemented and applied. The reconstruction of the MC iron energy spectrum using a dataset with signal and background events yielded in good results. The investigated formula for the expected background events in the whole FoV of $5\,^\circ$ contains information about the efficiency and PDF of the background events corresponding to the signal class. These are more exact with higher accuracies of the CNN_α and energy estimator. For the calculation of the reconstructed flux the mean energy must be defined by using the formula 3.4 which was originally defined in [60]. The reason is the unequal distribution of events within one energy bin due to the given gradient defined by the spectral index. Overall it can be concluded that the extended Forward Folding algorithm works well on MC events.
The application of real data PKS 2155-304 runs on the energy estimator resulted in a similar distribution as the real data like MC dataset. Whereas, the further application on CNN_α resulted in an unique distribution of the events. This behavior was not seen on the real data like MC dataset. Therefore the MC real data discrepancy could be a possible reason which was also seen in [58]. Due to the poor separation of the classifier the fit parameters did not match the ones determined by satellite experiments and balloon flights [68] [10]. In addition, the formula used for the background estimation presupposes that the spectra of the background is well defined. Further, only runs containing a weak or no source should be taken due to the high dispersion in the PDF of γ-rays which were reconstructed using the energy estimator trained on iron nuclei events. It can be concluded that for now it is not possible to reconstruct an iron spectrum using the Deep Learning approach for the classification of real data.

In summary, in this thesis a multi-classifier should be developed to distinguish heavier nuclei. The investigated DNNs for the classification task achieved similar accuracies of about 56 %. However, it was shown that it is possible to classify cosmic radiation using the Deep Learning approach. Further an energy estimator should be developed to reconstruct the energy of hadronic particles in order to create an iron energy spectrum because the standard analysis tools can only reconstruct gamma-like EAS. The energy $bias \pm 10\,\%$ resulted in a total range of 60 TeV for iron nuclei. Using the DNN on γ-rays it could compete with ImPACT if it is trained on a greater energy range. The reconstruction of a MC iron spectrum using the investigated formula for the background estimation in the Forward Folding algorithm resulted in similar pure power-law variables to the origin ones. An iron nuclei energy spectrum of real data was done but the fit parameter results were different to the assumed ones of [68] and [10].

Bibliography

[1] An open source machine learning framework for everyone. https://www.tensorflow.org/. Last accessed: 16.02.2019.

[2] Largest ever Cherenkov telescope sees first light. https://www.mpi-hd.mpg.de/hfm/HESS/pages/press/2012/HESS_II_first_light/. Last accessed: 07.06.2018.

[3] *scipy.optimize.curve_fit*. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html. Last accessed: 09.03.2019.

[4] H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, jun 2010.

[5] F. Aharonian, A. Akhperjanian, A. Bazer-Bachi, B. Behera, M. Beilicke, W. Benbow, D. Berge, K. Bernlöhr, C. Boisson, O. Bolz, et al. An exceptional very high energy gamma-ray flare of pks 2155–304. *The Astrophysical Journal Letters*, 664(2):L71, 2007.

[6] F. Aharonian et al. First ground-based measurement of atmospheric cherenkov light from cosmic rays. *Physical Review D*, 75(4), feb 2007.

[7] M. P. Alessandro De Angelis. *Introduction to Particle and Astroparticle Physics*. 2192-4791. Springer, EditionSecond edition, 2018.

[8] Alex Williams. Everything you did and didn't know about PCA. http://alexhwilliams.info/itsneuronalblog/2016/03/27/pca/, Mar. 2016. Last accessed: 19.01.2019.

[9] H. Allamy. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). 12 2014.

[10] A. Archer, W. Benbow, R. Bird, R. Brose, M. Buchovecky, V. Bugaev, M. Connolly, W. Cui, M. Daniel, A. Falcone, et al. Measurement of the iron spectrum in cosmic rays by veritas. *Physical Review D*, 98(2):022009, 2018.

[11] B. E. Bejnordi, M. Veta, and e. a. Paul Johannes van Diest. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22):2199, dec 2017.

[12] M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars. *Computational geometry: algorithms and applications*. Springer-Verlag TELOS, 2008.

[13] D. Berge, S. Funk, and J. Hinton. Background modelling in very-high-energy γ-ray?astronomy. *Astronomy & Astrophysics*, 466(3):1219–1229, apr 2007.

[14] K. Bernlöhr. Cosmic-ray air showers. `https://www.mpi-hd.mpg.de/hfm/CosmicRay/Showers.html`. Last accessed: 17.05.2018.

[15] K. Bernlöhr. Simulation of imaging atmospheric cherenkov telescopes with corsika and sim_telarray. *Astroparticle Physics*, 30(3):149–158, 2008.

[16] K. Bernlöhr, O. Carrol, R. Cornils, S. Elfahem, P. Espigat, S. Gillessen, G. Heinzelmann, G. Hermann, W. Hofmann, D. Horns, et al. The optical system of the hess imaging atmospheric cherenkov telescopes. part i: layout and components of the system. *Astroparticle Physics*, 20(2):111–128, 2003.

[17] M. Buscema. Back propagation neural networks. *Substance use & misuse*, 33(2):233–270, 1998.

[18] M. Bustamante et al. High-energy cosmic-ray acceleration. 2010.

[19] D. Casadei. Direct measurement of galactic cosmic ray fluxes with the orbital detector ams-02. 2003.

