

STUDIEN ZUM t -KANAL-UNTERGRUND IN DER
SUCHE NACH EINZEL-TOPQUARK-PRODUKTION
IM s -KANAL MIT DEM CMS-EXPERIMENT

Han Wang

Fakultät für Physik
Karlsruher Institut für Technologie (KIT)

BACHELORARBEIT

Referent: Prof. Dr. Thomas Müller

Gutachter: Dr. Thorsten Chwalek

Institut für Experimentelle Teilchenphysik

13. August 2018

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 13. August 2018

.....
(Han Wang)

Inhaltsverzeichnis

1. Einleitung	1
2. Theorie	3
2.1. Standardmodell	3
2.2. Top-Quark	6
2.2.1. Top-Quark-Paarproduktion ($t\bar{t}$)	6
2.2.2. Einzel-Top-Quark-Produktion	8
2.3. Simulation der Ereignisse	9
3. Experiment	11
3.1. Large Hadron Collider	11
3.2. Compact Muon Solenoid	13
4. Statistische Methoden	17
4.1. Tiefe neuronale Netze	17
4.1.1. Training des tiefen neuronalen Netzes	20
4.1.2. Parameter des tiefen neuronalen Netzes	20
4.2. Kolmogorov-Smirnov-Test	21
4.3. Signifikanz	22
5. Suche nach der Einzel-Top-Quark-Produktion im s -Kanal	23
5.1. Strategie	23
5.2. Studien zur Optimierung der Referenzanalyse	24
5.3. Studien in der 2-Jet-2-Tag-Region	25
5.3.1. Klassifizierung	28
5.3.2. Schlussfolgerung	29
5.4. Studien in der 2-Jet-1-Tag Region	30
5.4.1. Myon-Kanal	31
5.4.2. Elektron-Kanal	40
5.4.3. Kombination des Myon- und Elektron-Kanals	45
5.4.4. Schlussfolgerung	55
6. Zusammenfassung	57

A. Variablen für das DNN-Training in Myon-Kanal	59
B. Variablen für das DNN-Training im Elektron-Kanal	63
C. Variablen für das DNN-Training des kombinierten Myon- und Elektron-Kanals	67

1. Einleitung

Das Standardmodell der Teilchenphysik fasst die wesentlichen Erkenntnisse der Teilchenphysik zusammen und beschreibt die Eigenschaften der verschiedenen Elementarteilchen und ihre Wechselwirkungen. Quarks stellen einen wichtigen Bestandteil des Standardmodells dar. Das Top-Quark ist dabei mit einer Masse von 173,0 GeV [1] das massivste aller beobachteten Elementarteilchen. Wegen seiner hohen Masse müssen die Teilchen bei sehr hoher Energie zur Kollision gebracht werden, um das Top-Quark zu erzeugen. Es wurde erst im Jahr 1995 mit Experimenten am Tevatron entdeckt [2, 3], einem Teilchenbeschleuniger am Forschungszentrum Fermilab in den USA.

Seit 2010 können Top-Quarks regelmäßig am Large Hadron Collider (LHC) erzeugt werden, einem Teilchenbeschleuniger mit einem Umfang von 27 km, der von der Europäischen Organisation für Kernforschung (CERN) bei Genf betrieben wird. Dabei werden die Protonen in entgegengesetzter Richtung auf hohe Energien beschleunigt und zur Kollision gebracht. Die aus den Kollisionen stammenden Zerfallsprodukte werden durch Detektoren gemessen und zur Rekonstruktion der physikalischen Objekte verwendet. Einer der Detektoren ist der Compact Muon Solenoid (CMS), welcher fast alle bei jeder Kollision erzeugten stabilen Teilchen identifizieren kann. Top-Quarks werden bei Beschleunigern bevorzugt paarweise über die starke Wechselwirkung erzeugt. Einzelne Top-Quarks werden dagegen über die schwache Wechselwirkung erzeugt. Die Erzeugung einzelner Top-Quarks ist über verschiedene Prozesse möglich, wobei der s -Kanal-Prozess nur einen geringen Beitrag liefert, da der Wirkungsquerschnitt des s -Kanals kleiner als bei den anderen möglichen Prozessen ist. Trotzdem ist der s -Kanal-Prozess von besonderem Interesse, da eine mögliche Abweichung des Wirkungsquerschnitts von der Standardmodell-Vorhersage ein Indikator für Physik jenseits des Standardmodells sein könnte. In der vorliegenden Arbeit wird die Analyse zur Suche nach der Produktion einzelner Top-Quarks im s -Kanal verbessert. Für die Untersuchung dieses Prozesses werden die im Jahr 2016 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV gemessenen Daten verwendet, die einer integrierten Luminosität von $L_{\text{int}} = 35,9$ fb $^{-1}$ entsprechen. Um den s -Kanal-Prozess am LHC zweifelsfrei nachweisen zu können, muss dieser Prozess von Untergrundprozessen getrennt werden. Die

Arbeit fokussiert sich auf die Trennung von s -Kanal- und t -Kanal-Ereignissen, da die t -Kanal-Produktion der dominante Prozess zur Erzeugung einzelner Top-Quarks ist. Um eine gute Klassifizierung zu erhalten, wird in der Arbeit eine multivariate Analyseverfahren, die tiefen neuronalen Netze (Deep Neural Networks, DNNs), verwendet.

Im Folgenden werden zunächst in Kapitel 2 die theoretischen Grundlagen beschrieben. Dabei werden das Standardmodell und die relevanten Prozesse zur Erzeugung des Top-Quarks erklärt. Außerdem wird die Simulation der Ereignisse beschrieben. Anschließend wird in Kapitel 3 das experimentelle Umfeld für die Erzeugung des Top-Quarks vorgestellt. In Kapitel 4 werden die für die Klassifizierung verwendeten Methoden beschrieben. Dabei werden auch der Kolmogorov-Smirnov (KS)-Test und die Signifikanz erklärt, deren Werte das DNN-Training evaluiert. In Kapitel 5 wird die Optimierung der s -Kanal-Analyse beschrieben und das Ergebnis wird mit einer Referenzanalyse verglichen.

2. Theorie

Im Folgenden wird ein Überblick über die theoretischen Grundlagen gegeben, die für diese Arbeit notwendig sind. Es werden das Standardmodell und die unterschiedlichen Prozesse zur Erzeugung des Top-Quarks beschrieben.

2.1. Standardmodell

Das Standardmodell der Teilchenphysik ist eine relativistische Quantenfeldtheorie, welche die Eigenschaften und Wechselwirkungen von fundamentalen Teilchen beschreibt. Diese Theorie wurde in den 1960er und 1970er Jahren entwickelt. Sie sagte neue Teilchen voraus, die später von Beschleuniger-Experimenten entdeckt und verifiziert wurden.

Im Standardmodell werden grundsätzlich zwei Arten von Elementarteilchen unterschieden, Fermionen und Bosonen.

- Fermionen
Fermionen sind Elementarteilchen mit halbzahligem Spin. Sie unterliegen dem Pauli-Prinzip. Zwei oder mehrere Fermionen können also nach diesem Grundprinzip nicht den gleichen quantenmechanischen Zustand einnehmen. Sämtliche Materie besteht in ihren Grundbausteinen aus Fermionen. Folglich bestehen alle Nukleonen aus Fermionen und auch Elektronen sind Fermionen. Weiterhin unterscheidet man Fermionen in sechs Quarks und sechs Leptonen, die in drei Familien angeordnet werden (siehe Abb. 2.1).
- Bosonen
Bosonen sind Elementarteilchen mit ganzzahligem Spin. Von Fermionen unterscheiden sich die Bosonen darin, dass sie nicht dem Pauli-Prinzip unterliegen. Zu den Bosonen gehören vier Eichbosonen (siehe Abb. 2.1), die der Vermittlung der verschiedenen Wechselwirkungen dienen.

Drei Generationen
der Materie (Fermionen)

	I	II	III		
Masse	2,3 MeV	1,275 GeV	173,07 GeV	0	125,09 GeV
Ladung	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
Spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
Name	u up	c charm	t top	γ Photon	H Higgs Boson
	4,8 MeV	95 MeV	4,18 GeV	0	
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
Quarks	d down	s strange	b bottom	g Gluon	
	<2 eV	<0,19 MeV	<18,2 MeV	91,2 GeV	
	0	0	0	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
	ν_e Elektron- Neutrino	ν_μ Myon- Neutrino	ν_τ Tau- Neutrino	Z^0 Z Boson	
	0,511 MeV	105,7 MeV	1,777 GeV	80,4 GeV	
	-1	-1	-1	± 1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
Leptonen	e Elektron	μ Myon	τ Tau	W^\pm W Boson	Eichbosonen

Abbildung 2.1.: Standardmodell der Teilchenphysik. Es umfasst alle bekannten fundamentalen Teilchen. Es sind die drei Generationen der Fermionen in der ersten bis dritten Spalte, Eichbosonen in der vierten Spalte und das Higgs-Boson in der fünften Spalte dargestellt [4]. Dabei ist für jedes Teilchen Masse, Spin und Ladung angegeben.

Alle Elementarteilchen werden durch fundamentale Kräfte beeinflusst. Das Standardmodell beschreibt die elektromagnetische, starke und schwache Wechselwirkung. Die Schwerkraft ist nicht enthalten, aber sie kann auf mikroskopischen Skalen wegen ihrer geringen Stärke im Vergleich zu den anderen Kräften vernachlässigt werden. Die Wechselwirkungen des Standardmodells werden durch Quantenfeldtheorien beschrieben und ihre Vermittlung erfolgt über Eichbosonen. Es existieren drei Arten von Wechselwirkungen im Standardmodell:

- Elektromagnetische Wechselwirkung
Die elektromagnetische Wechselwirkung wird durch die Quantenelektrodynamik (QED) [5, 6] beschrieben. In dieser Theorie wird die elektromagnetische Wechselwirkung durch Photonen γ übertragen. Die Photonen besitzen keine Masse. Sie können nicht mit sich selbst wechselwirken, da sie keine Ladung tragen. Folglich können Photonen nur mit elektrischen geladenen Teilchen wechselwirken.

- **Starke Wechselwirkung**
Die starke Wechselwirkung wirkt in kleinen Abständen auf Quarks und Gluonen. Sie wird durch die $SU(3)_c$ Eichgruppe repräsentiert und durch die Theorie der Quantenchromodynamik (QCD) [7, 8] beschrieben. Die Eichbosonen dieser Wechselwirkung sind die Gluonen g , die masselos sind und keine elektrische Ladung tragen. Aufgrund der komplexen Symmetrie ist die Existenz von acht Gluonen notwendig. Sie tragen eine der drei Arten von Farbladungen (rot, grün oder blau) und eine der Antifarbladungen. Die QCD verbietet die Existenz von einzelnen freien farbgeladenen Objekten, sodass nur Mesonen (Quark-Antiquark-System) und Baryonen (System aus 3 Quarks) beobachtet werden können. Die Gluonen unterscheiden sich von Photonen, da sie mit sich selbst wechselwirken können, sodass Gluonen an alle farbgeladene Teilchen (Quarks und Gluonen) koppeln.
- **Schwache Wechselwirkung**
Die schwache Wechselwirkung wirkt in kleinen Abständen auf Quarks und Leptonen. Sie ist insbesondere an Zerfällen und Umwandlungen von Teilchen beteiligt und wird durch die elektroschwache Theorie [9] beschrieben. Dies ist eine Vereinigung der schwachen und elektromagnetischen Kraft. Die elektroschwache Theorie wird durch die kombinierte $SU(2)_L \otimes U(1)_Y$ Eichgruppe mit W^\pm - und Z^0 -Bosonen als massive Eichbosonen beschrieben.

Die Masse der Teilchen folgt im Standardmodell aus dem Higgs-Mechanismus [10]. Er beschreibt, dass die Masse von W^\pm -, Z -Bosonen und auch Fermionen durch Wechselwirkung mit einem weiteren Feld, dem Higgs-Feld, entsteht. Während das Higgs-Feld nicht direkt messbar ist, tritt bei seiner Existenz ein weiteres Elementarteilchen auf, das Higgs-Boson. Im Jahr 2012 wurde das Higgs-Boson durch die Experimente ATLAS und CMS am LHC entdeckt [11, 12]. Dadurch wird das Standardmodell vervollständigt. Es besitzt eine Masse von etwa 125 GeV und ist ein einzelnes Teilchen mit Spin 0. Mit der Entdeckung des Higgs-Bosons wurde der Higgs-Mechanismus verifiziert. Das Top-Quark hat eine besonders starke Kopplung an das Higgs-Feld, da es das massivste aller besonderer Elementarteilchen ist, und ist somit von beobachteter Bedeutung für die Untersuchung der Eigenschaften des Higgs-Bosons.

2.2. Top-Quark

Top-Quarks können in der Natur nicht direkt beobachtet werden, da die mittlere Lebensdauer des Top-Quarks aufgrund seiner großen Masse sehr gering ist. Sie beträgt etwa 10^{-25} s. Die fünf leichteren Quarks sind durch die starke Wechselwirkung nur in gebundenen Zuständen zu finden und können nicht als freie Teilchen existieren. Aber das Top-Quark wird nicht in gebundenen Zuständen beobachtet, da es zerfällt, bevor es einen gebundenen Zustand bilden kann, sodass nur seine Zerfallsprodukte gemessen werden können. Das Top-Quark kann nicht nur in Paaren erzeugt werden, sondern auch einzeln. Im Folgenden werden die beiden Prozesse jeweils erklärt.

2.2.1. Top-Quark-Paarproduktion ($t\bar{t}$)

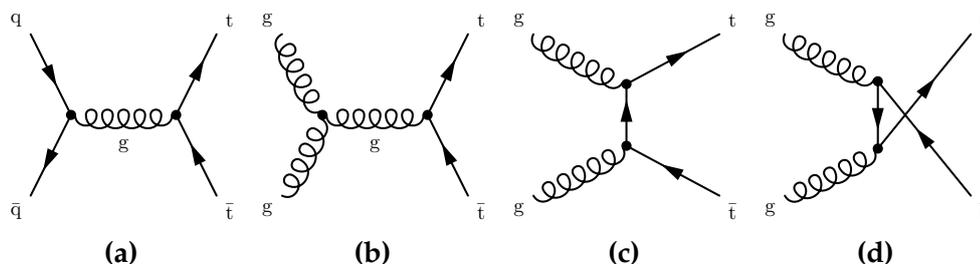


Abbildung 2.2.: Das Top-Quark-Paar entsteht über die starke Wechselwirkung. Im ersten Diagramm (a) ist die Produktion über die Auslöschung eines Quark-Antiquark-Paares dargestellt. Die anderen drei Diagramme (b-d) zeigen die Entstehung des Top-Quark-Paares über Gluon-Gluon-Fusion.

Der dominante Produktionsprozess für Top-Quarks ist die Paarproduktion über die starke Wechselwirkung. Top-Quark-Paare können entweder durch Quark-Antiquark-Vernichtung oder durch Gluon-Gluon-Fusion entstehen. Die relevanten Feynman-Diagramme sind in Abbildung 2.2 dargestellt.

- Im ersten Fall wird durch die Auslöschung eines Quark-Antiquark-Paares ein Top-Quark Paar über ein virtuelles Gluon produziert (siehe Abb. 2.2a).
- Im zweiten Fall vereinigen sich zwei reelle Gluonen zu einem virtuellen Gluon, das in ein Top-Quark Paar zerfällt (siehe Abb. 2.2b). Eine weitere Möglichkeit ist, dass die beiden einlaufenden Gluonen ein Top-Quark bzw. Top-Antiquark austauschen (siehe Abb. 2.2c, 2.2d).

Top-Quark-Paare können vollständig leptonisch, semileptonisch oder hadronisch zerfallen. Hadronische Zerfälle tragen in dieser Arbeit nicht zum Untergrund bei, da nur Ereignisse mit einem isolierten geladenen Lepton ausgewählt werden. Rein leptonische Zerfälle werden dagegen als relevanter Untergrund angesehen. Der Hauptuntergrund ist der semileptonische Top-Quark-Paar-Zerfall, da seine Charakteristik ähnlich zu der des s -Kanals ist (siehe Abb.2.3).

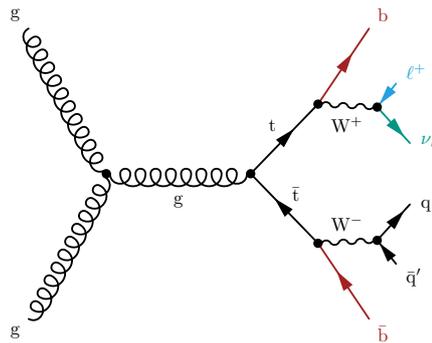


Abbildung 2.3.: Das Top-Quark-Paar entsteht über Gluon-Gluon-Fusion, wobei es semileptonisch zerfällt. Die Endprodukte bestehen aus einem geladenen Lepton, einem Neutrino, und vier Quarks. Zwei diese Quarks sind Bottom-Quarks.

2.2.2. Einzel-Top-Quark-Produktion

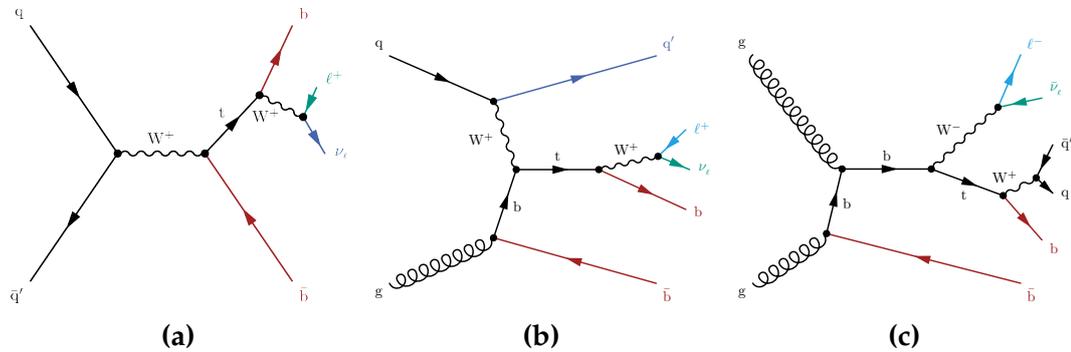


Abbildung 2.4.: Das einzelne Top-Quark entsteht über die elektroschwache Wechselwirkung. In der ersten Abbildung (a) ist der s -Kanal mit einem leptonischen Zerfall des W -Bosons dargestellt. Die zweite Abbildung (b) zeigt die Erzeugung des einzelnen Top-Quarks im t -Kanal. Die dritte Abbildung (c) zeigt den tW -Prozess, wobei das zusätzliche W -Boson hadronisch zerfällt.

Einzelne Top-Quarks werden über die elektroschwache Wechselwirkung erzeugt. Hierbei koppelt das Top-Quark an ein Bottom-Quark und ein W -Boson. Drei verschiedene einzelne Top-Quark-Prozesse existieren. Die relevanten Feynman-Diagramme sind in Abb. 2.4 dargestellt.

- s -Kanal
Zwei zum selben Isospin-Dublett gehörende Quarks werden zur Kollision gebracht. Dabei entstehen durch Austausch eines virtuellen W -Bosons ein Top-Quark und ein Bottom-Quark. Aufgrund der großen Masse des Top-Quarks zerfällt es direkt in ein W -Boson und ein weiteres Bottom-Quark. Das W -Boson zerfällt weiter in ein Lepton und ein entsprechendes Neutrino. Die Endprodukte des s -Kanals bestehen aus insgesamt zwei Bottom-Quarks, einem Lepton und einem Neutrino (siehe Abb. 2.4a).
- t -Kanal
Dieser Kanal ist der dominante Kanal für die Einzel-Top-Quark Produktion. Daher stellt der t -Kanal-Prozess in dieser Arbeit als den Hauptuntergrund dar. Das Bottom-Quark wird dabei durch Austausch eines virtuellen W -Bosons mit einem leichten Quark (Up- oder Down-Quark) in ein Top-Quark umgewandelt. Dadurch wird auch das leichte Quark in anderes leichte Quark umgewandelt. Das Top-Quark zerfällt wie beim s -Kanal-Prozess leptonisch.

- assoziiert tW -Produktion

In diesem Kanal werden ein Top-Quark und ein reelles W-Boson durch die Wechselwirkung eines Gluons und eines Bottom-Quarks erzeugt (siehe Abb. 2.4c). Der Endzustand besteht somit aus einem zusätzlichen W-Boson, das leptonisch oder in ein Quark-Antiquark-Paar zerfallen kann.

2.3. Simulation der Ereignisse

In der Analyse werden theoretischen Vorhersage mit den experimentellen Daten verglichen. Dies geschieht durch Simulation von Ereignissen mit der Monte-Carlo (MC)-Methode [13], die alle Berechnungen von den anfänglichen Zerfallsproduktionen bis zu den endgültig beobachtbaren Teilchen umfasst. MC-Ereignis-Generatoren berechnen für jeden gegebenen Prozess entsprechende Feynman-Diagramme. Jeder Prozess hat eine unendliche Anzahl an Diagrammen, da die zusätzlich hinzugefügten virtuellen Gluonen oder Photonen den Endzustand nicht verändern können. Die Amplitude der Diagramme verringert sich mit der Anzahl der Vertizes, wodurch der Prozess weniger wahrscheinlich ist. MC-Ereignis-Generatoren simulieren auch die Hadronisierung und den Partonschauer, da diese beiden Phänomene nach der Kollision von Teilchen im Detektor auftreten. Auf die von den MC-Ereignis-Generatoren simulierte Ereignisse wird das Programm GEANT4 [14] zur Simulation der Detektorantwort angewendet, sodass die simulierten Ereignisse mit den gemessenen Daten vergleichbar sind.

3. Experiment

Im Folgenden wird die experimentelle Umgebung beschrieben, mit der nach der Produktion einzelner Top-Quarks im s -Kanal gesucht wird. Zunächst wird ein Überblick über die verschiedenen Beschleunigeranlagen präsentiert, anschließend wird der Aufbau des CMS-Detektors erläutert.

3.1. Large Hadron Collider

Der Large Hadron Collider (LHC) [15] ist der größte und leistungsstärkste Teilchenbeschleuniger der Welt. Er hat einen Umfang von 27 km und befindet sich unterirdisch an der Grenze zwischen Frankreich und der Schweiz. Mit dem LHC wurde das Higgs-Boson erstmals nachgewiesen und das Standardmodell verifiziert.

