

Einsatz von neuronalen Netzwerken bei der Photon-Identifikation im ATLAS-Experiment

Jonas Alexander Lotz

Bachelorarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

Dezember 2023

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate kenntlich gemacht habe.

Bonn, 13.12.2023
Datum

Jonas Loh
Unterschrift

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachterin: Priv.-Doz. Dr. Tatjana Lenz

Danksagung

Ich bedanke mich bei Prof. Dr. Ian C. Brock dafür, dass ich meine Bachelorarbeit als Mitglied seiner Arbeitsgruppe schreiben durfte. Zudem möchte ich meinen Dank an Priv. Doz. Dr. Tatjana Lenz aussprechen, dass sie sich dazu bereit erklärt hat, meine Arbeit als Zweitgutachterin zu bewerten. Außerdem möchte ich mich bei Florian Kirfel für die vielen konstruktiven Gespräche und die insgesamt hervorragende Betreuung bedanken. Auch möchte ich mich bei Christian Kirfel für das Korrekturlesen meiner Entwürfe bedanken. Schließlich gilt mein Dank der gesamten Brock Gruppe, die mich als eines ihrer Mitglieder bei sich aufgenommen hat und mir bei vielen Gelegenheiten das Gefühl gegeben hat, ein Teil der Gruppe gewesen zu sein.

Inhaltsverzeichnis

1	Einleitung	1
2	Physikalischer Hintergrund und experimentelle Grundlage	3
2.1	Das Standardmodell der Teilchenphysik	3
2.1.1	Fermionen	3
2.1.2	Bosonen	5
2.2	Beschleunigerphysik	5
2.2.1	Einheiten	5
2.2.2	Kinematische Größen	6
2.3	Der Large Hadron Collider und das ATLAS Experiment	7
2.3.1	Das ATLAS Koordinatensystem	9
2.3.2	Innerer Detektor	10
2.3.3	Kalorimeter	11
2.3.4	Myon Spektrometer	12
2.3.5	ATLAS Triggersystem	12
3	Photonidentifikation beim ATLAS Experiment	15
3.1	Photon Isolation und Selektion	15
3.2	Monte Carlo Simulationen	16
3.3	Identifikationseffizienz	17
3.3.1	Matrix Methode	18
3.3.2	Closure Unsicherheit	19
4	Machine Learning Konzepte	21
4.1	Netzwerk-Struktur eines Binärklassifikators	21
4.2	Fehlerrückführung	23
4.3	Optimierung eines neuronalen Netzwerkes	24
4.3.1	Leistungsbewertung	24
4.3.2	Generalisierungsfähigkeit	25
5	Optimierung der Photon-Identifikation mittels neuronaler Netzwerke	27
5.1	Training und Optimierung des Binärklassifikators	27
5.1.1	Gesamte Ereignisanzahl in Signal und Untergrund Monte Carlo Dateien	28
5.1.2	Vergleich mit tight-ID Arbeitspunkt	29
5.1.3	Hyperparameter Optimierung	29

5.1.4	Narrow-Strip Binärklassifikatoren	30
5.1.5	Relaxed-Tight Binärklassifikatoren	31
5.1.6	Resultate der Netzwerkoptimierung	33
5.2	Optimierung der Cuts auf den Netzwerkoutput	33
5.2.1	Signal Efficiency s_E	33
5.2.2	Background Rejection b_R	36
5.3	Tight-ID Effizienz und closure Unsicherheit mit Messdaten	39
5.3.1	Tight-ID Effizienz	39
5.3.2	Closure Unsicherheit	42
6	Resultate und Fazit	45
A	Leistung der Binärklassifikatoren	47
B	Cuts auf die Outputs der neuronalen Netzwerke	49
C	Ereigniszahl in den Messdaten	57
	Literatur	59
	Abbildungsverzeichnis	63
	Tabellenverzeichnis	65

Einleitung

Die Frage nach den fundamentalen Bausteinen der Materie spielt seit jeher eine zentrale Rolle in der Physik. Bereits im antiken Griechenland existierte die Vorstellung, dass alle Stoffe aus definierten kleinsten Teilchen, den sogenannten Atomen, bestünden. Dank einer rasanten Entwicklung von theoretischen und experimentellen Methoden im 20. Jahrhundert konnte eine Vielzahl von zentralen Erkenntnissen gewonnen werden, die den Weg zu unserem aktuellen Verständnis der Teilchenphysik geebnet haben. Seit den 1970er Jahren können alle Grundbausteine uns bekannter Materie, sowie deren Interaktionen, mit Hilfe des Standardmodells der Teilchenphysik beschrieben werden.

Die Vorhersagen des Standardmodells konnten dank der Entwicklung moderner Teilchenbeschleunigeranlagen getestet und mit immer höherer Genauigkeit bestätigt werden. Die weltweit größte Beschleunigeranlage ist derzeit der Large Hadron Collider (LHC). Sie befindet sich an der Grenze zwischen Frankreich und der Schweiz, nahe der Stadt Genf. Zur Messung der physikalischen Prozesse, die bei der Teilchenkollision im LHC hervorgerrufen werden, werden sehr komplexe Detektorsysteme benötigt. Eines dieser Systeme ist der ATLAS-Detektor. Unter anderem konnte durch die Analyse der Daten von ATLAS das Higgs Boson H , dessen Existenz vom Standardmodell vorhergesagt wurde, im Jahr 2012 gefunden werden.

In der Regel beruht die Analyse von ATLAS Daten auf der Identifikation und Rekonstruktion der Teilchen, die bei der Teilchenkollision entstanden sind. Die Verbesserung der Teilchenidentifikation ist somit ein wichtiger Schritt zur Erzielung neuer Erkenntnisse über die Teilchenphysik, die von ATLAS beobachtet werden kann. In dieser Arbeit wird ein Ansatz zur Verbesserung der Photon-Identifikation (Photon ID) mit Hilfe von neuronalen Netzwerken untersucht. Dabei werden vier Binärklassifikatorkomplexe auf unterschiedliche Variablensätze trainiert, um mit ihnen Signal Ereignisse von Untergrund Ereignissen unterscheiden zu können. Die Outputs der Netzwerke werden in der Matrix Methode zur Photon ID verwendet.

Zunächst werden in Kapitel 2 die physikalischen Grundlagen, die fundamental für die in dieser Arbeit behandelten Themen sind, erläutert. Es wird zudem ein Überblick über den LHC und ATLAS gegeben. Darauf folgt eine Einführung in die Photon-Identifikation mit der Matrix Methode in Kapitel 3. In Kapitel 4 werden die grundlegenden Konzepte zum Thema Machine Learning, die in dieser Arbeit genutzt wurden, behandelt. Schließlich wird in Kapitel 5 der Ansatz zur Optimierung der Photon-Identifikation mittels neuronaler Netzwerke beschrieben und untersucht. Die Zusammenfassung der Ergebnisse dieser Arbeit wird in Kapitel 6 gegeben.

Physikalischer Hintergrund und experimentelle Grundlage

Dieses Kapitel gibt einen Überblick über die physikalischen Konzepte, die grundlegend für die in dieser Arbeit behandelten Themen sind. Zudem wird eine Einführung in das ATLAS (A Toroidal LHC ApparatuS) Experiment am Large Hadron Collider gegeben.

2.1 Das Standardmodell der Teilchenphysik

Das Standardmodell der Teilchenphysik ist das bisher erfolgreichste Modell zur Beschreibung physikalischer Prozesse, die auf subatomaren Skalen stattfinden. Es dient als Basis unseres aktuellen Verständnisses der Teilchenphysik. Im Standardmodell werden alle Grundbausteine uns bekannter Materie, sowie deren Interaktionen via der elektromagnetischen, starken und schwachen Wechselwirkungen (jedoch nicht der Gravitation) beschrieben. Die Wechselwirkungen zwischen Materieteilchen (*Fermionen*) werden im Standardmodell durch Austauschteilchen (*Bosonen*) vermittelt. Eine grafische Darstellung des Standardmodells ist in Abb. 2.1 zu sehen. In den folgenden Unterabschnitten werden die physikalischen Eigenschaften der beiden Teilchenarten kurz zusammengefasst. Für eine detaillierte Erklärung der Teilchenphysik und des Standardmodells sei auf [1] und [2] verwiesen.

2.1.1 Fermionen

Fermionen sind Teilchen, die einen halbzahligen Spin besitzen und dem Pauli Prinzip unterliegen. Das Pauli Prinzip besagt, dass je zwei Fermionen innerhalb eines Quantensystems niemals denselben Quantenzustand besetzen können. Das Standardmodell umfasst zwölf solcher Spin-1/2-Teilchen, sowie deren entsprechende Antiteilchen. Diese sogenannten *Antifermionen* haben die gleiche Masse wie die jeweiligen Fermionen, tragen jedoch entgegengesetzte Quantenzahlen. Die Fermionen können nach aufsteigender Masse in drei Generationen eingeteilt werden. Als *Quarks* werden Fermionen bezeichnet, die Farbladung (rot, grün oder blau) tragen und somit der starken Wechselwirkung unterliegen. Aufgrund der starken Wechselwirkung sind Quarks in der Natur nie als freie Teilchen zu finden, sondern bilden mit anderen Quarks einen farbneutralen Zustand (genannt *Hadron*). Dieser Prozess wird als *Hadronisierung* bezeichnet. Alleine das Top Quark kann, aufgrund seiner kurzen Lebensdauer von $\tau \approx 5 \cdot 10^{-25}$ s [5] nicht hadronisieren. Im Standardmodell sind sechs Quarks, mit

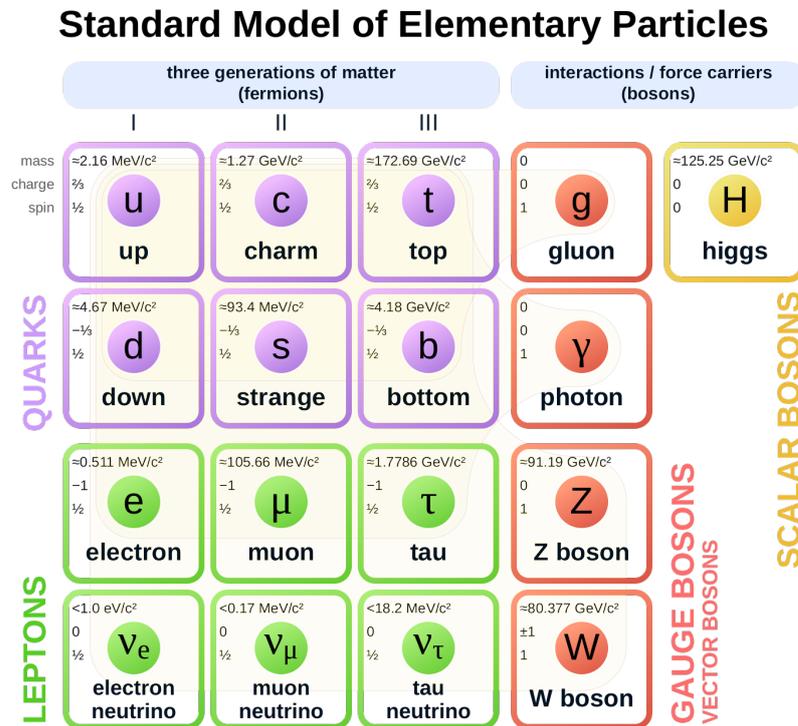


Abbildung 2.1: Das Standardmodell der Teilchenphysik mit seinen Bestandteilen. Die Grafik wurde von [3] entnommen und die Teilchenmassen auf die Version von der *Review of Particle Physics* Ausgabe von 2022 [4] angepasst.

unterschiedlichem *Flavor*, enthalten, die über ihre charakteristischen Quantenzahlen klassifiziert werden können. Diese Quantenzahlen sind die dritte Komponente des Isospin I_3 für Up und Down Quarks, Strangeness S für Strange Quarks, Charm C für Charm Quarks, Bottomness B für Bottom Quarks und Topness T für Top Quarks. Das Proton ist ein Hadron, welches aus zwei Up Quarks und einem Down Quark besteht ($p = uud$). Es ist das einzig stabile Hadron und bildet zusammen mit dem Neutron, welches aus zwei Down Quarks und einem Up Quark besteht ($n = udd$), die Bestandteile, aus denen die Atomkerne jedes chemischen Elementes zusammengesetzt sind. Durch Prozesse der schwachen Wechselwirkung können Quarks ihren Flavor wechseln. Übergänge innerhalb einer Generation werden in solchen Prozessen bevorzugt. Übergänge zwischen einer Generation sind um eine Größenordnung, Übergänge zwischen zwei Generationen sogar um zwei Größenordnungen unterdrückt.

Fermionen, die keine Farbladung tragen, werden als *Leptonen* bezeichnet. Im Standardmodell sind sechs Leptonen enthalten. Diese sind das Elektron e^- , Myon μ^- und Tauon τ^- , sowie deren Neutrinos ν_e , ν_μ , ν_τ . Anders als bei Quarks, sind Übergänge zwischen Leptonen generations-unabhängig. Neutrinos sind elektrisch neutral und interagieren somit nur mittels der schwachen Wechselwirkung. Zwar werden Neutrinos im Standardmodell als masselos vorausgesagt, die Beobachtung von Neutrinooszillationen zeigt jedoch, dass sie eine Masse besitzen müssen. Durch Messung des Betaspektrums beim Zerfall von Tritium ist es bisher möglich gewesen, eine obere Grenze für die Masse des Elektron-Antineutrinos

zu bestimmen [6]. Aus der experimentell gemessenen Breite der Z^0 -Resonanz konnte gezeigt werden, dass kein viertes leichtes Neutrino existieren kann. Jegliche Suche nach Quarks oder Leptonen weiterer Generationen war bisher vergeblich.

2.1.2 Bosonen

Bosonen sind Teilchen, die einen ganzzahligen Spin besitzen. Jede der drei elementaren Wechselwirkungen wird im Standardmodell durch den Austausch von Vektorbosonen mit Spin 1 übertragen. Das Boson, welches an alle elektrisch geladenen Teilchen koppelt und so für die Vermittlung der elektromagnetischen Wechselwirkung zuständig ist, ist das Photon γ . Das Photon ist selbst elektrisch neutral und hat keine Masse, wodurch die elektromagnetische Wechselwirkung eine unendliche Reichweite hat. Die Austauschbosonen der Schwachen Wechselwirkung sind die beiden geladenen W^\pm Bosonen sowie das elektrisch neutrale Z Boson. Aufgrund der hohen Masse ihrer Austauschbosonen, ist die schwache Wechselwirkung nur sehr kurzreichweitig. Die geladenen Austauschbosonen koppeln nur an linkshändige Fermionen und rechtshändige Antifermionen. Somit ist der geladene Strom der schwachen Wechselwirkung maximal paritätsverletzend. Der neutrale Strom der schwachen Wechselwirkung ist teilweise paritätsverletzend, da er in unterschiedlicher Stärke an rechts- und linkshändige Fermionen und Antifermionen koppelt. Die elektromagnetische und die schwache Wechselwirkung können zu einer einheitlichen Wechselwirkung, der sogenannten elektroschwachen Wechselwirkung, zusammengefasst werden.

Die starke Wechselwirkung wird über das masselose Gluon g übertragen, welches an die Farbladungen der Quarks koppelt. Gluonen tragen selbst Farbladung und wechselwirken somit untereinander. Dies beschränkt die Reichweite der starken Wechselwirkung, da schon bei sehr geringen Abständen die Energie des Farbfelds zur Erzeugung reeller Quark-Antiquark-Paare ausreicht. Freie Teilchen sind somit immer farbneutral.

Schließlich erfordert die Konsistenz des Standardmodells die Existenz eines elektrisch neutralen Skalarboson mit Spin 0, welches an die anderen Elementarteilchen proportional zu ihrer Masse koppelt. Dieses Teilchen ist das Higgs Boson H , dessen Existenz 2012 zum ersten Mal am LHC experimentell gezeigt werden konnte [7].

2.2 Beschleunigerphysik

Teilchenbeschleuniger sind heutzutage das wichtigste Instrument im Repertoire der experimentellen Teilchenphysik. Sie erlauben es Forschern Prozesse zu beobachten, die in räumlichen Größenordnungen kleiner als dem Durchmesser eines Protons, Zeitskalen kleiner als der Dauer eines Lichtblitzes und bei Energien stattfinden, wie sie nur in den extremsten Bereichen des Universums erreicht werden können. Dank der Entwicklung moderner Beschleunigeranlagen konnten die Vorhersagen des Standardmodells immer wieder getestet und mit immer größer werdender Genauigkeit bestätigt werden. Dieser Abschnitt führt alle benötigten Größen aus der Beschleunigerphysik ein, die für das weitere Verständnis dieser Arbeit notwendig sind.

2.2.1 Einheiten

Die in der Teilchenphysik gebräuchliche Einheit für Energie ist das Elektronvolt (eV). Das Elektronvolt ist die Energie, die ein Teilchen mit der Elementarladung $e = 1,602 \cdot 10^{-19}$ C nach Durchlaufen einer

Beschleunigungsspannung von 1 V hat:

$$1 \text{ eV} = 1,602 \cdot 10^{-19} \text{ J.} \quad (2.1)$$

Zur Vereinfachung von Rechnungen in der Teilchenphysik benutzt man oft ein System von physikalischen Größen, in dem Masse, Impuls, Energie, inverse Länge und inverse Zeit die gleichen Dimensionen haben. In diesem System kann man die Einheiten so definieren, dass für das reduzierte Plancksche Wirkungsquantum $\hbar = 1,0546 \cdot 10^{-34} \text{ J s}$ und die Lichtgeschwindigkeit $c = 2,9979 \cdot 10^8 \text{ m s}^{-1}$ folgende Relation gilt:

$$\hbar = c = 1. \quad (2.2)$$

2.2.2 Kinematische Größen

Schwerpunktsenergie

Der LHC ist ein Teilchenbeschleuniger vom Typ *Collider*. Bei dieser Art von Beschleunigern kreisen die Teilchen, die zur Kollision gebracht werden sollen, zunächst in entgegengesetzter Richtung innerhalb eines Speicherringes. Die kollidierenden Teilchen erreichen hierbei relativistische Energien, sodass $E_{1,2} \gg m_{1,2}$ gilt. Da beide Teilchen gleich stark beschleunigt werden, gilt zudem $E_1 = E_2$. Die Impulse der beiden Teilchen sind entgegengesetzt $\vec{p}_1 = -\vec{p}_2$. Energie und Impuls können für beide Teilchen zu Viererimpulsen $p_1 = (E_1, \vec{p}_1)$ und $p_2 = (E_2, \vec{p}_2)$ zusammengefasst werden. Man kann eine lorentzinvariante Energievariable s einführen, die das Quadrat der Schwerpunktsenergie bei Kollision der beiden Teilchen beschreibt:

$$s = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2E_1E_2 - 2\vec{p}_1\vec{p}_2. \quad (2.3)$$

Für die Schwerpunktsenergie in einem Collider gilt somit:

$$\sqrt{s} = 2E_1. \quad (2.4)$$

Die Schwerpunktsenergie eines Collider-Experiments ist, bei gleicher Strahlenergie, deutlich größer als bei einem Experiment mit ruhendem Target, weshalb Collider die bevorzugte Bauart für Experimente sind, die das Standardmodell in hohen Energiebereichen testen sollen.

Luminosität

Die Luminosität ist ein Maß für die Kollisionsrate der Teilchen pro Zeit und Fläche im Speicherring. Sie ist definiert als:

$$\mathcal{L} = \frac{N_a N_b \cdot jv/U}{4\pi\sigma_x\sigma_y}. \quad (2.5)$$

für eine Anzahl von j Teilchenpaketen mit jeweils N_a bzw. N_b Teilchen in einem Collider vom Umfang U , die mit der Geschwindigkeit v in einander entgegengesetzte Richtung umlaufen. Der Strahlquerschnitt im Kollisionspunkt der beiden Teilchenstrahlen wird über die Relation $A = 4\pi\sigma_x\sigma_y$ beschrieben. Hierbei wird angenommen, dass die Teilchen in horizontaler und vertikaler Richtung gaußförmig um die Strahlmitte verteilt sind. Die Größen σ_x und σ_y beschreiben die Standardabweichungen dieser Verteilungen. Um die Anzahl der Kollisionen pro Fläche in einem bestimmten

Zeitraum zu erhalten, kann man die Luminosität über diesen Zeitraum integrieren. Diese Größe wird dann als integrierte Luminosität bezeichnet.

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt. \quad (2.6)$$

Die Einheit, in der die Luminosität gemessen wird, ist das Barn ($1 \text{ b} = 1 \cdot 10^{-28} \text{ m}^2$).

Wirkungsquerschnitt

Der Wirkungsquerschnitt σ ist ein Maß für die Wahrscheinlichkeit einer Wechselwirkung zwischen zweier aufeinander treffender Teilchen. Er ist wie folgt definiert:

$$\sigma = \frac{\text{Zahl der Reaktionen pro Zeiteinheit}}{\text{Zahl der Strahlteilchen pro Zeiteinheit} \cdot \text{Zahl der Streuzentren pro Flächeneinheit}}. \quad (2.7)$$

Zusammen mit der integrierten Luminosität und dem Wirkungsquerschnitt kann die Gesamtmenge an Wechselwirkungen bestimmt werden:

$$N = \sigma \mathcal{L}_{\text{int}}. \quad (2.8)$$

2.3 Der Large Hadron Collider und das ATLAS Experiment

In dieser Arbeit werden Daten vom ATLAS Experiment analysiert. Dieser Abschnitt gibt einen kurzen Überblick über die wichtigsten Aspekte des LHC und ATLAS Detektors, die für das weitere Verständnis dieser Arbeit benötigt werden. Für eine detaillierte Erklärung des LHC und ATLAS Detektors wird auf [8] und [9] verwiesen.

Der Large Hadron Collider

Der Large Hadron Collider (LHC) ist der größte und leistungsstärkste Teilchenbeschleuniger der Welt. Er wurde im Jahr 2008 am Beschleunigerkomplex des CERN (Centre Européen pour la Recherche Nucléaire) in Genf in Betrieb genommen. Im 26,7 km langen Beschleunigerring des LHC können Protonkollisionen mit einer Schwerpunktsenergie von bis zu 13 TeV erzeugt werden. Um diese Schwerpunktsenergie zu erreichen, sind eine Vielzahl von Vorbeschleunigern notwendig. Der gesamte Beschleunigerkomplex des CERN ist in Abb. 2.2 dargestellt. Angefangen mit Linac2, einem Linearbeschleuniger, der die Protonen auf eine Energie von 50 MeV bringt, werden die Protonen schrittweise über das Proton Synchrotron Booster, das Proton Synchrotron und das Super Proton Synchrotron auf eine Energie von 450 GeV beschleunigt, bevor sie letztendlich in den Hauptbeschleuniger eingespeist werden. Die Protonen innerhalb des Beschleunigers sind in 2556 Bündel unterteilt. Die höchste Luminosität die damit im Jahr 2018 erreicht wurde, war $\mathcal{L} = 2,1 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [10]. Der LHC besteht aus acht geraden und acht gekrümmten Abschnitten. In den geraden Abschnitten werden die Protonen durch sogenannte *Radiofrequency Cavities* beschleunigt. Die Protonen werden von supraleitenden Dipolmagneten auf ihrer Kreisbahn gehalten. Die Magneten erreichen eine Magnetische Flussdichte von 8 T und müssen auf eine Temperatur von 1,9 K ($-271,3^\circ \text{C}$) heruntergekühlt werden [11]. An vier Punkten werden die Protonenstrahlen zur Kollision gebracht. Um jeden dieser Kollisionspunkte ist eines von vier Experimenten (ATLAS, CMS, LHCb und ALICE)

aufgebaut. Während ATLAS und CMS als Allzweckdetektoren dienen, und somit eine große Reihe von physikalischen Prozessen messen können, untersuchen ALICE Quark-Gluon Plasmas und LHCb Zerfälle von Hadronen, die ein Bottom oder Charm Quark enthalten.

