

# Studien zu neuronalen Netzen im Rahmen einer leptonischen Mono-top-Analyse

Studies on neural networks in the context  
of a leptonic mono-top analysis

Bachelorarbeit

Alina Klingel

An der Fakultät für Physik  
Institut für Experimentelle Teilchenphysik

Erstgutachter:	Prof. Dr. Ulrich Husemann
Zweitgutachter:	Dr. Michael Waßmer
Betreuender Mitarbeiter:	Sebastian Wieland

Karlsruhe, 05. August 2021



---

Diese Arbeit wurde vom Erstgutachter der Bachelorarbeit akzeptiert.

**Karlsruhe, den 05. August 2021**

.....  
(Prof. Dr. Ulrich Husemann)



---

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

**Karlsruhe, den 05. August 2021**

.....  
(Alina Klingel)



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Physikalische Grundlagen</b>	<b>3</b>
2.1	Das Standardmodell . . . . .	3
2.2	Dunkle Materie . . . . .	5
2.3	Das CMS-Experiment . . . . .	5
<b>3</b>	<b>Die Mono-top-Analyse</b>	<b>11</b>
3.1	Der Mono-top-Prozess . . . . .	11
3.2	Die Untergrundprozesse . . . . .	13
3.3	Klassische Ereignisselektion . . . . .	14
3.4	Ereignisselektion mit einem tiefen neuronalen Netz . . . . .	15
<b>4</b>	<b>Methodik</b>	<b>21</b>
4.1	Verwendete Daten und Netzwerkkonfiguration . . . . .	21
4.2	Auswertungsmethoden . . . . .	25
<b>5</b>	<b>Multiklassifikation von Signal und Untergründen mit neuronalen Netzen</b>	<b>29</b>
5.1	Die Eingabevariablen . . . . .	29
5.2	Ergebnisse des Trainings . . . . .	30
<b>6</b>	<b>Klassifizierung verschiedener Monotop-Signalhypothesen</b>	<b>35</b>
6.1	Klassifikation von Signalhypothesen im Bereich hoher Mediatormassen . . .	35
6.2	Weitere Klassifizierungsansätze . . . . .	37
6.3	Untersuchung des Verhaltens bei zu geringer Trennkraft . . . . .	39
6.4	Beobachtungen zur Verteilung der Anzahl der Jets . . . . .	40
6.5	Schlussfolgerungen . . . . .	43
<b>7</b>	<b>Vergleich der klassischen Methode mit der Klassifizierung durch neuronale Netze</b>	<b>45</b>
7.1	Training mit der transversalen Masse $m_T$ . . . . .	45
7.2	Diskriminatoren . . . . .	47
7.3	Reinheiten . . . . .	50
7.4	ROC-Kurven . . . . .	51
<b>8</b>	<b>Zusammenfassung</b>	<b>53</b>
	<b>Literatur</b>	<b>55</b>
	<b>Anhang</b>	<b>59</b>
A	Abkürzungen . . . . .	59
B	Wichtige Variablen . . . . .	59

C	Ergänzende Materialien zur Multiklassifikation von Signal und Untergründen	60
D	Materialien zur Klassifizierung verschiedener Mono-top-Signalthypothesen .	63
E	Materialien zur Untersuchung des Verhaltens bei geringer Trennkraft . . . .	71
F	Verteilung der fehlenden Transversalenergie gegen die des Abstandes zwischen Lepton und Myon aufgetragen . . . . .	73



# 1 Einleitung

Weltweit forschen Physiker an den elementaren Bausteinen unseres Universums und deren Wechselwirkungen mit dem Ziel zu verstehen, aus was unser Universum besteht und wie es funktioniert. Die Grundlagen dieser Bestrebungen sind in einer physikalischen Theorie zusammengefasst, dem Standardmodell der Teilchenphysik. Im Laufe der letzten Jahrzehnte konnte diese Theorie mit sich weiter steigernder Genauigkeit durch eine Vielzahl von Experimenten bestätigt werden. Gleichzeitig zeigt sich durch mehr und mehr Beobachtungen, dass diese Theorie keinesfalls vollständig ist. Daher ist es ein Bestreben der Teilchenphysik das Standardmodell zu erweitern, um neue Phänomene im Einklang mit den bisherigen Erkenntnissen erklären zu können.

Eine der Unzulänglichkeiten des Standardmodells ist das Fehlen eines Teilchens, welches die Dunkle Materie beschreiben könnte. Zahlreiche Beobachtungen der Kosmologie und Astrophysik deuten auf die Existenz dieser unsichtbaren Form der Materie hin, welche nur schwach mit der bekannten, für uns sichtbaren Materie wechselwirkt. Ein Beispiel dafür ist die Rotationsgeschwindigkeit von Galaxien, welche beruhend auf theoretischen Berechnungen, welche nur die Gravitation der sichtbaren Masse berücksichtigen, zu hoch ist[1]. Es kann angenommen werden, dass 25 % der im Universum enthaltenen Energie der Dunklen Materie zugeordnet werden muss.

Bisher waren Bestreben Dunkle Materie, die vermutlich mit Ausnahme der Gravitation kaum mit den bekannten Teilchen des Standardmodells wechselwirkt, experimentell nachzuweisen nicht von Erfolg gekrönt. Ein möglicher Ansatz Hinweise auf Physik jenseits des Standardmodells zu finden sind hochenergetische Streuexperimente, bei denen im Idealfall Dunkle-Materie-Teilchen unter Laborbedingungen erzeugt und nachgewiesen werden können. Dazu werden am Large Hadron Collider (LHC) am CERN die Produkte von Proton-Proton-Kollisionen mit Hilfe verschiedener Detektorsysteme, wie dem Compact Muon Solenoid (CMS)-Experiment, untersucht.

Diese Arbeit beschäftigt sich dabei mit einem speziellen Prozesses, bei dem neben Teilchen des Standardmodells auch Dunkle-Materie-Teilchen entstehen. Während erstere mit Hilfe des CMS-Experiment direkt detektiert werden können, wird gehofft letztere indirekt nachweisen zu können. Dazu wird nach der Produktion von Dunkle-Materie-Teilchen in Assoziation eines einzelnen Top-Quarks gesucht. Im Endzustand dieses leptonischen Monotop-Prozesses liegen dann neben einem Top-Quark auch Dunkle-Materie-Teilchen vor. Da letztere wie bereits erwähnt kaum mit der Materie des Standardmodells, und damit auch

dem Detektor, wechselwirken, besteht die Mono-top-Signatur neben den Zerfallsprodukten des Top-Quarks, die Teil des Standardmodells sind, aus fehlender Transversalenergie. Es gibt jedoch auch Teilchen des Standardmodells, wie die Neutrinos, welche den Detektor ebenfalls undetektiert verlassen und sich in der Signatur ihrer jeweiligen Prozesse nur durch fehlende Transversalenergie zeigen. Daher gibt es reine Standardmodell-Prozesse, deren Signaturen der des leptonischen Mono-top-Prozesses ähneln. Da diese Prozesse oft einen deutlich höheren Wirkungsquerschnitt aufweisen als der Mono-top-Prozess, muss für einen erfolgreichen Nachweis der Dunklen-Materie-Teilchen eine erfolgreiche Klassifizierung der Signal- und Untergrundprozesse sichergestellt werden.

Klassischerweise basiert die Trennung der Mono-top-Signalthypothesen von den dominanten Standardmodell-Untergrundprozessen auf verschiedenen Differenzierungskriterien der kinematischen Variablen.

Das Ziel dieser Arbeit ist zu überprüfen, ob diese Klassifizierung mit Hilfe eines neuronalen Netzes Potential dazu hat, Signal- und Untergrundprozesse effizient zu trennen. Dazu wird die Eigenschaft neuronaler Netze genutzt, Muster in Daten zu erkennen und dann eigenständig nach bestimmten vorgegebenen Kategorien anzureichern. Dies verspricht einen geringeren Verlust von Untergrundinformationen, was mit einer erhöhten Sensitivität bei der Suche nach dem Signal-Prozess einhergeht. Diese Studie wird auf simulierten Kollisionsereignissen durchgeführt.

Im Folgenden liefert Kapitel 2 einen kurzen Überblick über die physikalischen Grundlagen, also dem Standardmodell der Teilchenphysik, dem Konzept der Dunklen Materie und dessen Detektion, sowie dem LHC und dem dort ansässigen CMS-Experiment. In Kapitel 3 werden die in dieser Arbeit betrachteten Signal- und Untergrundprozesse kurz vorgestellt und es wird weiterhin auf die Klassifizierungsmethoden eingegangen, die Gegenstand dieser Arbeit sind. Im Zuge dessen wird der grundlegende Aufbau tiefer neuronaler Netze beschrieben. Im Anschluss werden in Kapitel 4 die verwendeten Datensätze, die Parameter der neuronalen Netze und die zur Auswertung und Bewertung der Klassifikation verwendeten Methoden vorgestellt.

Im Hauptteil der Arbeit wird in Kapitel 5 zunächst mit einem einfachen Ansatz überprüft, ob ein Netz mit den verwendeten Parametern und Variablen in der Lage ist, eine Klassifizierung erfolgreich durchzuführen. Anschließend wird in Kapitel 6 untersucht, wie sehr sich das Netz auf die Eigenschaften einer bestimmten Signalthypothese fixiert. Dazu wird das Training mit Mono-top-Datensätzen durchgeführt, die in verschiedenen Massenbereichen liegen, welche freie Parameter des untersuchten Modells sind. Um entsprechend dem Ziel dieser Arbeit eine Aussage zum Potential der Mono-top-Klassifikation mit neuronalen Netzen treffen zu können, erfolgt in Kapitel 7 ein quantitativer Vergleich der Leistungsfähigkeit der herkömmlichen Klassifizierungsmethode mit der Klassifikation durch neuronale Netze. Abschließend werden in Kapitel 8 die Ergebnisse und Erkenntnisse dieser Arbeit zusammengefasst.

## 2 Physikalische Grundlagen

In diesem Kapitel werden die physikalischen Grundlagen besprochen, die in dieser Arbeit benötigt werden. Darüber hinaus gibt das Kapitel einen Überblick über den Versuchsaufbau des Experiments, das Gegenstand dieser Arbeit ist.

In Abschnitt 2.1 wird ein kurzer Einblick in das Standardmodell der Teilchenphysik gegeben, welches die wesentlichen Erkenntnisse der Teilchenphysik in einem mathematischen Modell zusammenfasst. Die Theorie des Standardmodells bildet dabei die physikalische Grundlage dieser Arbeit. Ein umfassender Überblick zu dem Thema findet sich in [2].

Weiterhin wird in Abschnitt 2.2 eine kurze Übersicht zur Forschung und Detektion von Physik jenseits des Standardmodells, der Dunklen Materie, gegeben.

Abschließend wird in Abschnitt 2.3 auf das Experiment, den Compact-Muon-Solenoid-Detektor, auf dem die in dieser Arbeit verwendeten Datensätze basieren, eingegangen. In dieser Arbeit werden natürliche Einheiten verwendet. Damit gilt  $\hbar = c = 1$ .

### 2.1 Das Standardmodell

Das Standardmodell der Teilchenphysik beschreibt die bis heute bekannten fundamentalen Teilchen und deren Wechselwirkungen untereinander. Dabei beschreibt es drei der elementaren Grundkräfte der Physik. Diese sind die elektromagnetische, die schwache und die starke Wechselwirkung. Die vergleichsweise schwache Gravitation kann durch das Standardmodell nicht beschrieben werden.

Die im Standardmodell enthaltenen Elementarteilchen werden nach ihrem Spin in zwei unterschiedliche Typen sowie weitere Untergruppen unterteilt. Diese Kategorisierung ist in Abbildung 2.1 dargestellt.

Zum einen gibt es die Fermionen mit Spin  $1/2$ . Diese sind in drei Generationen aufgeteilt. Dabei besteht jede Generation aus je zwei Quarks und zwei Leptonen. Erstere tragen im Gegensatz zu den Leptonen eine nicht ganzzahlige elektrische Ladung und besitzen einen zusätzlichen Freiheitsgrad, eine der drei möglichen Farbladungen. Weiterhin kann man allen Quarks eine der Flavor-Quantenzahlen zuordnen, die sich nicht weiter klassifizieren lassen. In jeder der Gruppen gibt es drei der mit  $2/3$  Elementarladungen geladenen Quarks mit den Flavors Up, Charm oder Top. Außerdem gibt es drei Quarks mit den Flavors Down, Strange oder Bottom, deren elektrische Ladung  $-1/3$  der Elementarladung entspricht. Bei den Leptonen dagegen wird jeder der drei Generationen ein einfach negativ geladenes

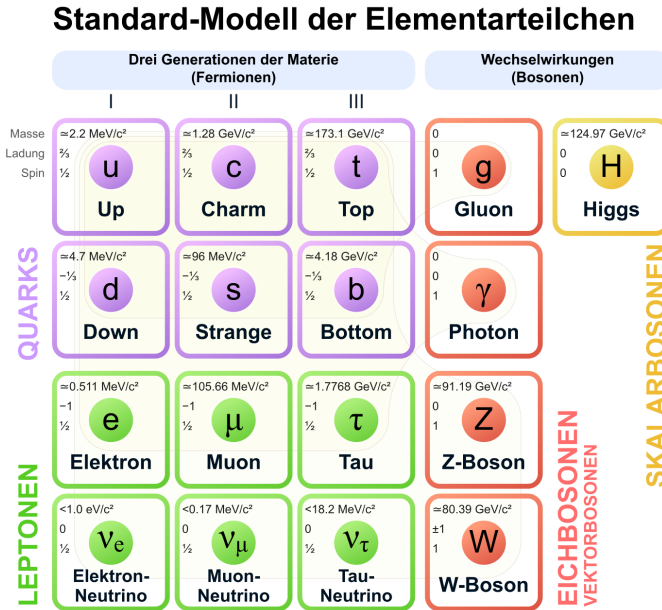


Abbildung 2.1: Das Standardmodell der Teilchenphysik mit allen darin vorhergesagten Elementarteilchen. Alle Ladungen sind in Einheiten von  $e$ , alle Spins in Einheiten von  $\hbar$  angeben. Entnommen aus [3].

Teilchen und ein zugehöriges neutrales Neutrino zugeordnet. Die Bezeichnungen für die negativ geladenen Teilchen sind Elektron, Muon und Tau, wonach auch die entsprechenden Neutrinos benannt sind. Weiterhin haben alle Leptonen und Quarks ein entsprechendes Antiteilchen mit gleicher Masse, jedoch entgegengesetzter elektrischer Ladung.

Zum anderen gibt es die Bosonen. Die Wechselwirkung zwischen Fermionen wird durch bosonische Austauschteilchen vermittelt. Vier von ihnen gehören der Gruppe der Eichbosonen an: Gluon, Photon, Z-Boson und W-Boson sind vektorartig und haben alle den Spin 1. Darüber hinaus gibt es das skalare Higgs-Boson, das den Spin 0 trägt.

Eine der drei vom Standardmodell berücksichtigten Kräfte ist die starke Wechselwirkung. Als Austauschteilchen für diese Interaktion dienen die farbgeladenen Gluonen, welche mit anderen farbbehafteten Teilchen, also Quarks und anderen Gluonen, wechselwirken. Dabei bewirken letztere durch ihre doppelte Farbladung einen Austausch der Quarkfarbe. Weiterhin müssen freie Teilchen nach einer wichtigen Eigenschaft der Quantenchromodynamik, dem Confinement der starken Wechselwirkung, immer nach außen hin farbneutral sein. Das resultiert in der Bindung zu Hadronen. So können sich beispielsweise drei Quarks oder ein Quark-Antiquark-Paar jeweils durch die Gluonen zu farbneutralen Baryonen bzw. Mesonen verbinden. Zu den Baryonen gehören beispielsweise das Proton und Neutron.

Die elektromagnetische Wechselwirkung dagegen basiert auf dem Austausch von Photonen zwischen elektrisch geladenen Teilchen. Zu diesen gehören die Quarks, geladene Leptonen und die  $W^\pm$ -Bosonen. Letztere sind außerdem zusammen mit den Z-Bosonen Austauschteilchen der schwachen Wechselwirkung. Diese koppelt an alle Quarks und Leptonen. Die Eichbosonen der schwachen Wechselwirkung sind im Gegensatz zu den Gluonen und Photonen stark massebehaftet. Bei hohen Energien kann die elektroschwache Vereinheitlichung angenommen werden, die in einer Kombination von elektromagnetischer und schwacher Wechselwirkung resultiert.

Eine Sonderstellung hat das Higgs-Boson inne, welches sich als gequantelte Anregung des Higgs-Feldes bezeichnen lässt. Durch Wechselwirkung mit eben diesem erhalten Fermionen, sowie W- und Z-Bosonen Masse.

## 2.2 Dunkle Materie

Mit dem hier kurz umrissenen Standardmodell der Teilchenphysik lassen sich viele Beobachtungen aus teilchenphysikalischen Experimenten erklären. Dennoch gibt es einige Unstimmigkeiten, wie die von Null verschiedene Masse der Neutrinos oder auch verschiedene kosmologische Experimente, die darauf hinweisen, dass die Theorie um das Standardmodell nur ein Teil des Universums erklären kann. Dieser Abschnitt orientiert sich an Ref. [4] und [5].

Ein Teil der so motivierten Physik jenseits des Standardmodells beschäftigt sich dabei mit der Suche nach der sogenannten Dunklen Materie. Das Konzept der Dunklen Materie beruht auf verschiedenen experimentellen Beobachtungen, die als Hinweis darauf interpretiert werden können, dass nur ein Teil der Materie im Universum sichtbar ist. Besagte Dunkle Materie emittiert keine messbare Strahlung, verursacht jedoch messbare Gravitation. Es wird davon ausgegangen, dass etwa zwei Drittel aller im Universum enthaltener Energie der sogenannten Dunklen Energie und etwa 25% der Dunklen Materie zugeordnet werden kann, für die es im Standardmodell keine Erklärung gibt. Da die im Standardmodell beschriebene baryonische Materie nur einen Anteil von etwa 5% des Universum ausmacht[6], kann so auch die Suche nach Dunkler Materie mit teilchenphysikalischen Experimenten motiviert werden. Weltweit gibt es etliche solcher Experimente, die sich mit der Suche nach Dunkler Materie und der Untersuchung ihrer Eigenschaften beschäftigen. Diese lassen sich nach den verwendeten Methoden in drei Kategorien einordnen. Alle beruhen dabei auf der Annahme, dass mögliche Dunkle Materie mit der Materie des Standardmodells wechselwirken kann.

Eine dieser drei Methoden ist die direkte Detektion, die auf der Suche nach Streuprozessen zwischen Standardmodell-Teilchen und Dunkler-Materie-Teilchen aus einem galaktischen Halo unserer Galaxie basiert. Dabei wird davon ausgegangen, dass letztere Teilchen an Atomkernen eines Targetmaterials streuen, was dann beispielsweise durch Szintillation oder Ionisierung detektiert werden kann. Ein Beispiel für ein solches Experiment ist EDELWEISS[7].

Die zweite Methode ist die indirekte Detektion. Diese basiert auf der Annihilation bzw. dem Zerfall von Dunkler-Materie-Teilchen in Standardmodell-Teilchen wie Photonen oder Elektron-Positron-Paaren in Regionen des Universums mit einer hohen Dichte an Dunkler Materie. Die entstehenden Standardmodell-Teilchen können dann mit Hilfe von bodengebundenen oder satellitengestützten Teleskopen detektiert werden. Beispiele dafür sind der FERMI-Satellit[8] und das IceCube-Experiment[9].

Ein weiterer Ansatz der Suche nach Dunkler Materie ist die direkte Erzeugung von eben dieser durch die Kollision von Teilchen des Standardmodells, insofern der Prozess erlaubt ist. Ein Beispiel für ein solches Experiment ist der LHC, der im nächsten Abschnitt vorgestellt wird. In Kapitel 3 wird näher auf diese Art der Detektion eingegangen.

Weiterhin kann angemerkt werden, dass die direkte und indirekte Suche nach Dunkler Materie und die kolliderbasierte Suche unterschiedliche Parameterbereiche abdecken und sich daher komplementär ergänzen.

## 2.3 Das CMS-Experiment

Ein Methode der experimentellen Teilchenphysik, um Hypothesen des Standardmodells zu überprüfen oder auch Hinweise für Physik jenseits des Standardmodells zu finden, sind

oft hochenergetische Streuexperimente. Dazu werden mit Hilfe von Teilchenbeschleunigern Teilchen stark beschleunigt und zur Kollision gebracht. Anschließend können die Zerfallsprodukte mit Hilfe komplexer Detektorsysteme untersucht und die gesammelten Daten analysiert werden. Das folgende Kapitel soll einen kurzen Überblick über den Large Hadron Collider (LHC) (vgl. Abschnitt 2.3.1) und das Compact Muon Solenoid-Experiment (CMS) (vgl. Abschnitt 2.3.2) geben. Nach einer kurzen Beschreibung des Aufbaus geht Abschnitt 2.3.3 auf das Koordinatensystem und die Kinematik am CMS-Detektor ein. Abschließend gibt Abschnitt 2.3.4 einen Ausblick auf die Vorgehensweise und Methoden der Analysen bei CMS.

### 2.3.1 Der Large Hadron Collider

Der momentan größte und hochenergetischste Teilchenbeschleuniger ist der am europäischen Kernforschungszentrum CERN bei Genf gelegene LHC[10]. Der LHC ist als Synchrotron der letzte in einer Kette von Beschleunigern am CERN. Innerhalb des 27 km langen Rings werden Protonen oder auch Bleiionen in gegenläufige Richtung auf nahezu Lichtgeschwindigkeit beschleunigt und anschließend bei verschiedenen Detektoren zur Kollision gebracht. Dabei werden Schwerpunktsenergien von bis zu 13 TeV erreicht. Die vier größten Detektorsysteme entlang des Hadronenspeicherrings sind LHCb (LHC beauty Experiment)[11], ALICE (A Large Ion Collider) [12], ATLAS (A Torodial LHC Apparatus)[13] und CMS[14]. Dabei ist der LHCb auf die Verletzung der CP-Symmetrie und auf die seltenen Zerfälle von B-Hadronen spezialisiert. An ALICE wird die starke Wechselwirkung untersucht. Dazu werden Schwerionen-Kollisionen bei sehr hohen Dichten und Temperaturen ausgeführt. Mit ATLAS und dem CMS dagegen werden viele unterschiedliche Phänomene untersucht. Dazu gehören Standardmodell-Messungen wie der Nachweis des Higgs-Bosons, aber auch Physik jenseits des Standardmodells wie die Suche nach Dunkler Materie. Da beide Detektorsysteme den gleichen physikalischen Zielen unter verschiedenen technischen Voraussetzungen folgen, lassen sich gemessene Daten überprüfen[13].

### 2.3.2 Der CMS-Detektor

Der Compact Muon Solenoid setzt sich aus mehreren Detektoren zusammen, die in zylindrischen Schichten um das Strahlrohr angeordnet sind. Wie sein Name impliziert, zeichnet er sich besonders durch seine trotz umfangreicher Funktionalität kompakte Bauweise aus und besitzt in seinem Inneren einen Solenoidmagneten und neben anderen Detektoren auch mehrere Myonenkammern. Die folgende Beschreibung des Aufbaus ist an Ref. [15] und [14] angelehnt. Eine visuelle Übersicht dazu bietet das in Abbildung 2.2 dargestellte Segment des Detektorquerschnittes.

Die innerste, direkt am Strahlrohr positionierte Schicht, besteht aus Silizium-Pixel- bzw. Silizium-Streifendetektoren. Zusammen bilden diese einen Spurdetektor, mit dem die Spur von geladenen Teilchen gemessen werden kann. Im Anschluss ist das elektromagnetische Kalorimeter positioniert. Es besteht aus Bleiwolframat-Kristallen, mit deren Hilfe die Energie von Photonen und Elektronen in Form von elektromagnetischen Schauern gemessen werden kann. Danach folgt ein hadronisches Kalorimeter, mit welchem die Energie beliebiger Hadronen bestimmt werden kann. Es besteht aus Messing oder Stahl, der mit Kunststoff-Szintillatoren oder Quarzfasern durchsetzt ist.

Im mittleren Teil des CMS-Detektors ist der namensgebende Solenoidmagnet positioniert. Seine Magnetfelder, die bis zu 4 T erreichen, zwingen alle geladenen Teilchen im Detektorsystem auf eine gekrümmte Bahn. Durch das Tracken der Spur der Teilchen lässt sich deren Krümmung und damit auch Impuls und Ladung bestimmen.

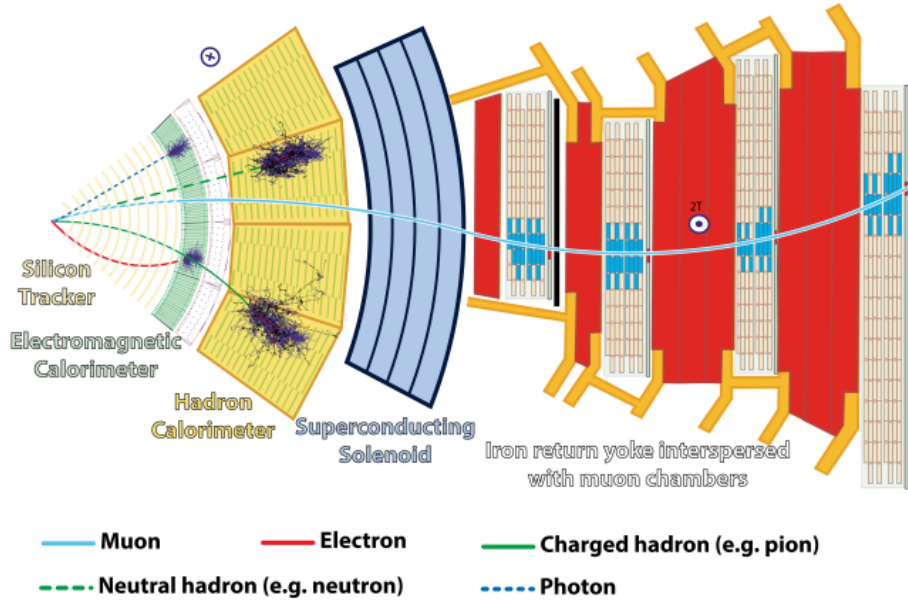


Abbildung 2.2: Schematische Darstellung eines Segments des CMS-Detektorsystems. Dieses setzt sich aus einigen Detektoren zusammen. Weiterhin sind auch die Spuren einiger mit dem Detektor wechselwirkender Teilchen angedeutet. Entnommen aus [16].

Im äußersten Abschnitt liegen vier Schichten von Myonenkammern. Diese bestehen aus Driftröhren, Kathodenstreifenkammern und Widerstandsplattenkammern. Mit ihnen lässt sich sehr präzise die Spur der bei der Kollision entstandenen Myonen bestimmen und damit deren Impuls rekonstruieren.

### 2.3.3 Kinematik

Das Koordinatensystem am CMS-Experiment ist üblicherweise so gewählt, dass der Kollisionspunkt der Protonen im Ursprung liegt. Die  $z$ -Achse ist so gewählt, dass sie parallel zum Teilchenstrahl liegt. Die  $x$ -Achse zeigt in Richtung des Mittelpunktes des LHC-Speicherrings, während die  $y$ -Achse vertikal nach oben gerichtet ist. Aufgrund der Symmetrie des Detektors werden üblicherweise anstatt der kartesischen Koordinaten  $(x, y, z)$  die zylindrischen Detektorkoordinaten  $(r, \eta, \phi)$  verwendet. Dabei ist  $r = \sqrt{x^2 + y^2}$  und  $\phi$  der Azimutalwinkel in der  $x$ - $y$ -Ebene. Die Pseudorapidität  $\eta$  ergibt sich aus dem Polarwinkel  $\theta$ , der im Bezug zur  $z$ -Achse steht (vgl. Gl.(2.1)).

$$\eta = -\ln \tan \left( \frac{\theta}{2} \right) \quad (2.1)$$

Eine weitere wichtige Größe für die Analyse von Detektordaten ist der Abstand zweier Punkte in der  $\eta\phi$ -Ebene

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \quad (2.2)$$

der abhängig von der Differenz der Azimutalwinkel  $\Delta\phi = \phi_2 - \phi_1$  und der Differenz der Pseudorapiditäten  $\Delta\eta = \eta_2 - \eta_1$  der beiden Punkte ist.

Wichtig für die indirekte Detektion von Teilchen wie Neutrinos oder auch möglichen Dunkel-Materie-Teilchen, die den Detektor ohne Wechselwirkung verlassen, ist besonders der Transversalimpuls  $\vec{p}_T = (p_x, p_y)^T$ . Dieser zweidimensionale Vektor entspricht der Komponente des Impulses in der  $xy$ -Ebene und steht senkrecht zur Strahlrichtung. Der Betrag

dieser Größe ist definiert als  $|\vec{p}_T| = p_T$ . Für einzelne Streuprodukte der Kollision ist  $p_T$  eine wichtige Messgröße in der Größenordnung einiger GeV bis zu einem TeV.

Es kann angenommen werden, dass der Transversalimpuls der Protonen vor der Kollision im Vergleich zu der Komponente des Impulses, die parallel zum Strahl liegt, vernachlässigbar klein ist. Da das Konzept der Impulserhaltung auch für Wechselwirkung von Dunkler Materie mit Materie des Standardmodells gültig ist, muss die Summe der Transversalimpulse der Kollisionsprodukte ebenfalls verschwindend gering sein.

Wenn jedoch während einer solchen Kollision Teilchen entstehen, die den Detektor unbemerkt verlassen, fehlt ein Bruchteil des Transversalimpulses. Diese Signatur wird als fehlende Transversalenergie

$$\vec{\cancel{E}}_T = - \sum_{i=1}^n \vec{p}_{T,i} \quad (2.3)$$

bezeichnet. Dabei ist  $n$  die Anzahl der detektierten Streuprodukten einer Kollision. Der Betrag dieser Summe des fehlenden Transversalimpulses bzw. der fehlenden Transversalenergie wird geschrieben als  $\cancel{E}_T$ .

Verwandt zur fehlenden Transversalenergie  $\cancel{E}_T$  ist der sogenannte hadronische Rückstoß

$$\vec{J}_T = \vec{\cancel{E}}_T + \vec{p}_T^{\text{Myonen/Elektronen/Photonen}}, \quad (2.4)$$

mit der Vektorsumme  $\vec{p}_T = \sum_{i=1}^{n(\text{Myonen/Elektronen/Photonen})} \vec{p}_{T,i}$  der Transversalimpuls aller mit hoher Rekonstruktionsqualität rekonstruierbaren Myonen, Photonen und Elektronen eines Ereignisses. Der hadronische Rückstoß beschreibt den der hadronischen Aktivität wegen entstehenden Transversalimpuls innerhalb eines Ereignisses mit Jets aus Myonen, Elektronen und Photonen produziert werden und dabei an den Jets abprallen können. Der hadronische Rückstoß dient zur Berücksichtigung des von der Dunklen-Materie weggetragenen Impulses.

### 2.3.4 Objektrekonstruktion bzw. Ereignissimulation

Durch das Erfassen der Zerfallsprodukte durch die verschiedenen in Abschnitt 2.3 vorgestellten Detektoren entstehen sehr große Datenmengen, die es zu analysieren gilt. Ziel dieser Analysen ist es, durch eine Vielzahl verschiedener Analysemethoden aus den Detektorsignalen Rückschlüsse auf die ursprünglichen, bei der Kollision entstanden Teilchenprozesse, ziehen zu können.

Im Folgenden wird kurz auf einen Aspekt der Objektrekonstruktion, die sogenannten Jets eingegangen. Damit soll ein besseres Verständnis für die in Kapitel 3 erläuterte Mono-top-Analyse ermöglicht werden.