Der LHC beschleunigt Protonen mit einer Schwerpunktenergie von 13 TeV und bringt sie zur Kollision. Bevor die Protonen in den LHC eingespeist werden, müssen sie beschleunigt werden (siehe Abb. 3.1). Sie werden zuerst durch den Linearbeschleuniger Linac2 auf 50 MeV beschleunigt. Anschließend wird der Protonenstrahl in den Super-Proton-Synchrotron-Booster (PSB) injiziert, der die Protonen auf 1,4 GeV beschleunigt, gefolgt vom Proton Synchrotron (PS), wo sie eine Energie von 25 GeV erreichen. Dann wird der Strahl zum Super-Proton-Synchrotron (SPS) weitergeleitet, wo er auf 450 GeV beschleunigt wird. Die Protonen werden schließlich gegenläufig in den LHC injiziert. Um den Protonenstrahl auf seiner Bahn innerhalb des Speicherrings zu halten, werden supraleitende Magnetspulen verwendet, welche zwei starke magnetische Dipolfelder erzeugen, die auf die geladenen Teilchen wirken. Nach ca. 20 Minuten erreichen die Protonen ihre maximale Energie von 6,5 TeV. Die Protonen werden an vier Orten innerhalb des LHC zur Kollision gebracht, an denen sich die Detektoren ATLAS, CMS, ALICE und LHCb befinden.

ATLAS und CMS sind Vielzweckdetektoren und haben ähnliche Ziele: Die Vermessung des Higgs-Bosons und die Suche nach neuen Teilchen sowie die

3. Experiment

Suche nach Dunkle-Materie. Allerdings unterscheiden sie sich in ihrem experimentellen Aufbau und sie werden von unabhängigen Arbeitsgruppen betrieben. Dadurch wird die Zuverlässigkeit der Messungen der beiden Detektoren überprüft. Das Ziel des ALICE-Experiments ist das tiefere Verständnis der starken Wechselwirkung in der Natur, indem das Quark-Gluon-Plasma studiert wird. Das LHCb-Experiment untersucht die Unterschiede zwischen Materie und Antimaterie, um Vorhersagen des Standardmodells zu überprüfen.

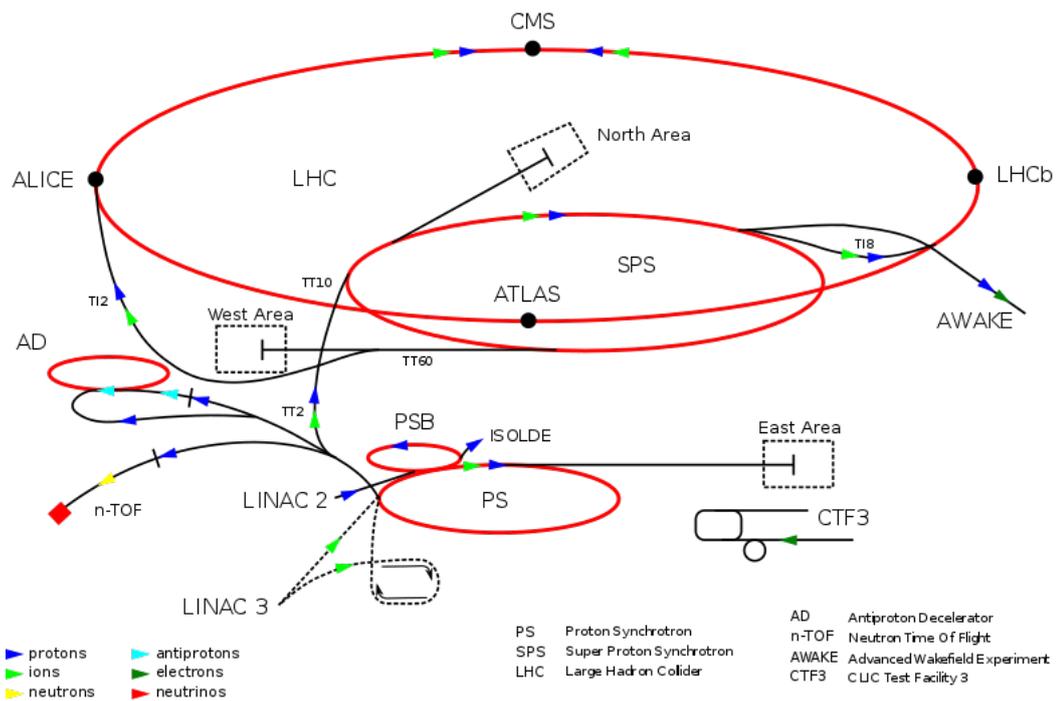


Abbildung 3.1.: Der Aufbau des LHC. Die Abbildung zeigt die verschiedenen Beschleuniger und Detektoren des LHC. Nach der Beschleunigung der Protonen werden sie in den LHC eingespeist. Die Protonenstrahlen fliegen in entgegengesetzter Richtung und werden später an vier Orten zur Kollision gebracht. [16]

Die Leistung eines Teilchenbeschleunigers wird durch die Luminosität L beschrieben. Sie beschreibt die Anzahl der Teilchenkollision pro Zeit und Fläche und wird durch die Ereignisrate \dot{N} bzw. den Wirkungsquerschnitt σ eines bestimmten Prozesses bestimmt

$$L = \frac{\dot{N}}{\sigma}. \quad (3.1)$$

Besitzt der Beschleuniger eine hohe Luminosität, werden mehr Ereignisse in einem bestimmten Zeitraum aufgezeichnet. Die Menge der aufgezeichneten Daten wird als integrierte Luminosität angegeben. Sie ergibt sich aus

$$L_{\text{int}} = \int_T L dt. \quad (3.2)$$

Im Jahr 2016 wurde am LHC eine Datenmenge von $35,9 \text{ fb}^{-1}$ mit dem CMS-Detektor gemessen. Die Anzahl der erwarteten Ereignisse N für einen bestimmten Prozess wird mit dessen Wirkungsquerschnitt berechnet:

$$N = L_{\text{int}} \cdot \sigma. \quad (3.3)$$

3.2. Compact Muon Solenoid

In dieser Arbeit geht es um die Analyse mit Daten des CMS-Experimentes [17]. Dabei werden die unterschiedlichen Teilchen, die bei der Kollision der Protonen entstehen, umfassend vermessen.

Ziel des Detektors ist es, Energie, Impuls, Ladung und Spur jedes Teilchens möglichst genau zu bestimmen. Dadurch können die physikalischen Ereignisse vollständig rekonstruiert werden. Die unterschiedlichen Teilchen werden anhand charakteristischer Eigenschaften im geeigneten Sub-Detektor nachgewiesen. Mit diesen Informationen lassen sich die Pseudorapidität η , der Transversalimpuls p_T und der Winkel ϕ von Teilchen ermitteln. Die Pseudorapidität η gibt dabei den Winkel ϕ einer Trajektorie relativ zur Strahlachse an:

$$\eta = -\ln\left[\tan\left(\frac{\phi}{2}\right)\right]. \quad (3.4)$$

Der Winkelbereich mit großen Werten für η wird als Vorwärtsrichtung bezeichnet. Wenn der Wert von η gleich Null ist, fliegt das Teilchen senkrecht zur

Strahlachse. Der transversale Impuls p_T beschreibt den Betrag des Impulses senkrecht zur Strahlachse,

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad (3.5)$$

wobei die z -Achse in Richtung des Protonenstrahls zeigt. In Abbildung 3.2 ist der Querschnitt des CMS-Detektors schematisch dargestellt. Im Folgenden werden die Komponenten des CMS-Detektors beschrieben.

- Spurdetektoren

Im Inneren des CMS-Detektors befinden sich die Spurdetektoren aus Silizium, mit denen die Bahnen geladener Teilchen bestimmt werden. Dabei kommen zwei verschiedene Spurdetektoren zum Einsatz. Die Pixeldetektoren liegen nahe am Kollisionspunkt und können die Position von Teilchen mit einer Genauigkeit von 0,01mm bestimmen. Die Streifendetektoren, die eine Genauigkeit von 0,1mm ermöglichen, vermessen die radiale Richtung. In den Spurdetektoren lassen sich die Spuren der elektrischen geladenen Teilchen nachweisen, indem kleine elektrische Signale durch die Erzeugung freier Ladungsträger induziert werden. Dieses Signal lässt sich weiter verstärken und zum Schluss detektieren.

- Kalorimeter

Die Spurdetektoren sind von dem elektromagnetischen (ECAL) und dem hadronischen Kalorimeter (HCAL) umgeben, welche für die Messung der Teilchenenergie verantwortlich sind. Wenn die in der Proton-Proton-Kollision entstandenen Teilchen mit dichter Materie wechselwirken, kann ein Teilchenschauer erzeugt werden. Das primäre Teilchen erzeugt bei dieser Wechselwirkung neue Teilchen mit niedrigerer Energie, die in gleicher Weise wieder wechselwirken können, bis die verbleibenden Teilchen keine ausreichende Energie zur Bildung weiterer Teilchen haben. Das ECAL misst die Teilchenenergie von Elektronen und Photonen mit hoher Genauigkeit durch Absorption dieser Teilchen in einem Szintillator-Material, welches aus etwa 80 000 Bleiwolframat-Kristallen (PbWO_4) besteht. Das HCAL umschließt das ECAL. Hadronen werden nach Durchfliegen des ECAL am HCAL gestoppt. Dabei entsteht ein hadronischer Schauer, der durch das HCAL detektiert werden kann. Das HCAL ist abwechselnd aus Absorberschichten und Szintillatorschichten aufgebaut. In den Absorberschichten finden die Wechselwirkungsprozesse von Hadronen statt. Dabei entstehende Teilchen werden durch die nachfolgende Szintillatorschicht

nachgewiesen.

Bei der Hadronisierung erzeugen Quarks kollimierte Strahlenbündel (Jets), welche durch die Informationen aus ECAL, HCAL und Spurdetektor rekonstruiert werden.

- **Magnetspule und Magnetjoch**

Die Magnetspule umgibt die Kalorimeter, sodass die geladenen Teilchen im Inneren des CMS-Detektors durch deren Magnetfeld abgelenkt werden. Durch Messung der Krümmung der Trajektorien wird das Verhältnis von Impuls und Ladung ermittelt. Der Magnet ist 13 m lang und besitzt einen Durchmesser von 6 m. Durch Abkühlung auf -270 °C bildet seine supraleitende Spule ein Magnetfeld mit einer Feldstärke von bis zu 4 T aus.

Das rote Magnetjoch umschließt die Magnetspule und sorgt für geschlossene Magnetfeldlinien. Dadurch werden Myonen, die die Myonkammern passieren, abgelenkt.

- **Myonkammern**

Die Myonkammern bilden die äußerste Schicht des CMS-Detektors. Sie decken eine Fläche von insgesamt $18\,000\text{ m}^2$ ab. Aufgrund ihrer hohen Masse im Vergleich zu Elektronen, wechselwirken Myonen nur schwach mit dem Detektormaterial und gelangen als einzige geladene Teilchen zu den Myonkammern. In den Myonkammern ionisieren die Myonen das Gas und lösen ein elektrisches Signal aus.

3. Experiment

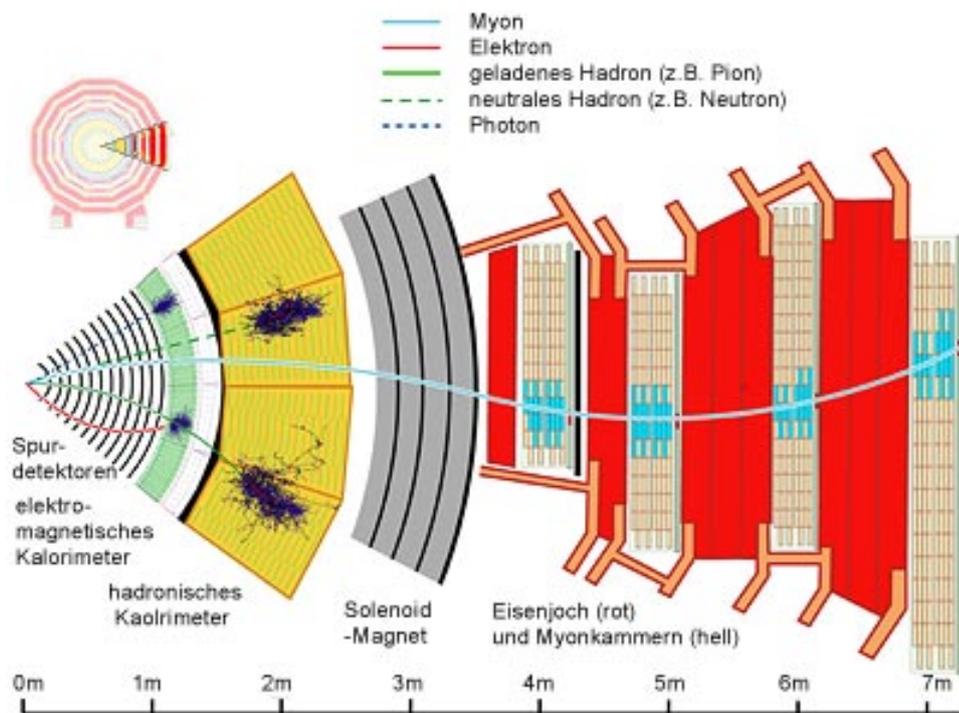


Abbildung 3.2.: Querschnitt des CMS-Detektors. Die Abbildung zeigt den Aufbau des CMS-Detektors. Die unterschiedlichen Teilchen lassen sich in den geeigneten Komponenten detektieren. [18]

4. Statistische Methoden

Aufgrund des geringen Wirkungsquerschnitts des s -Kanal-Prozesses bei Proton-Proton-Kollisionen sind spezielle Methoden erforderlich, um die s -Kanal-Produktion von den viel häufigeren Untergrundereignissen unterscheiden zu können. Hierfür wird eine multivariate Analyseverfahren, die tiefen neuronalen Netze (Deep Neural Networks, DNNs), verwendet. In diesem Kapitel werden die Grundlagen von DNNs erläutert, anschließend werden der Kolmogorov-Smirnov-Test, welche die Klassifizierung des DNNs bewertet, sowie die Signifikanz, welche die Wahrscheinlichkeit angibt, dass der s -Kanal-Prozess gemessen wurde, beschrieben.

4.1. Tiefe neuronale Netze

Bei multivariaten Analyseverfahren (MVAs) trainiert das Programm die Trennung von Signal und Untergrund mit einem bekannten Datensatz und lernt somit, unbekannte Daten zu klassifizieren. In den letzten Jahren hat sich gezeigt, dass ein DNN Vorteile in der Objektklassifizierung gegenüber etablierten multivariaten Methoden hat. Im Folgenden werden der Aufbau von DNNs und die zugehörigen Algorithmen [19] beschrieben.

DNNs sind inspiriert durch das menschliche Gehirn und bestehen aus Neuronen, auch Knoten genannt. Die Neuronen können die Informationen von anderen Neuronen aufnehmen und modifiziert an andere Neuronen weiterleiten oder als Endergebnis ausgeben. Sie sind schichtweise angeordnet. Zwischen den Schichten ist jedes Neuron der einen Schicht immer mit allen Neuronen der nächsten Schicht verbunden. Dabei ist der Ausgabewert des einen Neurons der Eingabewert des nächsten. Die Eingabewerte der Neuronen werden zunächst mit Gewichten w multipliziert und danach aufsummiert. Die Gewichte sind reelle Zahlen und in der Lage, die Bedeutung der jeweiligen Eingabewerte für die Ausgabewerte auszudrücken. Je stärker die Gewichtung, desto größeren Einfluss kann ein Neuron über die Verbindung auf ein anderes Neuron ausüben.

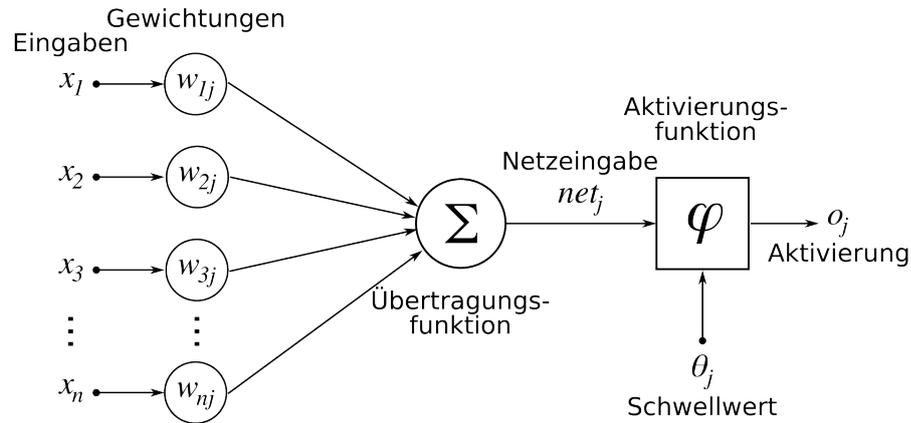


Abbildung 4.1.: Bearbeitung der Informationen an einem Neuron. Die Eingabewerte werden zunächst durch die Übertragungsfunktion gewichtet und summiert, anschließend wird aus dieser Summe die Netzeingabe des Neurons berechnet. Die Aktivierungsfunktion bestimmt schließlich den Ausgabewert des Neurons. Der Schwellwert ist ein konstantes Element und wird häufig weggelassen und durch ein zusätzliches Neuron mit einer konstanten Aktivierung von 1 ersetzt [20].

Dieser Einfluss wird durch ein lineares Modell beschrieben:

$$(y_1^{(l)}, \dots, y_n^{(l)} | w_{1j}^{(l)}, \dots, w_{nj}^{(l)}) \rightarrow \sum_{i=1}^n y_i^{(l)} w_{ij}^{(l)}, \quad (4.1)$$

wobei n die Anzahl der Neuronen in einer Schicht und l die Schicht bezeichnet. $y_i^{(l)}$ beschreibt ein Neuron i in Schicht l und $w_{ij}^{(l)}$ bezeichnet das Gewicht zwischen dem sendenden Neuron i und dem empfangenden Neuron j .

Um eine gute Klassifizierung zu erhalten, lassen sich die Gewichte zwischen Neuronen modifizieren. Allerdings kann eine kleine Änderung der Gewichte in dem Netzwerk zu einer Umkehr des Endergebnisses führen, da die Verbindung zwischen den Neuronen nur durch ein lineares Modell beschrieben wird. Die gewichteten und summierten Eingabewerte werden mittels einer Aktivierungsfunktion modifiziert. Durch diese Änderung der Eingabewerte wird eine kleine Abweichung der Ausgabe verursacht. In dieser Arbeit wird die Tangens-Hyperbolicus-Funktion verwendet. Dabei ergibt sich der Ausgabewert y_{NN} aus

$$y_{NN} = \tanh\left(\sum_{i=1}^n y_i^l w_{ij}^{(l)}\right). \quad (4.2)$$

Die in Abbildung 4.1 dargestellte Aktivierung des Neuronen wird durch obige Gleichung 4.2 beschrieben. Je nach Wert der Netzeingabe ist der Ausgabewert der Aktivierungsfunktion durch eine Verschiebung nahe 0 bzw. 1. Wenn sich

das Endergebnis nahe 0 befindet, wird das betrachtete Ereignis als Untergrund klassifiziert, umgekehrt wird es als Signalereignis erkannt. Der in Abbildung 4.1 beschriebene Ausgabewert des Neurons ist gültig für neuronale Netze, die nur eine Schicht zwischen den Eingabe- und Ausgabeneuronen besitzen. DNNs bestehen dagegen aus mehreren Schichten. Dennoch wird der Ausgabewert analog zu einem einfachen Netz bestimmt.

Abbildung 4.2 stellt die Struktur eines DNNs schematisch dar. Die Eingabeschicht nimmt die Informationen, in dieser Arbeit in Form von Verteilungen verschiedener kinematischer Größen, auf und gibt die gewichteten Eingangssignale an die Neuronen der ersten versteckten Schicht weiter. Zwischen der Eingabe- und Ausgabeschicht befindet sich in jedem neuronalen Netz mindestens eine versteckte Schicht. Theoretisch ist die Anzahl der versteckten Lagen in DNNs unbegrenzt. Allerdings bewirkt jede eingefügte versteckte Schicht einen Anstieg der benötigten Rechenleistung. Die Ausgabeschicht liegt hinter den versteckten Schichten und bildet die letzte Schicht des DNNs. Sie besteht aus einem einzelnen Neuron und stellt den Endpunkt des Informationsflusses in DNNs dar. Sie gibt das Ergebnis, in dieser Arbeit in Form von einer Verteilung, die bestimmt, wie signal- oder untergrundartig die Ereignisse sind, aus. In DNNs findet der Informationsfluss ausschließlich vorwärtsgerichtet von der Eingabeschicht über die versteckten Lagen zur Ausgabeschicht statt.

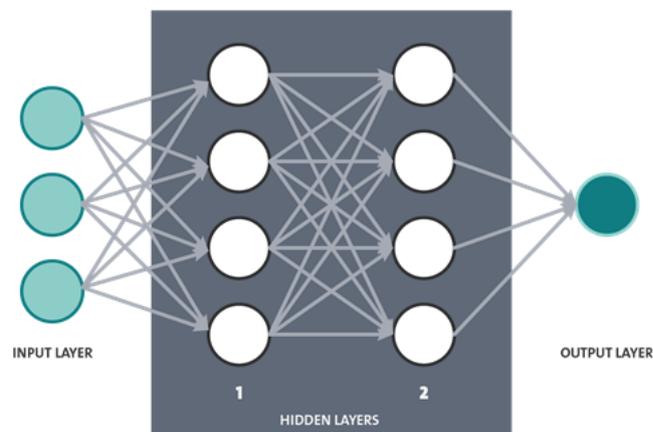


Abbildung 4.2.: Dieses vorwärts gekoppelte neuronales Netz besteht aus einer Eingabeschicht, zwei versteckten Schichten und einer Ausgabeschicht [21].

4.1.1. Training des tiefen neuronalen Netzes

Durch das Modifizieren der Gewichte wird das Endergebnis des DNNs geändert. Ziel ist es, die Gewichte so zu verändern, dass die Fehlerfunktion E minimiert wird. Die Fehlerfunktion gibt an, wie stark die mit den DNN erzielte Klassifikation der Ereignisse von der erwarteten Verteilung von Signal- und Untergrundereignissen abweicht. Sie ist wie folgt definiert:

$$E(\mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{w}) = \sum_{a=1}^N E_a(\mathbf{x}_a | \mathbf{w}) = \sum_{a=1}^N \frac{1}{2} (y_{\text{NN},a} - \hat{y}_a)^2 \quad (4.3)$$

wobei das Netz auf Basis von N Ereignissen $\mathbf{x}_a = (x_1, \dots, x_{n_{\text{var}}})$, wobei $a=1, \dots, N$, mit n_{var} unterschiedlichen Variablen trainiert wird. Für jedes Trainingsereignis a wird eine entsprechende Ausgabe $y_{\text{NN},a}$ des DNNs im Vergleich zum wahren Ausgabewert $\hat{y}_a \in \{0, -1\}$ berechnet. \mathbf{w} sind die einstellbaren Gewichte des Netzes, welche die Fehlerfunktion minimieren sollen. Durch Minimierung der Fehlerfunktion wird eine genauere Klassifikation des DNNs erreicht.