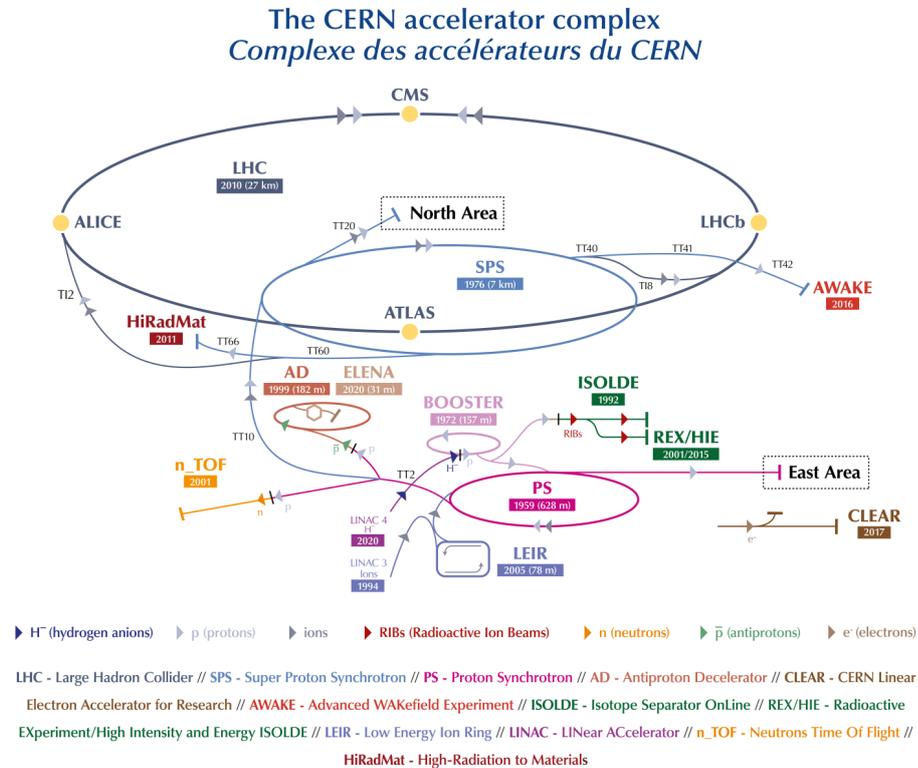


Abbildung 2.2: Der Beschleunigerkomplex des CERN [12].

Der ATLAS Detektor

Der ATLAS (A Toroidal LHC ApparatuS) Detektor ist ein Allzweckdetektor mit einer Länge von 46 m und einem Durchmesser von 25 m. Er ist zylindrisch um einen der vier Kollisionpunkte der Protonstrahlen des LHC herum aufgebaut. Der Detektor besteht aus vier übergeordneten Systemen, die in mehreren Schichten um den Kollisionspunkt herum angeordnet sind. Direkt am Wechselwirkungspunkt befindet sich der innere Detektor, der die Flugbahn elektrisch geladener Teilchen, die bei der Protonkollision entstehen, aufzeichnet. Darauf folgt das Kalorimetersystem. Es besteht aus einem Flüssig-Argon-Kalorimeter und einem Tile-Hadronkalorimeter, welche die Energien der Teilchen messen. Die äußerste Komponente des Detektors ist das Myon Spektrometer, das den Spurverlauf und Impuls der Myonen misst. Das Magnetsystem des Detektors besteht aus einem zentralen Solenoid-Magneten, welcher das 2 T starke Magnetfeld erzeugt, das den inneren Detektor durchsetzt, und dem Endkappen- und dem Barrel-Toroiden. Das Magnetfeld dient zur Krümmung der Flugbahn geladener Teilchen, wodurch der Detektor Informationen über deren elektrische Ladung und Impuls erhält. Eine computergenerierte Darstellung des gesamten ATLAS Detektorsystems ist in

Abb. 2.3 zu sehen.

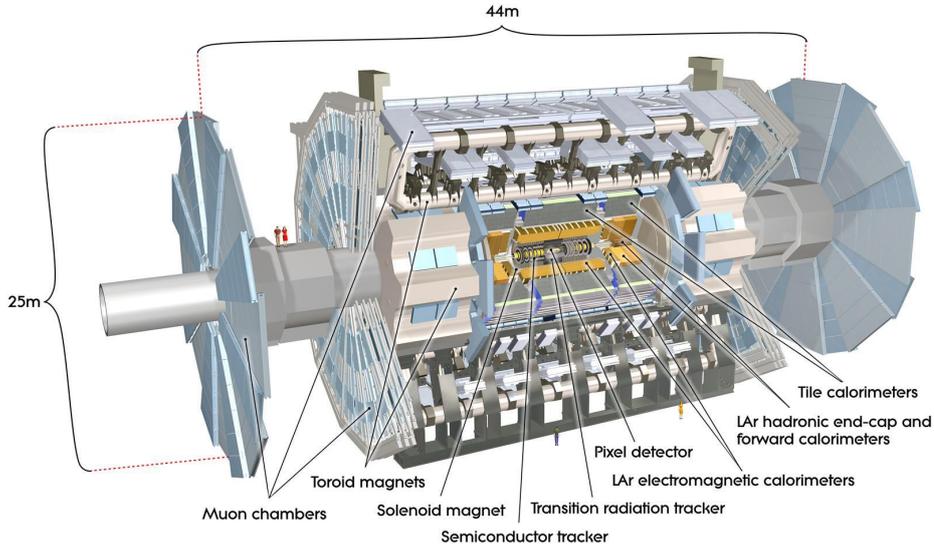


Abbildung 2.3: Computergenerierte Abbildung des ATLAS Detektorsystems [13].

2.3.1 Das ATLAS Koordinatensystem

Um die Beschreibung der Teilchenkollisionen zu vereinfachen, wird für den ATLAS Detektor ein spezielles Koordinatensystem genutzt. Der Kollisionspunkt stellt den Ursprung dieses Koordinatensystems dar. Die z -Achse verläuft entlang der horizontalen Strahlachse. Die x -Achse zeigt zum Mittelpunkt des LHC und spannt mit der vertikal nach oben zeigenden y -Achse die Transversalebene auf. In Zylinderkoordinaten bleibt der Azimutwinkel ϕ erhalten, der Polarwinkel θ wird jedoch von der sogenannten Pseudorapidität η ersetzt:

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (2.9)$$

Innerhalb der x - y -Ebene ist $\eta = 0$. In Richtung der positiven oder negativen Strahlachse ist $\eta = \pm\infty$. Ein Intervall $\Delta\eta$ ist lorentzinvariant unter einem Lorentz-Boost in z -Richtung. Abstände innerhalb der ϕ - η -Ebene werden durch ΔR ausgedrückt:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (2.10)$$

Der Impulsvektor der Teilchen kann somit im ATLAS-Koordinatensystem wie folgt ausgedrückt werden:

$$\vec{p} = \begin{pmatrix} p_X \\ p_Y \\ p_Z \end{pmatrix} = \begin{pmatrix} p_T \cdot \cos \phi \\ p_T \cdot \sin \phi \\ p_T \cdot \sinh \eta \end{pmatrix} \quad \text{wobei} \quad p_T = \sqrt{p_X^2 + p_Y^2}. \quad (2.11)$$

Für hochrelativistische Teilchen gilt $E \gg m$, weshalb der Betrag des Transversalimpulsvektors p_T näherungsweise der Transversalenergie E_T entspricht.

2.3.2 Innerer Detektor

Der innere Detektor besteht aus drei Subdetektoren, die vom 2 T starken magnetischen Feld des zentralen Solenoid-Magneten durchsetzt sind. Durch das magnetische Feld wird die Bewegung der elektrisch geladenen Teilchen, die aus der Protonenkollision resultieren, abgelenkt. Dies ermöglicht eine präzise Messung von elektrischer Ladung und Impuls der Teilchen, bei einer Pseudorapidität von $|\eta| < 2,5$. Abb. 2.4 zeigt eine computergenerierte Darstellung des inneren Detektors. Für einen knappen technischen Überblick über den Detektor, an dem sich auch die Beschreibungen in diesem Abschnitt orientieren, wird auf [14] verwiesen.

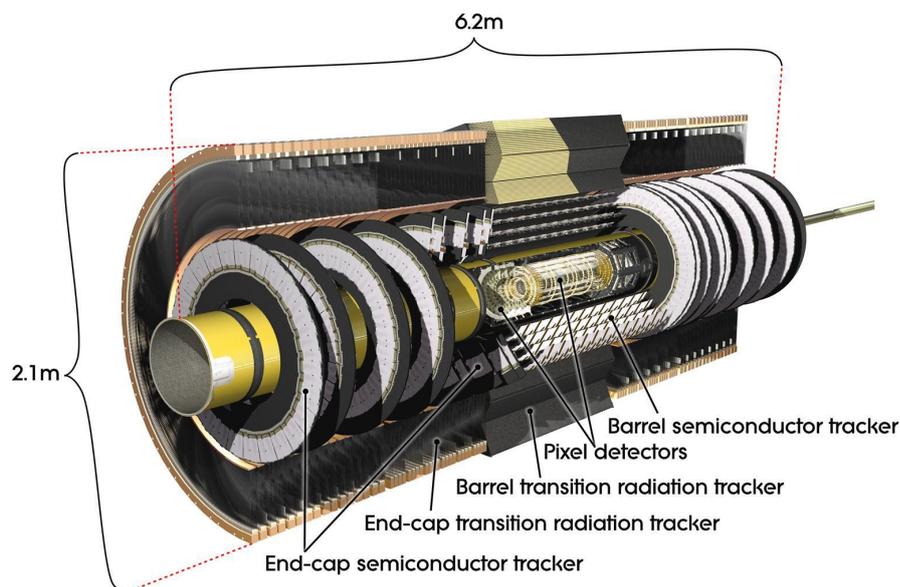


Abbildung 2.4: Computergenerierte Abbildung des inneren Detektorsystems [15].

Silizium-Pixeldetektor

Mit einem Abstand von nur 3,3 cm vom Teilchenstrahl ist der Pixeldetektor der erste Detektionspunkt in ATLAS. Geladene Teilchen, die vom Kollisionspunkt aus abgestrahlt werden, deponieren einen Bruchteil ihrer Energie im Detektor, der diese Signale mit einer Ortsauflösung von fast $10 \mu\text{m}$ messen kann. Dank seiner vier Schichten von 1736 Silizium Pixelsensor-Modulen kann so eine dreidimensionale Spur für die Teilchen rekonstruiert werden. Die erste Schicht des Pixeldetektors ist das *Insertable B-Layer* (IBL) [16]. Sie wurde erst im Jahr 2015 installiert, und steigert die Leistung des Detektors, um mit höhere Luminositäten nach Verbesserung des LHC arbeiten zu können.

Silizium-Streifendetektor

Der Silizium-Streifendetektor bildet die zweite Schale des inneren Detektors. Er besteht aus mehr als 4 000 Modulen mit mehr als $6 \cdot 10^6$ Streifen an Silizium-Sensoren. Er deckt eine größere Fläche ab, als der Pixeldetektor, und misst die Spur der Teilchen mit einer Genauigkeit von bis zu 25 μm .

Übergangsstrahlungsspurdetektor

Die letzte Schale des inneren Detektors wird vom Übergangsstrahlungsspurdetektor gebildet. Dieser besteht aus 300 000 Driftrohrkammern, die jeweils einen Durchmesser von nur 4 mm aufweisen und im Inneren einen 30 μm goldbeschichteten Wolframdraht im Mittelpunkt haben. Die Driftkammern enthalten eine Gasmischung, die von hindurchfliegenden Teilchen ionisiert werden kann. Dies resultiert in einem elektrischen Signal, welches vom Detektor aufgezeichnet werden kann. Dank der Übergangsstrahlung kann der Detektor, zusätzlich zum Spurverlauf, auch die Art des hindurchfliegenden Teilchens identifizieren.

2.3.3 Kalorimeter

Das Kalorimetersystem im ATLAS-Experiment ist für die Energiemessung der resultierendem Teilchen nach der Protonkollision zuständig. Ein knapper technischer Überblick, an dem sich dieser Unterabschnitt orientiert, ist in [17] zu finden. Die Kalorimeter sind dazu in der Lage, die meisten bekannten Teilchenarten innerhalb des Detektors zu stoppen, wobei Myonen und Neutrinos hierbei die Ausnahmen bilden. Über die unterschiedlichen Wege, wie die Teilchen ihre Energie innerhalb der Kalorimeter deponieren, kann auf deren Identität zurückgeschlossen werden. Elektromagnetische Kalorimeter können die Energie von Elektronen und Photonen messen. Für die Messung von Hadronen sind hingegen die hadronischen Kalorimeter zuständig. Das ATLAS Kalorimetersystem besteht aus einem Flüssig-Argon-Kalorimeter und einem Tile-Hadronkalorimeter. Die Kalorimetersysteme sind in Abb. 2.5 dargestellt.

Flüssig-Argon-Kalorimeter

Das Flüssig-Argon-Kalorimeter umgibt den inneren Detektor. Es besteht aus Schichten verschiedener Metalle (Wolfram, Kupfer oder Blei), in denen die eintreffenden Teilchen ihre kinetische Energie deponieren können. Die Energie wird dabei in sogenannte *Shower* oder Teilchen mit geringerer Energie umgewandelt. Diese können das flüssige Argon, das sich zwischen den Schichten befindet, ionisieren, was wiederum zu einem messbaren elektrischen Strom führt. Die Zentralregion des Kalorimeters ist dafür optimiert, Elektronen und Photonen zu messen.

Tile-Hadronkalorimeter

Das Tile-Hadronkalorimeter umgibt das Flüssig-Argon-Kalorimeter. Es misst die Energie derjenigen Hadronen, die nicht ihre gesamte Energie im Flüssig-Argon-Kalorimeter deponiert haben. Es besteht aus Schichten von Stahl und Kunststoffszintillatoren. Wenn Teilchen auf die Stahlschichten treffen, erzeugen sie einen Shower neuer Teilchen. Dieser trifft auf die Kunststoffszintillatoren, die wiederum Photonen erzeugen, welche in einen elektrischen Strom umgewandelt werden. Der Strom ist proportional

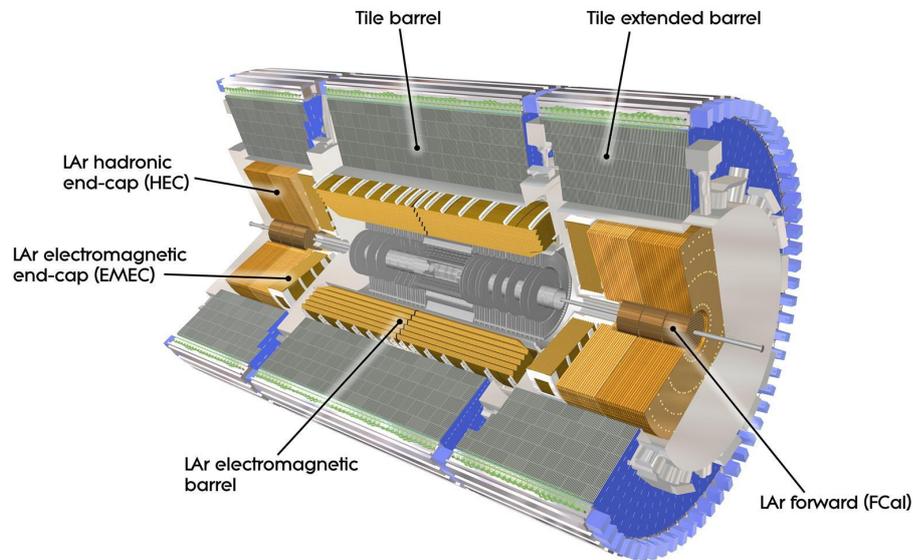


Abbildung 2.5: Computergenerierte Abbildung der Zentralregion des ATLAS-Detektors mit Hervorhebung des Kalorimetersystems [18].

zur Energie des ursprünglichen Teilchens. Das Tile-Hadronkalorimeter besteht aus 420 000 solcher Kunststoffszintillatoren.

2.3.4 Myon Spektrometer

Das Myon Spektrometer bildet die äußerste Komponente des Detektors. Es dient zur Identifikation und Impulsmessung der Myonen, die bei der Protonkollision entstehen. Myonen sind 200 mal schwerer als Elektronen, weshalb sie durch das ATLAS Kalorimetersystem hindurchfliegen können. Weiterführende Informationen zum Myon Spektrometer sind in [19] und [20] zu finden. Der Präzisionsdetektor des Spektrometers misst die Spur der Myonen innerhalb einer Pseudorapidität von $|\eta| < 2,7$ mithilfe mehrerer Schalen von Driftkammern. Dabei wird eine Präzision von $80 \mu\text{m}$ erreicht. In der Region der Endkappen ($2,0 < |\eta| < 2,7$) wird diese Messung von Kathodenstreifenkammern gemacht. Zusätzliche Komponenten mit hoher Zeitauflösung sorgen dafür, dass alle potentiell physikalisch relevanten Ereignisse vom Detektor herausgefiltert werden. Innerhalb einer Pseudorapidität von $|\eta| < 1,05$ werden hierfür sogenannte *Resistive Plate Chambers* (RPC) genutzt, während im Pseudorapiditätsintervall von $1,05 < |\eta| < 2,4$ *Thin Gap Chambers* (TGC) genutzt werden.

2.3.5 ATLAS Triggersystem

Das Triggersystem des ATLAS Detektors filtert aus den immensen Datenmengen, die bei der Messung der Protonkollisionen entstehen, die physikalisch interessanten Ereignisse heraus. Eine knappe technische Zusammenfassung ist in [21] zu finden. Die erste Stufe des Triggersystems ist der Hardware basierte Level 1 Trigger, der die Datenrate von 40 MHz auf 100 kHz reduziert. Die Entscheidung, welche Ereignisse herausgefiltert werden, basiert auf Informationen vom Myon Spektrometer und

Kalorimetersystem, in denen bestimmte Regionen identifiziert werden können, die von physikalischem Interesse sind. Die zweite Stufe wird durch den High Level Trigger gebildet. Dieser basiert auf offline rekonstruktions Algorithmen, die nach einer Zeitspanne von 300 ms eine Entscheidung treffen können, welche Ereignisse herausgefiltert werden sollen. Weiterführende Informationen finden sich in [22].

Photonidentifikation beim ATLAS Experiment

Ziel der Photoidentifikation (Photon ID) beim ATLAS Experiment ist es, Photonen im Detektor zu identifizieren, die direkt von der Protonenkollision herrühren (sogenannte *prompt* Photonen). Sie wird zur Messung verschiedener physikalischer Prozesse angewandt, in denen Photonen auftreten. Diese sind unter anderem die Analyse des Inclusive-Prompt-Photon- und Di-Photon-Wirkungsquerschnitts, Messungen verschiedener Eigenschaften des Higgs Bosons, welches in zwei Photonen zerfällt, oder das Setzen von Obergrenzen für Produktionsquerschnitte hypothetischer Prozesse, in denen Photonen im Endzustand auftreten. Prompt Photonen, die nicht von hadronischen Zerfällen her stammen, können in zwei Produktionskanäle unterteilt werden. Diese sind die Entstehung von Photonen bei harten Streuprozessen, vor allem via Quark-Gluon Compton Streuung ($qg \rightarrow \gamma q$) oder Quark-Antiquark Annihilation ($q\bar{q} \rightarrow \gamma g$), und Fragmentations-Photonen, die bei der Zersplitterung eines Partons bei hohem p_T entstehen. Um den Beitrag von Photonen zu vermeiden, die zum Beispiel bei dem Zerfall neutraler Hadronen wie dem π^0 und η -Meson entstehen, muss ein Isolationskriterium auf die Photonen angewandt werden. In ATLAS werden die Photonen anhand der Energiedepots innerhalb der Kalorimetersysteme und Informationen vom inneren Detektor rekonstruiert. Die Photon ID basiert auf einer Reihe von rechteckigen Schnitten (sogenannten *Cuts*) auf mehreren diskriminierenden Variablen, die aus der lateralen und longitudinalen Schauer Entwicklung im elektromagnetischen Kalorimeter und der *Shower Leakage Fraction* des Photons im hadronischen Kalorimeter berechnet werden. Eine Auflistung dieser sogenannten *shower-shape* Variablen ist in Tabelle 3.1 am Ende dieses Kapitels zu finden. Für eine detaillierte Beschreibung der Photon Rekonstruktion und ID in ATLAS wird auf [23] und [24] verwiesen. Bei der Photon Rekonstruktion und ID wird zwischen *konvertierten* Photonen und *unkonvertierten* Photonen unterschieden. Konvertiert bezeichnet hierbei, dass das Photon durch Interaktion mit Detektormaterial im inneren Detektor in ein Elektron-Positron Paar übergegangen ist. Unkonvertierte Photonen erreichen die Kalorimeter ohne solche Übergänge.

3.1 Photon Isolation und Selektion

Die Photon-Identifikationseffizienz ε wird durch über die Anzahl der Photonen bestimmt, die ein bestimmtes Isolationskriterium erfüllen. Dieses Kriterium ist definiert über die Transversalenergie innerhalb eines Kegels der Größe ΔR (vgl. Gleichung (2.10)) um einen Photonkandidaten herum. Die Transversalenergie besteht aus zwei Komponenten, der Kalorimeter-Isolation E_T^{iso} und der sogenannten

Track-Isolation p_T^{iso} . Die Kalorimeter-Isolation E_T^{iso} beschreibt die Summe aller Transversalenergien innerhalb der Kalorimetercluster, nachdem der die Energie des Photonkandidaten und sonstige pile-up Effekte subtrahiert wurden. Die Track-Isolation p_T^{iso} ist die Summe aller Transversalimpulse oberhalb von 1 GeV, mit Ausnahme möglicher konversions Spuren. Im Folgenden werden die zwei Isolationskriterien vorgestellt, die in dieser Arbeit zur Photonisation verwendet wurden. Für weitere Informationen sei auf [23] verwiesen. Das erste Kriterium zur Photonisation ist das *Loose*-Isolationskriterium:

$$E_T^{\text{iso}} \Big|_{\Delta R < 0,2} < 0,065 \cdot E_T \quad \text{und} \quad p_T^{\text{iso}} \Big|_{\Delta R < 0,2} < 0,05 \cdot E_T. \quad (3.1)$$

Das zweite Kriterium ist das *Tight*-Isolationskriterium:

$$E_T^{\text{iso}} \Big|_{\Delta R < 0,4} < 0,022 \cdot E_T + 2,45 \text{ GeV} \quad \text{und} \quad p_T^{\text{iso}} \Big|_{\Delta R < 0,2} < 0,05 \cdot E_T. \quad (3.2)$$

Da sich konvertierte und unkonvertierte Photonen sehr stark in den Variablen, die ihre Schauerentwicklung beschreiben, unterscheiden, müssen die Identifikationskriterien für die beiden Photonarten separat optimiert werden. Zudem ist die Geometrie des Kalorimetersystems für verschiedene Pseudorapiditätsintervalle sehr ungleichmäßig. Die Identifikationskriterien sind für jedes Intervall individuell optimiert worden. Die 6 Pseudorapiditätsintervalle, auf die die Photon ID optimiert ist, sind:

$$[0,00, 0,60), [0,60, 0,80), [0,80, 1,37), [1,52, 1,81), [1,81, 2,01), [2,01, 2,37).$$

Die Region $1,37 < |\eta| < 1,52$ wird hier aufgrund einer hohen Menge von Material, das nicht zur Messung beiträgt (*material upstream*), innerhalb des Kalorimeters nicht in Betracht gezogen. Die Transversalenergie E_T wird in 14 Intervalle unterteilt. Diese sind:

$$[25, 30), [30, 35), [35, 40), [40, 45), [45, 50), [50, 60), [60, 80), [80, 100), [100, 125), [125, 150), \\ [150, 175), [175, 250), [250, 350), [350, 1500] \quad (\text{Einheiten in GeV}).$$

Durch Vergrößerung der Intervallbreite bei höherer E_T , wird dafür gesorgt, dass genug Statistik in jedem Intervall für die Messung vorhanden ist.

3.2 Monte Carlo Simulationen

Monte Carlo Simulationen werden von einem Computerprogramm (einem *MC Event Generator*) generiert und können zum Vergleich von Vorhersagen aus theoretischen Modellen mit experimentellen Daten genutzt werden. Ihre Anwendung ist in der Teilchenphysik weit verbreitet. Im Falle vom LHC werden pp -Kollisionen basierend auf den Vorhersagen des Standardmodells simuliert. Da Teilchenkollisionen und Zerfälle stochastische Prozesse sind, existieren verschiedene Outputs für jedes simulierte Event. Zufällige Selektion dieser Outputs führt letztendlich zu einem Datenset, dass die zu erwartenden Prozesse akkurat beschreiben kann. Bei Prozessen, die sehr gut modelliert werden können, können Monte Carlo Simulationen zur Optimierung von Analysemethoden genutzt werden.

