

**Multivariate Analysis of a Top Quark Production  
with an associated Z Boson at the ATLAS  
Experiment corresponding to  $\sqrt{s} = 13$  TeV**

Piet Nogga

Bachelorarbeit in Physik  
angefertigt im Physikalischen Institut

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

Februar 2020

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

Bonn, .....  
Datum

.....  
Unterschrift

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachter: Priv.-Doz. Dr. Philip Bechtle

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical background</b>	<b>3</b>
2.1	The Standard Model . . . . .	3
2.2	The LHC and the ATLAS detector . . . . .	4
2.2.1	Large Hadron Collider . . . . .	4
2.2.2	ATLAS Detector . . . . .	5
2.3	Associated $tZq$ production and their most important backgrounds . . . . .	9
2.3.1	$tZq$ production . . . . .	9
2.3.2	Background events . . . . .	10
2.4	Artificial neural networks . . . . .	11
2.4.1	Structure and Principle of neural networks . . . . .	11
2.4.2	Evaluation methods . . . . .	14
<b>3</b>	<b>Multivariate Analysis</b>	<b>19</b>
3.1	Preprocessing . . . . .	19
3.2	Optimization of Hyperparameters . . . . .	23
3.2.1	Learning rate . . . . .	23
3.2.2	Structure of the neural network . . . . .	25
3.2.3	Validation fraction . . . . .	25
3.2.4	Dropout . . . . .	25
3.3	Results of the neural network . . . . .	25
<b>4</b>	<b>Comparison with NeuroBayes</b>	<b>31</b>
4.1	Network structure of NeuroBayes . . . . .	31
4.2	Comparison of Results . . . . .	31
<b>5</b>	<b>Conclusion and Outlook</b>	<b>37</b>
	<b>Bibliography</b>	<b>39</b>
<b>A</b>	<b>Variable Ranking with Keras</b>	<b>43</b>
<b>B</b>	<b>Keras network code</b>	<b>45</b>
	<b>List of Figures</b>	<b>47</b>

<b>List of Tables</b>	<b>49</b>
<b>Acknowledgements</b>	<b>51</b>

---

# Introduction

---

## Motivation

Mankind's curiosity has led to many great achievements. Starting from observations of nature, new technologies are often developed. And with the progress of these technologies further possibilities emerged to discover the unknown.

Over time, physicists explained the dynamics of macroscopic objects with the theory of classical mechanics and general relativity and continued to describe nature at microscopic scales via quantum mechanics. The latter development accumulated into the Standard Model (SM) of modern particle physics in the second half of the 20th century. Ever since, it was tested and confirmed by countless experiments.

Nevertheless, it is also well known that it does not explain a sufficient amount of matter/antimatter asymmetry and, of course, neither gravitation on microscopic scales. Aiming to further explore these mysteries, the precision of experiments has been enhanced and also the energy ranges of experiments have been expanded in the last decades. Especially, the extended energy ranges led to the confirmation of the top quark ( $t$ ) (1995) [1] and the Higgs boson ( $H$ ) (2012) [2]. Today, the most advanced high energy accelerator is the Large Hadron Collider (LHC) [3] at the European Organization for Nuclear Research (CERN) which went into operation in 2009. With its four sophisticated detectors, it provides a formidable instrument to validate the SM on the one hand, but, on the other hand, also gives access to analyze open questions in particle physics.

One of the four detectors at LHC is ATLAS [4]. As the energy of particle collisions are converted into mass, ATLAS is designed to detect the large amount of outgoing particles. The data collected largely contributes to the analysis done in particle physics. Its high precision measurements give insight into many SM parameters.

As nature forbids direct measurements of single quarks [5], free quarks are instead reconstructed with the produced particles via hadronization [6]. The hadrons produced are often from narrow cones of hadrons, so called jets, that can be identified [7] and allow reconstruction of the quark triggering it.

The heaviest fundamental particle, the top quark, produced many times during the runtime of the LHC, has been studied in various productions. The top quark plays a key role in high precision measurements of the SM. Due to its high mass, its lifetime is only  $1 \times 10^{-25}$  s. Therefore, unlike any other quark, it decays before the process of hadronization sets in and can be experimentally observed as a single quark without being bound into a hadron. The large mass also indicates that it has a large coupling to Higgs bosons. For that reason, the top quark seems promising in searches for physics

beyond the SM [8, 9].

But the identification of collisions producing top quarks and, especially, processes involving a top quark is not easy. First, the reconstruction of the jets is necessary. To identify rare associated top quark productions, like the associated  $tZq$  production [10], several of these jets have to be related to lighter quarks and, additionally, corresponding gauge bosons, like  $Z$  and  $W$ , have to be identified at the same time. Processes like the associated  $tZq$  production therefore only became accessible after upgrading to the record centre-of-mass energy of  $\sqrt{s} = 13$  TeV. These are of special interest to probe the coupling of the  $Z$  boson and the  $t$  quark. Additionally, this process can serve as an important background for rare  $tH$  productions. In order to learn more about this interesting production process, this thesis focuses on the identification of associated  $tZq$  events.

In order to cope with the challenging task of distinguishing events measured by ATLAS and relate the measured data to a process like associated  $tZq$ , one relies on recent developments in the field of computer science. First proposed in 1943 [11], machine learning gained much popularity over the years. Applications like image or speech recognition are well known to the general public. But artificial neural networks (NN) have also proven to be a useful tool to deal with the increasing complexity of physical tasks being of particular interest in particle physics. They are by now a standard tool for the identification of events in this field.

Also, associated  $tZq$  productions have already been investigated based on NNs. The corresponding multivariate analysis has thereby been implemented with NeuroBayes [12]. In this thesis, a different programming package, Keras [13], is used. It will be studied in detail whether the different implementation can keep up or even exceed the NeuroBayes performance for the identification of associated  $tZq$  events. In this case, experiments could be studied by different NN algorithms.

### Thesis overview

The thesis is structured as follows. The theoretical background to understand the physical and technical concepts behind this thesis is discussed in Chapter 2. This foundation is laid by giving an introduction to the SM, overviewing the LHC and ATLAS detector and explaining the principle of NNs, including methods to evaluate the success of the training procedure.

Chapter 3 describes the multivariate analysis done in this thesis. Before showing the discrimination results of this analysis, the training input has to be preprocessed and a set of user-defined parameters for the neural network has to be adjusted to the task of  $tZq$  identification. For this, the evaluation methods explained in Chapter 2 are taken into account.

A detailed comparison to the existing NeuroBayes artificial neural network is given in Chapter 4. Here, the training quality and network discrimination is considered. Finally, the overall performance of both networks are compared to answer the question if Keras is suitable for a long term NeuroBayes replacement.

The last chapter gives a conclusion and outlook of this thesis.

---

## Theoretical background

---

In order to understand the theoretical concepts of this thesis, the chapter first focuses on giving the reader a brief introduction to the Standard Model (SM) of particle physics. The Large Hadron Collider and the ATLAS detector with its components are discussed in the second section. Then, the  $tZq$  process and its most important backgrounds are explained. Lastly, this chapter covers the concept of machine learning and methods to evaluate neural networks.

### 2.1 The Standard Model

The SM of modern particle physics explains all known particles without any further substructure and their interactions with one another. So far, the predictions of the SM have been very precise and have been verified by many experiments.

The SM predicts exactly two types of fundamental particles. The fermions and the bosons. While matter consists only of fermions, the bosons are often referred to as the force carrier particles as they describe three of the four fundamental forces; The strong, weak and electromagnetic interactions excluding only the gravitational interaction.

The last missing pieces predicted by the SM have been discovered quite recently, including the top quark (1995) [1], the tau neutrino (2000) [14] and the Higgs boson (2012) [2] which was first detected by the ATLAS detector. Figure 2.1 shows the properties of the fundamental particles. Additionally, each particle has an associated antiparticle with the same spin and mass but opposite physical charges such as the electrical charge.




