[20] A. D. D. Rebuzzi, A. Rimoldi. High Energy Electromagnetic Shower Development in Air with Geant4. *ALT-PHYS-2001-010*, 2001.

[21] Dang Ha The Hien. A giude to receptive field arithmetic for Convolutional Neural Networks. `https://medium.com/mlreview/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks-e0f514068807`, Apr. 2017. Last accessed: 16.02.2019.

[22] T. Dozat. Incorporating nesterov momentum into adam. 2016.

[23] R. Engel, D. Heck, and T. Pierog. Extensive Air Showers and Hadronic Interactions at High Energy. *Annual Review of Nuclear and Particle Science*, 61(1):467–489, nov 2011.

[24] A. A. et al. VHEΓ-ray emission of PKS 2155–304: spectral and temporal variability. *Astronomy and Astrophysics*, 520:A83, sep 2010.

[25] J. A. et al. Correlation of the highest-energy cosmic rays with nearby extragalactic objects. *Science*, 318(5852):938–943, nov 2007.

[26] E. Fermi. On the origin of the cosmic radiation. *Phys. Rev.*, 75:1169–1174, Apr 1949.

[27] S. Funk. *A new population of very high-energy gamma-ray sources detected with H.E.S.S. in the inner part of the Milky Way.* PhD thesis, 2005.

[28] S. Funk. Space- and Ground-Based Gamma-Ray Astrophysics. *Annual Review of Nuclear and Particle Science*, 65:245–277, Oct. 2015.

[29] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[30] D. Groom et al. Cosmic Rays. *European Physical Journal*, C(15):1, 2000.

[31] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *null*, pages 1735–1742. IEEE, 2006.

[32] E. Halkiadakis. www.physics.rutgers.edu/~evahal/talks/tasi09/TASI_day3_school.pdf, 2009. Last accessed: 17.05.2018.

[33] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 1135–1143. Curran Associates, Inc., 2015.

[34] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[35] V. F. Hess. Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Physikikalische Zeitschrift*, 13:1084–1091, 1912.

[36] Hieu Pham. When should you use corss entropy loss and why ? https://www.quora.com/When-should-you-use-cross-entropy-loss-and-why, Sept. 2018. Last accessed: 23.02.2019.

[37] J. Hinton and W. Hofmann. Teraelectronvolt astronomy. *Annual Review of Astronomy and Astrophysics*, 47:523–565, 2009.

[38] W. Hofmann. The H.E.S.S. Telescopes. https://www.mpi-hd.mpg.de/hfm/HESS/pages/about/HESS_I_II/, July 2012. Last accessed: 23.02.2019.

[39] J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1875–1882, 2014.

[40] Y. Huang, S. Liu, J. Hu, and W. Deng. Metric-promoted siamese network for gender classification. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 961–966. IEEE, 2017.

[41] C. V. Jahn. *Systematic Studies of an On-O-Analysis of H.E.S.S. data in Search for a Signal from the Dark Matter Halo of the Milky Way*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2013.

[42] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[43] G. Koch, R. Zemel, and R. Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

[44] F. Krennrich. Gamma ray astronomy with atmospheric cherenkov telescopes: the future. *New Journal of Physics*, 11(11):115008, nov 2009.

[45] J. Matthews. A heitler model of extensive air showers. *Astroparticle Physics*, 22(5-6):387–397, 2005.

[46] J. McCarthy. Arthur Samuel: Pioneer in Machine Learning. `http://infolab.stanford.edu/pub/voy/museum/samuel.html`. Last accessed: 10.05.2018.

[47] R. Mewaldt. Cosmic Rays. *Macmillan Encyclopedia ofPhysics*, 1996.

[48] J. J. Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[49] H. Netzer. *The Physics and Evolution of Active Galactic Nuclei*. Cambridge University Press, 2009.

[50] Y. Ohira. Cosmic ray acceleration mechanism. In *Proceedings of the 7th International Workshop on Very High Energy Particle Astronomy in 2014 (VHEPA2014)*, page 011002, 2017.

[51] Oinkina. Understanding LSTM Networks. `http://colah.github.io/posts/2015-08-Understanding-LSTMs/`, Aug. 2015. Last accessed: 23.02.2019.

[52] B. Olson, I. Hashmi, K. Molloy, and A. Shehu. Basin hopping as a general and versatile optimization framework for the characterization of biological macromolecules. *Advances in Artificial Intelligence*, 2012:3, 2012.

[53] R. Parsons and J. Hinton. A monte carlo template based analysis for air-cherenkov arrays. *Astroparticle Physics*, 56:26–34, 2014.

[54] C. Patrignani et al. Review of Particle Physics, 2016-2017. *Chin. Phys. C*, 40(10):100001. 1808 p, 2016.

[55] P. Reynolds, C. Akerlof, M. Cawley, M. Chantell, D. Fegan, A. Hillas, R. Lamb, M. Lang, M. Lawrence, D. Lewis, et al. Survey of candidate gamma-ray sources at tev energies using a high-resolution cerenkov imaging system-1988-1991. *The Astrophysical Journal*, 404:206–218, 1993.

[56] A. Sallab, M. Abdou, E. Perot, and S. Yogamani. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, 2017(19):70–76, jan 2017.

[57] T. D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural networks*, 2(6):459–473, 1989.