3.3 Identifikationseffizienz

Zur Berechnung der tight-ID Effizienz aus Daten vom ATLAS Experiment wird in dieser Arbeit die Matrix Methode angewandt. Um eine effiziente Messung mit Hilfe der Matrix Methode zu erhalten, wird zuerst eine Vorselektion der Messdaten vorgenommen. Bei dieser werden alle Photonen mit einer Energie unterhalb von 25 GeV heraus sortiert. Zudem müssen die Photonen das Loose-Kriterium (Gleichung (3.1)) erfüllen. Die Gesamtzahl der Photonen N^T wird in *prompt* Photonen N^S und *fake* Photonen (also Untergrund Photonen, die nicht direkt von der Protonenkollision herrühren) N^B unterteilt. Durch Anwendung des Tight-Kriteriums erhält man die Anzahl der tight-ID Photonen N_{ID}^T , die erneut in prompt und fake Photonen unterteilt werden kann.

$$N^T = N^S + N^B, \quad (3.3)$$

$$N_{ID}^T = N_{ID}^S + N_{ID}^B. \quad (3.4)$$

Die tight-ID Effizienz ist definiert als das Verhältnis zwischen tight-ID prompt Photonen und allen prompt Photonen:

$$\varepsilon^{\text{tight-ID}} = \frac{N_{ID}^S}{N^S}. \quad (3.5)$$

Die Reinheit (*Purity*) des Signals ist das Verhältnis der Anzahl von prompt Photonen und der Gesamtzahl aller Photonen, jeweils vor und nach der Anwendung des tight-ID Kriteriums.

$$A = \frac{N^S}{N^T}, \quad (3.6)$$

$$P = \frac{N_{ID}^S}{N_{ID}^T}. \quad (3.7)$$

Mit diesen Purities kann die tight-ID Effizienz also wie folgt ausgedrückt werden:

$$\varepsilon^{\text{tight-ID}} = \frac{P \cdot N_{ID}^T}{A \cdot N^T}. \quad (3.8)$$

Um zwischen prompt und fake Photonen zu unterscheiden, werden Spur-Isolationskriterien (*track isolation criteria*) genutzt. Die Isolationskriterien schließen alle Photonen aus, die eine Spur mit $p_T > 1$ GeV innerhalb eines Kegels mit $0,1 < \Delta R < 0,4$ um das Photon Cluster herum haben. Spuren innerhalb des Kegels $\Delta R < 0,1$ werden ausgeschlossen, um die Selektion von Übergangsspuren zu vermeiden [23]. Die Gesamtzahl der Photonen, die diese Kriterien erfüllen, werden mit \hat{N}^T und \hat{N}_{ID}^T bezeichnet. Sie können analog zu Gleichung 3.3 und Gleichung 3.4 berechnet werden:

$$\hat{N}^T = \hat{N}^S + \hat{N}^B, \quad (3.9)$$

$$\hat{N}_{ID}^T = \hat{N}_{ID}^S + \hat{N}_{ID}^B. \quad (3.10)$$

Die Spurisolierungseffizienz wird aus dem Verhältnis der spurisolierten Photonen und der Gesamtzahl der Photonen berechnet:

$$\hat{\varepsilon} = \frac{\hat{N}^T}{N^T}, \quad (3.11)$$

$$\hat{\varepsilon}_{\text{ID}} = \frac{\hat{N}_{\text{ID}}^T}{N_{\text{ID}}^T}. \quad (3.12)$$

Die Spurisolierungseffizienzen für prompt und fake Photonen werden als $\hat{\varepsilon}^S$ und $\hat{\varepsilon}^B$ bezeichnet. $\hat{\varepsilon}_{\text{ID}}^S$ und $\hat{\varepsilon}_{\text{ID}}^B$ sind die entsprechenden Spurisolierungseffizienzen nach Anwendung des tight-ID Kriteriums. Mit Hilfe der Spurisolierungseffizienzen kann die Gesamtzahl der spurisolierten Photonen \hat{N} und die Anzahl der spurisolierten tight-ID Photonen \hat{N}_{ID} berechnet werden.

$$\hat{N}^T = \hat{\varepsilon}^S \cdot \hat{N}^S + \hat{\varepsilon}^B \cdot \hat{N}^B, \quad (3.13)$$

$$\hat{N}_{\text{ID}}^T = \hat{\varepsilon}_{\text{ID}}^S \cdot \hat{N}_{\text{ID}}^S + \hat{\varepsilon}_{\text{ID}}^B \cdot \hat{N}_{\text{ID}}^B. \quad (3.14)$$

Die Purities können somit über die Spurisolierungseffizienzen ausgedrückt werden.

$$A = \frac{N^S}{N^T} = \frac{\hat{\varepsilon} - \hat{\varepsilon}^B}{\hat{\varepsilon}^S - \hat{\varepsilon}^B}, \quad (3.15)$$

$$P = \frac{N_{\text{ID}}^S}{N_{\text{ID}}^T} = \frac{\hat{\varepsilon}_{\text{ID}} - \hat{\varepsilon}_{\text{ID}}^B}{\hat{\varepsilon}_{\text{ID}}^S - \hat{\varepsilon}_{\text{ID}}^B}. \quad (3.16)$$

Durch Einsetzen der Ausdrücke für die Spurisolierungseffizienzen in Gleichung 3.5, kann die tight-ID Effizienz wie folgt geschrieben werden:

$$\varepsilon^{\text{tight-ID}} = \frac{\frac{\hat{\varepsilon}_{\text{ID}} - \hat{\varepsilon}_{\text{ID}}^B}{\hat{\varepsilon}_{\text{ID}}^S - \hat{\varepsilon}_{\text{ID}}^B} \cdot N_{\text{ID}}^T}{\frac{\hat{\varepsilon} - \hat{\varepsilon}^B}{\hat{\varepsilon}^S - \hat{\varepsilon}^B} \cdot N^T}. \quad (3.17)$$

Die Werte von N_{ID}^T , N^T , $\hat{\varepsilon}$ und $\hat{\varepsilon}_{\text{ID}}$ können direkt aus den Messdaten des Experiments entnommen werden. Die Spurisolierungseffizienzen $\hat{\varepsilon}^S$, $\hat{\varepsilon}_{\text{ID}}^S$ von prompt Photonen können mit sehr hoher Genauigkeit durch Monte Carlo Simulationen ermittelt werden. Dagegen ist die Übereinstimmung der Spurisolierungseffizienzen $\hat{\varepsilon}^B$, $\hat{\varepsilon}_{\text{ID}}^B$ für fake Photonen zwischen Daten und Monte Carlo Simulation nicht exakt genug, um die Effizienzen direkt aus der Simulation zu extrahieren. Diese müssen deshalb über die Matrix Methode ermittelt werden.

3.3.1 Matrix Methode

Zur Berechnung von $\hat{\varepsilon}^B$ und $\hat{\varepsilon}_{\text{ID}}^B$ werden die Photonen in vier Regionen eingeteilt, indem man neun shower-shape Variablen aus Tabelle 3.1 in zwei Kategorien unterteilt. Die erste Kategorie sind die *narrow-strip* Variablen F_{side} , w_{η_1} , ΔE_S und E_{ratio} . Sie werden von der Energie im ersten Kalorimeter Layer berechnet, die nur geringfügig mit der Spurisolierung von Photonen korreliert sind. Die zweite Kategorie sind die *relaxed-tight* Variablen R_{had} , R_{η} , R_{ϕ} , w_{η_2} , $w_{s,\text{tot}}$ [25]. Sie beschreiben die Entwicklung des Photonen-Schauers im mittleren und letzten Layer des elektromagnetischen

Kalorimeters, sowie den Verlust in dem hadronischen Kalorimeter. Die Regionen der Matrix Methode sind zur Veranschaulichung in Abb. 3.1 dargestellt. Alle Photonen, die das tight-ID Kriterium erfüllen,



Abbildung 3.1: Schematische Darstellung der vier Regionen von der Matrix Methode (angefertigt mit [26]).

befinden sich in Region 1. Diese kann somit genutzt werden, um die Spurisolationseffizienz für fake Photonen, die das tight-ID Kriterium erfüllen, zu berechnen. Unter der Annahme, dass die narrow-strip Variablen nur gering mit dem Spurisolutionskriterium korreliert sind, können folgende Näherungen gemacht werden:

$$\hat{\varepsilon}_{\text{ID}}^{\text{B}} = \frac{\hat{N}_1^{\text{B}}}{N_1^{\text{B}}} \approx \frac{\hat{N}_3^{\text{B}}}{N_3^{\text{B}}}, \quad (3.18)$$

$$\hat{\varepsilon}^{\text{B}} = \frac{\hat{N}_{1+2+3+4}^{\text{B}}}{N_{1+2+3+4}^{\text{B}}} \approx \frac{\hat{N}_{2+3+4}^{\text{B}}}{N_{2+3+4}^{\text{B}}}. \quad (3.19)$$

3.3.2 Closure Unsicherheit

Bei der Messung der tight-ID gibt es sowohl statistische Unsicherheiten als auch systematische Unsicherheiten. Die für diese Arbeit relevanteste und speziell im Niederenergiebereich größte systematische Unsicherheit ist die so genannte *closure* Unsicherheit. Diese wird durch die Approximationen in Gleichung (3.18) und Gleichung (3.19) verursacht, da die narrow-strip Variablen tatsächlich nicht vollständig mit dem Spurisolutionskriterium unkorreliert sind. Die absolute Größe der Unsicherheit wird durch fake Photonen in Monte Carlo Simulationen abgeschätzt.

$$\Delta \hat{\varepsilon}_{\text{ID}}^{\text{B}} = \frac{|\hat{\varepsilon}_1^{\text{B}} - \hat{\varepsilon}_3^{\text{B}}|}{\hat{\varepsilon}_1^{\text{B}}}, \quad (3.20)$$

$$\Delta \hat{\varepsilon}_{\text{ID}}^{\text{B}} = \frac{|\hat{\varepsilon}_{1+2+3+4}^{\text{B}} - \hat{\varepsilon}_{2+3+4}^{\text{B}}|}{\hat{\varepsilon}_{1+2+3+4}^{\text{B}}}. \quad (3.21)$$

Tabelle 3.1: Shower-shape Variablen, die zur Klassifizierung von tight-ID Photonen genutzt werden [27].

Kategorie	Variablenname	Beschreibung
Hadronischer Verlust	R_{had_1}	Verhältnis von E_T im ersten sampling Layer des Hadronischen Kalorimeters zu E_T des EM Cluster (bei einer Pseudorapidität von $ \eta < 0,8$ oder $ \eta > 1,52$)
	R_{had}	Verhältnis von E_T im Hadronischen Kalorimeter zu E_T im EM Cluster (bei einer Pseudorapidität von $0,8 < \eta < 1,37$)
ECAL erstes Layer	w_{η_1}	Laterale Schauerweite, $\sqrt{\sum E_i(i - i_{max})^2 / \sum E_i}$, wo i über alle Streifen in einem Intervall $3 \times 2\eta \times \phi$ geht, und i_{max} der Index des Streifen mit der höchsten Energie, berechnet über drei Streifen im Umkreis des Streifen mit dem höchsten Energiedepot, ist.
	$w_{s,tot}$	Totale laterale Schauerweite, $\sqrt{\sum E_i(i - i_{max})^2 / \sum E_i}$, wo i über alle Streifen in einem Intervall $20 \times 2\eta \times \phi$ geht, und i_{max} der Index des Streifen mit der höchsten gemessenen Energie ist.
	F_{side}	Energie außerhalb des Kerns der drei zentralen Streifen, aber innerhalb der sieben Streifen, geteilt durch die Energie innerhalb der drei zentralen Streifen.
	ΔE_s	Differenz zwischen der Energie, die mit dem zweiten Maximum im Streifen Layer assoziiert ist und der Energie, die im Streifen mit dem minimalen Wert zwischen dem ersten und zweiten Maximum rekonstruiert wurde.
	E_{ratio}	Verhältnis der Energie Differenz zwischen dem maximalen Energiedepot und dem Energiedepot im zweiten Maximum im Cluster zur Summe dieser Energien.
	f_1	Verhältnis der Energie im ersten Layer zur totalen Energie des EM Cluster.
ECAL zweites Layer	R_η	Verhältnis der Energien in $3 \times 7\eta \times \phi$ Zellen und in 7×7 Zellen die sich um die Photon Cluster Position herum befinden.
	w_{η_2}	Laterale Schauerweite $\sqrt{\sum E_i \eta_i^2 / \sum E_i - (\sum E_i \eta_i / \sum E_i)^2}$, wobei E_i die Energie und η_i die Pseudorapidität der Zelle i ist, und die Summe innerhalb eines Intervalls von 3×5 Zellen berechnet wird.
	R_ϕ	Verhältnis der Energie in $3 \times 7\eta \times \phi$ Zellen zur die Energie in 3×7 Zellen die um die Photon Cluster Position herum Zentriert sind.

Machine Learning Konzepte

Dieses Kapitel gibt einen Überblick über die Konzepte aus dem Bereich *Machine Learning*, die in dieser Arbeit genutzt werden. Für eine detaillierte Einführung in dieses Forschungsgebiet wird auf [28] und [29] verwiesen. Der Oberbegriff *Machine Learning* umfasst alle künstlichen Systeme, die Muster und Gesetzmäßigkeiten aus Lerndaten erkennen können, und dann zuvor unbekannte Daten anhand dieser Muster beurteilen können. Für diese Arbeit wird ein solches System in Form eines künstlichen neuronale Netzwerks verwendet. Ähnlich zu den Neuronen im Gehirn eines Lebewesens sind bei einem solchen System die künstlichen Neuronen (*Knoten*) über ein komplexes Netzwerk von gewichteten Verbindungen untereinander verknüpft. Die Gewichtung der einzelnen Verbindungen wird während des Trainings dynamisch angepasst. Dies ermöglicht es dem Netzwerk, Muster in den Trainingsdaten zu erkennen. Nach Finalisierung des Lernprozesses ist das Netzwerk schließlich dazu in der Lage, selbstständig Entscheidungen auf zuvor unbekanntem Daten zu treffen.

4.1 Netzwerk-Struktur eines Binärklassifikators

Für diese Arbeit werden neuronale Netzwerke vom Typ Binärklassifikator eingesetzt. Mit dieser Art von Netzwerk ist es möglich, Eingabedaten in zwei Kategorien einzuteilen. In der Teilchenphysik kann ein solches Netzwerk beispielsweise genutzt werden, um prompt und fake Photonen aus einem gemeinsamen Datensatz herauszufiltern. Die wichtigste Komponente des Netzwerks sind die Knoten, die in mehreren Schichten angeordnet sind. Die erste Schicht, genannt *Eingabeschicht*, enthält so viele Knoten wie die Anzahl der Variablen, auf die das Netzwerk trainiert wird. Darauf folgt eine bestimmte Anzahl von Schichten, deren Menge an Knoten vom Benutzer festgelegt werden können. Diese bilden das Netzwerk, das die Eingabevariablen mit der Ausgabeschicht verbindet. Bei dem Binärklassifikator Netzwerk, das in dieser Arbeit verwendet wird, ist in der Ausgabeschicht genau ein Knoten enthalten. Dieser gibt einen Wert zwischen Null (fake Photon im Physik Beispiel) und Eins (prompt Photon). Eine schematische Darstellung von dieser Form von neuronalem Netzwerk ist in Abb. 4.1 zu sehen. Jeder Knoten in einem Netzwerk mit L Schichten und M Knoten pro Schicht enthält einen skalaren Wert a_m^l . Dieser setzt sich wie folgt zusammen. Die einzelnen Knoten innerhalb einer Schicht $l + 1$ sind mit allen anderen Knoten der benachbarten Schicht über Verknüpfungen mit einer Gewichtung w_{mn}^{l+1} , verbunden. Zunächst wird also die Summe über alle skalaren Werte der vorherigen Schicht, multipliziert mit dem Gewicht ihrer jeweiligen Verknüpfung, berechnet. Diese Summe kann zudem

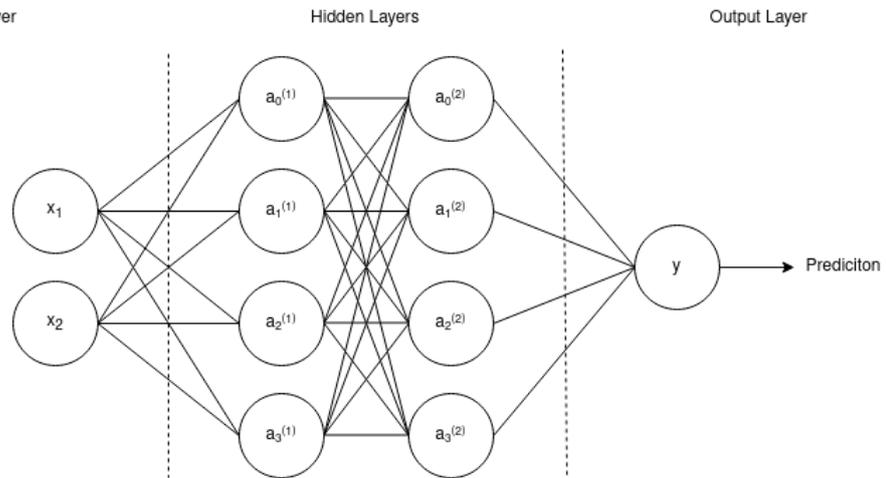


Abbildung 4.1: Schematische Darstellung eines neuronalen Netzwerks mit zwei Eingabevariablen und zwei versteckten Schichten (angefertigt mit [26]).

um einen konstanten Wert b_n^{l+1} , den sogenannten *Bias*, geschiftet werden. Schließlich wird das Resultat dieser Rechnung als Input für eine Aktivierungsfunktion σ verwendet. Die Aktivierungsfunktion kann den Wert des Knotens auf ein bestimmtes Intervall beschränken und eine zusätzliche nichtlineare Komponente einführen, die vom Netzwerk zur Beschreibung nichtlinearer Zusammenhänge genutzt werden kann.

$$a_n^{l+1} = \sigma \left(\sum_m^M w_{mn}^{l+1} \cdot a_m^l + b_n^{l+1} \right). \quad (4.1)$$

Die Aktivierungsfunktionen, die im neuronalen Netzwerk in dieser Arbeit genutzt werden, sind in Tabelle 4.1 angegeben. Für die letzte Schicht wird in einem Binärklassifikator meist die Sigmoid-Funktion genutzt, da diese das Intervall ihres Output Wertes auf $[0, 1]$ beschränkt.

Tabelle 4.1: Definitionen der genutzten Aktivierungsfunktionen.

Funktionsname	Definition
Sigmoid	$\sigma(x) = \frac{1}{1+\exp(-x)}$
Softmax	$\sigma(x)_j = \frac{\exp(x_j)}{\sum_{k=1}^K \exp(x_k)}$
Rectified Linear Unit (ReLU)	$\sigma(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases}$

4.2 Fehlerrückführung

In dieser Arbeit wird eine Methode aus dem überwachten Lernverfahren zum Training des neuronalen Netzwerkes angewandt. Überwachte Lernverfahren bezeichnen im Allgemeinen Methoden, bei denen die Netzwerke auf Daten trainiert werden, deren Ergebnisse bereits bekannt sind. Die Ergebnisse des Netzwerkes können somit beim Training mit den tatsächlichen Ergebnissen verglichen werden. Zu Beginn des Trainings wird das Netzwerk mit zufälligen Gewichtungen und Biases initialisiert. Das Netzwerk gibt in dieser Konfiguration im Regelfall nur sehr schlechte Resultate. Die Qualität der Resultate wird über eine Loss-Funktion quantifiziert. Um die Resultate des Netzwerkes zu verbessern, muss die Loss-Funktion vom Netzwerk minimiert werden. Die Loss-Funktion kann also als ein Maß für die Güte des Netzwerkes verstanden werden. Für Binärklassifikatoren wird die *Binary-Cross-Entropy* L_{BCE} (vgl. Gleichung (4.2)) als Loss-Funktion verwendet. Diese Funktion vergleicht die korrekten Werte $y \in \mathbb{Z} \cap [0, 1]$ mit dem Output des Netzwerkes $p \in [0, 1]$. Wenn der Ausgabewert des Netzwerkes nicht mit dem tatsächlichen Wert übereinstimmt, steigt der Wert dieser Funktion exponentiell an. Da beim Training der Funktionswert minimiert werden soll, ist dieses Verhalten besonders gut, um falsche Ergebnisse zu vermeiden.

$$L_{\text{BCE}} = -(\log(p) + (1 - y) \log(1 - p)). \quad (4.2)$$

Um das Minimum der Loss-Funktion zu finden, wird ihr negativer Gradient bezüglich der Parameter des Netzwerkes $G = -\nabla L$ berechnet. Die Berechnung des Gradienten wird also für jeden Parameter a_n^{l+1} , w_{mn}^{l+1} und b_n^{l+1} im Netzwerk durchgeführt. Diese Rechnung wird als Fehlerrückführung (*Back-propagation*) bezeichnet. Schließlich werden die Parameter in Richtung des Gradienten angepasst. Der Trainingsprozess des neuronalen Netzwerkes wird also durch die dynamische Anpassung der Gewichtungen und dem Bias durchgeführt. Die Anpassung der Parameter wird von einem Optimierungsalgorithmus vorgenommen. Der einfachste Weg, um die Parameter zu optimieren, ist über das Stochastische-Gradientenverfahren (*Stochastic Gradient Descent*, SGD). In diesem Verfahren werden die Parameter nach jedem Trainingsabschnitt i wie folgt angepasst:

$$\beta_f = \beta_i - \alpha \nabla L(x_i, y_i, \beta_i). \quad (4.3)$$

Hier bezeichnet β_i den ursprünglichen Parameterwert, β_f den aktualisierten Parameterwert, x_i die input Variablen und y_i deren Zielwert. Der Skalierungsfaktor α wird als Lernrate bezeichnet und beeinflusst wie stark die Parameter angepasst werden. Um starke Fluktuationen der Parameter zu vermeiden, kann der Mittelwert des Gradienten über eine bestimmte Anzahl von Trainingsabschnitten N_b (genannt *Batch Size*) genommen werden.

$$\beta_f = \beta_i - \alpha \left(\frac{1}{N_b} \sum_{i=0}^{N_b-1} \nabla L(x_i, y_i, \beta_i) \right). \quad (4.4)$$

Dies führt zu einer stabileren Konvergenz und beschleunigt den Trainingsprozess. Da die Lernrate α einmal zu Beginn des Trainingsprozesses festgelegt wurde, ist sie wohlmöglich nicht für jeden Trainingsabschnitt optimal. Um dem entgegenzuwirken, wird das *Momentum* m [30] eingeführt. Dieser Parameter ersetzt den Gradienten mit dem exponentiell gleitenden Durchschnitt des aktuellen (m_i)

und vorherigen Gradienten (m_p) wie folgt:

$$\beta_f = \beta_i - \alpha \cdot m_i \quad \text{mit} \quad m_i = \gamma \cdot m_p + (1 - \gamma) \left(\frac{1}{N_b} \sum_{i=0}^{N_b-1} \nabla L(x_i, y_i, \beta_i) \right). \quad (4.5)$$

Hier bezeichnet γ einen frei wählbaren Parameter und der Index p bezeichnet die Werte des vorherigen Gradienten. Ein Optimisierungsalgorithmus, der diese Momente nutzt, ist die *Adaptive Momentum Estimation* (kurz *Adam*) [31]. Adam kann eine individuelle parameterabhängige Lernrate selbstständig bestimmen. Die Lernrate wird mit Hilfe der ersten und zweiten Momentums des Gradienten bestimmt. Zusätzlich kann Adam den jeweiligen Gradienten über den exponentiell gleitenden Durchschnitt der Gradienten wie folgt anpassen:

$$m_t = \delta_1 m_p + (1 - \delta_1) \left(\frac{1}{N_b} \sum_{i=0}^{N_b-1} \nabla L(x_i, y_i, \beta_i) \right), \quad (4.6)$$

$$v_t = \delta_2 v_p + (1 - \delta_2) \left(\frac{1}{N_b} \sum_{i=0}^{N_b-1} \nabla L(x_i, y_i, \beta_i) \right)^2, \quad (4.7)$$

$$\beta_f = \beta_i - \frac{\alpha}{\sqrt{\hat{v}_i} + \varepsilon} \cdot \hat{m}_i \quad \text{mit} \quad \hat{m}_i = \frac{m_i}{1 - \delta_1} \quad \text{und} \quad \hat{v}_i = \frac{v_i}{1 - \delta_2}. \quad (4.8)$$

Hier bezeichnen \hat{m}_i und \hat{v}_i die Bias Korrekturen, die über Gleichung (4.6) und Gleichung (4.7) definiert sind. Die Parameter $\delta_{1,2}$ und ε werden zu Beginn des Trainings festgelegt [32]. Die Optimierung jedes Trainingsabschnitts im vollständigen Datensatz wird als *Epoche* bezeichnet.