The upper part of the figure shows the three generations of fermions, consisting of six leptons and six quarks, with their electrical charge, mass and possible interactions while the lower part describes the gauge bosons which are related to the fundamental forces. The quark masses are scheme dependent and are taken from Ref. [15].

The particle responsible for the electromagnetic interaction is the photon which couples to all electrically charged particles. The electromagnetic interaction between an orbital electron and the atomic nucleus keeps an atom stable.

The  $W^\pm$  and  $Z$  gauge bosons mediate the weak interaction. These particles are massive and act on all matter particles. This force is most commonly known for  $\beta$  decays like  $p \rightarrow n + e^+ + \nu_e$ .

Lastly, the gluon mediates the strong force. It only acts on color charged particles (quarks and the gluon itself). It is ultimately responsible for confining quarks into hadrons. The recently found Higgs

boson is an exception as it is the only spin 0 boson. Simplified, it explains why all elementary particles but the photon and gluon are massive. Historically, this was a problem in the SM (for the  $W$  and  $Z$  gauge bosons). The Higgs boson is also massive, hence it interacts with itself [16, 17].

Leptons				Quarks				
	Particle	Q	Mass	Interaction	Particle	Q	Mass	Interaction
First generation	electron ( $e^-$ )	-1	0.511 MeV	 	down ( $d$ )	-1/3	4.7 MeV	  
	neutrino ( $\nu_e$ )	0	< 2 eV		up ( $u$ )	+2/3	2.2 MeV	  
Second generation	muon ( $\mu^-$ )	-1	105.7 MeV	 	strange ( $s$ )	-1/3	96 MeV	  
	neutrino ( $\nu_\mu$ )	0	< 2 eV		charm ( $c$ )	+2/3	1.28 GeV	  
Third generation	tau ( $\tau^-$ )	-1	1776.9 MeV	 	bottom ( $b$ )	-1/3	4.18 GeV	  
	neutrino ( $\nu_\tau$ )	0	< 2 eV		top ( $t$ )	+2/3	173.1 GeV	  

Interaction	Boson	Strength	Spin	Mass	Higgs boson
Strong	Gluon ( $g$ )	1	1	0	<div style="border: 1px dashed orange; padding: 5px;">                     Mass 125.1 GeV Spin 0                 </div>
Weak	W boson ( $W^\pm$ )	$10^{-8}$	1	80.4 GeV	
	Z boson ( $Z^0$ )			91.2 GeV	
Electromagnetism	photon ( $\gamma$ )	$10^{-3}$	1	0	

Figure 2.1: Overview over the current Standard Model of modern particle physics [18]. The type of interaction is shown in the respective color. The Higgs interaction is displayed by the dotted orange line

With all the success the SM had in recent history, there is still research to do. As of today, the SM cannot explain gravitation. Furthermore, the model is inconsistent with the Lambda-cold dark matter ( $\Lambda$ -CDM) model of the universe and therefore it cannot explain dark matter [19] and the matter/antimatter asymmetry [20].

## 2.2 The LHC and the ATLAS detector

### 2.2.1 Large Hadron Collider

The LHC is a circular high energy particle collider located near Geneva. It was built by CERN between 1998 and 2008.

In order to collide proton beams (or heavy ion beams) at centre-of-mass energies up to 13 TeV (1 045 TeV), the particles must be sent through multiple other accelerators beforehand, as seen in Figure 2.2. Once in the LHC, the bunches are kept on a circular orbit by dipole magnets. Additionally, quadrupole magnets are used to focus the beam. A complex cooling system is installed to keep the



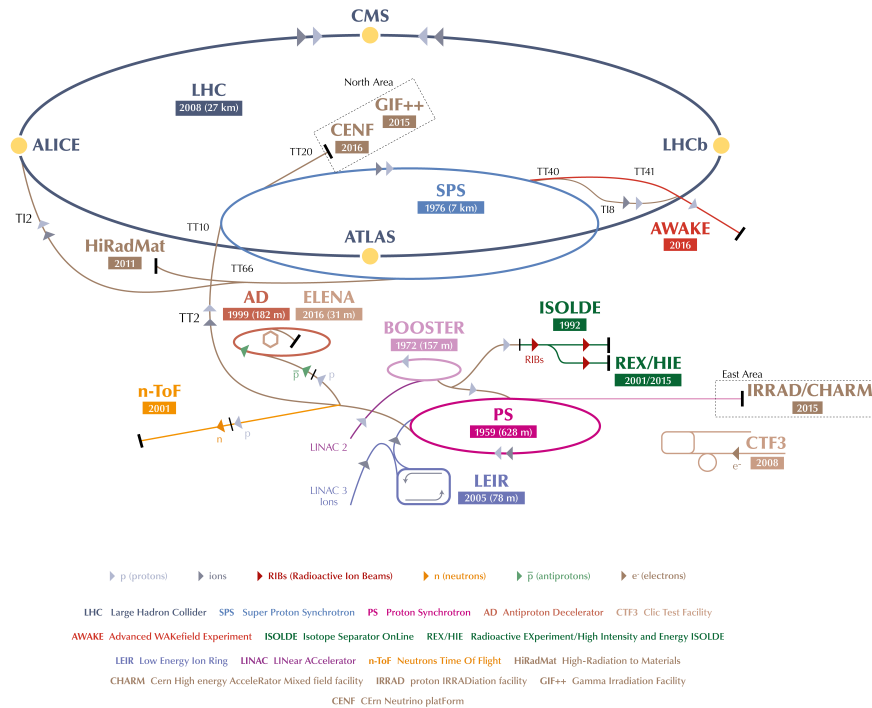


Figure 2.2: Overview of the accelerators of the Large Hadron Collider [21]. The protons (heavy ion beams) are first accelerated in LINAC2 (linear accelerator) and three circular accelerators (Proton Synchrotron Booster, Proton Synchrotron and Super Proton Synchrotron) before being injected into the LHC

superconducting magnets at their working temperature of 1.9 K. Using accelerating cavities and electromagnetic resonators, the protons are accelerated to high velocities being only  $3.1 \text{ m s}^{-1}$  slower than the speed of light.

After reaching their maximum energy, the two beams circulating in opposite directions intersect at one of four points where the main detectors analyze the collision. These detectors are ATLAS, CMS, ALICE and LHCb. This thesis uses Monte Carlo simulated data based on the ATLAS detector [22].

### 2.2.2 ATLAS Detector

In order to analyze the events occurring at the interaction points, a complex structure of detectors is necessary. The largest particle detector at CERN is the general purpose ATLAS detector with a length of 44 m and a diameter of 25 m weighing about 7 000 t. An overview of ATLAS and its components can be seen in Figure 2.3. The following information is taken from Ref. [4].

#### The ATLAS coordinate system

The interaction point of the event is taken as the origin of the ATLAS standard coordinate system. Then, the beam direction and the transverse plane of the beam line define the z-axis and the x-y plane,

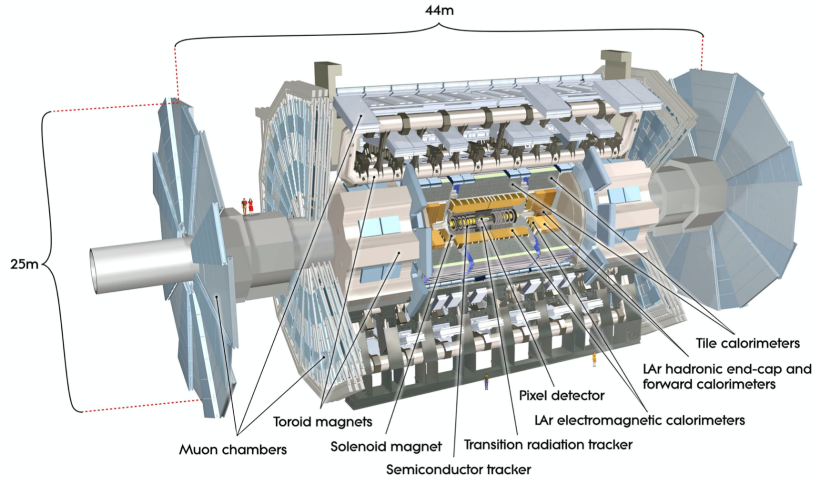


Figure 2.3: Overview of the structure of the ATLAS detector [4]

respectively. The positive  $x$ -axis points to the centre of the LHC while the positive  $y$ -axis is defined as pointing upwards and  $r = \sqrt{x^2 + y^2}$ . The transverse azimuthal angle  $\phi$  is measured around the  $z$ -axis, and the polar angle  $\theta$  is the angle from the  $z$ -axis. The more common way to describe the polar angle is by introducing the pseudorapidity  $\eta$ . It is defined as

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

Distances in the pseudorapidity-azimuthal angle space are then calculated by

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

Figure 2.4 shows the definition of the ATLAS coordinate system.

