

Untersuchung zur Differenzierung von Signalverläufen mit Hilfe von Neuronalen Netzen

Bachelorarbeit
zur Erlangung des akademischen Grades
Bachelor of Science (B.Sc.)
im Studiengang Physik

Alexander Doeker
alexander.doeker@student.uni-siegen.de
Matr. Nr. 1538218

16.08.2023

Universität Siegen
Naturwissenschaftlich Technische Fakultät
Department Physik

Inhaltsverzeichnis

1. Physikalische Grundlagen	4
1.1. Geschichte der kosmischen Strahlung	4
1.2. Kosmische Strahlung	4
1.2.1. Energiespektrum der primären kosmischen Strahlung	5
1.2.2. Ursprung der kosmischen Strahlung	6
1.2.3. GZK-Cutoff	7
1.2.4. Ausgedehnte Luftschauer (sekundäre kosmische Strahlung)	7
1.3. Cherenkov-Strahlung	10
2. Pierre Auger Observatorium	12
2.1. Geschichte	12
2.2. Aufbau des Pierre Auger Observatoriums	12
2.3. Fluoreszenzdetektor	13
2.4. Oberflächendetektor	15
2.4.1. Aufbau	15
3. Neuronale Netze (NN)	17
3.1. Geschichte der Neuronalen Netze	17
3.2. Allgemeiner Aufbau von Neuronalen Netzen	18
3.3. Arten von Neuronalen Netzen	21
4. Durchführung	25
4.1. Verwendeter Datensatz	25
4.2. Vorgehensweise	26
4.3. Aufbau des verwendeten NN	27
5. Auswertung	29
5.1. Optimierung des Neuronalen Netzes	29
5.1.1. Batch Größe	29
5.1.2. Anzahl Epochen	31
5.1.3. Validation Split	33
5.1.4. Kernel Größen	34
5.1.5. Anzahl an Filtern	35
5.1.6. Tiefe des NN	36
5.2. Grenzen des Neuronalen Netzes	37
6. Zusammenfassung & Ausblick	41
A. Anhang	42
A.1. Tabelle mit den Ergebnissen des ersten Teiles der Analyse	42
A.2. Tabelle mit den Ergebnissen des zweiten Teiles der Analyse	46
Literatur	48

Einführung

Das Pierre Auger Observatorium untersucht hochenergetische kosmische Strahlung. Dabei unterscheiden sich die ausgebildeten Schauer je nachdem was für ein Primärteilchen den Schauer induziert hat. Die einzelnen Schauerkomponenten erzeugen unterschiedliche Signalverläufe im Oberflächendetektor. In der nachfolgenden Arbeit soll untersucht werden, wie ähnlich zwei Signalverläufe sein dürfen, damit diese von einem Neuronalen Netz differenziert werden können. Als Grundlage dafür dienen generierte Landau Verteilungen, da diese den möglichen Signalverlauf in einem der Photomultipliern des Oberflächendetektors des Pierre Auger Observatoriums beschreiben. Dazu soll in einem ersten Schritt das verwendete Neuronale Netz optimiert werden. In einem zweiten Schritt danach werden die beiden generierten Datensätze schrittweise aneinander angeglichen, um zu untersuchen, wie klein die Unterschiede zwischen den beiden Verteilungen maximal sein dürfen.

1. Physikalische Grundlagen

1.1. Geschichte der kosmischen Strahlung

Die Erforschung der kosmischen Strahlung begann mit Victor Franz Hess, der 1912 bei Ballonflügen in einer Höhe von 5 km eine erhöhte Menge an ionisierter Strahlung nachweisen konnte (1). Diese wurde lange Zeit als Höhenstrahlung bezeichnet, da der Ursprung unbekannt war. Heute wissen wir, dass es sich um sekundäre kosmische Strahlung gehandelt hat. Für diese Entdeckung erhielt Hess 1936 den Nobelpreis. 1927 fotografierte Dimitry Skobelzyn kosmische Strahlung in einer Nebelkammer. 1928 sagte Paul Dirac Antimaterie als 2. Lösung der Wellengleichung voraus. Fünf Jahre später erbrachten Carl Anderson und Seth Neddermeyer den Nachweis für die Existenz von Antimaterie, in Form von Anti-Elektronen, Teilchen die in der Nebelkammer wie Elektronen aussehen, aber eine entgegengesetzte Ladung haben (2). Hierfür erhielt Carl Anderson 1936 den Nobelpreis zusammen mit Hess. Weitere fünf Jahre später entdeckten Anderson und Neddermeyer das Myon. Es ist geladen wie ein Elektron, hat aber die 200-fache Masse (3). Die Existenz von Höhenschauern geht auf Pierre Auger zurück, der diese 1938 in der Schweiz zusammen mit seinem Team nachwies. Kurz darauf, Anfang der 1940er Jahre konnte Marcel Schein beweisen, dass die primäre kosmische Strahlung zum Großteil aus Protonen besteht (4). Hierzu führte er eine Reihe von Ballonexperimenten in den obersten Schichten der Erdatmosphäre durch. Bis in die folgenden Jahre wurde kosmische Strahlung für verschieden Experimente verwendet, da dies die einzige Quelle an hochenergetischer Strahlung war, bis es in den 1950er Jahren Durchbrüche in der Entwicklung von Beschleunigern gab. 1949 postulierte Enrico Fermi einen möglichen Grund für die hohen Energien der primären kosmischen Strahlung: Die Beschleunigung von geladen Teilchen an kosmischen Plasmawolken mit geraden Schockfronten, z.B. wie sie bei Supernovae Explosionen entstehen (5). In den 1960er Jahren wurde das erste mal kosmische Strahlung mit einer Energie von mehr als 10^{20} eV nachgewiesen. Einige Jahre später postulierten die drei Physiker Kenneth Greisen, Vadem Kuzmin und Georgi Zatsepin, dass hoch energetische Protonen mit der kosmischen Mikrowellen-Hintergrundstrahlung wechselwirken und somit höchst energetische Protonen auf ca. $5 \cdot 10^{19}$ eV abgebremst werden (6). In den nächsten drei Jahrzehnten wurden verschiedene Experimente zur Untersuchung der kosmischen Strahlung gebaut, unter anderem das Haverah Park Experiment, welches aus 200 Wasser-Cherenkov-Detektoren bestand, oder das Fly's Eye Experiment in den USA, welches 1991 das bis dahin energiereichste Event mit einer Energie von $3 \cdot 10^{20}$ eV beobachten konnte (7). 1995 begann dann der Bau des bis heute größten Experimentes zur Untersuchung kosmischer Strahlung: das Pierre Auger Observatorium (8).

1.2. Kosmische Strahlung

Der nachfolgende Abschnitt basiert, solange nicht anders angegeben auf folgenden Quellen: (9), (10), (11).

Die kosmische Strahlung lässt sich in zwei Arten aufteilen, die primäre kosmische Strahlung und die sekundäre Strahlung. Die primäre kosmische Strahlung ist jene, die nicht

terrestrischen Ursprungs ist, also z.B. von der Sonne oder von anderen Sternen, Galaxien, usw. kommt. Die primäre kosmische Strahlung setzt sich aus geladener und aus nicht geladener Strahlung zusammen. Erstere besteht zu 2 % aus Elektronen, zu 87 % aus Protonen, zu 12 % aus Heliumkernen und zu ca. 1 % aus schwereren Atomkernen. Die Zusammensetzung der primären kosmischen Strahlung ähnelt der chemischen Zusammensetzung des Sonnensystems. Die nicht ionisierende Strahlung ist in erster Linie schwache elektromagnetische Strahlung. Außerdem gibt es einen kleinen Teil an elektrisch neutralen Teilchen, vor allem Neutrinos und Photonen. Diese werden von kosmischen Magnetfeldern nicht abgelenkt und eignen sich somit zur Richtungsbestimmung. Da Neutrinos eine Masse haben, auch wenn diese extrem gering ist, können Neutrinos durch Gravitationsfelder abgelenkt werden, dieser Effekt ist aber oft zu vernachlässigen. Die sekundäre kosmische Strahlung entsteht, wenn energiereiche primäre kosmische Strahlung auf die oberen Schichten der Erdatmosphäre trifft und mit diesen wechselwirkt.

1.2.1. Energiespektrum der primären kosmischen Strahlung

Der differentielle Fluss $\frac{d\Phi}{dE}$ der primären kosmischen Strahlung ist aufgrund von vielen verschiedenen Experimenten im Laufe der Jahre sehr gut bestimmt. Hierbei hat sich ergeben, dass der differentielle Fluss durch ein Potenzgesetz der Form:

$$\frac{d\Phi}{dE} \propto E^{-\gamma} \quad (1)$$

beschrieben werden kann, mit γ als Spektralindex. Das Spektrum erstreckt sich über 12 Größenordnungen von 10^9 eV bis hin zu Energien von 10^{20} eV und mehr. Der Fluss fällt mit steigender Energie über mehr als 32 Größenordnungen ab. Es gibt mehrere markante Punkte im Energiespektrum, an denen sich der Spektralindex ändert. Der erste dieser Punkte liegt bei einer Energie von ca. $5 \cdot 10^{15}$ eV und wird auch als Knie (engl. knee) bezeichnet, hier ändert sich der Spektralindex von 2,7 auf 3,1. Verschiedene Experimente weisen auf ein mögliches zweites Knie bei einer Energie von ca. $4 \cdot 10^{17}$ eV mit einer Änderung des Spektralindex von 3,0 auf 3,3 hin. Da andere Experimente, wie z.B. AGASA dieses zweite Knie jedoch nicht ausmachen konnten, bleibt dessen Existenz bis auf weiteres umstritten. Der nächste markante Punkt ist der Knöchel (engl. ankle). Dieser liegt bei einer Energie von ca. $5 \cdot 10^{18}$ eV. Hier ändert sich der Spektralindex wieder von 3,1 zurück zu 2,7. Dieser Knick lässt sich mit dem Übergang von galaktischen zu extragalaktischen Quellen begründen. Das letzte Merkmal des Spektrums liegt bei $5 \cdot 10^{19}$ eV und ist der GZK-cut-off. Dieser wurde schon in den 1960er Jahren postuliert, konnte aber bis zum Start des Pierre-Auger-Observatoriums aufgrund von unzureichenden Daten nicht mit letzter Sicherheit bestätigt werden. Das Spektrum ist in [Abbildung 1](#) dargestellt. Das energetisch untere Ende des Spektrums wurde zum großen Teil direkt, durch verschiedene Experimente, vermessen. Dies erfolgte über verschiedene Ballon- bzw. Satellitenexperimente. Je höher die Energie jedoch wird, desto seltener treffen Teilchen mit einer bestimmten Energie auf die Erde, somit werden immer größere Detektoren benötigt, um einen gleich bleibenden Quotienten von $\frac{\text{Ereignisse}}{\text{Zeit}}$ im Detektor beizubehalten. Da es für die Raumfahrt eine gewisse logistische Herausforderung darstellt, kilometergroße

Experimente im Weltraum zu errichten, wird der energetisch obere Teil des Spektrums nur indirekt untersucht. Hierfür werden großflächige Detektoren auf der Erdoberfläche gebaut, die dann die sekundäre kosmische Strahlung untersuchen.

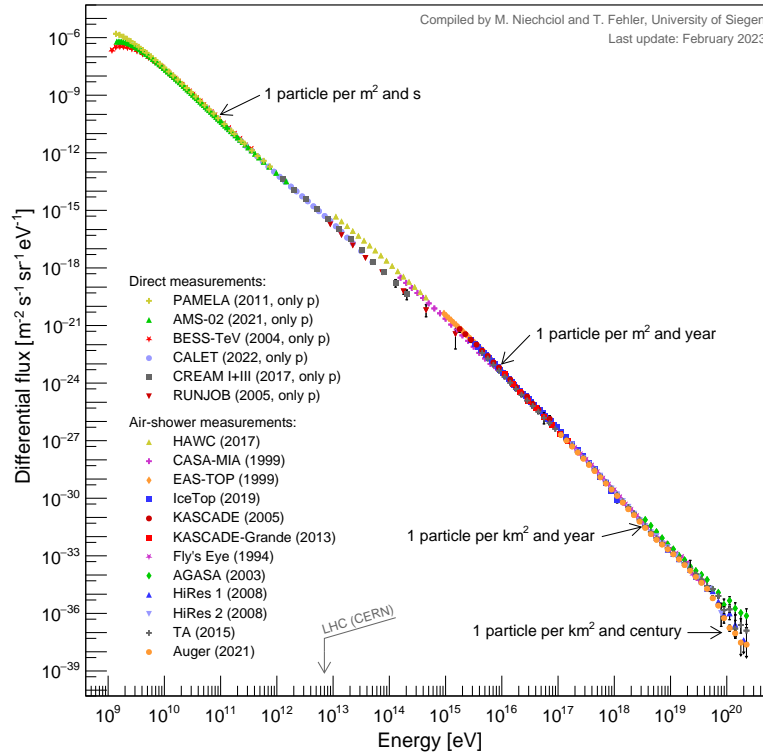


Abbildung 1: Spektrum der kosmischen Strahlung mit den Messungen verschiedener Experimente. (12)

1.2.2. Ursprung der kosmischen Strahlung

Es gibt verschiedene mögliche Quellen für kosmische Strahlung. Diese lassen sich grob in drei Gruppen einteilen: solare, galaktische und extra-galaktische kosmische Strahlung. Solare Strahlung ist die energetisch schwächste Strahlung, sie kommt, wie der Name schon vermuten lässt, von der Sonne. Die energetisch schwächste Strahlung der Sonne, ist der Sonnenwind, ein konsistenter Strom an vorwiegend Protonen, Elektronen und Alpha-Teilchen, mit einer Energie von einigen keV. Der Sonnenwind entsteht durch Beschleunigung von geladenen Teilchen in der äußersten Schicht der Sonne, der Corona. Des weiteren können auf der Oberfläche der Sonne Anomalien auftreten, in Form von Sonnenflecken. Diese thermisch kälteren Punkte weisen ein starkes lokales Magnetfeld auf. Durch die beim Auf- bzw. Abbau der magnetischen Felder entstehenden elektrischen Felder können die geladenen Teilchen des Sonnenplasmas auf Energien von bis zu 100 GeV beschleunigt werden.

Als mögliche Quelle galaktischer kosmischer Strahlung gelten unter anderem Supernovae. Hierbei können Teilchen direkt bei der Supernova Explosion entstehen oder aber mittels des Fermi-Mechanismus erster Ordnung beschleunigt werden. Dies geschieht, wenn sich die Schockwelle, die bei einer Supernova entsteht, in einem Plasma ausbreitet und die geladenen Teilchen in dem Plasma dabei beschleunigt/mitreißt. Eine weitere Art der Beschleunigung von geladenen Teilchen bietet der Fermi-Mechanismus zweiter Ordnung. Dieser besagt, dass über stochastische Prozesse einzelne geladene Teilchen in interstellaren Plasmawolken Energie aufnehmen können. Außerdem können geladene Teilchen in den extrem starken Magnetfeldern von Pulsaren auf hohe Energien beschleunigt werden.

Extragalaktische Quellen werden benötigt, um kosmische Strahlung mit Energien jenseits von 10^{18} eV zu erklären. Zu den möglichen Quellen, gehören z.B. AGNs. Ein AGN (aktiver galaktischer Kern) ist ein supermassives schwarze Loch, welches in Zentren von Galaxien vorkommt, und um das sich eine Akkretionsscheibe bildet. Senkrecht zu dieser Akkretionsscheibe entstehen Teilchen-Jets mit einer möglichen Teilchenenergie von über 10^{20} eV. Es werden noch weitere Quellen vermutet, wie z.B. Gamma Ray Bursts (GRBs), diese sind aber noch weitestgehend unerforscht.

1.2.3. GZK-Cutoff

Der GZK-Cutoff, benannt nach den drei Physikern Greisen, Zatsepin und Kuzmin, beschreibt die Wechselwirkung von höchst energetischen Protonen mit der kosmischen Hintergrundstrahlung. Erreicht ein Proton, welches durch den Weltraum propagiert, eine Energie von ca. 5×10^{19} eV kann dieses mit den Photonen der kosmischen Hintergrundstrahlung interagieren und über einen der folgenden beiden Kanäle wechselwirken.



Bei dieser Wechselwirkung verliert das Proton ca. 20 % seiner ursprünglichen Energie und wird in eine zufällige Richtung abgelenkt. Die Reaktionswahrscheinlichkeit sorgt dafür, dass höher energetische Teilchen nur eine Wegstrecke von ungefähr 100 Mpc zurücklegen können, ohne abgebremst zu werden. Siehe Abbildung 2. Das neutrale Pion kann in zwei Photonen zerfallen, die dann jeweils ca. 10 % der Energie des primären Photons besitzen. Dies ist eine mögliche Erklärung für die Existenz von ultra-hochenergetischen Photonen. (6)

1.2.4. Ausgedehnte Luftschauer (sekundäre kosmische Strahlung)

Die sekundäre kosmische Strahlung ist jene, die entsteht, wenn Teilchen der primären kosmischen Strahlung auf die Moleküle der Erdatmosphäre treffen und mit dieser wechselwirken. Infolge dieser Wechselwirkung entstehen neue Teilchen. Diese wechselwirken auch wieder und bilden so einen Luftschauer aus, der mit fortlaufender Ausbreitung in der Atmosphäre exponentiell an Teilchen dazu gewinnt. Hierbei lassen sich die erzeugten Teilchen in eine von drei Komponenten einordnen. Die hadronische, die myonische und die elektromagnetische Komponente. Siehe Abbildung 3.