4.3 Optimierung eines neuronalen Netzwerkes

Bestimmte Eigenschaften des neuronalen Netzwerkes werden vor Beginn des Trainings festgelegt und können für den gesamten Prozess nicht mehr geändert werden. Diese Eigenschaften werden als *Hyperparameter* bezeichnet. Dazu zählen die Anzahl an Schichten, die Anzahl an Knoten pro Schicht, die Aktivierungsfunktion, die Batch Size und die Anzahl der Epochen. Um die bestmögliche Leistung und Generalisierungsfähigkeit des Netzwerkes zu erhalten, müssen die Hyperparameter des Netzwerkes optimal gewählt werden.

4.3.1 Leistungsbewertung

Die Leistung eines Binärklassifikators wird anhand der sogenannten *Receiver Operating Characteristic* (ROC) Kurve gemessen. Der Binärklassifikator liefert einen kontinuierlichen Output-Wert zwischen Null und Eins. Man kann für einen Schwellwert die Falsch-Positiv-Rate (FPR, Gleichung (4.9)) und Richtig-Positiv-Rate (*True-Positive-Rate*, TPR, Gleichung (4.10)) berechnen. Variiert man diesen Schwellwert in einem bestimmten Intervall und berechnet dabei die FPR und TPR, so erhält man die

ROC-Curve.

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}, \quad (4.9)$$

$$\text{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \quad (4.10)$$

Die Fläche unterhalb der Kurve (*Area Under Curve*, AUC) wird zur Beurteilung der Leistung verwendet. Liefert das Netzwerk eine perfekte Klassifikation, ist die $\text{AUC} = 1$. Im schlechtesten Fall trifft das Netzwerk rein zufällige Entscheidungen. In diesem Fall ist die $\text{AUC} = 0,5$. Bei $\text{AUC} = 0$ wurde die Klassifikation vertauscht. Eine ROC-Kurve mit hervorragender Klassifikatorleistung ist exemplarisch in Abb. 4.2 dargestellt.

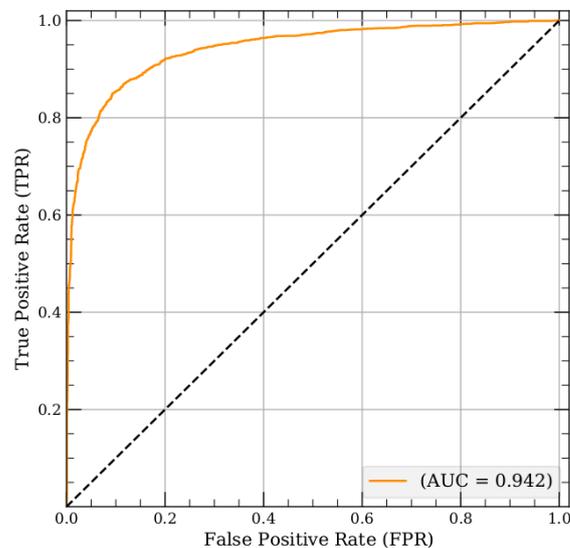


Abbildung 4.2: Darstellung einer ROC-Kurve mit sehr guter Klassifikatorleistung (Entnommen aus [33]).

4.3.2 Generalisierungsfähigkeit

Um die Generalisierungsfähigkeit des Netzwerkes zu untersuchen, wird ein Teil der Trainingsdaten dem Netzwerk vorenthalten und als Testdatenset verwendet. Bei einem Netzwerk, dessen Größe und Komplexität für das Datenset nicht gerechtfertigt ist, tritt sogenanntes *Overtraining* auf. *Overtraining* bedeutet, dass das Netzwerk lediglich das gesamte Trainingsdatenset auswendig gelernt hat und nicht die darunterliegenden Muster gefunden hat. Es gibt also ein starkes Missverhältnis zwischen den Resultaten des Netzwerkes auf dem Trainingsdatenset und dem Testdatenset. Ein Weg um *Overtraining* zu vermeiden ist es beispielsweise, bei jedem Trainingsschritt einen gewissen Prozentsatz von Knoten zu deaktivieren. Die deaktivierten Knoten werden dann bei der Berechnung der neuen Netzwerkparameter ignoriert. Man bezeichnet diese Methode als *Dropout* [34]. Sie reduziert die Koabhängigkeit zwischen zwei bestimmten Knoten und verringert den Fokus des Netzwerkes auf bestimmte Verbindungen. Eine weitere Methode, um die Geschwindigkeit und Stabilität des Trainings zu verbessern und *Overfitting* zu verhindern, bietet die *Batch Normalisation* [35]. Bei dieser Methode

wird der Input einer Schicht so skaliert, dass die Werte von jedem Batch einen Mittelwert von Null und eine Standardabweichung von Eins haben. Dies verhindert einen internal covariance shift, bei dem die Gewichtungen nach jeder Netzwerk Schicht eine neue Verteilung mit anderem Mittelwert und neuer Standardabweichung haben. Ohne Batch Normalisation müsste sich das Netzwerk in jedem Durchlauf auf neue statistische Eigenschaften einstellen, was das Training deutlich verlangsamen würde. Zudem würden Ausreißer in Trainingsdaten zu starken Anpassungen des Netzwerkes führen, was wiederum die Generalisierungsfähigkeit des Netzwerkes reduzieren würde. Es ist jedoch nicht immer möglich Overtraining komplett zu vermeiden. In diesen Fällen kann das Training frühzeitig gestoppt werden. Eine geeignete Methode, um einen frühzeitigen Stop zu veranlassen, ist die Entwicklung der Differenz zwischen der Loss-Funktion auf Trainingsdaten und Testdaten nach jeder Epoche zu überwachen. Um den Einfluss statistischer Fluktuationen zu vermeiden, wird eine Anzahl von Epochen festgelegt, über die sich die Loss-Funktion verschlechtern kann, bevor frühzeitig gestoppt wird. Eine naheliegende Strategie, um die Hyperparameter eines Netzwerkes zu optimieren, ist, innerhalb eines festgelegten Parameterintervalls über alle möglichen Kombinationen von Parametern zu scannen. Dieses Verfahren wird als *Gridsearch* bezeichnet. Die Kombination mit den besten Resultaten wird dann verwendet. Die Suche nach diesen Parametern kann parallelisiert werden und eignet sich somit zur Durchführung auf großen Rechenclustern.

Optimierung der Photon-Identifikation mittels neuronaler Netzwerke

Ziel dieser Arbeit ist es, die Photon-Identifikation (Photon ID) beim ATLAS Experiment mit Hilfe von neuronalen Netzwerken zu optimieren und dabei die Unsicherheiten zu verringern. Der Ansatz, der zu diesem Zweck in dieser Arbeit verfolgt wird, sieht vor, jeweils zwei Binärklassifikatornetzwerke für sowohl konvertierte Photonen als auch unkonvertierte Photonen zu trainieren. Jeweils ein Binärklassifikatornetzwerk wird auf die narrow-strip Variablen und ein separates Binärklassifikatornetzwerk auf die relaxed-tight Variablen aus Monte Carlo Simulationen trainiert. Die Netzwerke sollen dabei lernen, prompt Photonen von fake Photonen in Messdaten zu unterscheiden. Der Output der Netzwerke wird dann verwendet, um die Photonen aus den Messdaten in die korrekten Regionen (Abb. 3.1) der Matrix Methode (Abschnitt 3.3.1) einzuteilen. Es wird erwartet, dass insbesondere die closure Unsicherheit (Abschnitt 3.3.2), die bei der Photon ID mit der Matrix Methode eine bedeutende Rolle spielt, dank dieser Methode verringert werden kann. Die Vorgehensweise der Implementierung der Analyse mittels neuronaler Netzwerke wird in diesem Kapitel erläutert.

5.1 Training und Optimierung des Binärklassifikators

Die Binärklassifikator Netzwerke, die für diese Arbeit genutzt wurden, basieren auf dem Netzwerk, das im Rahmen der Masterarbeit von Florian Kirfel [33] implementiert wurde. Sie nutzen die Python Bibliothek *Tensorflow* [36] mit dem *Keras* [37] Interface. Die Netzwerke wurden auf Signal und Untergrund Ereignisse mit Daten aus Monte Carlo Simulationen trainiert. Die Ereignisse in den Monte Carlo Simulationen wurden mit *Pythia8* [38] generiert. Für die Simulation der Reaktion des Detektors wurde das *Geant4* [39] Toolkit verwendet.

Als Signal Ereignisse werden erstens γ + jets Ereignisse führender Ordnung, die von den Streuprozessen $qg \rightarrow \gamma q$ oder $q\bar{q} \rightarrow \gamma g$ her stammen, verwendet. Zweitens werden Fragmentations-Photonen, die bei der Zersplitterung eines Partons bei hohem p_T entstehen, als Signal Ereignisse genommen.

Die Untergrund Ereignisse bestehen aus fake Photonen, die durch jegliche $2 \rightarrow 2$ QCD Prozesse (mit Ausnahme der γ + jets Ereignisse durch Quark Fragmentation) produziert werden können.

Es wurden zwei Netzwerke auf die relaxed-tight Variablen und die narrow-strip Variablen trainiert. Eine Unterteilung in die verschiedenen η - und p_T -Intervalle aus Gleichung (3.1) wurde bei den Trainingsdaten nicht vorgenommen. Die Einteilung in konvertierte Photonen und unkonvertierte

Photonen wurde jedoch beibehalten, da sich die shower-shape Variablen für beide Photonarten stark voneinander unterscheiden.

5.1.1 Gesamte Ereignisanzahl in Signal und Untergrund Monte Carlo Dateien

Zum Training der Netzwerke wurden die zwei Datensätze mit allen Signal und allen Untergrund Monte Carlo Daten erzeugt. Die Anzahl der Ereignisse, die den Netzwerken im Untergrund- und Signal-Datensatz zur Verfügung stehen, sind in Tabelle 5.1 und Tabelle 5.2 angegeben. Aus den

Tabelle 5.1: Ereigniszahl pro η - und p_T -Intervall in allen Untergrund Dateien zusammengenommen.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	34997	12412	34623	17400	8034	13242	120708
$30 \leq p_T < 35$	20456	7668	21158	10550	4675	7950	72457
$35 \leq p_T < 40$	18553	6633	17995	9193	4304	7569	64247
$40 \leq p_T < 45$	11925	4417	12121	6170	2793	4906	42332
$45 \leq p_T < 50$	9386	3433	9617	5026	2114	3878	33454
$50 \leq p_T < 60$	16547	5847	16167	8432	3999	6958	57950
$60 \leq p_T < 80$	12003	4247	11853	6402	2968	5227	42700
$80 \leq p_T < 100$	3383	1241	3309	1828	889	1502	12152
$100 \leq p_T < 125$	1392	541	1484	765	387	689	5258
$125 \leq p_T < 150$	562	171	543	286	132	259	1953
$150 \leq p_T < 175$	239	94	228	129	66	108	864
$175 \leq p_T < 250$	178	76	216	95	67	115	747
$250 \leq p_T < 350$	42	19	51	19	13	26	170
$350 \leq p_T < 1500$	15	6	23	7	2	6	59
Total p_T	129678	46805	129388	66302	30443	52435	455051

Tabelle 5.2: Ereigniszahl pro η - und p_T -Intervall in allen Signal Dateien zusammengenommen.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	153330	50847	148896	73047	46488	82680	555288
$30 \leq p_T < 35$	119975	42437	128228	66330	39780	71852	468602
$35 \leq p_T < 40$	476742	157076	435421	207189	141891	265422	1683741
$40 \leq p_T < 45$	327972	108852	307301	149842	96543	172907	1163417
$45 \leq p_T < 50$	248452	85495	247152	122610	78079	144343	926131
$50 \leq p_T < 60$	988740	324943	911563	433649	300199	544601	3503695
$60 \leq p_T < 80$	1430943	473450	1343199	639187	444347	801801	5132927
$80 \leq p_T < 100$	858794	282097	819908	397397	265064	473836	3097096
$100 \leq p_T < 125$	390532	127950	370574	183342	122376	219439	1414213
$125 \leq p_T < 150$	577371	188947	544453	272729	188334	328905	2100739
$150 \leq p_T < 175$	718866	239075	694065	336342	226155	383493	2597996
$175 \leq p_T < 250$	678703	224515	650880	314956	205231	338961	2413246
$250 \leq p_T < 350$	1468474	485930	1401233	617324	386728	549571	4909260
$350 \leq p_T < 1500$	4545346	1448922	3644355	1090267	515212	486840	11730942
Total p_T	12984240	4240536	11647228	4904211	3056427	4864651	41697293

Tabellen geht klar hervor, dass die Anzahl der Ereignisse pro Intervall in den Untergrund Dateien deutlich geringer ist, als in den Signal Dateien. Insbesondere im hohen p_T -Bereich sind nur sehr wenige Untergrund Ereignisse vorhanden. Somit ist zu erwarten, dass die Klassifizierungsleistung der Binärklassifikatoren insbesondere im hohen p_T -Bereich abnehmen sollte.

5.1.2 Vergleich mit tight-ID Arbeitspunkt

Zum einfachen Vergleich zwischen der Klassifikationsleistung der neuronalen Netzwerke und der traditionellen Cut-basierten Methode, muss die Receiver Operating Characteristic, die die Leistung eines Binärklassifikators misst (siehe Abschnitt 4.3.1), modifiziert werden. Die modifizierte Form der ROC zeigt die Untergrund Aussortierung (*Background Rejection* b_R , Gleichung (5.2)) als Funktion der *Signal Efficiency* (s_E , Gleichung (5.1)).

$$s_E = \frac{\text{Signal Photonen, die korrekt identifiziert wurden}}{\text{Alle Signal Photonen}}, \quad (5.1)$$

$$b_R = 1 - \frac{\text{Untergrund Photonen, die als Signal identifiziert wurden}}{\text{Alle Untergrund Photonen}}. \quad (5.2)$$

An sich stellt diese Modifikation nur eine Vertauschung der beiden Achsen der ROC-Kurve dar. Der AUC Wert bleibt dabei unverändert. Der tight-ID Arbeitspunkt wird zum direkten Vergleich mit der Cut-basierten Methode in jeder Grafik mit angegeben. Der tight-ID Arbeitspunkt kann direkt aus den Monte Carlo Simulationen berechnet werden, indem man die gesamten tight-Selektionskriterien auf sie anwendet. In Abb. 5.1 ist eine ROC-Kurve zusammen mit einer modifizierten ROC-Kurve zum Vergleich dargestellt.

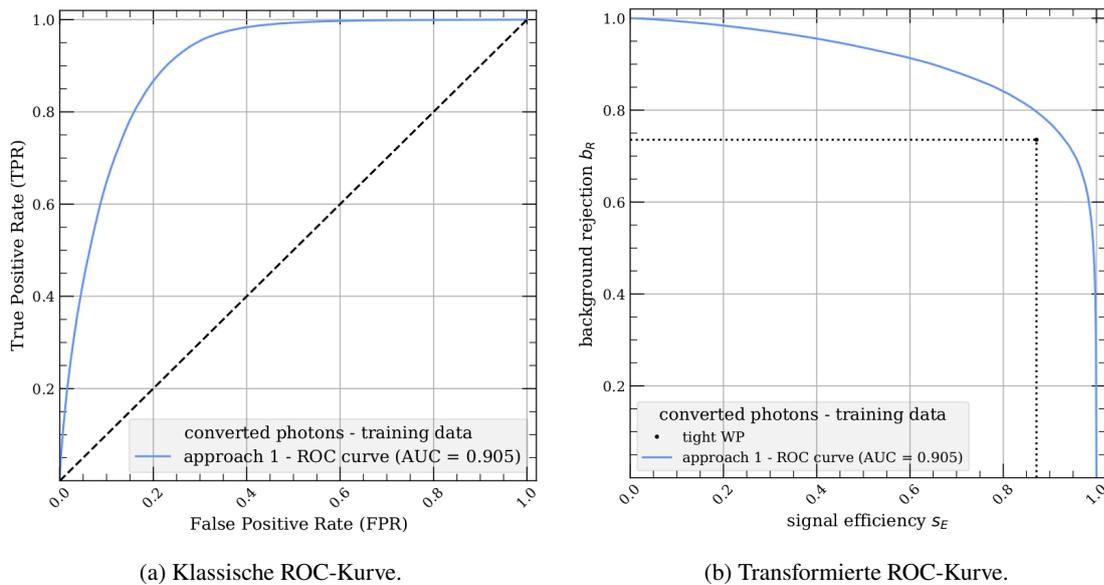


Abbildung 5.1: Darstellung der Transformation der ROC-Kurve zur Evaluation der Netzwerkleistung (Entnommen aus [33]).

5.1.3 Hyperparameter Optimierung

Wie bereits in Abschnitt 4.3 beschrieben wurde, ist die Optimierung der Netzwerk Hyperparameter ein essenzieller Schritt zur Erzielung guter Klassifizierungsleistung. In diesem Abschnitt werden die Parameter und Optimierungsstrategien für die neuronalen Netzwerke beschrieben. Alle Netzwerke in

dieser Arbeit nutzen die Sigmoid Aktivierungsfunktion in der Output Schicht (vgl. Tabelle 4.1), wodurch das Intervall der Output Werte auf $[0, 1]$ beschränkt ist. Die restlichen Schichten nutzen die ReLU Aktivierungsfunktion. Als Loss-Funktion wurde die Binary-Cross-Entropy (Gleichung (4.2)) verwendet und als Optimierungsalgorithmus wurde Adam (siehe Abschnitt 4.2) genutzt. Zur Vermeidung von Overtraining wurde die Methoden Dropout, Batch Normalisation und Early Stopping genutzt (vgl. Abschnitt 4.3.2). Die Aufteilung des Datensets in Trainingsdaten und Testdaten wurde mit einem Teilungsverhältnis von 60 zu 40 vorgenommen. Die optimale Anzahl der Schichten sowie die Anzahl der Knoten pro Schicht wurde über eine Gridsearch ermittelt. Die Parameterintervalle, die in der Gridsearch abgerastert wurden, sind in Tabelle 5.3 angegeben. Von den 280 Netzwerken, die im Rahmen

Tabelle 5.3: Hyperparameterbereich der Gridsearch.

Hyperparameter	Intervall	Schrittweite
Versteckte Schichten	[1, 10]	linearer Anstieg um 1
Knoten pro Schicht	[16, 112]	linearer Anstieg um 16

der Gridsearch trainiert wurden, wurden die vier Netzwerke, die die beste Klassifizierungsleistung aufzeigen konnten, ausgewählt. Zur Feinabstimmung der Hyperparameter wurde anschließend als Crosscheck eine Bayes'sche Optimierung [40] in einem kleinen Intervallbereich um die grob ermittelten Werte der Gridsearch herum, durchgeführt. Die Bayes'sche Optimierung konnte zu keiner Verbesserung der Klassifizierungsleistung beitragen, was als Verifikation der Güte der zuvor ermittelten Parameter angesehen werden kann. Es wurden schließlich die Hyperparameter, die durch die Gridsearch ermittelt wurden, für die finalen Netzwerke verwendet.

5.1.4 Narrow-Strip Binärklassifikatoren

Die finalen Hyperparameter für die Binärklassifikator Netzwerke für die narrow-strip Variablen sind in Tabelle 5.4 angegeben. Die Plots der modifizierten ROC-Kurven sind für die beiden Netzwerke in

Tabelle 5.4: Hyperparameter der narrow-strip Binärklassifikatoren für konvertierte und unkonvertierte Photonen.

Hyperparameter	Konvertierte Photonen	Unkonvertierte Photonen
Versteckte Schichten	2	2
Knoten pro Schicht	112	96
Lernrate	0.001	0.001
Dropout Prozentsatz	30 %	30 %
Batchgröße	16384	16384
Anzahl an Epochen	500	500

Abb. 5.2 dargestellt. Bei beiden Netzwerke liegen die Kurven für Trainingsdaten und Testdaten nahezu exakt übereinander. Overtraining konnte also in beiden Fällen vermieden werden. Obwohl die beiden Netzwerke auf einer geringeren Anzahl von Variablen trainiert wurden, als der tight-ID Arbeitspunkt zur Klassifikation der Photonen nutzt, erreichen beide Netzwerke eine Klassifizierungsleistung, die sehr nahe an den tight-ID Arbeitspunkt herankommt. Die Outputs aller Netzwerke für Signal und Untergrund Photonen sind in Anhang A zu finden. Die Plots der Loss Kurven für die Trainingsdaten und Testdaten sind in Abb. 5.3 dargestellt. Anhand der Loss Kurven kann gesehen werden, dass

5.1 Training und Optimierung des Binärklassifikators

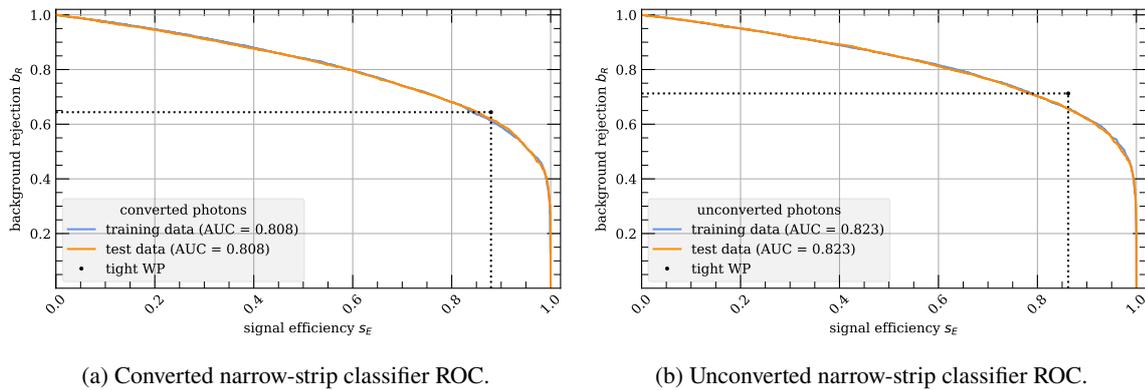


Abbildung 5.2: ROC-Kurven für narrow-strip Binärklassifikator.

das Training bei beiden Netzwerken nach deutlich weniger als 500 Epochen beendet wurde. Bei beiden Netzwerken wurde also Early Stopping verwendet. Bei geringer Anzahl an Epochen fällt bei beiden Netzwerken die Loss Kurve auf den Trainingsdaten deutlich schneller ab als auf den Testdaten. Die Loss Kurve für Trainingsdaten zeigt zudem höhere Fluktuationen als auf Testdaten, was darauf zurückzuführen ist, dass beim Training die Dropout Methode angewandt wurde.

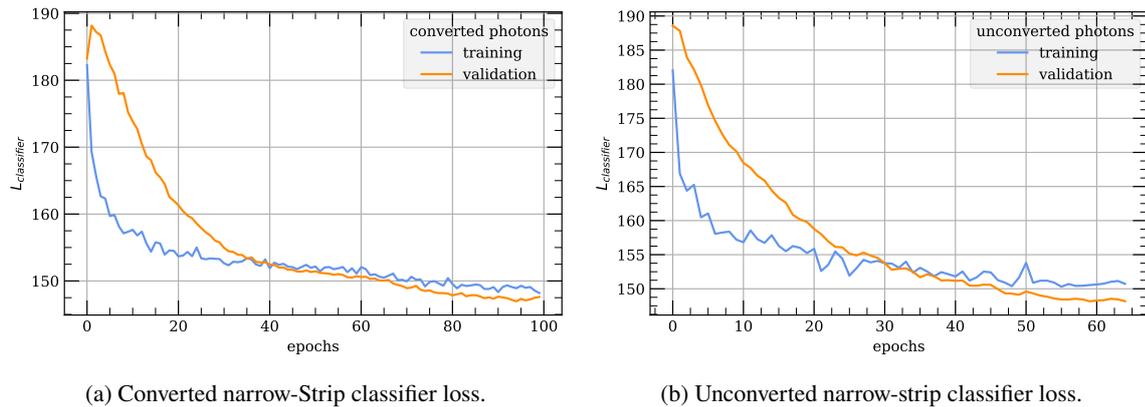


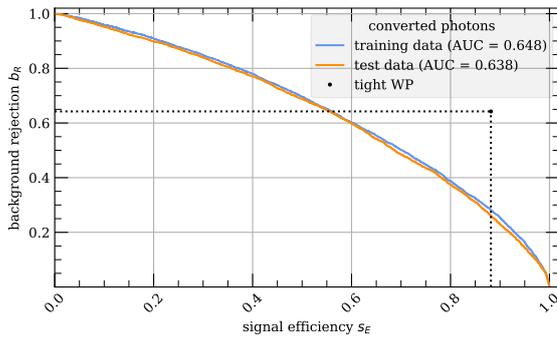
Abbildung 5.3: Loss-Kurven der narrow-strip Binärklassifikatoren.