Mittels des Gradientenverfahrens wird die Richtung zur Änderung der Gewichte ermittelt. Bei DNNs beginnt dieser Algorithmus entgegen dem Informationsfluss mit den Gewichten der Ausgabeschicht. Am Anfang werden die Gewichte $\mathbf{w}^{(1)}$ zufällig ausgewählt. Anschließend werden die Gewichte Schicht für Schicht aktualisiert, bis die Eingangsgewichte erreicht werden, indem sie um eine kleine Strecke im \mathbf{w} -Raum in Richtung $-\nabla_{\mathbf{w}} E$ verschoben werden, in der E am stärksten abnimmt:

$$\mathbf{w}^{(1+1)} = \mathbf{w}^{(1)} - \eta \cdot \nabla_{\mathbf{w}} E. \quad (4.4)$$

Dabei ist η die Lernrate. Dieser rückwärts propagierte Ablauf [22] wird Training genannt, da die Netze dadurch die Klassifizierung von Daten erlernen und ein präziseres Endergebnis ausgeben können.

4.1.2. Parameter des tiefen neuronalen Netzes

Um das DNN zu optimieren, lassen sich mehrere Bereiche seines Aufbaus modifizieren, zum Beispiel die Anzahl der Neuronen und Schichten des Netzes sowie die verwendete Aktivierungsfunktion. Weitere Parameter, die variiert werden können, sind die Lernrate, die Konvergenzschritte und die Trainingswiederholungen. Die Regularisierungsmethode, Dropout [23], wird zur Vermeidung des sogenannten Overtrainings, das Auswendiglernen von Datensätzen, verwendet. Die wichtigsten DNN-Parameter sind:

- Die Anzahl der Neuronen
Ein Netz mit zu wenigen versteckten Neuronen ist nicht flexibel genug, um Nichtlinearitäten in den Daten darzustellen. Bei zu vielen versteckten Neuronen können die Gewichte auf Null reduziert werden, sodass sich die erforderliche Rechenleistung erhöht.
- Die Lernrate
Die Lernrate beträgt normalerweise $\eta < 1$. Wenn die Lernrate niedrig ist, ist das Training zuverlässiger, aber die Optimierung nimmt viel Zeit in Anspruch, da die Schritte bis zum Minimum der Fehlerfunktion sehr klein werden. Wenn die Lernrate hoch ist, können die Gewichtsänderungen so groß sein, dass das Minimum der Fehlerfunktion bei der Optimierung überschritten wird.
- Die Konvergenzschritte (Convergence Steps)
Die Konvergenzschritte beschreiben, wie viele Trainingsschritte ohne Verbesserung der Minimierung durchgeführt werden müssen, bevor das Netzwerk als konvergent deklariert wird. Bei einer geringen Anzahl der Konvergenzschritte wird die Gefahr des Overtrainings verringert.
- Dropout
Beim Dropout wird beim Training des Netzwerks eine vorher bestimmte Anzahl von Neuronen in bestimmten versteckten Schichten des Netzwerks deaktiviert, sodass sie keinen Einfluss auf die weiteren Berechnungsschritte haben.

4.2. Kolmogorov-Smirnov-Test

Der Kolmogorov-Smirnov-Test (KS-Test) ist ein statistischer Test, mit dem die Stichprobenereignisse auf Übereinstimmung mit einer vermuteten Verteilung überprüft werden können.

Dabei wird jeweils eine empirische Verteilungsfunktion der DNN-Ausgabewerte des Trainingsdatensatzes und eines unabhängigen Testdatensatzes erstellt, um eine Klassifikation eines DNNs auf Overtraining zu testen. Die DNN-Ausgabewerte x_i der Ereignisse i werden nach ihrer Größe sortiert und die empirische Verteilungsfunktion wird durch

$$F(x) = \frac{\text{Anzahl der } x_i \text{ Werte } \leq x}{n} \quad (4.5)$$

konstruiert, wobei n die Anzahl der Ereignisse ist und $x \in \mathbb{R}$. $F(x)$ ist eine Treppenfunktion mit $F(x) = 0$ für $x < x_{\min}$ und $F(x) = 1$ für $x > x_{\max}$. Gesucht wird die maximale Differenz zwischen den beiden Verteilungen, auch Testgröße t genannt:

$$t = \sqrt{n} \cdot \max|F_{\text{train}} - F_{\text{test}}|. \quad (4.6)$$

Je kleiner dieser Wert ist, desto besser stimmen die beiden Verteilungen überein. Bei großen Werten von t unterliegen die Trainings- und Testdaten unterschiedlichen Verteilungen.

In dieser Arbeit wird anstelle von t die Wahrscheinlichkeit berechnet, einen Wert $t \leq t_0$ zu erhalten, wobei t_0 ein kritischer Wert ist und besagt, ob zwei Verteilungen miteinander kompatibel sind. Sie beschreibt die Ähnlichkeit zwischen den Trainings- und Testdaten mit einem Wert zwischen 0% und 100%. Wenn dieser Wert niedrig sind, sind die beiden Verteilungen nicht kompatibel. Somit ist es mithilfe des KS-Tests möglich, das DNN-Training auf Overtraining zu überprüfen.

4.3. Signifikanz

Bei einer Messung wird mittels der Signifikanz festgelegt, ob das Messergebnis eine Fluktuation des Untergrundes darstellt, also die Null-Hypothese erfüllt, oder ob die gemessenen Daten einem Signal entsprechen und somit die Signal + Untergrund-Hypothese erfüllen. Je höher die Signifikanz ist, desto besser ist die Trennung von Signal- und Untergrundeignissen. Die Verteilung vieler statistischer Tests basiert auf der Normalverteilung oder kann aus der Normalverteilung abgeleitet werden, sodass daraus die Signifikanz berechnet werden kann. Eine typische Eigenschaft der Normalverteilung ist, dass 68% ihrer Beobachtungen in den symmetrischen Bereich von 1σ und 95% in den Bereich von 2σ um dem Mittelwert fallen, wobei σ die Standardabweichung ist. In der Teilchenphysik spricht man von einer Evidenz, d.h. einer hohen Wahrscheinlichkeit, ein Signal gefunden zu haben, wenn die Signifikanz 3σ beträgt. Bei einer Signifikanz von 5σ spricht man von der Entdeckung eines Prozesses, da die Wahrscheinlichkeit einer Untergrundfluktuation lediglich 0,00006% beträgt.

5. Suche nach der Einzel-Top-Quark-Produktion im s -Kanal

Im Folgenden wird eine neue Analyse zur Verbesserung der Suche nach einzelnen Top-Quarks im s -Kanal vorgestellt. Diese Analyse basiert auf den Daten aus dem Jahr 2016, gemessen mit dem CMS-Detektor am LHC bei einer Schwerpunktsenergie von 13 TeV. Die gemessenen Daten entsprechen einer integrierten Luminosität L von $35,9 \text{ fb}^{-1}$.

Die grundsätzliche Strategie für die Suche nach dem s -Kanal-Prozess wird in Abschnitt 5.1 vorgestellt. Anschließend wird eine mögliche Optimierung zur Referenzanalyse [24] in Abschnitt 5.2 beschrieben. In den Abschnitten 5.3 und 5.4 wird die Durchführung der Klassifizierung mittels DNNs in unterschiedlicher Ereignisregionen beschrieben. Dabei werden die Klassifikationen aus verschiedenen Regionen verglichen. Zum Schluss lässt sich die Signifikanz der neuen Analyse ermitteln, um zu testen, ob sie die Referenzanalyse verbessert.

5.1. Strategie

Die Erzeugung einzelner Top-Quarks durch die elektroschwache Wechselwirkung ist wesentlich weniger wahrscheinlich als die Produktion von Top-Quark-Paaren über die starke Wechselwirkung. Darüberhinaus hat der s -Kanal den kleinsten Wirkungsquerschnitt der drei Einzel-Top-Quark-Produktionsmechanismen, die in Kapitel 2.2.2 beschrieben wurden. Daher ist es erforderlich, die s -Kanal-Ereignisse als Signal von dem viel häufigeren Untergrund zu trennen.

Zunächst werden die physikalischen Objekte mittels Daten, die mit dem CMS-Detektor gesammelt wurden, rekonstruiert. Besonders wichtig ist die Rekonstruktion von Quarks und ihre Zuordnung zu gebündelten Teilchenstrahlen, welche als Jets bezeichnet werden. Aufgrund des Confinements dürfen bei einer Kollision zwei Quarks nicht auseinandergerissen werden, sondern es werden

mehrere Quark-Antiquark-Paare erzeugt, welche einen Jet bilden. Jet, der von b-Tagging-Algorithmen als von einem Bottom-Quark stammend identifiziert wurde, wird 'b-tagged' (b-Tag) genannt. Anschließend werden basierend auf den Endzustand des s -Kanal-Prozesses Ereignisse mit einem isolierten geladenen Lepton sowie zwei b-Tags ausgewählt, um die Untergrundbeiträge zu reduzieren. Allerdings ist die Ereignis Selektion allein nicht ausreichend, um den s -Kanal-Prozess eindeutig vom Untergrund zu trennen. Zur Optimierung der Trennung werden multivariate Analysemethoden, DNNs, auf ausgewählte Ereignisregionen angewendet. Kinematische Variablen, deren Verteilungen sich in den Signal- und Untergrundprozessen unterscheiden, werden rekonstruiert und für das DNN-Training verwendet. Die durch das DNN-Training ermittelte Ereignisklassifikation wird anschließend auf alle für diese Analyse relevanten Untergrundprozesse angewendet. Es wird überprüft, ob die DNN-Ausgabeverteilung der gemessenen Daten mit der Verteilung der simulierten Ereignisse übereinstimmt. Zum Schluss wird die Signifikanz unter Berücksichtigung der systematischen Unsicherheiten berechnet, um festzustellen, ob die gemessenen Daten den gesuchten s -Kanal-Prozess beinhalten.

5.2. Studien zur Optimierung der Referenzanalyse

Durch die Auswahl der Ereignisregion wird der Beitrag der Untergrundeignisse reduziert. In der Referenzanalyse wird das DNN-Training in der sogenannten 2-Jet-2-Tag-Ereignisregion (2j2t-Region) durchgeführt. In dieser Region befinden sich nur Ereignisse, die zwei Jets aus einem Bottom-Quark sowie ein isoliertes Lepton aufweisen. Dadurch ist in dieser Region das Signal-zu-Untergrund-Verhältnis am größten. Der s -Kanal-Prozess wird als Signal und semileptonische und dileptonische $t\bar{t}$ -Prozesse als Untergrund für das DNN-Training verwendet, da die Top-Quark-Paarproduktion den Hauptuntergrund für die s -Kanal-Produktion darstellt. In dieser Arbeit werden dagegen t -Kanal-Ereignisse anstatt der $t\bar{t}$ -Produktion als Untergrund für das DNN-Training in der Referenzanalyse verwendet, da der t -Kanal der dominante Prozess zur Erzeugung einzelner Top-Quarks ist und somit schwierig vom s -Kanal-Prozess zu unterscheiden ist. Diese Arbeit fokussiert sich daher auf die Trennung vom s -Kanal- und t -Kanal-Prozess. Die Endprodukte des t -Kanal-Prozesses bestehen aus zwei Jets, wobei einer aus einem leichten Quark und der andere Jets aus einem Bottom-Quark stammt, einem geladenen Lepton sowie einem Neutrino. Folglich befinden sich die meisten t -Kanal-Ereignisse in der 2-Jet-1-Tag-Region (2j1t-Region). In dieser Arbeit werden DNNs jeweils in der 2j2t- und 2j1t-Region

trainiert. Anschließend werden die Verteilungen der DNN-Ausgabewerte auf ihre Trennung zwischen Signal und Untergrund sowie auf Overtraining überprüft. Die DNN-Klassifikation wird, wie in der Referenzanalyse, auf die Daten und Monte-Carlo-Ereignisse angewendet und es wird überprüft, ob die Verteilung der Daten und der Monte-Carlo-Ereignisse kompatibel sind. Zum Schluss wird diese neue Klassifikation, die entweder in der 2j1t-Region oder der 2j2t-Region trainiert wurde, mit der DNN-Klassifikation der Referenzanalyse in den Kontrollregionen für den $t\bar{t}$ -Prozess (3-Jet-1-Tag, 3-Jet-2-Tag) kombiniert, um die Signifikanz zu ermitteln. Die in dieser Arbeit bestimmte Signifikanz wird schließlich mit der Referenzanalyse verglichen.

5.3. Studien in der 2-Jet-2-Tag-Region

Die 2j2t-Region ist die Signalregion, da sie nur Ereignisse beinhaltet, welche den gleicher Endzustand wie der gesuchte s -Kanal-Prozess besitzen. Folglich ist die Anzahl simulierter t -Kanal-Ereignisse in dieser Region allein nicht ausreichend. Für das DNN-Training ist jedoch eine ausreichend große Menge an Daten notwendig, um eine gute Klassifikation zu erzielen. Für das DNN-Training in der 2j2t-Region werden daher neben den t -Kanal-Ereignissen auch der $t\bar{t}$ -Prozess als Untergrund verwendet.

Für das Training werden zunächst geeignete kinematische Variablen ausgewählt und rekonstruiert. Dabei stehen insgesamt 19 Variablen aus der Referenzanalyse zur Verfügung, die sensitiv auf die Trennung von s -Kanal- und $t\bar{t}$ -Ereignisse sind. Um die Klassifikation von s -Kanal- und t -Kanal-Ereignissen zu optimieren, werden zwei neue Variablen in das DNN-Training aufgenommen. Die erste kinematische Variable ist die Differenz der Pseudorapidität η zwischen dem Bottom-Jet, welcher aus dem W -Boson stammt, und dem Bottom-Jet aus dem Top-Quark, $|\eta(b_w) - \eta(b_t)|$. Da der Endzustand des t -Kanal-Prozesses nur einen Bottom-Jet besitzt, wird der Jet, der aus einem leichten Quark stammt (light-flavor Jet), dem W -Boson zugeordnet. Da dieser Jet ein hohes η besitzt ($|\eta| > 2,4$), unterscheidet sich diese rekonstruierte Variable von der im s -Kanal-Prozess. Die zweite eingefügte Variable ist die Ladung des Leptons. Unterschiede in dieser Variable kommen daher zustande, dass das aus dem Top-Quark zerfallende W -Boson unterschiedlich im s -Kanal- und t -Kanal-Prozess geladen sein kann. Nach DNN-Training wurde beobachtet, dass der Unterschied der Ladung des Leptons im s -Kanal- und t -Kanal-Prozess nicht sehr groß ist. Die insgesamt 21 Variablen, die nach ihrer Trennungsstärke im Training sortiert

5. Suche nach der Einzel-Top-Quark-Produktion im s-Kanal

sind, sind in Tabelle 5.1 dargestellt.

Tabelle 5.1.: Variablen für die Klassifikation von s-Kanal- und Untergrundereignissen in der 2j2t-Region. Der härteste Jet bezeichnet dabei den Jet mit dem größten Transversalimpuls.

Rang	Variable	Bezeichnung
1	$\Delta\phi(t_s, b_1)$	$\Delta\phi$ zwischen dem aus dem zweithärtesten Bottom (b)-Jet rekonstruierten Top-Quark und dem härtesten b-Jet
2	Fox Wolfram #0	Nullte Ordnung des Fox-Wolfram-Moments des Ereignisses
3	$\Delta\phi(t_l, b_s)$	$\Delta\phi$ zwischen dem aus dem härtesten b-Jet rekonstruierten Top-Quark und dem zweithärtesten b-Jet
4	m_T^W	Transversale Masse des rekonstruierten W-Bosons
5	$m(t)$	Invariante Masse des rekonstruierten Top-Quarks
6	$\Delta R(t, b_W)$	ΔR zwischen dem rekonstruierten Top-Quark und dem b-Jet aus dem zeitartigen W-Boson
7	$ p_T(t) - p_T(b_w) $	Absolute Differenz der transversalen Impulse des rekonstruierten Top-Quarks und des b-Jets aus dem zeitartigen W-Boson
8	$\Delta\phi(t, b_1)$	$\Delta\phi$ zwischen dem rekonstruierten Top-Quark und dem härtesten b-Jet
9	M_{lb2}	Invariante Masse des Leptons und des zweihärtesten b-Jets
10	p_T^{miss}	Fehlender transversaler Impuls
11	p_T^{bb}	Vektorielle Summe der transversalen Impulse der beiden b-Jets des Ereignisses
12	$\Delta\phi(b_l, b_s)$	$\Delta\phi$ zwischen den beiden b-Jets des Ereignisses
13	$ \eta(l) - \eta(b_1) $	Absolute Differenz der Pseudorapidität zwischen dem geladenen Lepton und dem härtesten b-Jet
14	$ \eta(t_s) - \eta(b_1) $	Absolute Differenz der Pseudorapidität zwischen dem aus dem zweithärtesten b-Jet rekonstruierten Top-Quark und dem härtesten b-Jet
15	$p_T(l)$	Transversalimpuls des geladenen Leptons
16	$ \eta(t) - \eta(b_w) $	Absolute Differenz der Pseudorapidität zwischen dem rekonstruierten Top-Quark und dem b-Jet aus dem W-Boson
17	H_T	Skalare p_T Summe aller b-Jets, des isolierten geladenen Leptons und p_T^{miss} des Ereignisses
18	$\cos\theta$	Kosinus des Winkels zwischen dem Lepton aus dem Zerfall des Top-Quarks und dem b-Jet des zeitartigen W-Bosons
19	Fox Wolfram #3	Dritte Ordnung des Fox-Wolfram-Moments des Ereignisses
20	$ \eta(b_w) - \eta(b_t) $	Differenz der Pseudorapidität zwischen dem Bottom-Jet aus dem W-Boson und dem Bottom-Jet aus dem Top-Quark
21	$q(l)$	Ladung des Leptons

5.3.1. Klassifizierung

Das DNN-Training wird mit 21 Variablen in der $2j2t$ -Region durchgeführt. Anhand des geladenen Leptons des Endzustandes lassen sich die Ereignisse in zwei Kategorien unterteilen, den Myon- und den Elektron-Kanal. Dabei werden zunächst nur die Ereignisse des Myon-Kanals mit einem DNN klassifiziert.

In dieser Arbeit wird studiert, welche Anzahl an welchen Lagen und an Neuronen pro Schicht das DNN-Training optimiert. Das Training des DNNs, dessen Struktur aus drei versteckten Schichten und 25, 50, 50 Neuronen pro Schicht besteht, weist dabei den höchsten Wert der Receiver-Operating-Characteristic (ROC)-Kurve von 0,788 auf. Die ROC-Kurve beschreibt die Beziehung zwischen der Effizienz von Signalauswahl gegenüber der Effizienz von Untergrundunterdrückung. Der Wert des ROC-Integrals stellt die Fläche unter der ROC-Kurve dar und liegt zwischen 0 und 1. Je größer dieser Wert ist, desto besser wird das Signal von Untergrund getrennt. Die anderen entscheidenden Parameter des DNNs sind in Tabelle 5.2 dargestellt.

Tabelle 5.2.: Die Parameter des DNNs für die Klassifikation der s -Kanal-Ereignisse, und der t -Kanal- und $t\bar{t}$ -Produktion in der $2j2t$ -Region. Das TMVA-Softwarepaket wird verwendet [19].

Parameter	Wert
Layout	TANH 25, TANH 50, TANH 50, LINEAR
LearningRate	0,01
Momentum	0,8
Repetitions	1
ConvergenceSteps	30
DropConfig	0.0+0,2+0,4+0.0
Multithreading	True

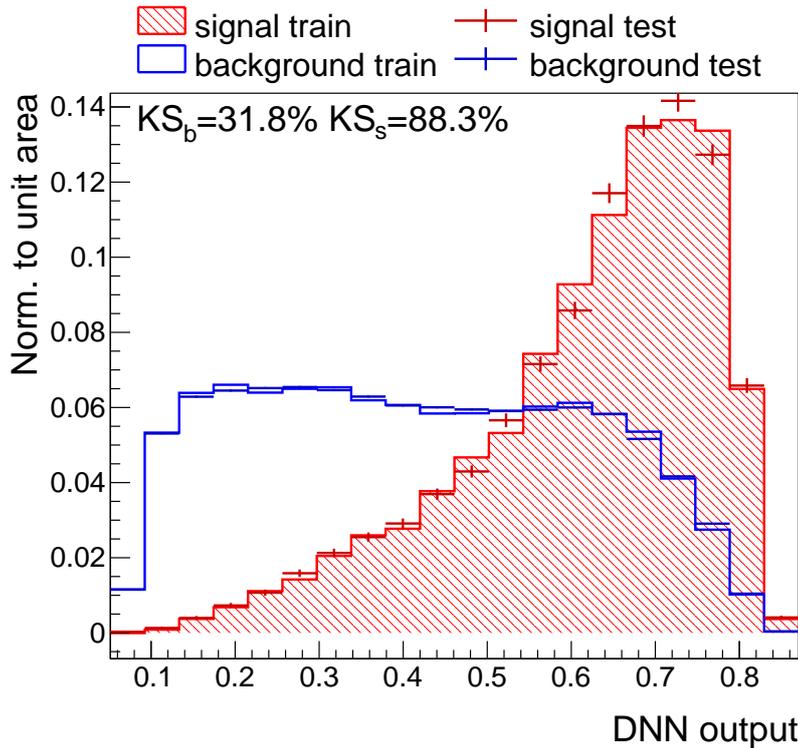


Abbildung 5.1.: Klassifikation der Ereignisse durch die DNN-Ausgabewerte in der 2j2t-Region. Die Testdaten stimmen gemäß den KS-Tests mit den Trainingsdaten überein. Jedoch ist das Signal nicht ausreichend vom Untergrund getrennt.

Die DNN-Ausgabevertellungen des Test- und Trainingsdaten und die entsprechenden Ergebnisse der KS-Tests sind in Abbildung 5.1, jeweils für Signal und Untergrund, dargestellt. Um Overtraining zu vermeiden, wird ein minimaler Wert von 10% für die KS-Tests gefordert. Es ist erkennbar, dass diese Trennung von Signal und Untergrund nicht ausreichend ist. Die Untergründereignisse sollten mehrheitlich einen DNN-Ausgabewert nahe 0 besitzen. Jedoch weist die Verteilung des Untergrunds kein Maximum in Richtung des Ausgabewertes Null auf, sondern entspricht einer flachen Verteilung.