Die durch Protonenkollisionen entstehenden Quarks und Gluonen rufen Hadronenschauer hervor, da sie wie in Abschnitt 2.1 beschrieben, der starken Wechselwirkung und damit auch dem Confinement unterliegen und daher Hadronen ausbilden, aus denen besagte Schauer resultieren. Die detektierten Teilchen des Hadronenschauers können mit Hilfe bestimmter Algorithmen zu Jets zusammengefasst werden. Aus diesen werden die Quarks und Gluonen rekonstruiert, die direkt nach der Kollision gebildet wurden. Dazu wird ausgenutzt, dass alle Teilchen des Schauers in Richtung des Partons fliegen, das den Schauer ausgelöst hat. Die grundlegende Rekonstruktion bzw. Identifikation der Teilchen aus den gemessenen Signalen in den Datensätzen wird mit Hilfe des **Particle-Flow-Algorithmus**[17] durchgeführt. Der am CMS für das Zusammenfassen der zuvor identifizierten einzelnen Teilchen hauptsächlich verwendete Algorithmus ist der **anti-k<sub>T</sub>-Algorithmus**[18]. In dieser Arbeit wird dabei ein Radiusparameter von  $R = 0,4$  verwendet. Die damit zusammengefassten Jets werden im Folgenden als AK4-Jets, oder kurz Jets, bezeichnet. Ein weiterer Algorithmus der



speziell für die Klassifizierung bzw. Identifizierung von Jets mit kleinerem Radiusparameter verwendet wird, die aus einem Bottom-Quark entstehen, ist der b-tagging-Algorithmus. Der b-tagging-Algorithmus ist in Ref. [19] beschrieben. Die so klassifizierten Jets werden als b-tagged Jets bezeichnet.

Um gemessene Daten mit theoretischen Daten zu vergleichen, oder auch um Analyseverfahren zu entwickeln oder zu verbessern, werden Simulationsdaten verwendet. Diese Datensätze werden mit Monte-Carlo-basierten Ereignisgeneratoren erzeugt. Die in dieser Arbeit verwendeten Datensätze sind ausschließlich Simulationsdaten. Informationen zu den Daten finden sich in Abschnitt 4.1.1.



## 3 Die Mono-top-Analyse

Wie bereits in Abschnitt 2.2 erwähnt, gibt es zahlreiche Hinweise für die Existenz Dunkler Materie. Daher gibt es weltweit eine Vielzahl von Experimenten deren Ziel es ist, die Natur Dunkler Materie besser zu verstehen.

Ein möglicher Ansatz Dunkle Materie zu detektieren liegt dabei darin, diese direkt zu erzeugen, sofern die Prozesse kinematisch erlaubt sind. Dazu werden am LHC Teilchen des Standardmodells zur Kollision gebracht. Somit kann am LHC Dunkle Materie möglicherweise erzeugt und nachgewiesen werden. Das folgende Kapitel ist angelehnt an Ref. [20] und [4].

Nach einer Vielzahl von Theorien kann ein hoher Betrag an fehlender Transversalenergie  $\cancel{E}_T$  Hinweis auf Physik jenseits des Standardmodells sein (vgl. Abschnitt 2.3.3). Eine möglicher Prozess, bei dem Dunkle-Materie-Teilchen entstehen, ist der sogenannte Mono-top-Prozess, der Gegenstand dieser Arbeit ist. In Abschnitt 3.1 wird auf dessen Signatur und Prozesseigenschaften eingegangen. Außerdem werden in Abschnitt 3.2 zwei weitere, reine Standardmodell-Prozesse mit ähnlicher Signatur vorgestellt, was sie zu wichtigen Untergrundprozessen macht. Diese sind namentlich der  $W$ +Jets- und der  $t\bar{t}$ -Prozess.

Weiterhin wird im Anschluss auf zwei Methoden eingegangen, den Mono-top-Prozess als Signal von Untergrundprozessen zu trennen und erfolgreich zu klassifizieren. Dabei wird in Abschnitt 3.3 zunächst ein Ansatz vorgestellt, der mit möglichst genauer Kenntnis der Eigenschaften der hier betrachteten Prozesse, eine Selektion durch verschiedene Schnitte und anderer Kriterien ermöglicht.

Gegenstand dieser Arbeit soll jedoch die Klassifizierung des Mono-top-Prozesses mit Hilfe von tiefen neuronalen Netzwerken sein. Diese haben sich aufgrund ihrer Leistungsfähigkeit in der Ereignisklassifikation großer Datensätze in der Teilchenphysik schon in vielen anderen Analysen als ebenbürtige, oder sogar überlegene Alternative zu klassischen Ansätzen erwiesen. In Abschnitt 3.4 wird auf die Funktionsweise eines solchen neuronalen Netzes eingegangen.

### 3.1 Der Mono-top-Prozess

Es ist anzunehmen, dass Dunkle-Materie-Teilchen mit der Materie des Detektors nicht messbar wechselwirken. Diese Hypothese stützt sich auf der Tatsache, dass Dunkle-Materie-Teilchen schon nachgewiesen worden wären, wenn sie elektromagnetisch oder stark wechselwirkten. Weiterhin ist ein gravitativer Nachweis bei der Detektion in einem Collider-Experiment auszuschließen, da die Gravitation auf kleinen Längenskalen sehr gering ist.

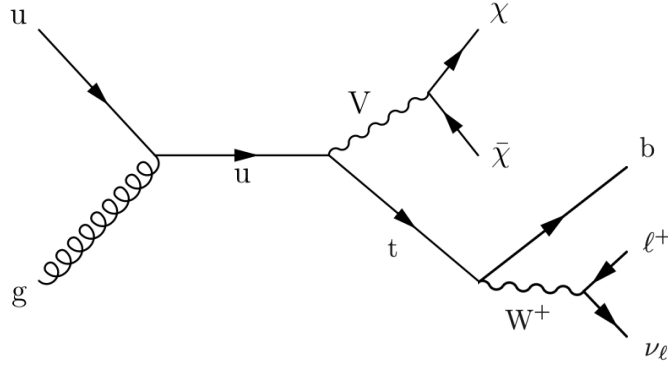
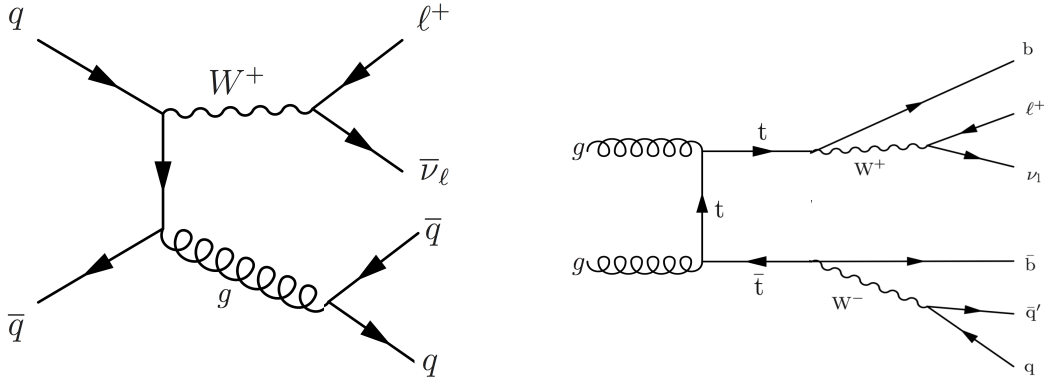


Abbildung 3.1: **Mögliches Feynman-Diagramm des Mono-top-Signalprozesses.**  
Entnommen aus ?? und angepasst mit ??.

Somit bleibt als letzte Möglichkeit der Interaktion der Dunkle-Materie-Teilchen mit den Detektoren die schwache Wechselwirkung bei sehr kleinen Wirkungsquerschnitten. Die in dieser Arbeit betrachteten Dunkle-Materie-Teilchen werden daher als *WIMPs* (engl. für *weakly interacting massive particles*) bezeichnet.

In Folge dessen sind Prozesse, bei denen nur Dunkle-Materie-Teilchen entstehen, nicht direkt nachweisbar. Deswegen kommen zum Nachweis dieser Teilchen nur Prozesse in Frage, bei denen zusätzlich ein Teilchen des Standardmodells als Endprodukt vorliegt. Am CMS-Experiment wird dazu unter anderem nach einem Prozess gesucht, bei dem als Endprodukt ein Top-Quark und zwei Dunkle-Materie-Teilchen entstehen. Diese wechselwirken wie in Abschnitt 2.3.3 erläutert nicht messbar mit dem Detektor und können daher nur indirekt durch beispielsweise einen hohen Betrag fehlender transversaler Energie  $\cancel{E}_T$  detektiert werden. Die Signatur für diesen Prozess besteht dann aus einem großen Betrag von  $\cancel{E}_T$  und aus einem Top-Quark, bzw. dessen Zerfallsprodukten.

Die Analyse in dieser Arbeit beruht auf einem effektiven Modell, welches die leptonische Mono-top-Signatur modellunabhängig beschreibt. Nach dieser Theorie kann die Signatur aus einem flavourändernden neutralen Strom resultieren, wobei *FCNCs* (engl. für *Flavor-changing neutral currents*) in niedrigster Ordnung im Standardmodell verboten sind. Nach dem hier betrachteten effektiven Modell wird der flavourändernde Strom über ein Vektor-Teilchen als Mediator vermittelt. Dieser, im Folgenden als Mono-top-Prozess bezeichnete Prozess, soll Gegenstand dieser Arbeit sein. Ein mögliches Feynman-Diagramm für den Prozess ist in Abbildung 3.1 dargestellt. Darin wird gezeigt wie das Up-Quark sich durch die Kopplung an dem Vektormediator  $V$  in ein Top-Quark umwandelt, wodurch sich netto der Flavor ändert. Der Vektormediator zerfällt im Anschluss in die zwei Dunkle-Materie-Teilchen  $\chi$  und  $\bar{\chi}$ . Das Top-Quark zerfällt in ein b-Quark und ein W-Boson. Ersteres bildet einen Jet aus während das W-Boson weiter in ein Lepton-Neutrino-Paar zerfällt. Die semileptonische Mono-top-Signatur, die in dieser Arbeit betrachtet wird, setzt sich damit aus der  $\cancel{E}_T$ , einem b-Jet und einem Lepton zusammen. Wie bereits zuvor in Abschnitt 2.3.4 erwähnt, lassen sich b-Jets experimentell mithilfe eines b-tagging-Algorithmus identifizieren. Dabei muss jedoch beachtet werden, dass dieser Algorithmus eine Fehlidentifikationsrate hat. Aus der eben beschriebenen Signatur lassen sich Selektionskriterien motivieren, auf die in Abschnitt 3.3 näher eingegangen wird.



(a) W+Jets, aus ?? entnommen und angepasst.

(b)  $t\bar{t}$ , aus ?? entnommen und angepasst.

Abbildung 3.2: **Mögliche Feynman-Diagramme der betrachteten Untergrundprozesse.** In (a) ist der W+Jets- und in (b) der  $t\bar{t}$ -Prozess dargestellt.

## 3.2 Die Untergrundprozesse

Bei Teilchenkollisionen am CMS-Detektor treten wie bereits erwähnt reine Standardmodell-Prozesse auf, die ähnliche Signaturen aufweisen wie der möglicherweise auftretende Mono-top-Prozess. Für eine erfolgreiche Klassifizierung des Signalprozesses ist ein gutes Verständnis dieser Untergrundprozesse essentiell. Im Folgenden wird auf zwei der wichtigsten Untergrundprozesse, den W+Jets- und den  $t\bar{t}$ -Prozess, eingegangen.

Der am häufigsten auftretende Untergrund-Prozess ist der W+Jets-Prozess. Auf Teilchen-niveau ist dessen Endzustand nicht derselbe wie der des Mono-top-Prozesses. Dennoch ist seine Signatur der Mono-top-Signatur sehr ähnlich, wie das Feynman-Diagramm des W+Jets-Prozesses in Abbildung 3.2a zeigt. Bei der Produktion des W-Bosons entstehen namensgebend auch Jets. Das W-Boson zerfällt in ein Lepton und das entsprechende Neutrino, welches vom Detektor nicht detektiert werden kann. Somit kommt eine Signatur mit einem geladenen Lepton und fehlender Transversalenergie  $\cancel{E}_T$  zustande. Weiterhin entstehen Jets, die aus einem leichten Quark oder einem Gluon stammen, die vom b-tagging-Algorithmus falsch als ein von einem b-Quark stammenden Jet identifiziert werden können. Als leicht werden die Quarks mit den Flavor Up, Down und Strange bezeichnet. In seltenen Fällen können auch Bottom-Quarks durch QCD-Wechselwirkung entstehen. Selbst wenn beide Möglichkeiten sehr selten auftreten, kann der W+Jets-Prozess wegen seines wesentlich größeren Wirkungsquerschnitts (vgl. Tabelle 4.1) den erwarteten Endzustand des Mono-top-Signalprozesses widerspiegeln.

Neben dem W+Jets-Prozess wird in dieser Arbeit außerdem die Produktion von Top-Quark-Antiquark-Paaren ( $t\bar{t}$ ) mit anschließendem semileptonischen Zerfall des Systems als weiterer wichtiger Untergrundprozess betrachtet. Ein Beispiel für ein Feynman-Diagramm eines  $t\bar{t}$ -Prozesses wird in Abbildung 3.2b dargestellt. Die entstehenden Top-Quarks zerfallen dabei jeweils in ein W-Boson und ein Bottom-Quark. Damit ergibt sich eine Signatur, die wie das Signal die b-tagging-Bedingung erfüllt. Beim semileptonischen Zerfall, bei dem angenommen wird, dass eines der W-Bosonen hadronisch und eines leptonisch zerfällt, kommt zu der Signatur weiterhin wieder ein geladenes Lepton und fehlende Transversalenergie  $\cancel{E}_T$  durch nicht detektierte Neutrinos hinzu. Wie zuvor beim W+Jets-Prozess ist auch der Wirkungsquerschnitt des  $t\bar{t}$ -Prozesses wesentlich höher als der des Mono-top-Prozesses (vgl. Tabelle 4.1), was ihn zu einem wichtigen Untergrund macht.

### 3.3 Klassische Ereignisselektion

Nachdem nun auf den Signal- und die beiden Untergrundprozesse eingegangen wurde, können nun auf Basis der Kenntnis dieser Prozesse Selektions-, bzw. Differenzierungskriterien besprochen werden. Dabei wird jedoch nur das Prinzip einer solchen Klassifizierung skizziert. Genau Werte und Spezifikationen finden sich beispielsweise in Ref.[20]. Dieser Abschnitt orientiert sich an Ref. [4] und [20].

Zunächst wird im Folgenden auf verschiedene Kriterien eingegangen, die zur Grundselektion verwendet werden. Im Anschluss werden verschiedene Differenzierungskriterien besprochen, die in drei orthogonalen Ereignisselektionen dazu dienen die Signal- und die beiden Untergrundregionen zu definieren. Letztere fungieren dabei als Kontrollregionen, die in der Analyse dazu verwendet werden können, die Untergründe innerhalb einer Maximum-Likelihood-Anpassung zu beschränken bzw. deren Unsicherheiten zu minimieren. Abschließend wird darauf eingegangen, wie die Sensitivität für die Klassifizierung des Signals weiter verbessert werden kann.

Zuvor muss jedoch eine weitere Größe, die sich aus dem Transversalimpuls des geladenen Leptons  $p_{T,l}$  und der fehlenden Transversalenergie  $\cancel{E}_T$  zusammensetzt, eingeführt werden. Die sogenannte transversale Masse ist definiert als

$$m_T = \sqrt{2p_{T,l}\cancel{E}_T(1 - \cos \Delta\phi(l, \vec{\cancel{E}}_T))}, \quad (3.1)$$

wobei  $\Delta\phi$  die Differenz der Azimutalwinkel zwischen dem Lepton und der fehlenden Transversalenergie  $\vec{\cancel{E}}_T$  bezeichnet. Die Verteilung der transversalen Masse  $m_T$  unterscheidet sich deutlich für den Signalprozess und die Untergrundprozesse (vgl. Abschnitt 7.2.2). Daher kann diese Größe als wichtiger Diskriminator verwendet werden.

Für die Grundselektion eignet sich, basierend auf den zuvor beschriebenen Prozessen, die Forderung nach beispielsweise genau einem isolierten Lepton, mindestens einem Jet mit großem Transversalimpuls und einem großen Betrag an fehlender Transversalenergie  $\cancel{E}_T$ . Durch die Grundselektion sollen leicht unterscheidbare Untergründe reduziert und der Signalprozess angereicht werden.

Zur Anreicherung der drei in dieser Arbeit betrachteten Prozesse können nun Kriterien verwendet werden, die auf den Unterschieden dieser Prozesse basieren. Das sind beispielsweise die Anzahl der Jets mit einem b-Tag und der Bereich der transversalen Masse  $m_T$ . Um die den Mono-top-Prozesses in der Signalregion anzureichern, kann für die Ereignisse genau ein Jet mit einem b-Tag mit einem besonders hohen Transversalimpuls wie beispielsweise 70 GeV, kein weiterer Jet mit einem Transversalimpuls größer 30 GeV und weiterhin eine transversale Masse die größer ist als beispielsweise 40 GeV ist, gefordert werden.

Für das Anreichern der Region des W+Jets-Prozesses dagegen kann ein Jet mit großem Transversalimpuls, z.B. 70 GeV, gefordert werden, der vom b-Tagger nicht als b-Jet identifiziert wird und darüber hinaus keine weiteren Jets mit einem hohen Transversalimpuls.

Für die dritte Region, die  $t\bar{t}$ -Region, bietet sich als Selektionskriterium für die Ereignisse die Anwesenheit von zwei Jets mit b-Tag an. Weiterhin ist es sinnvoll zu fordern, dass der Transversalimpuls des Jets mit dem höchsten Transversalimpuls größer als 70 GeV ist.

Wie bereits zuvor in Abschnitt 3.2 thematisiert haben die beiden Standardmodell-Prozesse einen deutlich größeren Wirkungsquerschnitt als der Mono-top-Prozess. Das gilt auch noch im Phasenraum von großer fehlender Transversalenergie  $\cancel{E}_T$  und Jets mit hohem Transversalimpuls  $p_T$ .

Daher ist es wichtig abschätzen zu können, wie viele der in der Mono-top-Region angereicherten Ereignisse tatsächlich dem Mono-top-Prozess zuzuordnen sind, da zu erwarten

ist, dass auch viele Untergrundereignisse, vor allem  $W$ +Jets und  $t\bar{t}$ , mitselektiert werden. Dazu wird ein gutes Verständnis der Untergründe benötigt. Eine mögliche Methode zu messen, wie viel mögliches Signal in den selektierten Daten tatsächlich enthalten ist, ist die Maximum-Likelihood-Methode. Diese ist näher in Ref. [21] erläutert. Die mit dieser Methode durchgeführten Messungen sind stark von den Unsicherheiten der Untergrundprozesse, wie beispielsweise die der Wirkungsquerschnitte, abhängig. Durch Minimierung dieser Unsicherheiten kann auch die Unsicherheit auf das Mono-top-Ergebnis minimiert werden.

Eine Möglichkeit die Unsicherheiten der Untergründe zu verringern sind die oben genannten Kontrollregionen um die Erwartungen bezüglich  $W$ +Jets und  $t\bar{t}$  in der Simulation zu präzisieren. Dabei werden Untergrund-Ereignisse, die die semileptonische Mono-top-Signatur, also die Signatur des Signal-Prozesses, zeigen, angereichert. Dabei wird die Selektion minimal variiert, indem die geforderte Anzahl der  $b$ -Tags angepasst wird. Bei einer Anpassung zu zwei Jets mit  $b$ -Tag werden folglich  $t\bar{t}$ -Ereignisse mit der Signal-Signatur und bei Null  $b$ -Tags  $W$ +Jets-Ereignisse mit der Signal-Signatur angereichert. Diese Ereignisse können dann in den Maximum-Likelihood-Anpassung mit aufgenommen werden. Die somit erreichte Verringerung der Unsicherheiten auf den Untergrund überträgt sich direkt auf die Signalregion mit einem  $b$ -Tag. Damit führt dieses Vorgehen auch zu einer Verbesserung der Sensitivität auf das Signal.

### 3.4 Ereignisselektion mit einem tiefen neuronalen Netz

Im Folgenden soll zunächst in Abschnitt 3.4.1 die Verwendung neuronaler Netze für die Klassifizierung motiviert werden, um dann in Abschnitt 3.4.2 auf die Funktionsweise und den Aufbau spezieller neuronaler Netze, den tiefen neuronalen Netzen, einzugehen. Letzterer Abschnitt orientiert sich dabei an Ref. [22], [23] und [24]. Die Spezifikationen der in dieser Arbeit verwendeten neuronalen Netze sind in Abschnitt 4.1.2 aufgeführt.

#### 3.4.1 Motivation

In dieser Arbeit sollen Datensätze mit einer großen Anzahl an Ereignissen (vgl. Abschnitt 4.1.1) in verschiedene Regionen angereichert, bzw. klassifiziert werden. Dabei zeigt der hier betrachtete Mono-top-Prozess, wie schon erwähnt, jedoch sehr große Ähnlichkeit zu anderen Standardmodell-Prozessen die zuvor in Abschnitt 3.2 als Untergründe eingeführt worden sind. Dieser Umstand erschwert die Zuordnung der vorliegenden Ereignisse anhand der kinematischen Variablen deutlich. Da neuronale Netze Informationen mehrerer Variablen zusammenfassen bzw. vereinigen können, kann sich die Klassifizierung dadurch verbessern. Während bei der klassischen Ereignisselektion effektiv die Anzahl der  $b$ -Tags die drei Regionen definieren wird dieses Kriterium bei der Multiklassifikation durch das Netz durch Informationen aus vielen Variablen ersetzt. Somit könnte eine reinere Anreicherung der drei jeweiligen Prozesse und somit auch eine größere Sensitivität für das Erkennen von möglichen Signalereignissen erreicht werden.

#### 3.4.2 Tiefe neuronale Netze

Im Folgenden wird kurz auf das allgemeine Prinzip der Klassifizierung mit einem neuronalen Netz eingegangen. Im Anschluss wird dann der grundlegende Aufbau einer speziellen Variante, den tiefen neuronalen Netzen (engl. *Deep Neural Networks*, kurz *DNNs*), beschrieben. Diese Variante findet Verwendung in den Analysen dieser Arbeit, da sie besonders geeignet für Klassifikations-Aufgaben ist. Weiterhin wird der Trainingsprozess vorgestellt und auf Möglichkeiten, das *Overfitting* zu verhindern eingegangen.

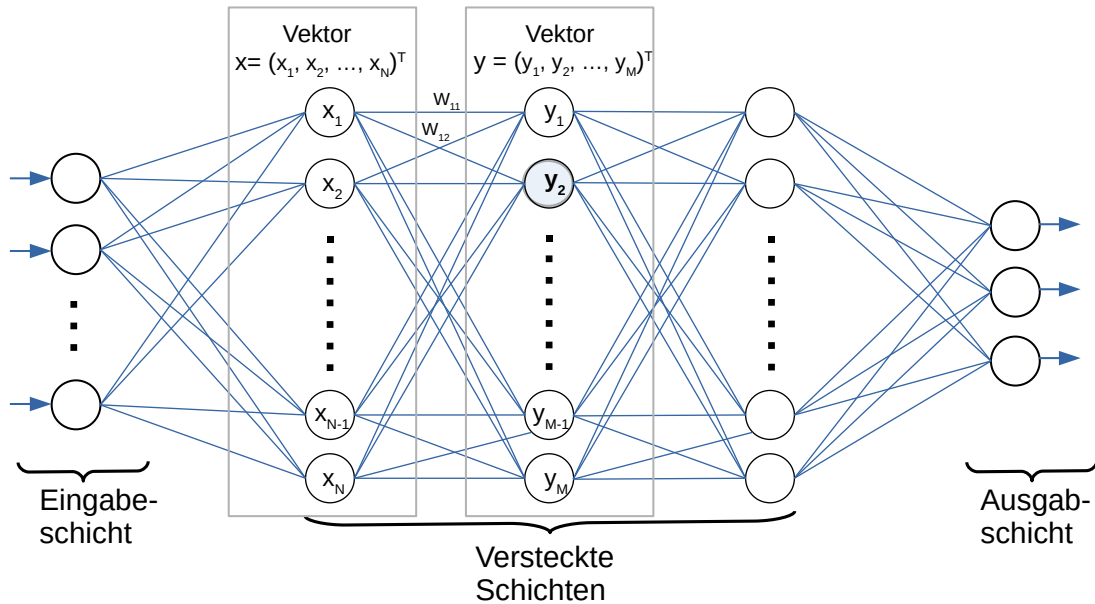


Abbildung 3.3: **Schematische Darstellung eines tiefen neuronalen Netzes.** Das dargestellte Netz besteht aus einer Eingabeschicht, drei versteckten Schichten und einer Ausgabeschicht. Jede Schicht besteht aus einem Vektor von Neuronen. Die benachbarten Schichten sind dabei jeweils vollständig miteinander verbunden. Die Pfeile und der Ein- bzw. Ausgabeschicht deuten den Informationsfluss durch das Netz an. In Abbildung 3.4 ist ein bestimmtes Neuron, z.B.  $y_2$  im Detail schematisch dargestellt.

## Prinzip

Analog zum Vorgehen in Abschnitt 3.3 ist die Idee der Ereignisklassifikation mit neuronalen Netzen, mithilfe von bestimmten Eigenschaften eines Ereignisses dessen Klasse vorherzusagen.

Um diese Vorhersagen leisten zu können, muss ein neuronales Netz zunächst trainiert werden. Dazu werden Simulationsdaten verwendet, bei denen die eigentliche Klasse schon bekannt ist. Letztere wird als wahre Klasse bezeichnet, während die vorhergesagte Klasse vom Netz aus den Eingabevariablen abgeleitet ist.

Für das Training wird der simulierte Datensatz üblicherweise zunächst in ein Trainings- und in einen Testdatensatz unterteilt. Mit Hilfe des Trainingsdatensatzes kann das Netz dann Korrelationen zwischen den Eingabevariablen und der wahren Ausgabeklasse lernen. Dabei kann es aber zu sogenanntem *Overfitting* kommen. Letzteres bedeutet, dass sich das Netz sehr genau den Trainingsdaten anpasst, jedoch nicht in der Lage ist zu verallgemeinern und neue Datensätze richtig zu klassifizieren. Um dies rechtzeitig zu erkennen und die Leistungsfähigkeit des Netzes im Allgemeinen zu bewerten kann der Testdatensatz verwendet werden. Darüber hinaus kann optional ein weiterer Teildatensatz zur unabhängigen Validierung der Trainingsresultate verwendet werden. Das kann sinnvoll sein, da oft nur so eine unverfälschte Abschätzung der Leistungsfähigkeit möglich ist, da der Testdatensatz oft verwendet wird, um abzuschätzen, wann der Trainingsprozess beendet werden kann. Wie bereits erwähnt sollen bei der Klassifizierung Ereignisse von einem DNN anhand ihrer verschiedenen Eingabevariablen einer bestimmten Klasse zugeordnet werden. Dabei muss jedoch beachtet werden, dass diese Zuordnung nie vollkommen fehlerfrei ablaufen kann. Daher ist es sinnvoll, als Ausgabe des Netzes nach dem Training eine Wahrscheinlichkeits-



verteilung für jede der Ausgabeklassen (in dieser Arbeit Mono-top, W+Jets und  $t\bar{t}$ ) zu wählen, die angibt, mit welcher Wahrscheinlichkeit das durch das Netz propagierte Ereignis den Klassen zugehörig ist.

### Aufbau

Ein neuronales Netz besteht üblicherweise aus mehreren Neuronen (engl. *neurons* oder auch *nodes*) die in verschiedenen Schichten (engl. *layers*) angeordnet sind. In der Eingabeschicht werden die zu untersuchenden Daten aufgenommen, während in der Ausgabeschicht die Ergebnisse der Vorhersage des Netzes abgelesen werden können. Ein tiefes neuronales Netz zeichnet sich durch mindestens eine versteckte Schicht (engl. *hidden layer*) von Neuronen aus, welche die Komplexität und somit auch die Extrapolationsfähigkeit des Netzes erhöhen können. Eine schematische Darstellung eines tiefen neuronalen Netzes ist in Abbildung 3.3 dargestellt.

Alle Eingabevariablen können dabei in einem Vektor zusammengefasst werden, wobei in der Eingabeschicht jede einzelne Variable als Eingabeneuron modelliert wird. Jedes dieser Eingabeneuronen gibt als Ausgabewert den Wert seiner Variable an die nächste Schicht, bzw. an alle der darin enthaltenen Knoten weiter. In den versteckten Schichten hat jedes Neuron der nächsten Schicht als Eingabe die Ausgaben aller Neuronen der vorherigen Schicht und eine Ausgabe, die sich aus dieser Eingabe und verschiedenen Parametern wie Gewichten zusammensetzt. In der letzten, als Ausgabeschicht bezeichneten Schicht, entspricht die Anzahl der Knoten  $k$  der Anzahl der Klassen, denen das Netz die Ereignisse zuordnen soll. In dieser Arbeit sind das z.B. die Prozesse Mono-top, W+Jets und  $t\bar{t}$ .

Um eine Rechenvorschrift für die Ausgabe jedes der Neuronen in den versteckten Schichten zu definieren, lassen sich die versteckten Schichten eines tiefen neuronalen Netzes als sogenannte *Dense Layers* beschrieben. Diese sind Funktionen  $f : \mathbb{R}^N \rightarrow \mathbb{R}^M$  welche  $N$  Neuronen  $x$ , dargestellt als ein Vektor in  $\mathbb{R}^N$  auf die nächsten  $M$  Neuronen  $y \in \mathbb{R}^M$  abbilden. Es gilt für jedes der nächsten Neuronen die Übertragungsfunktion

$$y = f(x) = \sigma(W \cdot x + b), \quad W \in \mathbb{R}^{N \times M}, \quad b \in \mathbb{R}^M, \quad \sigma : \mathbb{R} \rightarrow \mathbb{R}. \quad (3.2)$$

Die Parameter  $W$  und  $b$  werden dabei vom Netz gelernt, bzw. angepasst.  $W$  sind die Gewichte für die betrachteten Neuronen, während  $b$  der sogenannte *Bias*-Vektor ist. Letzteres ist ein weiteres, konstant auf 1 gesetztes Gewicht, welches als Verstärkungsfaktor mit allen Neuronen jeder der versteckten Schichten verbunden werden kann. Die Funktion  $\sigma$  ist die sogenannte Aktivierungsfunktion, die Nicht-Linearität in das Netz einführt. Dabei wird  $\sigma$  elementweise angewandt. Ohne sie könnte das Netz nur lineare Abbildungen modellieren. Die nicht-lineare Aktivierungsfunktion ist beliebig und wird meist empirisch gewählt. Eine schematische Detail-Ansicht eines bestimmten Neurons einer Schicht findet sich in Abbildung 3.4.