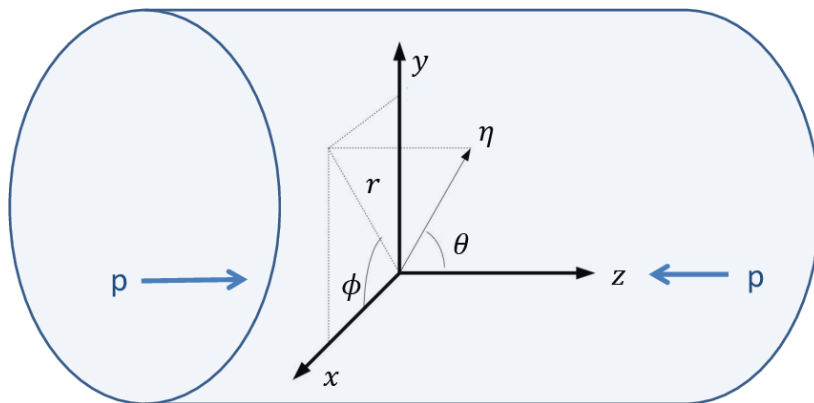


Figure 2.4: The ATLAS coordinate system [23]. The cylinder illustrates the ATLAS detector and  $p$  shows the colliding particles and their respective directions

### The Inner Detector

The innermost component of the ATLAS detector is the inner detector (ID). It is responsible for tracking electrically charged particles and measuring their momentum. By surrounding the inner detector with a magnetic field of 2 T, the direction of the curve reveals the particles charge and the curvature gives information about its momentum. The ID covers a range of  $|\eta| < 2.5$ . This wide tracking capability is achieved by layering the ID in several subdetectors. These can be seen in Figure 2.5. The pixel detectors are located at the interaction point, practical for identifying particles

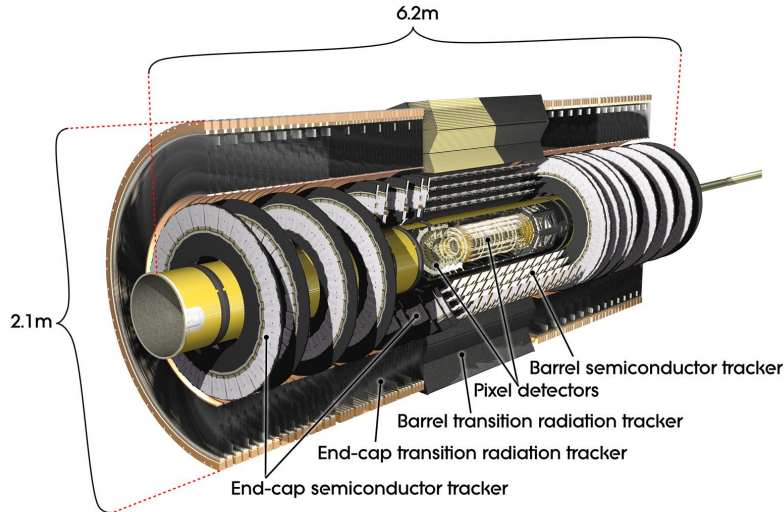


Figure 2.5: Schematic view of the subcomponents of the Inner Detector [24]

in dense environments. It consists of four barrel layers and three endcap disks with a total of 1744 modules. Each module scans an area of 2 by 6 centimeters and contains roughly 47000 pixels. The readout of each pixel is about 50 by 400 micrometers totaling about 80 million readout channels.

The semi-conductor tracker works similarly to the pixel detectors but rather relies on long, narrow strips and aims to detect a larger area. The strips measure  $80 \mu\text{m} \times 12 \text{ cm}$ . 6.3 million readout channels are used to analyze the four double layers of silicon strips.

The largest part of the ID is reserved for the transition radiation tracker taking up a volume of about  $12 \text{ m}^3$ . Drift tubes with a diameter of 4 mm and a length of 144 cm detect charged particles. Overall, the transition radiation tracker has 298000 straws.

### Electromagnetic Calorimeter

Measuring energy of particles is essential to particle identification. The electromagnetic calorimeter (ECAL) aims to absorb the energy of the particles interacting electromagnetically (charged particles and photons).

Electrons(or positrons) passing through matter will emit Bremsstrahlung when entering the field of the nucleus. Photons with energies of  $E_\gamma > 10 \text{ MeV}$  near the nucleus predominantly create an electron-positron pair via pair production. This leads to an electromagnetic shower. By alternating a lead or stainless steel absorber material and a liquid argon scintillator, the electromagnetic shower

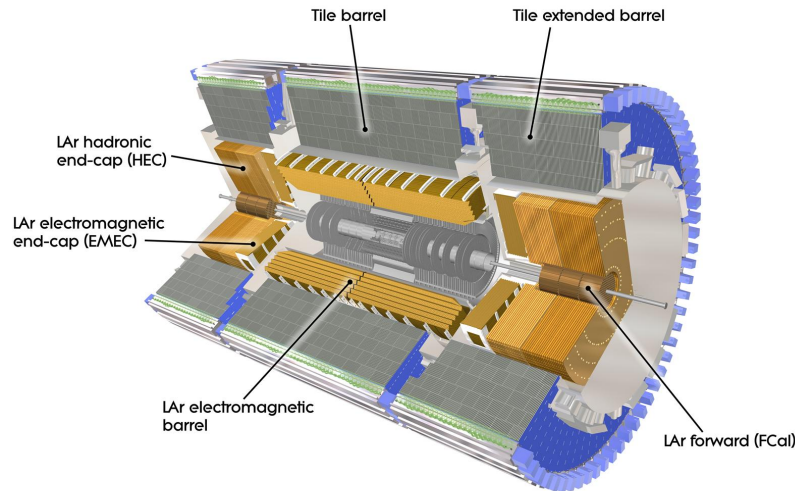


Figure 2.6: Overview of the ATLAS calorimeter system [25]. The Tile barrel and two Tile extended barrels form the Tile Calorimeter (TileCal).

is contained in the calorimeter as soon as the particle loses energy through ionization rather than Bremsstrahlung. In order to minimize the size of the ECAL, the absorber's material is chosen with a high atomic number  $Z$ , as the distance  $X_0$ , by which the particle's energy decreases by the factor  $1/e$ , is inverse proportional to  $Z$ .

The calorimeter is further divided into a barrel part (liquid argon (LAr) electromagnetic barrel) and two end cap components (LAr electromagnetic end-caps), covering the pseudorapidity of  $|\eta| < 1.475$  and  $1.375 < |\eta| < 3.2$ , respectively.

As heavier charged particles, like muons and hadrons, do not emit as much Bremsstrahlung as electrons, they will barely leave any traces in the ECAL.