5.3.2. Schlussfolgerung

Die optimierte Klassifikation wird auf die simulierten Ereignisse des s -Kanals, des t -Kanals und des $t\bar{t}$ -Prozesses angewendet. Damit lassen sich die Verteilungen der DNN-Klassifikation dieser Prozesse vergleichen. In Abbildung 5.2 wird beobachtet, dass die Verteilung des t -Kanals in der 2j2t-Region signalar-

tig ist, sodass der s -Kanal-Prozess vom t -Kanal-Untergrund nicht gut getrennt wird. Daher wird in dieser Arbeit getestet, ob ein DNN-Training in der $2j1t$ -Region, die von der t -Kanal-Produktion dominiert wird, eine bessere Trennung zwischen s - und t -Kanal-Prozess liefert.

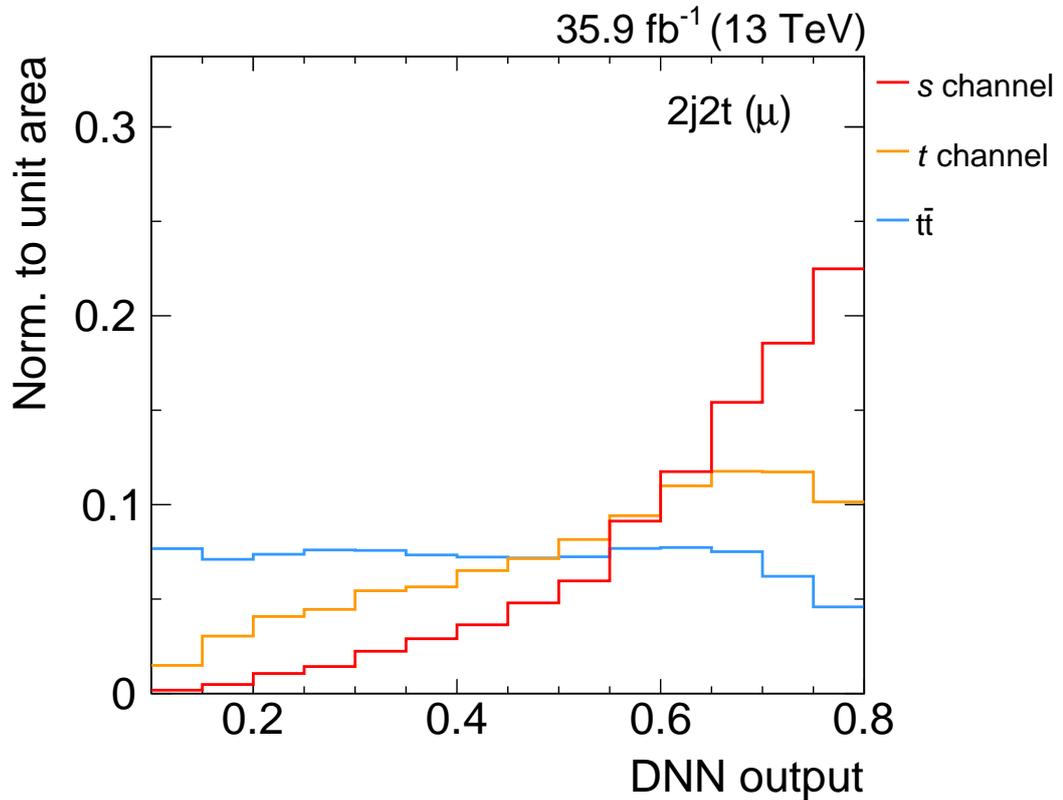


Abbildung 5.2.: DNN-Ausgabeverteilungen in der $2j2t$ -Region, für den Signal-Prozess sowie für den t -Kanal- und $t\bar{t}$ - Untergrund. Es wird beobachtet, dass die Verteilung des t -Kanals ähnlich der Verteilung des s -Kanals ist, sodass dieser Prozess schlecht vom Signalprozess trennbar ist.

5.4. Studien in der 2-Jet-1-Tag Region

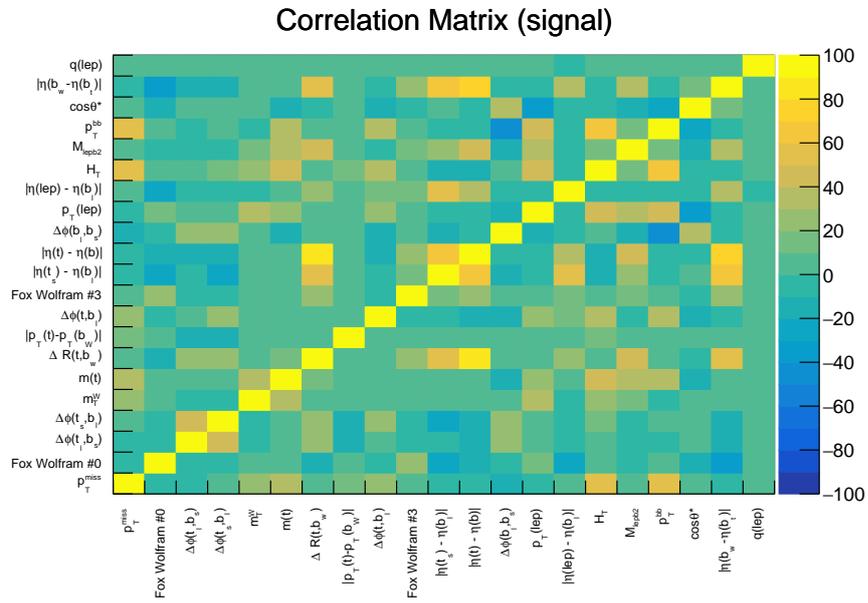
Da deutlich mehr simulierte t -Kanal-Ereignisse in der $2j1t$ -Region zur Verfügung stehen als in der $2j2t$ -Region, wird der t -Kanal-Prozess als einziger Untergrundprozess im DNN-Training verwendet.

5.4.1. Myon-Kanal

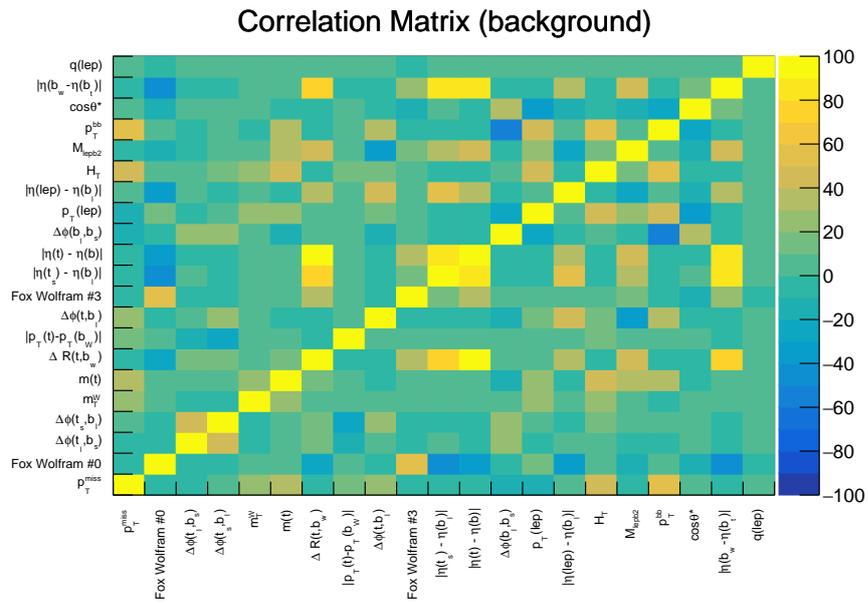
Zunächst wird die Modellierung der in Abschnitt 5.3 verwendeten Variablen getestet. Dabei werden Korrelationen zwischen den einzelnen Variablen und die Übereinstimmung der Daten mit simulierten Ereignissen untersucht. Variablen mit hoher Korrelation werden nicht im DNN-Training berücksichtigt, da sie keinen Informationsgewinn liefern und das Netzwerk eventuell instabil machen können. Das DNN-Training wird mit der gleichen Netz-Struktur und den gleichen Variablen wie im Myon-Kanal der 2j2t-Region durchgeführt. Durch die Berechnung der Korrelationsmatrix aller Eingangsvariablen, welche im DNN-Training automatisch durchgeführt wird, lässt sich die Korrelation zwischen zwei Variablen ablesen. In Abbildung 5.3 und 5.4 sind die Matrizen des Signals und des Untergrunds gezeigt. Die Nebendiagonalelemente geben Aufschluss über die Korrelation zweier unterschiedlicher Variablen. Je näher der Wert des Matrixelementes bei 100 % liegt, desto höher ist die Korrelation zwischen zwei Variablen. In Tabelle 5.3 sind vier Variablen, deren Korrelation größer als 80% ist, gelistet.

Tabelle 5.3.: Kinematische Variablen mit einer Korrelation von mehr als 80%. Die Variable mit größter Trennungsstärke wird beibehalten, die restlichen verworfen.

Variablen	Korrelation (%)
$\Delta R(\mathbf{t}, \mathbf{b}_w), \eta(\mathbf{t}) - \eta(\mathbf{b}_w) $	96,86
$ \eta(\mathbf{t}_s) - \eta(\mathbf{b}_1) , \eta(\mathbf{t}) - \eta(\mathbf{b}_w) $	89,92
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) , \eta(\mathbf{t}_s) - \eta(\mathbf{b}_1) $	91,71
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) , \eta(\mathbf{t}) - \eta(\mathbf{b}_w) $	91,13

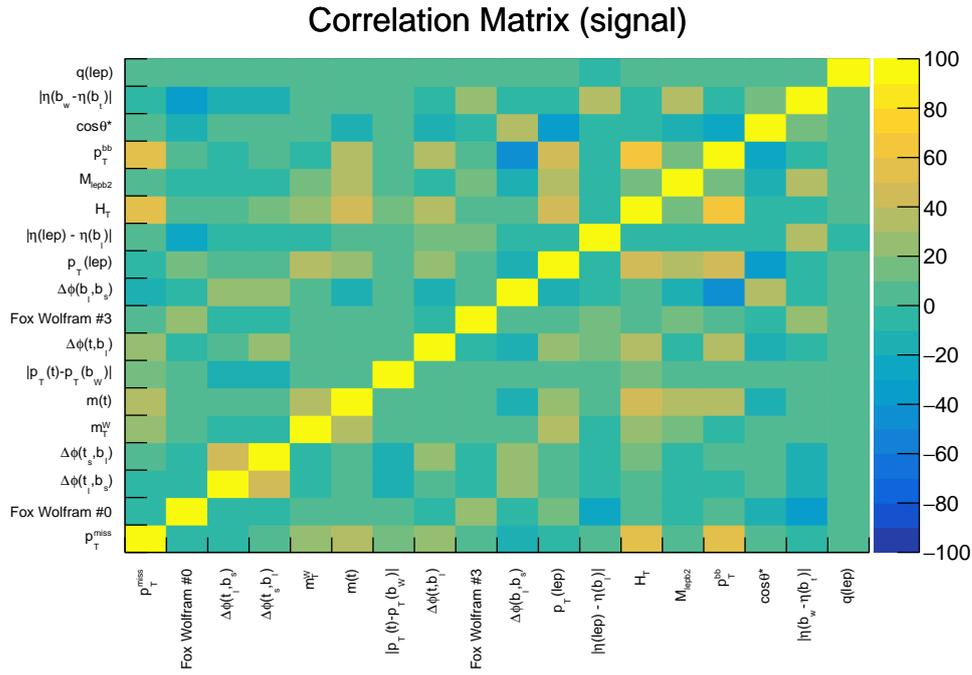


(a) Korrelationsmatrix des Signals in der 2j1t-Region

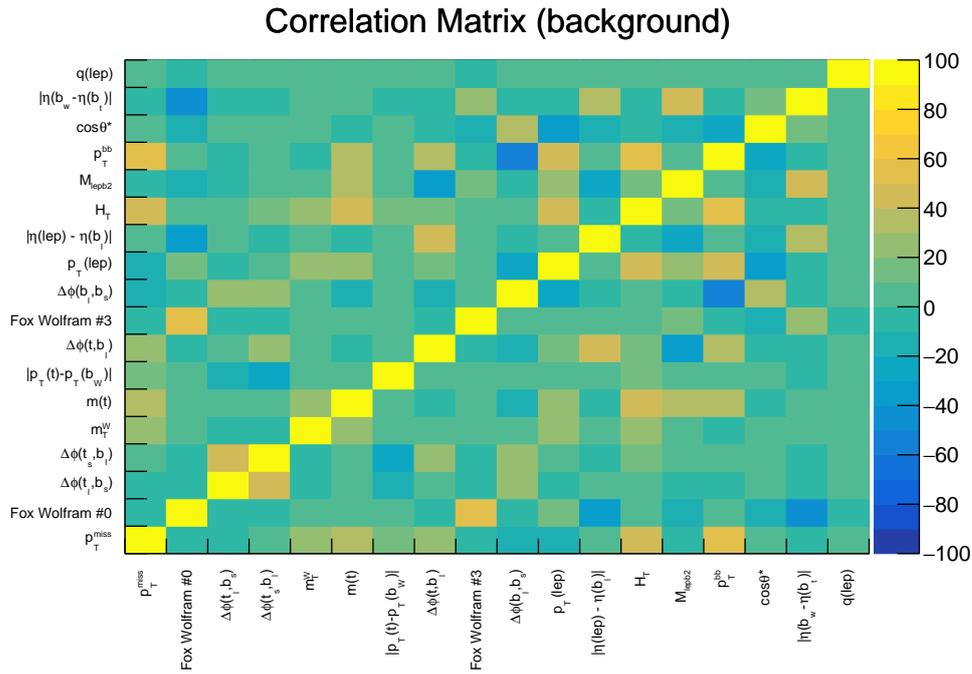


(b) Korrelationsmatrix des Untergrundes in der 2j1t-Region

Abbildung 5.3.: Die Korrelationsmatrizen des Signals und des Untergrundes vor Entfernung der hoch-korrelierten Variablen.



(a) Korrelationsmatrix des Signals in der 2j1t-Region



(b) Korrelationsmatrix des Untergrundes in der 2j1t-Region

Abbildung 5.4.: Die Korrelationsmatrizen des Signals und des Untergrundes nach Entfernung der hochkorrelierten Variablen.

Da die Variable $|\eta(\mathbf{b}_w) - \eta(\mathbf{b}_t)|$ die größten Beitrag zum DNN-Training liefert, wird diese Variable weiterhin verwendet und die drei anderen Variablen werden aus dem Training entfernt. Anschließend wird diese Klassifikation auf die vom DNN-Training unabhängigen Daten und auf alle simulierten Ereignisse, einschließlich sämtlicher Untergrundprozesse der Analyse angewendet. Es wird die Übereinstimmung zwischen Daten und simulierten Ereignissen für die verwendeten Eingangsvariablen und für die Verteilung der DNN-Ausgabe überprüft. Wenn die Daten innerhalb der Verteilung einer Variable von den simulierten Ereignissen um mehr als 10% abweichen, ist diese Variable nicht ausreichend gut modelliert und wird im Training nicht weiter berücksichtigt. Auf diese Weise lassen sich insgesamt fünf Variablen, $\Delta\phi(t_s - t_1)$, $\Delta\phi(t_1 - t_s)$, $|p_T(t) - p_T(\mathbf{b}_w)|$, H_T und Fox Wolfram #0 entfernen. Ihre Verteilungen sind in Abbildung 5.5 für die 2j1t-Region gezeigt.

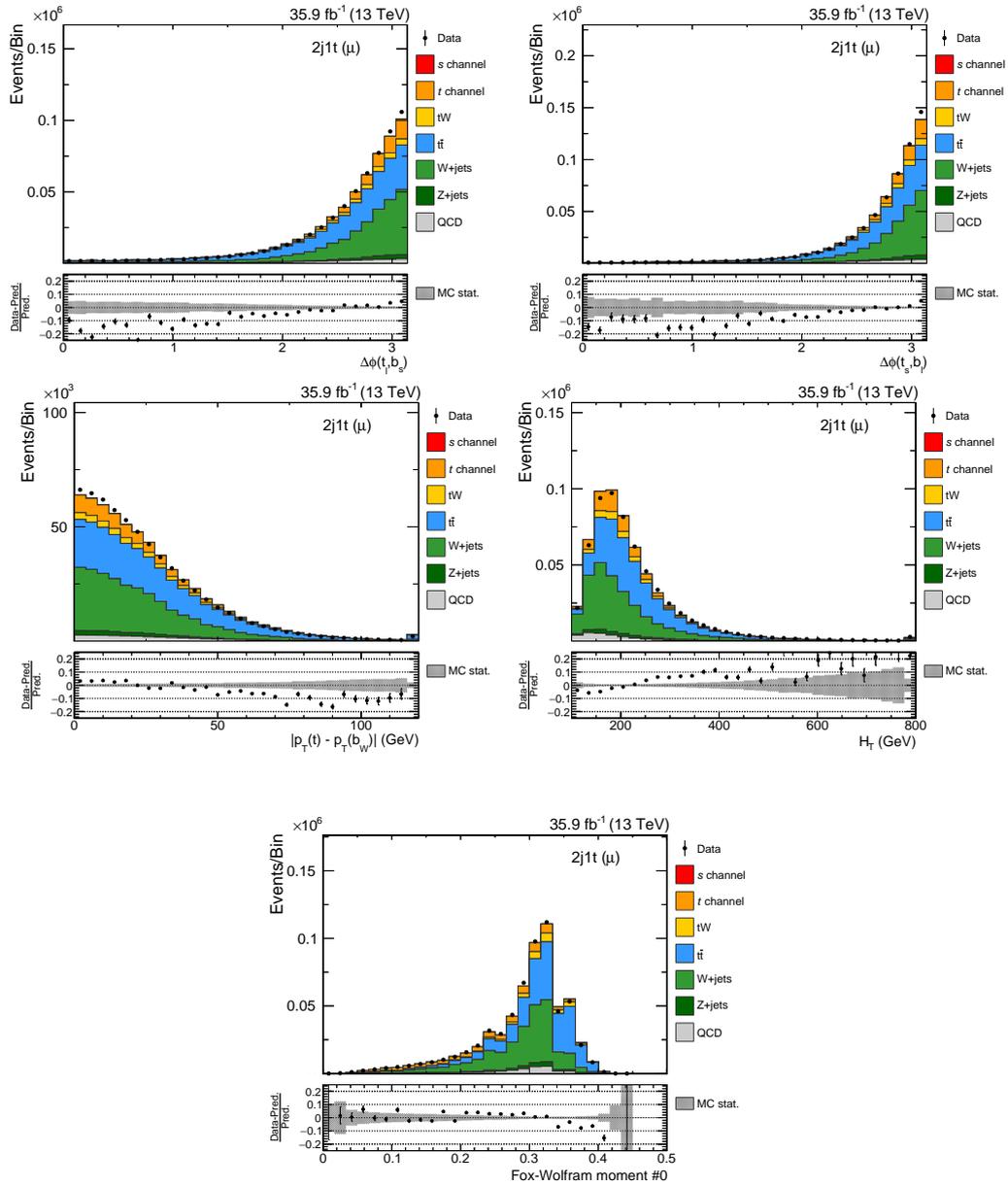


Abbildung 5.5.: Die Verteilung dieser Variablen weisen eine starke Abweichung von den gemessenen Daten auf.

Dadurch stehen lediglich 13 Variablen für das DNN-Training zur Verfügung. Um dennoch eine ausreichend Klassifikation durch den DNN zu erreichen, wird eine neue Variable ausgewählt und rekonstruiert. Diese neue Variable ist der Abstand $\Delta R(l, b_w)$ zwischen dem geladenen Lepton und dem b-Jet, welcher in der Rekonstruktion dem W-Boson zugeordnet wird. Dabei ist im s-Kanal-

Prozess die Pseudorapidität η des b -Jets des W -Bosons deutlich niedriger als der light-flavor Jet des t -Kanals-Prozesses, welcher in der Rekonstruktion dem W -Boson zugeordnet wird. Der Abstand ist definiert als

$$\Delta R(l, b_w) = \sqrt{\Delta\eta(l, b_w)^2 + \Delta\phi(l, b_w)^2}. \quad (5.1)$$

Somit ergeben sich unterschiedliche $\Delta R(l, b_w)$ -Werte für den s -Kanal- und t -Kanal-Prozess. Das DNN-Training wird mit dieser neuen Variable im Myon-Kanal der $2j1t$ -Region durchgeführt. Sie weist keine hohe Korrelation mit den anderen Variablen auf. Nach Anwendung dieses Trainings auf Daten und simulierte Ereignisse wird die Modellierung dieser neuen Variable überprüft. Sie zeigt eine gute Übereinstimmung zwischen Daten und simulierten Ereignissen (siehe Abb. 5.6).

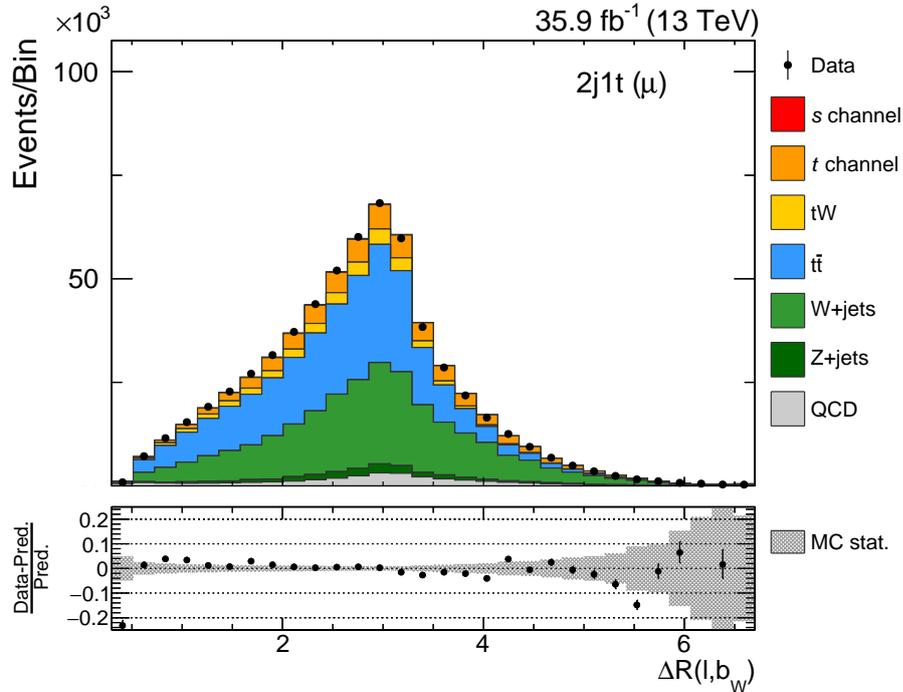


Abbildung 5.6.: Verteilung der neu eingeführten Variable $\Delta R(l, b_w)$ im Myon-Kanal der $2j1t$ -Region. Die Daten stimmen gut mit den simulierten Ereignissen überein.