5.1.5 Relaxed-Tight Binärklassifikatoren

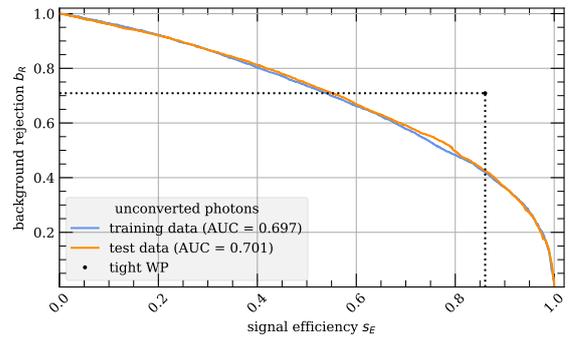
Die Hyperparameter für die Binärklassifikator Netzwerke für die relaxed-tight Variablen sind in Tabelle 5.5 angegeben und die Plots der modifizierten ROC-Kurven sind in Abb. 5.4 dargestellt. Auch hier liegen die Kurven für beide Netzwerke sehr nahe beieinander. Overtraining konnte also erneut vermieden werden. Die Klassifizierungsleistung beider Netzwerke liegt deutlich unterhalb des tight-ID Arbeitspunktes. Die Loss Kurven der beiden Netzwerke sind in Abb. 5.5 dargestellt. Erneut wurde Early Stopping verwendet, um das Training nach deutlich weniger als 500 Epochen zu beenden. In beiden Fällen ist die Loss Kurve auf den Testdaten stets unterhalb der Loss Kurve der Trainingsdaten. Insbesondere für konvertierte Photonen ist bei der Loss Kurve auf Testdaten kein starker Abfall zu

Tabelle 5.5: Hyperparameter der relaxed-tight-Binärklassifikatoren für konvertierte und unkonvertierte Photonen.

Hyperparameter	Konvertierte Photonen	Unkonvertierte Photonen
Versteckte Schichten	3	7
Knoten pro Schicht	80	112
Lernrate	0.001	0.001
Dropout Prozentsatz	30 %	30 %
Batchgröße	16384	16384
Anzahl an Epochen	500	500



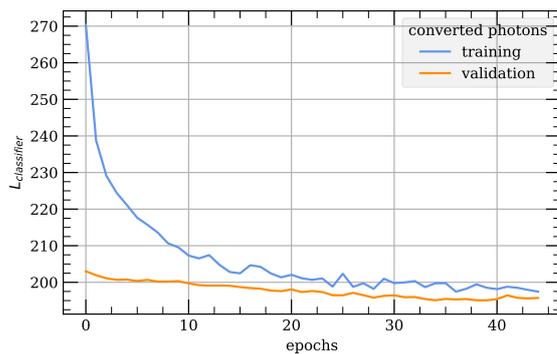
(a) Converted relaxed-tight classifier ROC.



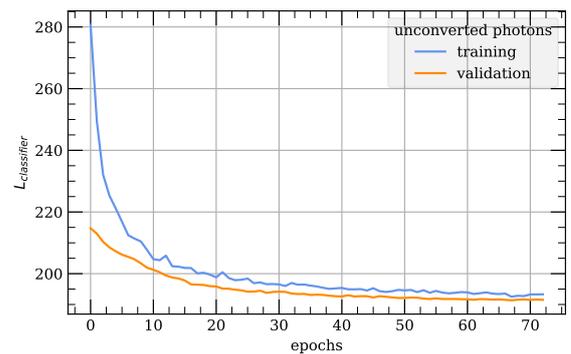
(b) Unconverted relaxed-tight classifier ROC.

Abbildung 5.4: ROC-Kurven für relaxed-tight Binärklassifikatoren.

beobachten. Für unkonvertierte Photonen ist ein Abstieg der Loss Kurve auf Testdaten zu sehen, jedoch viel geringer, als der Abstieg der Loss Kurve der Trainingsdaten. Auch hier sind aufgrund der Dropout Methode stärkere Fluktuationen in der Trainingskurve zu sehen.



(a) Converted relaxed-tight classifier loss.



(b) Unconverted relaxed-tight classifier loss.

Abbildung 5.5: Loss-Kurven der relaxed-tight Binärklassifikatoren.

5.1.6 Resultate der Netzwerkoptimierung

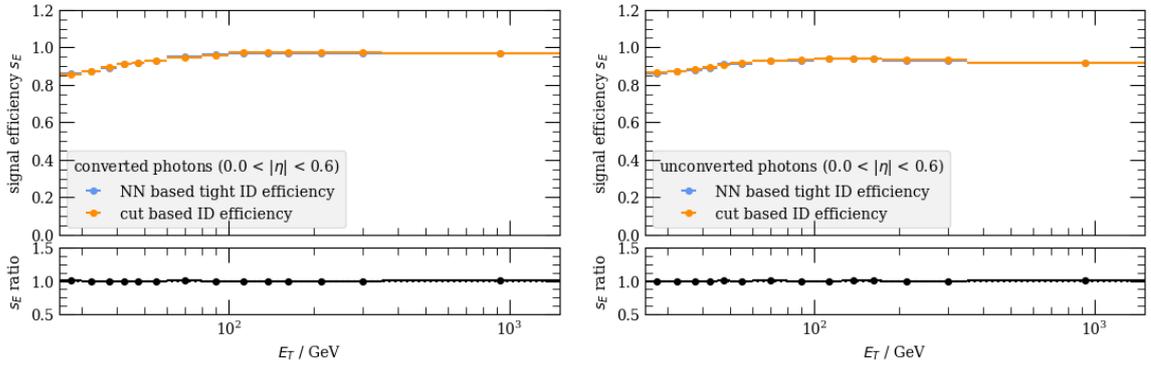
Es ist eine große Diskrepanz zwischen der Klassifizierungsleistung der narrow-strip Netzwerke und der relaxed-tight Netzwerke zu sehen. Während es den narrow-strip Netzwerken möglich ist, prompt Photonen von fake Photonen mit einer Genauigkeit nahe dem tight-ID Arbeitspunkt zu unterscheiden, zeigen die relaxed-tight Netzwerke eine deutlich schlechtere Klassifizierungsleistung. Es ist also davon auszugehen, dass die benötigten Informationen zur Unterscheidung von prompt und fake Photonen vor allem in den narrow-strip Variablen, also in der ersten Schicht des Kalorimeters, vorliegen.

5.2 Optimierung der Cuts auf den Netzwerkoutput

Wie in Abschnitt 5.1.3 beschrieben, schreibt der Binärklassifikator jedem Photon eine reelle Zahl im Intervall $[0, 1]$ als Output zu. Um diesen Zahlenwert für die weitere Analyse nutzen zu können, muss ein Schwellwert (*Cut*-Wert) definiert werden, der überschritten werden muss, damit das Photon als prompt Photon klassifiziert werden kann. Um eine optimale Klassifikation zu erhalten, muss dieser Cut-Wert für jedes Netzwerk in jedem η - und p_T -Intervall separat angepasst werden. Der Output der Netzwerke soll für die Einteilung der Photonen in die Regionen der Matrix Methode verwendet werden. Für jedes Photon muss also sowohl der Output des narrow-strip Binärklassifikators, als auch des relaxed-tight Binärklassifikators berücksichtigt werden. Zur Cut-Optimierung wurden für jedes η - und p_T -Intervall 100 verschiedene Kombinationen von Cuts auf die relaxed-tight und narrow-strip Netzwerk Outputs untersucht. Für jede dieser Kombinationen wurde anschließend die Signal Efficiency s_E und die Background Rejection b_R bestimmt. Zusätzlich wurden s_E und b_R mit direkten Cuts aus den Monte Carlo Daten bestimmt. Durch den Vergleich der Werte für Signal Efficiency und Background Rejection mit den tatsächlichen Werten aus den Monte Carlo Daten konnte für die neuronalen Netzwerke die optimalen Cut-Werte bestimmt werden. Die optimalen Cut-Werte für alle η - und p_T - Intervalle sind in Anhang B angegeben. Das Programm, mit dem diese Cut-Werte berechnet wurden, greift auf das TMVA-Toolkit [41] zurück.

5.2.1 Signal Efficiency s_E

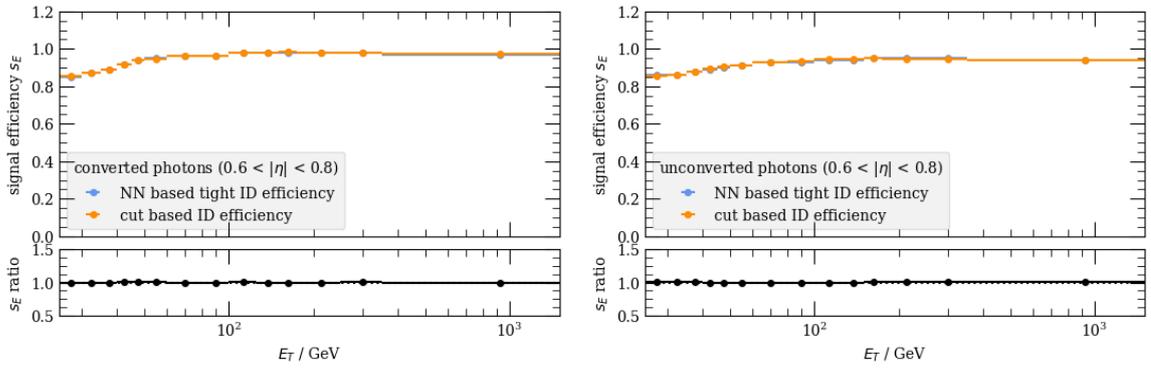
Die Signal Efficiencies in allen η - und p_T - Intervallen sind in Abb. 5.6 bis Abb. 5.11 dargestellt. Wie aus den Abbildungen hervorgeht, konnten die Signal Efficiencies aus den Monte Carlo Daten mit den optimierten Cuts auf die Outputs der neuronalen Netzwerke in allen η - und p_T - Intervallen nahezu exakt reproduziert werden.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

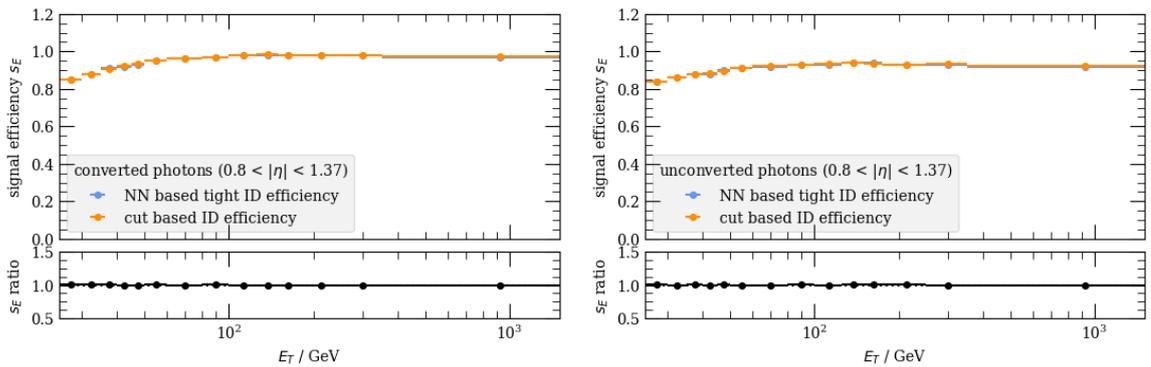
Abbildung 5.6: Signal Efficiency s_E für $0,0 < |\eta| < 0,6$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.7: Signal Efficiency s_E für $0,6 < |\eta| < 0,8$.

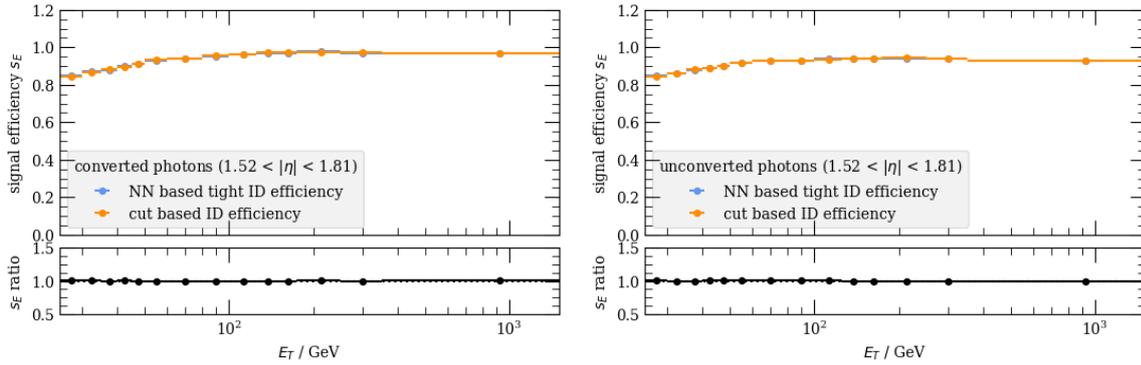


(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.8: Signal Efficiency s_E für $0,8 < |\eta| < 1,37$.

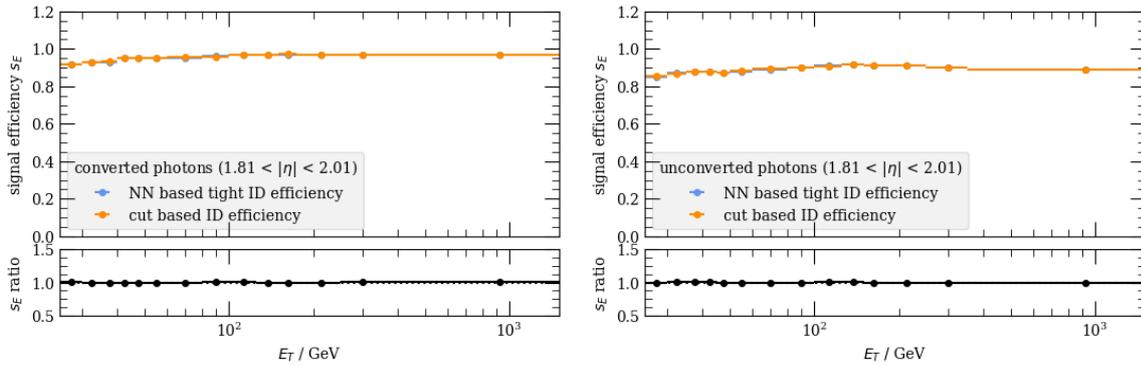
5.2 Optimierung der Cuts auf den Netzwerkoutput



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

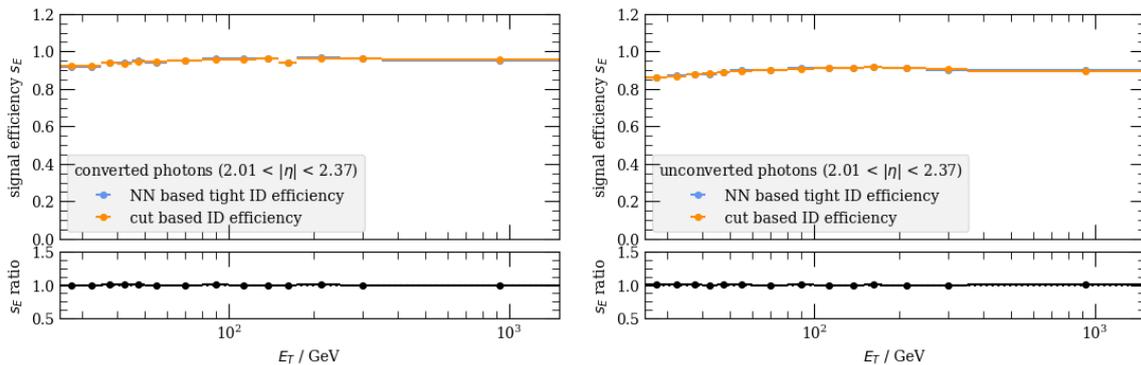
Abbildung 5.9: Signal Efficiency s_E für $1,52 < |\eta| < 1,81$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.10: Signal Efficiency s_E für $1,81 < |\eta| < 2,01$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.11: Signal Efficiency s_E für $2,01 < |\eta| < 2,37$.

5.2.2 Background Rejection b_R

Die Background Rejections für alle η - und p_T - Intervalle sind in Abb. 5.12 bis Abb. 5.17 dargestellt. Wie bereits in Abschnitt 5.1 vermutet, zeigt die Background Rejection der neuronalen Netzwerke insbesondere im hohen p_T -Bereich starke Abweichungen von den Werten aus den Monte Carlo Daten. Im geringen p_T -Bereich sind die Abweichungen zu den Werte aus den Monte Carlo Daten kleiner. Dort handelt es sich vor allem um geringfügige Fluktuationen um den Monte Carlo Wert herum. Dies lässt darauf schließen, dass die geringe Anzahl an Untergrund Ereignissen im hohen p_T -Bereich die Background Rejection der Binärklassifikatoren beeinträchtigt. Die Größe und Richtung der Abweichung ist von Intervall zu Intervall unterschiedlich.

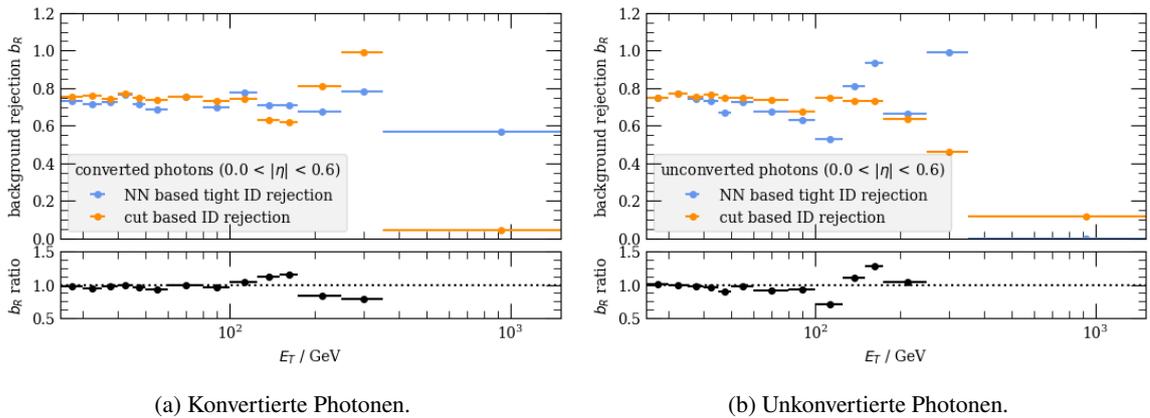


Abbildung 5.12: Background Rejection b_R für $0,0 < |\eta| < 0,6$.

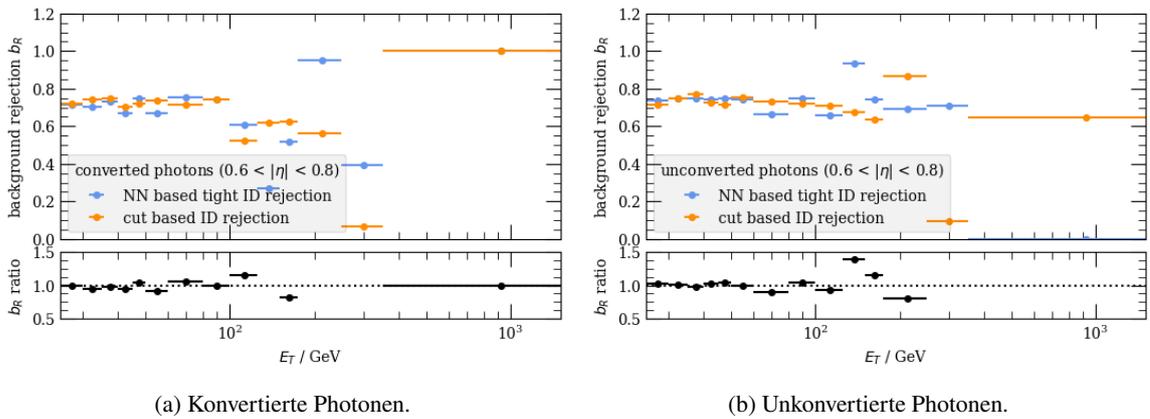
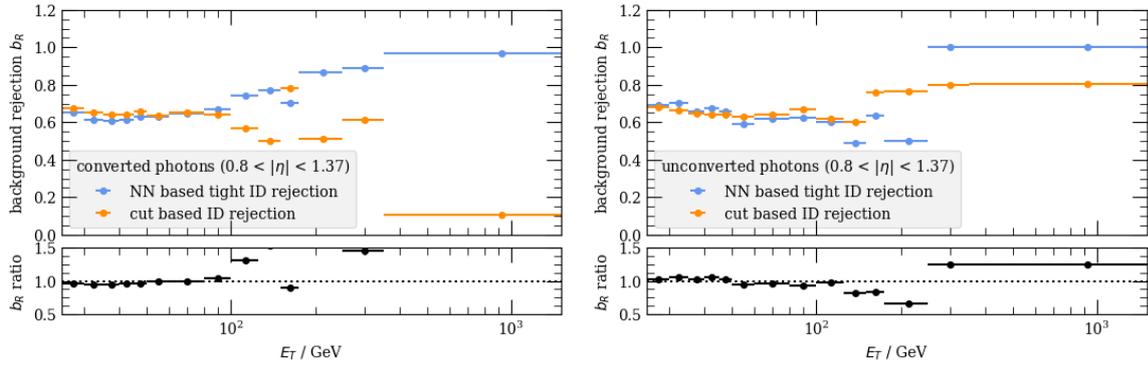


Abbildung 5.13: Background Rejection b_R für $0,6 < |\eta| < 0,8$.