### Hadronic Calorimeter

Due to their higher mass, particles interacting with the strong force pass the ECAL and their energies are therefore measured with the hadronic calorimeter. Hadrons passing matter will lose their energy through ionization or strong interaction with the material's nuclei. The resulting particles will also lose energy while being in the material. In consequence, a hadronic shower occurs.

The hadronic calorimeter is built of the tile calorimeter (TileCal), the LAr hadronic end-cap calorimeter and the LAr forward calorimeter.

The TileCal is responsible for the central part, covering pseudorapidity regions of  $|\eta| < 1.7$ . The calorimeter is structured like the ECAL but using steel as the absorber and plastic scintillating tiles as active material.

A liquid argon calorimeter with copper as an absorber measures the energy between  $1.5 < |\eta| < 3.2$ . This LAr hadronic end-cap calorimeter is located, as the name suggests, directly behind the end-cap electromagnetic calorimeter.

Lastly, the LAr forward calorimeter absorbs energy in the pseudorapidity range of  $3.1 < |\eta| < 4.9$ . It has several layers using first copper and then tungsten as absorber. Tungsten is predominantly used

to measure interactions of hadrons.

Figure 2.6 shows the calorimeter system and their respective parts.

### Muon Spectrometer

With muons being the only long lived particles not sufficiently interacting in the ID nor in the calorimeters for detection, an extra muon spectrometer is necessary to measure their transverse momentum. The spectrometer is designed to measure high momentum particles of 100 GeV with 3 percent accuracy.

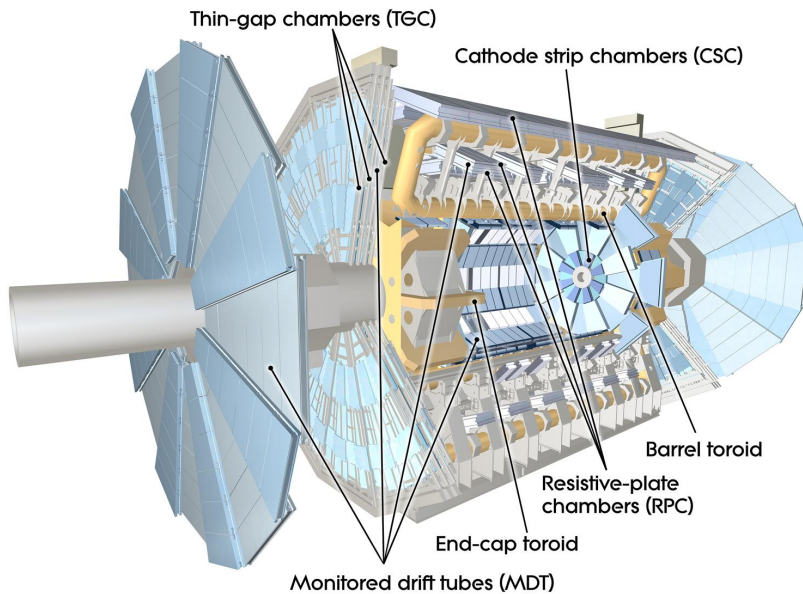


Figure 2.7: Structure of the Muon Spectrometer [26]

Similar to the ID, the muon spectrometer uses magnets to bend their trajectories up to  $|\eta| < 2.7$ . From the degree of curvature, one concludes the momentum of the muon. Additionally, it has its own muon trigger system further helping ATLAS identify muons with about 1 million readout channels. The complete structure is seen in Figure 2.7.

The only known particles not detected by ATLAS are neutrinos as they do not interact electromagnetically or strongly. Due to conservation of momentum, any net transverse momentum measured by the detector indicates missing transverse energy  $E_T^{\text{miss}}$  carried by undetected particles. Information about neutrinos is therefore extracted from  $E_T^{\text{miss}}$  [27].

## 2.3 Associated $tZq$ production and their most important backgrounds

### 2.3.1 $tZq$ production

This thesis focuses on the top quark production with an associated  $Z$  boson. This process has a small cross-section making it a challenging process to analyze. Since the upgrade to  $\sqrt{s} = 13$  TeV in 2015, a

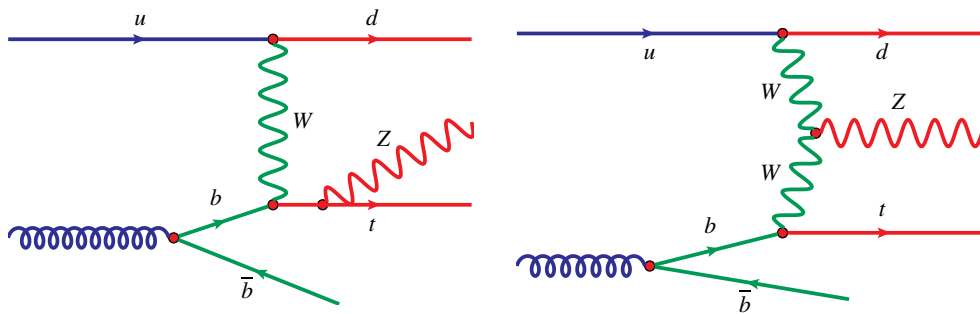


Figure 2.8: Example Feynman graphs for the associated production of a top quark and a  $Z$  boson

considerable amount of associated top quark productions have been detected [28]. The  $tZq$  production can, for example, provide information about the  $t\bar{t}Z$  or  $WWZ$  coupling or contribute to the background for the rare  $tH$  production. The Feynman diagrams for leading order (LO)  $tZq$  productions can be seen in Figure 2.8 and involve the weak interaction in the  $t$ -channel. The  $Z$  boson can either radiate from the  $W$  boson or in principle from any of the given quarks. Evidence of the first associated  $tZq$  production was found in 2018 by the ATLAS collaboration [10].

As the top quark and the  $Z$  boson have a very small lifetime, only the final state can be detected. It can consist of only hadrons or up to three leptons, depending on the decay of the top quark and the  $Z$  boson. Although the three lepton channel has the smallest branching ratio, this thesis only covers a leptonically decaying top quark and  $Z$  boson as this channel reduces the background dramatically [29].

A  $tZq$  final state therefore consists of three leptons, one  $b$  jet, one or two additional untagged jets, referred to as  $2j1b$  and  $3j1b$  signal regions, respectively. Jets are particle cones produced via hadronization of a quark or gluon and are reconstructed by ATLAS [7]. In case of the  $3j1b$  signal region, the forward jet is denoted as the untagged jet for which the invariant mass is largest. The other untagged jet is then labelled as radiation jet [22]. Furthermore, as neutrinos are involved (from the  $W$  boson which resulted from the top quark decay), the final state also has missing transverse energy  $E_T^{\text{miss}}$ .

### 2.3.2 Background events

Analyzing events involving a  $t$  quark or a  $Z$  boson means identifying and measuring the final state of the event. Consequently, other productions with a similar final state may be misinterpreted as an associated  $tZq$  production. These background events are discussed in the following.

#### Diboson

The largest background is the diboson production [22]. The produced bosons, especially the  $WZ$  productions may then decay leptonically and lead to an identical final state if additional jets are present. The  $WW$  and the  $ZZ$  productions contribute less [18]. However, a reconstruction failure is possible so that these events also occur in  $tZq$  selection. Example diboson Feynman graphs are displayed in Figures 2.9(a) and 2.9(b).



### $t\bar{t}V$ and $tWZ$

$t\bar{t}Z$  productions already have a  $t$  quark and  $Z$  boson and therefore have a similar final state as the  $tZq$  productions. This background then often passes selection. For the  $t\bar{t}H$  events, the  $H$  boson rarely decays leptonically [18] and has a small cross-section so that this background has a very small contribution. A Feynman graph for  $t\bar{t}Z$  productions is laid out in Figure 2.9(c).

The  $tWZ$  production has never been measured experimentally and has a very small predicted cross-section [30]. The  $tWZ$  contribution is still important but due to its similar kinematic nature to the  $t\bar{t}V$  productions, these productions are kept together [29]. A  $tWZ$  production Feynman graph is seen in Figure 2.9(d).