[58] I. Shilon, M. Kraus, M. Büchele, K. Egberts, T. Fischer, T. Holch, T. Lohse, U. Schwanke, C. Steppa, and S. Funk. Application of deep learning methods to analysis of imaging atmospheric cherenkov telescopes data. *Astroparticle Physics*, 105:44–53, feb 2019.

[59] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[60] Stefan Hoppe. *Emitters of VHE gamma-radiation as revealed by the H.E.S.S. Galactic plane survey.* PhD thesis, Ruperto-Carola University of Heidelberg, 2008.

[61] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton. On the importance of initialization and momentum in deep learning. *ICML (3)*, 28(1139-1147):5, 2013.

[62] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, et al. Review of particle physics. *Physical Review D*, 98(3):030001, 2018.

[63] A. Tharwat. Classification assessment methods. *Applied Computing and Informatics*, aug 2018.

[64] Tobias Fischer. Convolutional Neural Networks for H.E.S.S. Master's thesis, Friedrich-Alexander-Universität Erlangen-Nuernberg, 2018.

[65] M. Walter. Victor Hess und die Entdeckung der kosmischen Strahlung. `https://www.weltderphysik.de/gebiet/universum/kosmische-strahlung/entdeckung-der-kosmischen-strahlung/`, June 2012. Last accessed: 08.05.2018.

[66] T. C. Weekes, M. Cawley, D. Fegan, K. Gibbs, A. Hillas, P. Kowk, R. Lamb, D. Lewis, D. Macomb, N. Porter, et al. Observation of tev gamma rays from the crab nebula using the atmospheric cerenkov imaging technique. *The Astrophysical Journal*, 342:379–395, 1989.

[67] B. Wiebel. Chemical Composition in High Energy Cosmic Rays. *Bergische Universität-Gesamthochschule Wuppertal*, Apr. 1994.

[68] B. Wiebel-Sooth, P. L. Biermann, and H. Meyer. Cosmic rays vii. individual element spectra: prediction and data. *arXiv preprint astro-ph/9709253*, 1997.

[69] I. Wingerter-Seez. Particle Physics Instrumentation. *CERN Yellow Reports*, 2:295–314, 2017.

A. NN Metric

The metric gives information about the current performance of a classifier. The most common used metrics are the accuracy, precision, recall and specificity. In the case of an uneven distribution of training examples or focusing on the precision and recall the F-measure is used. The formulas were taken from [63].
The accuracy of a NN is calculated by

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

where the denominator is the total number of all events the classifier was trained on. The numerator is the number of events which were correctly predicted by the trained NN.
The precision of a NN is given by the ratio of TP to the sum over TP and FP considering one class

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

The recall describes the true positive rate of a classifier. In the binary case the recall describes the signal class but only applied on the problem explained in this thesis. The value can be determined by the following formula

$$Recall = \frac{TP}{TP + FN}$$

Specificity is the true negative ratio of the trained NN. It describes the "recall" of the background events in the binary classification.

$$Specificity = \frac{TN}{TN + FP}$$

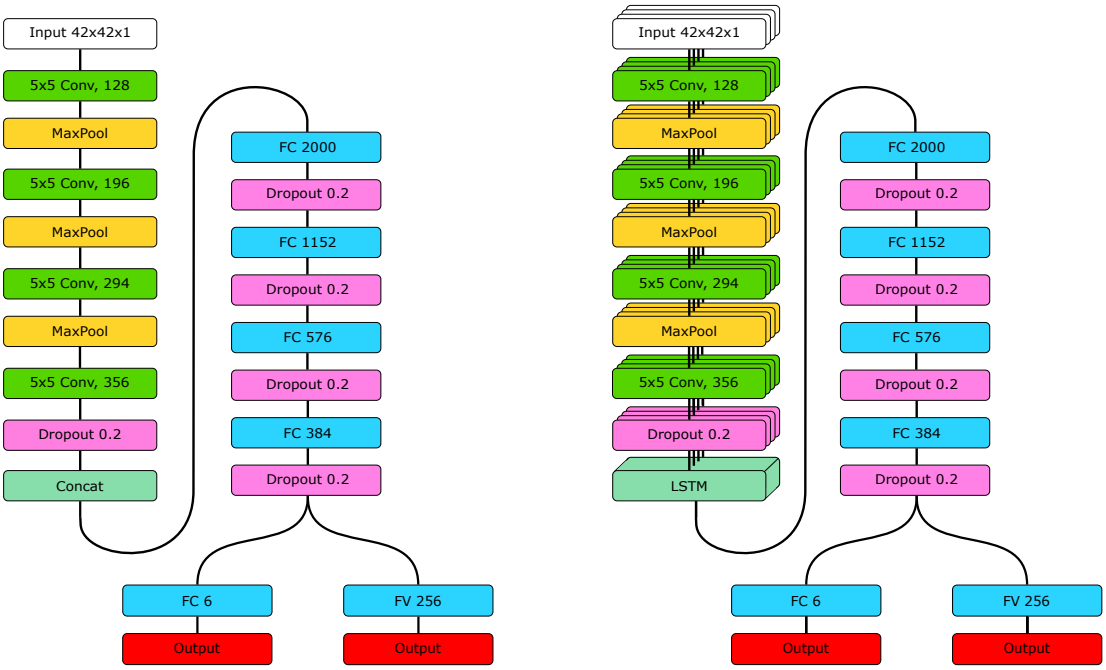
F_β-Score is important for a dataset which contains an uneven event distribution but also if the focus is on the precision and recall of a classifier. It is given by

$$F_\beta - Score = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{\beta^2 Precision + Recall} \ ,$$

where β is the weight value. In the case of $\beta = 1$ precision and recall are both equally taken into account. For this thesis the β value was set to 1, which is also known as F_1-Score.