Eine gängige Aktivierungsfunktion ist die ELU-Funktion (kurz für engl. *exponential linear unit*), die einen frei wählbaren Hyperparameter  $\alpha$  besitzt:

$$elu(y_j) = \begin{cases} \alpha(e^{y_j} - 1) & y_j \leq 0 \\ y_j & y_j > 0 \end{cases}. \quad (3.3)$$

Für die letzte Schicht, die Ausgabeschicht, wird jedoch für Klassifizierungsprobleme üblicherweise an Stelle der Aktivierungsfunktion die **SOFTMAX**-Funktion

$$\text{softmax}(y_j) = \frac{e^{y_j}}{\sum_l^k e^{y_l}} \quad (3.4)$$

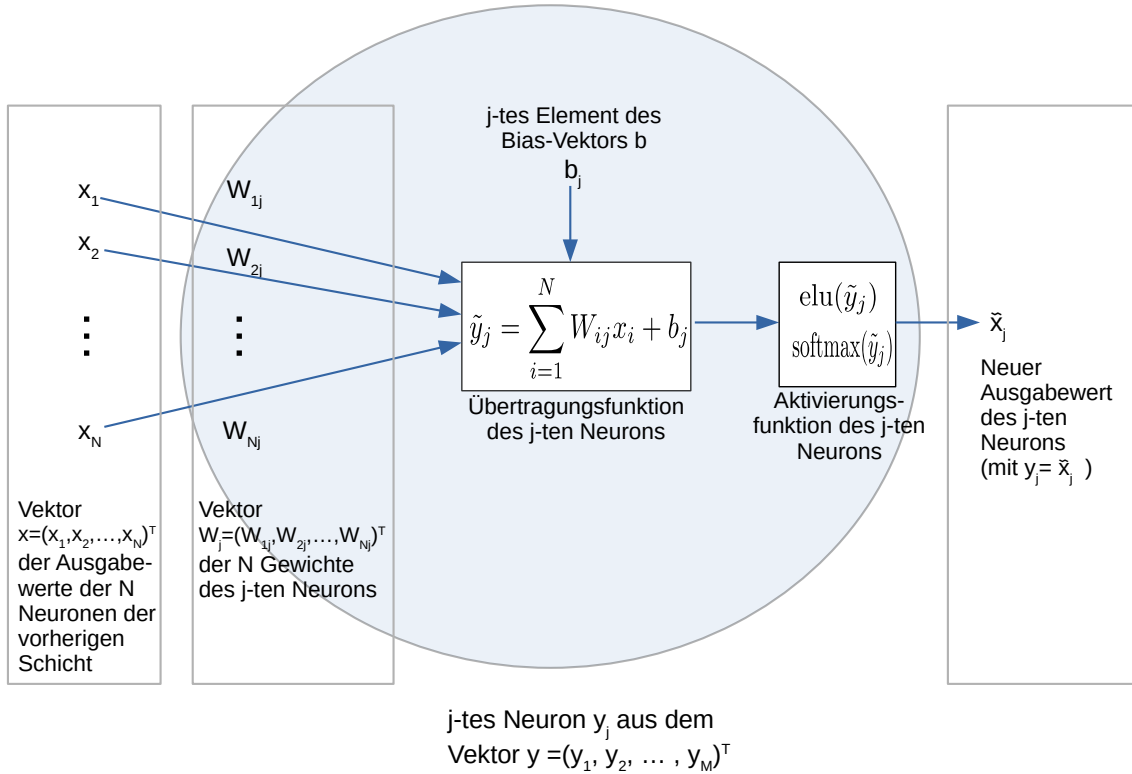


Abbildung 3.4: **Schematische Detail-Ansicht des Grundaufbaus eines  $j$ -ten Neurons einer Schicht.** Beispielsweise von dem in Abbildung 3.3 markierten Neuron  $y_2$  mit  $j = 2$ . Befindet sich das Neuron in einer verdeckten Schicht findet als Aktivierungsfunktion die ELU-Funktion Anwendung. Für den Fall, dass sich das Neuron in der Ausgangsschicht befindet wird stattdessen die SOFTMAX-Funktion angewendet. Darstellung in Anlehnung an Ref. [25].

verwendet. Wie bereits zuvor erwähnt finden sich in dieser Schicht, der Anzahl an zu unterscheidenden Kategorien  $k$  entsprechend,  $k$  Neuronen. Die SOFTMAX-Funktion bildet die Ausgaben  $y_j$  der einzelnen Neuronen dieser Schicht auf einen Wert zwischen 0 und 1 ab. Dabei ist die Summe dieser Ausgaben auf 1 normiert. Somit können die Ausgabewerte der Ausgabeneuronen als Wahrscheinlichkeiten für jede der Klassen interpretiert werden. Dabei entspricht die Vorhersage des Netzes bei der Kategorisierung in der Regel der Klasse, deren Ausgabeknoten den größten Wert hat.

### Der Trainingsprozess

Das Ziel des Trainings bzw. Lernprozesses ist es, die Parameter des Netzes so anzupassen, dass die wahre Klasse mit der Klasse des Neurons mit größter Wahrscheinlichkeit übereinstimmt. Um ein DNN zu trainieren braucht es ein Maß dafür, die sogenannte Kostenfunktion (engl. *loss function*), wie gut die Klassifizierung durch das Netz funktioniert, also wie gut die Vorhersage mit der wahren Klasse übereinstimmt. Dafür wird oft die als

$$\text{cat. crossentropy}(\hat{y}, y) = - \sum_i y_i \cdot \ln(\hat{y}_i) \quad (3.5)$$

definierte kategoriale Kreuzentropie (engl. *categorical crossentropy*) verwendet. Es entspricht  $\hat{y}$  dem Vektor der Werte der vorhergesagten Wahrscheinlichkeiten des Netzes für

jede Ausgabeklasse. Dabei nimmt  $y$  den Wert 1 für die wahre Klasse und 0 für alle anderen Klassen an. Je kleiner der Wert dieser Kostenfunktion ist, desto ähnlicher sind sich  $y$  und  $\hat{y}$ . Somit ist das Ziel des Trainings diese Funktion zu minimieren, indem die Parameter  $W$  in jeder Schicht entsprechend angepasst werden.

Um diese Minimierung zu erreichen werden Optimierungsalgorithmen verwendet, die auf einem sogenannten *Gradient-Descent*-Algorithmus basieren. Dieser nutzt die Tatsache, dass sowohl die versteckten Schichten als auch die Kreuzentropie nach den Parametern  $W$  und  $b$  differenzierbar sind. Dieser Gradient beschreibt die größtmögliche Minimierung der Kreuzentropie, bei der kleinstmöglichen Änderung der Parameter. Weiterhin wird der Gradient zur Ermittlung des nächsten Schrittes mit der sogenannten Lernrate multipliziert.

Für einen guten Verlauf des Trainings sollten zum Anpassen der Parameter möglichst viele Ereignisse verwendet werden. Wegen des hohen Zeit- bzw. Rechenaufwands werden die Parameter erst nach einer festgelegten Anzahl an Ereignissen, der sogenannten *batch size*, angepasst. Dadurch verändert sich die Dimension jeder Schicht, da der Vektor der Eingabevariablen und der aller Zwischenneuronen zu einer Matrix wird.

Besagter Zeitaufwand kann weiterhin dadurch minimiert werden, dass die Verlustfunktion jedes dieser Batches gemittelt wird und mit Hilfe eines sogenannten *Stochastic-Gradient-Descent*-Algorithmus nur dieser Mittelwert minimiert wird.

Als weitere Verbesserung zu den einfacheren Verfahren die auf Gradient-Descent beruhen, dienen Rückführungsalgorithmen. Diese führt die Parameteranpassung mit Hilfe eines iterativen Verfahrens aus, welches auch versucht vorherige Schritte mit einzubeziehen und die Gesamtzahl der benötigten Schritte reduziert. Ein häufig verwendeter Algorithmus dieser Art ist der **Adam**-Algorithmus.

Ein einfacher Durchlauf dieses Algorithmus durch alle Batches eines Datensatzes wird als Epoche bezeichnet. Dabei kann angemerkt werden, dass die Batches oft nach jedem Durchlauf zufällig neu aufgeteilt werden.

### Regularisierung

Wie bereits zuvor erwähnt ist ein häufig auftretendes Problem beim Training das sogenannte *Overfitting*, also dass das Netz zu stark an die Trainingsdaten angepasst wird und daher eine Verallgemeinerung auf neue Daten und damit die Klassifizierung sich im Laufe des Trainings verschlechtert.

Eine mögliche Strategie dem entgegenzuwirken ist die Verwendung von sogenannten *dropout layers*, die verhindern sollen, dass sich das Netz zu sehr von nur einigen wenigen Neuronen abhängig macht. Dazu setzten diese speziellen Schichten, die zwischen den versteckten Schichten liegen, einen festen Anteil der Neuronen 0 und passen den Rest entsprechend an, sodass die Summe aller Neuronen gleich bleibt.

Eine weitere Möglichkeit *Overfitting* zu verhindern, ist das Einführen eines zusätzlichen Regulierungsterms in die finale Kostenfunktion, welcher proportional zu einer bestimmten Norm der Parameter ist. Ein Beispiel dafür ist die  $L^2$ -Norm. Damit kann dieser Term dann einzelne Neuronen unterdrücken, welche eine wesentlich Größere Gewichtung als die Restlichen tragen. Somit kann der Trainingsprozess deutlich stabilisiert werden.



## 4 Methodik

Nachdem im letzten Kapitel tiefe neuronale Netze im Allgemeinen eingeführt wurden, soll in diesem Kapitel nun auf die praktische Umsetzung solcher Netze eingegangen werden. Die in dieser Arbeit durchgeführten Analysen basieren zum größten Teil auf der Open Source-Machine-Learning Framework-Sammlung `DRACO-MLFOY` [26]. Diese Sammlung bietet neben Werkzeugen für das Trainieren von klassifizierenden neuronalen Netzen für teilchenphysikalische Prozesse auch etliche weitere Anwendungen zur Auswertung und Untersuchung der Trainings-Ergebnisse.

In Abschnitt 4.1 wird zunächst auf die in dieser Arbeit verwendeten Simulationsdaten und die Konfigurationen der Netze eingegangen. Anschließend werden in Abschnitt 4.2 verschiedene Methoden vorgestellt, die im Folgenden zur Auswertung verwendet werden.

### 4.1 Verwendete Daten und Netzwerkkonfiguration

Vor dem Training mit einem neuronalen Netz muss zunächst ein Datensatz festgelegt werden. Dazu wird in Abschnitt 4.1.1 zunächst auf die Simulation und Aufbereitung der in dieser Arbeit verwendeten Datensätze eingegangen. Außerdem werden einige für die folgenden Analysen relevante Eigenschaften dieser Datensätze angegeben. Weiterhin müssen vor jedem Training die Hyperparameter und Eingabevariablen gewählt werden. In Abschnitt 4.1.2 werden die Hyperparameter, also Parameter des Netzes, die durch das Training nicht mehr angepasst werden, vorgestellt. In Abschnitt 4.1.3 werden abschließend die verwendeten Eingabevariablen genannt, sowie auf die Vorselektion eingegangen, die vor dem eigentlichen Training stattfindet.

#### 4.1.1 Informationen zu den verwendeten Daten

Die in dieser Arbeit verwendeten Simulationsdaten spiegeln die Bedingungen am CMS-Experiment während der Datennahme im Jahr 2018 wieder. Die Datensätze bestehen aus simulierten Daten von Teilchenkollisionen bei einer Schwerpunktsenergie von 13 TeV und werden mit Hilfe von Monte-Carlo-Ereignisgeneratoren erzeugt.

Wie in Kapitel 3 eingeführt, werden als Untergrund  $W$ +Jets- und  $t\bar{t}$ -Ereignisse betrachtet, wobei der Unterdatensatz der semileptonischen  $t\bar{t}$ -Ereignissen mit `POWHEGBOXV2@NLO QCD`[27] generiert wurde. Es ist zu beachten, dass wenn im Folgenden  $t\bar{t}$  erwähnt wird, immer der semileptonische, kurz „SL“,  $t\bar{t}$ -Prozess gemeint ist. Die Unterdatensätze der  $W$ +Jets-Ereignisse wurden als  $W$ -Produktion mit bis zu vier zusätzlichen Partonen auf

Tabelle 4.1: **Spezifikationen der in Kapitel 5 und 7 verwendeten Datensätze.** Die Signalthypothese hat eine Mediatormasse von  $M_\phi = 2495$  GeV und eine Dunkle Materie-Masse von  $M_\chi = 1250$  GeV. Die verwendeten Untergründe sind  $t\bar{t}$  semi-leptonisch und W+jets. Für W+Jets wurden alle verfügbaren Unterdatensätze aus dem Datensatz verwendet. Dabei handelt sich um insgesamt zehn Datensätze mit skalaren Summen der  $p_T$  aller Jets zwischen 70 und 2500 GeV und Impulsen ab 2500 GeV.

	Klasse	Wirkungsquerschnitt in pb	Anzahl der Ereignisse
Untergründe	$t\bar{t}$ sl	$3,65 \cdot 10^2$	$3,42 \cdot 10^6$
	W+Jets	$6,2 \cdot 10^4$	$5,33 \cdot 10^6$
Signal	Mono-top	$5,84 \cdot 10^{-4}$	$2,32 \cdot 10^5$

Matrixelement-Ebene in führender QCD-Ordnung mit `MadGraph5`[28] generiert. Die Mono-top-Signalthypothesen wurden in führender QCD-Ordnung mit `MadGraph5_aMC@NLO`[28] generiert, der dabei das sog. `DMsimp UFO Modell`[29][30] nutzt.

Für alle drei Prozesse wurde die Fragmentation und die Hadronisierung mithilfe von `Pythia8`[31] durchgeführt. Für die Detektorsimulation wurde `Geant4`[32] verwendet. In den Kapiteln 5 und 7 werden die in Tabelle 4.1 aufgeführten Datensätze für den Signal- und die beiden Untergrund-Prozesse verwendet. In Kapitel 6 werden nur verschiedene Signalthypothesen verwendet (vgl. Tabelle 4.2). Auf die Datensätze wurde eine Vorselektion angewendet, die vor dem Training erfolgt. Auf diese Selektion, die dazu dient den Untergrund zu reduzieren, wird in Abschnitt 4.1.3 eingegangen.

Tabelle 4.2: **Spezifikationen der in Kapitel 6 verwendeten Mono-top-Signal-Hypothesen.** Dabei wird unterschieden zwischen Trainings im Bereich hoher und niedriger Mediator-Massen und dem Training mit Datensätzen mit Mediator-Massen, die sich möglichst deutlich unterscheiden. In jedem der drei Fälle wurde jeweils einmal eine Multiklassifikation mit drei Klassen und ein Training mit zwei Datensätzen durchgeführt.

Kapitel	Massen-Bereich	Mediator-Masse $M_\phi$ in GeV	DM-Masse $M_\chi$ in GeV	Anzahl der Ereignisse	Wirkungsquerschnitt in pb
6.1	hoch	3000	1000	$2,32 \cdot 10^5$	$7,58 \cdot 10^{-4}$
		2500	750	$2,28 \cdot 10^5$	$2,81 \cdot 10^{-3}$
		750	150	$2,29 \cdot 10^4$	1,069
6.1	hoch	3000	1000	$2,32 \cdot 10^5$	$7,58 \cdot 10^{-4}$
		750	150	$2,29 \cdot 10^4$	1,069
6.2.1	niedrig	295	150	$1,49 \cdot 10^5$	2,35
		495	250	$1,82 \cdot 10^5$	$6,81 \cdot 10^{-1}$
		200	50	$1,19 \cdot 10^5$	$5,82 \cdot 10^1$
6.2.1	niedrig	495	250	$1,82 \cdot 10^5$	$6,81 \cdot 10^{-1}$
		500	150	$1,77 \cdot 10^5$	4,20
6.2.2	möglichst unterschiedlich	3000	1000	$2,32 \cdot 10^5$	$7,58 \cdot 10^{-4}$
		1500	150	$2,53 \cdot 10^4$	$5,65 \cdot 10^{-2}$
		200	50	$1,19 \cdot 10^5$	$5,82 \cdot 10^1$
6.2.2	möglichst unterschiedlich	3000	1000	$2,32 \cdot 10^5$	$7,58 \cdot 10^{-4}$
		200	50	$1,19 \cdot 10^5$	$5,82 \cdot 10^1$

Bei Betrachtung von Tabelle 4.1 zeigt sich, dass die Wirkungsquerschnitte der Untergründe einige Größenordnungen größer sind als der Signalthypothese. Weiterhin unterscheiden sich die Anzahl der enthaltenen Ereignisse um eine Größenordnung. Auch in Tabelle 4.2 sind deutliche Unterschiede in der Größe der Wirkungsquerschnitte und der Anzahl der Ereignisse zwischen den verschiedenen Signalthypothesen erkennbar. Dazu muss angemerkt werden, dass die durch die Simulation generierte Anzahl an Ereignissen zunächst nicht der Anzahl an Ereignissen entspricht, die physikalisch zu erwarten ist. Letztere wird durch den Wirkungsquerschnitt und die Luminosität bestimmt. Für das Training mit dem neuronalen Netz wird die generierte Ereignisanzahl so gewichtet, dass alle Datensätze bzw. Klassen beim Training gleichberechtigt sind.

#### 4.1.2 Netzkonfiguration

Die Architektur des Netzes entspricht dem in Abschnitt 3.4.2 beschriebenen Aufbau eines tiefen neuronalen Netzes. Bei der Wahl der Hyperparameter des Netzes wurden zunächst die Standardparameter verwendet. Diese sind in Tabelle 4.3 aufgelistet.

Neben der Eingabe- und Ausgabeschicht besitzt das Netz drei versteckte Schichten mit jeweils 100 Neuronen. Entsprechend der Anzahl der zu differenzierenden Klassen gibt es immer zwei oder drei Ausgangsneuronen.

Tabelle 4.3: **Hyperparameter der verwendeten DNNs.** Die folgenden Hyperparameter wurden für alle Multiklassifikation mit drei und zwei Klassen verwendet.

Hyperparameter	Wert
Architektur	3 Schichten mit je 100 Neuronen
Kostenfunktion	Categorical Cross-entropy
Dropout	0,5
Gewichtungs-Regularisierung	L <sup>1</sup> -Norm= 0 L <sup>2</sup> -Norm = $1 \cdot 10^{-5}$
Batchgröße	4096
Optimierungsmethode	ADAM
Lernrate	$1 \cdot 10^{-4}$
Aktivierungsfunktion	ELU
Aktivierungsfunktion Ausgabebayer	SOFTMAX
Earlystopping-Prozentanteil	0.05
Earlystopping-Epochen	100

Als Aktivierungsfunktion in den versteckten Schichten wurde ELU und in der Ausgabeschicht wie bereits zuvor in Abschnitt 3.4.2 erwähnt die SOFTMAX-Funktion verwendet.

Die verwendete Kostenfunktion ist die zuvor erläuterte Kreuzentropie mit einer Lernrate von  $1 \cdot 10^{-4}$ . Die darauf basierende Optimierungsmethode ist wie zuvor erwähnt der Adam-Algorithmus. Außerdem wird nach 10 Epochen ohne Verbesserung des Mittelwertes der Kostenfunktion auf Basis des Test-Datensatzes, des sogenannten *test loss*, das Training abgebrochen. Dasselbe gilt, wenn der Unterschied zwischen *test loss* und *trainings loss* 5 % übersteigt. Dabei ergibt sich letzterer aus den Trainings-Daten. Dabei werden 20 % des gesamt verfügbaren Datensatzes als Test-Datensatz verwendet.

Die Batchgröße wird zu 4096 festgelegt. Zur Vermeidung des Overfittings ist ein Dropout von 50 % nach jeder Schicht und eine L<sup>2</sup>-Norm von  $10^5$  gewählt. In Abschnitt 5.2 wird die

Anzahl der versteckten Schichten und deren Neuronenzahl variiert, um das Netz möglichst weiter zu optimieren.

#### 4.1.3 Eingabevariablen und Ereignisselektion

Für alle folgenden Studien werden die in Tabelle 4.4 aufgeführten Eingabevariablen verwendet. Dabei werden zunächst naiv die wichtigsten Variablen aufgrund der Prozesskinematik gewählt, um zu überprüfen, ob das neuronale Netz auf diese Weise in der Lage ist die drei betrachteten physikalischen Prozesse (vgl. Abschnitt 3.1) zu unterscheiden. Bewusst wurde dabei zunächst auf den Diskriminator der klassischen Klassifizierungsmethode, die transversale Masse  $m_T$ , verzichtet, um zu überprüfen, ob auch ohne  $m_T$  genügend Trennkraft aufgebracht werden kann.

Tabelle 4.4: **Verwendete Eingabevariablen.** Alle in der Beschreibung erklärten Variablen finden sich im Anhang in Tabelle B.1. Die transversale Masse  $m_T$  wurde bei den Trainings in Kapitel 5 zunächst nicht verwendet.

Eingabevariable	Beschreibung
Lepton_Pt[0], Lepton_Eta[0], Lepton_E[0]	$p_T$ , $\eta$ und $E$ des Leptons
Jet_Pt[0], Jet_Eta[0], Jet_E[0]	$p_T$ , $\eta$ und $E$ des führenden Jets
N_Jets	Anzahl der Jets im Ereignis
N_BTagsM	Anzahl der b-tagged Jets im Ereignis <sup>1</sup>
Evt_Pt_MET	fehlender Transversalimpuls $p_T(\cancel{E}_T)$
Hadr_Recoil_Pt	hadronischer Rückstoß des Ereignisses
DeltaPhi_AK4Jet_MET[0]	$\Delta\phi$ zw. führendem Jet und $\cancel{E}_T$
DeltaR_AK4Jet_LooseLepton[0]	$\Delta R$ zw. Lepton und führendem Jet
HT_AK4Jets	Skalare Summe der $p_T$ aller Jets $H_T$
M_W_transverse[0]	transversale Masse $m_T$

Im Folgenden wird anstelle des fehlenden Transversalimpulses die fehlende Transversalenergie  $\cancel{E}_T$  verwendet. Beide Größen werden als Äquivalent angesehen. Weiterhin sind mit der Bezeichnung Jets immer AK4-Jets[18] gemeint. Außerdem werden alle Eingabevariablen auf einen Mittelwert von Null und die Standardabweichung eins normiert. So wird sichergestellt, dass der Raum der Eingabevariablen nicht zu groß streut, also alle diese Variablen etwa denselben Wertebereich besitzen.

Vor der eigentlichen Klassifizierung wird durch eine Vorselektion schon ein Teil des Untergrundes extrahiert (vgl. Tabelle 4.5). Diese ist aus den in Kapitel 3 beschriebenen hier betrachteten Prozessen motiviert. So müssen alle Ereignisse genau ein Elektron und kein Myon oder aber kein Elektron und genau ein Myon enthalten. Außerdem müssen die Ereignisse, die in das Training mit eingehen mindestens einen Jet enthalten. Weiterhin muss die fehlende Transversalenergie  $\cancel{E}_T$  mindestens 100 GeV, die transversale Masse mindestens 40 GeV und der Transversalimpuls des führenden Jets mehr als 50 GeV betragen. Der Abstand  $\Delta R$  zwischen dem Lepton und dem führenden Jet muss weiterhin größer als 1,5 sein.

<sup>1</sup>Bei den b-tagged Jets handelt es sich um Jets, die von dem b-tagging-Algorithmus als solche klassifiziert werden, die aus einem Bottom-Quark entstehen. Der b-tagging-Algorithmus ist in [19] beschrieben.



Tabelle 4.5: **Auswahlkriterien der Vorselektion.** Die Auswahlkriterien sind durch die in Kapitel 3 beschriebenen Prozesse motiviert. Die Variablen sind in Tabelle B.1 im Anhang beschrieben.

Variable	Betrag
Anzahl der Myonen und Elektronen	genau eines von beiden
Anzahl der Jets	$\geq 1$
$\cancel{E}_T$	$\geq 100$ GeV
$m_T$	$\geq 40$ GeV
$p_T(\text{Jet}_F)$	$> 50$ GeV
$\Delta R(\text{Lep}, \text{Jet}_F)$	$> 1,5$

## 4.2 Auswertungsmethoden

Es gibt verschiedene Möglichkeiten, die Ergebnisse der Klassifizierung durch das neuronale Netz darzustellen und zu untersuchen. Weiterhin gibt es verschiedene Kriterien anhand derer die Leistungsfähigkeit des neuronalen Netzes bewertet und auch mit der anderer Klassifizierungsmethoden verglichen werden kann. Einige dieser Möglichkeiten und Kriterien, die in dieser Arbeit Verwendung finden, werden im Folgenden vorgestellt. Dabei kann angemerkt werden, dass mit Ausnahme der Wahrheitsmatrizen alle hier vorgestellten Methoden nicht nur für die Bewertung neuronaler Netze geeignet sind, sondern auch für die anderer Klassifizierungsmethoden.

### 4.2.1 Verteilungen der Eingabevariablen

Um einen ersten Eindruck von den Eigenschaften und vor allem von Trennkraft der Eingabevariablen zu bekommen, werden die Verteilungen der Eingabevariablen betrachtet. Dazu werden die Verteilungen der betrachteten Prozesse mit Histogrammen dargestellt. Es werden dabei jedoch nur die relativen Unterschiede untersucht, da für eine bessere Darstellung die Ereignisse vom Wirkungsquerschnitt bereinigt sind. Dazu wird die Mono-top-Region skaliert, der jeweilige Skalierungsfaktor findet sich immer in der Legende der Abbildung. Ein Beispiel dazu findet sich in Abbildung 5.1a.

### 4.2.2 Wahrheitsmatrizen

Eine schnelle Beurteilung der Qualität der Klassifizierung ermöglichen sogenannte Wahrheitsmatrizen (engl. *confusion matrix*). In Abbildung 5.2 ist ein Beispiel für eine solche Wahrheitsmatrix für eine Multiklassifikation mit drei Klassen dargestellt. In jedem der neun Felder steht jeweils der Anteil der wahren Ereignisse einer bestimmten Klasse, die von dem Netz einer der drei Klassen zugeordnet beziehungsweise vorhergesagt wurden. Dabei sind die wahren Klassen durch die Zeile festgelegt, die vom Netz vorhergesagten Klassen dagegen durch die Spalte. Die angegebenen Zahlenwerte sind auf die Gesamtzahl der Ereignisse der wahren Klasse der entsprechenden Zeile normiert. Damit ergibt die Summe aller Werte einer Zeile immer eins.

Bei der Betrachtung einer Wahrheitsmatrix ist die Diagonale von größtem Interesse. Auf ihr sind die sogenannten *True Positives* positioniert, also der Anteil der Ereignisse einer wahren Klasse, die das Netz richtig zugeordnet hat. So wurden in dem hier dargestellten Beispiel 82,2 % der wahren W+Jets richtig vom Netz klassifiziert. Die Felder abseits der Diagonalen geben an, welcher Anteil der entsprechenden wahren Klasse als die jeweilige falsch vorhergesagte Klasse vom Netz erkannt wurde. Diese Werte können einen Hinweis darauf geben, welche Klassen vom Netz nur schwer unterscheidbar sind.

### 4.2.3 Diskriminatorplots

Zur Visualisierung der Ergebnisse der einzelnen Ausgangsneuronen können Diskriminatorplots verwendet werden. Die Ausgabe der einzelnen Ausgabeneuronen ist jedoch nicht einfach die vorhergesagte Klasse, wie nach dem letzten Abschnitt angenommen werden könnte. Stattdessen wird beim Propagieren eines Ereignisses durch das neuronale Netz ein Wert zwischen null und eins ausgegeben. Bei einer Multiklassifikation gibt es mehrere Ausgabeneuronen, die mit der **SOFTMAX**-Funktion transformiert werden. Daher ergibt die Summe aller Ausgabeneuronen eins. Daher kann der Ausgabewert der einzelnen Ausgabeneuronen als Wahrscheinlichkeit interpretiert werden, dass ein bestimmtes Ereignis vom Netz einer bestimmten Klasse zugeordnet wird.

In einem Diskriminatorplot wird die Zuordnung von Ereignissen in die verschiedenen Klassen in Form eines Histogramms dargestellt. Die Histogrammierung erfolgt dabei über den Ausgabewert eines oder mehrerer Neuronen. Dabei trägt ein Ereignis nur zu dem Diskriminator des Ausgabeneurons bei, wenn es den größten Ausgabewert hat, also die Wahrscheinlichkeit der Zugehörigkeit zu der betrachteten Klasse am größten ist. Ein Beispiel dafür ist in Abbildung 7.4a dargestellt.

Als Diskriminator der klassischen Methode dient die Verteilung der transversalen Masse  $m_T$ . Zur Erzeugung der Diskriminatorplots wird zunächst die in Abschnitt 4.1 angegebene Vorselektion angewendet. Danach werden alle in den Datensätzen der drei in dieser Arbeit betrachteten Prozesse (vgl. Tabelle 4.1) enthaltenen Ereignisse jeweils nach der Anzahl der b-Tags angereichert. Wie zuvor in Kapitel 3 erläutert, sind in der Signatur des Mono-top-Prozesses ein b-Quark, bzw. der daraus resultierende b-Jet zu erwarten (vgl. Abbildung 3.1). Beim W+Jets-Prozess ist die Entstehung von b-Quarks sehr unwahrscheinlich, beim  $t\bar{t}$ -Prozess entstehen dagegen zwei b-Quarks (vgl. Abbildung 3.2a). Daher sind in der W+Jets-Signatur keine und in der  $t\bar{t}$ -Signatur zwei b-Jets zu erwarten. Da sich die Anzahl der entstehenden b-Quarks für alle drei Prozesse unterscheidet, eignet sich diese Größe also für eine einfache Klassifizierung.

Damit ergeben sich für die drei Datensätze jeweils drei Histogramme für die transversale Masse  $m_T$  mit allen Ereignissen mit keinem, einem und zwei oder mehr Jets mit b-Tag. Im Anschluss ergeben sich die Diskriminatoren für die verschiedenen Regionen, indem aus allen Datensätzen die Histogramme mit gleicher Anzahl an Jets mit b-Tag in einer Abbildung dargestellt werden (vgl. Abbildungen 7.6).

### 4.2.4 Reinheit

Eine Möglichkeit, den Erfolg einer Klassifizierung zu bewerten beziehungsweise zu vergleichen, ist die Reinheit (engl. *purity*). Dazu wird untersucht, wie rein die jeweiligen Klassen in den Prozessen sind, in denen sie angereichert werden sollen. Zur Berechnung der Reinheiten wird die Anzahl der Ereignisse einer Klasse, die richtig-positiv sind, durch die Gesamtanzahl der Ereignisse in dieser Klasse geteilt. In der Gesamtzahl der Ereignisse finden sich neben den richtig zugeordneten Ereignissen auch diejenigen, die falsch zugeordnet worden sind. In Abschnitt 7.3 wird diese Rechnung an einem Beispiel erläutert.

### 4.2.5 ROC-Kurven

ROC-Kurven (kurz für engl. *Receiver-Operating-Characteristic-Curves*) geben einen Anhaltspunkt dazu, wie gut die Klassifizierungsmethode dazu geeignet ist, eine bestimmte Klasse, zum Beispiel ein Signal, vom Rest, dem Untergrund, zu trennen. Damit enthalten sie im Grunde ähnliche Informationen wie die zuvor vorgestellten Reinheiten. Für das Erstellen einer ROC-Kurve wird der Anteil der Ereignisse, die einer bestimmten Klasse fälschlicherweise zugeordnet worden sind, die Falsch-Positiv-Rate, gegen den Anteil der

Ereignisse, der richtig klassifiziert wurde, die Richtig-Positiv-Rate, aufgetragen. Wie bereits zuvor erwähnt entspricht jeder Ausgabewert der Ausgangsknoten der Wahrscheinlichkeit, dass das Ereignis einer bestimmten Klasse angehört. Dabei lässt sich ein Arbeitspunkt, also ein Schwellwert für diesen Ausgabewert, festlegen, ab dem das Ereignis der entsprechenden Klasse zugeordnet wird. Jeder dieser Schwellwerte entspricht einem Punkt bestehend aus Falsch-Positiv- und Richtig-Positiv-Rate. Alles dieser Punkte zusammen ergeben die ROC-Kurve. Die Richtig-Positiv-Rate wird auch als Signal-Effizienz und die Falsch-Positiv-Rate als Untergrundunterdrückung bezeichnet.

Beispiele für solche ROC-Kurven finden sich in Abbildung 7.7. Für diese ROC-Kurven sind als Signal der Mono-top-Prozess und als Untergründe die W+Jets- und  $t\bar{t}$ -Prozesse gewählt. Verläuft eine ROC-Kurve entlang der Diagonalen, die die Enden beider Achsen verbindet, bedeutet das, dass die Klassifizierung vollkommen zufällig ist.