### $t\bar{t}$ and $tW$

The cross-section of this production is orders of magnitude higher than  $tZq$  productions and gives a considerable background. A  $t\bar{t}$  pair production can be seen in Figure 2.9(e).

The  $t\bar{t}$  and  $tW$  productions are grouped due to the almost negligible contribution of  $tW$  and the difficulty to separate both samples [29]. For the sake of completeness, a Feynman graph of  $tW$  productions is given in Figure 2.9(f).

To separate signal ( $tZq$  productions) and background events efficiently, it is necessary so study the kinematic variables of the final states. A table of such variables can be found in Section 3.1.

The MC data used below will contain a sufficient amount of these backgrounds to test the identification capabilities of the NN.

## 2.4 Artificial neural networks

### 2.4.1 Structure and Principle of neural networks

One aim of the thesis is to discriminate signal and background. For this, a NN is implemented. This subsection lays the foundation to understand how a NN works.

Principally, a NN consists of several layers; An input layer, hidden layers and an output layer. The computation unit inside a layer is the neuron, also called the node. The width of the network, i.e. the number of nodes in a specific layer, varies. The input layer represents the knowledge before training. In this thesis, they are given by the kinematic variables of the event. As the output aims to discriminate signal and background, it is chosen to be between 0 and 1. The number of nodes in the hidden layers determines the complexity of the network and has to be adjusted to the difficulty of the task. Figure 2.10 shows a schematic structure of a basic NN.

The value of the nodes in the hidden layers depends on all nodes  $x'_1, x'_2, \dots, x'_n$  in the previous layer. They are calculated as follows:

$$x_j = F \left( b_j + \sum_{i=1}^n w_{ij} \cdot x'_i \right)$$

where  $b_j$  denotes a bias term to further increase the degree of freedom.  $w_{ij}$  corresponds to a weight which is multiplied to the value of a node in the previous layer. The activation function  $F(x)$  makes the system nonlinear and is crucial to the network [32]. This function depends on what one wants to achieve. For this thesis, the hidden layers use  $F(x) = \max(0, x)$ , also called Rectified Linear Unit

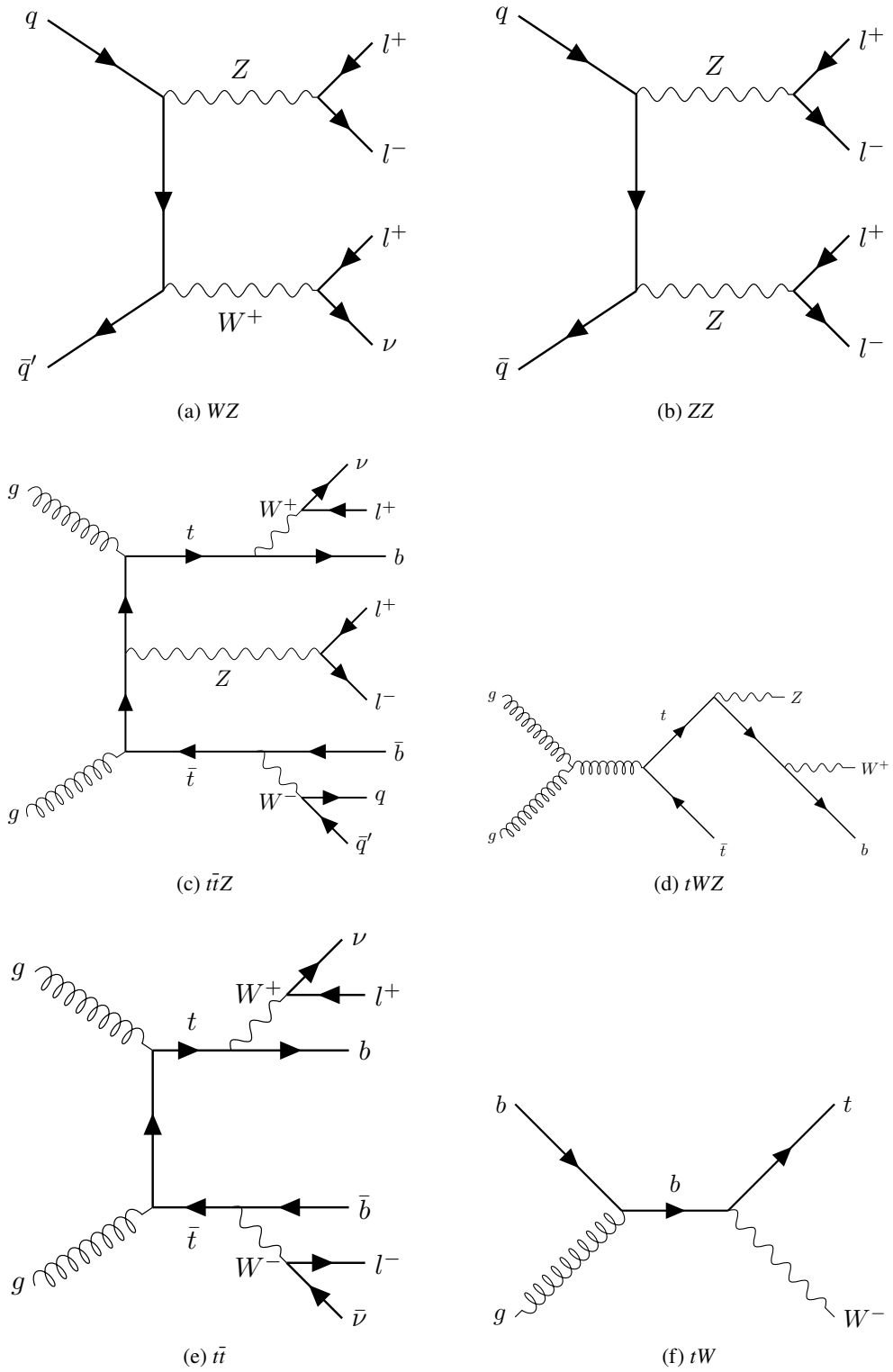


Figure 2.9: Feynman graphs of the background sources considered in  $tZq$  multivariate analysis [31]



(ReLU) [11]. The output layer is handled differently. Here, the Sigmoid activation function is used to give the NN its output in the range of 0 and 1 [33]. The Sigmoid function is given by

$$F(x) = \frac{1}{1 + e^{-x}}$$

For a NN to train a model, the output of a training sample are compared to a known data set (supervised

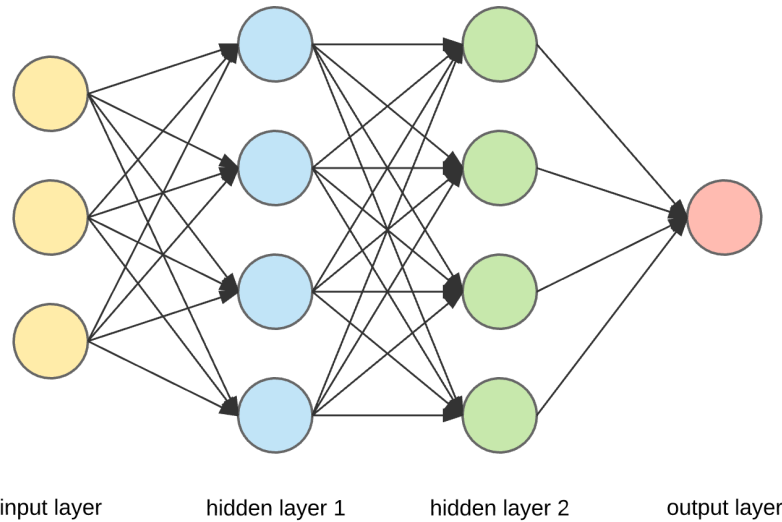


Figure 2.10: The Structure of a simple neural network. The circles represent nodes and are aligned to model the different layers. The arrows indicate the dependencies of the nodes [34]

learning). It constantly calculates a loss value, which is a measure of how well the network performs. For binary classifications, loss values  $L_p$  are defined with the binary crossentropy function [35]

$$L_p(y_i) = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))]$$

where  $y_i$  is the label of the known sample (1 for signal and 0 for background) and  $p(y_i)$  denotes the value of the output node.