B. Multi- and Multi-task Classification

B.1. Multi-Classification Architecture



(a) $SCNN_{\alpha}$ **(b)** $SCNN$

Figure B.1. – *Illustration of Siamese-classification networks having the a similar structure as the ones used for the multi-classification task except the last fully connected layer which is divided into a classification output and a feature vector output. Siamese networks consist of two identical towers. In this figure only one tower of each architecture is shown. The convolutional layers are marked with the color green. They are initialized with a kernel size* 5×5 *having* 128 *nodes. The maximum pooling layers have stride* 2 *and use padding same. After the convolutional layers a dropout of 0.2 is inserted after each layer to prevent over-fitting. a) The channel approach convolves over the stack of CT1-4 images treated as one RBG*α *image in the first layer. b) Images are provided one after another to the network sharing their weights. They get concatenated after the first dropout and convolution layers.*

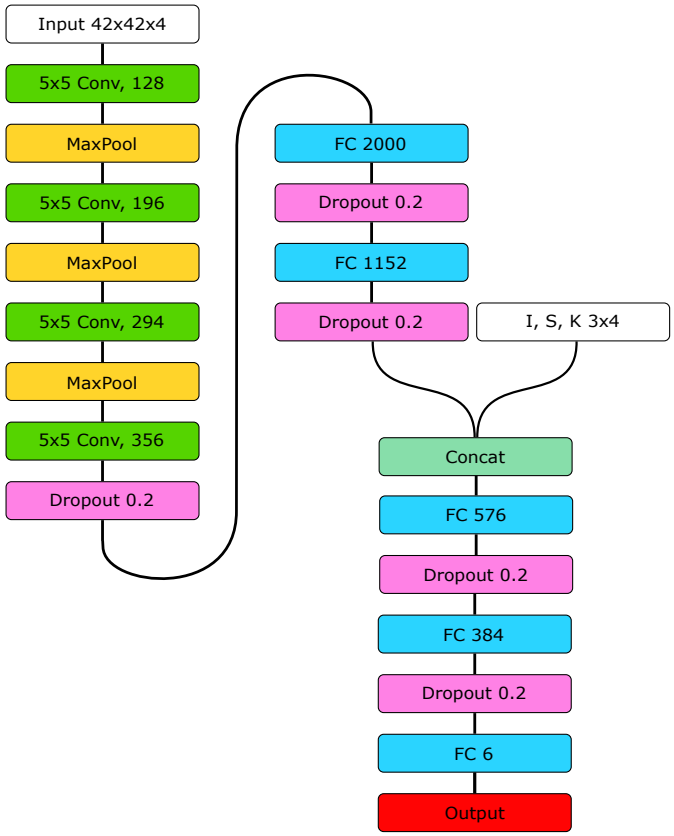


Figure B.2. – *Schematic structure of the multi-classifier CNN_{SK} using the channel approach. The input images consist of 42 × 42 pixels. The convolutional layers are initialized with a kernel size of 5×5. The number of nodes is e.g. in the first layer* 128*. Between the convolutional layers a maximum pooling layer is inserted. After the convolutional part a dropout of 0.2 is applied. The fully connected tower is divided into two parts. After the second fully connected layer and third dropout of 0.2 the size, skewness and kurtosis of the CT1-4 images are concatenated to the feature vector. The last two fully connected layers contain information about the image features and the additional parameters of the EAS per telescope. The output is a ζ-value for each class.*

The hyperparameters used in the multi-classification task are shown in table B.1. The X means that for this NN task these hyperparameter is not needed.

Table B.1. – *Used hyperparameter in multi-classification.*

Hyperparameter	**CNN$_{SK}$**	**CNN$_{\alpha}$**	**CNN**	**SCNN$_{\alpha}$**	**SCNN**
β	0.995	0.995	0.995	0.999	0.995
No. epochs per decay	3	3	3	7	6
α decay factor	0.5	0.5	0.5	0.5	0.5
α_{init}	0.001	0.0005	0.0005	0.0008	0.0008
staircase	True	False	False	True	True
m	X	X	X	1.0	1.7
λ_s	X	X	X	4	3

B.2. CNN$_{\alpha}$ Results

The results of the CNN$_{\alpha}$ with reference to the rows of the confusion matrix 5.12 are shown in the figures B.3, B.4 and B.5. The rows are plotted energy resolved to investigate the classifiers decision for each energy bin. The overall distribution of events looks similar in each figure. One interesting characteristic is that the FN of class Gamma are only labeled as class Proton and also with increasing charge Z the number of FP per class increases.