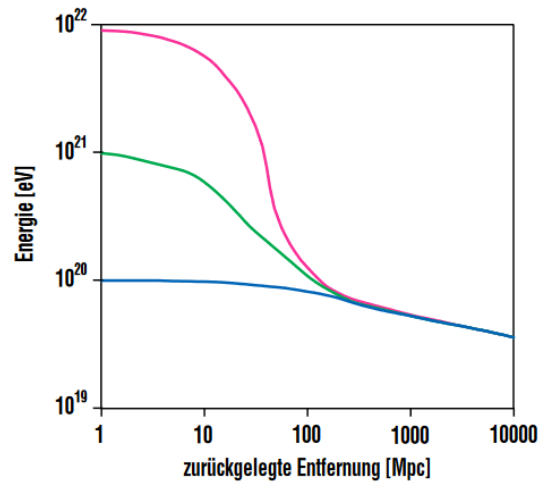


Abbildung 2: Darstellung des GZK-Cutoffs. Drei Protonen, mit verschiedener Anfangsenergie werden nach ca. 100 Mpc auf den Schwellenwert von ca. 5×10^{19} eV abgebremst. (13)

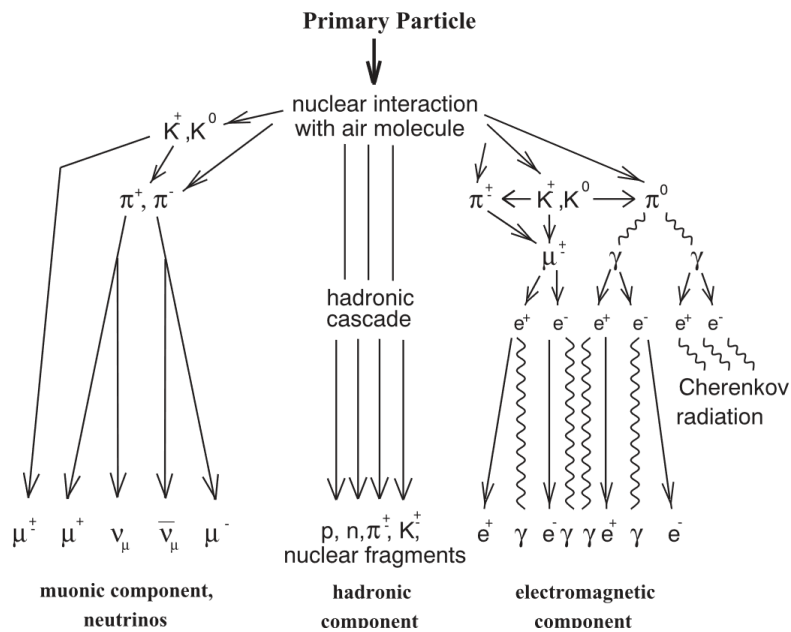


Abbildung 3: Schematische Darstellung der drei Komponenten eines Luftschauers, jeweils mit den dazugehörigen erzeugten Teilchen. (14)

Die elektromagnetische Komponente des Schauers besteht aus den Teilchen die elektromagnetisch wechselwirken, also Photonen, Elektronen und Positronen. Die beiden zugrunde liegenden physikalischen Vorgänge, die zur Ausbildung des Schauers beitragen, sind die Paarerzeugung und die Bremsstrahlung. Ein einfaches, aber dennoch gutes Modell um den elektromagnetischen Schauer zu beschreiben, ist das Heitler-Modell. Das Modell beruht auf der Annahme, dass nach einer mittleren Strahlungslänge X_0 entweder ein Elektron per Bremsstrahlung ein Photon erzeugt oder ein Photon per Paarerzeugung zu einem Elektron-Positron Paar wird. Dabei geht jeweils die Energie des primären Teilchens E_0 zu gleichen Teilen auf die beiden neuen Teilchen über und somit verdoppelt sich die Anzahl der Teilchen pro Schauer mit jeder zurückgelegten Strahlungslänge. Dies geht so lange weiter, bis die Energie der Teilchen unter die kritische Energie E_c , die notwendig für weitere Reaktionen ist, fällt. Die Energie, die die Teilchen nach k zurückgelegten Strahlungslängen haben ist:

$$E = \frac{E_0}{2^k} \quad (3)$$

Die kritische Energie wird nach k_c Schritten erreicht:

$$E_c = \frac{E_0}{2^{k_c}} \Rightarrow k_c = \frac{\ln\left(\frac{E_0}{E_c}\right)}{\ln 2} \quad (4)$$

Somit ergibt sich für die Tiefe des Schauer-Maximums folgendes:

$$X_{max}^{EM} = X_0 k_c = X_0 \frac{\ln\left(\frac{E_0}{E_c}\right)}{\ln 2} \quad (5)$$

Und für die maximale Anzahl an Teilchen im Schauer:

$$N_{max}^{EM} = 2^{k_c} = \frac{E_0}{E_c} \quad (6)$$

Somit hängt die Ausdehnung des elektromagnetischen Schauers lediglich von der Primärenergie des einfallenden Teilchens ab und hat folgende Proportionalitäten:

$$X_{max}^{EM} \propto \ln(E_0) \quad \text{und} \quad N_{max}^{EM} \propto E_0 \quad (7)$$

Die hadronische Komponente ist deutlich komplexer zu beschreiben, da diese aus zusammengesetzten Teilchen besteht, die auch mittels der starken Wechselwirkung untereinander wechselwirken können. Daher werden hadronische Schauer nicht, wie bei den elektromagnetischen Schauern durch die Strahlungslänge X_0 charakterisiert, sondern durch die Wechselwirkungslänge λ , die proportional zu der Massenzahl A und invers proportional zu der Avogadro-Konstante N_A , dem Wirkungsquerschnitt σ und der Dichte ρ ist.

$$\lambda = \frac{A}{N_A \times \sigma \times \rho} \quad (8)$$

Um nun den hadronischen Schauer zu beschreiben, kann auch wieder das Heitler-Modell verwendet werden, jedoch nur in einer modifizierten Variante. Beim hadronischen Schauer entstehen, nach einer Wechselwirkungslänge nicht zwei sondern $3n$ Pionen, welche sich gleichmäßig auf π^- , π^+ und π^0 aufteilen. Die π^0 zerfallen fast sofort in zwei Photonen,

$$\pi^0 \rightarrow \gamma + \gamma \quad (9)$$

welche dann, wie beschrieben den elektromagnetischen Teil des Schauers induzieren. Solange die geladenen Pionen genügend Energie besitzen, können diese nach einer weiteren Wechselwirkungslänge λ neue Pionen erzeugen. Fällt die Energie der geladenen Pionen unter eine kritische Energie von $E_{\text{krit}}^\pi = 20 \text{ GeV}$, besitzen die Pionen nicht mehr genügend Energie, um weitere Teilchen zu erzeugen, daher zerfallen sie in Myonen und die dazugehörigen Neutrinos.

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu \end{aligned} \quad (10)$$

Das gerade dargestellte Modell beschreibt die Wechselwirkung von Protonen mit der Luft zu Pionen. Pionen sind allerdings nicht die einzigen Teilchen, die bei dieser Wechselwirkung entstehen können. Nur zu ca. 90 % entstehen hierbei Pionen, zu 10 % entstehen andere Teilchen, wie z.B. Kaonen. Auch die Primärteilchen sind nur zu ca. 85 % Protonen, der Rest sind schwerere Atomkerne, deren Wechselwirkung noch komplexer ist und auf die an dieser Stelle nicht näher eingegangen wird.

Die myonische Komponente des Schauers entsteht in erster Linie durch den Zerfall von durch die starke Wechselwirkung erzeugten Mesonen. Die Dichte der Atmosphäre nimmt mit zunehmender Höhe exponentiell ab, das heißt, dass die Wechselwirkungsrate von Mesonen mit geringer werdender Höhe stark zunimmt. Somit haben die Mesonen nur in den oberen Schichten der Atmosphäre den nötigen Freiraum, um zu zerfallen. Also bildet sich der Großteil des myonischen Schauers in der oberen Schicht der Atmosphäre aus. Da Myonen kaum Energie durch Bremsstrahlung verlieren und auch nicht der starken Wechselwirkung unterliegen, erreichen die energiereicheren Myonen ungehindert die Erdoberfläche. Die energiearmen Myonen zerfallen allerdings aufgrund ihrer kurzen Lebensdauer bereits in der Atmosphäre. (15)

1.3. Cherenkov-Strahlung

Um nun die Teilchen, die die Erdoberfläche erreichen zu detektieren, gibt es verschiedene Möglichkeiten. Eine davon ist, die Teilchen per Cherenkov-Strahlung zu detektieren. Auf dieser Methode basiert der Oberflächendetektor des Pierre Auger Observatoriums. Siehe 2.4. Wenn ein geladenes Teilchen durch ein elektrisch isolierendes Medium fliegt, ionisiert es beim Vorbeifliegen die Atome des Mediums. Bei der Relaxation werden elektromagnetische Wellen erzeugt, die im Normalfall destruktiv miteinander interferieren. Hat das einfallende Teilchen nun eine höhere Geschwindigkeit, als die Lichtgeschwindigkeit in dem Medium, kommt es nicht mehr zur destruktiven Interferenz, sondern es bildet sich eine Wellenfront aus, ähnlich wie beim Überschallkegel.

$$v \geq \frac{c_0}{n} \quad (11)$$

Hierbei ist v die Geschwindigkeit des einfallenden Teilchens, c_0 die Lichtgeschwindigkeit im Vakuum und n der Brechungsindex des Mediums. Siehe Abbildung 4. Der Winkel

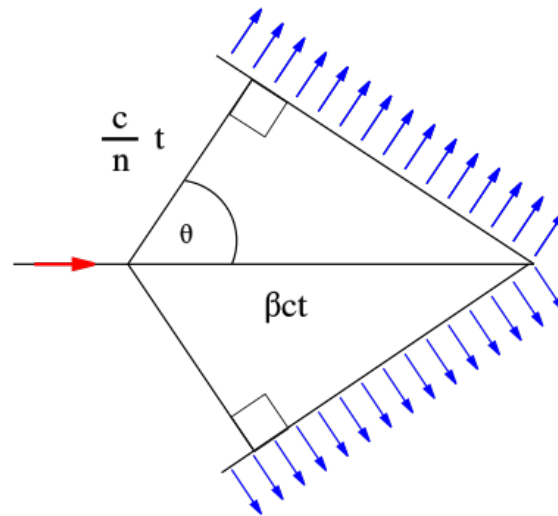


Abbildung 4: Darstellung des Cherenkov-Effekts, mit der Ausbreitung des Cherenkov-Kegels. (16)

unter dem das Cherenkov-Licht abgestrahlt wird, berechnet sich wie folgt:

$$\cos(\theta_c) = \frac{1}{\beta \times n} \quad (12)$$

mit $\beta = \frac{v}{c_0}$ als die auf die Lichtgeschwindigkeit normierte Geschwindigkeit des einfallenden Teilchens.

2. Pierre Auger Observatorium

Das Kapitel basiert, sofern nicht anders angegeben auf den folgenden Quellen: (8), (17), (18)

Das Pierre Auger Observatorium steht in Argentinien und wurde gebaut, um kosmische Strahlung zu untersuchen. Das Ziel des Projektes ist es, kosmische Strahlung am energetisch oberen Ende der Skala zu studieren, jenseits einer Energie von 10^{17} eV. Niedrig energetische kosmische Strahlung ist bereits gut verstanden, während höchst energetische kosmische Strahlung noch viele Fragen aufwirft. Das Pierre Auger Projekt versucht, Fragen nach dem Ursprung und der Beschaffenheit der kosmischen Strahlung zu beantworten und dringt dabei in Bereiche des Universums vor, die nie ein Mensch zuvor gesehen hat.

2.1. Geschichte

Die erste Idee zum Bau des Pierre Auger Observatoriums entstand 1992 und ging auf den Physik-Nobelpreisträger Jim Cronin und auf Alan Andrew Watson zurück. Benannt ist das Observatorium nach dem 1899 geborenen Physiker Pierre Victor Auger, der sich sein Leben lang mit der kosmischen Strahlung beschäftigte und 1938 die Existenz von ausgedehnten Luftschauern entdeckte. Die konkrete Planung des Baus begann 1995 und dauerte fünf Jahre. Zur Jahrtausendwende begann, nachdem kurz vorher die Pierre Auger Kollaboration gegründet wurde, der Bau des Observatoriums. Vier Jahre später, nachdem die ersten hundert Oberflächendetektoren installiert worden waren, fing das Pierre Auger Observatorium an erste Daten aufzunehmen. Die Kollaboration veröffentlichte drei Jahre später das erste Paper, in dem die räumliche Korrelation zwischen den 27 höchsten energetischen Ereignissen und verschiedener AGNs bestätigt wurde. Im darauf folgenden Jahr wurde die erste Bauphase des Observatoriums abgeschlossen. Das Observatorium besteht nun aus 1660 Oberflächendetektoren und 4 Standorten für die 24 Fluoreszenz Teleskope. Im gleichen Jahr, 2008, bestätigt das Pierre Auger Observatorium, dass der Fluss der kosmischen Strahlung bei den höchsten Energien stark unterdrückt ist. Diese Beobachtungen sind konsistent mit der GZK-Vorhersage. In den folgenden Jahren werden obere Limits für den Fluss für Photonen bzw. Neutrinos mit Energien größer als 10^{18} eV veröffentlicht. 2015 wird eine Anisotropie bei höchst energetischer kosmischer Strahlung festgestellt. Außerdem beginnt in diesem Jahr der Ausbau mit dem sogenannten AugerPrime Projekt. Bei diesem werden die Stationen des Oberflächendetektors unter anderem um einen Plastik Scintillator Detektor und eine Radioantenne erweitert.

2.2. Aufbau des Pierre Auger Observatoriums

Das Pierre Auger Observatorium besteht aus einem Array aus 1660 Wasser-Cherenkov-Detektoren, welche auf einer Fläche von 3000 km^2 in einem Abstand von 1,5 km aufgebaut sind. Das Observatorium steht auf der Hochebene Pampa Amarilla in der Provinz Mendoza und liegt im Schnitt 1400 Meter über N.N. Der Hauptstandort, inklusive Kontrollraum, ist in der Stadt Malargüe gelegen. Die ganze Hochebene wird von 27 Luft-Fluoreszenz Teleskopen, welche auf vier Stationen aufgeteilt sind, überschaut. 2015

wurde das Auger Observatorium mittels des AugerPrime Upgrades signifikant verbessert. Zu den Verbesserungen gehört die Erweiterung der Oberflächendetektoren um einen weiteren PMT, eine Radioantenne und einen Myonen Detektor.

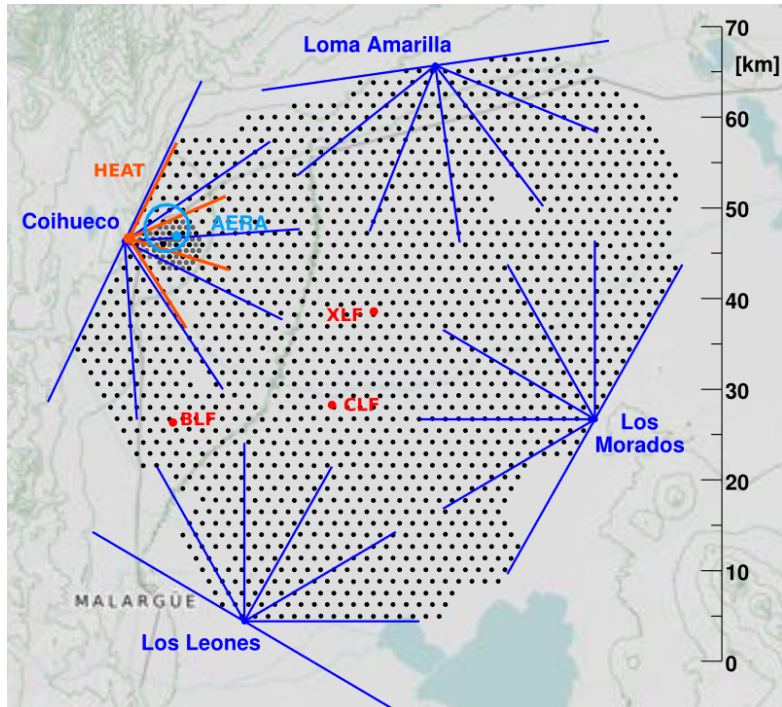


Abbildung 5: Layout des Observatoriums: jeder Punkt repräsentiert ein Tank des Oberflächendetektors. Die Fluoreszenz-Teleskope stehen an den 4 Standorten: Los Leones, Coihueco, Loma Amarilla & Los Morados und überblicken das ganze Gelände. Der Standort Coihueco beherbergt zusätzlich noch drei weitere Teleskope (HEAT). Außerdem ist hier ein Bereich mit geringerer Dichte an Cherenkov Detektoren (750 m Abstand) und das Radio Antennen Array (AERA) aufgebaut. Unten links ist die Stadt Malargüe zu sehen, hier ist die Operationszentrale des Pierre Auger Observatoriums angesiedelt. (18)

2.3. Fluoreszenzdetektor

Der Fluoreszenzdetektor besteht aus vier Anlagen, die an den vier Standorten: Los Leones, Coihueco, Loma Amarilla und Los Morados stehen. Jede Anlage besteht aus 6 Fluoreszenz Teleskopen, die jeweils ein Blickfeld von $30^\circ \times 30^\circ$ haben. Somit ergibt sich ein gesamter Blickwinkel von 180° . Jeder Detektor besteht aus einem Spiegel mit einer Spiegelfläche von 13m^2 , welcher in kleinere Segmente unterteilt ist, um die Produktion und den Transport zu vereinfachen. Es existieren zwei Designs für die Spiegel: die Hälfte der Spiegel besteht aus 36 rechteckigen Spiegeln aus anodisiertem Aluminium, die andere Hälfte der Spiegel besteht aus 60 hexagonalen Glasspiegeln. Vor der Öffnung der Teleskope befindet sich

ein Filter, welcher nur UV Licht durchlässt. Im Bereich von 310 nm bis 390 nm wird ca. 50 % Licht durchgelassen und im Bereich von 330 nm bis 380 nm ca. 80 %. Dies dient der Unterdrückung von Lichtverschmutzung. Der eigentliche Detektor besteht aus einer 420 Pixel großen Kamera. Aufgrund der enormen Empfindlichkeit der Kamera kann eine Beobachtung mit den Fluoreszenz Teleskopen nur in mondlosen Nächten stattfinden. An dem Standort in Coihueco ist neben dem normalen Fluoreszenz Teleskopen noch das HEAT Projekt aufgebaut (High Elevation Auger Telescope). Dieses besteht aus drei Teleskopen, die ähnlich funktionieren wie die anderen Fluoreszenz Teleskope. Allerdings besteht hier die Besonderheit, dass diese für einen höheren Bereich in der Atmosphäre konzipiert sind, somit sind sie in der Lage auch schon Teilchen mit geringer Energie (10^{17} eV) zu detektieren. Die anderen Fluoreszenz Teleskope können Teilchen erst ab einer Energie von ca. 10^{18} eV detektieren.