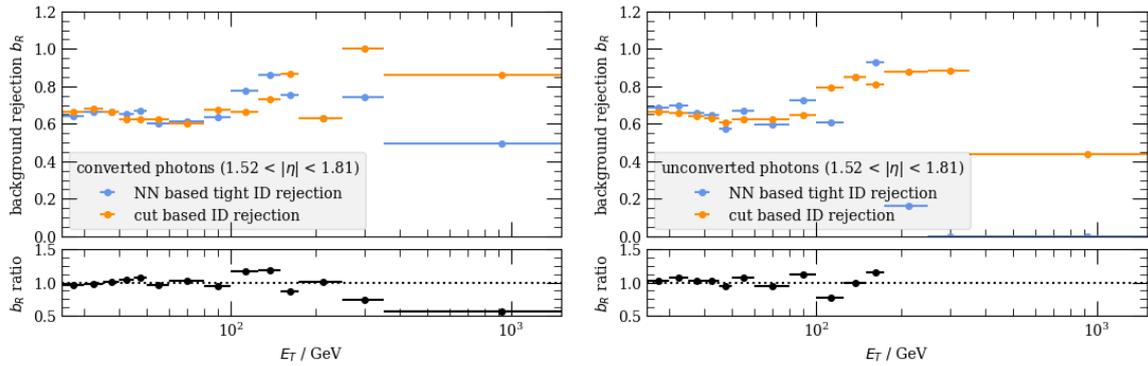
5.2 Optimierung der Cuts auf den Netzwerkoutput



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

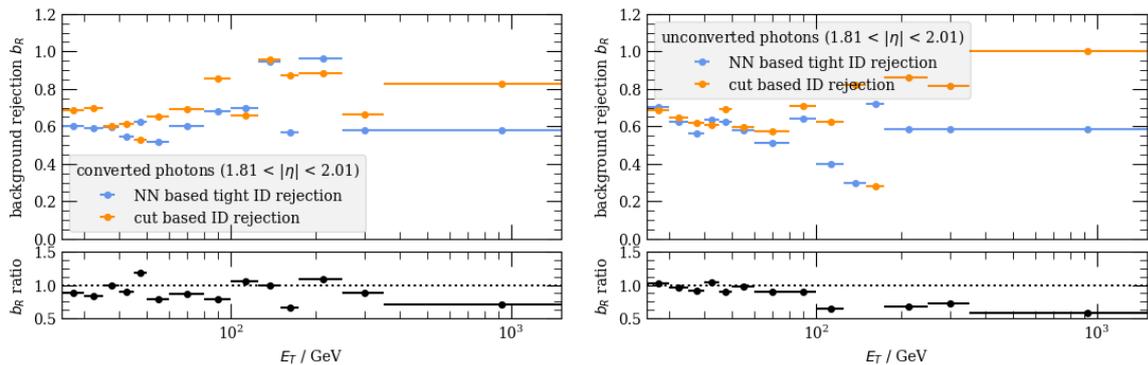
Abbildung 5.14: Background Rejection b_R für $0,8 < |\eta| < 1,37$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

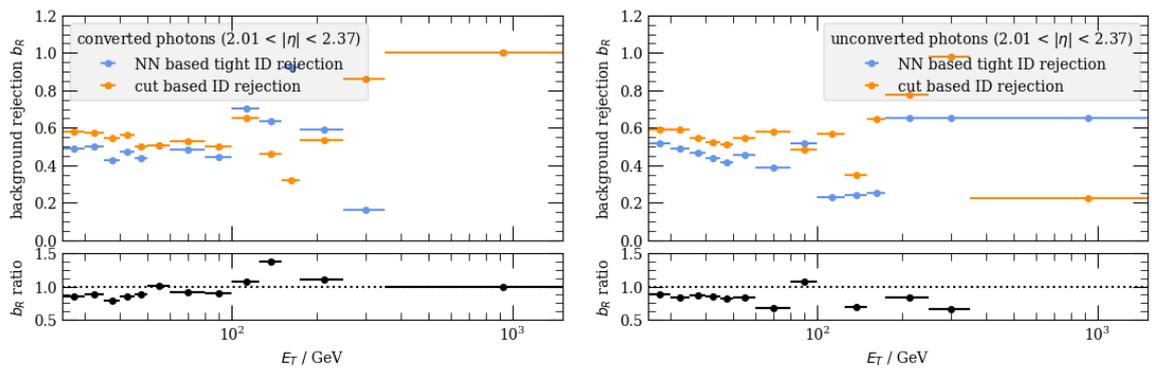
Abbildung 5.15: Background Rejection b_R für $1,52 < |\eta| < 1,81$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.16: Background Rejection b_R für $1,81 < |\eta| < 2,01$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.17: Background Rejection b_R für $2,01 < |\eta| < 2,37$.

5.3 Tight-ID Effizienz und closure Unsicherheit mit Messdaten

Nach Training und Optimierung der Binärklassifikatoren und Optimierung der Cuts auf die Outputs können die Netzwerke genutzt werden, um die tight-ID Effizienzen und closure Unsicherheiten in Messdaten zu bestimmen. In dieser Arbeit werden vorselektierte ATLAS-Daten von 2015 bis 2018 zur Photon ID verwendet. Die Gesamtzahl der Ereignisse in den Datensätzen der Jahre von 2015 bis 2018 sind in Anhang C zu finden. Die Berechnung der tight-ID Effizienzen und closure Unsicherheiten in Messdaten wurde zunächst basierend auf der traditionellen Matrix Methode, ohne die Ergebnisse der Binärklassifikatoren, durchgeführt. Dies dient dazu, Referenzwerte für die tight-ID Effizienzen und closure Unsicherheiten, berechnet in Daten mit der Matrix Methode, zu erhalten. Es ist anzumerken, dass die Spurisolierungseffizienzen $\hat{\epsilon}^b$ und $\hat{\epsilon}_{\text{ID}}^b$ (vgl. Abschnitt 3.3), die in der Matrix Methode zur Berechnung von $\epsilon^{\text{tight-ID}}$ benutzt werden (siehe Gleichung (3.17)) mit ansteigender Transversalenergie E_T monoton fallen und sogar Null werden können. In Intervallen mit hoher E_T führt dies in der Matrix Methode zu unphysikalischen Werten für $\epsilon^{\text{tight-ID}}$. Dies ist ein allgemein bekanntes Problem bei der Matrix Methode und es existieren komplexe Verfahren, um zum Beispiel via der Bootstrap Methode physikalisch sinnvolle Werte zu erhalten. Für diese Bachelorarbeit wurde der Ansatz gewählt, via mehrerer if-else Abfragen die Effizienzen in den betroffenen Intervallen auf Null zu setzen. Die so berechneten Werte für die tight-ID Effizienz und closure Unsicherheit mit der traditionellen Cut-basierten Analyseverfahren zeigen keinerlei Unterschiede zu den Werten, die in der ATLAS-internen Dokumentation [42] bei diesen Datensätzen angegeben wurden. Zum Vergleich mit der Cut-basierten Matrix Methode wurde die Analyse nochmal mit den Ergebnissen der Binärklassifikatoren durchgeführt. Dabei wurden die Outputs der Klassifikatoren verwendet, um die Photonen in die Regionen der Matrix Methode zu einzuteilen. Mit diesen Regionen wurde die Matrix Methode erneut durchgeführt und die tight-ID Effizienz und closure Unsicherheit berechnet.

5.3.1 Tight-ID Effizienz

Die tight-ID Effizienzen der traditionellen Cut-basierten Methode (in den Grafiken als *cut-based* angegeben) und der Methode mit dem Binärklassifikator Output (in den Grafiken als *NN-based* angegeben) sind in Abb. 5.18 bis Abb. 5.23 dargestellt. Die tight-ID Effizienzen, die mit dem Output der Binärklassifikatoren berechnet wurden, zeigen keinen großen Unterschied zu den tight-ID Effizienzen von der traditionellen Cut-basierten Methode. In Abb. 5.18(b) und Abb. 5.19(b) ist über eine breite Anzahl von p_T -Intervallen eine leichte Erhöhung der tight-ID Effizienz zu erkennen. Betrachtet man die Abbildungen in höheren η -Intervallen, so ist dort in den meisten p_T -Intervallen eher eine geringfügige Verringerung der tight-ID Effizienzen zu sehen. Alleine bei den Plots in Abb. 5.22(b) und Abb. 5.23(b) kann man relativ starke Unterschiede zur traditionellen Cut-basierten Methode sehen. Man muss jedoch berücksichtigen, dass diese Plots die tight-ID Effizienzen für unkonvertierte Photonen in hohen η -Intervallen zeigen. Bei diesen η -Intervallen treten in hohen p_T -Intervallen nur sehr wenig fake Photonen auf, die vom Netzwerk klassifiziert werden können (vgl. Tabelle 5.1). Die Unterschiede zwischen der traditionellen Cut-basierte Methode und der Methode mit den Binärklassifikatoren sind somit auf eine schlechte Optimierung der Netzwerke, auf Grund geringer Statistik in den jeweiligen Intervallen, zurückzuführen.

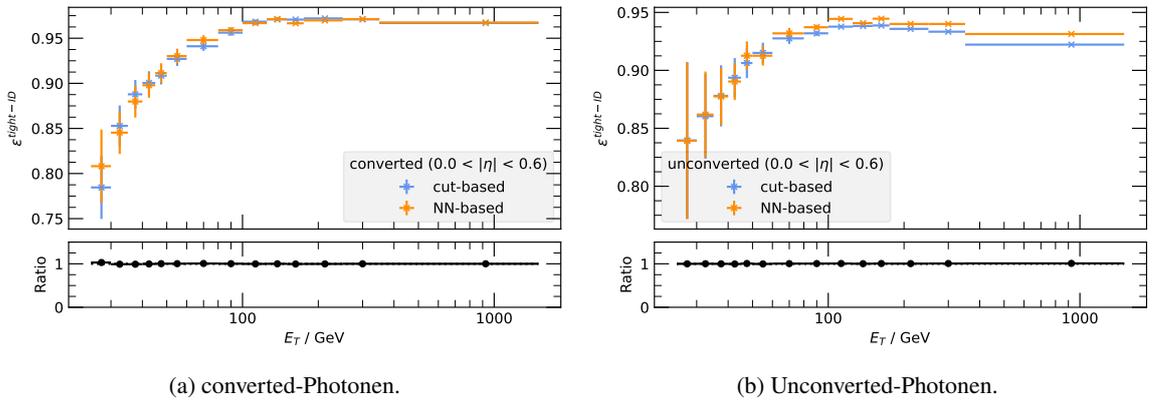


Abbildung 5.18: Tight-ID Effizienz $\varepsilon^{\text{tight-ID}}$ für $0,0 < |\eta| < 0,6$.

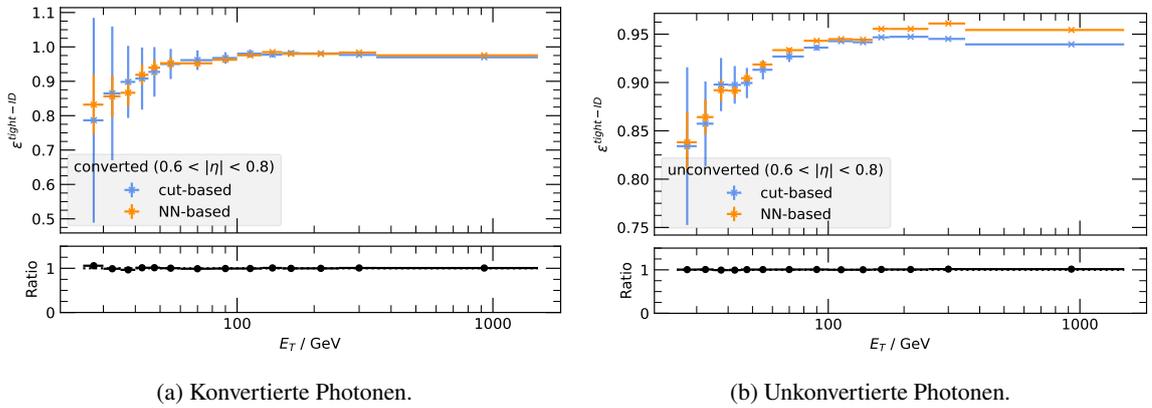


Abbildung 5.19: Tight-ID Effizienz $\varepsilon^{\text{tight-ID}}$ für $0,6 < |\eta| < 0,8$.

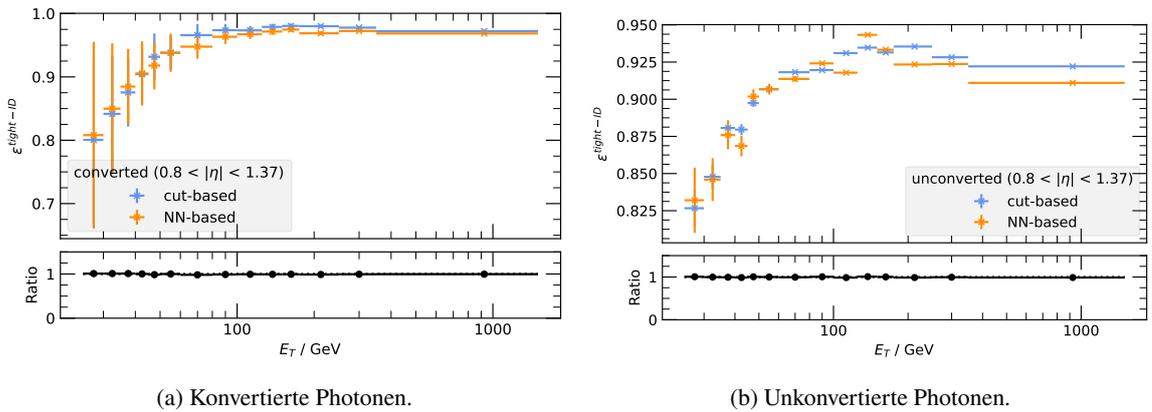


Abbildung 5.20: Tight-ID Effizienz $\varepsilon^{\text{tight-ID}}_{b_R}$ für $0,8 < |\eta| < 1,37$.

5.3 Tight-ID Effizienz und closure Unsicherheit mit Messdaten

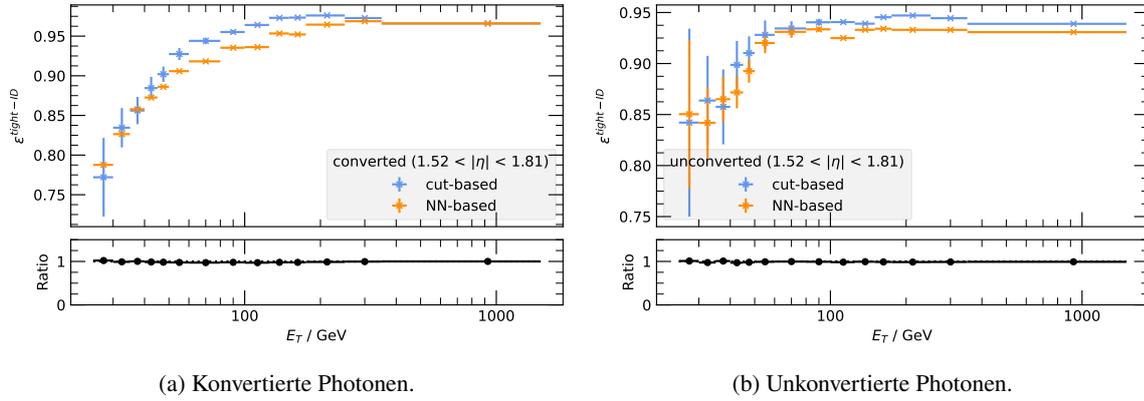


Abbildung 5.21: Tight-ID Effizienz $\epsilon^{\text{tight-ID}} b_R$ für $1,52 < |\eta| < 1,81$.

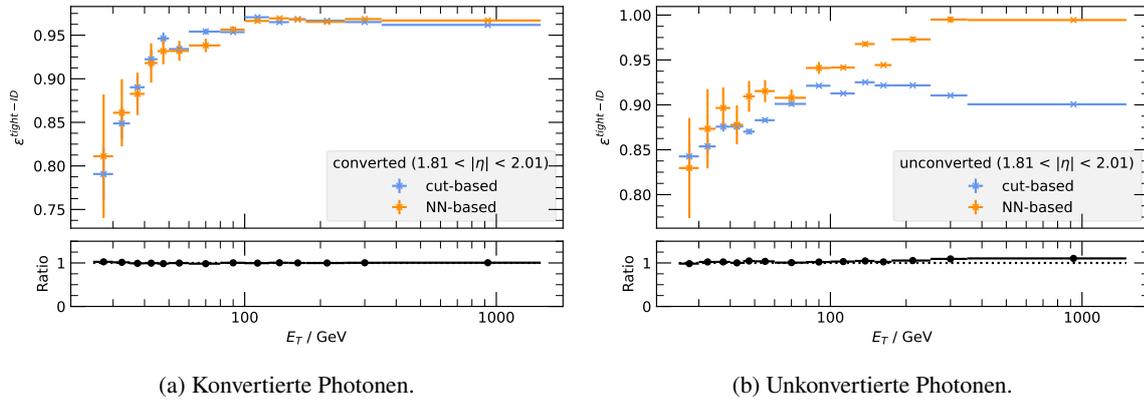


Abbildung 5.22: Tight-ID Effizienz $\epsilon^{\text{tight-ID}}$ für $1,81 < |\eta| < 2,01$.

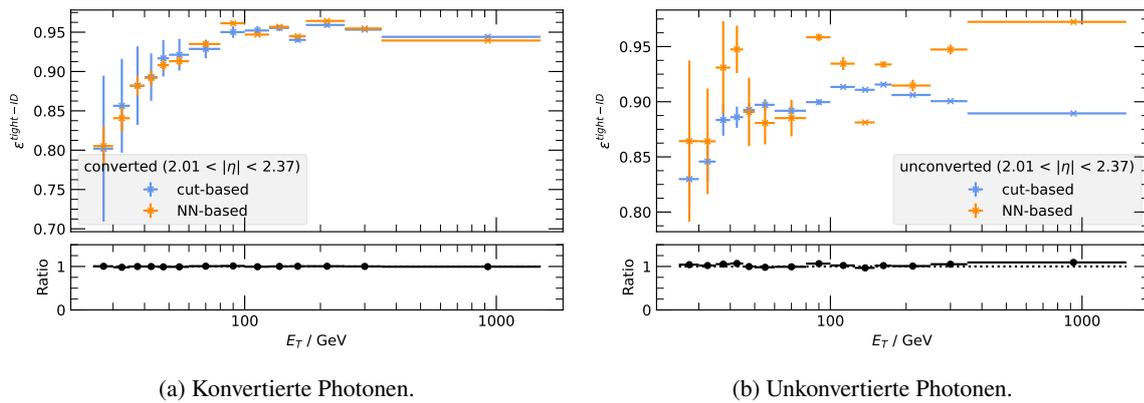


Abbildung 5.23: Tight-ID Effizienz $\epsilon^{\text{tight-ID}}$ für $2,01 < |\eta| < 2,37$.

5.3.2 Closure Unsicherheit

Die closure Unsicherheiten sind in Abb. 5.24 bis Abb. 5.29 für beide Methoden dargestellt. Das Hauptziel, das mit der Analysemethode via der Binärklassifikatoren verfolgt wurde, war es, die closure Unsicherheit bei gleich bleibender Signal Efficiency bei der Photon ID zu minimieren (siehe Abschnitt 3.3). Dieses Ziel konnte in mehreren η -Intervallen erfüllt werden. Die Unsicherheiten in Abb. 5.25 und Abb. 5.27 konnten sowohl für converted als auch unkonvertierte Photonen erkennbar reduziert werden. Zudem sind in Abb. 5.26(a) und Abb. 5.29(a) sehr starke Verringerungen der closure Unsicherheiten bei konvertierten Photonen zu betrachten. In anderen η -Intervallen konnten die closure Unsicherheiten der traditionellen Cut-basierten Methode mit den neuronalen Netzwerken reproduziert werden, wie zum Beispiel in Abb. 5.24. In wiederum anderen η -Intervallen wurde die closure Unsicherheit durch die neuronalen Netzwerke erhöht (vgl. Beispielsweise Abb. 5.28). Insgesamt sind also die Ergebnisse der closure Unsicherheiten, die für die verschiedenen η - und p_T -Intervalle in dieser Arbeit erhalten werden konnten, sehr unterschiedlich. Eine grundsätzliche Verbesserung scheint möglich zu sein, würde jedoch weitere Optimierung und Analyse pro η - und p_T -Intervall benötigen.

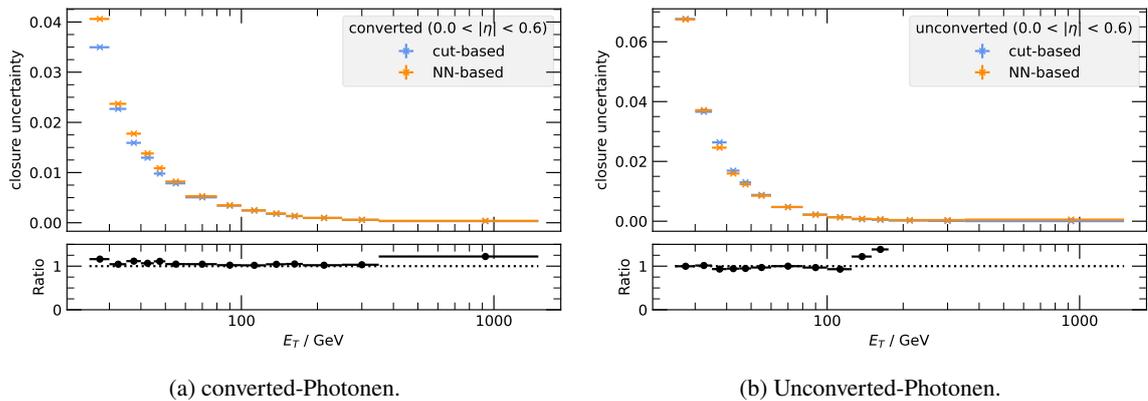


Abbildung 5.24: Closure Unsicherheit für $0,0 < |\eta| < 0,6$.

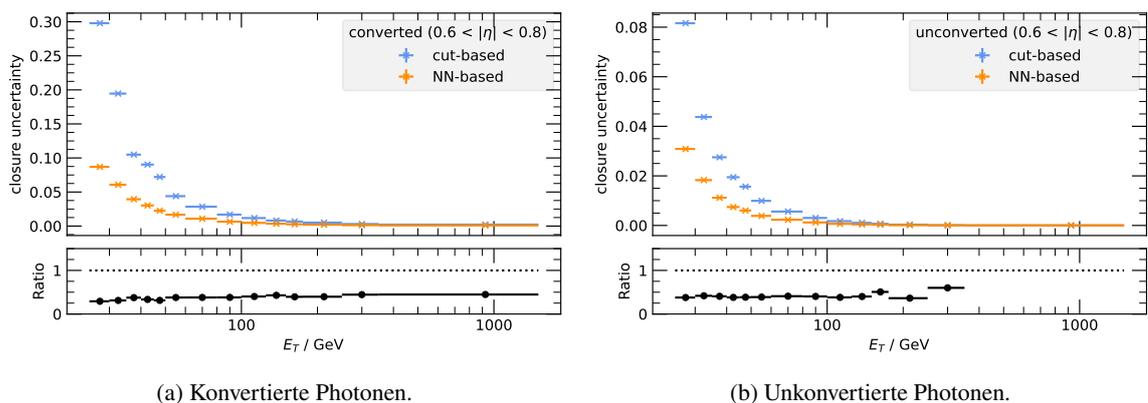


Abbildung 5.25: Closure Unsicherheit für $0,6 < |\eta| < 0,8$.

5.3 Tight-ID Effizienz und closure Unsicherheit mit Messdaten

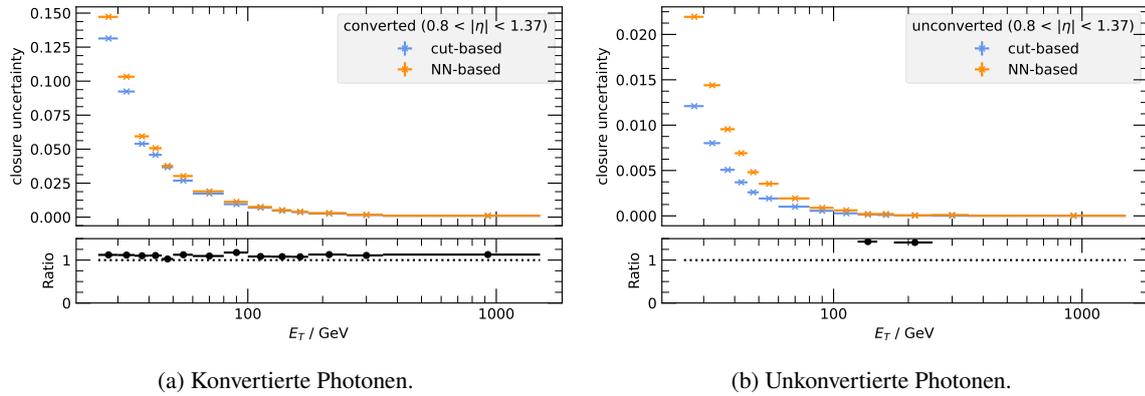


Abbildung 5.26: Closure Unsicherheit für $0,8 < |\eta| < 1,37$.

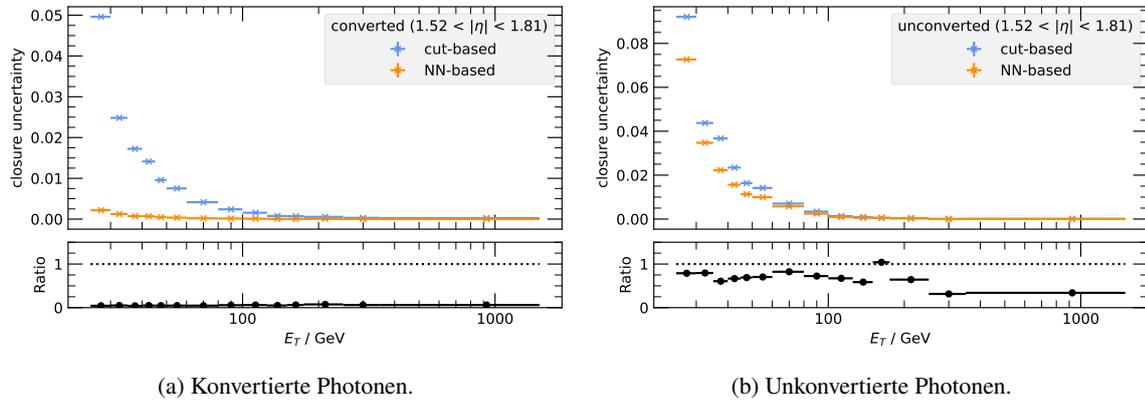


Abbildung 5.27: Closure Unsicherheit für $1,52 < |\eta| < 1,81$.