Je weiter die Kurve sich von dieser Diagonalen entfernt, also je größer die unter ihr eingeschlossene Fläche ist, desto besser funktioniert die Klassifizierung. Der sogenannte ROC-AUC-Wert (engl. *Area Under Curve*) entspricht dem Integral einer ROC-Kurve. Analog zu den vorherigen Erläuterungen zur ROC-Kurve bedeutet ein ROC-AUC-Wert von 0,5, dass das Netz die Ereignisse völlig willkürlich den verschiedenen Klassen zuordnet. Bei einem Wert von eins dagegen klassifiziert das Netz alle Ereignisse korrekt. Somit enthält der ROC-AUC-Wert eine Aussage über die Leistungsfähigkeit des Netzes beziehungsweise der Klassifizierung. Je näher der ROC-AUC-Wert an eins ist, desto besser funktioniert die Klassifizierung.



## 5 Multiklassifikation von Signal und Untergründen mit neuronalen Netzen

Quintessenz dieser Arbeit ist es, mit einem einfachen Ansatz zu überprüfen, ob ein neuronales Netz das Potential dazu hat, Mono-top-Signalthypothesen effizienter von Untergrundprozessen zu trennen, als es mit der in Abschnitt 3.3 vorgestellten klassischen Methode möglich ist. Sollte sich der Ansatz mit dem neuronalen Netz als vielversprechend erweisen, kann diese Arbeit Grundlage für weiterführende Studien sein, die dieses Konzept weiter ausarbeiten.

Dazu muss jedoch zunächst überprüft werden, ob ein neuronales Netz mit den verwendeten Eingabevariablen (vgl. Abschnitt 4.1.3) in der Lage ist, die Klassifizierung erfolgreich durchzuführen. In Abschnitt 5.1 werden zunächst die Eigenschaften der Eingabevariablen und deren Bedeutung für das Training untersucht. Weiterhin wird in Abschnitt 5.2 das Ergebnis eines ersten Trainings und eine Variation der Hyperparameter anhand von Wahrheitsmatrizen untersucht.

Die Analysen in diesem Kapitel wurden an einer Signalthypothese mit Mediatormasse  $M_\phi = 2495$  GeV und Dunkler Materie-Masse von  $M_\chi = 1250$  durchgeführt. Weitere Informationen zu den in diesem Kapitel verwendeten Datensätzen und Hyperparametern des Netzes finden sich in Abschnitt 4.1.

### 5.1 Die Eingabevariablen

Noch vor dem ersten Training werden zunächst die Verteilungen der Eingabevariablen erzeugt, um einen ersten Eindruck von deren Relevanz zu erhalten. Weiterhin kann so überprüft werden, ob die statistischen Verteilungen der Eingabevariablen der physikalischen Erwartung entsprechen. Alle Verteilungen finden sich im Anhang C.1. Es zeigt sich, dass sich die meisten Verteilungen der kinematischen Variablen für die drei Klassen Mono-top, W+Jets und  $t\bar{t}$  semileptonisch deutlich unterscheiden. Im Folgenden wird auf zwei charakteristische Verteilungen eingegangen.

Abbildung 5.1a zeigt die  $\cancel{E}_T$ -Verteilungen der drei betrachteten Prozesse. Wie nach Abschnitt 3 zu erwarten, fallen die Verteilungen für die Prozesse W+Jets und  $t\bar{t}$  zu steigenden Werten hin stark ab. Dagegen zeigt die Mono-top-Verteilung auch bei großen Werten der fehlenden Transversalenergie  $\cancel{E}_T$  einen größeren relativen Beitrag. An diesem Umstand lässt sich der in Abschnitt 3.1 beschriebene Einfluss der Dunklen-Materie-Teilchen

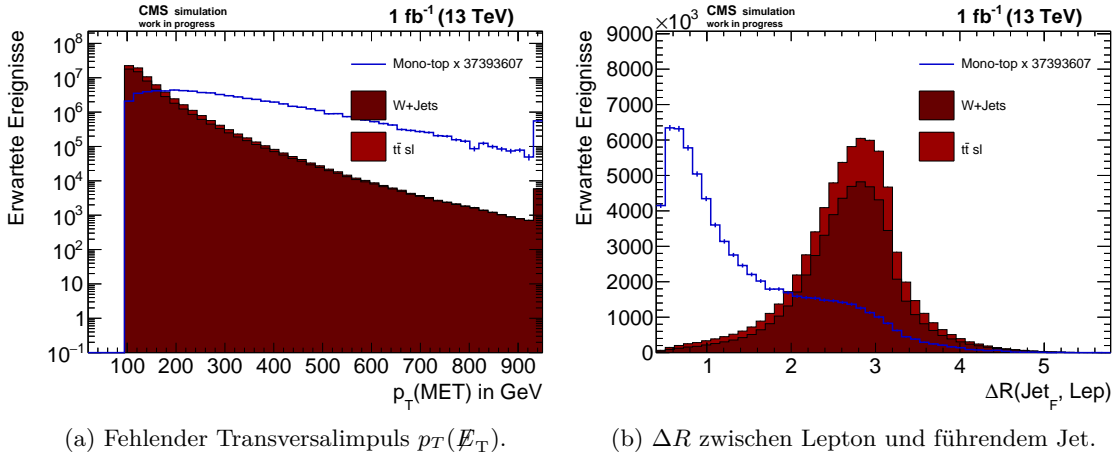


Abbildung 5.1: **Verteilungen zweier Eingabevariablen.** In (a) ist der fehlende Transversalimpuls  $p_T(\text{MET})$ , der als Äquivalent der fehlenden Transversalenergie  $\cancel{E}_T$  betrachtet wird, und in (b)  $\Delta R$  zwischen dem Lepton und dem führenden Jet dargestellt.

erkennen. Somit könnte die Variable  $\cancel{E}_T$  ein guter Diskriminator sein.

Die Betrachtung der in Abbildung 5.1b gezeigten  $\Delta R$ -Verteilung zwischen dem führenden Jet und dem Lepton lässt darauf schließen, dass diese Eingabevariable ebenfalls eine gute Trennkraft hat. Führender Jet bedeutet dabei, dass der Jet den größten Transversalimpuls  $p_T$  hat. Der räumliche Abstand  $\Delta R$  zwischen dem führenden Jet und dem Lepton ist für die beiden Standardmodell-Prozesse wesentlich größer als für den Mono-top-Prozess. Das lässt sich durch die große Masse bzw. durch den großen Impuls der neben dem Top-Quark entstehenden Dunkel-Materie-Teilchen erklären. Aufgrund von Impulserhaltung in der transversalen Ebene des Detektors erhält auch das Top-Quark, das dann in ein Bottom-Quark und ein W-Boson zerfällt, einen größeren Impuls (vgl. Abbildung 3.1). Das Lepton kommt aus dem W-Boson. Durch Hadronisierung resultiert das schon erwähnte Bottom-Quark in einem Jet. Sowohl das Lepton als auch das Bottom-Quark sind wegen des größeren Impulses kollimiert. Somit ist weiterhin zu erwarten, dass sich der Abstand  $\Delta R$  bei höheren Massenpunkten verkleinert.

Die oben genannten Vermutungen zur Relevanz der Variablen  $\cancel{E}_T$  und  $\Delta R$  werden im folgenden durch ein Ranking der Eingabevariablen überprüft. Dazu wird aus 16 Trainingsläufen mit jeweils 25 Epochen der relative Mittelwert der Gewichte der Variablen in der ersten Schicht bestimmt. Das daraus resultierende Ranking findet sich im Anhang in Abbildung C.8. Es zeigt sich, dass wie erwartet die fehlende Transversalenergie  $\cancel{E}_T$  eine hohe Trennkraft hat. An zweiter Stelle folgt der hadronische Rückstoß der jeweiligen Prozesse. Der Abstand  $\Delta R$  zwischen führenden Jets und dem Lepton dagegen findet sich entgegen der naiven Erwartung in der unteren Hälfte des Rankings.

Eine mögliche Erklärung dazu ist, dass diese Größe stark korreliert ist zu anderen Eingabevariablen mit einer guten Trennkraft. Wenn das Netz also Größen wie die fehlende Transversalenergie  $\cancel{E}_T$  und den hadronischen Rückstoß stark verwendet, kommt es auch zu einer impliziten Nutzung von  $\Delta R$ . Damit macht dann eine explizite Nutzung von  $\Delta R$  keinen Unterschied mehr.

## 5.2 Ergebnisse des Trainings

Zunächst wird das Netz mit 100 Epochen trainiert. Wie in Abschnitt 4.2.2 beschrieben sind Wahrheitsmatrizen eine Möglichkeit, die Ergebnisse des Trainings darzustellen und zu



Abbildung 5.2: **Wahrheitsmatrix.** Training mit einem Mono-top-Signal und den W+Jets- und  $t\bar{t}$ -Untergründen mit 100 Epochen. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben.

bewerten. An der in Abbildung 5.2 dargestellten Wahrheitsmatrix ist zu erkennen, dass die Trennung in den Signal-Prozess Mono-top und die beiden Untergrund-Prozesse W+Jets und  $t\bar{t}$  sehr gut funktioniert. So schafft es das Netz jeweils über 80 % der Ereignisse richtig zuzuordnen. Die Mono-top-Klassifizierung funktioniert mit einer Wahrscheinlichkeit von 89,2 % am besten. Das entspricht den Erwartungen, da das Netz darauf ausgelegt ist, das Signal vom Untergrund zu trennen. So gibt es unter den kinematischen Variablen wie bereits erwähnt einige, bei denen zu erkennen ist, dass sich das Signal deutlich vom Untergrund unterscheidet (vgl. Anhang C.1). Dazu gehören auch die zuvor beschriebene fehlende Transversalenergie  $\cancel{E}_T$  und der hadronische Rückstoß. Daher ist unter Verwendung dieser Variablen eine gute Leistung des Netzes zu erwarten.

Weiterhin wird untersucht, ob durch eine Variation der Anzahl von Schichten und Neuronen in den Hyperparametern des Netzes eine Leistungsverbesserung erreicht werden kann. Dazu werden die Parameter verändert, das Netz für 25 Epochen trainiert und die resultierenden Wahrheitsmatrizen betrachtet. In der Standardkonfiguration, die bisher für das Training verwendet wird, gibt es drei Schichten mit je 100 Neuronen (vgl. Tabelle 4.3). Eine leichte Verbesserung in der Trennkraft des Netzes lässt sich bei 5 Schichten mit je 300 Neuronen anhand der Wahrheitsmatrix ablesen (vgl. Abbildung 5.3). Um dem Vergleich mehr Aussagekraft zu verleihen wird das Training mit 25 Epochen und der Variation in der Konfiguration mit fünf Schichten und je 300 Neuronen 15 Mal wiederholt. Beim Vergleich der resultierenden Wahrheitsmatrizen mit denen aus der Standardkonfiguration zeigt sich, dass die leichte Verbesserung der Trennkraft relativ stabil ist. Da diese Verbesserung beispielsweise bei der Klassifizierung des Signals nur zwischen einem halben und einem Prozent liegt, wird davon abgesehen, die Konfigurationsänderung für die folgenden Trainings zu übernehmen. Die in Tabelle 4.3 aufgeführten Hyperparameter werden also beibehalten.

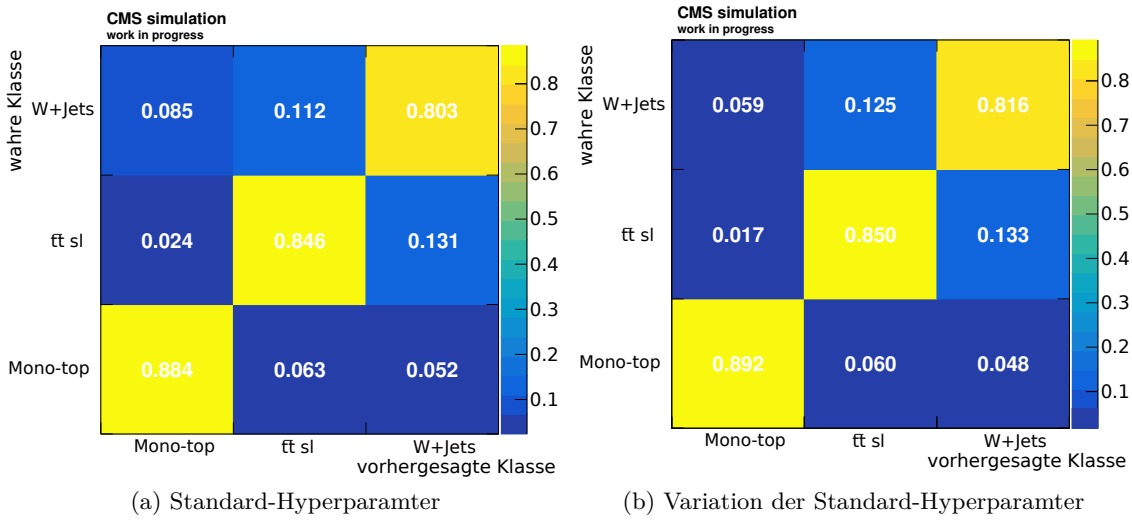


Abbildung 5.3: **Wahrheitsmatritzen im Vergleich.** In (a) abgebildet ist die Standard-Hyperparameter mit drei versteckten Schichten und jeweils 100 Neuronen und in (b) die Variation mit fünf versteckten Schichten und jeweils 300 Neuronen.

Außerdem zeigt sich bei der Betrachtung des Variations-Plots der wichtigsten Eingabevariable, der fehlenden Transversalenergie  $\cancel{E}_T$ , dass das neuronale Netzwerk sehr sensitiv auf diese ist. Zur Erzeugung des in Abbildung 5.4 gezeigten Variations-Plots werden zunächst die Mittelwerte der Eingabewerte der  $\cancel{E}_T$  und die Standardabweichung der  $\cancel{E}_T$ -Verteilung  $\sigma$  ermittelt. Die Mittelwerte werden dann um  $\pm 2\sigma$  variiert und gegen die Änderung der Ausgabewerte für die Neuronen vom Signal Mono-top und dem Untergrund aus W+Jets und  $t\bar{t}$  aufgetragen.

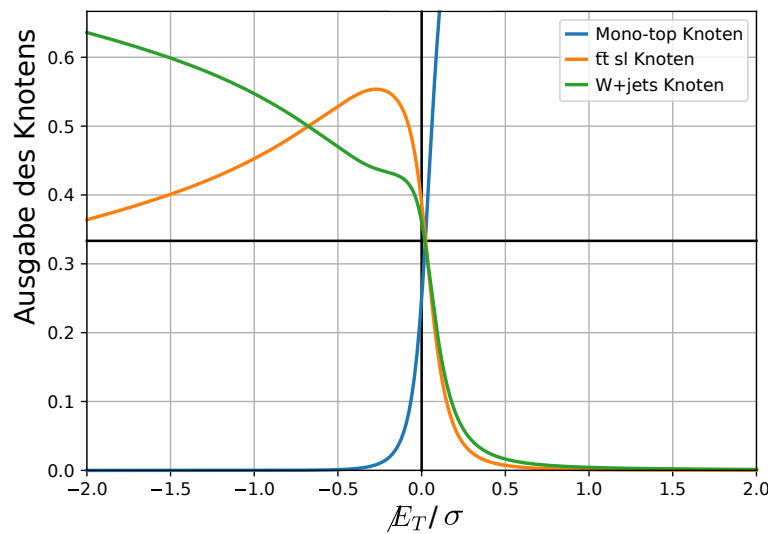


Abbildung 5.4: **Variationsplot der fehlenden Transversalenergie  $\cancel{E}_T$ .** Aus dem Training mit 100 Epochen mit einem Mono-top-Signal und den W+Jets- und  $t\bar{t}$ -Untergründen.



Die Vorhersage für das Mono-top-Neuron steigt durch eine leicht erhöhte fehlende Transversalenergie  $E_T$  sehr steil an, während die für die Untergrund-Neuronen steil abfallen. Das unterstreicht einerseits die Wichtigkeit und starke Trennkraft der  $E_T$  für das neuronale Netz. Andererseits kann die große Sensitivität des Netzes auf die  $E_T$ , welche in starker Korrelation zu der Masse steht, unter Umständen auch zu einer starken Sensitivität des Netzes auf den Massenpunkt, bzw. die Masse führen. Dies kann dazu führen, dass die Netzausgabe verzerrt wird. Wird das Netz mit einer anderen Signalthypothese ausgewertet, könnte die Leistungsfähigkeit des Netzes darunter leiden, wenn es dann nicht optimal auf die Masse abgestimmt ist. Um diese Beobachtungen weiter zu quantifizieren wird im nächsten Kapitel 6 untersucht, wie stark die Eigenschaften eines bestimmten Massenpunktes die Leistung des Netzes beeinflusst.



## 6 Klassifizierung verschiedener Monotop-Signalhypothesen

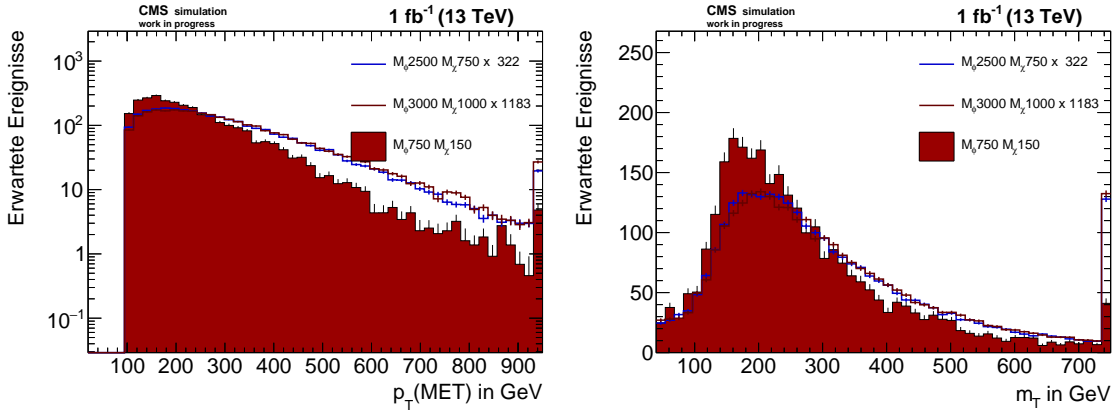
Wie bereits zuvor erwähnt, scheint das neuronale Netz sehr sensitiv auf die fehlende Transversalenergie  $\cancel{E}_T$  zu sein. Diese Größe steht in enger Korrelation mit der Masse des Mediators  $M_\phi$ . Daher ist es interessant zu untersuchen, wie stark sich das Netz auf die Eigenschaften eines bestimmten Massenpunktes, also einer bestimmten Signalhypothese, fixiert. Um dem nachzugehen wird im Folgenden überprüft, ob das Netz für verschiedene Signalhypothesen eine Trennkraft aufbringen kann. Zunächst wird mit Hypothesen mit sehr ähnlichen  $\cancel{E}_T$ -Verteilungen trainiert um die Klassifizierung so schwer wie möglich zu machen. Dieser Versuch wird in Abschnitt 6.1 einmal mit großen Mediatormassen und einmal mit niedrigen Massenpunkten in Abschnitt 6.2.1 durchgeführt. Im Anschluss werden diese Trainings in Abschnitt 6.2.2 für möglichst weit auseinander liegende Massenpunkte wiederholt. Darüber hinaus wird in Abschnitt 6.3 das Verhalten des Netzes im Zusammenhang mit ununterscheidbaren Signalhypothesen genauer untersucht. Weiterhin wird in Abschnitt 6.4 der Einfluss der Mediatormasse auf die Anzahl Jets untersucht. In Abschnitt 6.5 wird über die Erkenntnisse aus diesem Kapitel reflektiert.

### 6.1 Klassifikation von Signalhypothesen im Bereich hoher Mediatormassen

Zunächst wird das Netz mit drei Signalhypothesen mit Mediatormassen im hochenergetischen Bereich trainiert (vgl. Tabelle 6.1). Die Auswahl der Mono-top-Datensätze richtet sich dabei nach deren  $\cancel{E}_T$ -Verteilungen. Durch ähnliche Verteilungen wird für die Klassifizierung eine möglichst anspruchsvolle Ausgangssituation gewählt.

#### 6.1.1 Betrachtung der Eingabevariablen

Abbildung 6.1a zeigt, dass die  $\cancel{E}_T$ -Verteilungen der beiden Datensätze mit Mediatormassen  $M_\phi = 3000$  GeV und  $M_\phi = 2500$  GeV kaum unterscheidbar sind. Auch die Verteilungen der übrigen kinematischen Eingabevariablen unterscheiden sich kaum für die beiden höheren Massenpunkte. Der Massenpunkt  $M_\phi = 750$  GeV dagegen unterscheidet sich deutlicher (vgl. Anhang D.1). Die transversale Masse wird nun als zusätzliche Eingabevariable verwendet, da diese ähnlich zum fehlenden Transversalimpuls durch die Masse des Mediators beeinflusst wird. Diese Variable wird bei der klassischen Methode zur Trennung des Mono-top-Signals vom Untergrund verwendet (vgl. Abschnitt 3.3). Wie bereits erwähnt wurde



(a) Fehlender Transversalimpuls bzw. fehlende Transversalenergie  $E_T$ .

(b) Transversale Masse  $m_T$ .

Abbildung 6.1: **Verteilungen zweier Eingabevariablen für drei Signalthypothesen im Bereich hoher Mediatormassen.** In (a) ist die der fehlenden Transversalenergie  $E_T$  abgebildet und in (b) die der transversalen Masse  $m_T$ .

sie zuvor nicht als Eingabevariable verwendet, um eine deutlichere Differenzierung zu der klassischen Methode zu ermöglichen. Die Betrachtung des  $m_T$ -Verlaufes zeigt jedoch, dass auch hier Datensätze im Bereich hoher Mediatormassen kaum unterscheidbar sind (vgl. Abbildung 6.1b). Daher ist anzunehmen, dass das Netz selbst mit der zusätzlichen Variable für zwei der Massenpunkte kaum Trennkraft aufbringen kann, da die beiden Klassen zu ähnlich sind.

Während die Eingabevariablen zuvor unter dem Aspekt der möglichen Trennkraft des Netzes betrachtet worden sind, fällt über diesen Aspekt hinaus, die Verteilung der Anzahl der Jets auf, die sich entgegen der naiven Erwartung des Prozesses verhält. Dieser Umstand wird näher in Abschnitt 6.4 beschrieben und untersucht.

Tabelle 6.1: **Massen der verwendete Mono-top-Signalthypothesen.** Training im Bereich hoher Mediatormassen.

Klassifikation	Massen-Bereich	Mediatormasse $M_\phi$ in GeV	DM-Masse $M_\chi$ in GeV	Anzahl der Ereignisse
Multi	hoch	3000	1000	$2, 32 \cdot 10^5$
		2500	750	$2, 28 \cdot 10^5$
		750	150	$2, 29 \cdot 10^4$
Binär	hoch	3000	1000	$2, 32 \cdot 10^5$
		2500	750	$2, 28 \cdot 10^5$

### 6.1.2 Ergebnisse des Trainings

Das Netz wird zunächst mit den drei in Tabelle 6.1 aufgeführten Signalthypothesen für jeweils 50 Epochen trainiert. Dazu werden einmal die in Abschnitt 4.1.3 eingeführten Eingabevariablen verwendet. In einem weiteren Training wird zusätzlich die transversale Masse  $m_T$  verwendet um zu überprüfen, ob so eine verbesserte Klassifizierung erreicht werden kann.

Die in Abschnitt 6.1.1 getroffenen Annahmen bestätigen sich, wie die Betrachtung der in Abbildung 6.2a dargestellten Wahrheitsmatrizen zeigt: Die Klasse des Massenpunktes

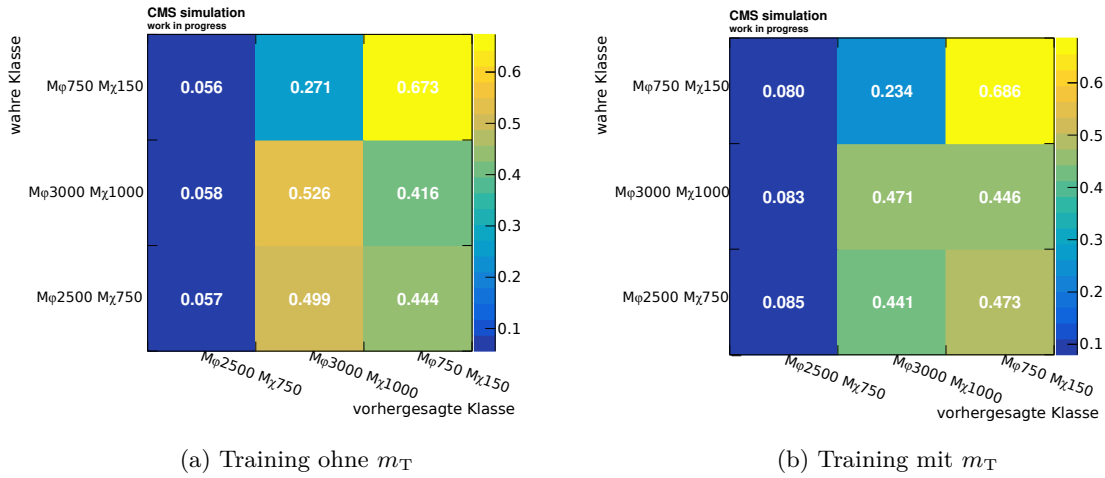


Abbildung 6.2: **Wahrheitsmatrizen.** Training mit drei Mono-top-Signalthypothesen mit jeweils 50 Epochen im Bereich hoher Mediatormassen. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. In (a) wird im Gegensatz zu (b) ohne die Eingabevariable transversale Masse  $m_T$  trainiert.

$M_\phi = 2500$  GeV ist deutlich unterrepräsentiert. Auch mit der zusätzlichen Eingabevariable  $m_T$  verbessert sich die Klassifizierung kaum (vgl. Abbildung 6.2b). Auch durch eine Verdopplung der Anzahl der Epochen verbessert sich die Zuordnung nicht.

Dieser Umstand kann damit erklärt werden, dass das Netz zwischen den beiden höheren Massenpunkten nicht unterscheiden kann und daher tendenziell alles der Klasse des Massenpunktes  $M_\phi = 3000$  GeV zuordnet. Diese Besonderheit tritt auch in allen folgenden Trainings mit drei Signalthypothesen auf. Die hier genannte Interpretation für das Verhalten des Netzes wird dabei auch im Folgenden beibehalten. In Abschnitt 6.3 wird die Zuordnung in eine Klasse bei zu wenig Trennkraft genauer untersucht.

Durch ein weiteres Training mit einem binären Klassifikator wird die Klassifizierung der Signalthypothesen  $M_\phi = 3000$  GeV und  $M_\phi = 2500$  GeV genauer untersucht. Die dabei verwendeten Konfigurationen sind in Abschnitt 4.1 aufgeführt. Es ergibt sich ein ROC-AUC-Wert von 0,52. Damit zeigt sich eindeutig, dass die Kinematik der beiden Datensätze zu ähnlich für eine Klassifizierung ist. Dennoch kann eine falsche Interpretation der Eingabevariablen ausgeschlossen werden, da der ROC-AUC-Wert nicht unter 0,5 liegt. Eine mögliche Erklärung für die Ununterscheidbarkeit der kinematischen Variablen kann in den hohen Mediatormassen der verwendeten Signalthypothesen liegen. Dadurch kann ein Plateau-Effekt entstehen, wodurch der CMS-Detektor nicht in der Lage ist die verschiedenen Hypothesen kinematisch zu unterscheiden. Um diese Annahme zu überprüfen wird das Training im Folgenden mit Signalthypothesen mit niedrigeren Mediatormassen wiederholt.

## 6.2 Weitere Klassifizierungsansätze

Nachdem zuvor die Klassifizierung Signalthypothesen im Bereich hoher Mediatormassen untersucht wurde, wird nun in Abschnitt 6.2.1 die Studie im Bereich niedriger Massen wiederholt. Im Anschluss wird in Abschnitt 6.2.2 die Trennkraft des Netzes für Signal-Datensätze betrachtet, die weiter auseinander liegen. So können Rückschlüsse auf die Leistungsfähigkeit eines Netzes gezogen werden, das beispielsweise mit einem Massenpunkt im Bereich niedriger Massen trainiert wird und dann auf ein Signal am anderen Ende des Energie-Spektrums trifft.

### 6.2.1 Klassifikation von Signalthypothesen im Bereich niederer Mediatormassen

Im Folgenden wird in einem Training mit drei Signalthypothesen mit Mediatormassen in einem Bereich niederer Massen überprüft, ob eine bessere Unterscheidung als zuvor in dem hohen Massen erreicht werden kann. Wie zuvor in Abschnitt 6.1 werden wieder Massenpunkte mit möglichst ähnlichen  $E_T$ -Verteilungen für das Training gewählt. Die Massen der drei Signalthypothesen sind in Tabelle 6.2 aufgelistet.

Eine Betrachtung der Verteilung aller Eingabevariablen zeigt, dass sich hier alle drei Hypothesen sehr ähneln (vgl. Anhang D.2). Dennoch fällt auf, dass sich die beiden ähnlichsten Verteilungen etwas mehr unterscheiden als bei den Eingabevariablen der in Abschnitt 6.1 untersuchten Datensätze. Das könnte die zuvor erwähnte Annahme mit dem Plateau-Effekt bestätigen. Weiterhin gibt es hier im Gegensatz zu den zuvor untersuchten Datensätzen im Bereich hoher Mediatormassen keine Signalthypothese, welche sich in ihrer Kinematik deutlicher von den anderen abhebt. Deswegen ist keine bessere Klassifizierung zu erwarten. Das Training bestätigt diese Annahme, wie die entsprechende Wahrheitsmatrix in Abbildung D.23a im Anhang zeigt. Dabei wird im Training die transversale Masse  $m_T$  wieder mit in die Eingabevariablen aufgenommen.

Analog zu dem Vorgehen in Abschnitt 6.1 wird das Training nun mit zwei Signalthypothesen mit niedrigen Mediatormassen wiederholt, die eine sehr ähnliche Kinematik aufweisen. Die entsprechenden Massenpunkte sind in Tabelle 6.2 aufgeführt. Das Training ergibt einen ROC-AUC-Wert von 0,51. Damit zeigt sich deutlich, dass das Netz die beiden nicht unterscheiden kann, die Zuordnung der beiden Klassen also zufällig ist.

Tabelle 6.2: **Massen der verwendeten Mono-top-Signalthypothesen.** Training im Bereich niedriger Mediatormassen.

Klassifikation	Massen-Bereich	Mediatormasse $M_\phi$ in GeV	DM-Masse $M_\chi$ in GeV	Anzahl der Ereignisse
Multi	niedrig	295	150	$1,49 \cdot 10^5$
		495	250	$1,82 \cdot 10^5$
		200	50	$1,19 \cdot 10^5$
Binär	niedrig	495	250	$1,82 \cdot 10^5$
		500	150	$1,77 \cdot 10^5$

### 6.2.2 Klassifikation von Signalthypothesen mit möglichst unterschiedlicher Kinematik

Um die Bedingungen für eine richtige Unterscheidung zu verbessern wird ein Training mit drei Mediatormassepunkten durchgeführt, deren Mediatormassen sich möglichst deutlich unterscheiden (vgl. Tabelle 6.3). Erwartungsgemäß ähneln sich die Verläufe der beiden Signalthypothesen mit höheren Mediatormassen sehr, aber dennoch etwas weniger, als die bei dem Training im hohen Massenbereich in 6.1. Der Punkt mit geringerer Mediatormasse hebt sich deutlich von den anderen beiden ab. Somit ist zu erwarten, dass die Trennung von dieser funktioniert, das Netz jedoch nicht zwischen den beiden Massenpunkten mit hohen Mediatormassen unterscheiden kann. Diese Annahme bestätigt sich im Training (vgl. Abbildung 6.3a). Somit lässt sich vermuten, dass das Netz die Signalthypothesen unterscheiden kann, insofern sie sich in ihrer Kinematik genug unterscheiden.