Anschließend wird die Struktur des DNN optimiert, um eine Erhöhung der Fläche unter der ROC-Kurve zu erreichen. Ein maximaler Wert von 0,789 wird erreicht, wenn die Parameter des DNN wie in Tabelle 5.4 beschrieben gewählt werden. Die Test- und Trainingsverteilungen des Signals und des Untergrundes sowie die Werte der KS-Tests sind in Abbildung 5.7 gezeigt. Es ist kein Overtraining feststellbar. Die Trennung zwischen Signal und Untergrund ist deutlich besser als in der 2j2t-Region, da die Verteilung des Untergrunds ein Maximum nahe am Wert 0 aufweist. Diese Klassifikation des DNN wird auf die simulierten s -Kanal- und t -Kanal-Ereignisse und zusätzlich auf $t\bar{t}$ -Ereignisse angewendet, um die unterschiedlichen Verteilungen der DNN-Ausgabewerte zu vergleichen. In Abbildung 5.8 wird eine deutlichen Trennung zwischen der s -Kanal- und t -Kanal-Produktion beobachtet. Jedoch ist die Verteilung der $t\bar{t}$ -Ereignisse in der 2j1t-Region ähnlich der Verteilung des s -Kanal-Prozesses. Da aber der $t\bar{t}$ -Prozess bereits in der 2j2t-Region gut vom Signal getrennt wird, stellt die beobachtete Ähnlichkeit mit Signal in der 2j1t-Region kein Problem dar. Zum Schluss wird diese Klassifikation auf alle simulierten Ereignisse und Daten angewendet. Eine Übereinstimmung zwischen Daten und Monte-Carlo-Ereignissen wird in Abbildung 5.9 beobachtet. Dabei wären sie besser kompatibel, wenn die Systematischen Unsicherheiten in der Analyse berücksichtigt würden. Die im Myon-Kanal verwendeten Variablen werden nach ihrer Trennstärke sortiert in Tabelle 5.5 dargestellt.

Da der s -Kanal-Prozess vom Untergrund im Myon-Kanal der 2j1t-Region gut getrennt wird, wird nun im Elektron-Kanal überprüft, wie gut der s -Kanal-Prozess vom t -Kanal-Untergrund trennbar ist.

Tabelle 5.4.: Optimierte Parameter für das DNN-Training im Myon-Kanal der 2j1t-Region.

Parameter	Wert
Layout	TANH 25, TANH 50, TANH 50,LINEAR
LearningRate	0,01
Momentum	0,8
Repetitions	1
ConvergenceSteps	20
DropConfig	0.0+0,5+0,5+0.0
Multithreading	True

5. Suche nach der Einzel-Top-Quark-Produktion im s-Kanal

Tabelle 5.5.: Die Variablen sind nach ihrer Trennstärke für das DNN-Training im Myon-Kanal der 2j1t-Region sortiert.

Variable	Myon-Kanal
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) $	1
$\Delta R(l, \mathbf{b}_W)$	2
$ \eta(l) - \eta(\mathbf{b}_1) $	3
M_{lb2}	4
Fox Wolfram #3	5
$\cos\theta$	6
$p_T(l)$	7
p_T^{bb}	8
m_T^W	9
$m(t)$	10
p_T^{miss}	11
$\Delta\phi(\mathbf{b}_1, \mathbf{b}_s)$	12
$\Delta\phi(\mathbf{t}, \mathbf{b}_1)$	13
$q(l)$	14

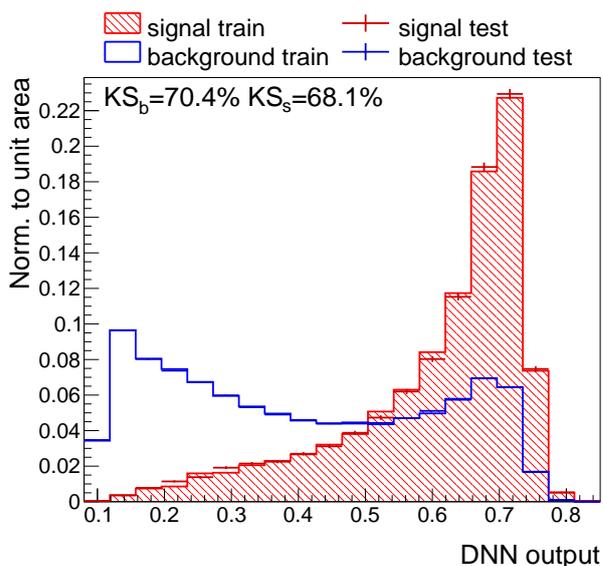


Abbildung 5.7.: Klassifikation der Ereignisse durch das DNN im Myon-Kanal der 2j1t-Region. Die Verteilung des Untergrunds weist in Nähe des Wertes 0 ein Maximum auf.

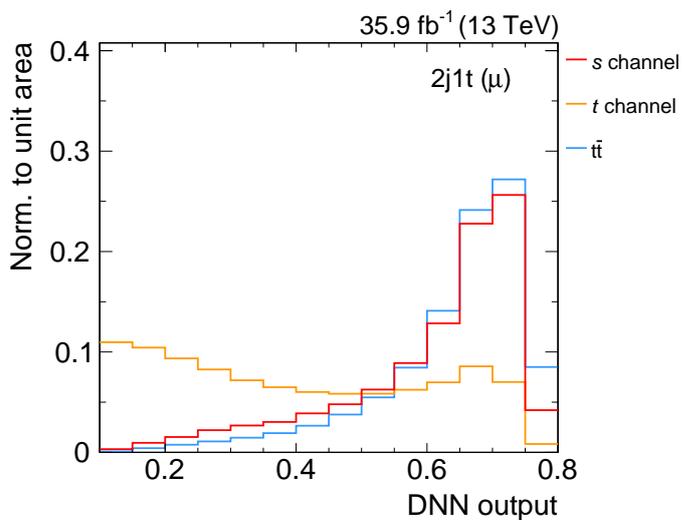


Abbildung 5.8.: DNN-Ausgabeverteilungen für verschiedene Prozesse. Der s -Kanal-Prozess wird in der 2j1t-Region der Myon-Kanal deutlich vom t -Kanal-Untergrund getrennt. Dagegen weist der $t\bar{t}$ -Prozess eine signalartige Verteilung der DNN-Ausgabewerte auf.

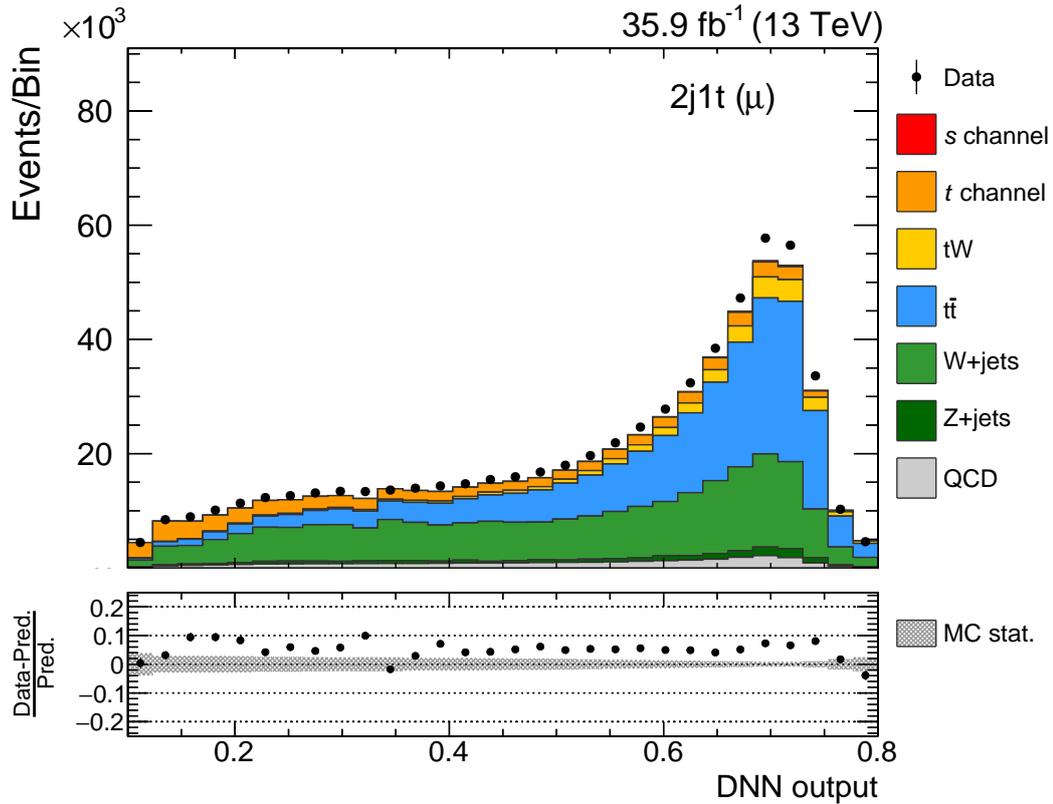


Abbildung 5.9.: DNN-Ausgabe im Myon-Kanal der 2j1t-Region. Die Daten stimmen relativ gut mit den simulierten Ereignissen überein. Sie weichen innerhalb der Verteilung von den simulierten Ereignissen um kleiner als 10% ab.

5.4.2. Elektron-Kanal

Im Elektron-Kanal wird zunächst die in Abschnitt 5.4.1 eingeführte Variable $\Delta R(l, b_w)$ rekonstruiert, danach wird das DNN-Training mit der gleichen Netzstruktur wie im Myon-Kanal durchgeführt. Nach Anwendung der Klassifikation wird jedoch eine relativ große Abweichung zwischen Daten und simulierten Ereignissen für diese Variable $\Delta R(l, b_w)$ beobachtet (siehe Abb. 5.10). Dadurch wird diese Variable für das DNN-Training in Elektron-Kanal wieder verworfen. Somit sind nur 13 Variablen für das DNN-Training verfügbar. Um die Gefahr des Overtrainings zu verringern, wird die Zahl der Konvergenzschritte auf 15 reduziert und der Anteil der Neuronen, bei denen Dropout angewendet wird, erhöht. Die Parameter für dieses DNN-Training sind in Tabelle 5.6 gelistet. Nach Durchführung des DNN-Trainings werden die Verteilungen des Signals und des Untergrunds auf ihre Trennung und auf Overtraining untersucht (siehe

Abb. 5.11). Die Verteilungen sind ähnlich zu denen im Myon-Kanal. Der Wert des KS-Tests für das Signal verringert sich auf 18,8%, da im Elektron-Kanal der 2j1t-Region weniger Ereignisse vorliegen als im Myon-Kanal. Allerdings ist der KS-Wert größer als 10%, somit liegt noch kein Overtraining vor. Die Fläche unter ROC-Kurve beträgt in dieser Klassifikation 0,789. In Abbildung 5.12 zeigt die DNN-Ausgabeverteilungen für s -, t - und $t\bar{t}$ Prozess. Zum Schluss wird dieses DNN-Training auf alle relevanten Untergrundprozesse angewendet. Die Daten stimmen mit den simulierten Ereignissen gut überein (siehe Abb. 5.13).

Tabelle 5.6.: Optimierte Parameter für das DNN-Training im Elektron-Kanal der 2j1t-Region.

Parameter	Wert
Layout	TANH 25, TANH 50, TANH 50,LINEAR
LearningRate	0,01
Momentum	0,8
Repetitions	1
ConvergenceSteps	15
DropConfig	0.0+0,6+0,5+0.0
Multithreading	True

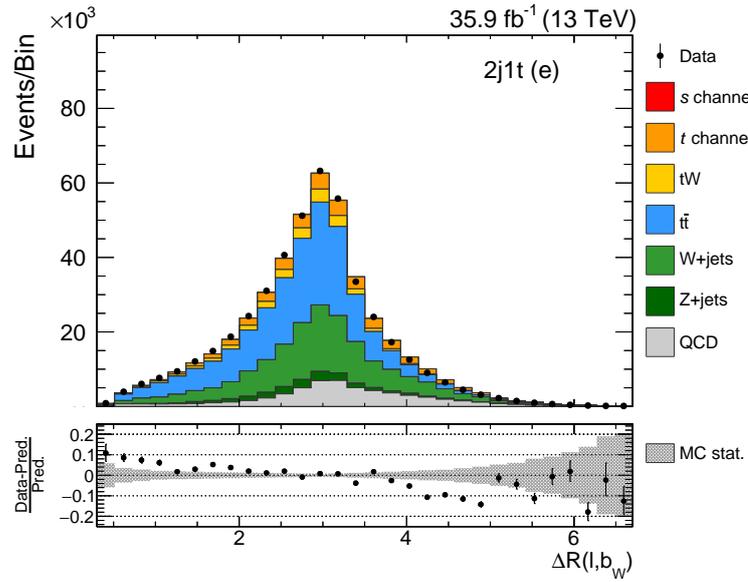


Abbildung 5.10.: Verteilung der neu eingeführten Variable $\Delta R(l, b_w)$ im Elektron-Kanal der 2j1t-Region. Die Daten weichen von den simulierten Ereignissen im Bereich von $4 \leq \Delta R(l, b_w) \leq 6,5$ stark ab.

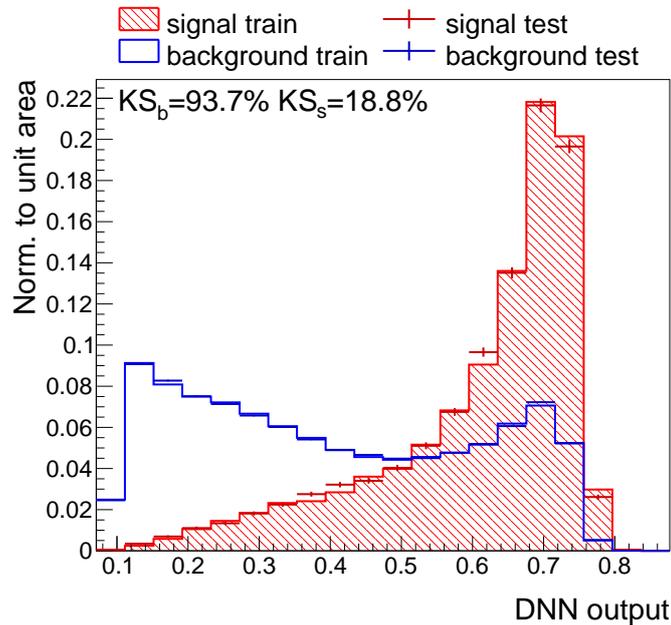


Abbildung 5.11.: Klassifikation der Ereignisse durch den DNN im Elektron-Kanal der 2j1t-Region. Die Verteilung des Untergrunds weist in Nähe des Wertes 0 ein Maximum auf.

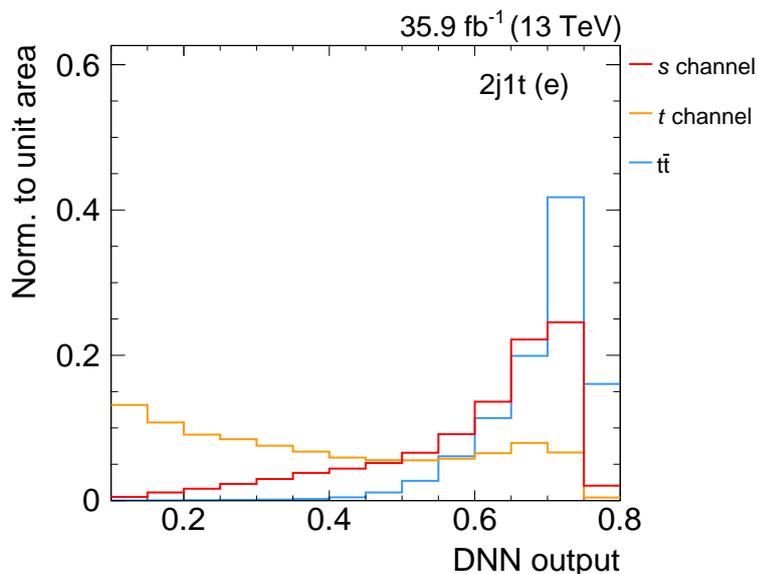


Abbildung 5.12.: DNN-Ausgabeverteilungen für verschiedene Prozesse. Der s -Kanal-Prozess wird in der 2j1t-Region der Elektron-Kanal deutlich vom t -Kanal-Untergrund getrennt. Dagegen weist der $t\bar{t}$ -Prozess eine signalartige Verteilung der DNN-Ausgabewerte auf.

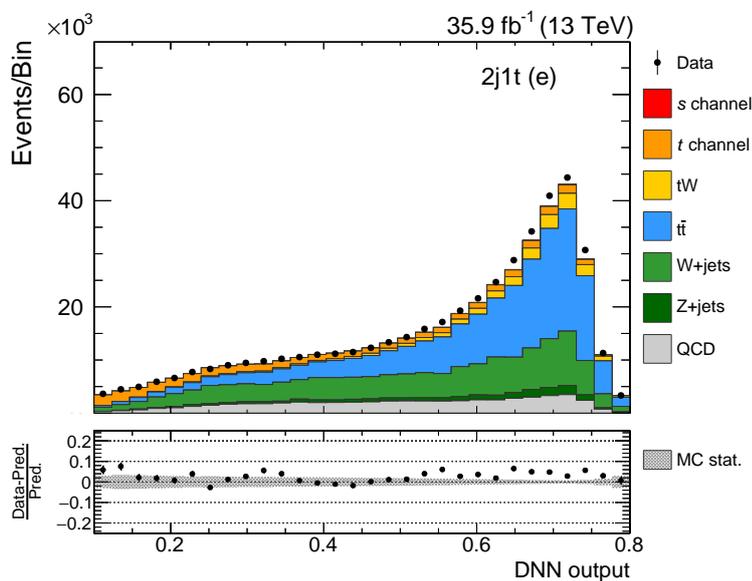


Abbildung 5.13.: Nach Anwendung des DNN-Trainings wird eine gute Übereinstimmung zwischen Daten und simulierten Ereignissen im Elektron-Kanal der 2j1t-Region beobachtet.

Die im Elektron-Kanal verwendeten Variablen werden nach ihrer Trennstärke sortiert in Tabelle 5.7 dargestellt. Im Vergleich zu Tabelle 5.5 wird beobachtet, dass die Reihenfolge der Variablen im Myon-Kanal und Elektron-Kanal im Wesentlichen identisch ist. Die Verteilungen der Eingangsvariablen für den Myon- und Elektron-Kanal befinden sich jeweils in Anhang A und B.

Tabelle 5.7.: Die Variablen sind nach ihrer Trennstärke für das DNN-Training im Elektron-Kanal der 2j1t-Region sortiert.

Variable	Elektron-Kanal
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) $	1
$\Delta R(l, \mathbf{b}_w)$	-
$ \eta(l) - \eta(\mathbf{b}_1) $	2
M_{lb2}	3
Fox Wolfram #3	4
$\cos\theta$	5
$p_T(l)$	10
p_T^{bb}	6
m_T^W	7
$m(t)$	8
p_T^{miss}	11
$\Delta\phi(\mathbf{b}_l, \mathbf{b}_s)$	12
$\Delta\phi(t, \mathbf{b}_1)$	9
$q(l)$	13

5.4.3. Kombination des Myon- und Elektron-Kanals

In den beiden vorherigen Abschnitten hat sich gezeigt, dass sich der s -Kanal-Prozess vom t -Kanal-Untergrund sowohl im Myon- als auch im Elektron-Kanal gut trennen lässt. Nun wird das DNN-Training im kombinierten Myon- und Elektron-Kanal durchgeführt mit dem Ziel, durch eine höhere Anzahl an Ereignissen eine bessere Trennung im DNN-Training zu erzielen.

Hauptuntergrund: t -Kanal

Für diese Studie werden zunächst nur die Ereignisse des t -Kanal-Prozesses als Untergrund für das DNN-Training berücksichtigt. Da im kombinierten Myon- und Elektron-Kanal eine ausreichende Anzahl an Ereignissen zur Verfügung steht, wird das DNN mit der gleichen Netz-Struktur wie im Myon-Kanal trainiert (siehe Tabelle 5.4). Aufgrund der großen Abweichung zwischen den Daten und den simulierten Ereignissen der Variable $\Delta R(l, b_w)$ im Elektron-Kanal wird für das DNN-Training im kombinierten Myon- und Elektron-Kanal diese Variable nicht berücksichtigt. Insgesamt 13 Variablen stehen für das Training zur Verfügung, sie sind in Tabelle 5.8 nach ihrem Beitrag zum Training sortiert. Nach Durchführung des DNN-Trainings beträgt die Fläche unter der ROC-Kurve 0,788. In Abb. 5.14 sind die Verteilungen des Signals und Untergrunds für die Trainings- und Testdaten gezeigt. Da signifikant mehr Ereignisse für das DNN-Training verfügbar sind, erhöhen sich die KS-Werte des Signals und des Untergrunds. Die Testdaten weisen somit eine hohe Übereinstimmung mit dem Trainingsdaten auf. Das DNN-Training wird jeweils auf die gemessenen Daten und simulierten Ereignisse im Myon- und Elektron-Kanal angewendet. In Abbildung 5.15 lässt sich erkennen, dass in den beiden Kanälen die Daten und Monte-Carlo-Ereignisse übereinstimmen. Die Verteilungen aller für das Training verwendeten Variablen sind in Anhang C jeweils für den Myon-Kanal und Elektron-Kanal gezeigt.

5. Suche nach der Einzel-Top-Quark-Produktion im s -Kanal

Tabelle 5.8.: Die Variablen sind nach ihrer Trennstärke für das DNN-Training im kombinierten Myon- und Elektron-Kanal der 2j1t-Region sortiert. Hierbei wurde nur der t -Kanal-Prozess als Untergrund verwendet.