Abbildung 6: Außenansicht eines der Fluoreszenz Teleskop Anlagen. (17)

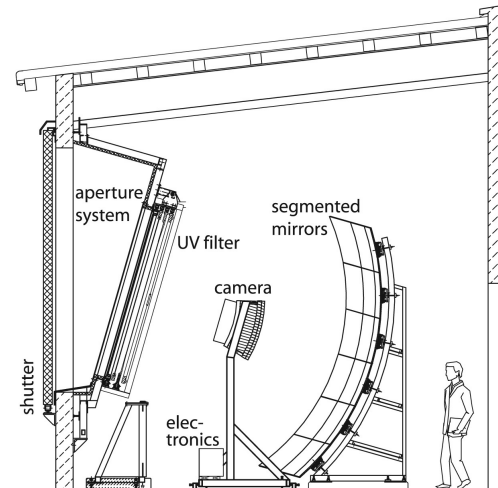


Abbildung 7: Innerer Aufbau eines der Teleskope. Schematisch dargestellt befindet sich ganz außen die Blende, danach folgt der UV-Filter. Das Licht wird danach vom Spiegel in die Kamera reflektiert. (17)

2.4. Oberflächendetektor

Der Oberflächendetektor besteht aus 1660 einzelnen Detektoren, die in einem Abstand von 1500 m auf einer Fläche von 3000 km² aufgebaut sind. Das gesamte Areal wird von vier Standorten aus, jeweils mit sechs Fluoreszenz Teleskopen überschaut. Jede einzelne Station ist selbstversorgend und wetterunabhängig. Somit hat der Oberflächendetektor ein Uptime von über 99 %. Zu jeder Station gehört, als entscheidende Komponente, der Wasser Cherenkov Detektor (WCD), welcher aus hochreinem Wasser besteht, in dem die Teilchen der ausgedehnten Luftschaer (engl. extensive air shower (EAS)) Cherenkov-Licht erzeugen, wenn sie diesen durchqueren. Des weiteren gehört zu jeder Station, seit dem Auger Prime Upgrade auch eine Radioantenne und ein Szintillator Detektor. Letzterer misst ebenfalls, wie der Wasser Cherenkov Detektor, die Energie der niedergehenden Luftschaer. Durch die unterschiedlichen Messverfahren lassen sich Rückschlüsse auf den Myonengehalt der Luftschaer ziehen.

2.4.1. Aufbau

Jede Station des Oberflächendetektors besteht aus einem Tank aus Polyethylen, der 12000 l hochreines Wasser beinhaltet. Die Tanks haben einen Durchmesser von 3.6 m und eine Höhe von anderthalb Meter, mit einer Wassertiefe von 1.2 m. Der Tank selber ist von innen mit einer diffus reflektierenden Schicht überzogen, damit das, durch die Luftschaer erzeugte Cherenkov Licht besser von den Photomultipliern (PMT) detektiert werden kann. Diese PMTs sitzen im Deckel vom Wassertank, drei große Photomultiplier und, seit dem Auger Prime Upgrade noch ein weiterer deutlich kleinerer Photomultiplier. Die Photomultiplier haben eine Zeitauflösung von 25 ns, also eine Auslesefrequenz von 40 MHz. Außen am Tank befindet sich die Aufbewahrungsbox für die gesamte Elektronik und die WLAN Antenne, die für die Kommunikation der Station mit den Nachbarstationen und der Datenweiterleitung an die Hauptstation verantwortlich ist. Mit dem Auger Prime Upgrade wurden die Stationen um eine Radioantenne und einen Szintillator Detektor erweitert. Jede Station wird über ein Solarpanel versorgt, welches, in Kombination mit einem Batteriepack vollkommen ausreicht, um die gesamte Elektronik mit Strom zu versorgen. Der Stromverbrauch beläuft sich auf knapp 10 W im Durchschnitt.

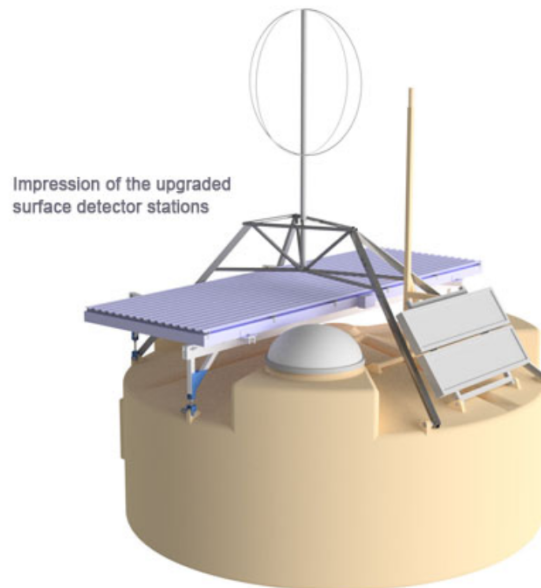


Abbildung 8: Hier ist als Renderbild eine Station des Oberflächendetektors dargestellt. Oben auf dem Wassertank in der Mitte befindet sich der Szintillator Detektor und die Radioantenne. Vorne auf dem Tank sind das zweiteilige Solarmodul und die WLAN Antennen zu sehen. Unter der Kuppel befindet sich, wettergeschützt, die Elektronik. (8)

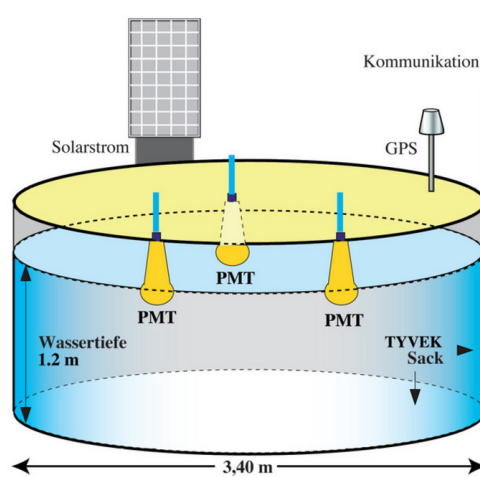


Abbildung 9: Innerer Aufbau des Wassertanks, mit den drei PMTs, dem Solarpanel und der Antenne, vor dem Auger Prime Upgrade. (19)

3. Neuronale Netze (NN)

Das Kapitel basiert, sofern nicht anders angegeben auf den folgenden Quellen: (20), (21) Ein neuronales Netz (oder auch künstliches neuronales Netz, kurz kNN) ist ein datenverarbeitendes System, welches in Aufbau und Funktionsweise dem menschlichem Gehirn nachempfunden ist. Hierbei besteht es, ähnlich wie sein biologisches Vorbild, aus Neuronen, die miteinander verbunden sind und Daten verarbeiten und weiterleiten können. Die Anwendungsmöglichkeiten heutzutage sind extrem vielfältig und reichen von Bild- und Mustererkennung über Wettervorhersage und Textvervollständigung bis hin zu medizinischen Diagnosen und autonomem Fahren. Hierbei kommen verschiedenste Arten von NN zum Einsatz, die gängigsten werden in 3.3 näher erläutert.

3.1. Geschichte der Neuronalen Netze

Das erste Mal erwähnt wurden Neuronale Netze von dem Neuropsychologen Warren McCulloch und dem Logiker Walter Pitts im Jahre 1943. Die Dartmouth Conference von 1956 gilt als Geburtsstunde des Forschungsfeldes Künstliche Intelligenz. Im darauf folgenden Jahr stellte Frank Rosenblatt einen ersten Neurocomputer vor, der in der Lage war, auf 20x20 Pixeln Ziffern zu erkennen. Eine praktische Anwendung fand ein Netz zur Echo-Unterdrückung in der Telefontechnik. Ende der 1960er Jahre fand die Anwendung von Neuronalen Netzen erst mal ein Ende, da die Rechenleistung der damaligen Computer nicht ausreichte, um komplexere Netze umzusetzen. Dennoch wurde in den folgenden Jahren viel an der Theorie zu NN geforscht. So stellte 1974 Paul Werbos einen Algorithmus vor, den er Backpropagation nannte. Dieser ist bis heute ein wesentlicher Bestandteil aller Neuronalen Netze. In den 1980er Jahren entstand ein neuer, allerdings recht kurzer, Aufschwung von NN. Hier wurden NN z.B. zur Proteinstruktur- vorhersage eingesetzt. Auch hier begrenzte wieder die eingeschränkte Rechenleistung eine weitere Verbreitung von NN. Erst im Jahr 2009 rückten NN wieder in den Vordergrund, nachdem die Forschungsgruppe um Jürgen Schmidhuber mit ihren rekurrenten NN eine Reihe von Mustererkennungs-Wettbewerben gewann. Zwei Jahre später optimierte Dan Ciresan seine NN dahin, dass er sie auf GPUs trainieren konnte. Dadurch gelang es ihm, sehr komplexe tiefe NN zu trainieren, die in der Lage waren, in einigen Mustererkennungs-Aufgaben übermenschliche Genauigkeit zu erzielen. Da nun endlich Computer über die nötige Rechenleistung verfügten, bzw. man in der Lage war, NN auf ganzen Supercomputer-Clustern zu trainieren, erfuhren NN einen riesigen Aufschwung und sind seitdem überall in unserem Leben in Verwendung. Die letzte große Errungenschaft im Bereich Neuronaler Netze gelang der Firma OpenAI, die mit ihrer KI ChatGPT ein Chatprogramm vorstellten, das in der Lage ist, wie ein Mensch auf verschiedenste Fragen zu antworten. Nur Tage nach der ersten Veröffentlichung von ChatGPT, fand dieses bereits weltweit großes Interesse.

3.2. Allgemeiner Aufbau von Neuronalen Netzen

Das einfachste Neuronale Netz besteht aus einem einzelnen Neuron. Solch ein Netz wird auch schon mal als Perceptron bezeichnet. Dieses bekommt als Input einen Vektor \vec{x} mit Daten. Diese werden jeweils mit einem Gewicht m_i multipliziert und danach aufsummiert. Typischerweise wird noch ein Bias b hinzu addiert, ein Wert, der unabhängig vom Input ist. Der Output eines Neurons erfolgt über eine Aktivierungsfunktion $y = f(z)$ mit $z = \sum x_i \times m_i + b$. Die einfachste Aktivierungsfunktion ist eine Stufenfunktion. Eine andere oft genutzte Aktivierungsfunktion ist die ReLU (Rectified Linear Activation Funktion). Diese lautet: $f(z) = \max(0, z)$. Nun kann man den Aufbau um einige Neuronen erweitern. Diese bekommen die selben Daten, wie das erste Neuron, haben aber andere Parameter, also andere Gewichte, anderen bias und einen andern Schwellenwert, als das erste Neuron und können somit andere Informationen aus den Daten ziehen. Somit hat man jetzt ein einfaches Neuronales Netz, mit dem klassischen Aufbau von: Input-schicht, (hidden)-Layer, Output-schicht. Graphisch dargestellt in Abbildung 10.

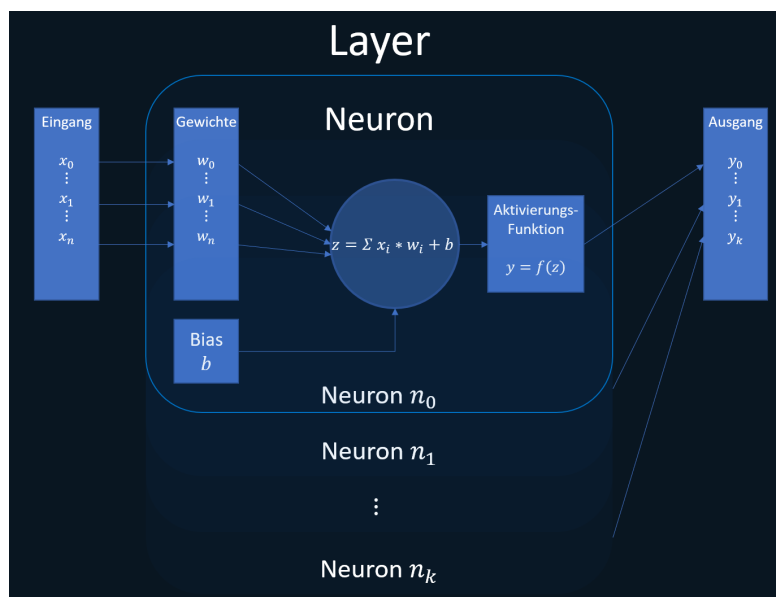


Abbildung 10: Aufbau eines einfachen Neuronalen Netzes, mit Input, hidden Layer und Output.

Das Trainieren eines Neuronalen Netzes bezeichnet den Vorgang, die einzelnen Parameter des Netzes iterativ solange anzupassen, bis das Netz in der Lage ist, die gewünscht Aussage anhand der eingegebenen Daten zu treffen. Dafür wird noch ein weiteres mathematisches Hilfsmittel benötigt: die Backpropagation. (22) Diese erlaubt es, den Fehler der Ausgabe des Neuronalen Netzes in entgegengesetzter Richtung durch das Netz zurück zu propagieren und so die Parameter der einzelnen Neuronen anzupassen. Damit bei einem erneuten vorwärts Durchlaufen des Netzes der Fehler der Ausgabe minimiert wird. Der Algorithmus lässt sich in mehrere Schritte aufteilen.

1. Als erstes wird der Datensatz in das NN gegeben und die Ausgabe berechnet.
2. Danach wird der Fehler der Ausgabe bestimmt, hier wird z.B. der quadratische Abstand zwischen getätigter und erwarteter Ausgabe genommen.
3. Nun wird der Fehler durch das Netz zurückpropagiert, indem die Kettenregel der Ableitung genommen wird, um den Einfluss jedes Parameters auf den Fehler zu bestimmen.
4. Als letztes werden die Parameter entsprechend der berechneten Ableitung angepasst, um den Fehler zu minimieren.

Der ganze Vorgang wird solange wiederholt, bis die gewünschte Vorhersagegenauigkeit erreicht ist. Zu Punkt zwei ist noch zu erwähnen, dass es verschiedene Arten gibt, den Fehler zu berechnen. Für Regressionsprobleme wird typischerweise der Mean-square-error (MSE) verwendet, für Klassifizierungsprobleme wird typischerweise die Crossentropy verwendet. Dabei ist die Formel für die Crossentropy:

$$CE(P, Q) = - \sum P(x) \times \log(Q(x)) \quad (13)$$

mit $P(x)$ als tatsächliche Wahrscheinlichkeit für das Ereignis x und $Q(x)$ als vorhergesagte Wahrscheinlichkeit für das Ereignis x . Summiert wird über alle möglichen Klassen. Außerdem gibt es verschiedene Namen für diese Fehlerfunktionen: loss-function, cost-function oder manchmal auch error-function. Auch wenn diese Namen mehr oder weniger Synonyme für ein und dieselbe Art von Funktion sind, werden sie teilweise in verschiedenen Quellen unterschiedlichen Gruppen von Fehlerfunktionen zugeordnet. Um das Ganze zumindest etwas anschaulicher zu machen folgt ein Beispiel: Nehmen wir als Grundlage ein ganz einfaches Neuronales Netz (Siehe Abb. 11). Für den ersten Schritt, werden die Daten vorwärts durch das Netz durchgegeben, somit berechnen sich die outputs der einzelnen Neuronen wie folgt:

$$\begin{aligned} h_1 &= f(z_1) \text{ mit } z_1 = i_1 \times w_1 + i_2 \times w_2 + b_1 \\ h_2 &= f(z_2) \text{ mit } z_2 = i_1 \times w_3 + i_2 \times w_4 + b_2 \end{aligned} \quad (14)$$

mit der Aktivierungsfunktion $f(z)$.

$$\begin{aligned} o_1 &= f(z_3) \text{ mit } z_3 = h_1 \times w_5 + h_2 \times w_6 \\ o_2 &= f(z_4) \text{ mit } z_4 = h_1 \times w_7 + h_2 \times w_8 \end{aligned} \quad (15)$$

Als zweites wird der Fehler der Ausgabe bestimmt.

$$E_{total} = \sum \frac{1}{2} (o_{i,true} - o_{i,pred})^2 = E_{o_1} + E_{o_2} \quad (16)$$

Danach wird die Abhängigkeit des Fehlers von den einzelnen Parametern, mithilfe der Kettenregel bestimmt:

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial o_1} \times \frac{\partial o_1}{\partial z_3} \times \frac{\partial z_3}{\partial w_5} \quad (17)$$

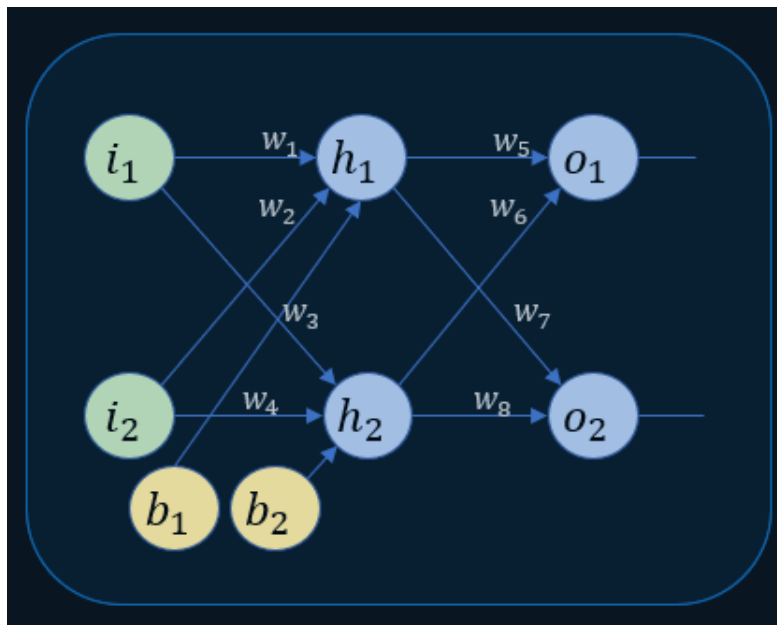


Abbildung 11: Aufbau eines einfachen Neuronales Netzes mit zwei input nodes, zwei hidden nodes und zwei output nodes.