Diese Erwartung lässt sich im Training mit zwei Massenpunkten überprüfen, deren Mediatormassen sich deutlich unterscheiden. Die verwendeten Signalthypothesen finden sich in

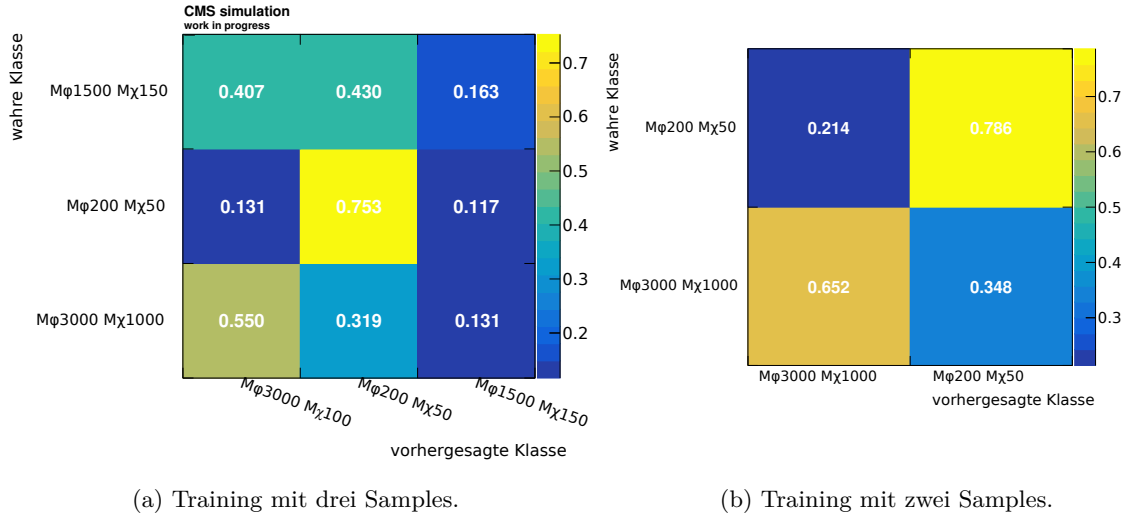


Abbildung 6.3: **Wahrheitsmatrizen.** Training mit Mono-top-Signalthypothesen mit Mediatormassen die sich deutlich unterscheiden mit jeweils 50 Epochen. In (a) werden drei Datensätze verwendet und in (b) zwei. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. Es wurde zusätzlich mit der Eingabevariable transversale Masse  $m_T$  trainiert.

Tabelle 6.3. Wie bereits vermutet ist das Netz in der Lage die beiden Massenpunkte zu klassifizieren. So werden die  $M_\phi = 200$  GeV Ereignisse zu 78,6 % richtig erkannt, während die  $M_\phi = 3000$  GeV Ereignisse mit 65,2 prozentiger Wahrscheinlichkeit richtig zugeordnet werden (vgl. Abbildung 6.3b).

Tabelle 6.3: **Massen der verwendeten Mono-top-Signalthypothesen.** Training mit Massenpunkten, deren Mediatormassen sich möglichst deutlich unterscheiden.

Klassifikation	Massen-Bereich	Mediatormasse $M_\phi$ in GeV	DM-Masse $M_\chi$ in GeV	Anzahl der Ereignisse
Multi	weit auseinander	3000	1000	$2,32 \cdot 10^5$
		1500	150	$2,53 \cdot 10^4$
		200	50	$1,19 \cdot 10^5$
Binär	weit auseinander	3000	1000	$2,32 \cdot 10^5$
		200	50	$1,19 \cdot 10^5$

## 6.3 Untersuchung des Verhaltens bei zu geringer Trennkraft

Wie bereits zuvor erwähnt, zeigt sich im Training mit drei Signalthypothesen, dass das Netz Ereignisse kinematisch zu ähnlicher Klassen einer der beiden zuordnet, wenn es nicht genügend Trennkraft aufbringen kann. Um dem nachzugehen werden verschiedene Aspekte untersucht.

Zunächst wird der Einfluss der Anzahl der in den Datensätzen enthaltenen Ereignisse betrachtet. Wie in Abschnitt 4.1.1 erwähnt, sind die Signalthypothesen so skaliert, dass sie die gleiche physikalische Rate haben. Dabei könnte sich die entsprechenden Datensätze deutlich in der Anzahl der Ereignisse unterscheiden. Es zeigt sich, dass in allen Trainings mit drei Signalthypothesen immer einer Klasse auf der Diagonalen der Wahrheitsmatrix über 64% zugeordnet wird. Letztere scheint sich in den Verteilungen der Eingabevariablen meist deutlicher von den anderen beiden Hypothesen abzuheben. Die anderen beiden

Klassen, die sich besonders ähneln, zeigen auf der Diagonalen jeweils ein Feld mit etwa 50 % und ein Feld mit etwa 0,5-1% . Es fällt auf, dass bei allen Trainings, die mit drei Signalthypothesen durchgeführt werden, bei den ähnlichen Klassen die einprozentige Klasse immer weniger Ereignisse enthält, als die 50-prozentige Klasse. Beispiele dazu sind in Tabelle E.2 im Anhang aufgeführt. Es kann weiterhin festgestellt werden, dass die über 64-prozentige Klasse, also die Klasse die von allen drei am stärksten klassifiziert wird, nicht von der Anzahl der Ereignisse beeinflusst scheint. So wird einerseits beim Training im hohen Mediatormassen-Bereich beispielsweise die  $M_\phi = 750$  GeV-Klasse am stärksten klassifiziert, obwohl sie eine Größenordnung weniger Ereignisse enthält als die anderen Klassen. Dagegen ist andererseits beim Training im niedrigen Massen-Bereich die  $M_\phi = 200$  GeV-Klasse am stärksten repräsentiert, die kaum weniger Ereignisse als die mit den meisten Ereignissen enthält. So kann vermutet werden, dass nur bei Datensätzen mit besonders ähnlicher Kinematik die Zahl der Ereignisse von Bedeutung ist. Bei dieser Situation scheint das Netz, wenn es sich nicht genügend Trennkraft aufbringen kann, die Klasse mit auch nur wenig mehr Ereignissen zu bevorzugen.

Weiterhin kann angemerkt werden, dass die Leistung des Netzes für die beiden ähnlichen Klassen im einzelnen nicht stabil ist. So zeigt sich bei einer 17-fachen Wiederholung des Trainings mit den drei Datensätzen im Bereich hoher Mediatormassen, dass die Werte der Wahrheitsmatrix fluktuieren. Drei der 17 generierten Wahrheitsmatrizen heben sich deutlich von den übrigen ab. So liegen die Werte auf der Diagonalen bei den beiden besonders ähnlichen Klassen bei etwa 20 % für die  $M_\phi = 2500$  GeV-Klasse mit weniger Ereignissen und bei etwa 30 % bei der  $M_\phi = 3000$  GeV-Klasse, in der mehr Ereignisse vorliegen (vgl. Abbildung E.31a). Das steht im Gegensatz zu den häufiger auftreten Werten von etwa einem bzw. etwa 50 %. Dennoch scheint die Summe der beiden beschriebenen Diagonalelementen immer etwa 50 % zu ergeben (vgl. Abbildung E.31b). Somit zeigt sich, dass die Summe aus beiden Elementen sowie das Element der unterschiedlichen  $M_\phi = 750$  GeV-Klasse dennoch relativ stabil ist. Bei den beiden Klassen die nicht unterschieden werden können, scheint die Zuordnung zufällig zu sein.

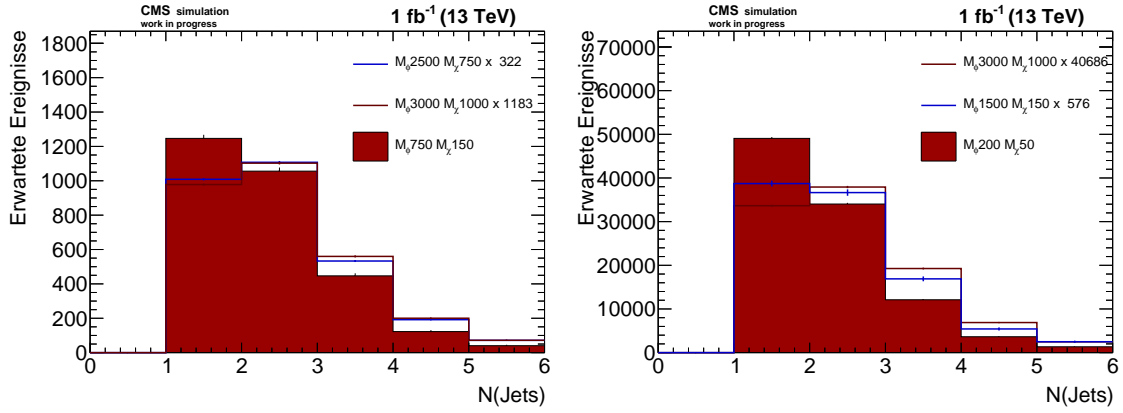
Ein weiterer Erklärungsversuch, warum gerade eines der beiden ähnlichen Datensätzen in der Klassifizierung bevorzugt wird, ist, dass ein Datensatz höher gewichtet wird. Ein Grund für eine solche höhere Gewichtung könnte beispielsweise ein kleinerer Wirkungsquerschnitts sein. Um diese Annahme zu überprüfen wird das Gewicht auf eins gesetzt und das Ergebnis des Trainings mit einem mit normaler Gewichtung verglichen. Bei dem Training mit möglichst weit auseinander liegenden Massenpunkten beispielsweise verschieben sich die Prozentualen-Werte der beiden ähnlichen Klassen um etwa ein Prozent (vgl. Abbildung E.32). Es zeigt sich, dass so kaum eine Veränderung der Klassifizierung erreicht wird.

Abschließend kann festgestellt werden, dass die Anzahl der in einem Datensatz enthaltenen Ereignisse bei sehr geringer Trennkraft die Klassifizierung des Netzes beeinflussen können, während die Gewichtung durch die Wirkungsquerschnitte kaum eine Rolle zu spielen scheint. Weiterhin ist die Leistung des Netzes dabei nicht stabil. Es bedarf weiterer Studien um die genaue Ursache bzw. Mechanismen zu verstehen, die zu diesen Besonderheiten des Netzes führen. Für diese Arbeit ist es jedoch ausreichend festzustellen, dass das Netz, wenn zwei Klassen zu ähnlich sind und daher nicht genügend Trennkraft aufgebracht werden kann, die meisten Ereignisse einer dieser Klasse zuordnet.

## 6.4 Beobachtungen zur Verteilung der Anzahl der Jets

Bei der Betrachtung der Eingabevariablen über den Aspekt der möglichen Trennkraft für das Netz hinaus, fallen die Verteilungen für die Anzahl der Jets  $N_{\text{Jets}}$  bei hohen Mediatormassen auf. Wie zuvor in Abschnitt 3.1 erläutert ist in führender Ordnung für den





(a) Verteilung von  $N_{\text{Jets}}$  im Bereich hoher Mediator-massen. (b) Verteilung von  $N_{\text{Jets}}$  mit möglichst unterschiedlichen Mediator-massen.

Abbildung 6.4: **Verteilungen der Anzahl der Jets.** In (a) für Mediator-massen im hohen Massenbereich und in (b) für Mediator-massen mit möglichst unterschiedlichen Massen.

Mono-top-Prozess ein Jet aus einem b-Quark zu erwarten, wobei das b-Quark aus dem Zerfall des t-Quarks stammt.

Für die  $N_{\text{Jets}}$ -Verteilung der in Abschnitt 6.1 betrachteten Massenkpunkte im Bereich hoher Mediator-massen zeigt sich eine Verschiebung für die beiden höheren Massenkpunkte  $M_\phi = 2500$  GeV und  $M_\phi = 3000$  GeV (vgl. Abbildung 6.4a). So enthalten die meisten Ereignisse dieser Verteilungen nicht wie beim Massenkpunkt  $M_\phi = 750$  GeV, einen Jet, sondern zwei Jets. Auch ist der relative Anteil der Ereignisse mit drei und vier Jets in den Verteilungen der höheren Mediator-massen deutlich höher.

In der  $N_{\text{Jets}}$ -Verteilung der in Abschnitt 6.2.2 betrachteten Massenkpunkte, deren Mediator-massen sich möglichst deutlich unterscheiden, zeigt sich, dass bei dem Massenkpunkt  $M_\phi = 200$  GeV mit einer niedrigen Mediator-masse deutlich mehr Ereignisse einen Jet enthalten als zwei oder mehr Jets (vgl. Abbildung 6.4b). Bei dem Massenkpunkt  $M_\phi = 1500$  GeV dagegen haben etwa gleich viele Ereignisse einen bzw. zwei Jets, wobei jedoch wie beim Massenkpunkt  $M_\phi = 1500$  GeV, bei dem es mehr Ereignisse mit einem Jet gibt, deutlich mehr Ereignisse bei drei und mehr Jets vorliegen, als bei dem Massenkpunkt  $M_\phi = 200$  GeV.

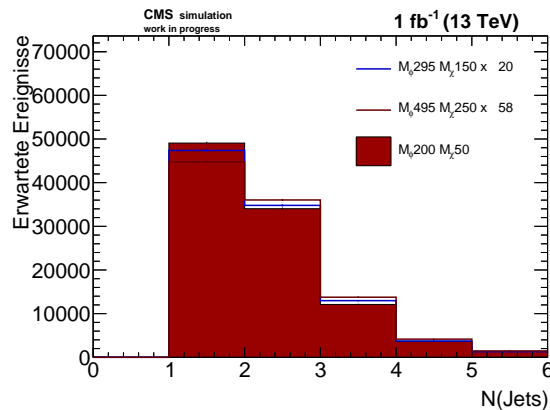


Abbildung 6.5: **Verteilungen der Anzahl der Jets.** Für Mediator-massen im niedrigen Massenbereich.

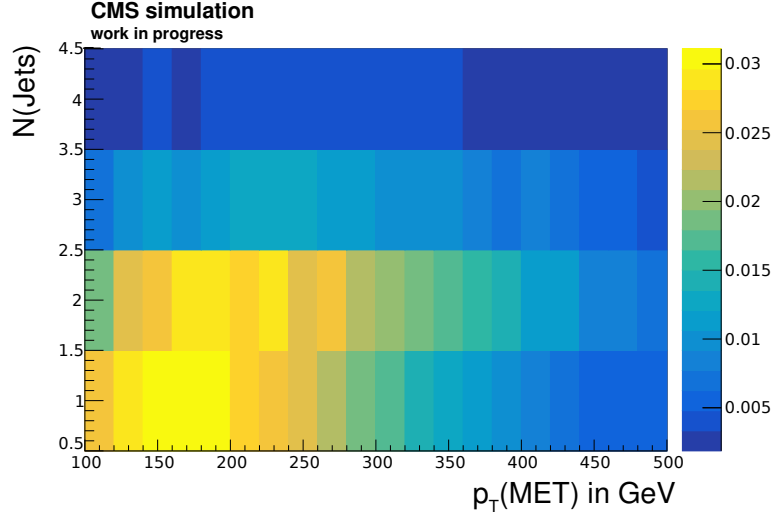


Abbildung 6.6: **Zweidimensionaler Plot.** Die Verteilung der fehlenden Transversalenergie  $\cancel{E}_T$  ist gegen die Anzahl der Jets  $N_{\text{Jets}}$  aufgetragen. Die betrachtete Mono-top-Signalthypothese hat dabei die Massen  $M_\phi = 3000$  GeV und  $M_\chi = 1000$  GeV. Bereitgestellt durch Michael Waßmer.

Auch bei anderen, in Abschnitt 6.2.1 untersuchten Massenpunkten, die im Bereich niedriger Mediatormassen liegen, zeigt sich in den  $N_{\text{Jets}}$ -Verteilungen, die alle am meisten Ereignisse mit einem Jet haben, dass die relative Anzahl der Ereignisse, die nicht mehr als ein Jet enthalten, mit geringer Mediatormasse größer wird (vgl. Abbildung 6.5).

Nach diesen Beobachtungen kann angenommen werden, dass die relative Anzahl der Ereignisse die einen Jet enthalten mit steigender Mediatormasse abnimmt, während die relative Anzahl der Ereignisse die zwei oder mehr Jets enthalten zunimmt. Der Wendepunkt scheint dabei bei einer Mediatormasse von etwa  $M_\phi = 1500$  GeV zu sein.

### Erklärungsansatz

Eine mögliche Erklärung für diese Verschiebung in den  $N_{\text{Jets}}$ -Verteilungen für hohe Mediatormassen findet sich in der Betrachtung des Mono-top-Prozesses. Wie bereits zuvor erwähnt ist in führender Ordnung ein Jet mit hohem Transversalimpuls zu erwarten, der aus einem b-Quark stammt. Dieses b-Quark ist dabei neben einem Lepton und einem Neutrino ein Zerfallsprodukt des t-Quarks, welches zusammen mit den beiden Dunklen-Materie-Teilchen entsteht. Das t-Quark erfährt dabei einen Rückstoß gegen die beiden Dunkle-Materie-Teilchen und erhält dadurch einen Lorentzboost. Die aus dem t-Zerfall entstehenden Teilchen werden kollimiert.

Bei hohen Mediatormassen tragen die aus dem Mediator entstehenden Dunkle-Materie-Teilchen einen höheren Impuls, was nach dem Rückstoß auch mit einem höheren Impuls des leichteren t-Quarks einhergeht. Je stärker der Impuls des t-Quarks ist, desto höher ist dessen Boost und desto mehr verringert sich der Abstand der Zerfallsprodukte b-Quark und Lepton. Dieser Abstand  $\Delta R$  ist jedoch ein Kriterium für die in der Analyse berücksichtigten Jets. Wenn sich der Abstand zwischen Lepton und Jet aus dem b-Quark durch den großen Boost bei hohen Mediatormassen zu sehr verringert und kleiner als  $\Delta R=0,4$  wird, sich Jet und b-Quark also überlappen, so fällt der Jet aus der Selektion und wird nicht mehr berücksichtigt. Somit könnte es zu einer Unterdrückung der Jets erster Ordnung bei hohen Mediatormassen kommen. Die Annahme das Transversalimpuls der Dunkle-Materie-Teilchen, also die fehlende Transversalenergie  $\cancel{E}_T$ , mit der Anzahl der Jets korreliert, lässt

sich durch die Betrachtung des 2D-Plots von  $N_{\text{Jets}}$  und  $\cancel{E}_T$  bekräftigen (vgl. Abbildung 6.6). Mit einer höheren Transversalenergie  $\cancel{E}_T$  nimmt auch der Beitrag zu einer Jetanzahl von zwei deutlich zu.

Darüber hinaus lässt sich der Zusammenhang zwischen kleinerem  $\Delta R$  zwischen Lepton und führendem Jet bei größerer fehlender Transversalenergie mit dem 2D-Plot von  $\cancel{E}_T$  und  $\Delta R$  belegen (vgl. Abbildung 6.7). Bei dem Lepton handelt es sich dabei entweder um ein Elektron oder Myon.

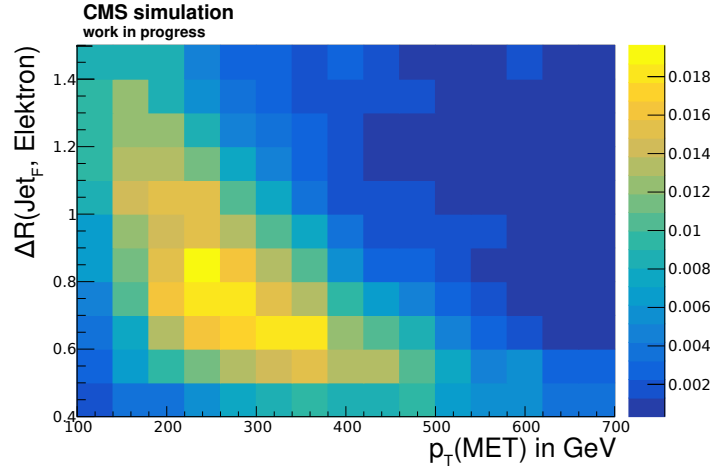


Abbildung 6.7: **Zweidimensionaler Plot.** Die Verteilung der fehlenden Transversalenergie  $\cancel{E}_T$  ist gegen die des Abstands  $\Delta R$  zwischen Elektron und führendem Jet aufgetragen. Der Plot für die Verteilung des Abstandes  $\Delta R$  zwischen Myon und Jet findet sich in Abbildung F.33 im Anhang. Die betrachtete Mono-top-Signalthypothese hat dabei die Massen  $M_\phi = 3000$  GeV und  $M_\chi = 1000$  GeV. Bereitgestellt durch Michael Waßmer.

## 6.5 Schlussfolgerungen

Im Gegensatz zu den Erwartungen wird klar, dass das Netz trotz seiner Sensitivität auf die Variable der fehlenden transversalen Energie  $\cancel{E}_T$  nicht auf die Eigenschaften der involvierten Massenpunkte fokussiert zu sein scheint. Es zeigt sich, dass das Netz nur eine Trennkraft aufbringen kann, wenn die Massen der Signalthypothesen weit auseinander liegen und sich somit kinematisch deutlich unterscheiden. Dieser Umstand kann positiv bewertet werden, da so ein Netz vermutlich mit einem beliebigen Massenpunkt trainiert werden und viele andere Massenpunkte damit erkannt werden können. Damit kann bei der Klassifizierung ein konsistenteres Ergebnis erreicht werden.

Es ist anzumerken, dass trotz einer Aufnahme der transversalen Masse  $m_T$  in die Eingabevariablen keine deutliche Verbesserung erreicht werden kann. Daher ist es von Interesse zu untersuchen ob diese Variable auch bei der Trennung eines Mono-top-Signals vom Untergrund eine bessere Leistungsfähigkeit erreichen kann. Daher wird im nächsten Kapitel in einem quantitativen Vergleich zwischen der klassischen Methode und einem neuronalen Netz auch die Verwendung der transversalen Masse als Eingabevariable mit in Betrachtung genommen.

Darüber hinaus kann neben den Analysen zur Trennkraft des Netzes für unterschiedliche Signalthypothesen ein Einfluss der fehlenden Transversalenergie  $\cancel{E}_T$  auf die Anzahl der gut rekonstruierbaren Jets beobachtet werden. Um diesen Sachverhalt zu erklären, bietet es sich an, weiter Studien dazu anzusetzen.



## 7 Vergleich der klassischen Methode mit der Klassifizierung durch neuronale Netze

In Kapitel 5 konnte gezeigt werden, dass ein neuronales Netz gut dazu geeignet ist, Signal und Untergründe nach einer Mono-top-Hypothese zu klassifizieren. Im folgenden soll nun die Leistung der zuvor in Kapitel 3.3 vorgestellten klassischen Methode quantitativ mit der der Klassifizierung durch ein neuronales Netz verglichen werden.

Die dafür gewählten Kriterien wurden zuvor in Abschnitt 4.2 eingeführt. Im Folgenden werden in Abschnitt 7.2 die Diskriminatoren betrachtet. In Abschnitt 7.3 erfolgt ein Vergleich der Reinheiten. Abschließend wird in Abschnitt 7.4 auf die ROC-Kurven eingegangen.

Für diesen Vergleich wird mit dem in Kapitel 5 verwendeten Massenpunkt mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV gearbeitet. Da dieser im Bereich hoher Mediator-Massen liegt, resultieren daraus sehr kleine Wirkungsquerschnitte. Damit sind recht kleine Werte für die im Folgenden betrachteten Diskriminatoren und Reinheiten zu erwarten. Für die Diskriminatoren müssen sie daher für eine bessere Darstellung skaliert werden.

Da die klassische Methode auf einem Schnitt der transversalen Masse  $m_T$  beruht, wurde bei den Versuchen in Kapitel 5 diese Variable nicht verwendet, um eine deutlichere Differenzierung der Methoden zu ermöglichen. So konnte festgestellt werden, dass die Klassifizierung auch ohne  $m_T$  erfolgreich ist. Nun werden nicht nur die beiden Methoden verglichen sondern auch untersucht, wie groß der Einfluss der transversalen Masse tatsächlich ist. Daher wird in Abschnitt 7.1 die Leistungsfähigkeit des Netzes mit und ohne  $m_T$  anhand der schon in Kapitel 5 verwendeten Kriterien untersucht.

### 7.1 Training mit der transversalen Masse $m_T$

Wie bereits erwähnt wird neben dem eigentlichen Vergleich der klassischen und der auf dem Netz basierenden Methode auch der Einfluss der transversalen Masse  $m_T$  auf das Netz untersucht. Dabei wird zunächst in Abbildung 7.1 die Verteilung von  $m_T$  betrachtet, um einen ersten Eindruck ihrer Trennkraft zu erhalten. Wie nach Abschnitt 3.1 zu erwarten, tendiert die Verteilung der Untergrundprozesse  $W$ +Jets und  $t\bar{t}$  zu niedrigeren transversalen Massen  $m_T$  und fällt dann ab etwa  $m_T = 100$  GeV stark ab, während der Mono-top Prozess erst bei etwa 200 GeV sein Maximum erreicht. Daher ist von dieser Eingabevariable, auch

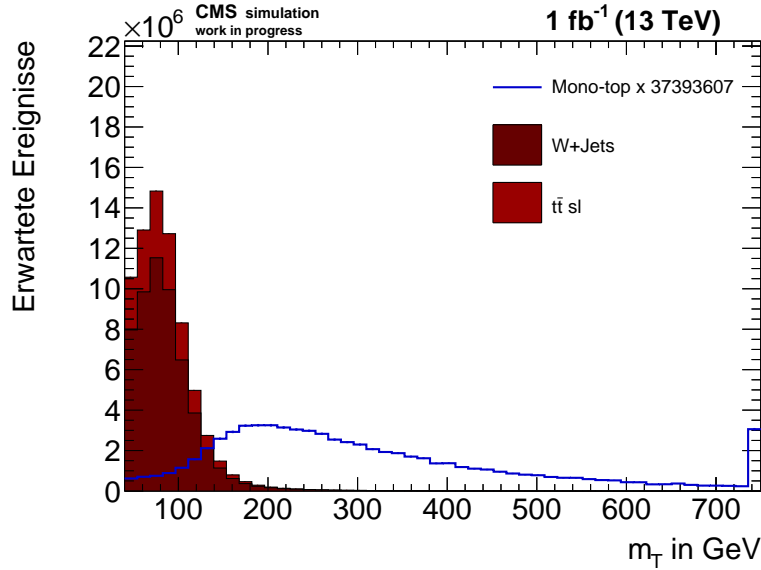


Abbildung 7.1: Verteilung der Eingabevariable transversale Masse  $m_T$ .

erkennbar anhand ihrer Verteilung, eine gute Trennkraft zu erwarten. Die Relevanz von  $m_T$  für die Klassifizierung wird anhand eines Rankings der Eingabevariablen untersucht. Im Anschluss werden darüber hinaus zwei Wahrheitsmatrizen aus einem Training mit und ohne  $m_T$  verglichen.

Das Ranking der Eingabevariablen (vgl. Abbildung 7.2) bestätigt diese Hypothese. Analog zum Vorgehen in Abschnitt 5.1 wird aus 16 Trainingsläufen mit jeweils 25 Epochen der relative Mittelwert der Gewichte der Variablen der ersten Schicht bestimmt. Wie zu erwarten steht dabei die transversale Masse an erster Stelle.

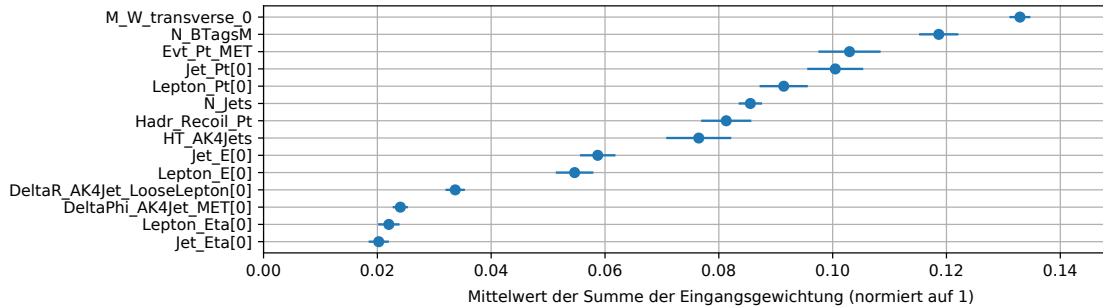


Abbildung 7.2: **Ranking der Eingabevariablen.** Zu sehen ist eine Liste der Eingabevariablen, die nach der Summe der Gewichte der ersten Schicht geordnet ist. Im Gegensatz zum Ranking in Abbildung C.8 wurde hier für die Trainings die transversale Masse  $m_T$  mit berücksichtigt. In Tabelle 4.4 finden sich eine Beschreibung der hier aufgeführten Variablen.

Bei der Betrachtung der Wahrheitsmatrizen aus den Trainings mit und ohne  $m_T$  zeigt sich jedoch, dass sich die Klassifizierung des Netzes kaum verbessert (vgl. Abbildung 7.3). So zeigt sich für die Mono-top-Klassifizierung nur eine Verbesserung von etwa 1,3 %. Daher ist anzunehmen, dass die Trennkraft, die zuvor vor allem durch die fehlende transversale Energie  $\cancel{E}_T$  und den hadronischen Rückstoß aufgebracht wurde, nun durch  $m_T$  aufgebracht

wird. Da die Größen korrelieren, macht die explizite Nutzung von  $m_T$  keine signifikanten Unterschiede mehr.

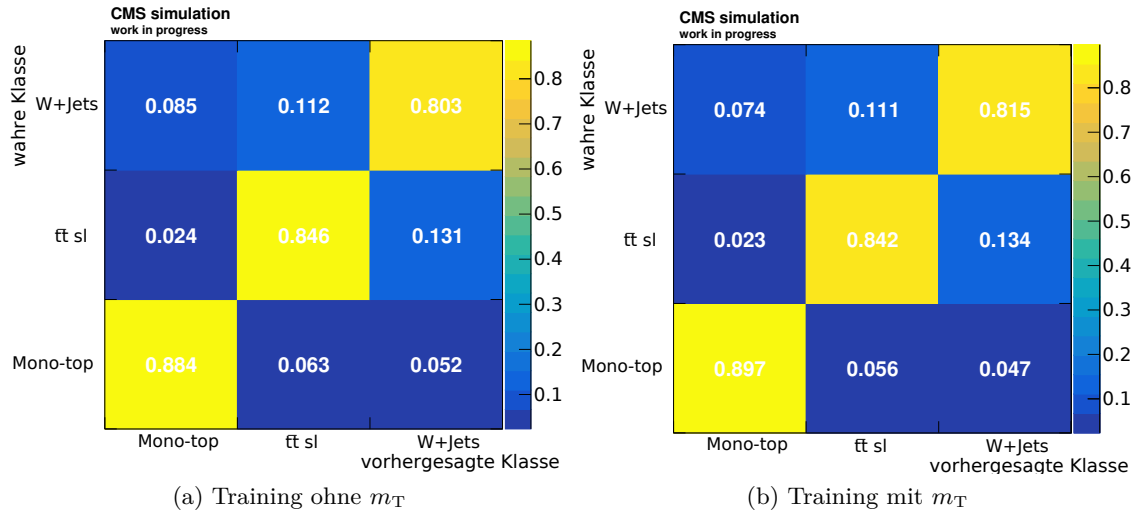


Abbildung 7.3: **Wahrheitsmatrizen im Vergleich.** Training mit jeweils 25 Epochen mit einem Mono-top-Signal und den W+Jets- und  $t\bar{t}$ -Untergründen. In (a) wurde im Gegensatz zu (b) die transversale Masse nicht als Eingabevariable berücksichtigt.

Beim folgenden Vergleich der Diskriminatoren, Reinheiten und ROC-Kurven ist jeweils nur eine minimale Verbesserung für das Miteinbeziehen der transversalen Masse erkennbar. Daher wird die Idee jeweils die Klassifikation der Netze mit und ohne der Verwendung von  $m_T$  der klassischen Methode gegenüberzustellen verworfen. Es wird bei dem Vergleich stattdessen nur auf die Größen aus dem Netz eingegangen, die unter Verwendung der transversalen Masse berechnet worden sind.