This value is essential for the training of the NN. The principle of every NN is to minimize  $L_p(y_i)$  by updating the weights  $w_{ij}$  on a batch. The update frequency is chosen beforehand by defining a batch size.

The value of the updated weight  $w'_{ij}$  can be computed with the Stochastic Gradient Descent (SGD) [36]:

$$w'_{ij} = w_{ij} - \eta \cdot \nabla_{w_{ij}} L_p(y_i)$$

where  $\eta$  corresponds to the learning rate. One must have  $\eta > 0$ . This algorithm is also known as backpropagation [37]. One computes the gradient of the loss with respect to the weights in order to minimize the error.

The SGD method is only one out of many optimizers. The NN in this thesis is implemented with the Adam (Adaptive Moment) optimizer [38]. In contrast to SGD, Adam changes the learning rate for

every parameter in the batch. The idea is to adapt the learning rate by a running average by both the gradient and the second moment of the gradient. If  $t$  indicates the current training iteration, then the running averages are calculated by

$$m_{w_{ij}}^{(t+1)} = \beta_1 m_{w_{ij}}^{(t)} + (1 - \beta_1) \nabla_{w_{ij}} L_p^{(t)}(q)$$

$$v_{w_{ij}}^{(t+1)} = \beta_2 v_{w_{ij}}^{(t)} + (1 - \beta_2) \left( \nabla_{w_{ij}} L_p^{(t)}(q) \right)^2$$

where  $m^0 = v^0 = 0$ .  $\beta_1$  and  $\beta_2$  are recommended to be 0.9 and 0.99, respectively [38]. Further defining

$$\hat{m}_{w_{ij}} = \frac{m_{w_{ij}}^{(t+1)}}{1 - (\beta_1)^{t+1}}$$

$$\hat{v}_{w_{ij}} = \frac{v_{w_{ij}}^{(t+1)}}{1 - (\beta_2)^{t+1}}$$

the update on the weight is then given by

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\hat{m}_{w_{ij}}}{\sqrt{\hat{v}_{w_{ij}} + \epsilon}}$$

$\epsilon$  is a number introduced to avoid division by zero. The default value of the Adam optimizer is  $\epsilon = 1 \times 10^{-8}$  [38].

## 2.4.2 Evaluation methods

Evaluation is an important step in machine learning as the training can fail and therefore predict incorrectly. The most common ways to double-check the model are as follows:

### Evaluation of loss curve

As the NN passes through an epoch, i.e. running the entire dataset, the average loss value of that epoch can be calculated. Plotting the epochs against the loss value gives the loss curve. It visualizes the learning process of the model. An example can be found in Figure 2.11(a). If the curve behaves as expected, the loss should decrease monotonically. In order to get an unbiased result from the model, another, to the model unknown sample (validation sample) also passes through the final network. The shape of the corresponding validation curve should not differ significantly from the training curve. For that reason, it is important to plot both validation and training curves.

Under some circumstances, the model can overtrain [39]. This usually happens if the network is chosen to be too complex, e.g. too many nodes or hidden layers. Then, the NN may learn non-existing patterns in the training dataset and fails to predict any future samples applied to the model. Typically, the loss of the validation sample then starts increasing again while the training error continues to decrease. An exemplary overtrained model is seen in Figure 2.11(b).

One way to prevent overtraining is by introducing dropout. It randomly ignores the output of the selected nodes, along with the incoming and outgoing connections, on a batch in training. This

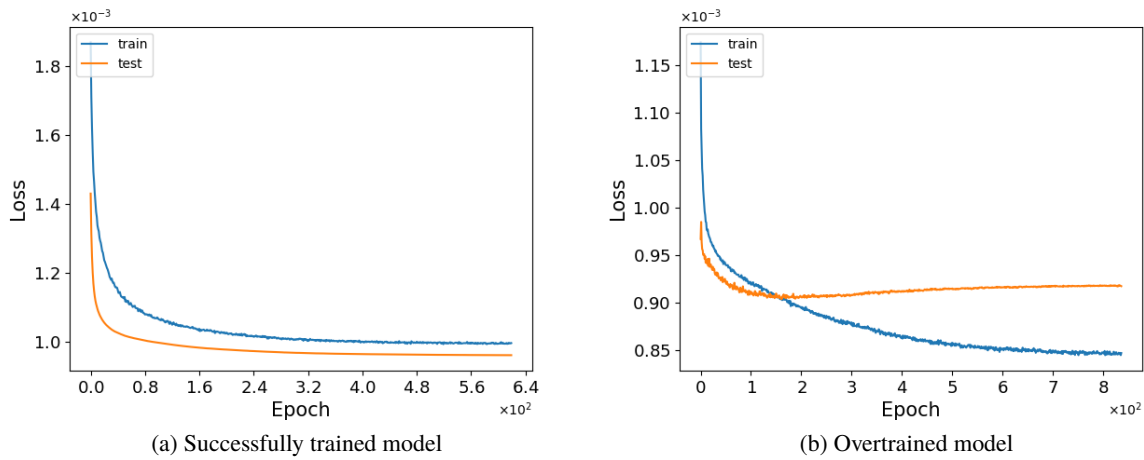


Figure 2.11: Loss curves of different models. The blue and orange line show the training and test samples, respectively

forces the NN to adapt during each batch making it less probable to overtrain. Nodes in a layer are denoted with a fixed probability that they are ignored. This is called the dropout  $d$ . For the validation sample, all computation units are used. Instead, the output values are scaled with  $d$  to match the expected output at training times [40].

### Accuracy metric

As the loss value does not represent the probability of predicting an event correctly, the accuracy metric is defined. It is given by the ratio of correctly predicted events and total events, i.e.

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}$$

where P refers to total signal, and N to total background events. TP and TN then correspond to correctly predicted signal and background events, respectively. Here, an event is associated as signal ( $tZq$ ) if the NN output reads  $O_{\text{NN}} \geq S_T = 0.5$ .

Figure 2.12(a) displays an accuracy curve and its expected shape. In this example the accuracy approaches about 78%, corresponding to the probability the model predicts signal or background correctly. For a perfect model, i.e. a flawless separation is achieved, the accuracy is 100%. Nevertheless, this evaluation method might not be sufficient as it is strongly dependent on the ratio of signal and background samples. If the amount of background samples were many orders of magnitude higher than the signal events, a NN always choosing background would yield a high accuracy.

This gives rise to develop other methods which are independent of the signal-background event yield ratio. One is discussed below.