Dies gilt für die Gewichte der output Schicht. Für die Gewichte des hidden Layers wird das ganze etwas komplizierter:

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial h_1} \times \frac{\partial h_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1} \quad (18)$$

Mit:

$$\frac{\partial E_{total}}{\partial h_1} = \frac{\partial E_{o_1}}{\partial h_1} + \frac{\partial E_{o_2}}{\partial h_1} \quad (19)$$

Hierbei ist:

$$\frac{\partial E_{o_1}}{\partial h_1} = \frac{\partial E_{o_1}}{\partial z_3} \times \frac{\partial z_3}{\partial h_1} \quad (20)$$

Mit:

$$\frac{\partial E_{o_1}}{\partial z_3} = \frac{\partial E_{o_1}}{\partial o_1} \times \frac{\partial o_1}{\partial z_3} \quad (21)$$

Da

$$z_3 = w_5 \times h_1 + w_6 \times h_2 \quad (22)$$

ist, folgt:

$$\frac{\partial z_3}{\partial h_1} = w_5 \quad (23)$$

Somit ist:

$$\frac{\partial E_{o_1}}{\partial h_1} = \frac{\partial E_{o_1}}{\partial o_1} \times \frac{\partial o_1}{\partial z_3} \times w_5 \quad (24)$$

Damit folgt für die gesamte Gleichung:

$$\Rightarrow \frac{\partial E_{total}}{\partial w_1} = \left(\frac{\partial E_{o_1}}{\partial o_1} \times \frac{\partial o_1}{\partial z_3} \times w_5 + \frac{\partial E_{o_2}}{\partial o_2} \times \frac{\partial o_2}{\partial z_4} \times w_7 \right) \times \frac{\partial h_1}{\partial z_1} \times \frac{\partial z_1}{\partial w_1} \quad (25)$$

Um nun die Gewichte anzupassen, kommt ein weitere Algorithmus ins Spiel: der sogenannte Gradient Descent Algorithmus. Dieser versucht iterativ die Gewichte so anzupassen, dass die loss-Funktion minimiert wird. Siehe Abbildung 12 Hierbei geht der Algorithmus schrittweise in Richtung des stärksten Abfalls. Da der normale Gradient descent Algorithmus oft in lokalen Minima hängen bleibt, gibt es verschiedene optimierter Algorithmen. Der am häufigsten verwendete ist der Adam Algorithmus (kommt von adaptive momentum estimator) (23). Dabei zeichnet sich der ADAM Algorithmen dadurch aus, dass dieser die learning rate basierend auf dem Momentum der Gradienten anpasst und somit eine schnelle und effiziente Konvergenz ermöglicht.

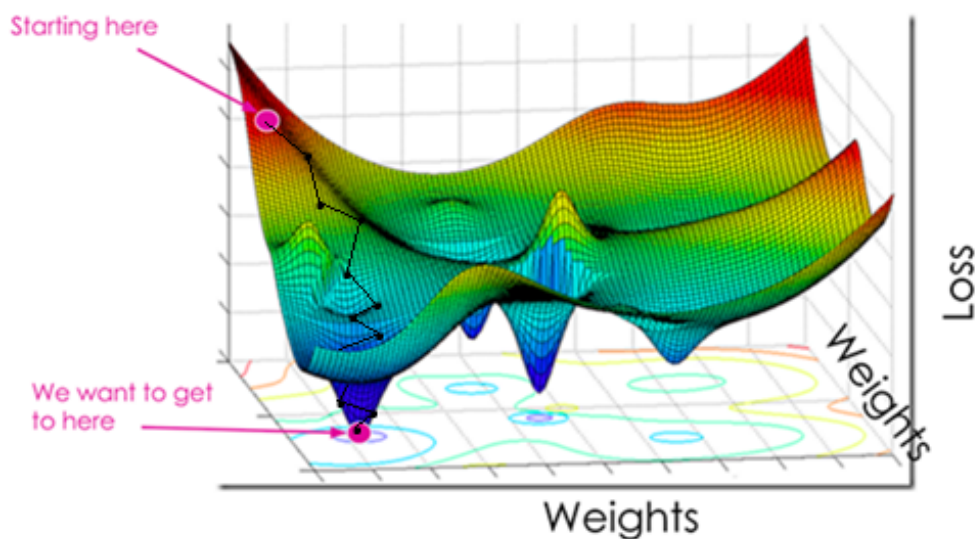


Abbildung 12: Schematische Darstellung der Loss-Funktion in Abhängigkeit von den Gewichten. (24, (modifiziert))

3.3. Arten von Neuronalen Netzen

Das klassische Feed-forward NN, oder auch Fully Connected NN genannt, da alle Neuronen einer Ebene mit allen Neuronen der nachfolgenden Ebene verbunden sind, wurde bereits im Kapitel zuvor ausführlich beschrieben. Da neben gibt es noch zwei weitere Arten

von Neuronalen Netzen, die von besonderer Bedeutung sind: zum einen Convolutional Neuronal Networks (CNN) und zum anderen Recurrent Neuronal Networks (RNN). Convolutional Neuronal Networks eignen sich besonders gut zur Bilderkennung, da sie in der Lage sind, Formen zu lernen. Hierbei wird eine zweidimensionale Faltungsmatrix (auch Filter oder Kernel genannt) über das Bild gelegt. Die Ausgabe der einzelnen Neuronen berechnet sich jetzt aus dem Produkt der Faltungsmatrix mal den Pixelwerten des Bildausschnittes. Somit reagiert ein einzelnes Neuron gleich bei einer ähnlichen Form, auch wenn diese an komplett anderer Stelle im Bild ist. Fügt man nun mehrere Convolutional Layer hintereinander, kann ein NN auch komplexe Formen lernen, wie z.B. einen Hund von einer Katze zu unterscheiden, oder ein Auto von einem Schiff. Siehe Abb. 13. Ein weiterer Layer, der typischerweise zusammen mit dem Convolutional Layer

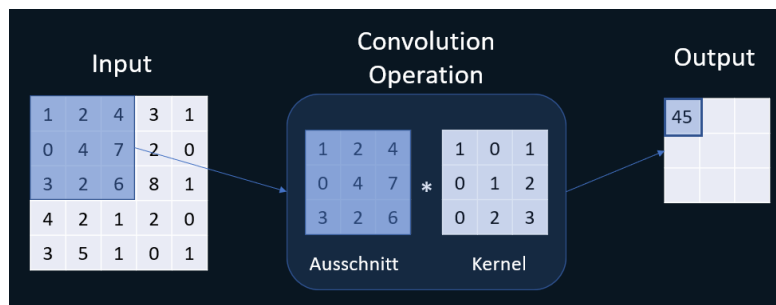


Abbildung 13: Darstellung eines 3x3 Convolutional Layers.

zum Einsatz kommt ist der Pooling Layer, bzw. im besonderen der Max-Pooling Layer. Dieser hat als Aufgabe, die Dimensionalität der Daten zu reduzieren und somit die Anzahl an Parametern im Netz zu verringern. Dafür nimmt der Algorithmus, im Falle des Max-Pooling Layers aus einem Bildausschnitt jeweils den Maximalwert. Vergleiche Abbildung 14. Auch wenn hierdurch das Eingabebild kleiner wird, verliert das Netz normalerweise nicht an Genauigkeit, sondern gewinnt typischerweise an Robustheit und beugt so dem Problem des Overfittings vor. Overfitting beschreibt dabei, den Fall, dass das Neuronale Netz auf dem Trainingsdatensatz gut performt, jedoch Probleme hat neue, nicht gesehene Daten zuzuordnen. Da das NN die Trainingsdaten auswendig lernt, anstatt die Eigenschaften der Daten zu lernen. Außerdem wird die Trainingszeit dadurch verringert, dass die Anzahl an Parametern reduziert wird und es wird einfacher, tiefere Netze zu bauen.

Eine weitere wichtige Art von Neuronalen Netzen sind Recurrent Neuronal Networks. Die Besonderheit bei diesen Netzen ist, dass einzelne Neuronen einer Schicht Verbindungen zu anderen Neuronen derselben oder einer vorherigen Schicht haben. Dadurch kann ein Netz temporäre Zusammenhänge lernen. Dies ist beispielsweise für Textvervollständigungen oder Wettervorhersagen nützlich. Bei einem einfachen RNN ist ein einzelnes Neuron mit sich selbst verknüpft und erhält somit die letzte Ausgabe als zusätzliche Eingabe. Vergleiche Abbildung 15. Diese Art von Neuronalem Netz eignet sich gut, um direkte temporäre Zusammenhänge zu lernen. Möchte man allerdings auch temporär nicht benachbarte Ereignisse miteinbeziehen, braucht man eine erweiterte Art des RNNs, bei

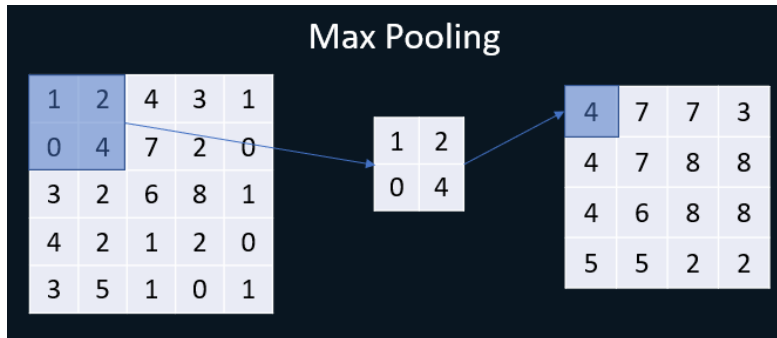


Abbildung 14: Darstellung eines 2×2 Max-Pooling Layers.

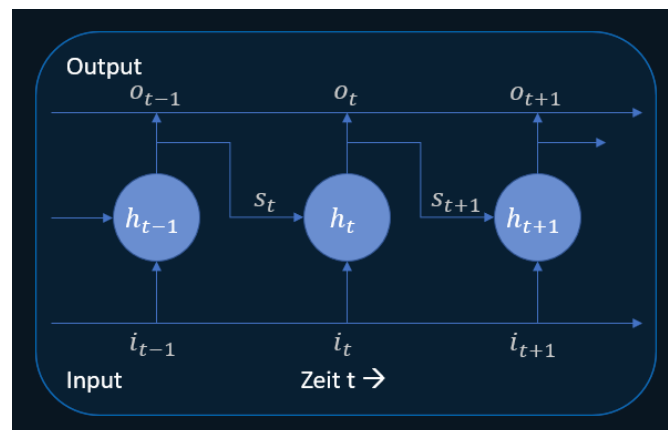


Abbildung 15: Zeitlich entfaltete Darstellung eines einzelnen Neurons in einem RNN. Hier wird die Ausgabe eines Neurons als weitere Eingabe beim nächsten Durchlauf hinzugegeben.

dem auch Informationen über einen längeren Zeitraum behalten werden. Hierzu wurde der LSTM Layer entwickelt (long short term memory). Ein LSTM Neuron besteht aus vier Hauptkomponenten:

1. Dem Input-Gate, das regelt, welche Informationen reinkommen.
2. Dem Forget-Gate, das regelt, welche Informationen vergessen werden sollen.
3. Dem eigentlichen Cell-State, in dem alle Informationen gespeichert sind.
4. Und dem Output-Gate, das regelt, welche Informationen weitergegeben werden sollen.

Als erstes wird das Input-Gate und das Forget-Gate berechnet. Danach wird der neue Zellenzustand berechnet, indem die bereits gespeicherten Informationen zusammen mit den neu hinzugekommenen Informationen verarbeitet werden. Zum Schluss wird das Output-Gate aktiviert, das bestimmt, welche Daten weiter gegeben werden. Somit kann ein LSTM relevante Informationen auch über einen längeren Zeitraum hinweg behalten. Dies ist dargestellt in Abbildung 16.

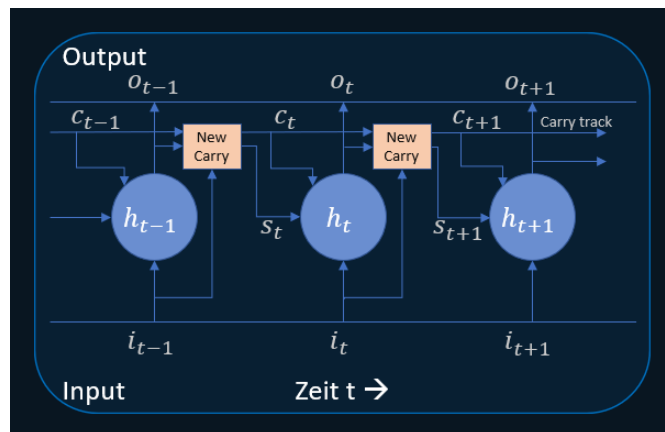


Abbildung 16: Zeitlich entfaltete Darstellung eines einzelnen LSTM Neurons.

4. Durchführung

4.1. Verwendeter Datensatz

Wenn geladene Teilchen der hoch energetischen sekundären kosmischen Strahlung auf das Wasser in einem der Tanks des Oberflächendetektors treffen, erzeugen diese einen Cherenkov-Kegel entlang der Flugbahn der Teilchen. Das so entstehende Cherenkov-Licht wird von der reflektierenden Unterseite des Tanks zurück in die Photomultipliern reflektiert. Diese wiederum werden mit einer Wiederholrate von 40 MHz ausgelesen und in ein digitales Signal umgewandelt. Die Signalamplitude wird typischerweise in „Vertical Equivalent Muon“ (VEM) angegeben. Dies entspricht dem Signal, das ein einzelnes vertikal durch den Tank gehendes Myon erzeugt. Die von einem Teilchenschauer in einem der Tanks deponierte Energie folgt einer Landau Verteilung. Hierbei erzeugen die verschiedene Komponenten eines Schauers verschiedene Landau Verteilungen. Die myonische Komponente eines Schauers kann die Erdatmosphäre nahezu ungehindert durchqueren. So erreicht diese als erstes den Detektor und erzeugt dabei auch den höchsten Ausschlag. Erst danach erreicht die elektromagnetische Komponente den Detektor. Zusammen mit der elektromagnetischen Komponente, die aus den hadronische Zerfällen erzeugt wurde. Etwas früher, aber ähnlich schwach wie die elektromagnetische Komponente aus hadronischen Zerfällen, erreicht die elektromagnetischen Komponente aus Myonen-Zerfällen den Detektor (siehe Abb. 17).

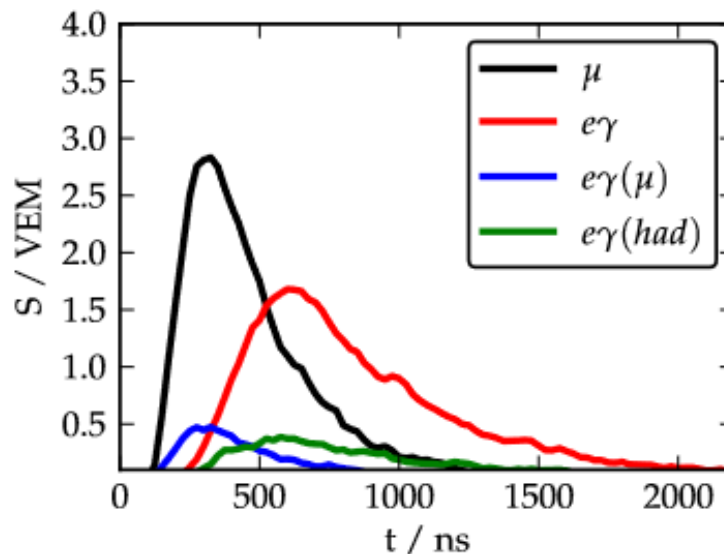


Abbildung 17: Simulierte Signale der einzelnen Komponenten eines Schauers in einem der Photomultipliern (18).

Dementsprechend sind die Daten, die für die nachfolgende Analyse verwendet werden, generierte Landau Verteilungen, mit festen Parametern, auf die eine Signalfuktuation und ein Untergrundrauschen addiert wurden. Dabei ist die Form der Landau Verteilung

mithilfe von drei Parameter anpassbar. Die drei Parameter sind: *width*, die Breite der Landau Verteilung, *amp*, die maximale Amplitude und *pos*, die Position des Maximums der Verteilung. Dabei wurde die Landau-Funktion wie folgt implementiert:

$$Landau(x) = amp \cdot e^{\frac{1}{width}} \cdot e^{\frac{-1}{width} \cdot ((x-pos) + e^{-1 \cdot (x-pos)})} \quad (26)$$

Ein Beispiel für die generierten Verteilungen ist in Abbildung 18 zu finden. Dabei ist auf der Y-Achse die Amplitude aufgetragen und auf der X-Achse die Zeit, die in einzelne Bins eingeteilt ist. Wobei jeder Bin ein Zeitfenster von 25 ns darstellt, entsprechend der Taktfrequenz der Ausleseelektronik.