## 7.2 Diskriminatoren

Zunächst sollen die verschiedenen Klassifizierungsmethoden anhand ihrer Diskriminatoren bewertet werden. Dazu werden in Abschnitt 7.2.1 zunächst die Diskriminatoren des Netzes und im Anschluss in Abschnitt 7.2.2 die der klassischen Methode betrachtet.

### 7.2.1 Diskriminatoren des neuronalen Netzes

In Abschnitt 4.2 wird beschrieben, wie die Diskriminatoren des Netzes erzeugt werden. Sie ergeben sich aus einem Training mit 25 Epochen und den in Abschnitt 4.1 beschriebenen Eingabevariablen und Konfigurationen. Dabei wird einmal mit und einmal ohne die transversale Masse trainiert. Da sich die Diskriminatoren kaum unterscheiden, wird hier im Folgenden nur auf die aus dem Training mit der transversalen Masse  $m_T$  eingegangen. Letztere sind in Abbildung 7.4 einmal mit linearer und einmal mit logarithmischer Skala abgebildet.

Die Betrachtung des Mono-top-Diskriminators zeigt, dass Mono-top-Ereignisse für einen Diskriminatorwert von eins stark angereichert sind. Auch W+Jets-Ereignisse zeigen hier eine erhöhte Anreicherung. Ein Abgleich mit den logarithmisch-skalierten Diskriminatoren zeigt, dass das Netz jedoch einen höheren Anteil von Mono-top-Ereignissen als W+Jets-Ereignissen anreichert. Dabei muss jedoch der Wirkungsquerschnitt der Prozesse beachtet

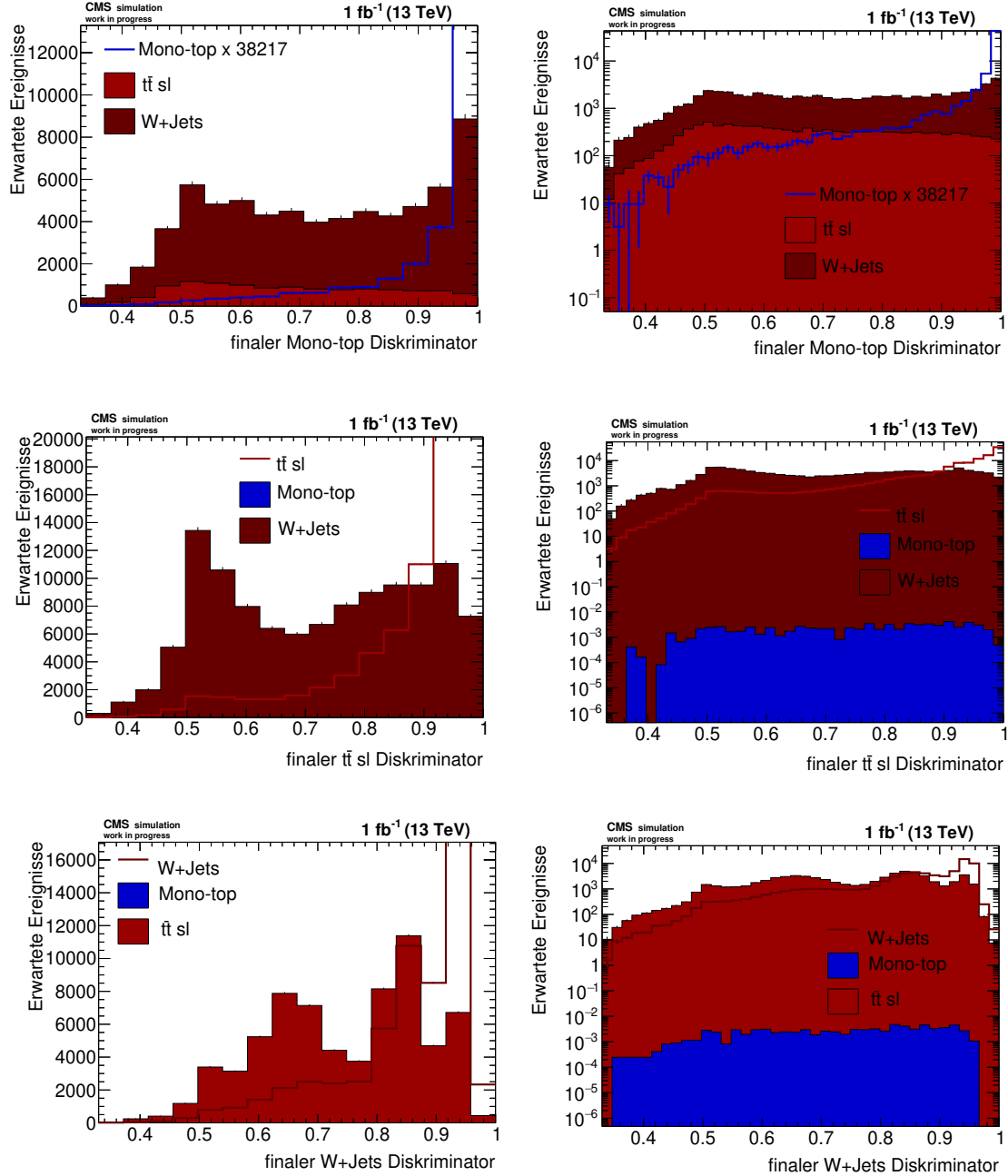


Abbildung 7.4: **Diskriminatoren des neuronalen Netzes.** Aus einem Training mit der Mono-top-Signalhypothese mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV über 25 Epochen. Auf der linken Seite finden sich die Diskriminatoren mit linearer Skalierung und rechts mit logarithmischer. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. Es wurde zusätzlich mit der Eingabevariable transversale Masse  $m_T$  trainiert.



werden. Da der Wirkungsquerschnitt des Mono-top-Prozesses sehr viel kleiner ist als der des W+Jets-Prozesses ist der Mono-top-Prozess trotz des höheren relativen Anteils sehr schwer zu messen. Das äußert sich in großen Unsicherheiten und geringer Sensitivität auf das Mono-top-Signal. Weiterhin kann festgestellt werden, dass die meisten W+Jets-Ereignisse für höhere Diskriminatorwerte angereichert sind. Die  $t\bar{t}$ -Ereignisse sind über die gesamte Verteilung gleichmäßig verteilt und ergeben so einen konstanten Untergrund.

Insgesamt zeigt der Mono-top-Diskriminator, dass das Netz seine Aufgabe entsprechend gut erfüllt, da auf der rechten Seite deutlich mehr Signal als Untergrund angereichert ist. Selbiges ergibt auch die Betrachtung der beiden Untergrund-Diskriminatoren. Bei  $t\bar{t}$  und W+Jets findet sich ebenfalls jeweils eine starke Anreicherung im Bereich des maximalen Diskriminatorwerts eins. Somit zeigt das Netz nicht nur gute Voraussetzungen um Mono-top-Ereignisse richtig zu klassifizieren, sondern auch die für eine richtige Erkennung der Untergrund-Ereignisse W+Jets und  $t\bar{t}$ .

### 7.2.2 Diskriminatoren der klassischen Methode

Wie in Abschnitt 4.2 beschrieben dient als Diskriminator der klassischen Methode die transversale Masse  $m_T$ . Wie zuvor beschrieben werden nach der Vorselektion (vgl. Abschnitt 4.1) aus dem Mono-top-,  $t\bar{t}$ -Datensatz und aus den W+Jets-Datensätzen jeweils die Ereignisse nach der Anzahl Jets mit b-Tag selektiert. Somit ergeben jeweils drei Histogramme für die drei Prozesse, für welche genau ein, zwei oder mehrere oder kein mit b-Tag getaggtten Jets gefordert worden sind. Die Untergründe sind kumulativ dargestellt.

Für den Mono-top-Diskriminator werden alle Histogramme mit den Ereignissen für die genau ein b-Tag verlangt wurde in einem Plot dargestellt. Der verwendete Mono-top-Datensatz liegt im Bereich hoher Mediator-Massen. Daraus resultiert ein geringer Wirkungsquerschnitt weswegen die Anzahl der Ereignisse mit der Summe aller Untergrund-Prozesse zur besseren Darstellung skaliert wird.

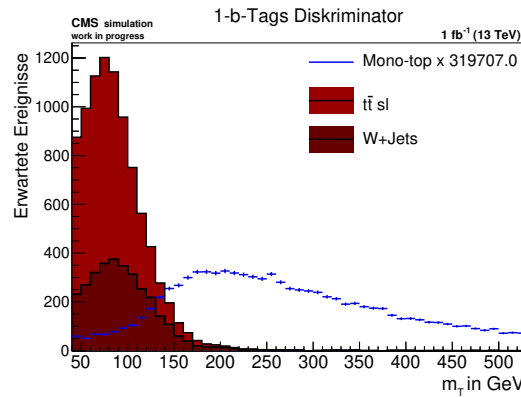


Abbildung 7.5: **Mono-top-Diskriminator der klassischen Methode.** Für die Mono-top-Signal-Hypothese mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV mit den Untergrundprozessen W+Jets und  $t\bar{t}$ .

Um den  $t\bar{t}$ -Diskriminator zu erhalten werden entsprechend alle Histogramme für Ereignisse mit Jets mit genau zwei oder mehreren b-Tags in einem Plot dargestellt, für den W+Jets-Diskriminator alle Histogramme für Ereignisse mit Jets mit keinem b-Tag. Somit ergeben sich die in Abbildung 7.6 dargestellten Diskriminatoren.

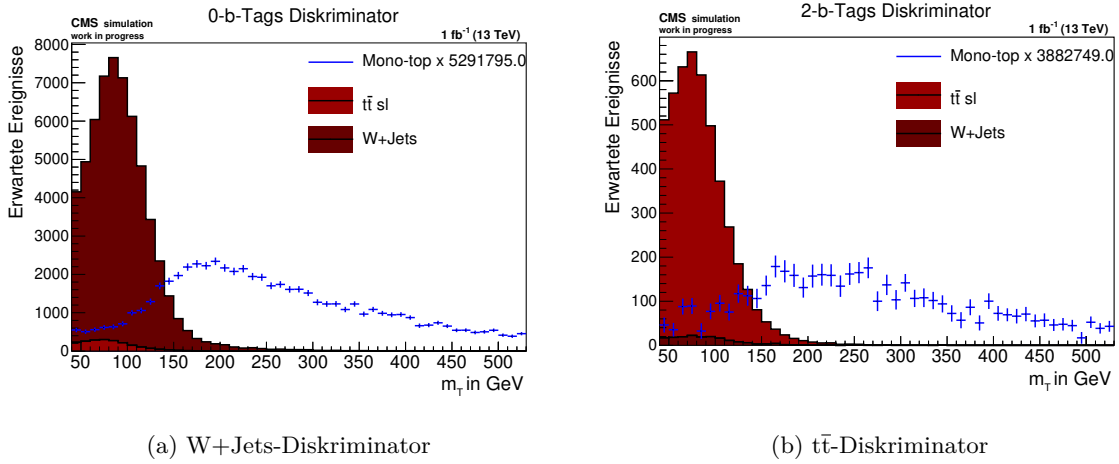


Abbildung 7.6: **Untergrund-Diskriminatoren der klassischen Methode.** Für die Mono-top-Signal-Hypothese mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV mit den Untergrundprozessen W+Jets und  $t\bar{t}$ .

Die Diskriminatoren zeigen deutlich die kinematischen Unterschiede der Untergründe und des Mono-top-Prozesses. Diese wurden zuvor in Kapitel 3 motiviert. So ist das Signal hauptsächlich zu hohen transversalen Massen angereichert, während die Untergründe nur im Bereich niedrigerer  $m_T$  angereichert sind und dann stark abfallen. Weiterhin zeigt sich deutlich, dass ein Schnitt bei einer bestimmten transversalen Masse zwar dazu geeignet ist, das Signal von den Untergründen zu trennen, aber sich nicht ohne weiteres für die Klassifizierung der Untergründe untereinander eignet. Das zeigt sich daran, dass die Verteilung bei allen Diskriminatoren gleich ist, unabhängig von der Anzahl der getaggtten Jets. Es ändert sich nur die Komposition der ausgewählten Ereignisse. Weiterhin verliert man bei einem Schnitt zur Anreicherung des Mono-top-Signals viel des Untergrundverständnisses.

### 7.3 Reinheiten

Eine gute Möglichkeit, die beiden Klassifizierungsmethoden zu vergleichen, findet sich in der sogenannten *Purity*. Wie bereits in Abschnitt 4.2.4 erwähnt gibt diese Größe einen Eindruck von der Reinheit der Kategorien in denen sie angereichert sind.

Zur Berechnung der Reinheiten der klassischen Methode werden die zuvor in Abschnitt 7.2.2 beschriebenen Histogramme verwendet. Für die Reinheit in der Mono-top-Klassifizierung in der Mono-top-Region wird die Anzahl der Ereignisse im Histogramme für Ereignisse mit genau einem Jet mit b-Tag des Mono-top-Datensatzes durch die Summe der Ereignisse der drei Histogramme für Ereignisse mit genau einem Jet mit b-Tag aus den drei Prozessen geteilt. Die Berechnung der Reinheiten für das Netz erfolgt analog. Alle Werte finden sich in Tabelle 7.1, dem jeweiligen klassischen Pendant gegenübergestellt.

Es zeigt sich, dass der Wert der Reinheiten für die Anreicherung der Mono-top-Ereignisse in der Mono-top-Region für die Klassifizierung mit dem Netz um eine ganze Größenordnung besser ist als mit der klassischen Methode. Interessanterweise scheint die Klassifizierung der Untergrund-Prozesse mit der klassischen Methode gleich gut oder sogar besser zu funktionieren, als mit dem Netz. Das ist eine interessante Beobachtung, da angenommen wurde, dass das Netz die Klassifizierung der Untergründe besser bewältigen sollte, da es die Klassifizierung ohne Schnitte durchführt und daher weniger Informationen verliert. Es wäre interessant, den hier angebrachten Vergleich für Signalthypothesen in anderen

Massenbereichen zu wiederholen und außerdem die Klassifizierung der Untergründe weiter zu untersuchen.

Tabelle 7.1: **Reinheiten der klassischen und DNN-Methode im Vergleich.** Die kleine Größenordnung der Werte in der Mono-top-Region lassen sich auf den kleinen Wirkungsquerschnitt des  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV Massenpunktes zurückführen.

Region		Klasse	Reinheit	
Klassisch	DNN		Klassisch	DNN
1 bTag	Mono-top	Mono-top	$3.128 \cdot 10^{-6}$	$2,226 \cdot 10^{-5}$
0 bTags	W+Jets	W+Jets	0,965	0,958
$\geq 2$ bTags	$t\bar{t}$	$t\bar{t}$	0,962	0.79

## 7.4 ROC-Kurven

Zuletzt werden die zuvor in Abschnitt 4.2.5 eingeführten ROC-Kurven der beiden Methoden betrachtet. Dabei wird die Falsch-Positiv-Rate gegen die Richtig-Positiv-Rate aufgetragen. Im Folgenden wird bei der klassischen Klassifizierungsmethode nur auf die Mono-top ROC-Kurve eingegangen, da die transversale Masse  $m_T$  wie bereits erwähnt dafür geeignet ist Mono-top-Signal von den beiden hier betrachteten Untergründen zu trennen, jedoch nicht dazu die Untergrund-Prozesse voneinander zu trennen.

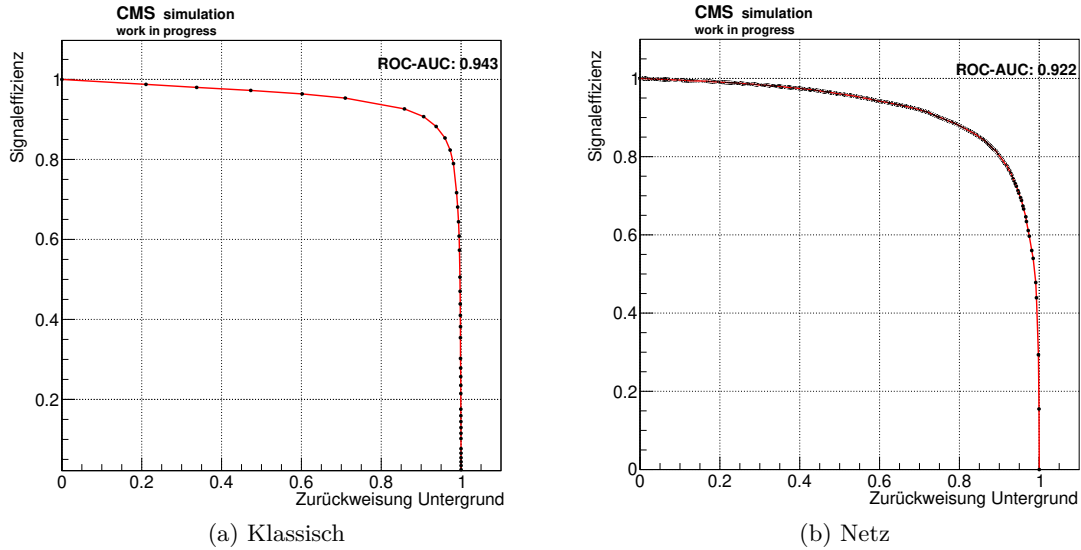


Abbildung 7.7: **Mono-top-ROC-Kurven beider Methoden im Vergleich.** Für die Mono-top-Signalthypothese mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV. In (a) ist die ROC-Kurve der klassischen Methode zu sehen, in (b) die des DNN.

Die ROC-Kurven für Mono-top als Signal beider Klassifizierungsmethoden sind in Abbildung 7.7 gezeigt. Der AUC-Wert und der Verlauf der ROC-Kurve ist bei der klassischen Methode etwas besser zu bewerten. In beiden Fällen liegen die Kurven deutlich über der 0,5-Geraden, was eine gut funktionierende Klassifizierung kennzeichnet.

Die AUC-Werte der ROC-Kurven von W+Jets und  $t\bar{t}$  des Netzes liegen ebenfalls jeweils über dem 0,5-Wert, aber deutlich weniger nah an dem Punkt perfekter Klassifikation als die

beiden zuvor vorgestellten Mono-top-ROC-Kurven (vgl. Abbildung 7.8). Das gilt vor allem für die Kurve von  $t\bar{t}$ . Somit kann anhand der ROC-Kurven abgelesen werden, dass die Klassifikation der Signalthypothese besser zu funktionieren scheint als die der Untergründe. Diese Beobachtung wurde im letzten Abschnitt bereits durch die Reinheiten bestätigt.

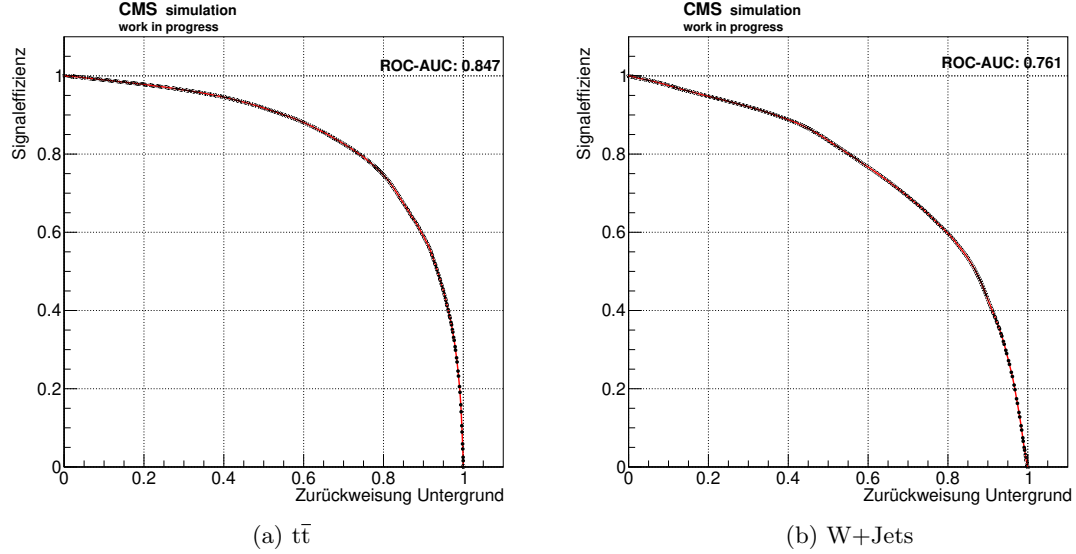


Abbildung 7.8: **Untergrund-ROC-Kurven des Netzes.** Für die Mono-top-Signal-Hypothese mit  $M_\phi = 2495$  GeV und  $M_\chi = 1250$  GeV. In (a) ist die ROC-Kurve von  $t\bar{t}$  abgebildet, in (b) die von W+Jets.

## 8 Zusammenfassung

Die Suche nach Dunkler Materie gestaltet sich herausfordernd, da bisher nur wenig über ihre Eigenschaften bekannt ist. Hinzu kommt, dass sie neben der Gravitation wenn überhaupt nur sehr schwach mit der bisher bekannten Materie des Standardmodells wechselwirkt und daher mit den Detektorsystemen der Teilchenphysik nicht direkt detektierbar ist. Ein möglicher Ansatz, Dunkle-Materie-Teilchen dennoch nachzuweisen, ist die mögliche Produktion solcher Teilchen durch Proton-Proton-Kollisionen am LHC, dem derzeit größten und hochenergetischsten Teilchenbeschleuniger der Welt, und dem anschließenden Nachweis durch spezielle Signaturen mit einem großen Betrag fehlender Transversalenergie z.B. im CMS-Detektor. Dabei ist jedoch ein tiefgreifendes Verständnis der wesentlich dominanteren, Standardmodell-Untergründe, die eine sehr ähnliche Signatur aufweisen, unerlässlich.

Diese Arbeit beschäftigt sich mit der auf der leptonischen Mono-top-Signatur basierten Suche nach Dunkler Materie am CMS-Detektor. Diese Signatur setzt sich aus den Produkten eines leptonischen Top-Quark-Zerfalls und hoher transversaler Masse zusammen. Dabei wurde an simulierten Datensätzen untersucht, inwieweit sich neuronale Netze zur Klassifizierung von Mono-top-Ereignissen und den wichtigsten Untergründen, den  $W$ +Jets- und  $t\bar{t}$ -Ereignissen, eignen. Für diese Studie standen neben den Simulationsdaten der Untergründe auch eine Vielzahl verschiedener Signalthypothesen mit variierenden Mediator- und Dunkle-Materie-Teilchen-Massen zur Verfügung. Der Bereich dieser Massen lag zwischen 50 und 3000 GeV.

Dazu wurde zunächst mit einem einfachen Ansatz ein neuronales Netz mit den Datensätzen einer Mono-top-Signalthypothese und denen der entsprechenden Untergründe implementiert. In mehreren Trainings zeigte sich in den entsprechenden Wahrheitsmatrizen, dass in den drei betrachteten Klassen Mono-top,  $W$ +Jets und  $t\bar{t}$  stets über 80% der Ereignisse richtig klassifiziert wurden. Damit bestätigt sich, dass das Netz genügend Trennkraft für eine erfolgreiche Klassifizierung aufbringen kann.

Weiterhin konnte festgestellt werden, dass das neuronale Netz eine starke Sensitivität auf die fehlende Transversalenergie zeigt, welche die Variable mit der stärksten Trennkraft war. Da die fehlende Transversalenergie in enger Korrelation zu der Masse des Dunkle-Materie-Teilchens steht, wurde im Anschluss in einer Studie untersucht, in wie weit sich dieser Umstand auf die Sensitivität bezüglich der speziellen Mediatormasse der betrachteten Signalthypothese auswirkt. Dazu wurde das Netz mit unterschiedlichen Signalthypothesen

ohne die Untergrund-Datensätze trainiert. Dabei zeigte sich, dass das Netz im Training mit den meisten Hypothesen kaum Trennkraft aufbringen kann. Nur im Training mit zwei Signalthypothesen, deren Massen sich im Rahmen der verfügbaren Datensätze möglichst deutlich unterschieden, war deren Kinematik unterschiedlich genug für eine Differenzierung durch das Netz. Somit konnte festgestellt werden, dass sich das Netz im Gegensatz zu den Erwartungen nicht auf die Eigenschaften der involvierten Massenpunkte festzulegen scheint. Das ist positiv zu werten, da ein mit einem bestimmten Massenpunkt trainiertes Netz viele Signalthypothesen mit Mediatoren in anderen Masse-Bereichen vom Untergrund trennen kann. Somit kann bei der Klassifizierung ein konsistenteres Ergebnis erreicht werden.

Darüber hinaus konnte weiterhin ein Zusammenhang zwischen der fehlenden Transversalenergie, bzw. der Mediatormasse einer Signal-Hypothese und der Anzahl der in den Ereignissen enthaltenen Jets festgestellt werden. So zeigte sich in den Verteilungen der Anzahl der Jets bei hohen Mediatormassen mehr Ereignisse mit zwei, anstatt den erwarteten ein Jets. Die Ursache für diese Abweichung von den Erwartungen sollte mit weiteren Studien untersucht werden, da sich dieser Umstand auf die Signalsensitivität auswirken kann. Dies gilt vorallem auch für die klassische Methode, bei deren Ereignisselektion für den Mono-top Prozess genau ein Jet mit hohem Transversalimpuls gefordert wird. Dieser Jet könnte bei einer Überlappung von Lepton und Jet durch einen großen Boost unterdrückt werden.

Zuletzt wurde die Leistungsfähigkeit der Mono-top-Klassifizierung quantitativ mit einer klassischen Analyse als Benchmark verglichen. Für letztere ist die transversale Masse ein wichtiger Diskriminator. Es zeigte sich, dass die Reinheit der Anreicherung der Mono-top-Ereignisse in der Mono-top-Region mit dem Netz um eine Größenordnung besser ist als mit der klassischen Methode. In allen anderen betrachteten Kategorien zeigte sich die Leistung des Netzes etwa gleich stark oder etwas schlechter.

Darüber hinaus wurde der Einfluss der transversalen Masse untersucht, dem wichtigsten Diskriminator der klassischen Methode. Dabei zeigte sich, dass eine zusätzliche Verwendung dieser Variable im Training keine signifikante Verbesserung bringt. Das lässt sich darauf zurückführen, dass die aus verschiedenen anderen Variablen zusammengesetzte transversale Masse stark mit anderen schon verwendeten Größen korreliert ist. Damit äußert sich ein Vorteil der Klassifizierung durch tiefe neuronale Netze, welche die Informationen aus verschiedenen Variablen selbstständig vereinigt nutzen können.

Zusammenfassend lässt sich sagen, dass sich der Ansatz der Klassifikation von Mono-top-Signalthypothesen und deren Untergrundprozessen als vielversprechend erweist und somit Potential für Folgestudien zur weiteren Ausarbeitung dieser Methode zeigt. Weiterhin bietet es sich an, die quantitative Einschätzung der Trennkraft des Netzes auch mit Signalthypothesen in anderen Massenpunkten zu wiederholen, da sich die Leistungsfähigkeit gegebenenfalls ändern könnte. Dabei sind jedoch keine deutlichen Unterschiede zu erwarten, da der Signalprozess im Wesentlichen immer dieselben kinematischen Eigenschaften zeigt.

# Literatur

- [1] V. C. Rubin und J. Ford W. Kent. „Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions“. In: *apj* 159 (Feb. 1970), S. 379. DOI: 10.1086/150317.
- [2] D. J. Griffiths. *Introduction to elementary particles*. New York: Wiley, 1987. ISBN: 0-471-61544-7; 0-471-60386-4.
- [3] Wikimedia Commons. *File:Standard Model of Elementary Particles-de.svg*. 2020. URL: [https://commons.wikimedia.org/w/index.php?title=File:Standard\\_Model\\_of\\_Elementary\\_Particles-de.svg&oldid=521452056%7D](https://commons.wikimedia.org/w/index.php?title=File:Standard_Model_of_Elementary_Particles-de.svg&oldid=521452056%7D) (besucht am 09.06.2020).
- [4] M. Waßmer. „Search for the production of Dark Matter in hadronic mono-top signatures with the CMS experiment“. Diss. Karlsruhe Institute of Technology (KIT), 2021. DOI: 10.5445/IR/1000129768.
- [5] S. Wieland. „Unfolding in a Search for Dark Matter with Monojet Signatures“. Masterthesis. Karlsruhe Institute of Technology (KIT), 2019.
- [6] N. Aghanim u. a. „Planck 2018 results“. In: *Astronomy & Astrophysics* 641 (Sep. 2020), A6. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: <http://dx.doi.org/10.1051/0004-6361/201833910>.
- [7] E. Armengaud u. a. „Performance of the EDELWEISS-III experiment for direct dark matter searches“. In: *Journal of Instrumentation* 12.08 (Aug. 2017), P08010–P08010. DOI: 10.1088/1748-0221/12/08/p08010. URL: <https://doi.org/10.1088/1748-0221/12/08/p08010>.
- [8] W. B. Atwood u. a. „THE LARGE AREA TELESCOPE ON THE FERMI GAMMA-RAY SPACE TELESCOPE MISSION“. In: *The Astrophysical Journal* 697.2 (Mai 2009), S. 1071–1102. DOI: 10.1088/0004-637x/697/2/1071. URL: <https://doi.org/10.1088/0004-637x/697/2/1071>.
- [9] R. Abbasi u. a. „The design and performance of IceCube DeepCore“. In: *Astroparticle Physics* 35.10 (2012), S. 615–624. ISSN: 0927-6505. DOI: <https://doi.org/10.1016/j.astropartphys.2012.01.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0927650512000254>.
- [10] O. S. Brüning u. a. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: <http://cds.cern.ch/record/782076>.
- [11] A. A. Alves Jr. u. a. „The LHCb Detector at the LHC“. In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.
- [12] The ALICE Collaboration. „The ALICE experiment at the CERN LHC“. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002. URL: <https://doi.org/10.1088/1748-0221/3/08/s08002>.

- [13] The ATLAS Collaboration. „The ATLAS Experiment at the CERN Large Hadron Collider“. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: <https://doi.org/10.1088/1748-0221/3/08/s08003>.
- [14] The CMS Collaboration. „The CMS experiment at the CERN LHC“. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- [15] *The CMS Experiment at CERN*. 2020. URL: <https://cms.cern/detector> (besucht am 09.06.2020).
- [16] D. Barney. „CMS Detector Slice“. CMS Collection. Jan. 2016. URL: <https://cds.cern.ch/record/2120661>.
- [17] A. Sirunyan u. a. „Particle-flow reconstruction and global event description with the CMS detector“. In: *Journal of Instrumentation* 12.10 (Okt. 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003. URL: <https://doi.org/10.1088/1748-0221/12/10/p10003>.
- [18] M. Cacciari, G. P. Salam und G. Soyez. „The anti-kt jet clustering algorithm“. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), S. 063–063. DOI: 10.1088/1126-6708/2008/04/063. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [19] A. Sirunyan u. a. „Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV“. In: *Journal of Instrumentation* 13.05 (Mai 2018), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. URL: <https://doi.org/10.1088/1748-0221/13/05/p05011>.
- [20] The CMS Collaboration. *Search for monotop in the muon channel in proton-proton collisions at  $\sqrt{s} = 8$  TeV*. Techn. Ber. Geneva: CERN, 2016. DOI: CMS-PAS-B2G-15-001. URL: <https://cds.cern.ch/record/2143260>.
- [21] D. Cousineau, S. Brown und A. Heathcote. „Fitting distributions using maximum likelihood: Methods and packages“. In: *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36 (Dez. 2004), S. 742–56. DOI: 10.3758/BF03206555.
- [22] I. Goodfellow, Y. Bengio und A. Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [23] S. D. Schaub. „Classification of ttH Events using Lorentz Layer Networks“. Bachelorarbeit. Karlsruhe Institute of Technology (KIT), 2021.
- [24] H. Lüdemann. *Untersuchung verschiedener Deep-Learning-Konzepte in der ttH-Analyse am CMS- Experiment*. Bachelorarbeit. 2019.
- [25] Wikimedia Commons. *File:NeuronModel-de.svg*. 2010. URL: [https://de.wikipedia.org/wiki/K%C3%BCnstliches\\_neuronales\\_Netz#/media/Datei:NeuronModel\\_deutsch.svg%7D](https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#/media/Datei:NeuronModel_deutsch.svg%7D) (besucht am 04.08.2020).
- [26] J. van der Linden et al. *DRACO-MLfoy*. 2021. URL: <https://github.com/kit-cn-cms/DRACO-MLfoy> (besucht am 19.07.2021).
- [27] S. Alioli u. a. „A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX“. In: *Journal of High Energy Physics* 2010.6 (Juni 2010). ISSN: 1029-8479. DOI: 10.1007/jhep06(2010)043. URL: [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043).
- [28] J. Alwall u. a. „MadGraph 5: going beyond“. In: *Journal of High Energy Physics* 2011.6 (Juni 2011). ISSN: 1029-8479. DOI: 10.1007/jhep06(2011)128. URL: [http://dx.doi.org/10.1007/JHEP06\(2011\)128](http://dx.doi.org/10.1007/JHEP06(2011)128).