Variable	Kombination mit Untergrund t -Kanal
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) $	1
$\Delta R(l, \mathbf{b}_W)$	-
$ \eta(l) - \eta(\mathbf{b}_1) $	2
M_{lb2}	3
Fox Wolfram #3	4
$\cos\theta$	5
$p_T(l)$	7
p_T^{bb}	6
m_T^W	8
$\mathbf{m}(t)$	9
p_T^{miss}	10
$\Delta\phi(\mathbf{b}_l, \mathbf{b}_s)$	12
$\Delta\phi(t, \mathbf{b}_1)$	11
$q(l)$	13

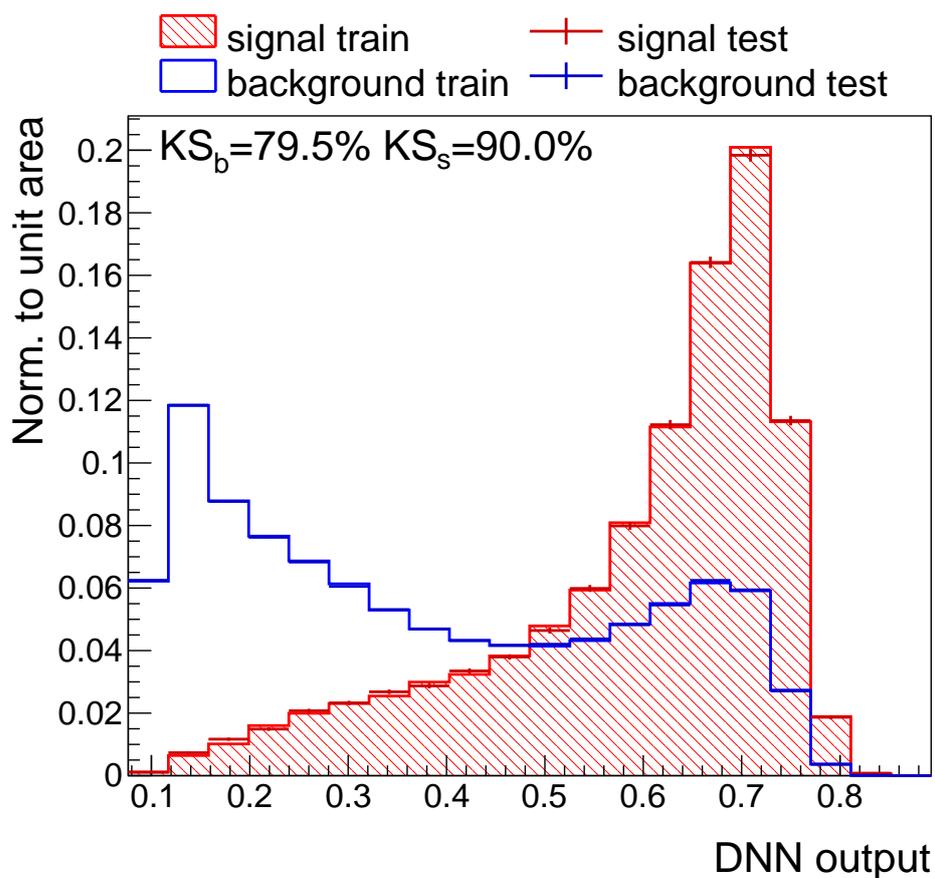
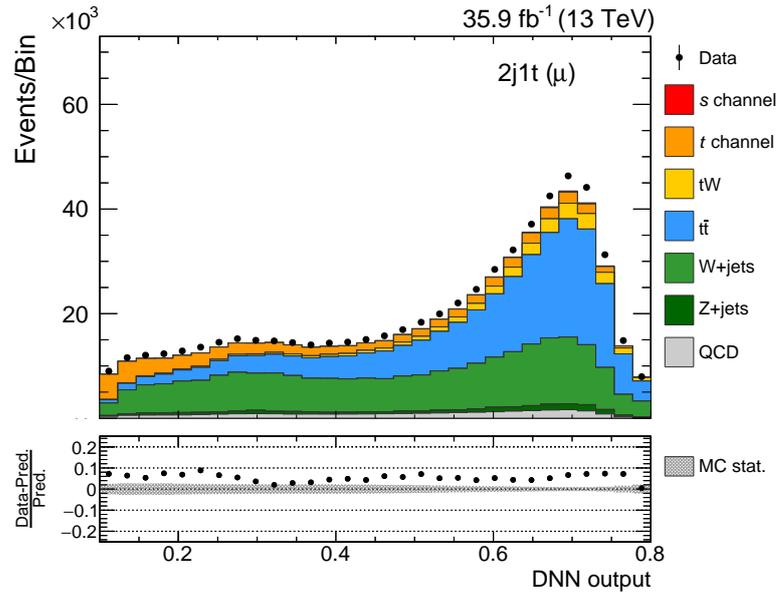
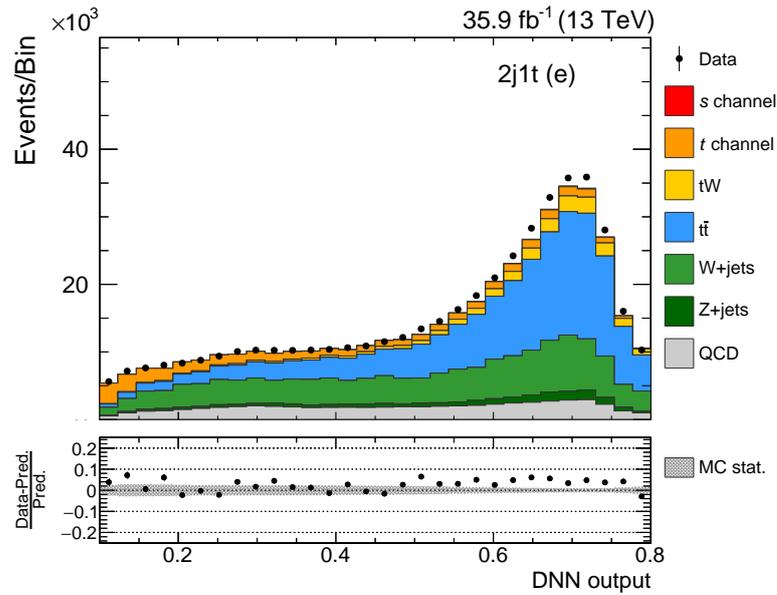


Abbildung 5.14.: Klassifikation der Ereignisse durch das DNN im kombinierten Myon- und Elektron-Kanal der 2j1t-Region. Die Verteilung des Untergrunds weist in Nähe des Wertes 0 ein Maximum auf.



(a) Myon-Kanal



(b) Elektron-Kanal

Abbildung 5.15.: Die Daten stimmen mit den simulierten Ereignissen überein. Sie weichen innerhalb der Verteilung von den simulierten Ereignissen um kleiner als 10% ab. Die Verteilung der DNN-Ausgabewerte besitzen in beiden Kanäle einen ähnlichen Verlauf.

Berechnung der Signifikanz

Um diese Analyse mit der Referenzanalyse, in der die Klassifikation durch ein DNN, das in der 2j2t-Region mit dem $t\bar{t}$ -Prozesses als Untergrund trainiert wurde, zu vergleichen, wird die Signifikanz des s -Kanal-Prozesses durch eine Anpassung der DNN-Ausgabe-Verteilungen in allen relevanten Regionen, 2-Jet-1-Tag, 2-Jet-2-Tag, 3-Jet-1-Tag und 3-Jet-2-Tag, berechnet. Hierbei wird für die 2j1t-Region die Klassifikation aus dieser Arbeit verwendet. Diese Klassifikation wird daher auf alle simulierten Ereignisse, die entsprechend der systematischen Unsicherheiten variiert sind, angewendet. Es existieren systematische Unsicherheiten, welche aus der Theorie stammen, sowie Unsicherheiten experimenteller Natur. Theoretische Unsicherheiten stammen aus dem Modell, welches für die Berechnung der Signal- und Untergrund-Prozesse verwendet wurde. Experimentelle Unsicherheiten werden beispielsweise durch das begrenzte Auflösungsvermögen des Detektors verursacht. Die in dieser Arbeit mittels der neu eingeführten Klassifikation des s -Kanal- und t -Kanal-Prozesses in der 2j1t-Region führt zu einer Signifikanz von 2,4, welche somit genauso hoch ist wie in der Referenzanalyse.

Hauptuntergrund: t -Kanal und $t\bar{t}$

Da sich die Signifikanz in der Analyse mit dem t -Kanal-Prozess als alleiniger Untergrund im DNN-Training nicht erhöht, wird nun im DNN-Training des kombinierten Myon- und Elektron-Kanals die $t\bar{t}$ -Produktion als weiterer Untergrundprozess berücksichtigt, wobei die gleichen Variablen verwendet werden. Die Parameter für die DNN-Struktur sind in Tabelle 5.9 dargestellt. Nach Durchführung des Trainings zeigt sich, dass sich die Reihenfolge der Variablen bezüglich der Trennstärke im Vergleich zum Training, welches nur den t -Kanal-Prozess berücksichtigt, stark verändert hat (siehe Tabelle 5.10). Die DNN-Ausgabeverteilungen von Signal und Untergrund sind in Abbildung 5.16 gezeigt. Die Fläche unter der ROC-Kurve beträgt lediglich 0,735. Außerdem besitzt die Untergrundverteilung kein ausgeprägtes Maximum nahe des Wertes 0. Somit wird der s -Kanal-Prozess vom Untergrund nicht gut getrennt. In Abb. 5.17 sind die DNN-Ausgabeverteilungen der beiden Untergrundprozesse $t\bar{t}$ und t -Kanal im Vergleich zu der Verteilung des Signals dargestellt, jeweils im Myon-Kanal und im Elektron-Kanal. Es wird beobachtet, dass beide Untergrundprozesse im Bereich von 0 bis 0,6 eine ähnliche Tendenz wie der s -Kanal-Prozess aufweisen. Daher ist es schwierig, den s -Kanal-Prozess von den Hauptuntergründen, t -Kanal und $t\bar{t}$, in der 2j1t-Region mit diesen DNN zu tren-

nen. Zusätzlich wird das DNN-Training auf seine Modellierung im Vergleich zu den Daten im Myon-Kanal und Elektron-Kanal überprüft. Im Bereich von 0,6 bis 0,8 (siehe Abb. 5.18) ist eine große Abweichung zwischen Daten und simulierten Ereignissen sowohl im Myon- als auch im Elektron-Kanal erkennbar.

Zum Schluss wird die Signifikanz berechnet, wobei in der $2j1t$ -Region diese neue Klassifikation verwendet wird. Sie beträgt 2,5 und ist somit leicht erhöht gegenüber der Analyse, welche nur t -Kanal als Untergrundprozess im Training verwendet. Da jedoch die Daten mit den simulierten Ereignissen nicht übereinstimmen und der Wert unter der ROC-Kurve vergleichsweise gering ist, wird dieses DNN nicht verwendet.

Tabelle 5.9.: Parameter für das DNN-Training zur Klassifikation der Ereignisse im kombinierten Myon- und Elektron-Kanal der 2j1t-Region, mit t -Kanal und $t\bar{t}$ als Untergrund.

Parameter	Wert
Layout	TANH 25, TANH 50, TANH 50,LINEAR
LearningRate	0,01
Momentum	0,8
Repetitions	1
ConvergenceSteps	20
DropConfig	0.0+0,6+0,6+0.0
Multithreading	True

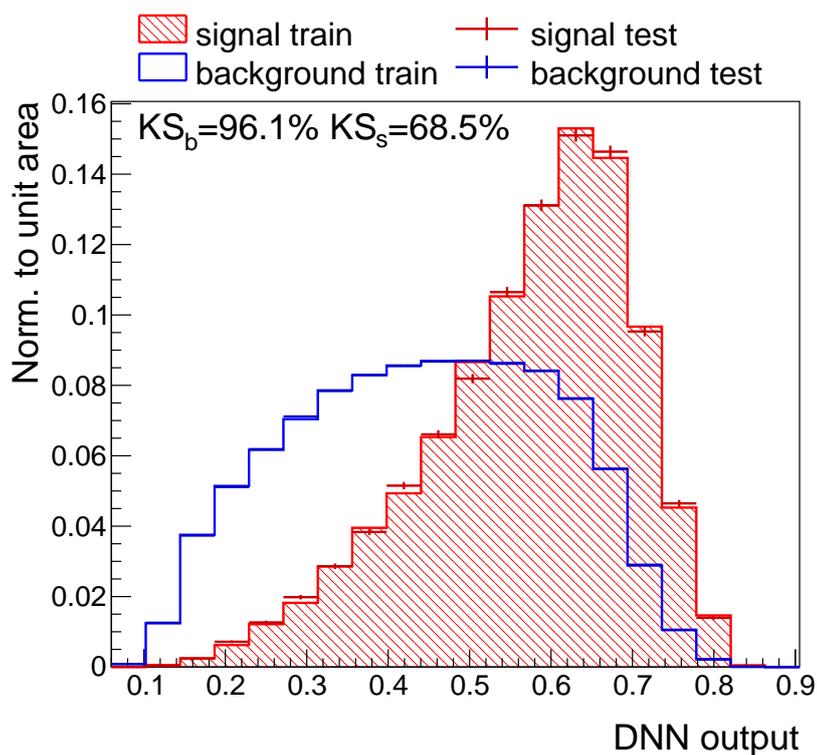
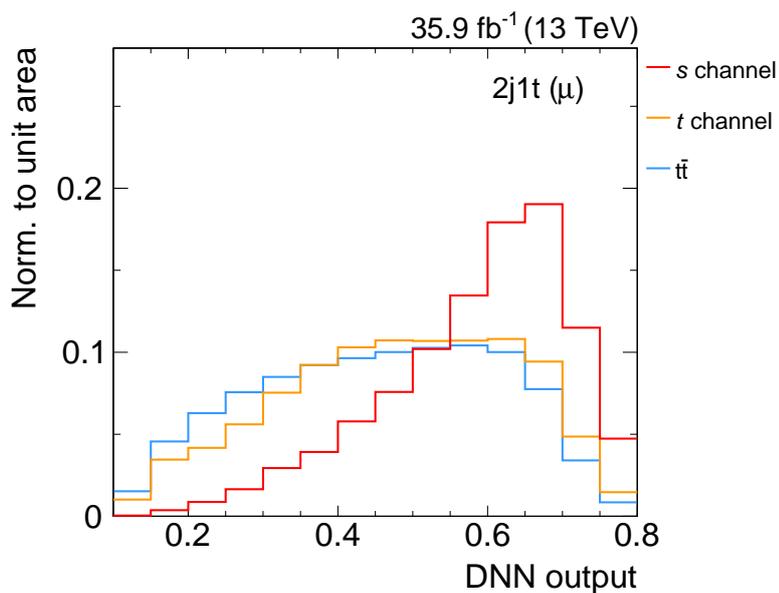


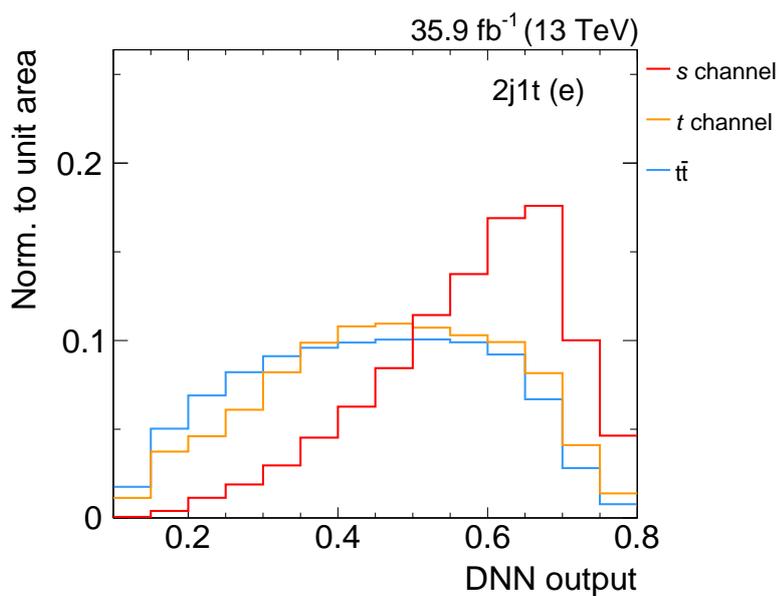
Abbildung 5.16.: Klassifikation der Ereignisse durch das DNN, mit t -Kanal und $t\bar{t}$ als Hauptuntergrund in der 2j1t-Region. Die Verteilung des Untergrunds weist kein Maximum nahe des Wertes 0 auf.

Tabelle 5.10.: Die Variablen sind nach ihrer Trennstärke für das DNN-Training im kombinierten Myon- und Elektron-Kanal der $2j1t$ -Region, mit t -Kanal- und $t\bar{t}$ -Prozess als Untergrund, sortiert.

Variable	Kombination mit Untergrund t -Kanal und $t\bar{t}$
$ \eta(\mathbf{b}_w) - \eta(\mathbf{b}_t) $	12
$\Delta R(l, \mathbf{b}_W)$	-
$ \eta(l) - \eta(\mathbf{b}_1) $	2
$M_{l\mathbf{b}_2}$	1
Fox Wolfram #3	11
$\cos\theta$	8
$p_T(l)$	6
p_T^{bb}	7
m_T^W	5
$\mathbf{m}(t)$	3
p_T^{miss}	10
$\Delta\phi(\mathbf{b}_l, \mathbf{b}_s)$	4
$\Delta\phi(t, \mathbf{b}_1)$	9
$q(l)$	13

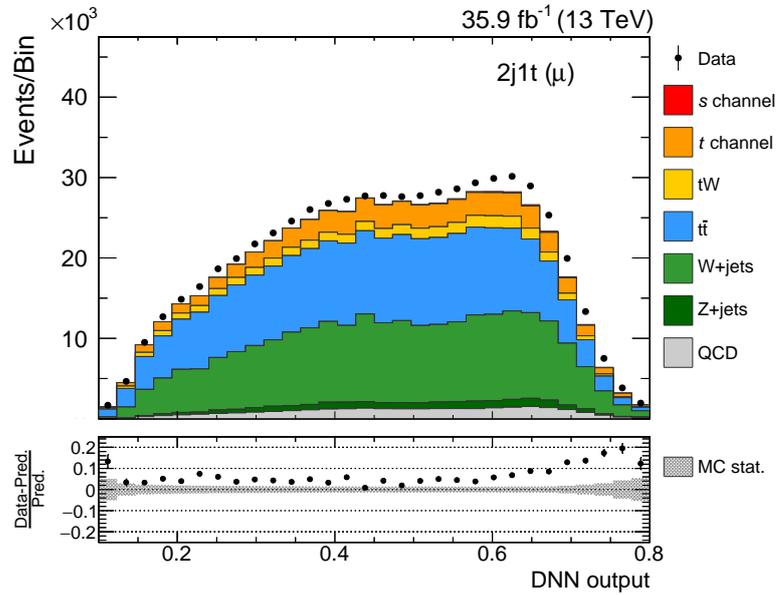


(a) Myon-Kanal

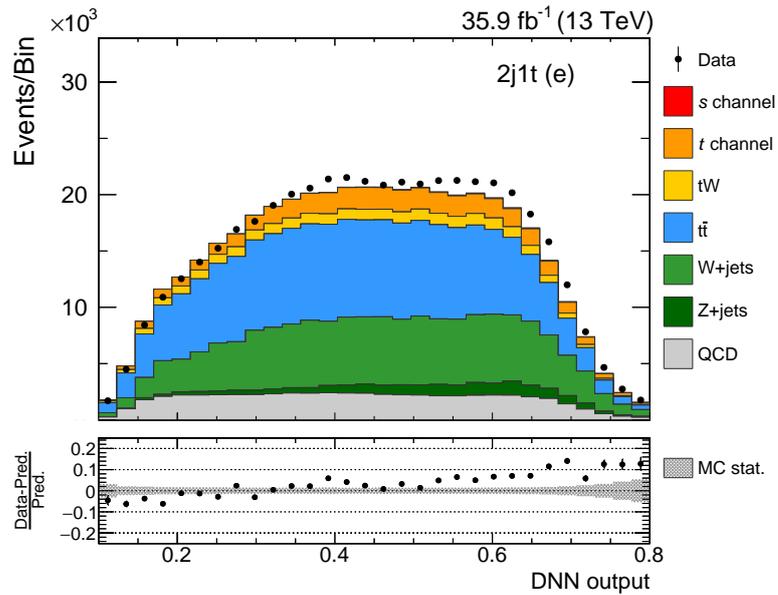


(b) Elektron-Kanal

Abbildung 5.17.: DNN-Ausgabevertellungen für den s -Kanal-, t -Kanal- und $t\bar{t}$ -Prozess. Die Verteilungen des t -Kanal- und des $t\bar{t}$ -Prozesses weisen eine hohe Ähnlichkeit zueinander auf und besitzen einen flachen Verlauf.



(a) Myon-Kanal



(b) Elektron-Kanal

Abbildung 5.18.: DNN-Ausgabeverteilung von Daten und Simulation. Die simulierten Ereignisse weichen im Bereich von 0,6 bis 0,8 sehr stark von der gemessenen Daten ab.

5.4.4. Schlussfolgerung

In der 2j1t-Region wird nur die t -Kanal-Produktion als Hauptuntergrund für das DNN-Training verwendet, da nach Anwendung dieser Klassifikation eine gute Übereinstimmung zwischen Daten und simulierten Ereignissen vorliegt und die DNN-Ausgabeverteilungen des Untergrunds und des Signals eine gute Trennung aufweisen.

6. Zusammenfassung

Um die Produktion einzelner Top-Quarks im s -Kanal nachzuweisen wurden tiefe neuronale Netze (DNN) zur Klassifikation benutzt. Dabei wurde der gesuchte s -Kanal-Prozess beim DNN-Training als Signal angenommen. Da je nach Wahl der Untergrundprozesse unterschiedliche Gewichte im DNN-Training für die Ereignisse berechnet wurden, ist die Ereignisklassifikation von der Wahl der Hauptuntergründe abhängig. Um zu entscheiden, welche Klassifikation geeignet ist, mussten verschiedene Kriterien erfüllt werden: Sie musste einen hohen Wert unter der ROC-Kurve, d.h. eine gute Trennung, aufweisen, und die Modellierung zwischen Daten und Monte-Carlo-Ereignissen musste für die DNN-Ausgabeverteilung übereinstimmen.