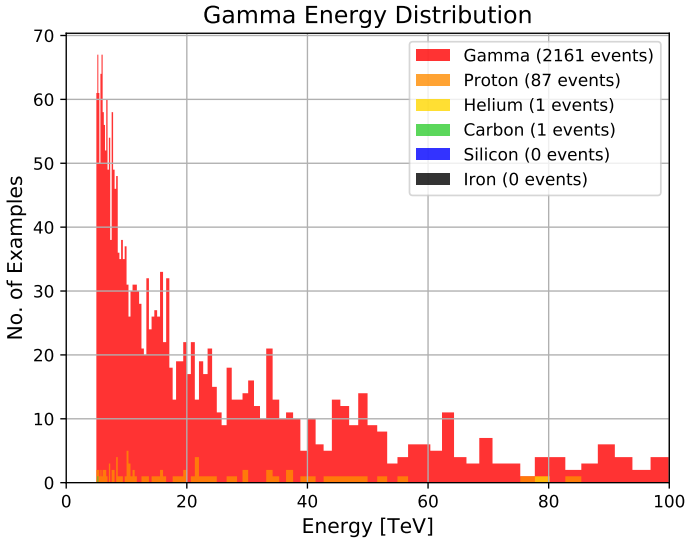
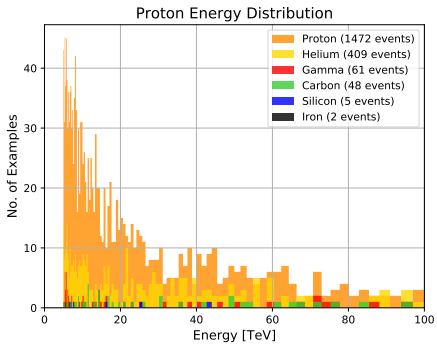
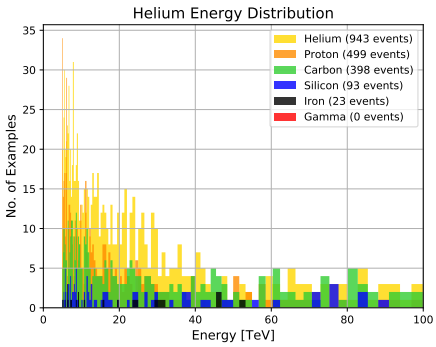


Figure B.3. – *Energy distribution of all as class Gamma labeled events. The classifier has* 2161 *events of class Gamma predicted as Gamma.* 87 H^+ *events are wrongly predicted as γ-ray EAS shower event. Two heavier nuclei are FP.*

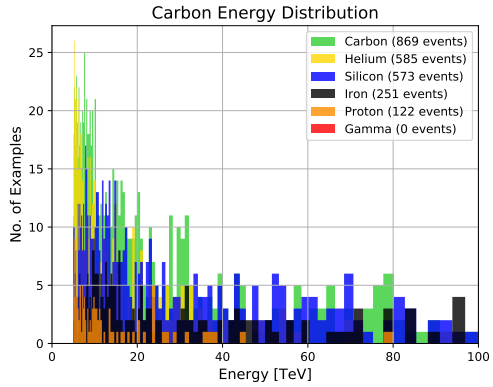


(a) *Proton*

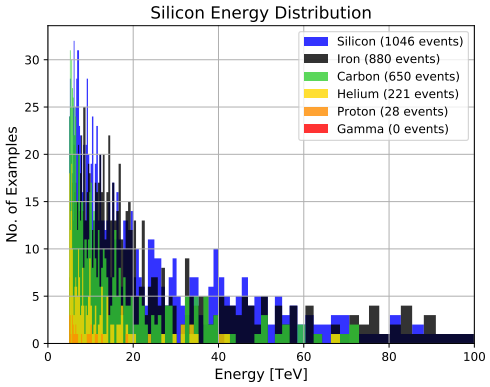


(b) *Helium*

Figure B.4. – *The energy distribution of the Proton or Helium. a) The FP of the class Proton decrease with increasing energy. The ranking of the FP is lead by Helium with in total 409 events misclassified as Proton. The number of FP of the other classes in sum is smaller than the number of FP of class Helium. Overall events of class Proton are labeled more often as H^+ than FP H^+ of the other classes. b) Here the total number of FP exceeds the number of TP. However, in nearly every energy bin more events were correctly predicted as class Helium.*



(a) *Carbon*



(b) *Silicon*

Figure B.5. – *The energy distribution of the classes Carbon and Silicon. a) The number of TPs of class Carbon exceeds in the energy region 28-32 TeV and 72-80 TeV the number of FPs. Thus, in the most energy bins the number of FPs is greater. The amount of FP events of the classes Helium and Silicon is the greatest. b) The energy distribution of the TP of class Silicon is sustained for each energy bin. The most FP are of class Iron with in total 880 events followed by class Carbon. The other classes are the minority.*

B.3. Multi- and Multi-task Classifier Results

Table B.2. – *Performance results of CNN_{SK} reaching a total accuracy of 57.76 %.*

Class	Precision	Recall	F_1-Score
γ	96.33 %	98.02 %	97.17 %
H^+	73.52 %	70.48 %	71.97 %
He^+	49.39 %	43.83 %	46.45 %
C^+	38.29 %	38.21 %	38.25 %
Si^+	37.36 %	40.37 %	38.81 %
Fe^+	52.55 %	55.61 %	54.04 %

Table B.3. – *Performance results of CNN reaching a total accuracy of 55.87 %.*

Class	Precision	Recall	F_1-Score
γ	97.14 %	97.66 %	97.40 %
H^+	71.65 %	67.33 %	69.42 %
He^+	45.52 %	42.53 %	43.97 %
C^+	38.12 %	40.77 %	39.40 %
Si^+	35.25 %	42.21 %	38.42 %
Fe^+	52.37 %	44.69 %	48.23 %