- [29] *DMsimp: Simplified dark matter models*. 2020. URL: <https://feynrules.irmp.ucl.ac.be/wiki/DMsimp> (besucht am 17.07.2021).
- [30] J. Andrea, B. Fuks und F. Maltoni. „Monotops at the LHC“. In: *Physical Review D* 84.7 (Okt. 2011). ISSN: 1550-2368. DOI: 10.1103/physrevd.84.074025. URL: <http://dx.doi.org/10.1103/PhysRevD.84.074025>.
- [31] T. Sjöstrand, S. Mrenna und P. Skands. „A brief introduction to PYTHIA 8.1“. In: *Computer Physics Communications* 178.11 (2008), S. 852–867. ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2008.01.036>. URL: <https://www.sciencedirect.com/science/article/pii/S0010465508000441>.
- [32] S. Agostinelli u.a. „Geant4—a simulation toolkit“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), S. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.



# Anhang

## A Abkürzungen

**MET** Fehlende Transversalenergie  $\cancel{E}_T$

**Jet<sub>F</sub>** Führender Jet

**Jets<sub>b</sub>** b-tagged Jets

**Lep** Lepton

**sl** semileptonisch

**ROC-Kurve** Receiver-Operating-Characteristic-Curves, Auftragung der Falsch-Positiv-Rate gegen die Richtig-Positiv-Rate

**ROC-AUC** Area Under Curve, Integral einer ROC-Kurve

**DNN** Deep neural Network

## B Wichtige Variablen

Tabelle B.1: **Wichtige Variablen.** Auflistung der wichtigsten Variablen mit ihren Bedeutungen.

Variablen	Bedeutung	Eingeführt in Abschnitt
$p_T$	Transversalimpuls	2.3.3
$m_T$	transversale Masse	3.3
$\eta$	Pseudorapidität	2.3.3
$E$	Energie	
$N$	Anzahl	
$\cancel{E}_T$	fehlende Transversalenergie	2.3.3
$\Phi$	Azimutalwinkel	2.3.3
$\Delta R$	Abstand in der $\eta$ - $\phi$ -Ebene des Detektors	2.3.3
$\vec{U}_T$	Hadronischer Rückstoß	2.3.3
$H_T$	Skalare Summe der $p_T$	

## C Ergänzende Materialien zur Multiklassifikation von Signal und Untergründen

### C.1 Verteilungen der Eingabevariablen

Im Folgenden sind die Verteilungen der Eingabevariablen des neuronalen Netzes zu sehen, auf die in Abschnitt 5.1 verwiesen wird. Eine Beschreibung der aufgeführten Variablen findet sich in Tabelle 4.4 bzw. Tabelle B.1. Die verwendete Mono-top-Signalthypothese hat die Massen  $M_\phi = 2495$  und  $M_\chi = 1250$ .

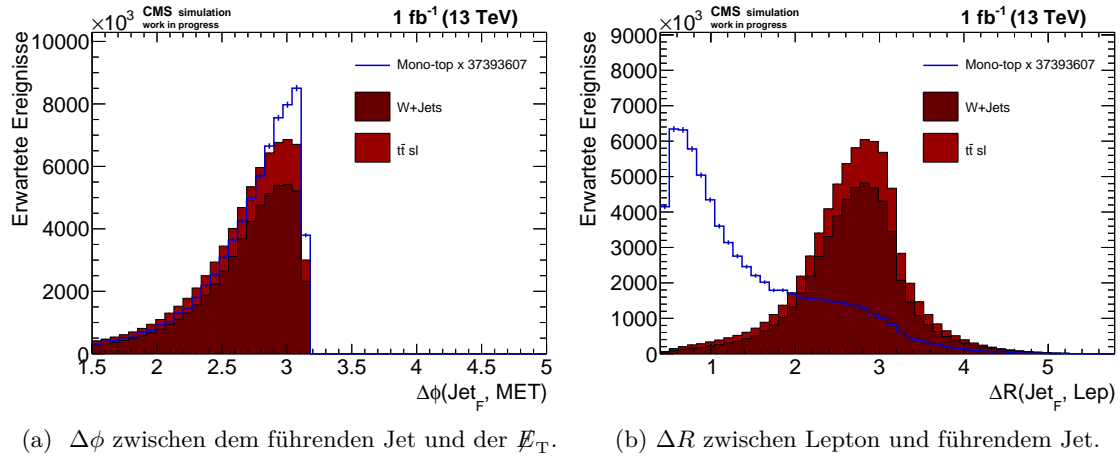


Abbildung C.1: Verteilungen von zwei der 15 Eingabevariablen.

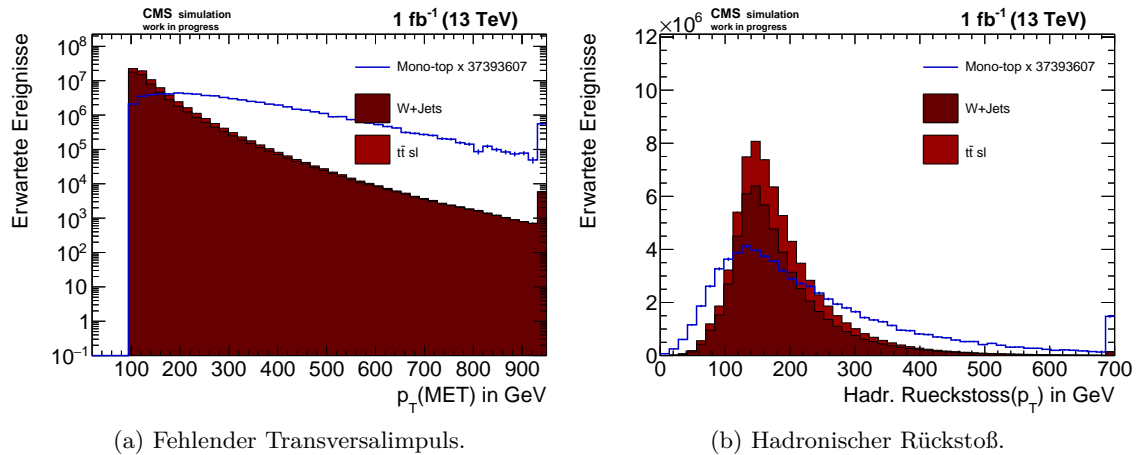
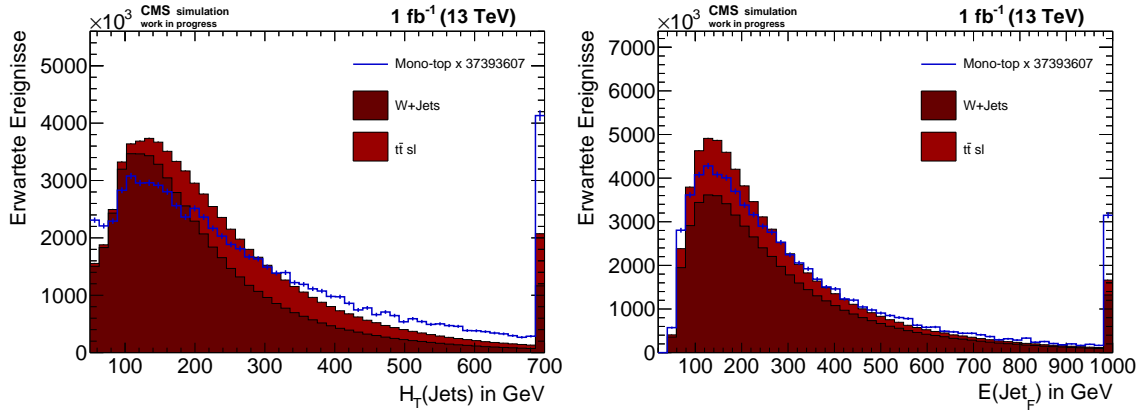


Abbildung C.2: Verteilungen von zwei der 15 Eingabevariablen.

(a) Skalare Summe der  $p_T$  aller Jets.

(b) Energie des führenden Jets.

Abbildung C.3: Verteilungen von zwei der 15 Eingabevariablen.

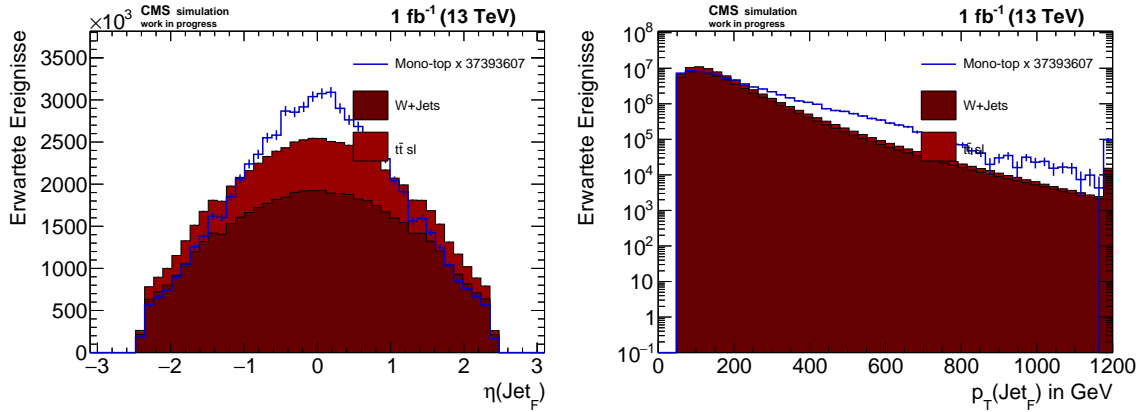
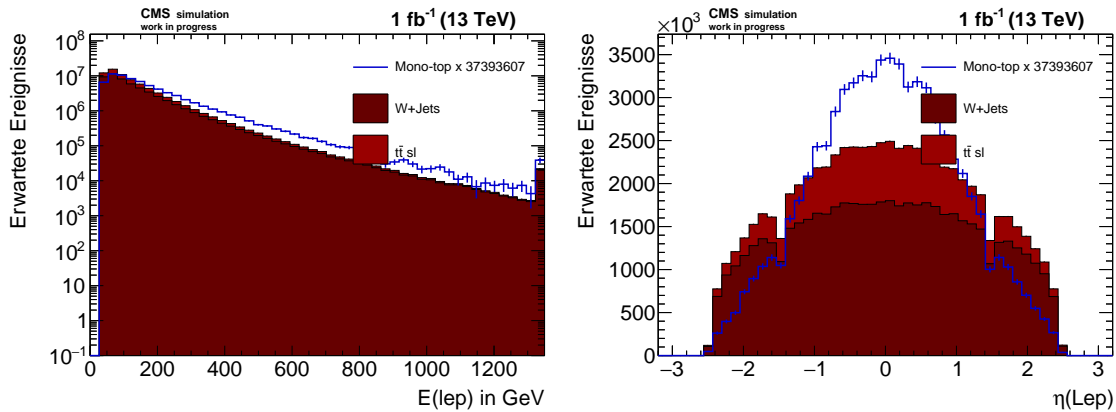
(a) Pseudorapidity  $\eta$  des führenden Jets.(b) Transversalimpuls  $p_T$  des führenden Jets.

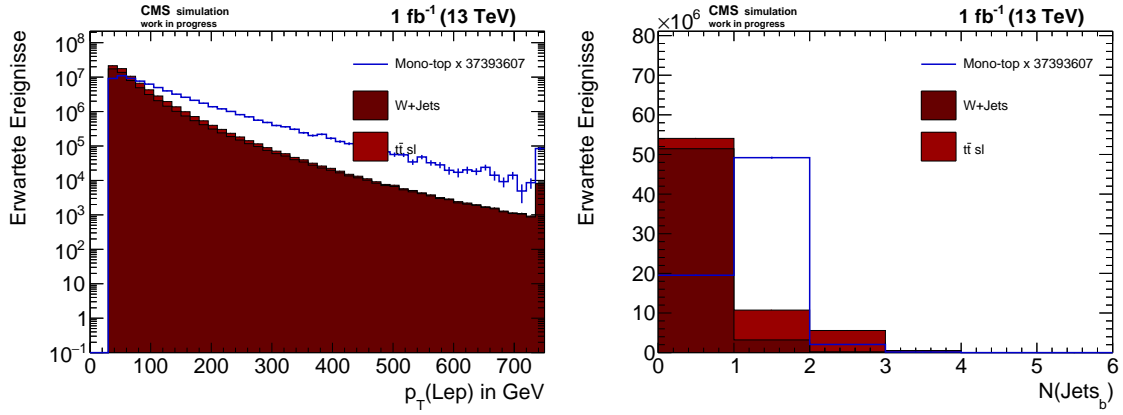
Abbildung C.4: Verteilungen von zwei der 15 Eingabevariablen.



(a) Energie des Leptons.

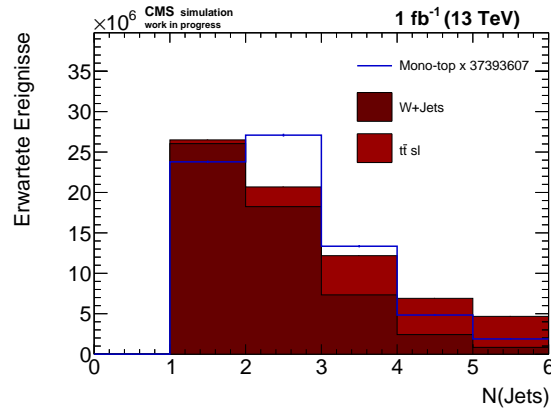
(b) Pseudorapidity  $\eta$  des Leptons.

Abbildung C.5: Verteilungen von zwei der 15 Eingabevariablen.

(a) Transversalimpuls  $p_T$  des Leptons.

(b) Anzahl der b-tagged Jets.

Abbildung C.6: Verteilungen von zwei der 15 Eingabevariablen.



(a) Anzahl der Jets.

Abbildung C.7: Verteilung von einer der 15 Eingabevariablen.

## C.2 Ranking der Eingabevariablen

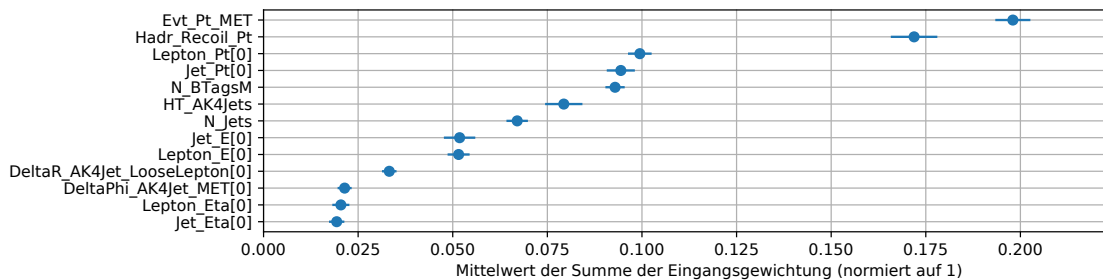


Abbildung C.8: **Ranking der Eingabevariablen.** Zu sehen ist eine Liste der Eingabevariablen, die nach der Summe der Gewichte der ersten Schicht geordnet ist. In Tabelle 4.4 findet sich eine Beschreibung der hier aufgeführten Variablen.

## D Materialien zur Klassifizierung verschiedener Mono-top-Signalthypothesen

### D.1 Eingabevariablen für das Training mit Signalthypothesen mit Mediatoren im Bereich hoher Mediatormassen

Im Folgenden sind die Verteilungen der Eingabevariablen des neuronalen Netzes zu sehen, auf die in Abschnitt 6.1 verwiesen wird. Eine Beschreibung der aufgeführten Variablen findet sich in Tabelle 4.4 bzw. Tabelle B.1.

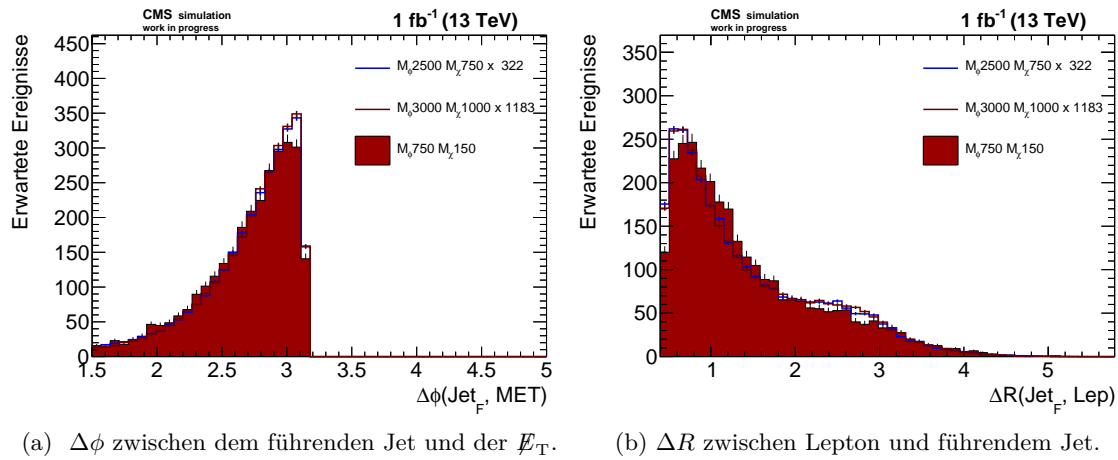


Abbildung D.9: Verteilungen von zwei der 15 Eingabevariablen.

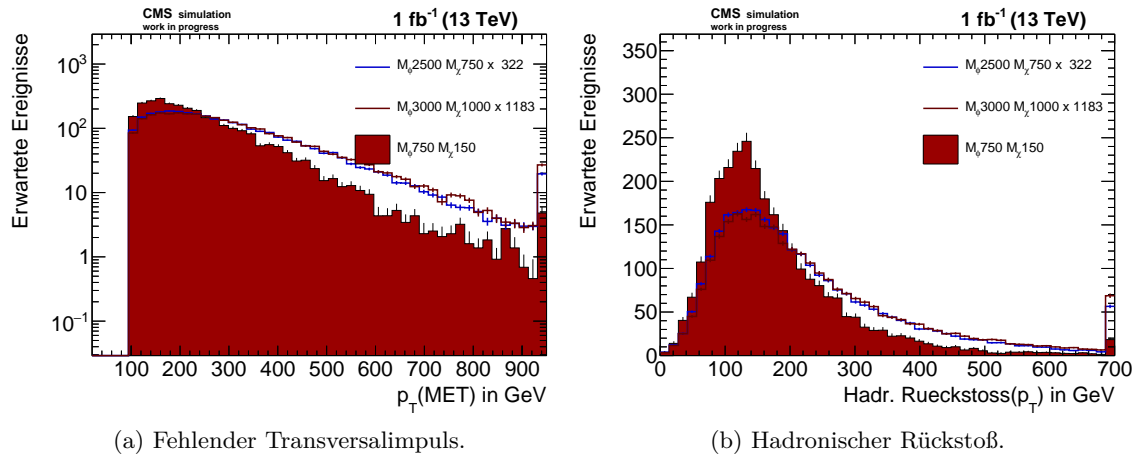
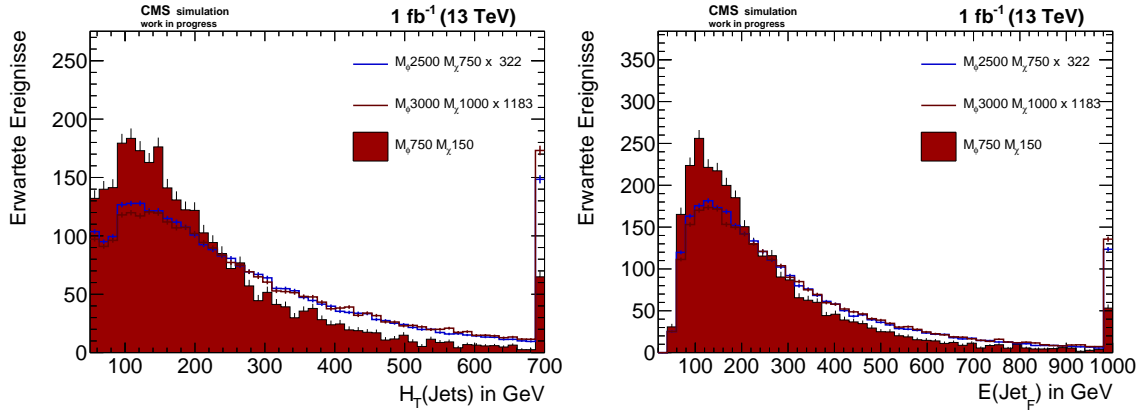


Abbildung D.10: Verteilungen von zwei der 15 Eingabevariablen.

(a) Skalare Summe der  $p_T$  aller Jets.

(b) Energie des führenden Jets.

Abbildung D.11: Verteilungen von zwei der 15 Eingabevariablen.

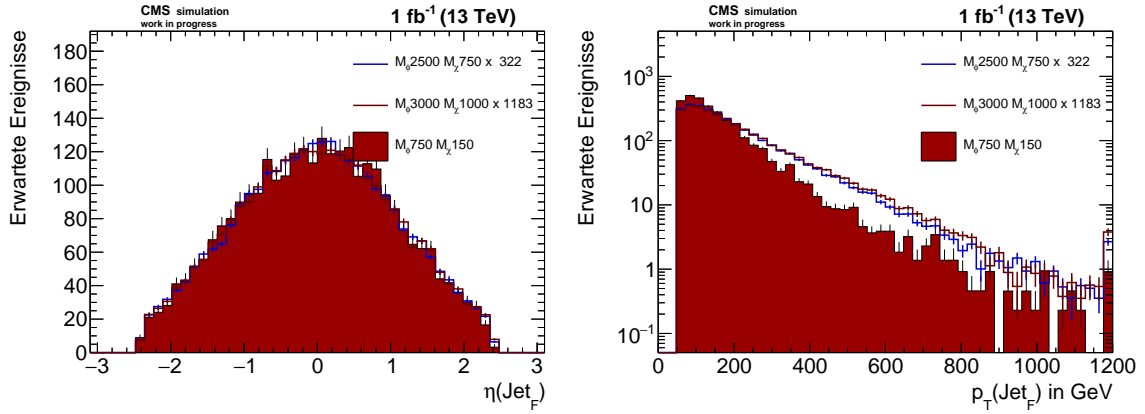
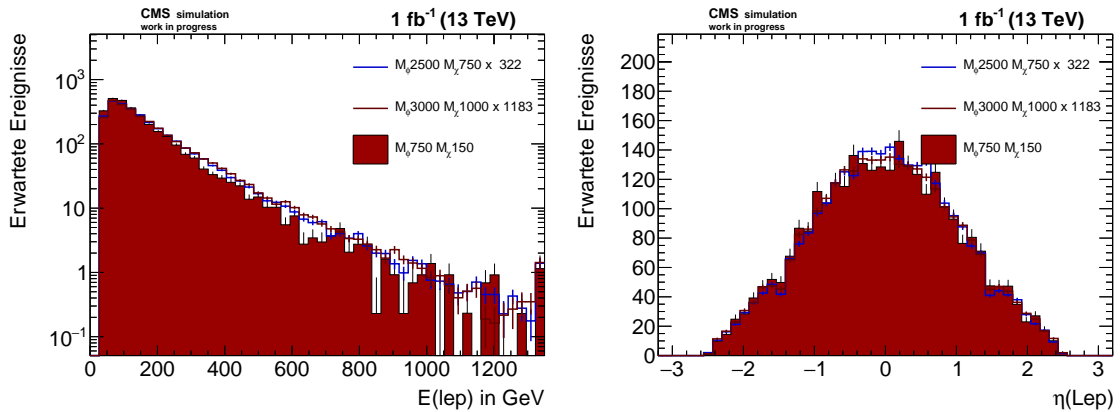
(a) Pseudorapidity  $\eta$  des führenden Jets.(b) Transversalimpuls  $p_T$  des führenden Jets.

Abbildung D.12: Verteilungen von zwei der 15 Eingabevariablen.



(a) Energie des Leptons.

(b) Pseudorapidity  $\eta$  des Leptons.

Abbildung D.13: Verteilungen von zwei der 15 Eingabevariablen.



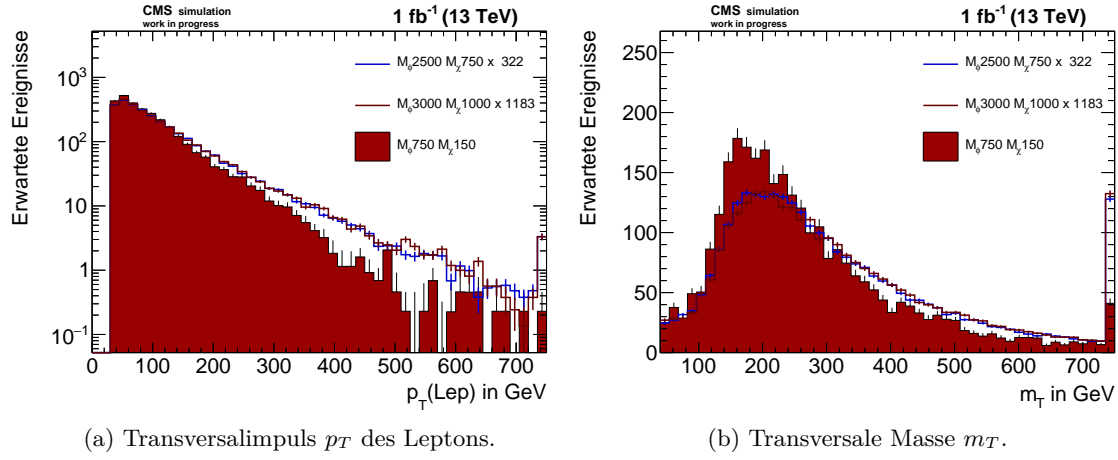


Abbildung D.14: Verteilungen von einer der 15 Eingabevariablen und der zusätzlich verwendeten transversalen Masse.

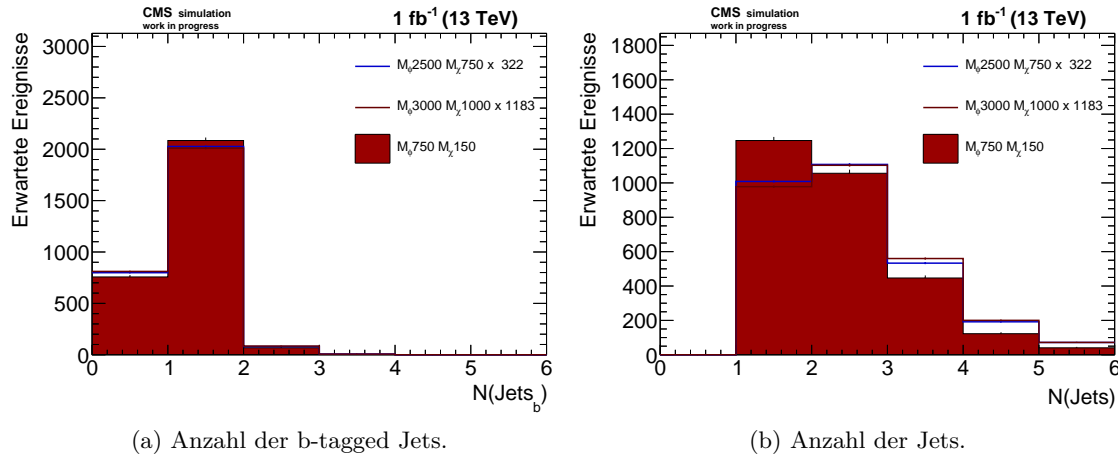


Abbildung D.15: Verteilung von zwei der 15 Eingabevariablen.

## D.2 Eingabevariablen für das Training im Bereich niedriger Mediatormassen

Im Folgenden sind die Verteilungen der Eingabevariablen des neuronalen Netzes zu sehen, auf die in Abschnitt 6.2.1 verwiesen wird. Eine Beschreibung der aufgeführten Variablen findet sich in Tabelle 4.4 bzw. Tabelle B.1.

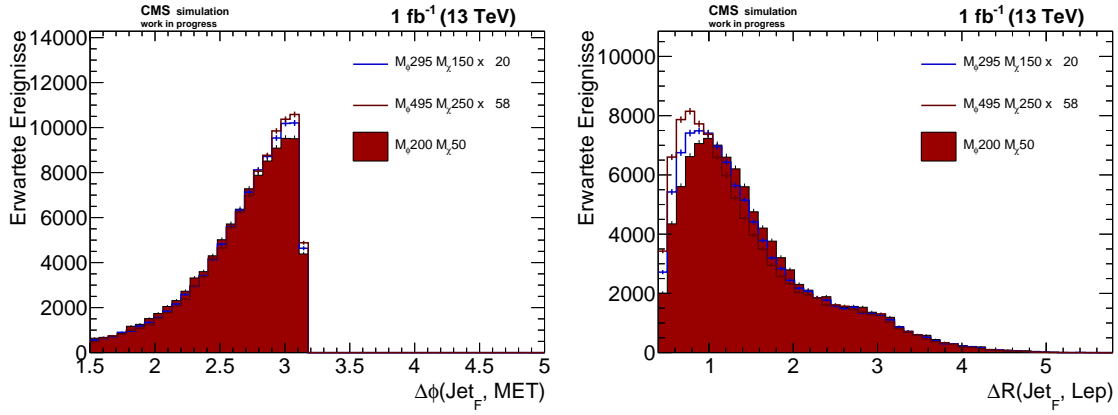
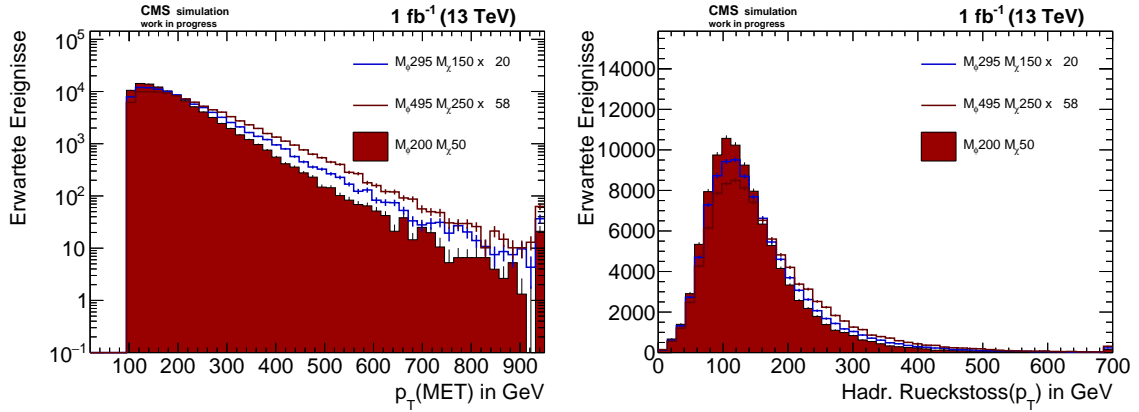
(a)  $\Delta\phi$  zwischen dem führenden Jet und der  $\cancel{E}_T$ .(b)  $\Delta R$  zwischen Lepton und führendem Jet.