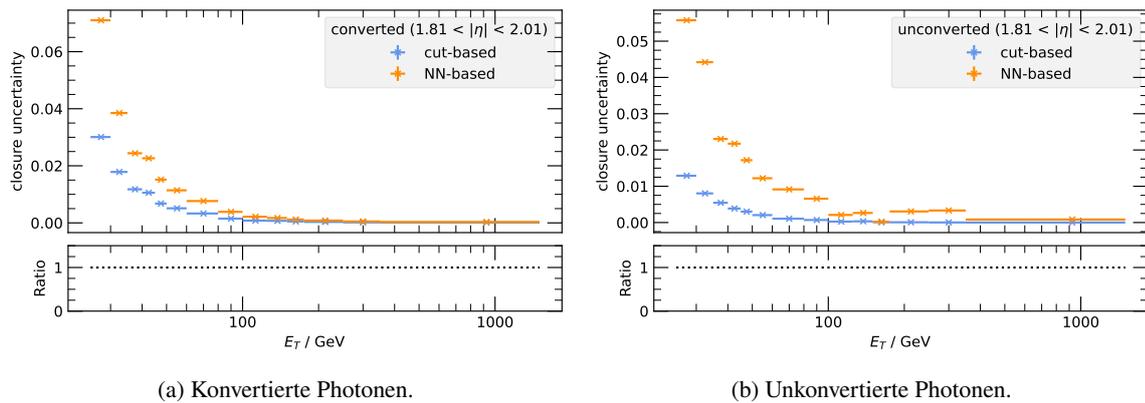
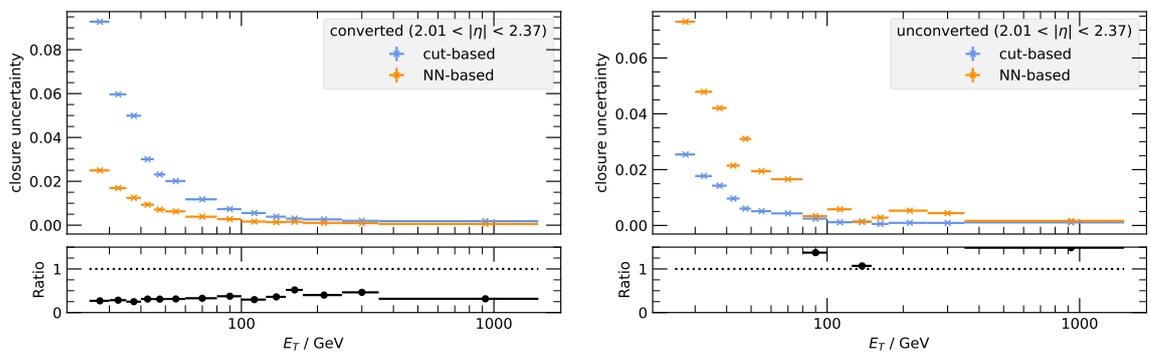


Abbildung 5.28: Closure Unsicherheit für $1,81 < |\eta| < 2,01$.



(a) Konvertierte Photonen.

(b) Unkonvertierte Photonen.

Abbildung 5.29: Closure Unsicherheit für $2,01 < |\eta| < 2,37$.

Resultate und Fazit

In dieser Arbeit konnte die Verwendung von Binärklassifikatornetzwerken in Kombination mit der Matrix Methode zur Photon ID in Messdaten erfolgreich implementiert und getestet werden. Zunächst wurden die Grundlagen aus der Beschleuniger- und Teilchenphysik, die zum Verständnis dieser Arbeit benötigt werden, knapp beschrieben. Zudem wurde ein Überblick über die Komponenten und Funktionsweise des LHC und des ATLAS-Detektors geben. Weitergehend wurde eine Einführung in die Photon ID via der Matrix Methode gegeben. Schließlich wurden die Schlüsselkonzepte aus dem Bereich Machine Learning, die in dieser Arbeit angewandt wurden, vorgestellt und erklärt.

Nach Einführung dieser grundlegenden Konzepte konnten vier Binärklassifikatornetzwerke auf die narrow-strip und relaxed-tight Variablen von converted und unconverted Photonen trainiert und zur Identifikation von prompt und fake Photonen verwendet werden. Zur Optimierung der Hyperparameter der Netze wurde im Vorfeld eine Gridsearch mit einer Bayes'schen Optimierung als Crosscheck durchgeführt. Es ist dabei hervorzuheben, dass die narrow-strip Binärklassifikatoren eine Genauigkeit nahe dem tight-ID Arbeitspunkt zur Unterscheidung von prompt und fake Photonen erreichen konnten. Hingegen war die Leistung der relaxed-tight Binärklassifikatoren fernab dem tight-ID Arbeitspunkt, sodass weiterführende Nachforschungen angestellt werden sollten, ob man zukünftig eine Methode zur Photon ID, welche nur auf den narrow-strip Variablen basiert, entwickeln könnte. In diesem Falle wäre es möglich komplett auf die Matrix Methode zu verzichten. Zudem wäre es auch notwendig, das Verhalten der neuronalen Netzwerke in jedem η - und p_T -Intervall zu testen und individuell zu optimieren.

Der Output der Binärklassifikatoren konnte verwendet werden, um die Photonen aus Messdaten in die Regionen der Matrix Methode einzuteilen. Die Cuts auf den Netzwerk Output wurden dabei im Vorfeld unter Rücksichtnahme der Signal Efficiency und Background Rejection optimiert.

Die Matrix Methode konnte schließlich sowohl über die traditionelle Cut-basierte Variante, als auch mit den Outputs der neuronalen Netzwerke ausgeführt werden. Es wurden in beiden Fälle die tight-ID Effizienz und closure Unsicherheit für jedes η - und p_T -Intervall berechnet. Das Hauptziel dieser Analyse, die Minimierung der closure Unsicherheit unter Beibehaltung der Signal Efficiency gegenüber der traditionellen Cut-basierten Variante, konnte in mehreren η -Intervallen erreicht werden. In diesen Intervallen ist eine klare Verbesserung der closure Unsicherheiten zu sehen. In anderen η -Intervallen konnte keine Verbesserung erzielt werden. Zum genauen Verständnis weshalb die Methode mit den neuronalen Netzwerken in anderen η -Intervallen keine Verringerung der closure Unsicherheit erzielen konnte, bedarf es weiterer Nachforschung. Zur weiteren Nutzung neuronaler

Netzwerke für die Reduzierung der closure Unsicherheit in der Photon ID via der Matrix Methode, könnte zudem der Ansatz verfolgt werden, Intervalle mit geringer Ereigniszahl zusammenzufügen, um so die Statistik für das Training der neuronalen Netzwerke zu erhöhen.

Leistung der Binärklassifikatoren

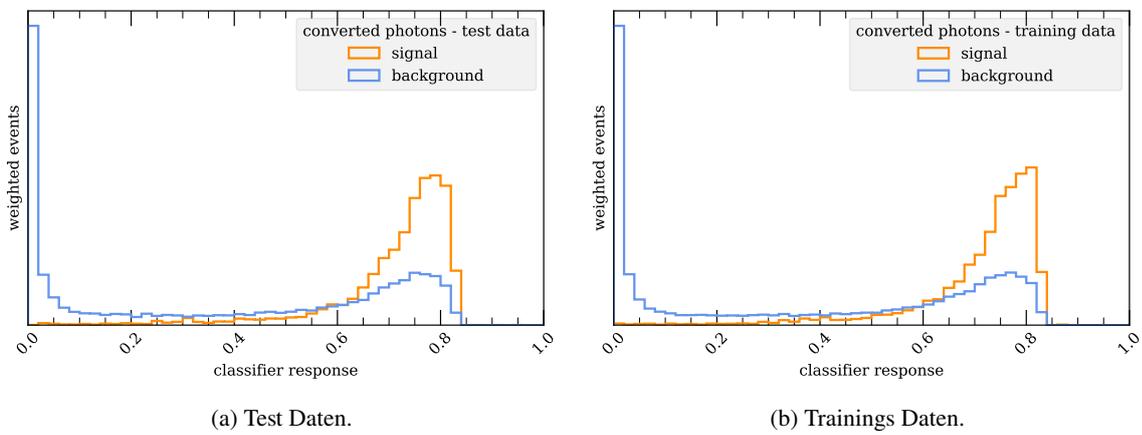


Abbildung A.1: Leistung des narrow-strip Binärklassifikators für konvertierte Photonen.

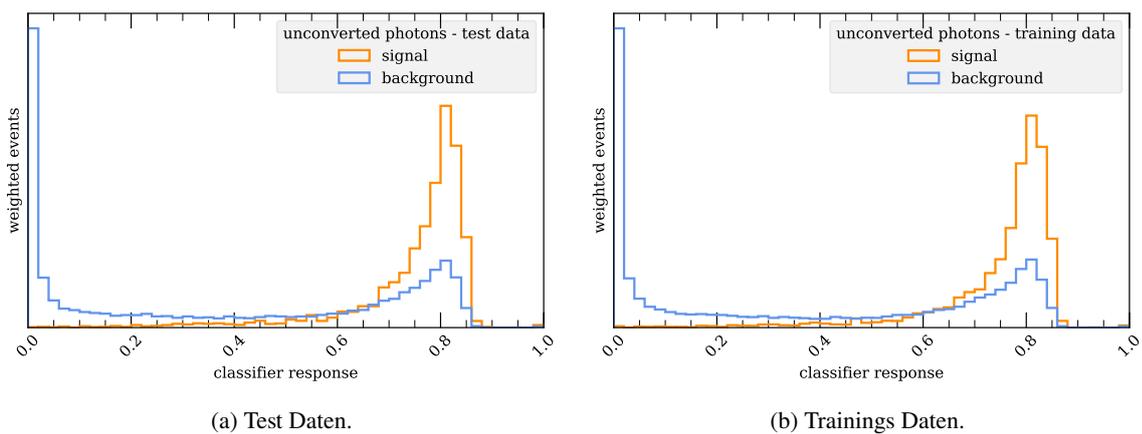


Abbildung A.2: Leistung des narrow-strip Binärklassifikators für unkonvertierte Photonen.

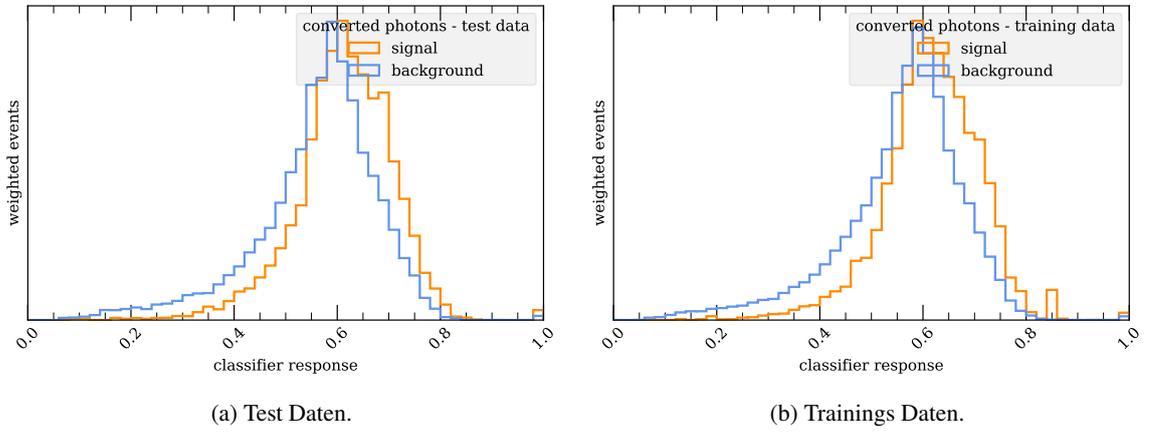


Abbildung A.3: Leistung des relaxed-tight Binärklassifikators für konvertierte Photonen.

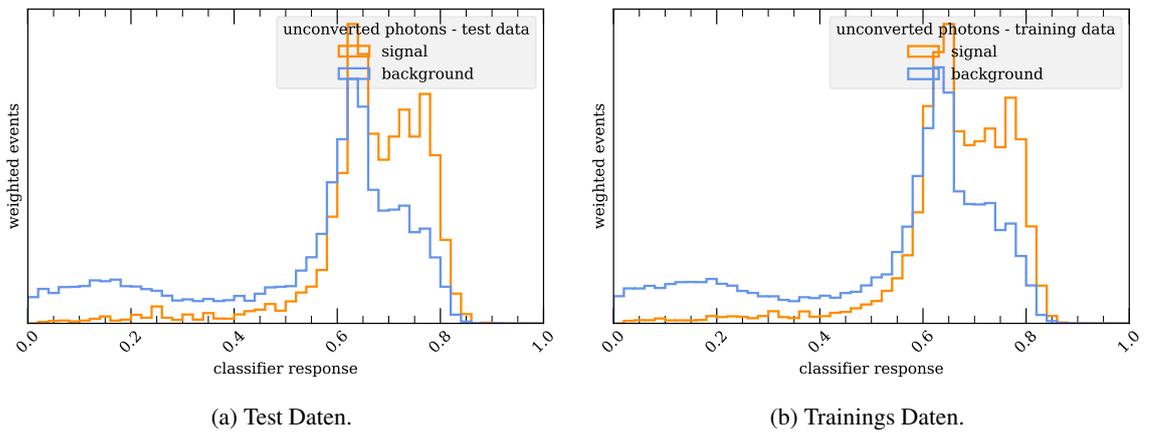


Abbildung A.4: Leistung des relaxed-tight Binärklassifikators für unkonvertierte Photonen.

Cuts auf die Outputs der neuronalen Netzwerke

Tabelle B.1: Cuts auf die narrow-strip Variablen bei $0,0 < |\eta| < 0,6$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,705824	1,000000e+30	7,378880e-01
30,0	35,0	1,000000e+30	0,713566	1,000000e+30	7,481916e-01
35,0	40,0	1,000000e+30	0,703960	1,000000e+30	7,423949e-01
40,0	45,0	1,000000e+30	0,712301	1,000000e+30	7,441564e-01
45,0	50,0	1,000000e+30	0,714964	1,000000e+30	7,368499e-01
50,0	60,0	1,000000e+30	0,715018	1,000000e+30	7,417951e-01
60,0	80,0	1,000000e+30	0,704458	1,000000e+30	7,389497e-01
80,0	100,0	1,000000e+30	0,723378	1,000000e+30	4,089187e-01
100,0	125,0	1,000000e+30	0,725300	1,000000e+30	6,206395e-01
125,0	150,0	1,000000e+30	0,712022	1,000000e+30	8,458570e-02
150,0	175,0	1,000000e+30	0,617028	1,000000e+30	1,455075e-01
175,0	250,0	1,000000e+30	0,733284	1,000000e+30	4,345459e-01
250,0	350,0	1,000000e+30	0,733505	1,000000e+30	2,716942e-01
350,0	1500,0	1,000000e+30	0,114155	1,000000e+30	-1,000000e+30

Tabelle B.2: Cuts auf die narrow-strip Variablen bei $0,6 < |\eta| < 0,8$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,656635	1,000000e+30	6,495857e-01
30,0	35,0	1,000000e+30	0,665101	1,000000e+30	6,431621e-01
35,0	40,0	1,000000e+30	0,618647	1,000000e+30	5,953100e-01
40,0	45,0	1,000000e+30	0,632242	1,000000e+30	6,572347e-01
45,0	50,0	1,000000e+30	0,652229	1,000000e+30	6,694878e-01
50,0	60,0	1,000000e+30	0,647917	1,000000e+30	6,499230e-01
60,0	80,0	1,000000e+30	0,602362	1,000000e+30	6,877277e-01
80,0	100,0	1,000000e+30	0,628002	1,000000e+30	6,200331e-01
100,0	125,0	1,000000e+30	0,629960	1,000000e+30	6,853327e-01
125,0	150,0	1,000000e+30	0,532839	1,000000e+30	5,449841e-01
150,0	175,0	1,000000e+30	0,617562	1,000000e+30	2,101077e-01
175,0	250,0	1,000000e+30	0,630065	1,000000e+30	6,711556e-01
250,0	350,0	1,000000e+30	0,511815	1,000000e+30	6,435463e-01
350,0	1500,0	1,000000e+30	0,526382	1,000000e+30	-1,000000e+30

Tabelle B.3: Cuts auf die narrow-strip Variablen bei $0,8 < |\eta| < 1,37$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,576968	1,000000e+30	0,562218
30,0	35,0	1,000000e+30	0,581122	1,000000e+30	0,545708
35,0	40,0	1,000000e+30	0,568291	1,000000e+30	0,524731
40,0	45,0	1,000000e+30	0,578419	1,000000e+30	0,570216
45,0	50,0	1,000000e+30	0,570769	1,000000e+30	0,526866
50,0	60,0	1,000000e+30	0,559571	1,000000e+30	0,543107
60,0	80,0	1,000000e+30	0,539174	1,000000e+30	0,543212
80,0	100,0	1,000000e+30	0,517256	1,000000e+30	0,396324
100,0	125,0	1,000000e+30	0,544932	1,000000e+30	0,603355
125,0	150,0	1,000000e+30	0,516543	1,000000e+30	0,350136
150,0	175,0	1,000000e+30	0,531778	1,000000e+30	0,563961
175,0	250,0	1,000000e+30	0,213244	1,000000e+30	0,479132
250,0	350,0	1,000000e+30	0,342369	1,000000e+30	0,605043
350,0	1500,0	1,000000e+30	0,497420	1,000000e+30	0,476955

Tabelle B.4: Cuts auf die narrow-strip Variablen bei $1,52 < |\eta| < 1,81$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,507232	1,000000e+30	0,435238
30,0	35,0	1,000000e+30	0,516854	1,000000e+30	0,492176
35,0	40,0	1,000000e+30	0,497432	1,000000e+30	0,492176
40,0	45,0	1,000000e+30	0,465259	1,000000e+30	0,481457
45,0	50,0	1,000000e+30	0,447056	1,000000e+30	0,368012
50,0	60,0	1,000000e+30	0,486567	1,000000e+30	0,352584
60,0	80,0	1,000000e+30	0,491940	1,000000e+30	0,272320
80,0	100,0	1,000000e+30	0,320169	1,000000e+30	0,358825
100,0	125,0	1,000000e+30	0,270225	1,000000e+30	0,212978
125,0	150,0	1,000000e+30	0,442254	1,000000e+30	0,316036
150,0	175,0	1,000000e+30	0,269652	1,000000e+30	0,314417
175,0	250,0	1,000000e+30	0,167594	1,000000e+30	0,504183
250,0	350,0	1,000000e+30	0,412046	1,000000e+30	0,546123
350,0	1500,0	1,000000e+30	0,412046	1,000000e+30	0,558479

Tabelle B.5: Cuts auf die narrow-strip Variablen bei $1,81 < |\eta| < 2,01$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,617388	1,000000e+30	0,699428
30,0	35,0	1,000000e+30	0,611395	1,000000e+30	0,668799
35,0	40,0	1,000000e+30	0,623681	1,000000e+30	0,647917
40,0	45,0	1,000000e+30	0,606807	1,000000e+30	0,667209
45,0	50,0	1,000000e+30	0,607150	1,000000e+30	0,462125
50,0	60,0	1,000000e+30	0,623148	1,000000e+30	0,110499
60,0	80,0	1,000000e+30	0,633262	1,000000e+30	0,147920
80,0	100,0	1,000000e+30	0,463238	1,000000e+30	0,107918
100,0	125,0	1,000000e+30	0,600922	1,000000e+30	0,186081
125,0	150,0	1,000000e+30	0,610818	1,000000e+30	0,180133
150,0	175,0	1,000000e+30	0,614060	1,000000e+30	0,591129
175,0	250,0	1,000000e+30	0,612135	1,000000e+30	0,071080
250,0	350,0	1,000000e+30	0,526382	1,000000e+30	0,009973
350,0	1500,0	1,000000e+30	0,526382	1,000000e+30	0,338022

Tabelle B.6: Cuts auf die narrow-strip Variablen bei $2,01 < |\eta| < 2,37$.

pt_lower	pt_upper	converted_ns_upper	converted_ns_lower	unconverted_ns_upper	unconverted_ns_lower
25,0	30,0	1,000000e+30	0,672748	1,000000e+30	7,385847e-01
30,0	35,0	1,000000e+30	0,684338	1,000000e+30	7,352831e-01
35,0	40,0	1,000000e+30	0,672223	1,000000e+30	6,625057e-01
40,0	45,0	1,000000e+30	0,678821	1,000000e+30	6,527749e-01
45,0	50,0	1,000000e+30	0,671516	1,000000e+30	1,762132e-01
50,0	60,0	1,000000e+30	0,694402	1,000000e+30	4,861687e-01
60,0	80,0	1,000000e+30	0,674057	1,000000e+30	3,575754e-01
80,0	100,0	1,000000e+30	0,579761	1,000000e+30	4,844399e-02
100,0	125,0	1,000000e+30	0,680279	1,000000e+30	1,881513e-01
125,0	150,0	1,000000e+30	0,627683	1,000000e+30	-1,000000e+30
150,0	175,0	1,000000e+30	0,444276	1,000000e+30	1,990297e-01
175,0	250,0	1,000000e+30	0,546935	1,000000e+30	3,266394e-01
250,0	350,0	1,000000e+30	0,526382	1,000000e+30	1,392753e-01
350,0	1500,0	1,000000e+30	0,665693	1,000000e+30	1,392753e-01

Tabelle B.7: Cuts auf die relaxed-tight Variablen bei $0,0 < |\eta| < 0,6$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	0,287204	1,000000e+30	0,444835
30,0	35,0	1,000000e+30	0,379868	1,000000e+30	0,326016
35,0	40,0	1,000000e+30	0,488843	1,000000e+30	0,502466
40,0	45,0	1,000000e+30	0,375986	1,000000e+30	0,505634
45,0	50,0	1,000000e+30	0,340974	1,000000e+30	0,370269
50,0	60,0	1,000000e+30	0,480430	1,000000e+30	0,539210
60,0	80,0	1,000000e+30	0,538170	1,000000e+30	0,302663
80,0	100,0	1,000000e+30	0,470855	1,000000e+30	0,682456
100,0	125,0	1,000000e+30	0,543618	1,000000e+30	0,673899
125,0	150,0	1,000000e+30	0,577413	1,000000e+30	0,721556
150,0	175,0	1,000000e+30	0,608216	1,000000e+30	0,729231
175,0	250,0	1,000000e+30	0,412570	1,000000e+30	0,733307
250,0	350,0	1,000000e+30	0,428996	1,000000e+30	0,755601
350,0	1500,0	1,000000e+30	0,687471	1,000000e+30	0,770501

Tabelle B.8: Cuts auf die relaxed-tight Variablen bei $0,6 < |\eta| < 0,8$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	0,443194	1,000000e+30	3,938472e-01
30,0	35,0	1,000000e+30	0,241961	1,000000e+30	5,013613e-01
35,0	40,0	1,000000e+30	0,499998	1,000000e+30	5,454213e-01
40,0	45,0	1,000000e+30	0,488122	1,000000e+30	5,323340e-01
45,0	50,0	1,000000e+30	0,356528	1,000000e+30	4,915924e-01
50,0	60,0	1,000000e+30	0,452501	1,000000e+30	5,589519e-01
60,0	80,0	1,000000e+30	0,533872	1,000000e+30	2,643851e-01
80,0	100,0	1,000000e+30	0,508869	1,000000e+30	6,141099e-01
100,0	125,0	1,000000e+30	0,500381	1,000000e+30	5,395164e-01
125,0	150,0	1,000000e+30	0,374227	1,000000e+30	6,480131e-01
150,0	175,0	1,000000e+30	0,566588	1,000000e+30	6,717123e-01
175,0	250,0	1,000000e+30	0,484263	1,000000e+30	4,668832e-01
250,0	350,0	1,000000e+30	0,526133	1,000000e+30	6,131937e-01
350,0	1500,0	1,000000e+30	0,521144	8,090735e-01	-1,000000e+30