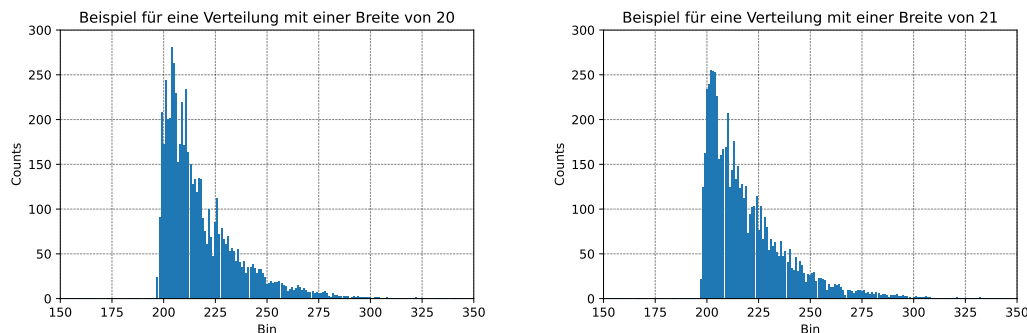


Abbildung 18: Zwei Beispiele für die verwendeten generierten Verteilungen. Einmal mit einer Breite von 20 und einmal mit einer Breite von 21.

4.2. Vorgehensweise

Die nachfolgende Analyse besteht aus zwei Teilen. Im ersten Teil soll ein Neuronales Netz entwickelt werden, welches die beiden oben beschriebenen Verteilungsklassen differenzieren können soll. Danach wird das Neuronale Netz hinsichtlich verschiedener Hyperparameter optimiert. Im zweiten Teil soll geprüft werden was die Grenzen des im ersten Teil optimierten Netzes sind. Wie ähnlich die beiden Verteilungsklassen sein dürfen, damit sie noch unterschieden werden können.

Die Analyse wird vollständig in Python 3.8.12 (25) durchgeführt. Der Generator für die Landau Verteilungen wurde freundlicherweise von Dr. Philip Rühl zur Verfügung gestellt. Für das Neuronale Netz wird das Python Paket tensorflow in der Version 2.6.0 verwendet (26). Trainiert wird das Neuronale Netz auf einer GPU. Einer NVIDIA RTX 3070 mit der Treiberversion 512.15 und CUDA 11.6. Der zweite Teil der Analyse wird auf der selben Hardware, aber mit neueren Treibern, Version 536.67 und CUDA 12.2 durchgeführt. Für die Datenauswertung werden die Python Pakete numpy 1.21.1 (27), pandas 1.3.3 (28) & matplotlib 3.4.3 (29) verwendet.

4.3. Aufbau des verwendeten NN

Das verwendete Neuronale Netz wird durch verschiedene Hyperparameter definiert. Dies sind Parameter, die festlegen, wie das Neuronale Netz aufgebaut ist und wie es sich beim Trainieren verhält. Die Hyperparameter werden vor dem Trainieren festgelegt und ändern sich währenddessen nicht. Im ersten Teil der Analyse werden verschiedene Hyperparameter variiert, um das bestmögliche Neuronale Netz zu finden. Einige der Hyperparameter bleiben aber auch unverändert. So folgt in jedem NN nach den Convolutional-Layers ein Global Average Pooling Layer und als Output ein Fully Connected Layer mit der Softmax Aktivierungsfunktion und zwei Neuronen, für die beiden möglichen Klassen. Der Vorteil der Softmax Funktion ist, dass sie eine Wahrscheinlichkeitsverteilung über alle Output-Klassen z_i erzeugt. Dabei ist die Softmax Funktion definiert als:

$$\text{softmax}(z)_i = \exp(z_i) / \sum_{k=1}^n \exp(z_k)$$

So ist die Summe über alle Klassen 1 und man erhält die Wahrscheinlichkeiten, dass ein Input zur Klasse i gehört. Außerdem wird als Loss-Funktion die Categorical-Crossentropy genommen, da es sich um ein Klassifizierungsprobleme mit mehreren Klassen handelt. Als Optimierungsfunktion wird der ADAM (Adaptive Moment Estimation) Algorithmus verwendet, da sich dieser als (fast immer) gute Wahl für Neuronale Netze erwiesen hat (23). Der ADAM Algorithmus verwendet eine learning rate von 0.001. Diese bleibt auch gleich, außer es bildet sich beim Lernen ein Plateau aus, dann wird die learning rate schrittweise auf ein Zehntel des ursprünglichen Wertes reduziert. Dabei beschreibt die learning rate wie groß die Schritte sind, die die Optimierungsfunktion macht. Ist die learning rate zu klein macht das Trainieren nur langsame Fortschritte und man läuft Gefahr in einem lokalen Minimum hängen zu bleiben. Ist die learning rate zu groß macht man zwar sehr schnell Fortschritte beim Trainieren, läuft aber Gefahr den idealen Punkt zu überschreiten (siehe Abb. 12).

Neben den Hyperparametern, die unverändert bleiben gibt es die, die in dieser Analyse untersucht werden. Dazu gehören: Die Anzahl an trainierten Epochen, die festlegt, wie oft der ganze Trainingsdatensatz beim Trainieren durchlaufen wird. Die Größe der batch-size, die bestimmt, wie viele Daten auf einmal, während eines forwardpass durch das NN propagiert werden. Eine größere batch-size führt typischerweise zu einem schnelleren Trainieren, erhöht aber den währenddessen notwendigen Speicherbedarf. Ein weiterer untersuchter Parameter ist der Validation Split. Dieser bestimmt, wie viele Daten aus dem Trainingsdatensatz während des Trainierens für das eigentliche Training und wie viele für das Validieren des Trainings verwendet werden. Der Trainingsdatensatz wird nach jeder Epoche erneut entsprechend des Validations Splits aufgeteilt. Hinzukommen drei Hyperparameter für den Aufbau des Convolutional Teils des Neuronalen Netzes. Der erste davon ist die Tiefe des Neuronalen Netzes, also wie viele Ebenen das Neuronale Netz besitzt. Je mehr Ebene das NN hat, desto besser sollte die Genauigkeit der Klassifizierung werden, im Gegenzug geht dafür die Trainingsdauer hoch. Jedes Convolutional Layer hat zwei Hyperparameter, die untersucht werden. Als erstes die Anzahl an Filtern, diese bestimmt, wie viele Merkmale ein Convolutional Layer aus dem Datensatz extrahieren

kann. Als zweiter Parameter, die Größe dieser Filter, bzw. die Länge dieser extrahierten Merkmale, da hier eindimensionale Convolutional Layer verwendet werden.

Diese Hyperparameter werden hinsichtlich drei verschiedener Merkmale untersucht. Zum einen die runtime, also wie lange das Trainieren des Neuronalen Netzes gedauert hat. Zum anderen, wie gut das fertig trainierte Neuronale Netz einen Testdatensatz klassifizieren kann, auf das das Neuronale Netz während des Trainings keinen Zugriff hatte. Dazu wird die Genauigkeit und der Loss des Neuronalen Netzes auf dem Testdatensatzes ermittelt. Dabei beschreibt der Loss die mathematische Abweichung von den wahren Werten zu den vom Neuronalen Netz vorhergesagten Werten.

Der gesamte Datensatz umfasst 20000 Landau Verteilungen, jeweils 10000 von jeder der beiden Klassen. Hiervon werden 18000 für das Trainieren des Neuronalen Netzes verwendet. Dieser Teil wird dann während des Trainings in Trainings- und Validierungsdaten aufgeteilt. Die anderen 2000 Daten sind der Testdatensatz, dieser wird nach dem Training einmal verwendet, um das Neuronale Netz zu testen. Von den sechs zu untersuchenden Parametern wird jeweils einer variiert, während die anderen unverändert bleiben. Die Standardparameter sind in [Abbildung 19](#) aufgeführt.

Epochs	batch_size	val_split	depth	filters	kernel_size
200.0	16.0	0.1	5.0	64.0	3.0

Abbildung 19: Standard Parameter des Neuronalen Netzes.

5. Auswertung

5.1. Optimierung des Neuronalen Netzes

5.1.1. Batch Größe

Mit zunehmender Größe der Batch Größe wird ein leichter Anstieg des Testloss und eine Verringerung der Testgenauigkeit erwartet. Dies lässt sich dadurch erklären, dass mit zunehmender Größe der Batch Größe weniger oft die Parameter im Neuronalen Netz via Backpropagation angepasst werden. Da die Backpropagation sehr rechenintensiv ist, wird erwartet, dass im gleichen Maße, wie die Batch Größe erhöht wird, die Trainingszeit abnimmt. Der Testloss und die Testgenauigkeit verhalten sich wie erwartet: bei kleinen Batch sizes gibt es kaum merkliche Unterschiede. Erst ab einer gewissen Batch Größe nimmt der Testloss nachweisbar zu und die Testgenauigkeit ab. Für die Trainingszeit wird erwartet, dass mit einer Verdoppelung der Batch Größe sich die Trainingszeit halbiert. Dies trifft auch für kleine Batch Größen zu. Für große Batch Größen ist zu beobachten, dass sich die Trainingszeit nicht mehr halbiert bei einer Verdoppelung der Batch Größe. Dies lässt sich dadurch erklären, dass hier nicht mehr die Backpropagation für den Großteil der Trainingszeit verantwortlich ist, sondern die Forwardpropagation, welche unabhängig von der Batch Größe ist. Siehe Grafik 20. Hierbei wurden für jeden Messwert drei Trainingsdurchläufe durchgeführt und Mittelwert und Standardabweichung aufgetragen. Kleinere Abweichungen lassen sich dadurch erklären, dass zwar identische Datensätze verwendet werden, jedoch die Parameter vom Neuronalen Netz jedesmal zufällig initialisiert werden.

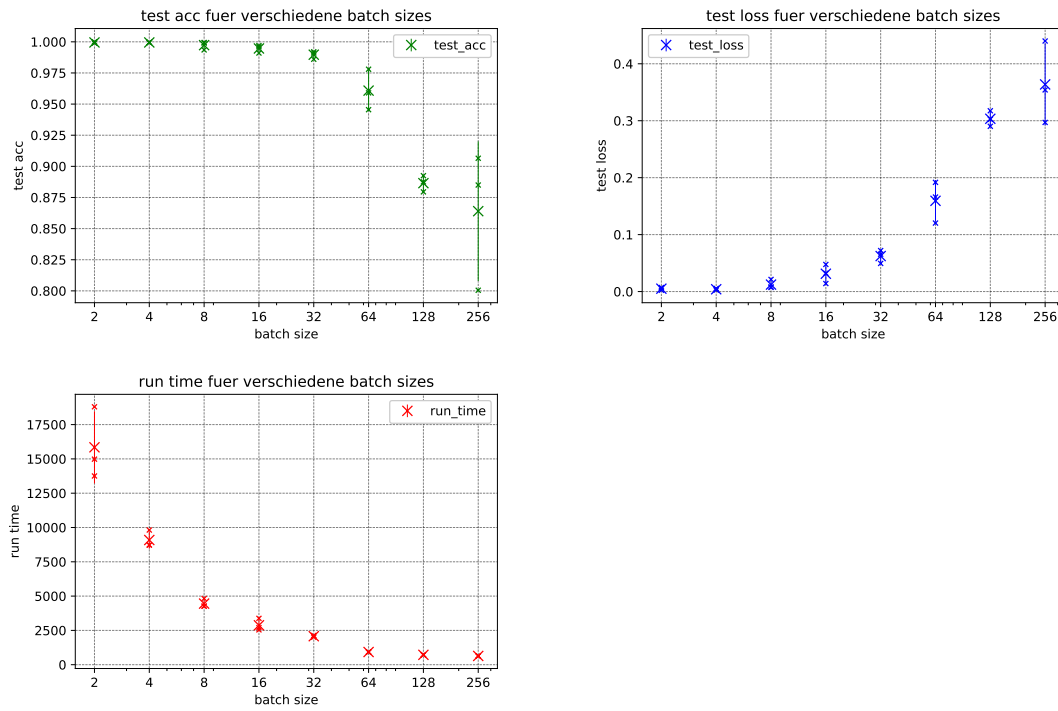


Abbildung 20: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Batch Größen.

5.1.2. Anzahl Epochen

Für die kleinste Anzahl an Trainings Epochen wird ein hoher Testloss und eine Testgenauigkeit von etwas über 0.5 erwartet. Mit zunehmender Anzahl an Epochen sollte der Testloss asymptotisch gegen 0 und die Testgenauigkeit sollte asymptotisch gegen 1 gehen. Für die Trainingszeit ist zu erwarten, dass diese linear mit der zunehmenden Anzahl an Epochen auch länger wird. Dies lässt sich auch bestätigen, so dass für die Trainingszeit einer einzelnen Epoche folgt, dass diese 11 Sekunden lang ist. Bei der Testgenauigkeit sieht man, dass diese nach ca. 50 Epochen bereits ihr Maximum von über 99% Genauigkeit erreicht, der Testloss jedoch noch um das Drei- bis Vierfache weiter abfällt, bis dieser ein Minimum von 0.03 bei 200 Epochen erreicht (siehe Abb. 21, 22).

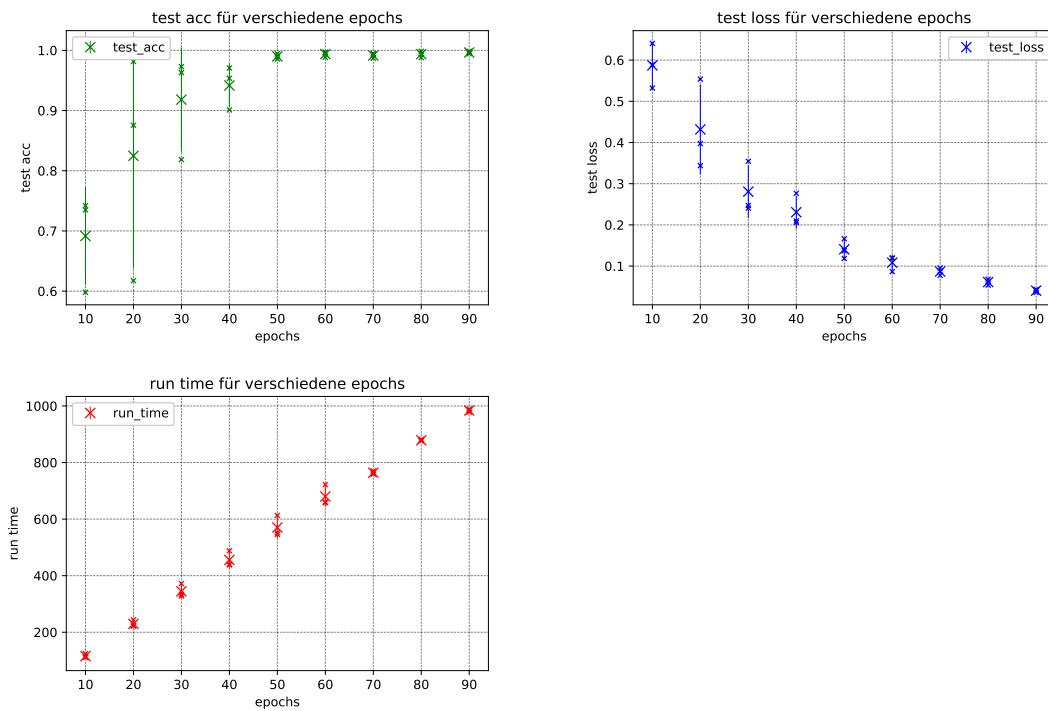


Abbildung 21: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Anzahl an Trainingsdurchläufen.

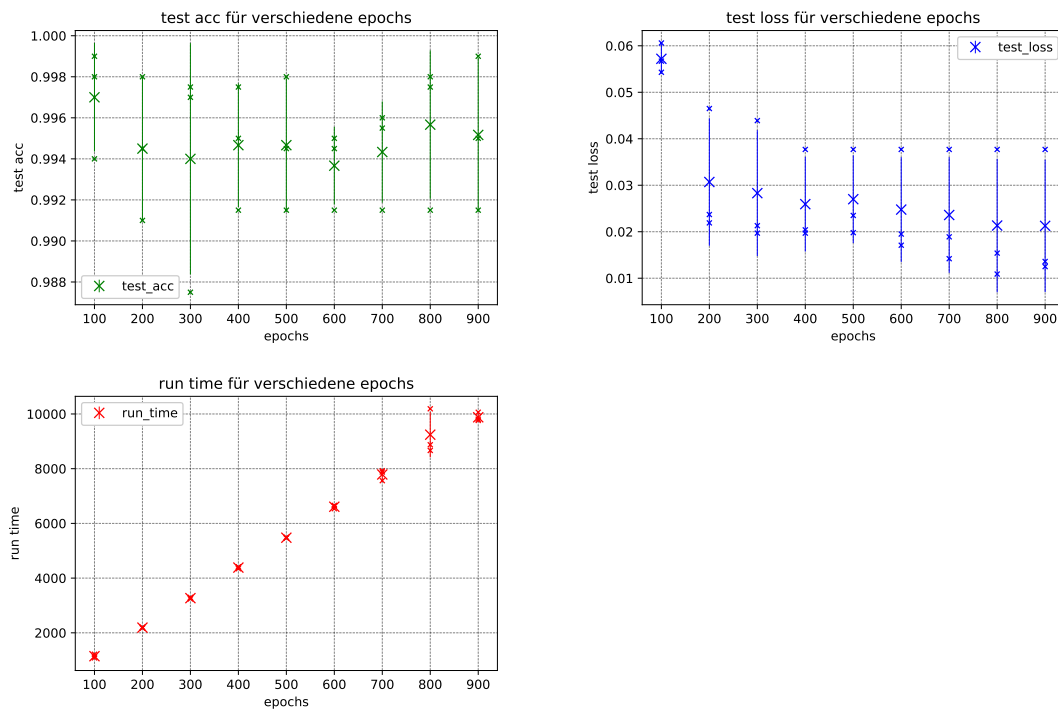


Abbildung 22: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Anzahl an Trainingsdurchläufen.