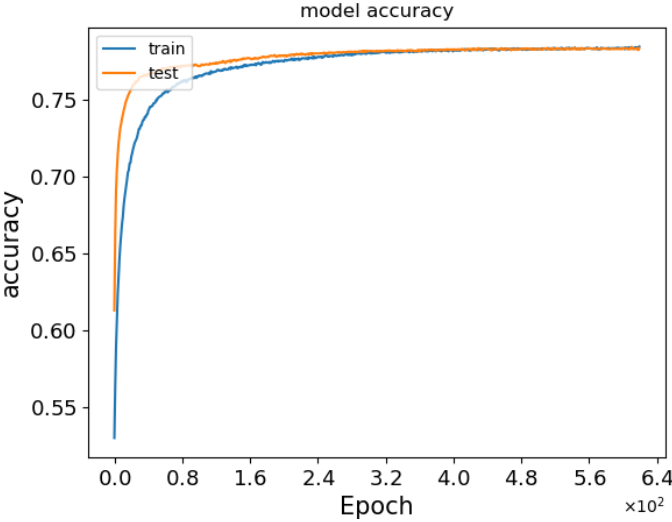
### Receiver Operating Characteristic Curve

In a binary classification, P and N are used to calculate the true positive rate (TPR) and false positive rate (FPR). These are defined as

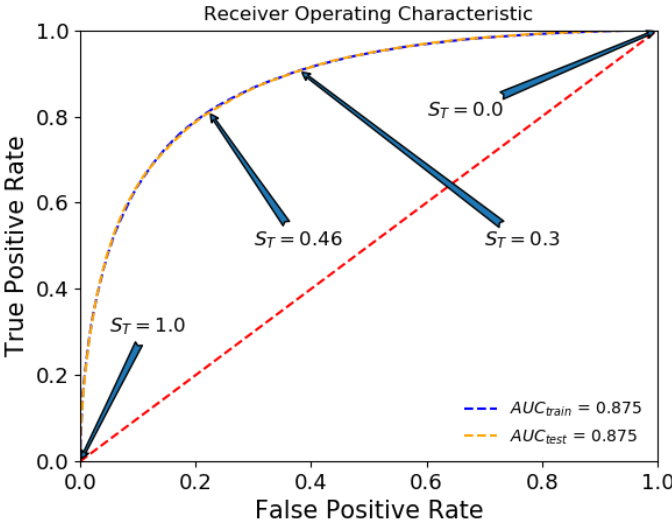
$$\text{TPR} = \frac{\text{TP}}{\text{P}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{N}}$$

where FP is the number of falsely predicted signal events. Plotting TPR against FPR for different signal thresholds  $S_T$  results in the Receiver Operating Characteristic Curve (ROC-Curve). The ROC-Curve is another method of evaluating the trained model. A perfect model detects all true signal events with a FPR of zero. If the ROC-Curve is near the diagonal, this can be interpreted as randomly guessing events. The Area under the Curve (AUC) gives an estimation of the performance [41] of the model so that a perfect model has an AUC value of 1. An example ROC-Curve and the corresponding AUC value is laid out in Figure 2.12(b).



(a)



(b)

Figure 2.12: Further evaluation methods of neural networks. On the top and bottom, the accuracy metric and the Receiver Operating Characteristic Curve are shown, respectively. For the ROC-Curve, the arrows denote the FPR-TPR pairs at different signal thresholds  $S_T$ . The corresponding AUC value is also displayed.



---

## Multivariate Analysis

---

The artificial neural network used in this thesis is programmed with Keras [13], an open source library written in Python. This chapter gives insight into the implementation of the NN by first discussing the preprocessing of the kinematic variables used for the network. Furthermore, every problem one attempts to solve with machine learning requires a good set of hyperparameters. Their impact and optimization is examined in the second section of the chapter. Section 3.3 gives an overview of the results of this analysis.

### 3.1 Preprocessing

The samples used for training are Monte Carlo (MC) simulated and correspond to ATLAS data taken from  $pp$  collisions at  $\sqrt{s} = 13$  TeV with an integrated luminosity of  $139 \text{ fb}^{-1}$  in the period of 2015 to 2018. Details about the simulated data can be found in Ref. [22]. To match the simulation with the data taken from ATLAS, the samples were reweighted. Furthermore, the thesis only considers the  $3j1b$  signal region.

In order to choose the variables for training, the variables are omitted step by step and the loss of correlation to the target (0 for background and 1 for signal) due to its removal is calculated. Only significant (correlation loss  $C_L > 2.5\%$ ) variables are used for training. For the variables in this analysis the ranking is done in Ref. [29] and listed in Table 3.1. The corresponding distributions are displayed in Figures 3.1, 3.2 and 3.3.

The neural network might perform poorly if the values of the variables are of different magnitude. For that reason, the data is normalized with the Standard Scaler from Scikit [42]. The standard score  $z$  is calculated as

$$z = \frac{x - u}{s}$$

with the sample  $x$ , the mean of the training samples  $u$  and the standard deviation  $s$ . The test sample is then fitted and transformed in that manner.

The MC weights are also rescaled with respect to the large amount of background samples compared to  $tZq$  three lepton events so that the signal is weighted stronger in training. With the variables prepared for training, the NN can be built using the Keras Sequential model [13]. The samples are split, using a fraction for training and the rest for validation data. Training stops after the loss does not decrease by  $1 \times 10^{-6}$  for 100 epochs. This is done by using the Keras EarlyStopping module. Keras

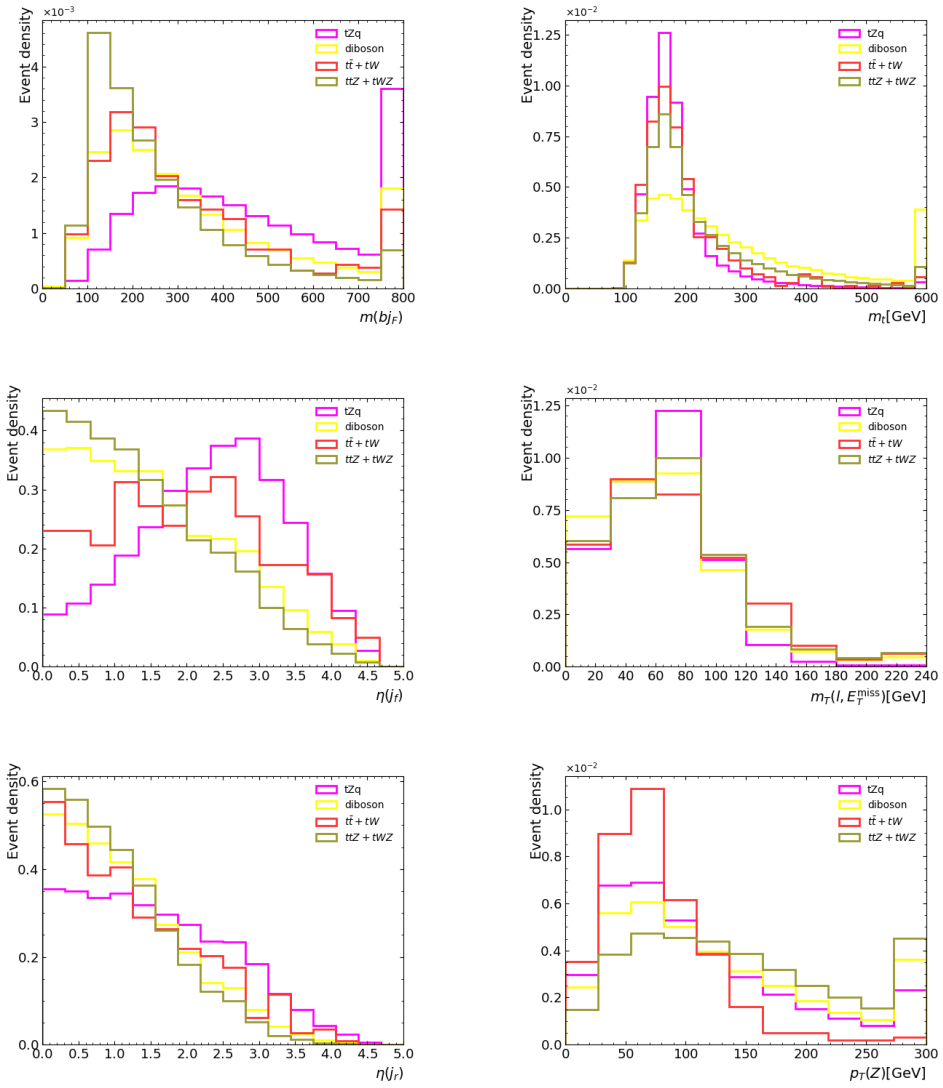


Figure 3.1: Kinematic distributions in order of their significance in the  $3j1b$  signal region, normalized area to unity