Tabelle B.9: Cuts auf die relaxed-tight Variablen bei $0,8 < |\eta| < 1,37$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	0,337827	1,000000e+30	0,225540
30,0	35,0	1,000000e+30	0,248396	1,000000e+30	0,374192
35,0	40,0	1,000000e+30	0,351578	1,000000e+30	0,377840
40,0	45,0	1,000000e+30	0,241366	1,000000e+30	0,360996
45,0	50,0	1,000000e+30	0,317632	1,000000e+30	0,460000
50,0	60,0	1,000000e+30	0,314505	1,000000e+30	0,501657
60,0	80,0	1,000000e+30	0,502899	1,000000e+30	0,537449
80,0	100,0	1,000000e+30	0,539261	1,000000e+30	0,604751
100,0	125,0	1,000000e+30	0,525543	1,000000e+30	0,483900
125,0	150,0	1,000000e+30	0,548706	1,000000e+30	0,620008
150,0	175,0	1,000000e+30	0,387873	1,000000e+30	0,606030
175,0	250,0	1,000000e+30	0,583245	1,000000e+30	0,635210
250,0	350,0	1,000000e+30	0,599126	1,000000e+30	0,626396
350,0	1500,0	1,000000e+30	0,589418	1,000000e+30	0,660089

Tabelle B.10: Cuts auf die relaxed-tight Variablen bei $1,52 < |\eta| < 1,81$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	0,148421	1,000000e+30	0,442804
30,0	35,0	1,000000e+30	0,174313	1,000000e+30	0,408579
35,0	40,0	1,000000e+30	0,312821	1,000000e+30	0,358175
40,0	45,0	1,000000e+30	0,346110	1,000000e+30	0,249800
45,0	50,0	1,000000e+30	0,416409	1,000000e+30	0,546845
50,0	60,0	1,000000e+30	0,256205	1,000000e+30	0,560258
60,0	80,0	1,000000e+30	0,211077	1,000000e+30	0,586817
80,0	100,0	1,000000e+30	0,510974	1,000000e+30	0,591523
100,0	125,0	1,000000e+30	0,525647	1,000000e+30	0,606087
125,0	150,0	1,000000e+30	0,097160	1,000000e+30	0,613798
150,0	175,0	1,000000e+30	0,542235	1,000000e+30	0,619514
175,0	250,0	1,000000e+30	0,552942	1,000000e+30	0,033737
250,0	350,0	1,000000e+30	0,352535	1,000000e+30	0,305999
350,0	1500,0	1,000000e+30	0,359268	1,000000e+30	0,285304

Tabelle B.11: Cuts auf die relaxed-tight Variablen bei $1,81 < |\eta| < 2,01$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	0,122028	1,000000e+30	0,334397
30,0	35,0	1,000000e+30	0,314732	1,000000e+30	0,381957
35,0	40,0	1,000000e+30	0,293293	1,000000e+30	0,547489
40,0	45,0	1,000000e+30	0,444554	1,000000e+30	0,542852
45,0	50,0	1,000000e+30	0,421229	1,000000e+30	0,622864
50,0	60,0	1,000000e+30	0,392924	1,000000e+30	0,630774
60,0	80,0	1,000000e+30	0,370074	1,000000e+30	0,633316
80,0	100,0	1,000000e+30	0,565260	1,000000e+30	0,637542
100,0	125,0	1,000000e+30	0,535127	1,000000e+30	0,637362
125,0	150,0	1,000000e+30	0,489060	1,000000e+30	0,637896
150,0	175,0	1,000000e+30	0,549799	1,000000e+30	0,302977
175,0	250,0	1,000000e+30	0,544869	1,000000e+30	0,644680
250,0	350,0	1,000000e+30	0,616161	1,000000e+30	0,646704
350,0	1500,0	1,000000e+30	0,616161	1,000000e+30	0,644680

Tabelle B.12: Cuts auf die relaxed-tight Variablen bei $2,01 < |\eta| < 2,37$.

pt_lower	pt_upper	converted_rt_upper	converted_rt_lower	unconverted_rt_upper	unconverted_rt_lower
25,0	30,0	1,000000e+30	2,378410e-01	1,000000e+30	2,971774e-01
30,0	35,0	1,000000e+30	2,440015e-01	1,000000e+30	3,233045e-01
35,0	40,0	1,000000e+30	3,305761e-01	1,000000e+30	4,859628e-01
40,0	45,0	1,000000e+30	3,505841e-01	1,000000e+30	5,073753e-01
45,0	50,0	1,000000e+30	2,869629e-01	1,000000e+30	5,705416e-01
50,0	60,0	1,000000e+30	3,949315e-01	1,000000e+30	5,593785e-01
60,0	80,0	1,000000e+30	4,390844e-01	1,000000e+30	5,819564e-01
80,0	100,0	1,000000e+30	4,897655e-01	1,000000e+30	5,952132e-01
100,0	125,0	1,000000e+30	4,189441e-01	1,000000e+30	5,981401e-01
125,0	150,0	1,000000e+30	5,067831e-01	6,522164e-01	-1,000000e+30
150,0	175,0	1,000000e+30	5,788593e-01	1,000000e+30	5,976924e-01
175,0	250,0	1,000000e+30	5,482632e-01	1,000000e+30	6,074823e-01
250,0	350,0	1,000000e+30	5,969852e-01	1,000000e+30	6,116361e-01
350,0	1500,0	9,213611e-01	-1,000000e+30	1,000000e+30	6,116361e-01

Ereigniszahl in den Messdaten

Tabelle C.1: Anzahl der Ereignisse in Daten 2015.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	60238	20945	55242	25832	13256	23584	199097
$30 \leq p_T < 35$	67106	23232	60627	27671	15106	27264	221006
$35 \leq p_T < 40$	33871	11962	30827	14548	7925	14194	113327
$40 \leq p_T < 45$	49326	17162	44247	20838	11446	20959	163978
$45 \leq p_T < 50$	48974	16931	43302	19872	11297	20841	161217
$50 \leq p_T < 60$	67989	23516	60951	28586	15844	29694	226580
$60 \leq p_T < 80$	120536	41270	108102	50182	28910	54452	403452
$80 \leq p_T < 100$	121639	41668	110922	52387	30076	57029	413721
$100 \leq p_T < 125$	104343	35733	95578	45990	26555	50035	358234
$125 \leq p_T < 150$	283585	97436	261964	122844	73250	137267	976346
$150 \leq p_T < 175$	128639	43163	120662	60145	33695	61869	448173
$175 \leq p_T < 250$	119230	40835	112922	56933	31498	56990	418408
$250 \leq p_T < 350$	28549	9803	27838	13280	7128	12132	98730
$350 \leq p_T < 1500$	9258	3163	8748	3783	1765	2393	29110
Total p_T	1243283	426819	1141932	542891	307751	568703	4231379

Tabelle C.2: Anzahl der Ereignisse in Daten 2016.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	157469	54904	145349	65735	33465	60554	517476
$30 \leq p_T < 35$	176661	61196	160340	72696	37379	69433	577705
$35 \leq p_T < 40$	89282	31249	80893	37534	19539	36353	294850
$40 \leq p_T < 45$	134283	46215	120721	52984	29119	55870	439192
$45 \leq p_T < 50$	144685	49637	129215	57615	31303	60062	472517
$50 \leq p_T < 60$	249946	85868	224873	101545	55372	106635	824239
$60 \leq p_T < 80$	384862	131097	344775	155198	86929	169111	1271972
$80 \leq p_T < 100$	340066	115914	307934	141282	79645	155831	1140672
$100 \leq p_T < 125$	286590	97364	261091	119868	68997	134035	967945
$125 \leq p_T < 150$	716245	242216	652334	277778	176417	336544	2401534
$150 \leq p_T < 175$	1336013	455182	1236715	572744	332027	628754	4561435
$175 \leq p_T < 250$	1246311	425782	1171800	563670	311738	576316	4295617
$250 \leq p_T < 350$	299543	103277	285638	133110	72221	121308	1015097
$350 \leq p_T < 1500$	98667	33547	92771	38707	18099	24584	306375
Total p_T	5660623	1933448	5214449	2390466	1352250	2535390	19086626

Tabelle C.3: Anzahl der Ereignisse in Daten 2017.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	139314	48516	129692	62902	30150	51136	461710
$30 \leq p_T < 35$	161792	56098	147921	71127	35369	61111	533418
$35 \leq p_T < 40$	82255	28481	74938	36871	18424	31840	272809
$40 \leq p_T < 45$	118658	41036	107978	51524	26823	47485	393504
$45 \leq p_T < 50$	118188	40451	107027	51051	26586	47910	391213
$50 \leq p_T < 60$	164095	56580	150285	72013	37931	68338	549242
$60 \leq p_T < 80$	208800	71355	190111	90734	49166	89173	699339
$80 \leq p_T < 100$	169962	57762	156736	76138	41540	76261	578399
$100 \leq p_T < 125$	142392	48372	131423	63661	35836	64722	486406
$125 \leq p_T < 150$	645506	216220	591796	250286	163241	294932	2161981
$150 \leq p_T < 175$	1823516	616583	1707366	825942	466986	830305	6270698
$175 \leq p_T < 250$	1702905	577485	1619226	822646	439883	759767	5921912
$250 \leq p_T < 350$	412233	139332	397599	195675	101135	162378	1408352
$350 \leq p_T < 1500$	135112	45822	128751	56436	25477	32654	424252
Total p_T	6024728	2044093	5640849	2727006	1498547	2618012	20553235

Tabelle C.4: Anzahl der Ereignisse in Daten 2018.

p_T -Intervalle	$0,0 < \eta < 0,6$	$0,6 < \eta < 0,8$	$0,8 < \eta < 1,37$	$1,52 < \eta < 1,81$	$1,81 < \eta < 2,01$	$2,01 < \eta < 2,37$	Total η
$25 \leq p_T < 30$	158250	55221	147557	68901	34174	57936	522039
$30 \leq p_T < 35$	177853	61342	162976	76627	38391	66796	583985
$35 \leq p_T < 40$	90528	31353	82621	39449	20011	34974	298936
$40 \leq p_T < 45$	136718	47338	124989	58263	30489	54086	451883
$45 \leq p_T < 50$	154541	53184	141635	65358	34802	61978	511498
$50 \leq p_T < 60$	306432	105415	280370	131168	70312	125820	1019517
$60 \leq p_T < 80$	354642	121701	322245	154186	83225	150855	1186854
$80 \leq p_T < 100$	221241	75114	203164	97484	53852	97340	748195
$100 \leq p_T < 125$	167903	57169	155907	75463	41868	75506	573816
$125 \leq p_T < 150$	828576	276929	761081	319293	208104	373604	2767587
$150 \leq p_T < 175$	2391464	808809	2245741	1078608	611657	1077530	8213809
$175 \leq p_T < 250$	2236499	757584	2128237	1069334	574706	985908	7752268
$250 \leq p_T < 350$	541148	183294	522123	255580	132257	209810	1844212
$350 \leq p_T < 1500$	177908	60345	169371	73074	33197	42696	556591
Total p_T	7943703	2694798	7448017	3562788	1967045	3414839	27031190

Literatur

- [1] B. Povh, K. Rith, C. Scholz, F. Zetsche und W. Rodejohann, *Teilchen und Kerne, Eine Einführung in die physikalischen Konzepte*, 9. Aufl., Springer Spektrum, 2014 (siehe S. 3).
- [2] M. Thomson, *Modern Particle Physics*, 1. Aufl., Cambridge University Press, 2013 (siehe S. 3).
- [3] Wikimedia Commons, *File:Standard Model of Elementary Particles.svg*,
URL: https://en.m.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (besucht am 15. 12. 2023) (siehe S. 4).
- [4] P. D. Group u. a., *Review of Particle Physics*,
Progress of Theoretical and Experimental Physics **2022** (2022) 083C01, ISSN: 2050-3911,
eprint: <https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf>,
URL: <https://doi.org/10.1093/ptep/ptac097> (siehe S. 4).
- [5] Particle Data Group, *61. Top Quark*,
URL: <https://pdg.lbl.gov/2021/reviews/rpp2021-rev-top-quark.pdf> (besucht am 13. 11. 2023) (siehe S. 3).
- [6] KATRIN Collaboration, *Direct neutrino-mass measurement with sub-electronvolt sensitivity*,
Nature Physics **18** (2022) 6 (siehe S. 5).
- [7] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, *Physics Letters B* **716** (2012) 29,
arXiv: [1207.7214v2](https://arxiv.org/abs/1207.7214v2) (siehe S. 5).
- [8] Brüning, Oliver Sim and Collier, Paul and Lebrun, P and Myers, Stephen and Ostojic, Ranko and Poole, John and Proudlock, Paul, *LHC Design Report*, (2004) 548,
URL: <https://cds.cern.ch/record/782076> (besucht am 23. 11. 2023) (siehe S. 7).
- [9] Lanni, Francesco and Pontecorvo, Ludovico, *Technical Design Report: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade*, 2020,
URL: <https://cds.cern.ch/record/2719855/> (besucht am 23. 11. 2023) (siehe S. 7).
- [10] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*,
URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Publications_and_Conference_Resu (besucht am 23. 11. 2023) (siehe S. 7).

- [11] CERN, *Cryogenics: Low temperatures, high performance*,
URL: <https://www.home.cern/science/engineering/cryogenics-low-temperatures-high-performance> (besucht am 16. 12. 2023) (siehe S. 7).
- [12] Mobs, Esma,
The CERN accelerator complex in 2019 Complexe des accélérateurs du CERN en 2019,
URL: <https://cds.cern.ch/record/2684277> (besucht am 23. 11. 2023) (siehe S. 8).
- [13] Pequenaio, Joao, *Computer generated image of the whole ATLAS detector*,
URL: <https://cds.cern.ch/record/1095924> (besucht am 23. 11. 2023) (siehe S. 9).
- [14] ATLAS Collaboration, *The inner Detector*, URL:
<https://atlas.cern/Discover/Detector/Inner-Detector> (besucht am 24. 11. 2023)
(siehe S. 10).
- [15] Pequenaio, Joao, *Computer generated image of the ATLAS inner detector*,
URL: <https://cds.cern.ch/record/1095926> (besucht am 23. 11. 2023) (siehe S. 10).
- [16] B. Abbott u. a., *Production and integration of the ATLAS Insertable B-Layer*,
Journal of Instrumentation **13** (2018) T05008, ISSN: 1748-0221,
URL: <http://dx.doi.org/10.1088/1748-0221/13/05/T05008> (siehe S. 10).
- [17] ATLAS Collaboration, *Calorimeter*, URL:
<https://atlas.cern/Discover/Detector/Calorimeter> (besucht am 24. 11. 2023)
(siehe S. 11).
- [18] Pequenaio, Joao, *Computer Generated image of the ATLAS calorimeter*,
URL: <https://cds.cern.ch/record/1095927> (besucht am 23. 11. 2023) (siehe S. 12).
- [19] ATLAS Collaboration,
Performance of the muon spectrometer alignment in 2017 and 2018 data, 2021,
URL: <https://cds.cern.ch/record/2753329/files/?ln=de> (besucht am 24. 11. 2023)
(siehe S. 12).
- [20] ATLAS Collaboration, *Muon Spectrometer*,
URL: <https://atlas.cern/Discover/Detector/Muon-Spectrometer> (besucht am
24. 11. 2023) (siehe S. 12).
- [21] ATLAS Collaboration, *Trigger and Data Acquisition*, URL:
<https://atlas.cern/Discover/Detector/Trigger-DAQ> (besucht am 24. 11. 2023)
(siehe S. 12).
- [22] zur Nedde, Martin, *The Run-2 ATLAS Trigger System: Design, Performance and Plan*, 2016,
URL: <https://cds.cern.ch/record/2238679/?ln=de> (besucht am 24. 11. 2023)
(siehe S. 13).
- [23] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, *Eur. Phys. J. C* **79** (2019) 55,
arXiv: [1810.05087](https://arxiv.org/abs/1810.05087) (siehe S. 15–17).
- [24] ATLAS Collaboration,
Topological cell clustering in the ATLAS calorimeters and its performance in Run 1,
Eur. Phys. J. C **77** (2017) 100, arXiv: [1603.02934](https://arxiv.org/abs/1603.02934) (siehe S. 15).

-
- [25] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, *The European Physical Journal C* **76** (2016) 42 (siehe S. 18).
- [26] draw.io AG, *draw.io*, URL: <https://www.drawio.com/> (besucht am 17. 12. 2023) (siehe S. 19, 22).
- [27] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *Journal of instrumentation* **14** (2019) 70, arXiv: 1908.00005 (siehe S. 20).
- [28] Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron, *Deep Learning, Das umfassende Handbuch : Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze*, 1. Aufl., MITP-Verlags GmbH & Co. KG, 2018 (siehe S. 21).
- [29] Chollet, François, *Deep Learning mit Python und Keras : Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*, 1. Aufl., MITP-Verlags GmbH & Co. KG, 2018 (siehe S. 21).
- [30] B. Polyak, *Some methods of speeding up the convergence of iteration methods*, *Ussr Computational Mathematics and Mathematical Physics* **4** (1964) 1 (siehe S. 23).
- [31] Kingma, Diedrik P. and Ba, Jimmy Lei, *Adam: A method for stochastic optimization*, (2014), arXiv: 1412.6980 (siehe S. 24).
- [32] Karim, Raimi, *10 Stochastic Gradient Descent Optimisation Algorithms + Cheatsheet*, URL: <https://towardsdatascience.com/10-gradient-descent-optimisation-algorithms-86989510b5e9> (siehe S. 24).
- [33] Kirfel, Florian, *Application of neural networks for photon identification using the ATLAS detector*, 2021, URL: <https://web.physik.uni-bonn.de/group/view.php?&group=3&lang=en&c=t&id=46> (siehe S. 25, 27, 29).
- [34] Srivastava, Nitish and Hinton, Geoffrey and Krizhevsky, Alex and Sutskever, Ilya and Salakhutdinov, Ruslan, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *Journal of Machine Learning Research* **15** (2014) 30, URL: <https://jmlr.org/papers/v15/srivastava14a.html> (siehe S. 25).
- [35] Ioffe, Sergey and Szegedy, Christian, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, (2015) 11, arXiv: 1502.03167 (siehe S. 25).
- [36] Google Brain Team, *TensorFlow*, URL: <https://www.tensorflow.org/> (besucht am 10. 12. 2023) (siehe S. 27).
- [37] Chollet, François, *Keras: Deep Learning for humans*, URL: <https://keras.io/> (besucht am 10. 12. 2023) (siehe S. 27).
- [38] T. Sjöstrand, *The Pythia event generator: Past, present and future*, *Computer Physics Communications* **246** (2020) 106910, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2019.106910> (siehe S. 27).
- [39] S. Agostinelli u. a., *GEANT4-a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment* **506** (2003) 250 (siehe S. 27).

- [40] Koehrsen, Will,
A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning,
URL: <https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f> (besucht am 12. 12. 2023) (siehe S. 30).
- [41] A. Hoecker u. a., *TMVA - Toolkit for Multivariate Data Analysis*, 2009,
arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) [[physics.data-an](https://arxiv.org/archive/physics)] (siehe S. 33).
- [42] B. Wendland, J. Erdmann und K. Koreniger,
Measurement of the photon identification efficiency with the matrix method using 139 fb-1 of proton-proton collision data at $\sqrt{s} = 13\text{TeV}$ collected by the ATLAS detector,
URL: <https://cds.cern.ch/record/2712917/files/ATL-COM-PHYS-2020-212.pdf>
(besucht am 19. 12. 2023) (siehe S. 39).

Abbildungsverzeichnis

2.1	Standardmodell der Teilchenphysik.	4
2.2	CERN Beschleunigerkomplex.	8
2.3	ATLAS Detektorsystem.	9
2.4	ATLAS innerer Detektor.	10
2.5	ATLAS Kalorimetersystem.	12
3.1	Regionen der Matrix Methode.	19
4.1	Binärklassifikator Struktur.	22
4.2	ROC-Kurve mit sehr guter Klassifikatorleistung.	25
5.1	Transformation der ROC-Kurve.	29
5.2	ROC-Kurven narrow-strip Netzwerke.	31
5.3	Loss Funktion narrow-strip Netzwerke.	31
5.4	ROC-Kurven relaxed-tight Netzwerke.	32
5.5	Loss-Kurven relaxed-tight Netzwerke.	32
5.6	s_E bei $0,0 < \eta < 0,6$	34
5.7	s_E bei $0,6 < \eta < 0,8$	34
5.8	s_E bei $0,8 < \eta < 1,37$	34
5.9	s_E bei $1,52 < \eta < 1,81$	35
5.10	s_E bei $1,81 < \eta < 2,01$	35
5.11	s_E bei $2,01 < \eta < 2,37$	35
5.12	b_E bei $0,0 < \eta < 0,6$	36
5.13	b_E bei $0,6 < \eta < 0,8$	36
5.14	b_E bei $0,8 < \eta < 1,37$	37
5.15	b_E bei $1,52 < \eta < 1,81$	37
5.16	b_E bei $1,81 < \eta < 2,01$	37
5.17	b_E bei $2,01 < \eta < 2,37$	38
5.18	$\varepsilon^{\text{tight-ID}}$ bei $0,0 < \eta < 0,6$	40
5.19	$\varepsilon^{\text{tight-ID}}$ bei $0,6 < \eta < 0,8$	40
5.20	$\varepsilon^{\text{tight-ID}}$ bei $0,8 < \eta < 1,37$	40
5.21	$\varepsilon^{\text{tight-ID}}$ bei $1,52 < \eta < 1,81$	41
5.22	$\varepsilon^{\text{tight-ID}}$ bei $1,81 < \eta < 2,01$	41
5.23	$\varepsilon^{\text{tight-ID}}$ bei $2,01 < \eta < 2,37$	41
5.24	Closure Unsicherheit bei $0,0 < \eta < 0,6$	42

5.25	Closure Unsicherheit bei $0,6 < \eta < 0,8$.	42
5.26	Closure Unsicherheit bei $0,8 < \eta < 1,37$.	43
5.27	Closure Unsicherheit bei $1,52 < \eta < 1,81$.	43
5.28	Closure Unsicherheit bei $1,81 < \eta < 2,01$.	43
5.29	Closure Unsicherheit bei $2,01 < \eta < 2,37$.	44
A.1	NS-Klassifikator auf konvertierten Photonen.	47
A.2	NS-Klassifikator auf unkonvertierten Photonen.	47
A.3	RT-Klassifikator auf konvertierten Photonen.	48
A.4	RT-Klassifikator auf unkonvertierten Photonen.	48

Tabellenverzeichnis

3.1	Shower-shape Variablen.	20
4.1	Aktivierungsfunktionen.	22
5.1	Untergund Ereigniszahl.	28
5.2	Signal Ereigniszahl.	28
5.3	Hyperparameter Gridsearch.	30
5.4	Hyperparameter narrow-strip Netzwerk.	30
5.5	Hyperparameter relaxed-tight Netzwerk.	32
B.1	Narrow-Strip Klassifikator Cuts $0,0 < \eta < 0,6$	49
B.2	Narrow-Strip Klassifikator Cuts $0,6 < \eta < 0,8$	50
B.3	Narrow-Strip Klassifikator Cuts $0,8 < \eta < 1,37$	50
B.4	Narrow-Strip Klassifikator Cuts $1,52 < \eta < 1,81$	51
B.5	Narrow-Strip Klassifikator Cuts $1,81 < \eta < 2,01$	51
B.6	Narrow-Strip Klassifikator Cuts $2,01 < \eta < 2,37$	52
B.7	Relaxed-Tight Klassifikator Cuts $0,0 < \eta < 0,6$	52
B.8	Relaxed-Tight Klassifikator Cuts $0,6 < \eta < 0,8$	53
B.9	Relaxed-Tight Klassifikator Cuts $0,8 < \eta < 1,37$	53
B.10	Relaxed-Tight Klassifikator Cuts $1,52 < \eta < 1,81$	54
B.11	Relaxed-Tight Klassifikator Cuts $1,81 < \eta < 2,01$	54
B.12	Relaxed-Tight Klassifikator Cuts $2,01 < \eta < 2,37$	55
C.1	Anzahl der Ereignisse in Daten 2015.	57
C.2	Anzahl der Ereignisse in Daten 2016.	57
C.3	Anzahl der Ereignisse in Daten 2017.	58
C.4	Anzahl der Ereignisse in Daten 2018.	58