5.1.3. Validation Split

Für den Validation Split ist zu erwarten, dass ab einer gewissen Größe des Validierungsdatensatzes nicht mehr genug Daten im Trainingsdatensatz bleiben, um das Neuronale Netz adäquat zu trainieren. Dementsprechend sollte die Testgenauigkeit fallen und der Testloss steigen. Die Trainingszeit sollte linear mit größer werdendem Validation Split fallen, da die Validierungsdaten weiterhin die Forwardpropagation durchlaufen, aber nicht mehr Teil der Backpropagation sind. Dies zeigen auch die aufgenommenen Daten. Die Große Varianz der Trainingszeit bei einigen Messungen lässt sich dadurch erklären, dass während dieser Durchläufe an dem Computer gearbeitet wurde und somit nicht die volle Rechenkapazität für das Trainieren des NN verwendet werden konnte. Der Testloss und die Testgenauigkeit verhalten sich anders als erwartet. Beide Werte bleiben über sämtliche Validation Splits konstant. Daraus lässt sich schlussfolgern, dass auch 10 % der Daten ausreichen um effektiv die beiden Klassen voneinander zu differenzieren (siehe Abb. 23).

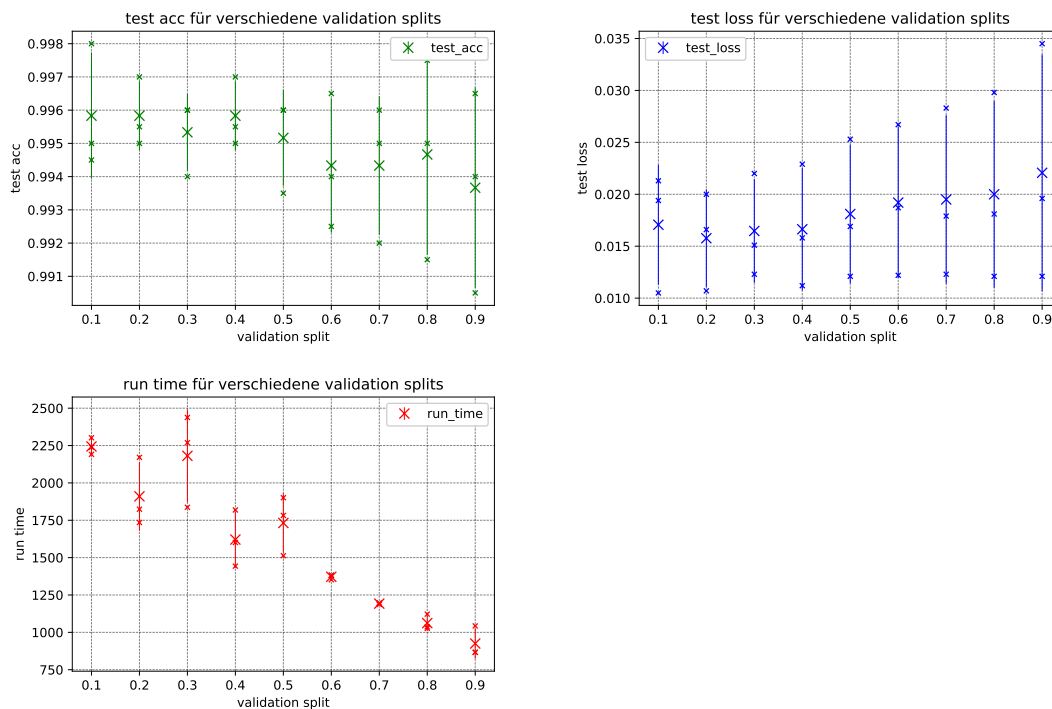


Abbildung 23: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Aufteilungen von Trainingsdaten und Validierungsdaten.

5.1.4. Kernel Größen

Bei der Kernel Größe ist zu erwarten, dass ein Kernel der Größe Eins nicht in der Lage ist Muster zu erkennen, somit sollte hierfür eine hoher Testloss und eine niedrige Testgenauigkeit zu erwarten sein. Da es in den Daten keine größeren zusammenhängende Strukturen gibt, sollte eine relativ kleine Kernel Größe ein ideales Ergebnis erzielen. Mit zunehmender Kernel Größe sollte danach das Ergebnis zunehmend schlechter werden, da das NN weniger gut relevante Punkte differenzieren kann. Da mit größer werdender Kernel Größe auch die Anzahl an trainierbaren Parametern zunimmt, sollte die Trainingszeit auch leicht ansteigen. Wie zu erwarten liefert ein Kernel mit einer Länge von eins kein gutes Ergebnis. Dennoch ist die Testgenauigkeit mit ca. 87 % überraschend hoch. Die Kernel Größen zwei und drei liefern ein gleich gutes Ergebnis. Für größere Kernel Größen deutet sich der erwartete schlechter werdende Trend an. Die Trainingszeit nimmt mit zunehmender Kernel Größe zu. Auch hier wurde zwischenzeitlich an dem Computer gearbeitet, so dass die Trainingszeit nicht ganz so schön linear ist, wie zu erwarten war (siehe Abb. 24).

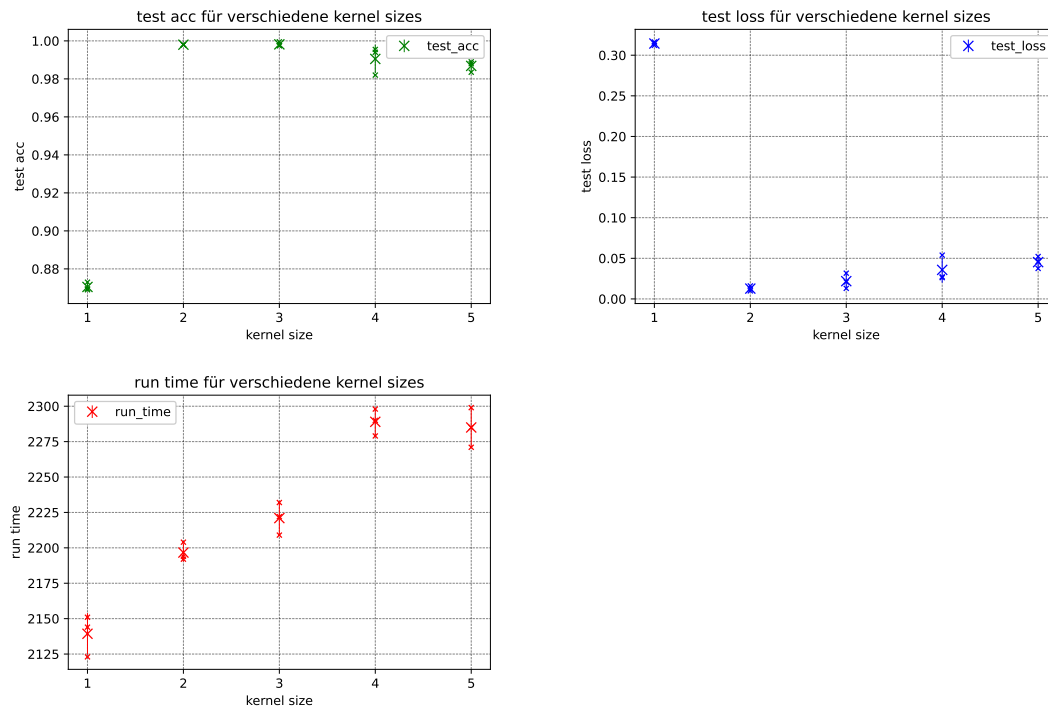


Abbildung 24: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Kernel Größen.

5.1.5. Anzahl an Filtern

Für die Anzahl an Filtern ist zu erwarten, dass mit zunehmender Anzahl an Filtern der Testloss und die Testgenauigkeit wieder asymptotisch gegen ihre ideal Werte streben. Da mit zunehmender Anzahl an Filtern die Anzahl an trainierbaren Parametern exponentiell steigt ist auch zu erwarten, dass die Trainingszeit exponentiell ansteigt. Hier verhält sich das Neuronale Netz wie erwartet. Mit zunehmender Anzahl an Filtern steigt die Testgenauigkeit und der Testloss sinkt. Selbst bei einer Filteranzahl von vier liefert das NN überraschend gute Ergebnisse mit einer Testgenauigkeit von ca. 96%. Auch die Trainingszeit nimmt, wie zu erwarten, exponentiell zu (siehe Abb. 25).

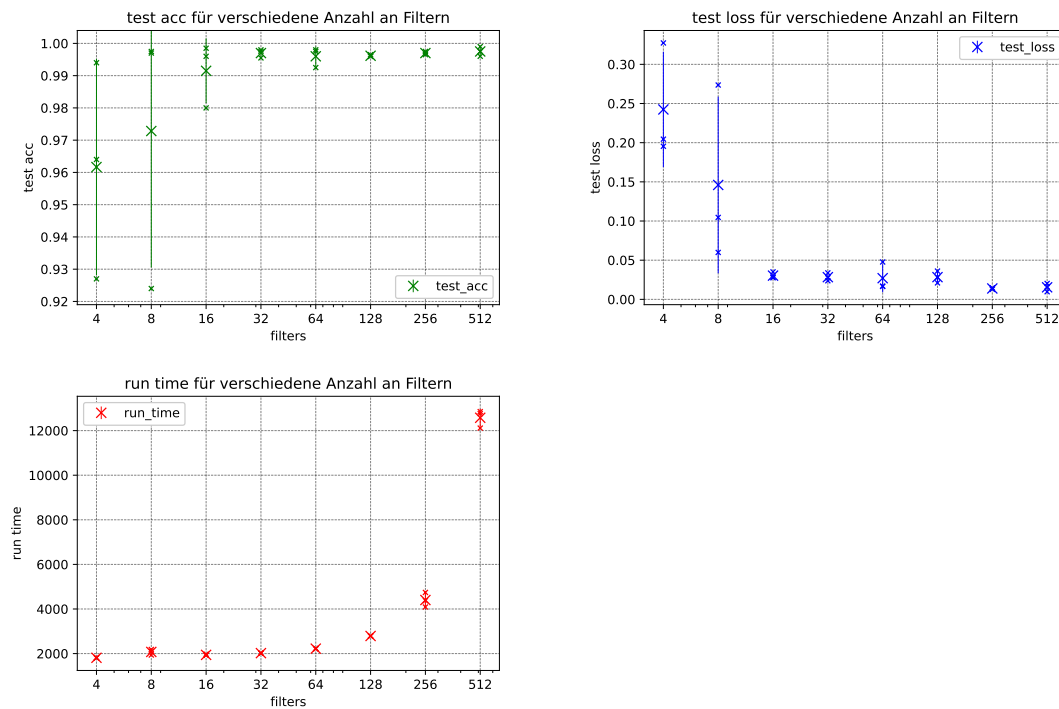


Abbildung 25: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Anzahl an Filtern.

5.1.6. Tiefe des NN

Auch für die Tiefe des Neuronalen Netzes ist zu erwarten, dass der Testloss für eine zunehmende Anzahl an Schichten gegen 0 geht und die Testgenauigkeit gegen 1 geht. Die Trainingszeit sollte linear mit der Anzahl an Schichten zusammenhängen, da jede Schicht aus der gleichen Anzahl an Parametern besteht. Der Testloss verhält sich genau so wie erwartet und auch die Trainingszeit verhält sich wie erwartet sehr linear. Bei der Testgenauigkeit fällt auf, dass diese bei einem relativ niedrigem Wert anfängt, bei ca. 67% und dann sehr steil gegen 1 geht. Ab einer Tiefe von drei Schichten ist keine nennenswerte Verbesserung mehr möglich, da die Testgenauigkeit hier bereits ihr Maximum von ca. 95.5% Genauigkeit erreicht hat. Dies bedeutet also, dass auch für ein sehr simples Problem wie dieses hier, eine gewisse Mindestanzahl an Ebenen in einem Neuronalen Netz notwendig sind, für eine erfolgreiche Klassifizierung (siehe Abb. 26).

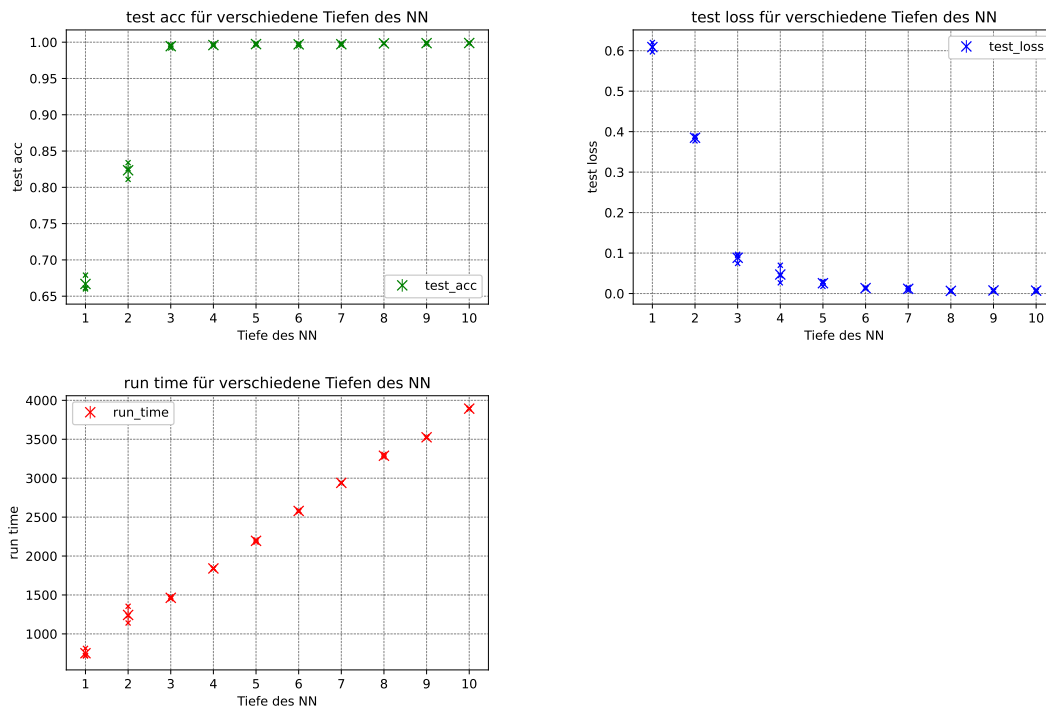


Abbildung 26: Testgenauigkeit, Testloss und Trainingszeit des NN in Sekunden für verschiedene Anzahl an Ebenen des NN.

5.2. Grenzen des Neuronalen Netzes

Um die Grenzen des Neuronalen Netzes auszureizen, wurde die eine Verteilung stückweise an die andere Verteilung angeglichen. Dazu wurden die drei Parameter, die verwendet wurden, um die Landau-Verteilungen zu generieren, nacheinander angepasst.

Zuerst wurde der Parameter, welcher die Breite der Landau Verteilung bestimmt, variiert. Wie in Abbildung (27 c) zu sehen ist, liegen die Mittelwerte pro Bin der beiden Verteilungen sehr dicht beieinander, so dass es einem menschlichen Betrachter schwerfällt, die beiden einzelnen Verteilungen (Siehe Abbildung 27 d) voneinander zu differenzieren. Hier schafft es das Neuronale Netz dennoch in knapp 95% der Fälle, die Verteilungen richtig den beiden Klassen zuzuordnen. Werden die Verteilungen nun noch ähnlicher zueinander, kommt auch das Neuronale Netz an seine Grenzen.

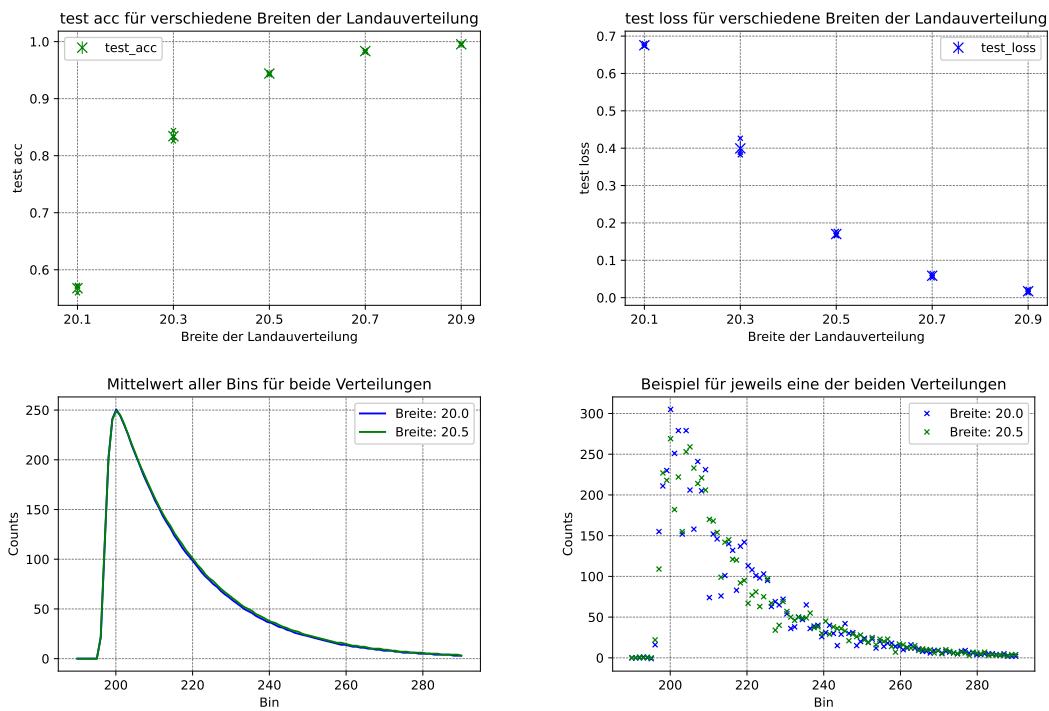


Abbildung 27: Testgenauigkeit, Testloss des NN für verschiedene Breiten der Landau-Verteilungen. Mittelwert pro Bin für beide Verteilungen und Beispiel jeweils eines Datensatzes.