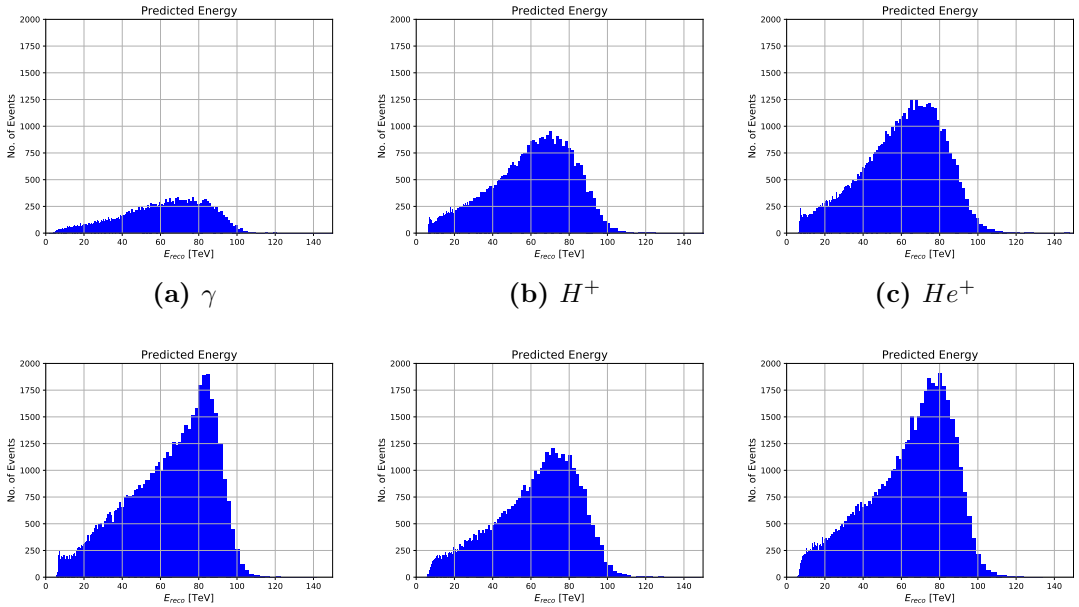
Table B.4. – *Performance results of $SCNN_{\alpha}$ reaching a total accuracy of 54.22 %.*

Class	Precision	Recall	F_1-Score
γ	96.42 %	95.86 %	96.14 %
H^+	69.77 %	63.68 %	66.59 %
He^+	41.67 %	48.87 %	44.99 %
C^+	35.04 %	41.85 %	38.15 %
Si^+	36.52 %	35.73 %	36.12 %
Fe^+	52.62 %	39.33 %	45.02 %

Table B.5. – *Performance results of SCNN reaching a total accuracy of 55.43 %.*

Class	**Precision**	**Recall**	**F_1-Score**
γ	97.15 %	96.76 %	96.96 %
H^+	71.68 %	65.17 %	68.27 %
He^+	45.56 %	45.23 %	45.39 %
C^+	37.87 %	42.35 %	39.98 %
Si^+	35.57 %	45.18 %	39.80 %
Fe^+	53.12 %	37.89 %	44.23 %

C. Energy reconstruction



(a) γ **(b)** H^+ **(c)** He^+

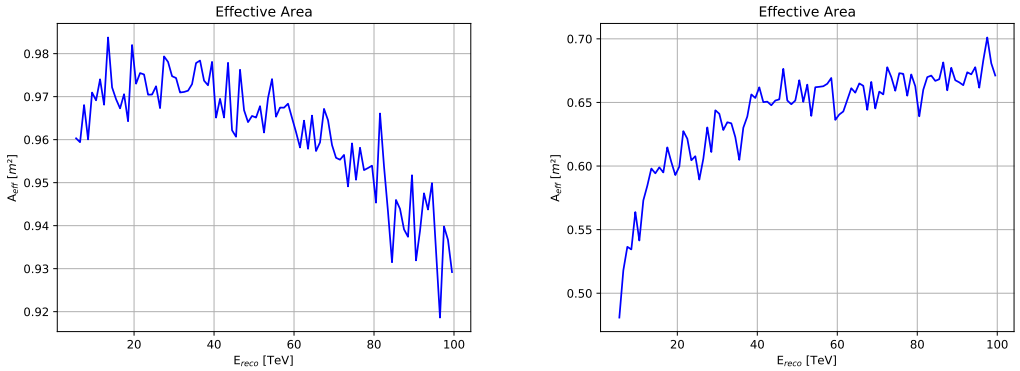
(d) C^+ **(e)** Si^+ **(f)** Fe^+

Figure C.1. – *Reconstructed energy resolved histograms of all classes. The total number of γ-ray, H^+ and Si^+ events is 26575, 50492 and 50800. All other energy estimators are tested on 66437 examples. One common fact is the decreasing predicted number of events in the energy range of about 90 TeV to 100 TeV. The energy ranges were adjusted to the reconstructed energy interval of the Fe^+ events.*