In der Referenzanalyse wurde hierfür der $t\bar{t}$ -Prozess als Hauptuntergrund für das DNN-Training in der $2j2t$ -Region verwendet. In dieser Arbeit lag der Fokus jedoch auf der Trennung des s -Kanal-Prozesses von einem weiteren Untergrundprozess, der Einzel-Top-Quark-Produktion im t -Kanal. Dabei wurden zunächst t -Kanal und $t\bar{t}$ als Untergrundprozesse beim Training in der $2j2t$ -Region verwendet, da diese Region die Signalregion ist. Es zeigte sich jedoch, dass die Form des t -Kanal-Prozesses signalartig war und zu wenige simulierte Ereignisse zum Training vorhanden waren. Daher wurde eine weitere Studie in der $2j1t$ -Region durchgeführt, mit dem t -Kanal-Prozess als Untergrund, da dort der t -Kanal-Prozess angereichert ist. Nach der Wahl und Rekonstruktion der kinematischer Variablen wurde das DNN-Training zunächst getrennt im Myon-Kanal und Elektron-Kanal durchgeführt. In beiden Kanälen ist eine gute Trennung des Signals vom Untergrund erkennbar und es liegt kein Overtraining vor. Anschließend wurde das DNN im kombinierten Myon- und Elektron-Kanal trainiert, um eine bessere Klassifikation durch mehr Ereignisse zu erreichen. Diese Klassifikation wurde auf Daten und Monte-Carlo-Ereignisse angewendet, wobei eine gute Übereinstimmung beobachtet wird (siehe Abb. 5.15). Zum Schluss wurde die Signifikanz in allen Ereignisregionen der Analyse berechnet, wobei für alle Regionen außer der $2j1t$ -Region der DNN aus der Referenzanalyse verwendet wurde. Obwohl sich die Signifikanz nicht erhöhte, bietet diese Analyse eine höhere Übereinstimmung der DNN-Ausgabeverteilungen für Da-

ten und Simulation in der $2j1t$ -Region an. Eine weitere Studie wurde in der $2j1t$ -Region durchgeführt, in der neben dem t -Kanal-Prozess auch $t\bar{t}$ -Ereignisse im DNN-Training berücksichtigt wurden. Hierbei zeigte sich jedoch, dass die DNN-Ausgabeverteilung für Daten und Simulation nicht gut übereinstimmt. Somit lässt sich mit einem DNN, das s -Kanal- von t -Kanal-Ereignissen in der $2j1t$ -Region unterscheiden kann, die Referenzanalyse am besten optimieren.

Um die Klassifikation von Signal und Untergrund weiter zu optimieren, könnte man für jede Ereignisregion jeweils die dominierenden Ereignisse als Untergrund im DNN-Training verwenden, sodass die Produktion einzelner Top-Quarks im s -Kanal leichter nachweisbar wäre.

A. Variablen für das DNN-Training in Myon-Kanal

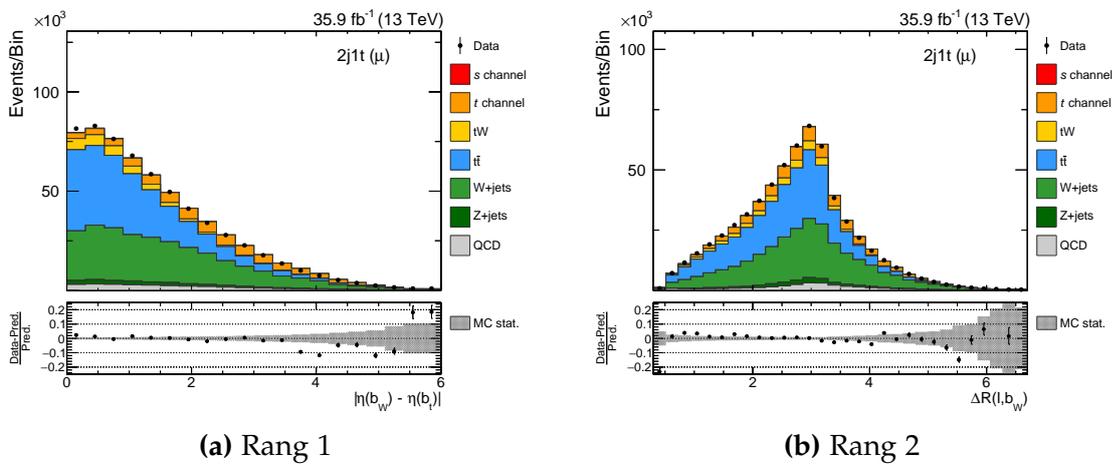


Abbildung A.1.: Die Verteilung der verwendeten Variablen für die Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

A. Variablen für das DNN-Training in Myon-Kanal

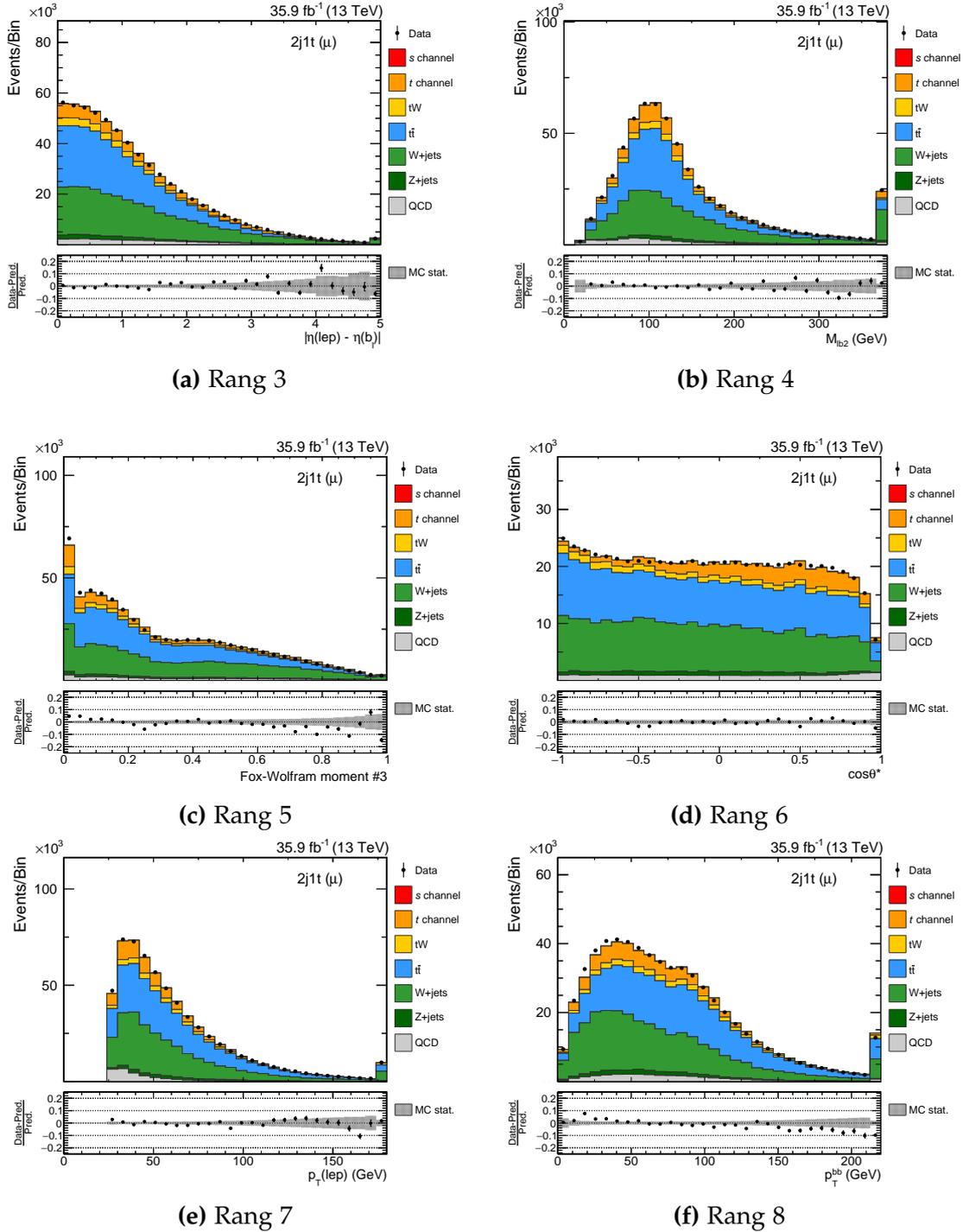
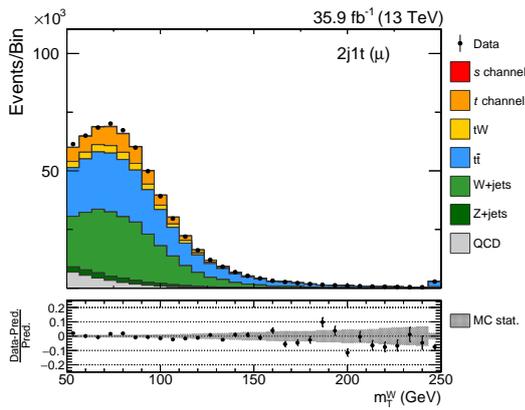
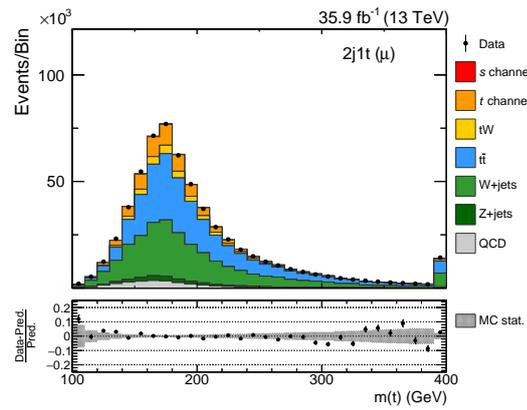


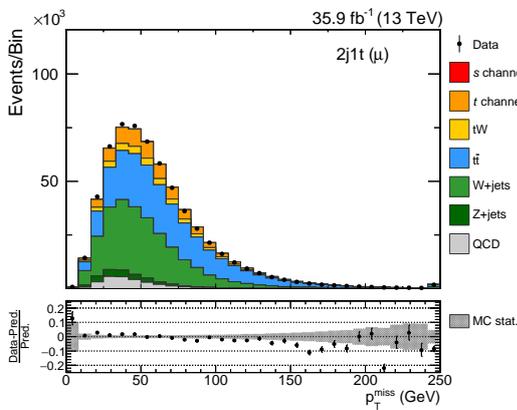
Abbildung A.2.: Die Verteilung der verwendeten Variablen für die Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.



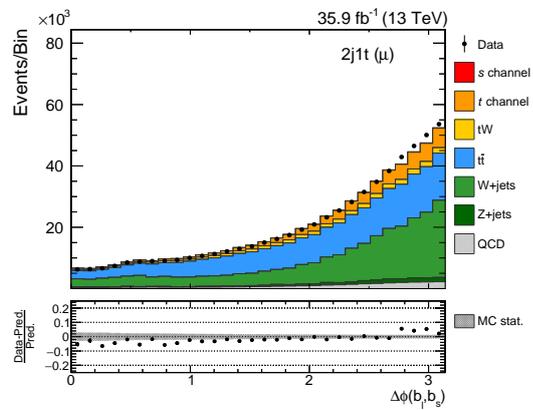
(a) Rang 9



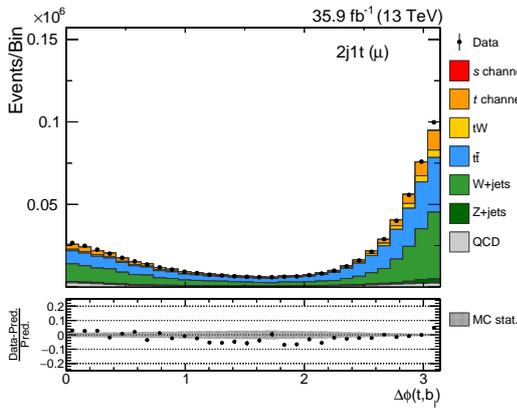
(b) Rang 10



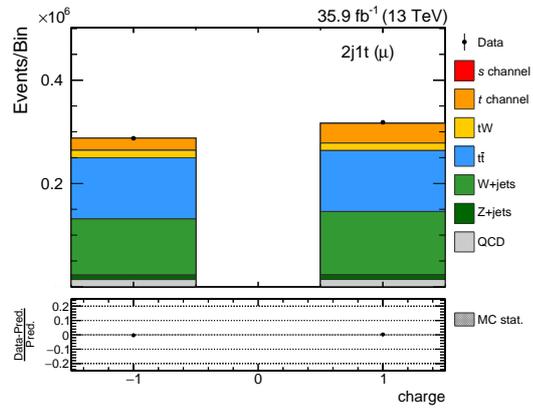
(c) Rang 11



(d) Rang 12



(e) Rang 13



(f) Rang 14

Abbildung A.3.: Die Verteilung der verwendeten Variablen für die Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

B. Variablen für das DNN-Training im Elektron-Kanal

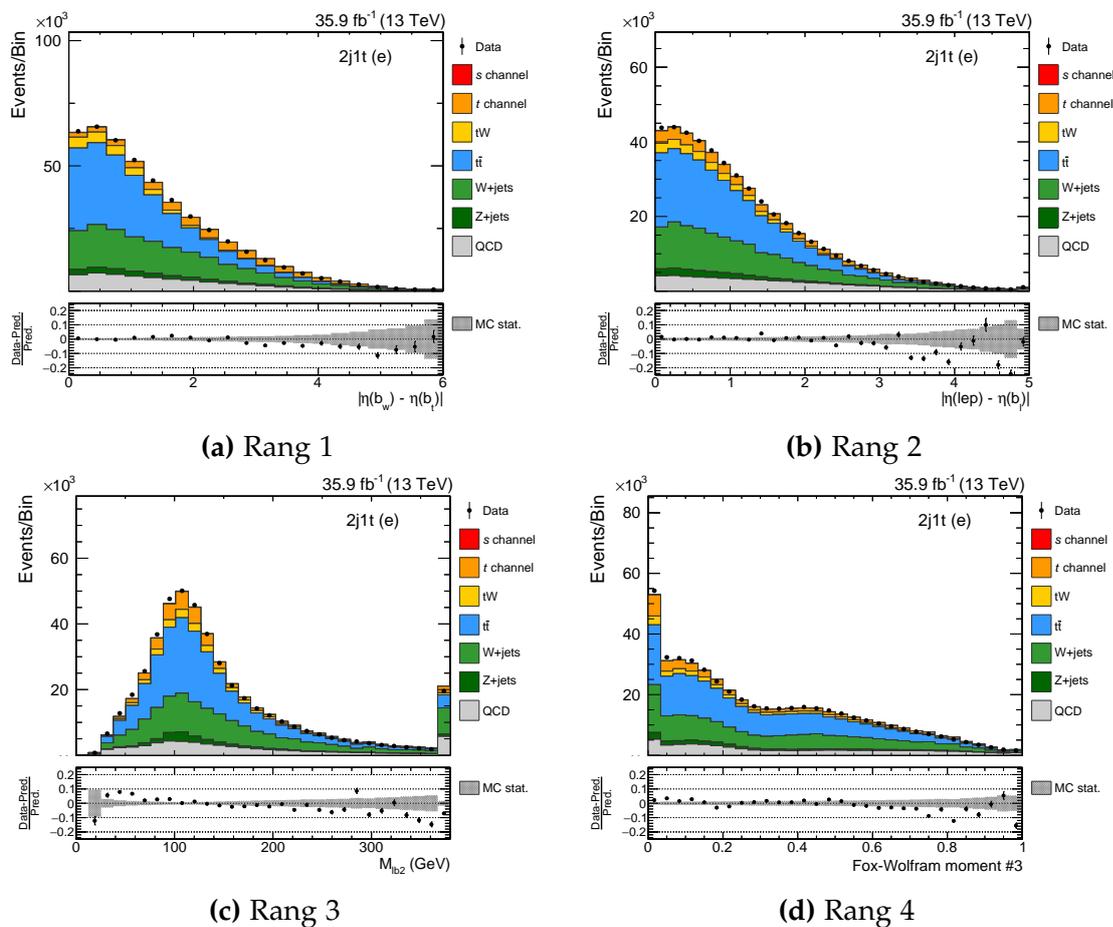


Abbildung B.1.: Die Verteilung der verwendeten Variablen für die Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

B. Variablen für das DNN-Training im Elektron-Kanal

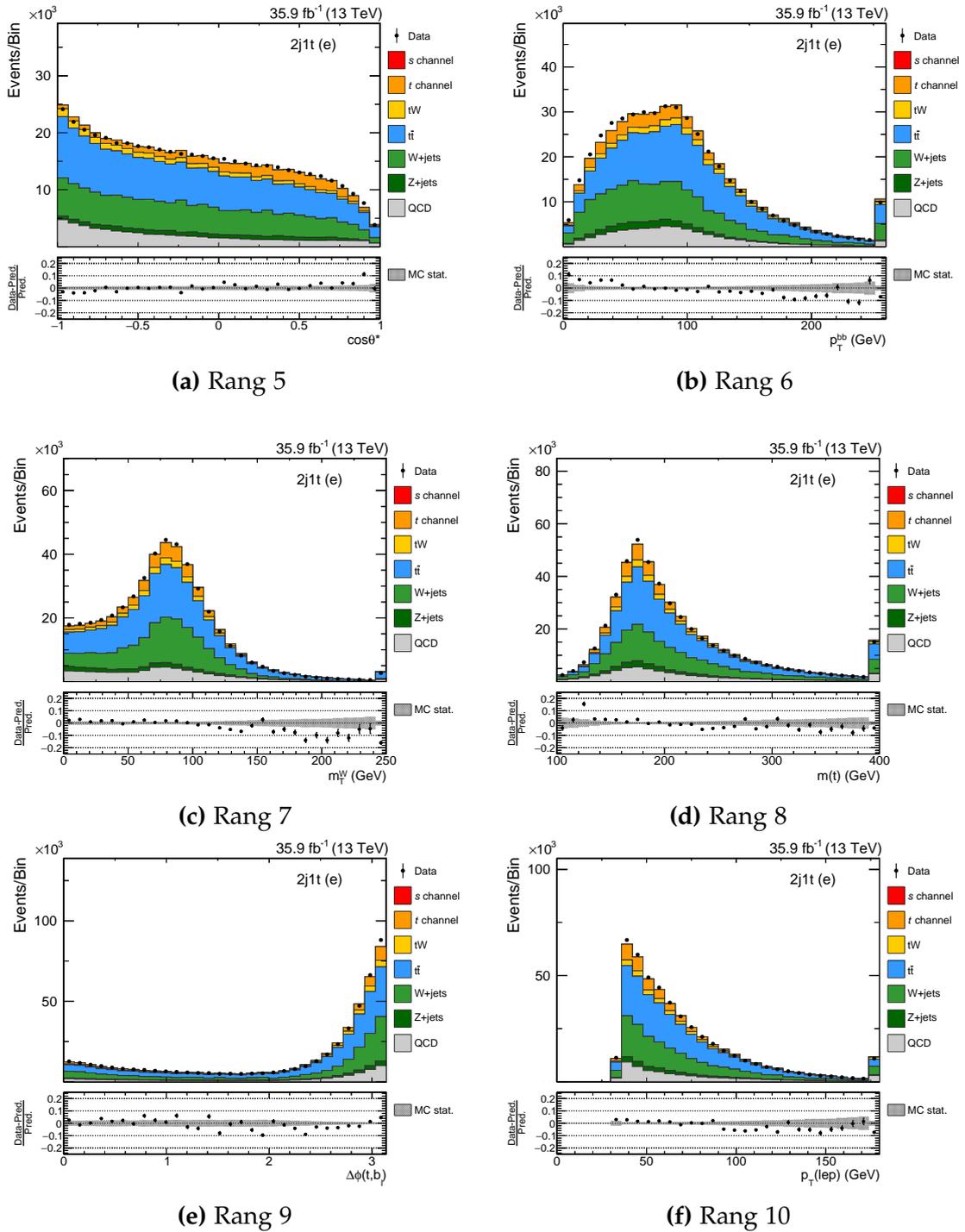
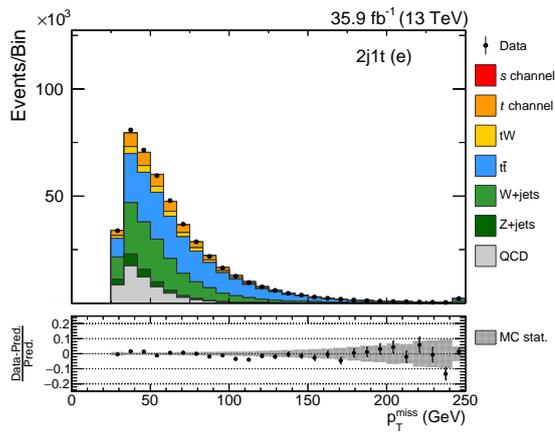
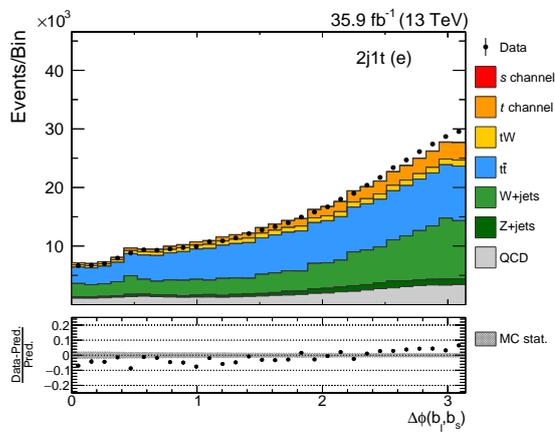


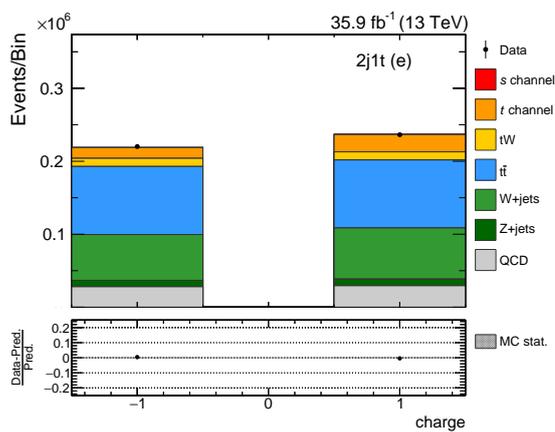
Abbildung B.2.: Die Verteilung der verwendeten Variablen für die Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.