Als zweites wurde die Amplitude der einen Verteilung verändert. Auch hier ist der Unterschied zwischen den beiden Verteilungen recht gering, aber für einen Menschen ist der Unterschied etwas einfacher zu erkennen als bei den verschiedenen Breiten der Verteilungen. Wird nun die Differenz zwischen den Maxima der Amplituden größer, fällt auch dem Neuronalen Netz die Differenzierung einfacher (siehe Abb. 28).

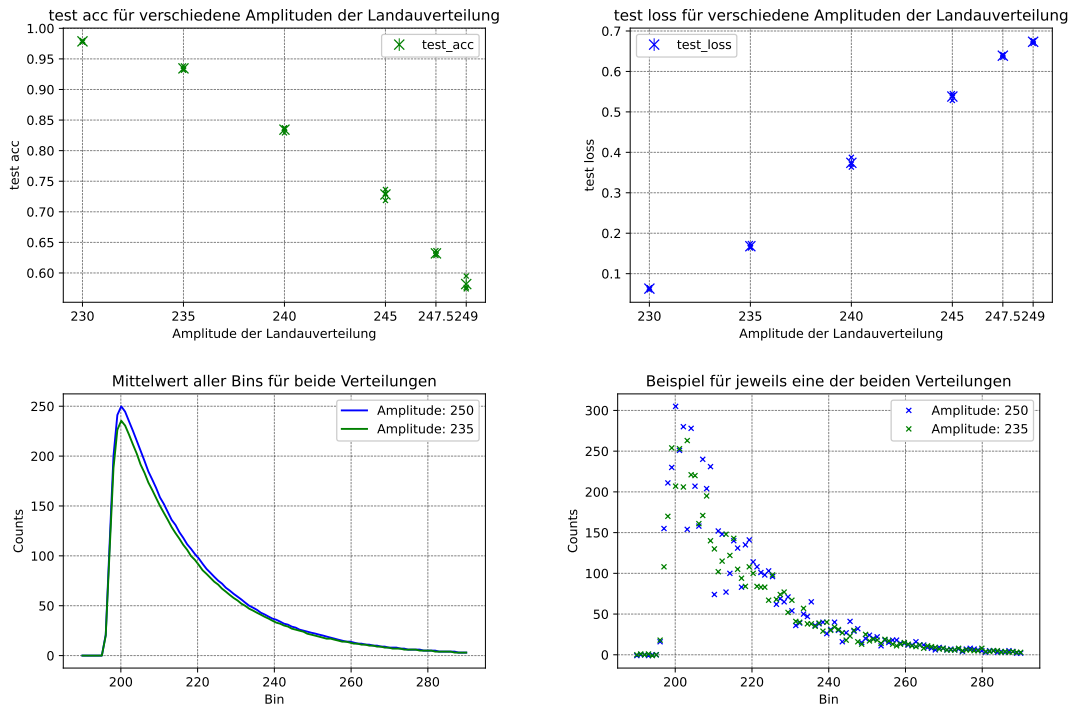


Abbildung 28: Testgenauigkeit, Testloss des NN für verschiedene Amplituden der Landau-Verteilungen. Mittelwert pro Bin für beide Verteilungen und Beispiel jeweils eines Datensatzes.

Als letzter Parameter der Landau Verteilung wurde die Position des Mittelwertes verschoben. Wie in Abbildung (29 c & d) zu sehen, sind diese sehr einfach mit dem Auge zu unterscheiden. Somit ist zu erwarten, dass das Neuronale Netz keine Probleme hat, die Verteilungen zu differenzieren. Diese Annahme stellt sich aber überraschenderweise als falsch heraus. Das Neuronale Netz hat sehr wohl ein Problem damit, die beiden Verteilungen voneinander zu unterscheiden. Selbst wenn die beiden Verteilungen soweit auseinanderliegen, dass sie sich fast gar nicht mehr überlappen, schafft es das Neuronale Netz nicht, diese ansonsten gleichen Verteilungen erfolgreich zu unterscheiden. Erklären lässt sich dies dadurch, dass ein reines Convolutional Neuronales Netz verwendet wird. Bei diesem versuchen die Kernels in jeder Ebene wiederkehrende Muster in den Daten zu erkennen, dafür laufen die Kernels jeweils über den kompletten Datenbereich. Somit können die Kernel minimale Unterschiede zwischen zwei Formen erkennen, jedoch lernen die Kernels nichts über die Position, an der sie eine Form gefunden haben. Dadurch hält das Neuronale Netz zwei gleiche Verteilungen für identisch, auch wenn diese zeitlich zueinander verschoben sind.

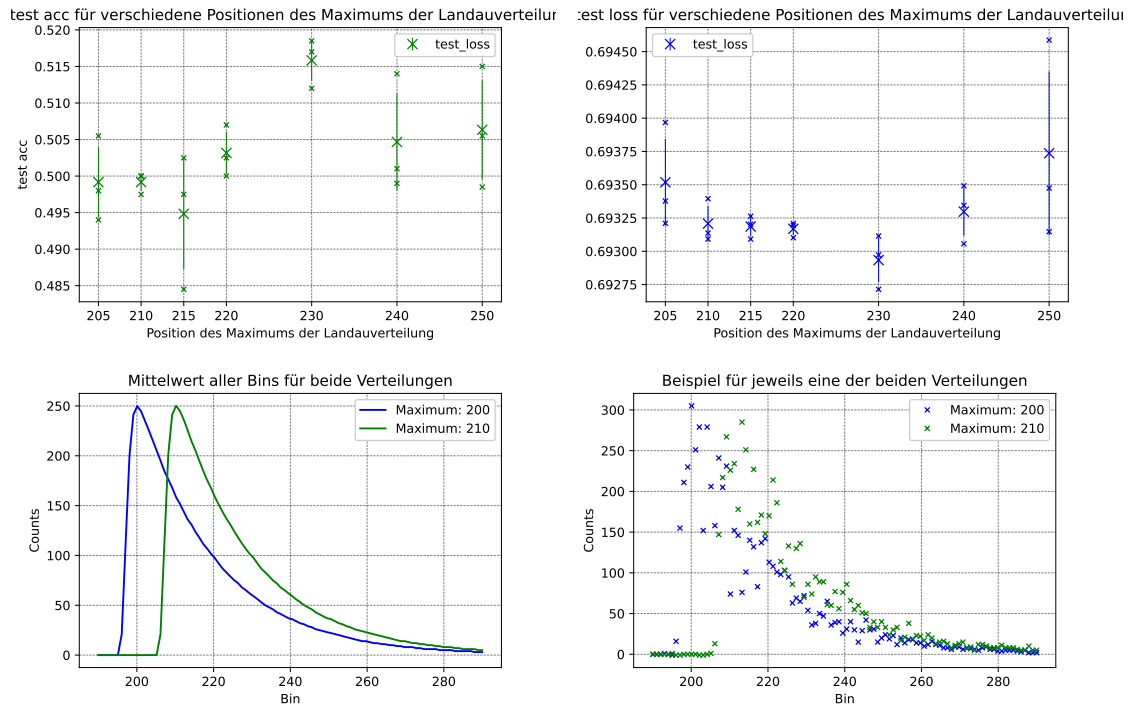


Abbildung 29: Testgenauigkeit, Testloss des NN für verschiedene Positionen des Maximum der Landau-Verteilungen. Mittelwert pro Bin für beide Verteilungen und Beispiel jeweils eines Datensatzes.

Auch bei starkem Erhöhen der Signalfuktuation und des Untergrundrauschens kann das Neuronale Netz immer noch ohne Probleme mit einer Genauigkeit von über 99 % die beiden Verteilungen unterscheiden (siehe Abb. 30).

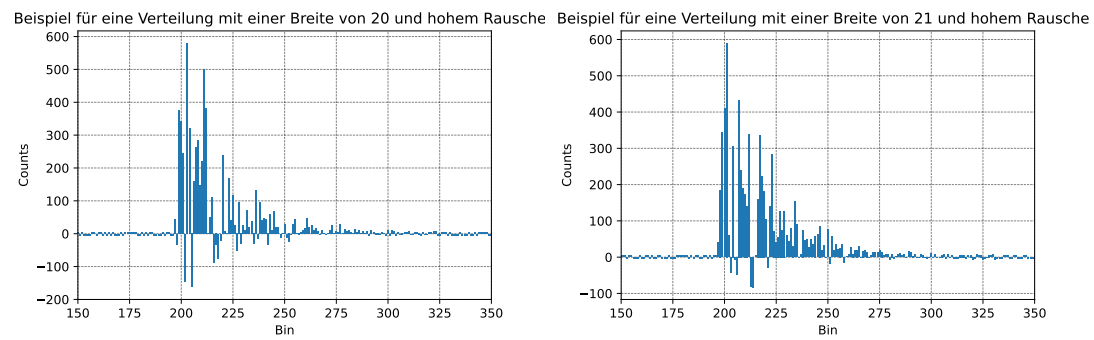


Abbildung 30: Beispiele für zwei Verteilungen mit hoher Signalfuktuation und höherem Untergrundrauschen.

6. Zusammenfassung & Ausblick

In dieser Analyse wurde untersucht, wie ähnlich zwei generierte Verteilungen zueinander sein können, damit diese von einem Convolutional Neuronales Netz sicher differenziert werden können. Hierfür wurde das Verhalten des Neuronales Netzes bei Änderung einzelner Hyperparameter untersucht. Allerdings konnten bei dieser Analyse längst nicht alle möglichen Hyperparameter untersucht werden. So wurde beispielsweise eine Änderung der Parameter der Optimierungsfunktion gar nicht untersucht. Außerdem wurde ein reines Convolutional Neuronales Netz verwendet. Oftmals kann ein Neuronales Netz noch bessere Ergebnisse erzielen, wenn auf den Convolution Teil des Neuronales Netzes ein Fully Connected Teil folgt. Auch wurden keine Recurrent Neuronales Netze untersucht, obwohl diese auch sehr gut für das Analysieren von zeitlich zusammenhängenden Daten geeignet sind. Das hängt unter anderem damit zusammen, dass RNNs im Gegensatz zu CNNs schlecht parallelisierbar sind und somit das Trainieren sehr viel länger dauern kann. Da auch so die reine Rechenzeit für die Analyse sich auf fast 200 h beläuft, konnten nur einige Hyperparameter untersucht werden. Dennoch hat der zweite Teil der Analyse gezeigt, dass das Convolutional Neuronales Netz sehr gut die zwei Verteilungen voneinander differenzieren kann, selbst dann, wenn diese sehr ähnlich zueinander sind. Als problematisch erweist sich, dass das Neuronales Netz nicht in der Lage ist zwei Verteilungen zu differenzieren, wenn diese an sich identisch sind, aber zueinander zeitlich verschoben sind. Dies ist ein Problem, da wie aus 17 zu entnehmen ist, die einzelnen Komponente eines Schauers mit einer gewissen zeitlichen Verzögerung den Detektor erreichen. Dieses Problem muss behoben werden, bevor dieser Ansatz weiter verfolgt werden kann. Abgesehen von diesem Punkt hat sich dieser Ansatz aber als vielversprechend erwiesen und als gute Grundlage, um darauf aufzubauen. Danach wäre der nächste Schritt, dass Neuronales Netz mithilfe von Monte Carlo Simulationen zu trainieren und zu optimieren, um dieses so entwickelte Neuronales Netz danach auf echten Daten zu testen.

A. Anhang

A.1. Tabelle mit den Ergebnissen des ersten Teiles der Analyse

NN	epochs	batch size	val split	depth	filters	kernel size	samples	run time	test loss	test acc
1	200	256	0.1	5	64	3	20000	646	0.4400	0.8005
1	200	256	0.1	5	64	3	20000	633	0.3541	0.8850
1	200	256	0.1	5	64	3	20000	634	0.2970	0.9065
2	200	128	0.1	5	64	3	20000	715	0.2902	0.8875
2	200	128	0.1	5	64	3	20000	725	0.3175	0.8795
2	200	128	0.1	5	64	3	20000	719	0.3026	0.8925
3	200	64	0.1	5	64	3	20000	925	0.1918	0.9455
3	200	64	0.1	5	64	3	20000	923	0.1203	0.9780
3	200	64	0.1	5	64	3	20000	924	0.1660	0.9590
4	200	32	0.1	5	64	3	20000	2093	0.0656	0.9910
4	200	32	0.1	5	64	3	20000	2131	0.0494	0.9920
4	200	32	0.1	5	64	3	20000	2004	0.0720	0.9860
5	200	16	0.1	5	64	3	20000	3378	0.0476	0.9910
5	200	16	0.1	5	64	3	20000	2701	0.0322	0.9960
5	200	16	0.1	5	64	3	20000	2538	0.0141	0.9970
6	200	8	0.1	5	64	3	20000	4261	0.0082	0.9995
6	200	8	0.1	5	64	3	20000	4274	0.0072	0.9990
6	200	8	0.1	5	64	3	20000	4825	0.0213	0.9935
7	200	4	0.1	5	64	3	20000	9815	0.0026	0.9995
7	200	4	0.1	5	64	3	20000	8722	0.0055	0.9995
7	200	4	0.1	5	64	3	20000	8727	0.0040	0.9995
8	200	2	0.1	5	64	3	20000	18797	0.0071	0.9990
8	200	2	0.1	5	64	3	20000	14981	0.0023	1.0000
8	200	2	0.1	5	64	3	20000	13755	0.0058	0.9995
9	200	16	0.1	5	64	3	20000	2302	0.0105	0.9980
9	200	16	0.1	5	64	3	20000	2239	0.0213	0.9950
9	200	16	0.1	5	64	3	20000	2191	0.0194	0.9945
10	200	16	0.2	5	64	3	20000	1735	0.0107	0.9970
10	200	16	0.2	5	64	3	20000	1824	0.0200	0.9955
10	200	16	0.2	5	64	3	20000	2171	0.0166	0.9950
11	200	16	0.3	5	64	3	20000	1837	0.0123	0.9960
11	200	16	0.3	5	64	3	20000	2269	0.0220	0.9940
11	200	16	0.3	5	64	3	20000	2438	0.0151	0.9960
12	200	16	0.4	5	64	3	20000	1443	0.0112	0.9970
12	200	16	0.4	5	64	3	20000	1819	0.0229	0.9955
12	200	16	0.4	5	64	3	20000	1601	0.0158	0.9950

13	200	16	0.5	5	64	3	20000	1513	0.0121	0.9960
13	200	16	0.5	5	64	3	20000	1901	0.0253	0.9935
13	200	16	0.5	5	64	3	20000	1782	0.0169	0.9960
14	200	16	0.6	5	64	3	20000	1352	0.0122	0.9965
14	200	16	0.6	5	64	3	20000	1381	0.0267	0.9925
14	200	16	0.6	5	64	3	20000	1383	0.0187	0.9940
15	200	16	0.7	5	64	3	20000	1197	0.0123	0.9960
15	200	16	0.7	5	64	3	20000	1193	0.0283	0.9920
15	200	16	0.7	5	64	3	20000	1187	0.0179	0.9950
16	200	16	0.8	5	64	3	20000	1122	0.0121	0.9975
16	200	16	0.8	5	64	3	20000	1037	0.0298	0.9915
16	200	16	0.8	5	64	3	20000	1027	0.0181	0.9950
17	200	16	0.9	5	64	3	20000	1043	0.0121	0.9965
17	200	16	0.9	5	64	3	20000	866	0.0345	0.9905
17	200	16	0.9	5	64	3	20000	867	0.0196	0.9940
18	100	16	0.1	5	64	3	20000	1084	0.0567	0.9940
18	100	16	0.1	5	64	3	20000	1217	0.0606	0.9990
18	100	16	0.1	5	64	3	20000	1134	0.0543	0.9980
19	200	16	0.1	5	64	3	20000	2173	0.0465	0.9910
19	200	16	0.1	5	64	3	20000	2200	0.0219	0.9980
19	200	16	0.1	5	64	3	20000	2198	0.0237	0.9945
20	300	16	0.1	5	64	3	20000	3241	0.0439	0.9875
20	300	16	0.1	5	64	3	20000	3296	0.0197	0.9975
20	300	16	0.1	5	64	3	20000	3265	0.0213	0.9970
21	400	16	0.1	5	64	3	20000	4344	0.0377	0.9915
21	400	16	0.1	5	64	3	20000	4402	0.0197	0.9975
21	400	16	0.1	5	64	3	20000	4410	0.0204	0.9950
22	500	16	0.1	5	64	3	20000	5466	0.0377	0.9915
22	500	16	0.1	5	64	3	20000	5493	0.0198	0.9980
22	500	16	0.1	5	64	3	20000	5473	0.0235	0.9945
23	600	16	0.1	5	64	3	20000	6659	0.0377	0.9915
23	600	16	0.1	5	64	3	20000	6537	0.0171	0.9950
23	600	16	0.1	5	64	3	20000	6632	0.0195	0.9945
24	700	16	0.1	5	64	3	20000	7562	0.0377	0.9915
24	700	16	0.1	5	64	3	20000	7925	0.0142	0.9955
24	700	16	0.1	5	64	3	20000	7878	0.0189	0.9960
25	800	16	0.1	5	64	3	20000	8666	0.0377	0.9915
25	800	16	0.1	5	64	3	20000	10189	0.0109	0.9980
25	800	16	0.1	5	64	3	20000	8875	0.0154	0.9975
26	900	16	0.1	5	64	3	20000	9764	0.0377	0.9915
26	900	16	0.1	5	64	3	20000	9809	0.0136	0.9950
26	900	16	0.1	5	64	3	20000	10063	0.0125	0.9990
27	10	16	0.1	5	64	3	20000	124	0.6404	0.7350