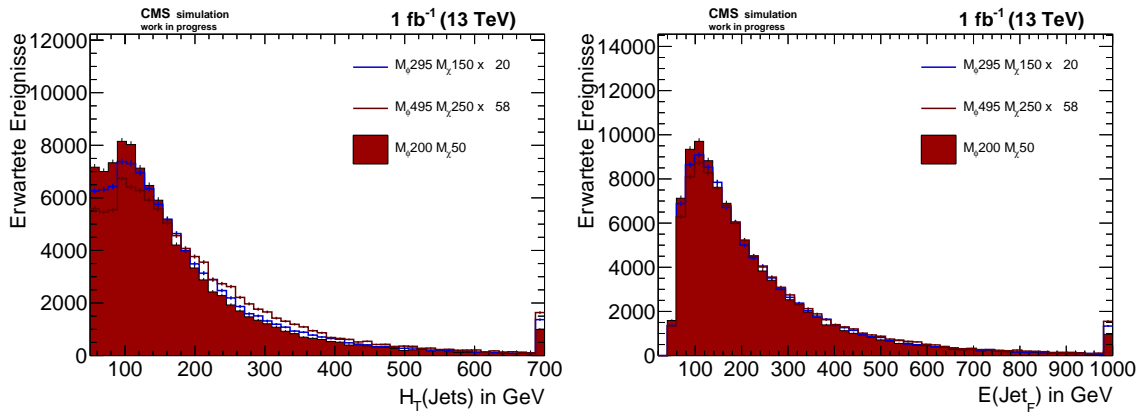
Abbildung D.16: Verteilungen von zwei der 15 Eingabevariablen.



(a) Fehlender Transversalimpuls.

(b) Hadronischer Rückstoß.

Abbildung D.17: Verteilungen von zwei der 15 Eingabevariablen.

(a) Skalare Summe der  $p_T$  aller Jets.

(b) Energie des führenden Jets.

Abbildung D.18: Verteilungen von zwei der 15 Eingabevariablen.

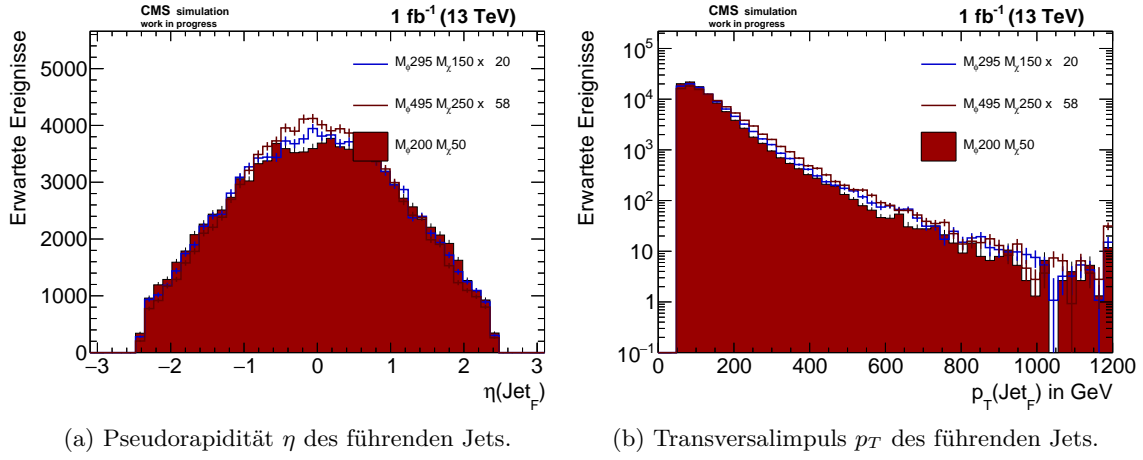


Abbildung D.19: Verteilungen von zwei der 15 Eingabevariablen.

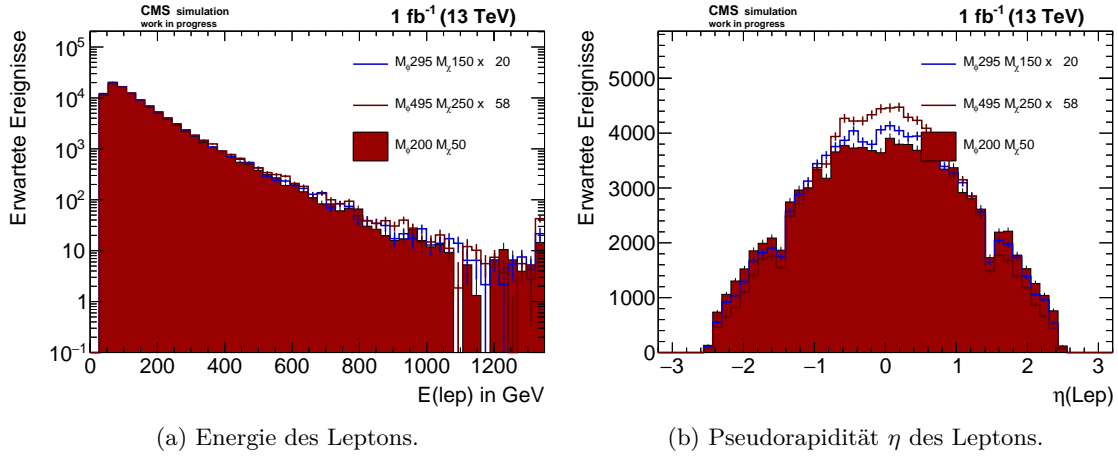


Abbildung D.20: Verteilungen von zwei der 15 Eingabevariablen.

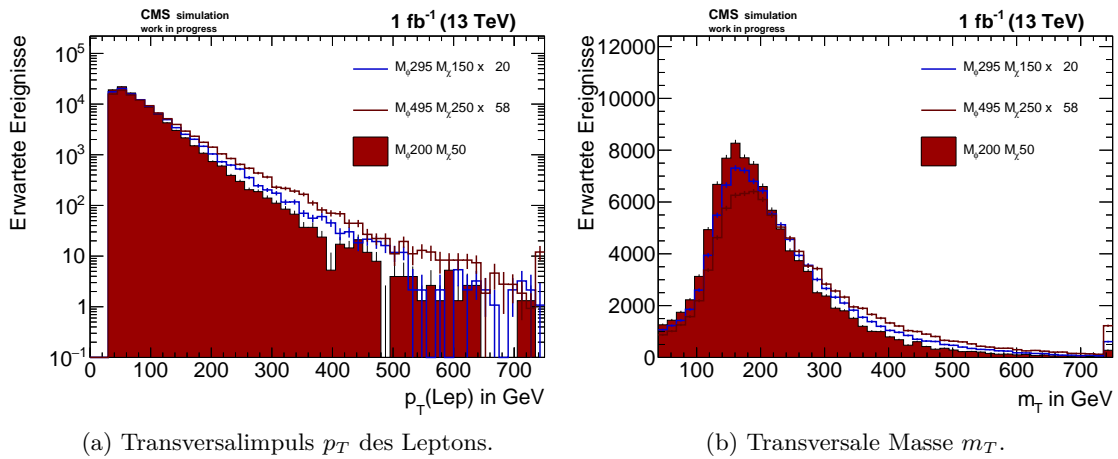
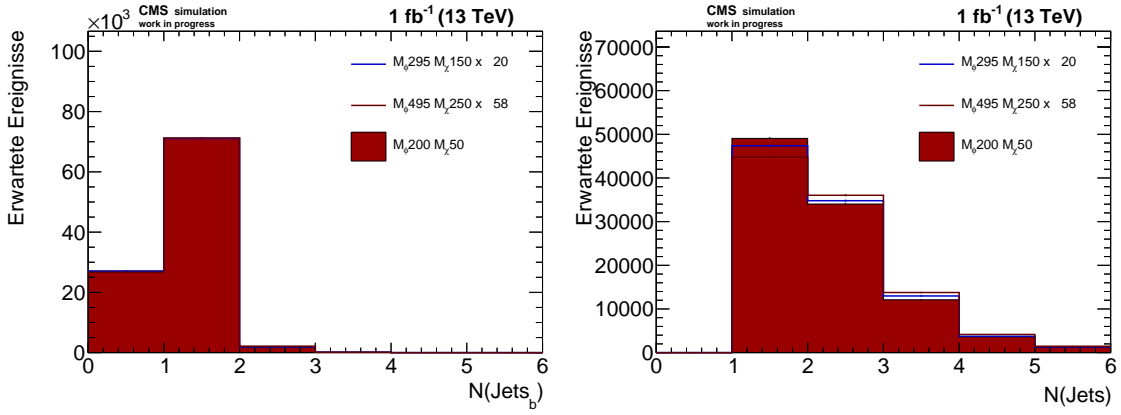


Abbildung D.21: Verteilungen von einer der 15 Eingabevariablen und der zusätzlich verwendeten transversalen Masse.



(a) Anzahl der b-tagged Jets.

(b) Anzahl der Jets.

Abbildung D.22: Verteilung von zwei der 15 Eingabevariablen.

### D.3 Wahrheitsmatrix für ein Training mit drei Mono-top-Datensätzen im Bereich niedriger Mediatormassen

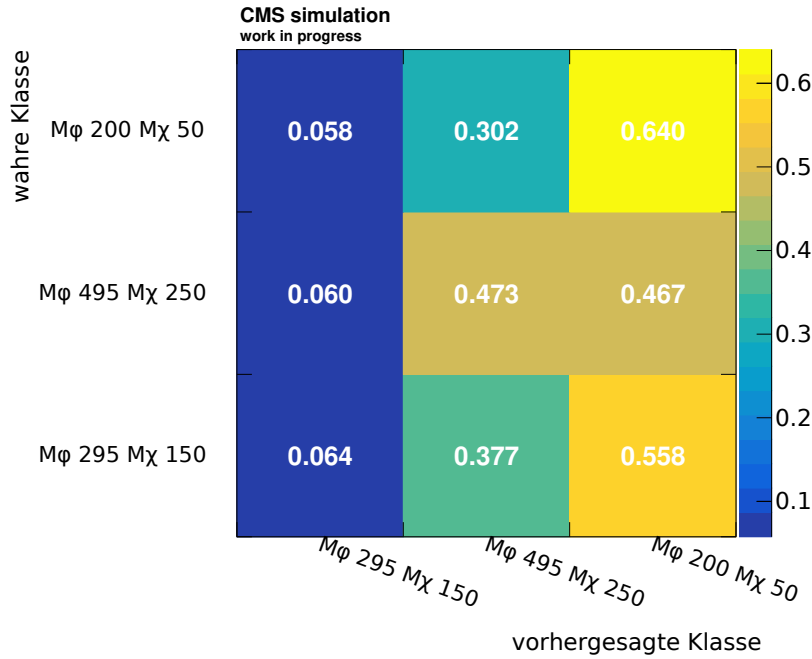


Abbildung D.23: **Wahrheitsmatrix** Training mit drei Mono-top-Signalthypothesen im niedrigen Mediatormassen-Bereich mit jeweils 50 Epochen. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. Es wurde zusätzlich mit der Eingabevariable transversale Masse  $m_T$  trainiert.

### D.4 Eingabevariablen für das Training mit weit auseinander liegenden Massepunkten

Im Folgenden sind die Verteilungen der Eingabevariablen des neuronalen Netzes zu sehen, auf die in Abschnitt 6.2.2 verwiesen wird. Eine Beschreibung der aufgeführten Variablen findet sich in Tabelle 4.4 bzw. Tabelle B.1.

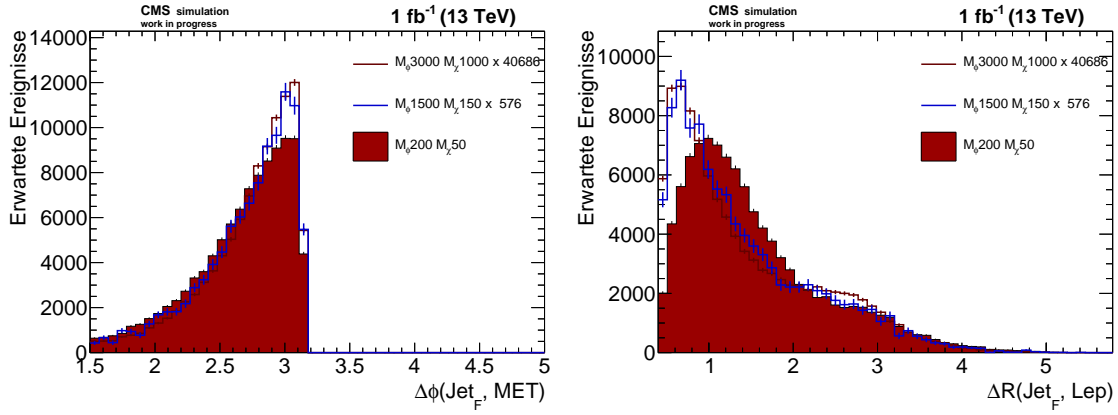
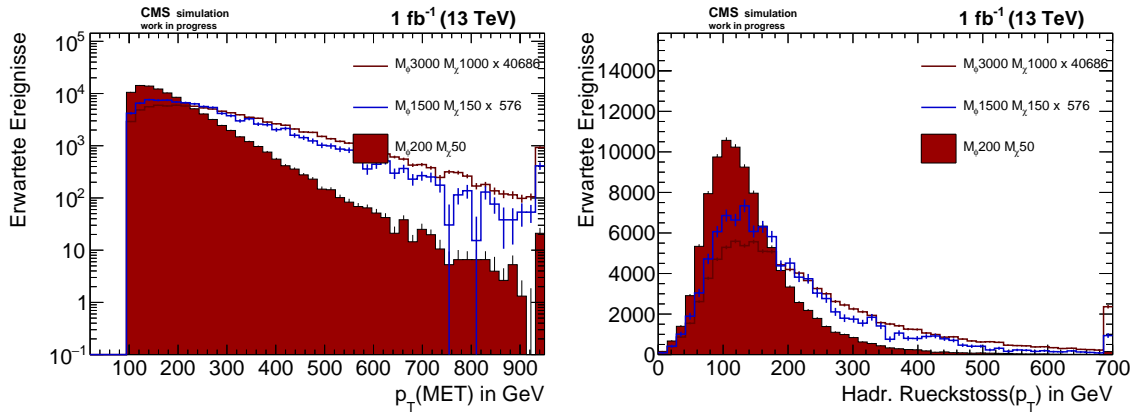
(a)  $\Delta\phi$  zwischen dem führenden Jet und der  $\cancel{E}_T$ .(b)  $\Delta R$  zwischen Lepton und führendem Jet.

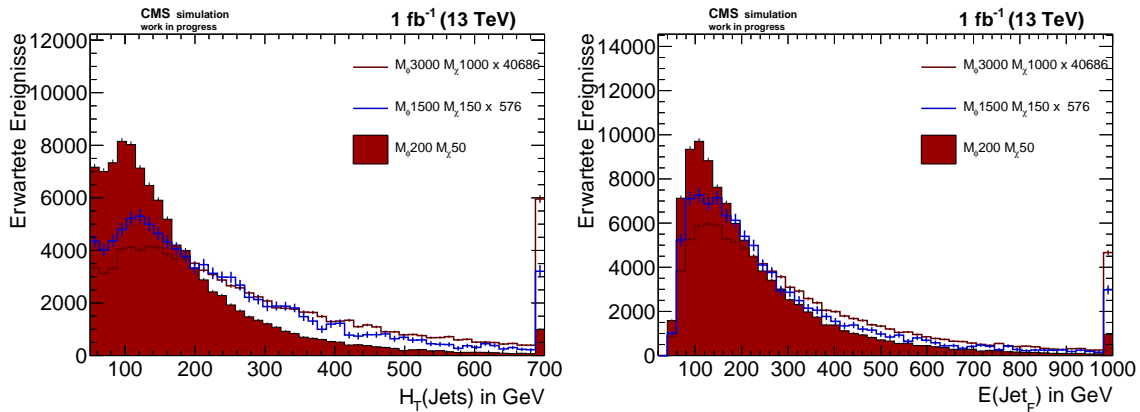
Abbildung D.24: Verteilungen von zwei der 15 Eingabevariablen.



(a) Fehlender Transversalimpuls.

(b) Hadronischer Rückstoß.

Abbildung D.25: Verteilungen von zwei der 15 Eingabevariablen.

(a) Skalare Summe der  $p_T$  aller Jets.

(b) Energie des führenden Jets.

Abbildung D.26: Verteilungen von zwei der 15 Eingabevariablen.

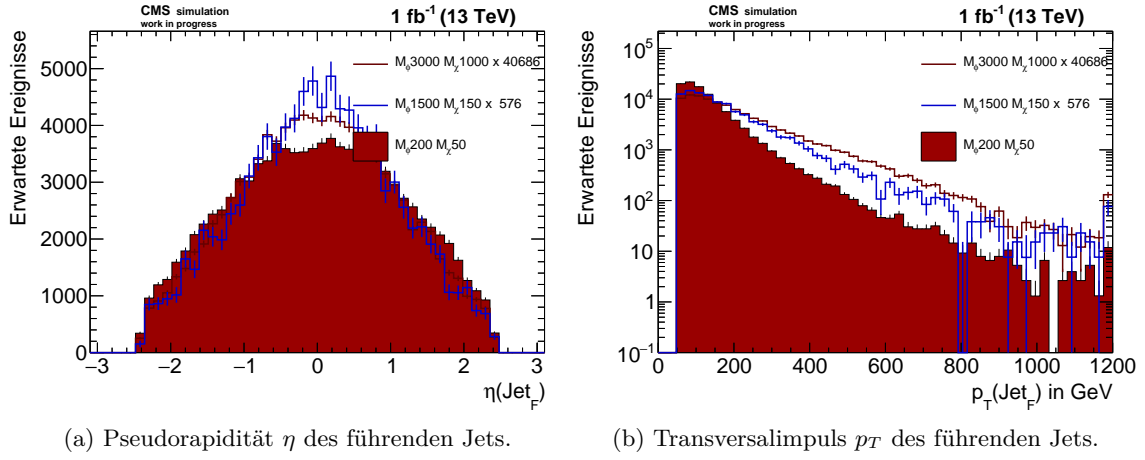


Abbildung D.27: Verteilungen von zwei der 15 Eingabevariablen.

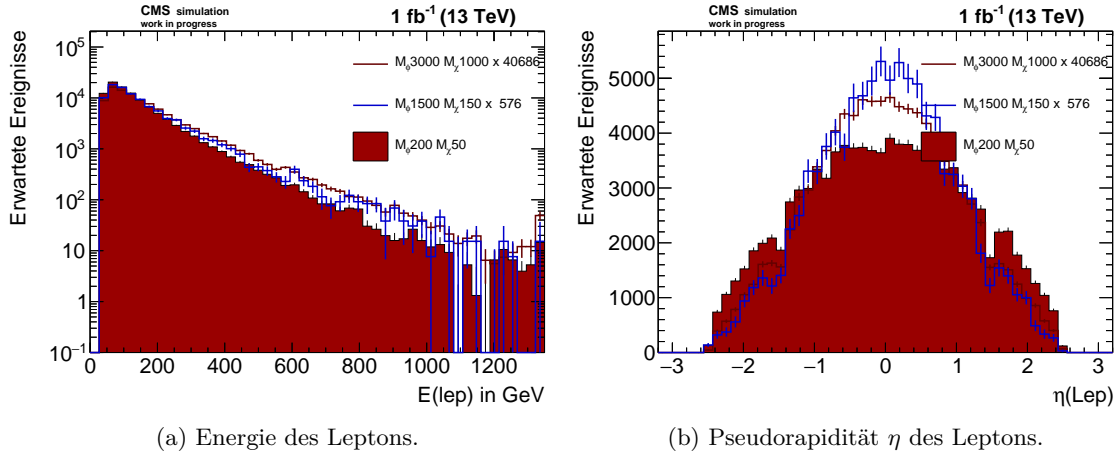


Abbildung D.28: Verteilungen von zwei der 15 Eingabevariablen.

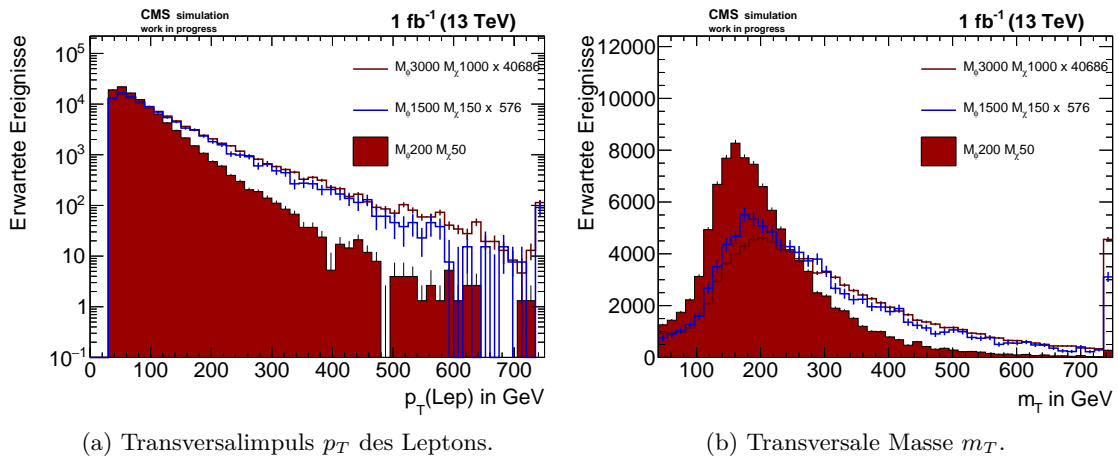
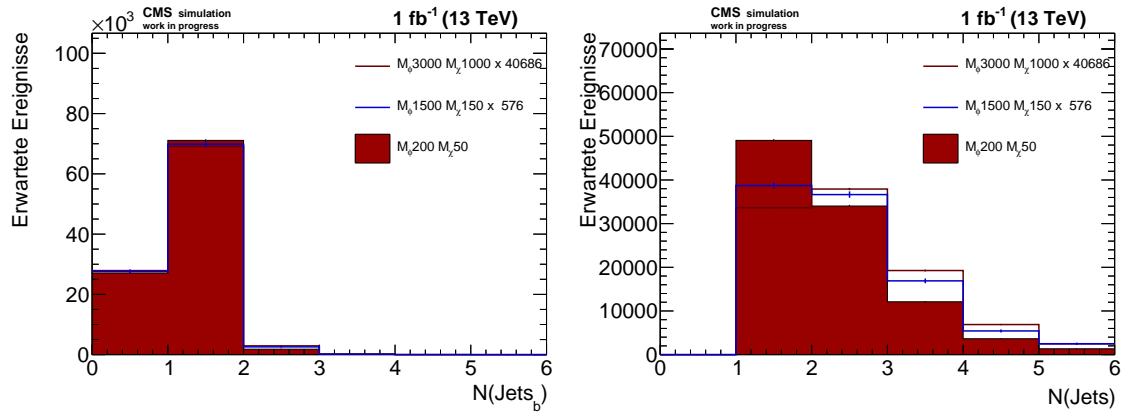


Abbildung D.29: Verteilungen von einer der 15 Eingabevariablen und der zusätzlich verwendeten transversalen Masse.



(a) Anzahl der b-tagged Jets.

(b) Anzahl der Jets.

Abbildung D.30: Verteilung von zwei der 15 Eingabevariablen.

## E Materialien zur Untersuchung des Verhaltens bei geringer Trennkraft

### E.1 Gegenüberstellung der Anzahl der im Training verwendeten Ereignisse mit den entsprechenden Werten auf der Diagonalen der Wahrheitsmatrix

Tabelle E.2: Massen der verwendeten Mono-top-Signalthypothesen und Ergebnisse der Diagonalen der Wahrheitsmatrix Training im Bereich hoher, niederer Mediatormassen und mit möglichst weit auseinander liegenden Massensepunkten.

Abschnitt	Massen-Bereich	Mediatormasse $M_\phi$ in GeV	DM-Masse $M_\chi$ in GeV	Anzahl der Ereignisse	Diagonale der Wahrheitsmatrix
6.1	hoch	3000	1000	$2,32 \cdot 10^5$	52,6 %
		2500	750	$2,28 \cdot 10^5$	0,57 %
		750	150	$2,29 \cdot 10^4$	67,3 %
6.2.1	niedrig	295	150	$1,49 \cdot 10^5$	0,64 %
		495	250	$1,82 \cdot 10^5$	47,3 %
		200	50	$1,19 \cdot 10^5$	64,64 %
6.2.2	weit auseinander	3000	1000	$2,32 \cdot 10^5$	55,0 %
		1500	150	$2,53 \cdot 10^4$	1,63 %
		200	50	$1,19 \cdot 10^5$	75,3 %

## E.2 Fluktuation in den Wahrheitsmatrizen bei der Multiklassifikation zu ähnlicher Signalthypothesen

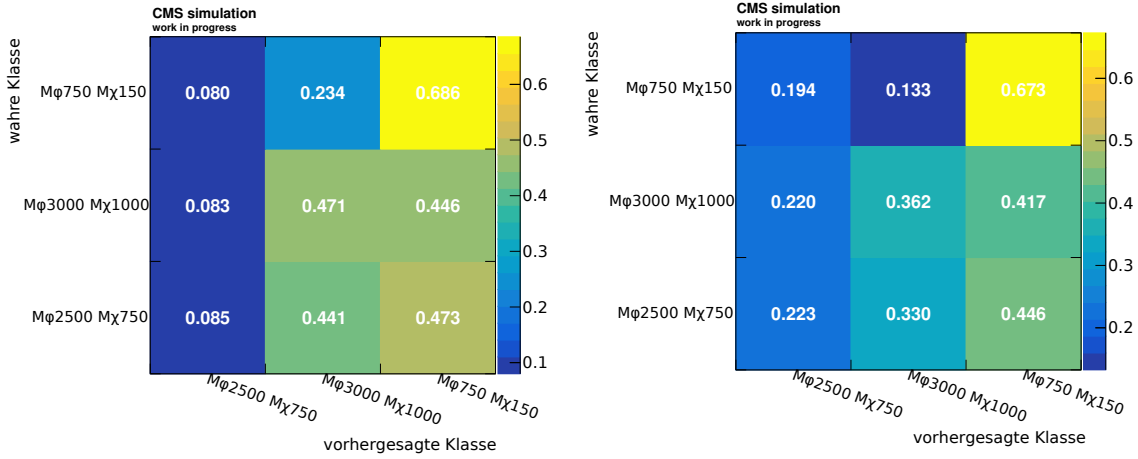
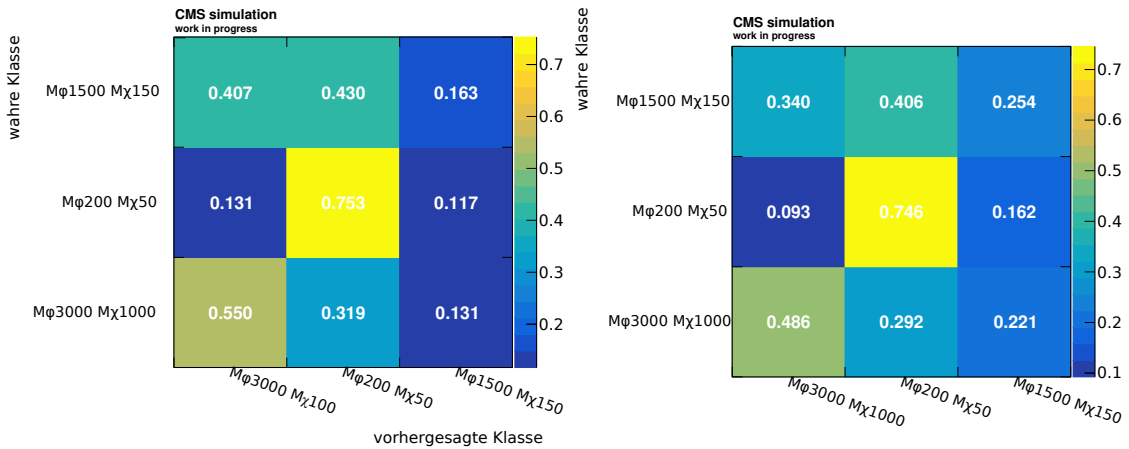


Abbildung E.31: **Instabilität des Trainings bei zu ähnlichen Signalthypothesen.**

Wahrheitsmatrizen aus 17 Trainings mit drei Mono-top-Signalthypothesen mit möglichst hohen Mediatormassen mit jeweils 50 Epochen. Die in (a) dargestellte Wahrheitsmatrizen entspricht dem häufigsten Fall. Seltener ergeben sich Wahrheitsmatrizen wie in (b) dargestellt. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. Es wurde zusätzlich mit der Eingabevariable transversale Masse  $m_T$  trainiert.

## E.3 Wahrheitsmatrizen zur Untersuchung des Einflusses der Wirkungsquerschnitts Gewichtung



(a) Unveränderte Gewichtung.

(b) Gewichtung für den Wirkungsquerschnitt auf ein gesetzt.

Abbildung E.32: **Wahrheitsmatrizen.** Training mit drei Mono-top-Signalthypothesen mit Mediatormassen die möglichst weit auseinander liegen mit jeweils 50 Epochen. In (a) wird die normale Gewichtung für den Wirkungsquerschnitt verwendet und in (b) ist diese auf eins gesetzt. Die für das Training verwendete Architektur ist in Abschnitt 4.1 beschrieben. Es wurde zusätzlich mit der Eingabevariable transversale Masse  $m_T$  trainiert.



## F Verteilung der fehlenden Transversalenergie gegen die des Abstandes zwischen Lepton und Myon aufgetragen

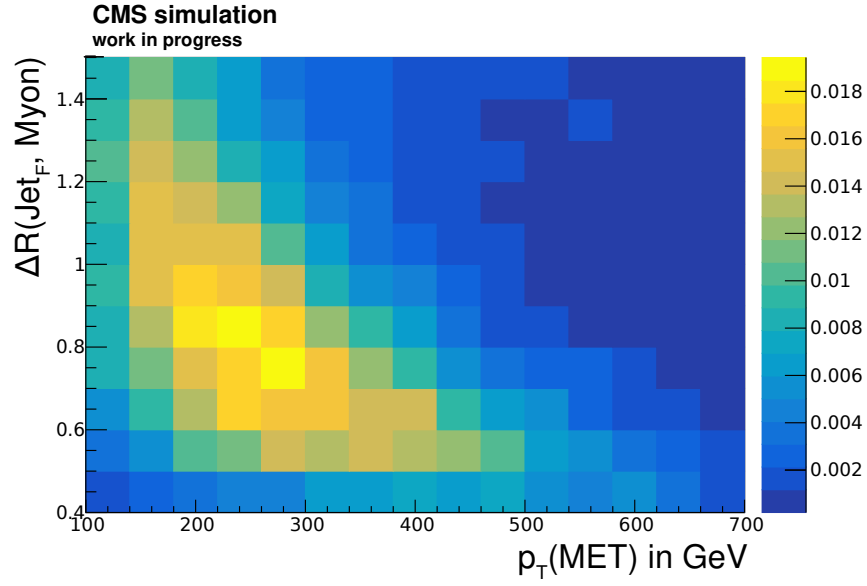


Abbildung F.33: **Zweidimensionaler Plot.** Die Verteilung der fehlenden Transversalenergie  $\cancel{E}_T$  ist gegen die des Abstandes  $\Delta R$  zwischen Myon und führendem Jet aufgetragen. Die betrachtete Mono-top-Signalhypothese hat die Massen  $M_\phi = 3000$  GeV und  $M_\chi = 1000$  GeV. Bereitgestellt durch Michael Waßmer.