D. Forward Folding

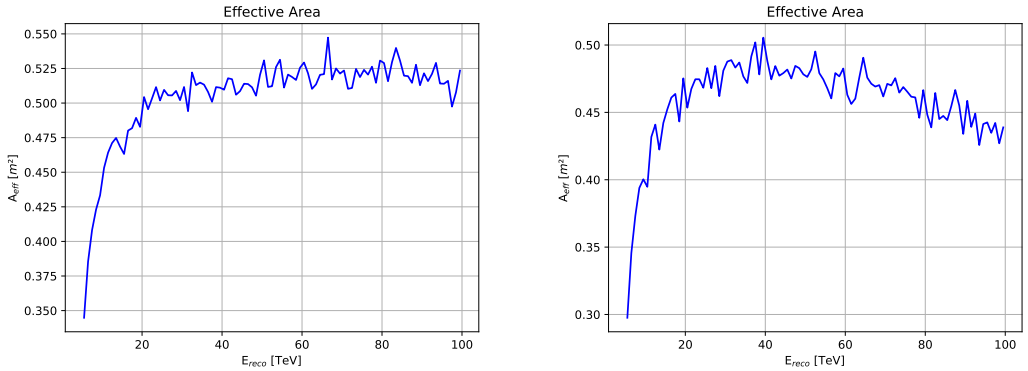
The Forward Folding algorithm was used to reconstruct the iron energy spectrum of MCs and real data. As mentioned in the introduction, it is possible to do this for all other classes the multi-classifier CNN_α was trained on. Hence, the effective area for all classes are shown.

D.1. Effective Areas



(a) γ **(b)** H^+

Figure D.1. – *Effective area of γ-ray and H^+ nuclei MC events. The effective area of class Gamma is higher than the one of class Proton in each energy bin. a) The effective area of the class Gamma is above 90 % over the whole reconstructed energy range. For higher energies it decreases from $\approx$ 97 % to $\approx$ 92 %. An effective area near to one is desired. b) The effective area of the class Proton is under 50 % at 5 TeV and increases rapidly to 65 %. Afterwards it stays almost constant.*



(a) He^+ **(b)** C^+

Figure D.2. – *The effective areas of the classes Helium and Carbon. Both show a similar course. In the energy region of 5 TeV to 20 TeV the effective area increases rapidly. Then it stays nearly constant with relatively small fluctuations compared to the rapid increase in the first energy bins.*

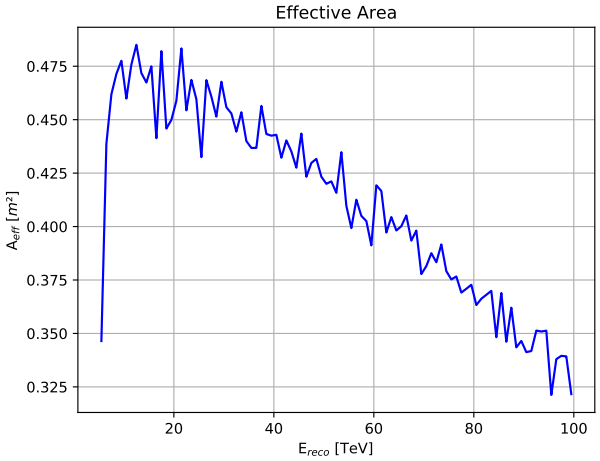
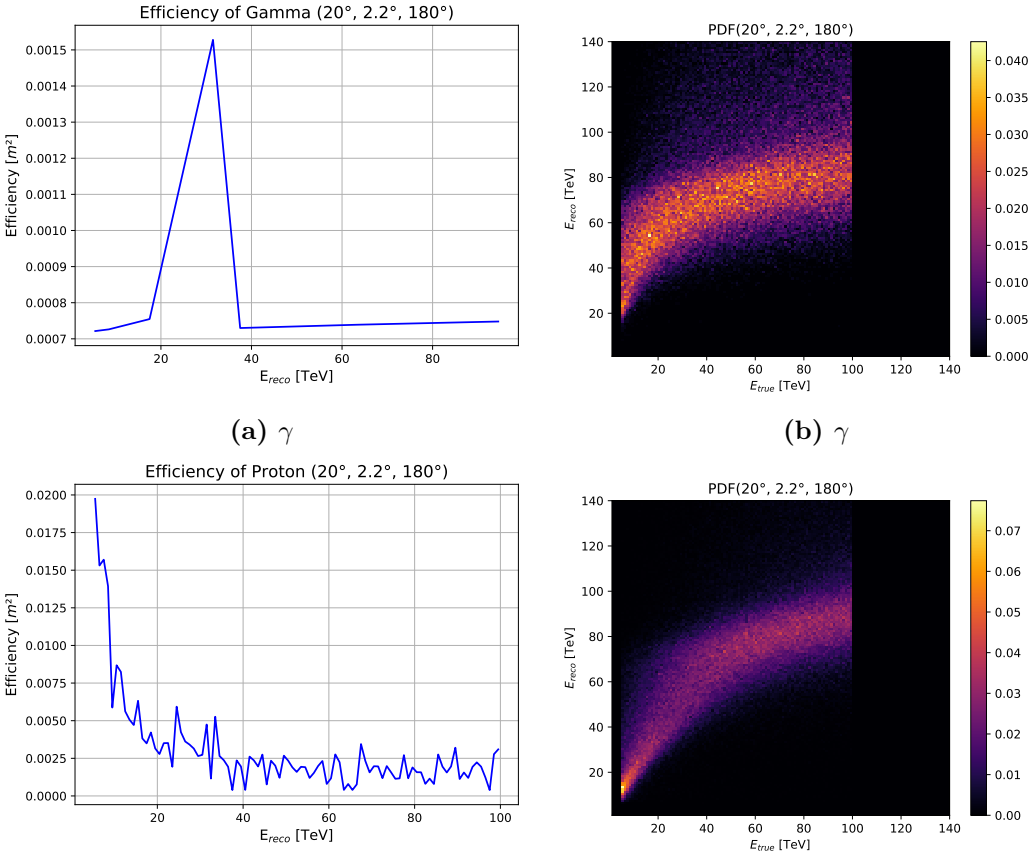