(a) Rang 11



(b) Rang 12



(c) Rang 13

Abbildung B.3.: Die Verteilung der verwendeten Variablen für die Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

C. Variablen für das DNN-Training des kombinierten Myon- und Elektron-Kanals

In Myon-Kanal

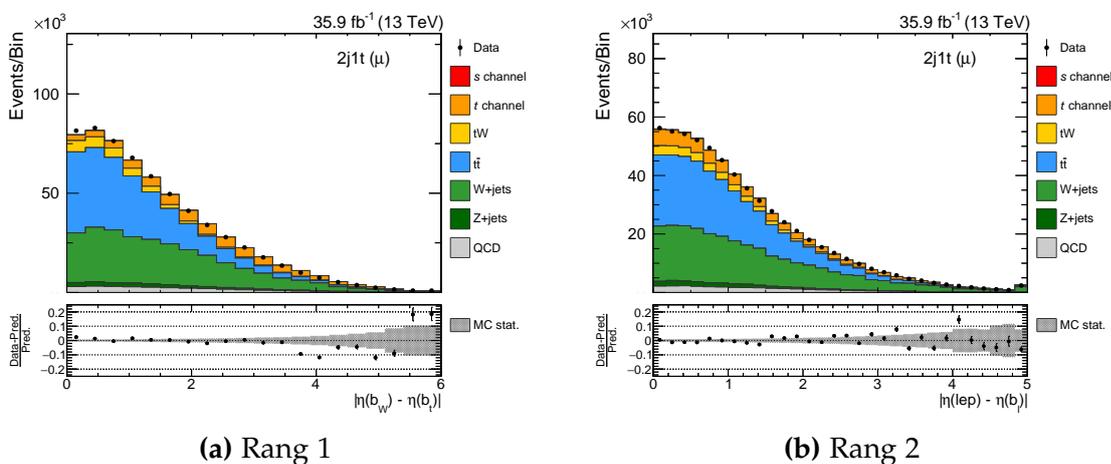


Abbildung C.1.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

C. Variablen für das DNN-Training des kombinierten Myon- und Elektron-Kanals

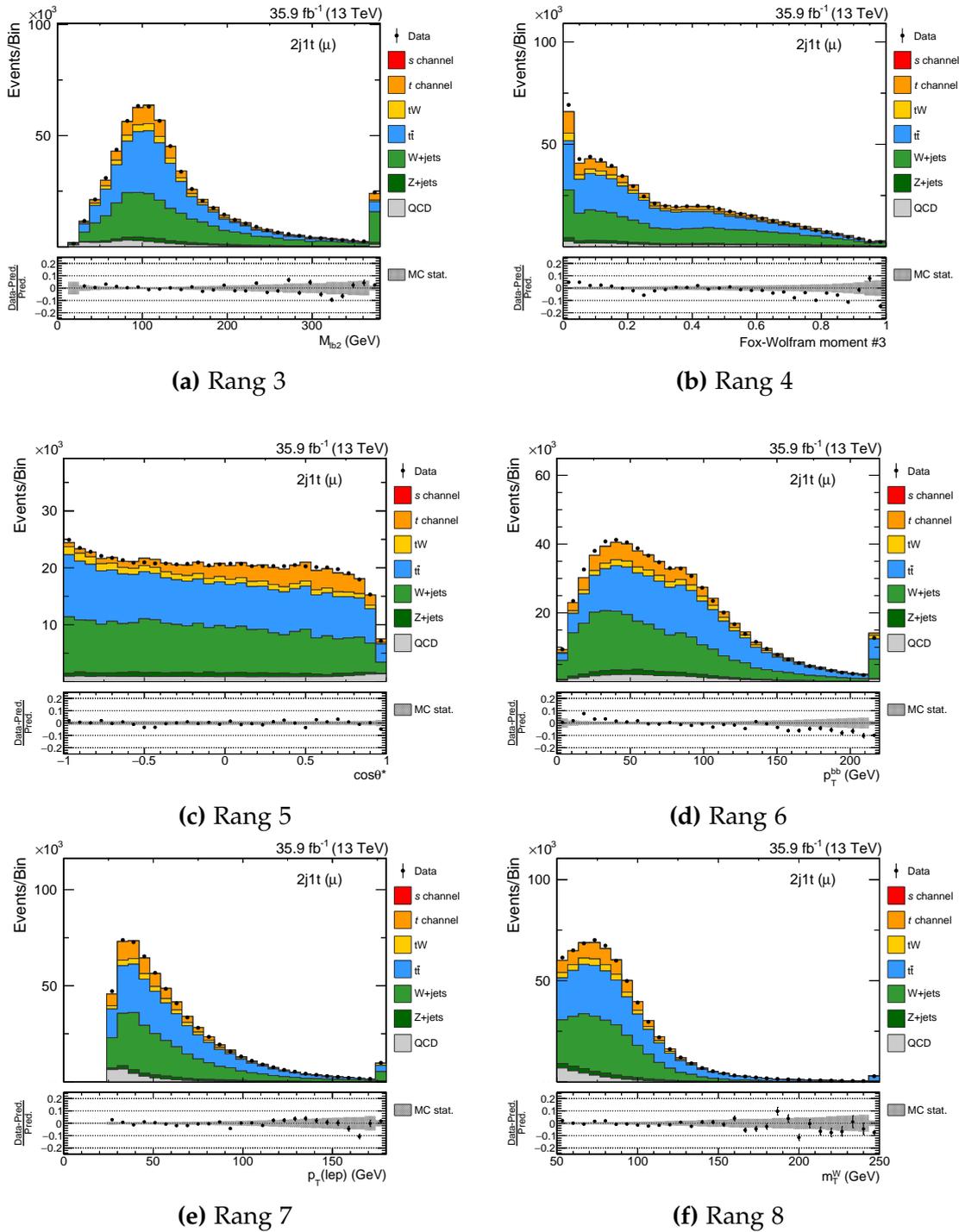


Abbildung C.2.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

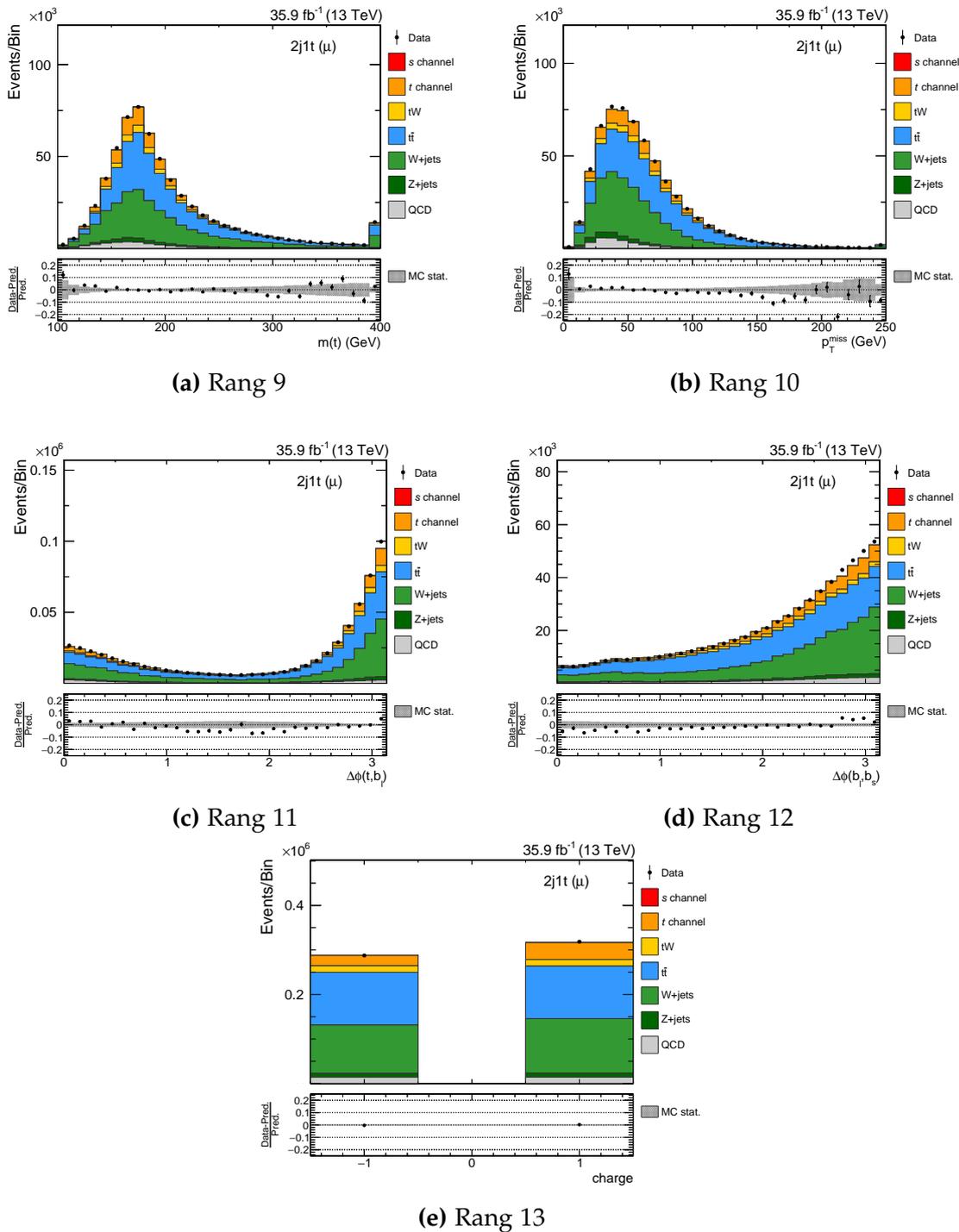


Abbildung C.3.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Myon-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

In Elektron-Kanal

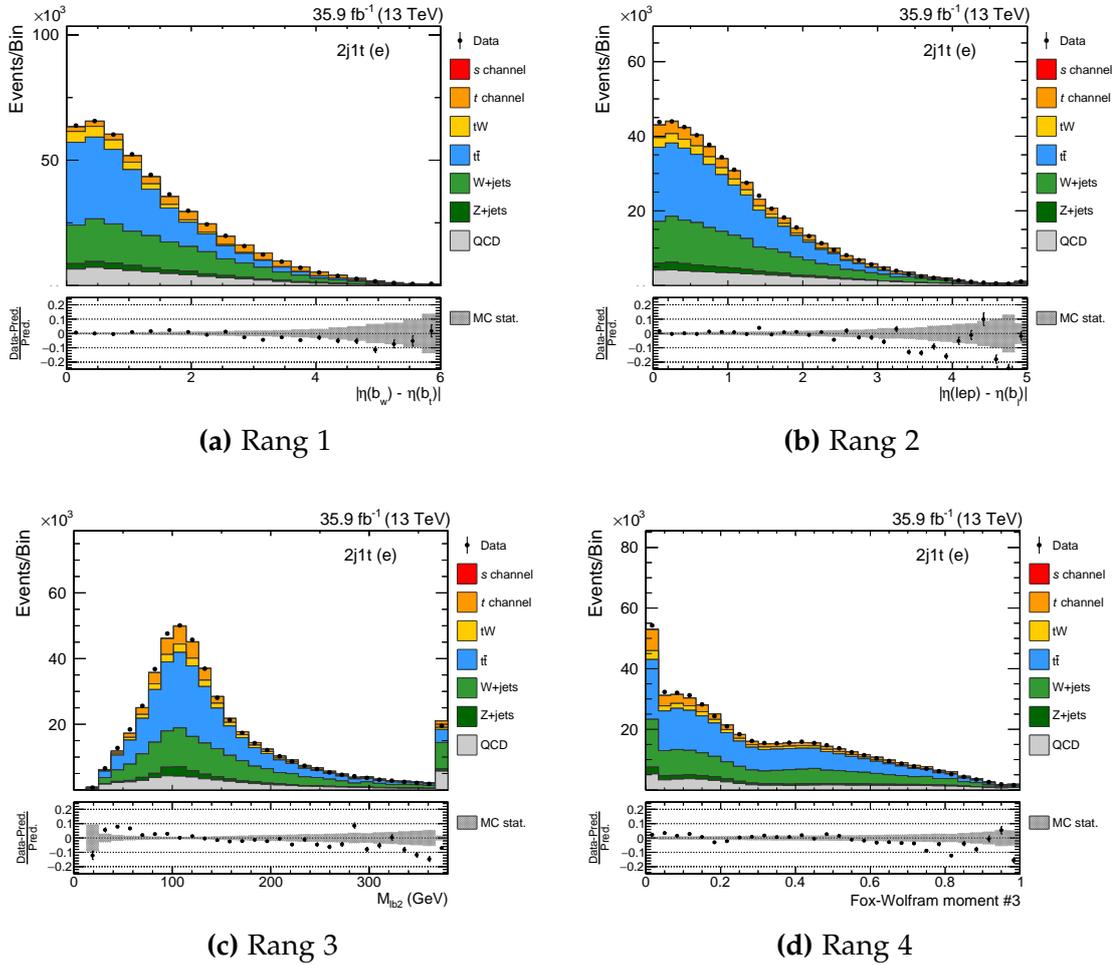
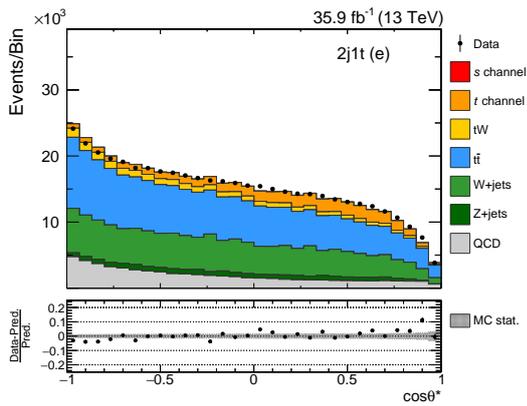
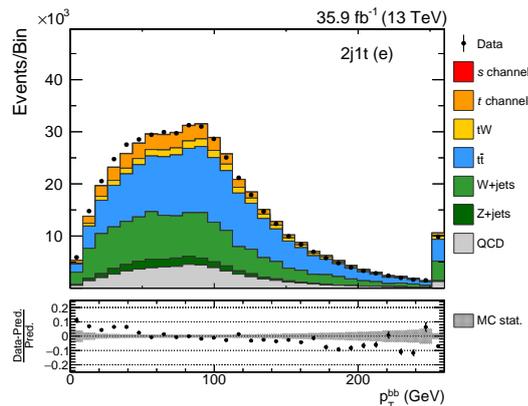


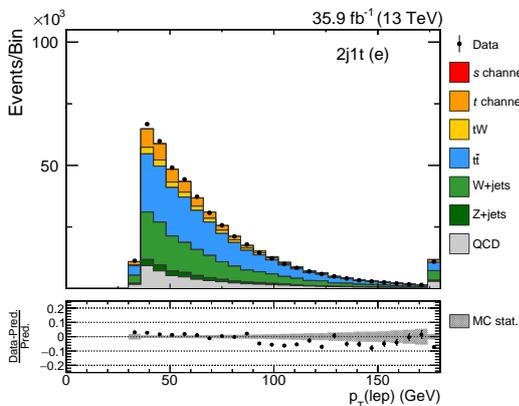
Abbildung C.4.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.



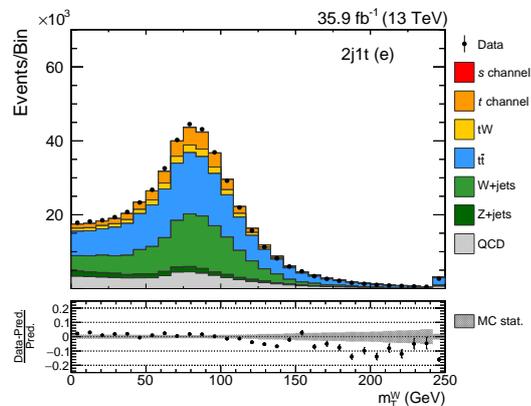
(a) Rang 5



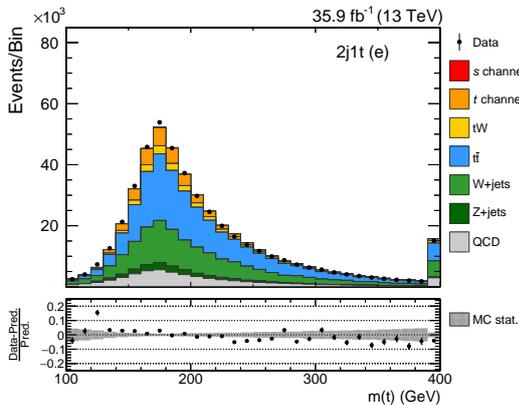
(b) Rang 6



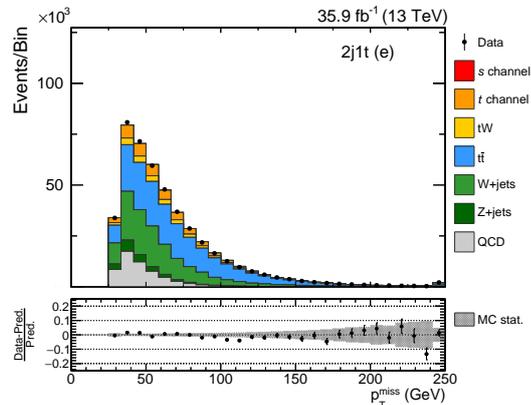
(c) Rang 7



(d) Rang 8

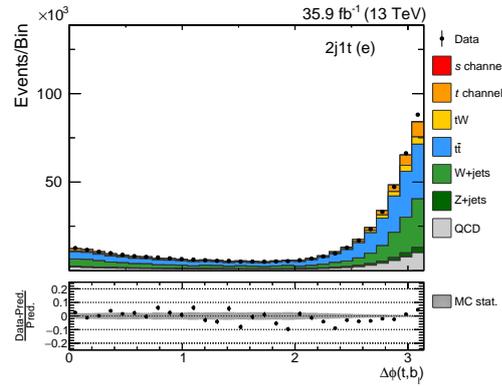


(e) Rang 9

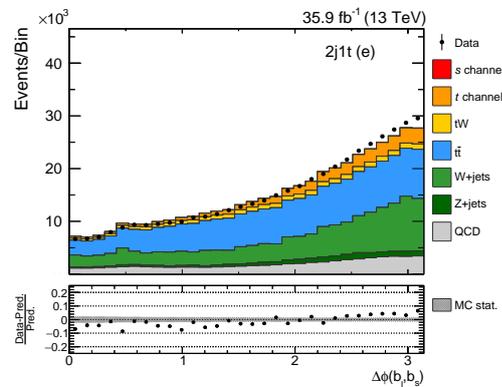


(f) Rang 10

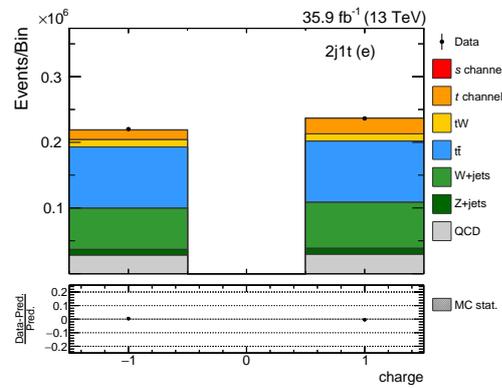
Abbildung C.5.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.



(a) Rang 11



(b) Rang 12



(c) Rang 13

Abbildung C.6.: Die Verteilung der verwendeten Variablen für die kombinierte Klassifikation im Elektron-Kanal der 2j1t-Region, sortiert nach ihrer Trennstärke im DNN-Training. Die Simulation wird so skaliert, dass sie den beobachteten Ereignissen in Daten entspricht. Gute Übereinstimmung zwischen Daten und simulierten Ereignissen wird beobachtet.

Literaturverzeichnis

- [1] M.Tanabashi (Particle Data Group), “2018 Review of Particle Physics,” <http://pdglive.lbl.gov/Particle.action?node=Q007&init=0>, 2018. Online; aufgerufen 08.06.2018.
- [2] V. M. Abazov *et al.*, “Observation of Single Top Quark Production,” *Phys. Rev. Lett.*, vol. 103, p. 092001, 2009.
- [3] T. Aaltonen *et al.*, “First Observation of Electroweak Single Top Quark Production,” *Phys. Rev. Lett.*, vol. 103, p. 092002, 2009.
- [4] Wikipedia Commons, “Standard Model of Elementary Particles , 2010.,” <https://commons.wikimedia.org/w/index.php?curid=11307906>, 2018. Online; aufgerufen 08.06.2018.
- [5] J. Schwinger, “On quantum-electrodynamics and the magnetic moment of the electron,” *Phys. Rev.*, vol. 73, pp. 416–417, 1948.
- [6] J.Schwinger, “Quantum electrodynamics. i.a covariant formulation,” *Phys. Rev.*, vol. 74, pp. 1439–1461, 1949.
- [7] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, “Advantages of the Color Octet Gluon Picture,” *Phys. Lett. B*, vol. 47, pp. 365–368, 1973.
- [8] A. S. Kronfeld and C. Quigg, “Resource Letter: Quantum Chromodynamics,” *Am. J. Phys.*, vol. 78, pp. 1081–1116, 2010.
- [9] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry,” *Phys. Rev. D*, vol. 2, pp. 1285–1292, 1970.
- [10] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.
- [11] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, pp. 1–29, 2012.

- [12] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.
- [13] N. Metropolis and S. Ulam, “The Monte Carlo method,” *Journal of the American statistical association*, vol. 44, pp. 335–341, 1949.
- [14] S. Agostinelli, “GEANT4: A Simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A*, vol. 506, 2002.
- [15] L. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, no. 08, p. S08001, 2008.
- [16] Wikipedia Commons, “Position des Proton Synchrotron (PS) am CERN ,2011.,” <https://commons.wikimedia.org/w/index.php?curid=15291088>, 2018. Online; aufgerufen 14.06.2018.
- [17] S. Chatrchyan *et al.*, “The CMS Experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008.
- [18] Wikipedia Commons, “CMS: Riesenmagnet auf Myonsuche,” <http://www.scinexx.de/dossier-bild-414-10-11508.html>, 2018. Online; aufgerufen 14.06.2018.
- [19] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS*, vol. ACAT, p. 040, 2007.
- [20] Wikipedia Commons, “Schema eines künstlichen Neurons,” https://de.wikipedia.org/wiki/K%C3%BCnstliches_neuronales_Netz#/media/File:ArtificialNeuronModel_deutsch.png, 2018. Online; aufgerufen 20.06.2018.
- [21] Wikipedia Commons, “künstliche neuronale Netze,” <https://www.stemmer-imaging.de/de/tech-tipps/machine-learning-and-deep-learning>, 2018. Online; aufgerufen 20.06.2018.
- [22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Neurocomputing: Foundations of research,” ch. Learning Representations by Back-propagating Errors, pp. 696–699, MIT Press, Cambridge, MA, USA, 1988.

- [23] P. Baldi and P. Sadowski, "The dropout learning algorithm," *Artif. Intell.*, vol. 210, pp. 78–122, May 2014.
- [24] T. Chwalek, N. Faltermann, D. Müller, T. Müller, and F. Schenkel, "Search for the single top s-channel production at 13 TeV," *CMS Note AN-17-216 (v5)*, 2018.