27	10	16	0.1	5	64	3	20000	112	0.5886	0.5980
27	10	16	0.1	5	64	3	20000	111	0.5321	0.7420
28	20	16	0.1	5	64	3	20000	244	0.3974	0.8755
28	20	16	0.1	5	64	3	20000	222	0.3438	0.9815
28	20	16	0.1	5	64	3	20000	223	0.5538	0.6175
29	30	16	0.1	5	64	3	20000	372	0.2407	0.9630
29	30	16	0.1	5	64	3	20000	328	0.2473	0.9730
29	30	16	0.1	5	64	3	20000	334	0.3544	0.8185
30	40	16	0.1	5	64	3	20000	488	0.2767	0.9010
30	40	16	0.1	5	64	3	20000	441	0.2058	0.9535
30	40	16	0.1	5	64	3	20000	438	0.2097	0.9705
31	50	16	0.1	5	64	3	20000	613	0.1664	0.9855
31	50	16	0.1	5	64	3	20000	546	0.1185	0.9945
31	50	16	0.1	5	64	3	20000	552	0.1374	0.9900
32	60	16	0.1	5	64	3	20000	722	0.0866	0.9960
32	60	16	0.1	5	64	3	20000	658	0.1189	0.9885
32	60	16	0.1	5	64	3	20000	660	0.1198	0.9970
33	70	16	0.1	5	64	3	20000	759	0.0878	0.9870
33	70	16	0.1	5	64	3	20000	770	0.0953	0.9925
33	70	16	0.1	5	64	3	20000	763	0.0777	0.9950
34	80	16	0.1	5	64	3	20000	877	0.0680	0.9880
34	80	16	0.1	5	64	3	20000	877	0.0539	0.9985
34	80	16	0.1	5	64	3	20000	882	0.0620	0.9945
35	90	16	0.1	5	64	3	20000	978	0.0433	0.9965
35	90	16	0.1	5	64	3	20000	985	0.0363	0.9945
35	90	16	0.1	5	64	3	20000	989	0.0416	0.9985
36	200	16	0.1	1	64	3	20000	716	0.6202	0.6615
36	200	16	0.1	1	64	3	20000	725	0.6095	0.6600
36	200	16	0.1	1	64	3	20000	814	0.5970	0.6790
37	200	16	0.1	2	64	3	20000	1355	0.3772	0.8340
37	200	16	0.1	2	64	3	20000	1235	0.3887	0.8110
37	200	16	0.1	2	64	3	20000	1140	0.3889	0.8255
38	200	16	0.1	3	64	3	20000	1461	0.0970	0.9965
38	200	16	0.1	3	64	3	20000	1477	0.0937	0.9920
38	200	16	0.1	3	64	3	20000	1452	0.0744	0.9950
39	200	16	0.1	4	64	3	20000	1844	0.0699	0.9950
39	200	16	0.1	4	64	3	20000	1843	0.0265	0.9975
39	200	16	0.1	4	64	3	20000	1835	0.0446	0.9955
40	200	16	0.1	5	64	3	20000	2199	0.0296	0.9985
40	200	16	0.1	5	64	3	20000	2208	0.0294	0.9960
40	200	16	0.1	5	64	3	20000	2179	0.0176	0.9975
41	200	16	0.1	6	64	3	20000	2586	0.0123	0.9980
41	200	16	0.1	6	64	3	20000	2570	0.0127	0.9950

41	200	16	0.1	6	64	3	20000	2581	0.0153	0.9980
42	200	16	0.1	7	64	3	20000	2941	0.0074	0.9985
42	200	16	0.1	7	64	3	20000	2939	0.0124	0.9970
42	200	16	0.1	7	64	3	20000	2940	0.0154	0.9960
43	200	16	0.1	8	64	3	20000	3287	0.0052	0.9985
43	200	16	0.1	8	64	3	20000	3306	0.0076	0.9980
43	200	16	0.1	8	64	3	20000	3272	0.0070	0.9985
44	200	16	0.1	9	64	3	20000	3521	0.0055	0.9995
44	200	16	0.1	9	64	3	20000	3523	0.0089	0.9975
44	200	16	0.1	9	64	3	20000	3532	0.0085	0.9990
45	200	16	0.1	10	64	3	20000	3892	0.0047	0.9990
45	200	16	0.1	10	64	3	20000	3899	0.0080	0.9990
45	200	16	0.1	10	64	3	20000	3889	0.0089	0.9985
46	200	16	0.1	5	64	1	20000	2144	0.3147	0.8690
46	200	16	0.1	5	64	1	20000	2123	0.3163	0.8730
46	200	16	0.1	5	64	1	20000	2151	0.3122	0.8695
47	200	16	0.1	5	64	2	20000	2194	0.0160	0.9980
47	200	16	0.1	5	64	2	20000	2192	0.0123	0.9980
47	200	16	0.1	5	64	2	20000	2204	0.0099	0.9980
48	200	16	0.1	5	64	3	20000	2232	0.0130	0.9995
48	200	16	0.1	5	64	3	20000	2222	0.0317	0.9975
48	200	16	0.1	5	64	3	20000	2209	0.0209	0.9975
49	200	16	0.1	5	64	4	20000	2298	0.0272	0.9955
49	200	16	0.1	5	64	4	20000	2279	0.0262	0.9940
49	200	16	0.1	5	64	4	20000	2290	0.0539	0.9820
50	200	16	0.1	5	64	5	20000	2271	0.0461	0.9890
50	200	16	0.1	5	64	5	20000	2299	0.0375	0.9880
50	200	16	0.1	5	64	5	20000	2285	0.0522	0.9835
51	200	16	0.1	5	4	3	20000	1802	0.3273	0.9270
51	200	16	0.1	5	4	3	20000	1814	0.2045	0.9640
51	200	16	0.1	5	4	3	20000	1820	0.1954	0.9940
52	200	16	0.1	5	8	3	20000	1936	0.2736	0.9240
52	200	16	0.1	5	8	3	20000	2185	0.1046	0.9975
52	200	16	0.1	5	8	3	20000	2103	0.0599	0.9970
53	200	16	0.1	5	16	3	20000	1977	0.0283	0.9985
53	200	16	0.1	5	16	3	20000	1924	0.0275	0.9800
53	200	16	0.1	5	16	3	20000	1926	0.0351	0.9960
54	200	16	0.1	5	32	3	20000	2017	0.0338	0.9955
54	200	16	0.1	5	32	3	20000	2018	0.0275	0.9980
54	200	16	0.1	5	32	3	20000	2026	0.0237	0.9975
55	200	16	0.1	5	64	3	20000	2219	0.0477	0.9925
55	200	16	0.1	5	64	3	20000	2233	0.0165	0.9975
55	200	16	0.1	5	64	3	20000	2214	0.0169	0.9980

56	200	16	0.1	5	128	3	20000	2798	0.0280	0.9960
56	200	16	0.1	5	128	3	20000	2781	0.0210	0.9960
56	200	16	0.1	5	128	3	20000	2789	0.0361	0.9965
57	200	16	0.1	5	256	3	20000	4089	0.0146	0.9965
57	200	16	0.1	5	256	3	20000	4375	0.0128	0.9975
57	200	16	0.1	5	256	3	20000	4743	0.0143	0.9970
58	200	16	0.1	5	512	3	20000	12852	0.0097	0.9990
58	200	16	0.1	5	512	3	20000	12765	0.0161	0.9960
58	200	16	0.1	5	512	3	20000	12113	0.0203	0.9975

A.2. Tabelle mit den Ergebnissen des zweiten Teiles der Analyse

NN	Land- au Breite	Land- au Ampli- tude	Land- au Positi- on	sgn- fluc	blfluc amp	blfluc prob	run time	test loss	test acc
59	20.1	250	200	0.2	1	0.3	2801	0.6799	0.5595
59	20.1	250	200	0.2	1	0.3	2563	0.6735	0.5710
59	20.1	250	200	0.2	1	0.3	2576	0.6730	0.5725
60	20.3	250	200	0.2	1	0.3	2627	0.3893	0.8325
60	20.3	250	200	0.2	1	0.3	2541	0.3823	0.8440
60	20.3	250	200	0.2	1	0.3	2561	0.4265	0.8260
61	20.5	250	200	0.2	1	0.3	2639	0.1657	0.9420
61	20.5	250	200	0.2	1	0.3	2563	0.1772	0.9445
61	20.5	250	200	0.2	1	0.3	2537	0.1679	0.9455
62	20.7	250	200	0.2	1	0.3	2854	0.0644	0.9815
62	20.7	250	200	0.2	1	0.3	2565	0.0582	0.9840
62	20.7	250	200	0.2	1	0.3	2555	0.0532	0.9845
63	20.9	250	200	0.2	1	0.3	2946	0.0128	0.9965
63	20.9	250	200	0.2	1	0.3	2566	0.0165	0.9955
63	20.9	250	200	0.2	1	0.3	2546	0.0221	0.9940
64	20.0	230	200	0.2	1	0.3	2636	0.0640	0.9785
64	20.0	230	200	0.2	1	0.3	2658	0.0651	0.9775
64	20.0	230	200	0.2	1	0.3	2540	0.0601	0.9800
65	20.0	235	200	0.2	1	0.3	3015	0.1745	0.9315
65	20.0	235	200	0.2	1	0.3	2547	0.1635	0.9385
65	20.0	235	200	0.2	1	0.3	2639	0.1657	0.9335
66	20.0	240	200	0.2	1	0.3	2641	0.3635	0.8370
66	20.0	240	200	0.2	1	0.3	3055	0.3876	0.8290
66	20.0	240	200	0.2	1	0.3	2557	0.3711	0.8370
67	20.0	245	200	0.2	1	0.3	2667	0.5283	0.7365
67	20.0	245	200	0.2	1	0.3	2551	0.5451	0.7300

67	20.0	245	200	0.2	1	0.3	3045	0.5403	0.7185
68	20.0	247	200	0.2	1	0.3	2669	0.6348	0.6365
68	20.0	247	200	0.2	1	0.3	2984	0.6400	0.6310
68	20.0	247	200	0.2	1	0.3	2552	0.6419	0.6285
69	20.0	249	200	0.2	1	0.3	2561	0.6746	0.5740
69	20.0	249	200	0.2	1	0.3	2642	0.6763	0.5765
69	20.0	249	200	0.2	1	0.3	3082	0.6692	0.5950
70	20.0	250	205	0.2	1	0.3	2676	0.6932	0.4940
70	20.0	250	205	0.2	1	0.3	2601	0.6940	0.5055
70	20.0	250	205	0.2	1	0.3	2556	0.6934	0.4980
71	20.0	250	210	0.2	1	0.3	2572	0.6931	0.5000
71	20.0	250	210	0.2	1	0.3	2612	0.6931	0.5000
71	20.0	250	210	0.2	1	0.3	2599	0.6934	0.4975
72	20.0	250	215	0.2	1	0.3	2575	0.6932	0.4975
72	20.0	250	215	0.2	1	0.3	2599	0.6933	0.4845
72	20.0	250	215	0.2	1	0.3	2626	0.6931	0.5025
73	20.0	250	220	0.2	1	0.3	2713	0.6931	0.5025
73	20.0	250	220	0.2	1	0.3	2616	0.6932	0.5000
73	20.0	250	220	0.2	1	0.3	2727	0.6932	0.5070
74	20.0	250	230	0.2	1	0.3	2699	0.6927	0.5170
74	20.0	250	230	0.2	1	0.3	2715	0.6931	0.5185
74	20.0	250	230	0.2	1	0.3	2609	0.6930	0.5120
75	20.0	250	240	0.2	1	0.3	2636	0.6935	0.4990
75	20.0	250	240	0.2	1	0.3	2587	0.6933	0.5010
75	20.0	250	240	0.2	1	0.3	2608	0.6931	0.5140
76	20.0	250	250	0.2	1	0.3	2647	0.6931	0.4985
76	20.0	250	250	0.2	1	0.3	2582	0.6935	0.5150
76	20.0	250	250	0.2	1	0.3	2852	0.6946	0.5055
77	20.0	250	200	1.0	5	1.0	2670	0.0341	0.9935

Literatur

1. V. F. Hess, *Physikalische Zeitschrift* **13**, 1084–1091 (1912).
2. C. D. Anderson, *Nobel Lecture*, Nobel Prize Outreach AB (<https://www.nobelprize.org/prizes/physics/1936/anderson/lecture/>).
3. R. Weiss, L. Brown, *Physics Today* **53**, 144 (Okt. 2000).
4. M. Schein, W. P. Jesse, E. O. Wollan, *Physical Review* **59**, 615–615, (<https://link.aps.org/doi/10.1103/PhysRev.59.615>) (7 Apr. 1941).
5. E. Fermi, *Physical Review* **75**, 1169–1174 (1949).
6. K. Greisen, *Physical Review Letter* **16**, 748–750, (<https://link.aps.org/doi/10.1103/PhysRevLett.16.748>) (17 Apr. 1966).
7. T. Fitoussi, G. Medina-Tanco, J. C. D’Olivo, *Journal of Cosmology and Astroparticle Physics* **2020**, 042–042, (<https://doi.org/10.1088%2F1475-7516%2F2020%2F01%2F042>) (Jan. 2020).
8. The Pierre Auger Collaboration, *Pierre Auger Observatory* (<https://www.auger.org/>).
9. H. Kolanoski, *Einführung in die Astroteilchenphysik*, (2009/10), (<https://www-zeuthen.desy.de/~kolanosk/astro0910/skripte/astro.pdf>).
10. D. H. Perkins, *Particle Astrophysics* (Oxford University Press, second edition, 2009).
11. D. Elsässer, W. Rhode, *Einführung in die Astroteilchenphysik*, Vorlesungsscript, nicht veröffentlicht, (2016).
12. M. Niechciol, T. Fehler, *Energy spectrum plot*, private Kommunikation, Feb. (2023).
13. H. Blümer, C. K. Guerard, *Nachrichten - Forschungszentrum Karlsruhe* **33**, 95–102 (2001).
14. B. G. Keilhauer, *Wissenschaftliche Berichte. FZKA* **6958** (2004).
15. J. Matthews, *Astroparticle Physics* **22**, 387–397, (<https://www.sciencedirect.com/science/article/pii/S0927650504001598>) (2005).
16. A. Horvath, *The geometry of the Cherenkov radiation* (<https://upload.wikimedia.org/wikipedia/commons/6/6b/Cherenkov.svg>).
17. The Pierre Auger Collaboration, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **798**, 172–213, (<https://www.sciencedirect.com/science/article/pii/S0168900215008086>) (2015).
18. The Pierre Auger Collaboration, *The Pierre Auger Observatory Upgrade - Preliminary Design Report*, (2016), arXiv: [1604.03637](https://arxiv.org/abs/1604.03637) (astro-ph.IM).
19. J. Blümer, *Innerer Aufbau eines Tscherenkow-Tanks* (<https://www.weltderphysik.de/gebiet/universum/kosmische-strahlung/detektoren/pierre-auger-observatorium>).

20. F. Chollet, *Deep Learning with Python* (Manning Publications Co., 2017).
21. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, <http://www.deeplearningbook.org> (MIT Press, 2016).
22. M. Mazur, *A Step by Step Backpropagation Example* (<https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example>).
23. D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, (2017), arXiv: 1412.6980 (cs.LG).
24. M. Kalirane, *Gradient Descent vs. Backpropagation: What's the Difference?* (<https://www.analyticsvidhya.com/blog/2023/01/gradient-descent-vs-backpropagation-whats-the-difference/>).
25. G. Van Rossum, F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
26. Martín Abadi u. a., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, (2015), (<https://www.tensorflow.org/>).
27. C. R. Harris u. a., *Nature* **585**, 357–362, (<https://doi.org/10.1038/s41586-020-2649-2>) (Sep. 2020).
28. T. pandas development team, *pandas-dev/pandas: Pandas*, Version 1.3.3, (2020), (<https://doi.org/10.5281/zenodo.3509134>).
29. J. D. Hunter, *Computing in Science & Engineering* **9**, 90–95 (2007).

Danksagung

Zuletzt möchte ich mich noch bei einigen Personen bedanken, die mich beim Schreiben dieser Arbeit unterstützt haben.

- Als erstes und im Besonderen möchte ich mich bei Prof. Dr. Risse bedanken. Dafür, dass er mir ermöglicht hat, diese spannende Arbeit zu schreiben und für die besondere Unterstützung in der letzten Phase der Arbeit.
- Ich möchte mich auch bei Dr. Philip Rühl bedanken, der mich in der ersten Zeit beim Schreiben dieser Arbeit betreut und begleitet hat.
- Darüber hinaus möchte ich meinen Eltern zum einen dafür danken, dass sie mir das Physikstudium ermöglicht haben und zum anderen für das Korrekturlesen dieser Arbeit und einiger hilfreicher Anmerkungen insbesondere bezüglich der sprachlichen Formulierungen.
- Als letztes möchte ich mich auch noch bei Maïke Czyborra bedanken, die mich als meine Freundin während des Schreibens dieser Arbeit und auch sonst immer unterstützt hat.

Vielen Dank!

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelor-Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate und Ergebnisse anderer kenntlich gemacht habe.

.....
Ort, Datum

.....
Unterschrift