Figure D.3. – *Effective area of Si^+ nuclei events. Within the energy range of 5-13 TeV the effective area is increasing and reaches its maximum efficiency at $E_{reco} = 13$ TeV. After the peak it decreases steadily with fluctuations on top.*

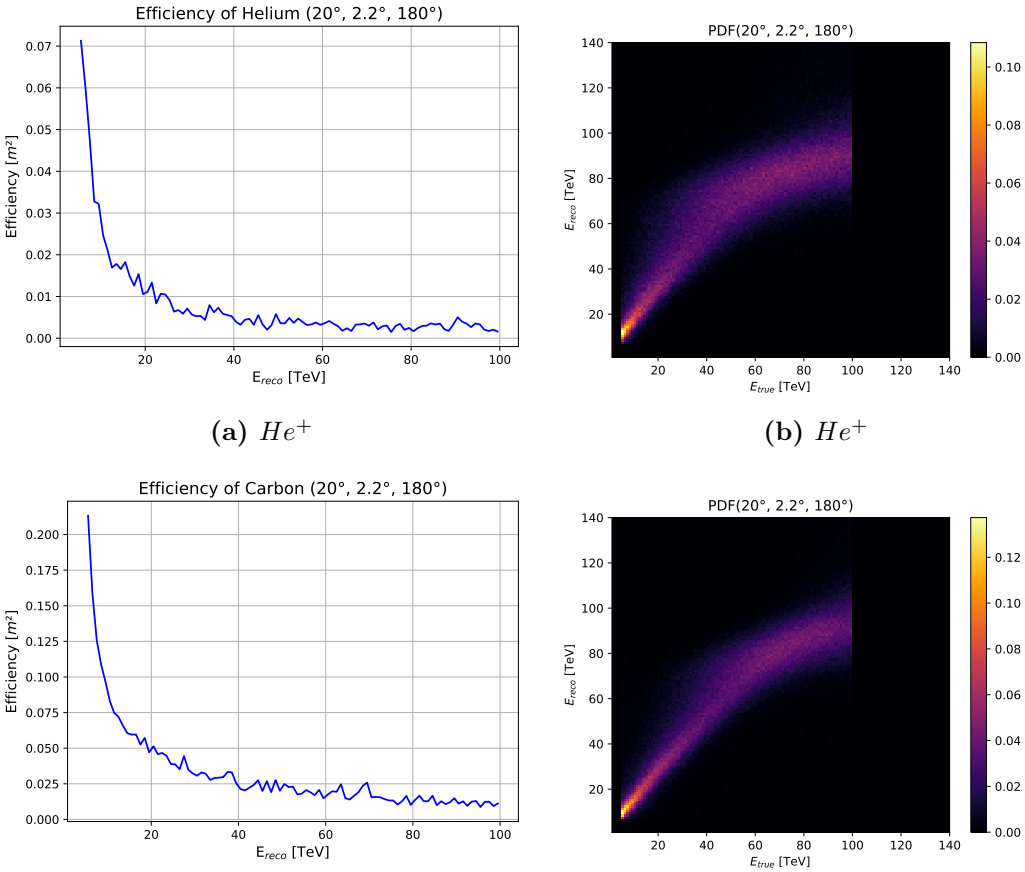
D.2. Efficiencies and PDFs of the Background Classes



(a) γ **(b)** γ

(c) H^+ **(d)** H^+

Figure D.4. – *The figures show the efficiency and PDFs of the background classes Gamma and Proton. a) The efficiency of γ-rays has a unique course. It has increased values up to $0.00152\,m^2$ in the energy interval of 18 TeV to 38 TeV. Overall the efficiency of the class Gamma on Iron is near to zero. b) The PDF of background γ-ray events entries containing probabilities are wide scattered in each energy bin. This means the energy estimator trained on Fe^+ events is almost guessing the energy of events from class Gamma. c) The efficiency class Proton on Iron is increased in the first energy bins. Afterwards it decreases and is fluctuating around the value $\sim 0.0017\,m^2$. d) The PDF of class Proton scatters wider in the lower energy bins and are more overestimated in their predicted value compared to the true energy. In the higher energy bins the events are more likely to be predicted correctly in comparison to the lower energies.*



(a) He^+ **(b)** He^+

(c) C^+ **(d)** C^+

Figure D.5. – *The figures show the efficiency and PDFs of the background classes Helium and Carbon. a) The efficiency of class Helium on Iron is relatively low. In the lower energy bins it has an increased value of 0.072 m^2. b) The PDF entries are more likely to be overestimated in the lower and middle part of the energy bins. c) The efficiency of class Carbon on Iron follows the same course as a). It has its highest peak at 5 TeV having a value higher than 0.2 m^2. d) The PDF of class Carbon has less dispersion compared to b). The uncertainty is higher in the energy region of 40 TeV to 60 TeV.*

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (inklusive Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Erlangen, den 18.03.2019

_____

Christina Hillig