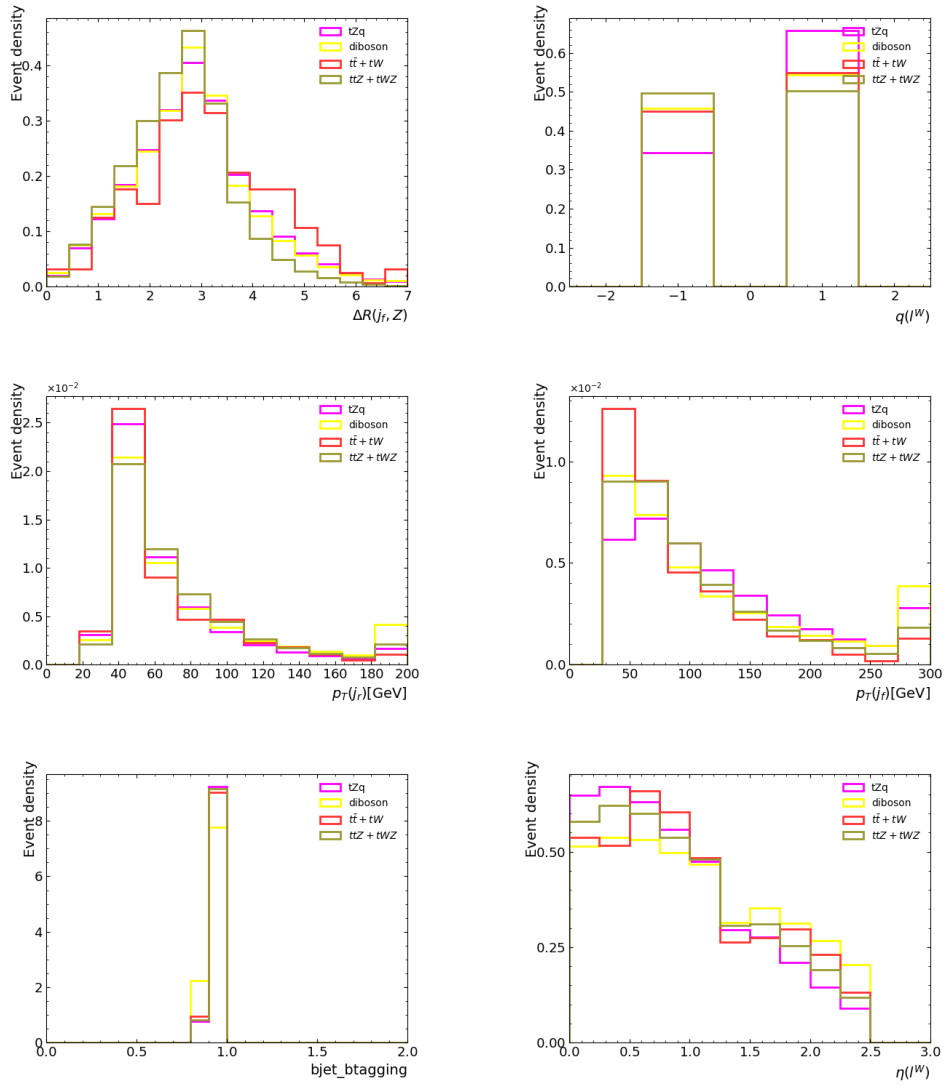


Figure 3.2: Kinematic distributions in order of their significance in the  $3j1b$  signal region, normalized area to unity

Variable	Corr. loss [%]	Definition
$m_{bj_f}$	43.0	Invariant mass of the $b$ -jet and the forward jet
$m_t$	31.8	Reconstructed mass of the top quark
$ \eta(j_f) $	17.4	Absolute pseudorapidity of the forward jet
$m_T(\ell, E_T^{\text{miss}})$	12.3	Transverse mass of $W$ boson
$ \eta(j_r) $	12.0	Absolute pseudorapidity of the radiation jet
$p_T(Z)$	10.2	Transverse momentum of the reconstructed $Z$ boson
$\Delta R(j_f, Z)$	8.1	$\Delta R$ between the forward jet and the $Z$ boson
$q(\ell^W)$	7.8	Electric charge of the lepton from the $W$ boson decay
$p_T(j_r)$	6.7	Transverse momentum of the radiation jet
$p_T(j_f)$	5.6	Transverse momentum of the forward jet
$b$ -tagging score	5.2	$b$ -tagging score of the $b$ -jet
$ \eta(\ell^W) $	4.2	Absolute pseudorapidity of the $W$ -boson lepton
$ \eta(Z) $	3.9	Absolute pseudorapidity of the reconstructed $Z$ boson
$p_T(\ell^W)$	3.5	Transverse momentum of the $W$ boson lepton
$p_T(W)$	2.5	Transverse momentum of the reconstructed $W$ boson

Table 3.1: Kinematic variables used for training in the NeuroBayes multivariate analysis ordered by their importance [29]. The  $b$ -tagging score denotes an estimation of how well the  $b$ -jet is identified

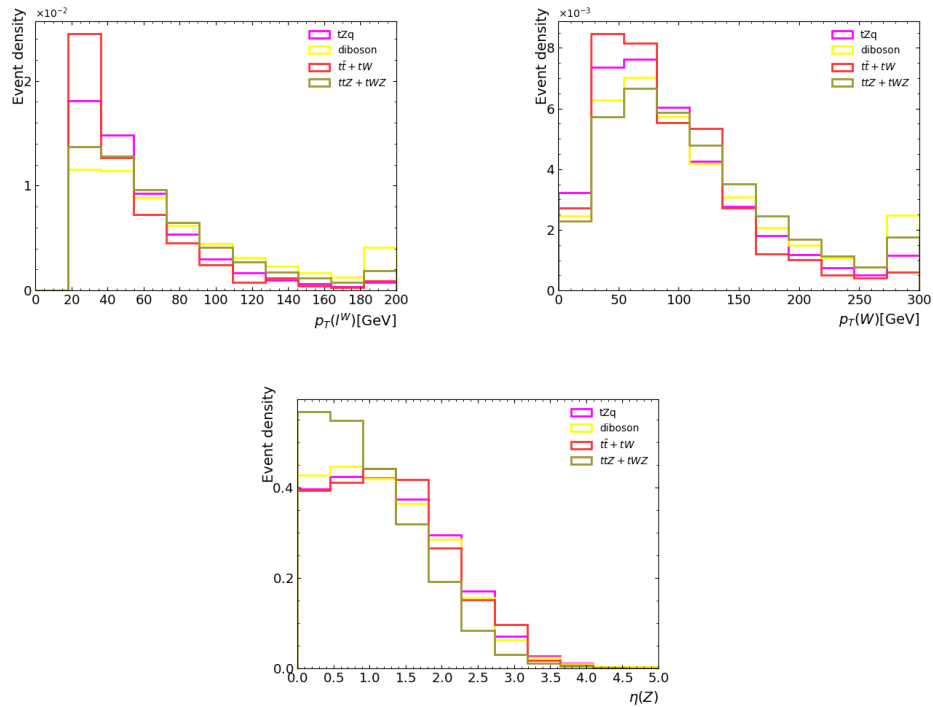


Figure 3.3: Kinematic distributions in order of their significance in the  $3j1b$  signal region, normalized area to unity

then restores the training weights of the best model, i.e. the lowest loss.

## 3.2 Optimization of Hyperparameters

In order to train properly, the hyperparameters need to be tuned carefully. Optimization is realized by evaluating the model with the methods explained in Section 2.4.2. The strategy for optimization is as follows. An initial set of parameters, shown in Table 3.2, is chosen beforehand and is then adjusted step by step. This ensures the best possible outcome.

	Parameter	Value
Learning rate	$\eta$	$1 \times 10^{-2}$
Number hidden layers	$n_{\text{layer}}$	3
Number nodes in hidden layers	$n_{\text{nodes}}$	30
Dropout	$d$	0.3
Validation fraction	$v$	0.1

Table 3.2: Initial set of parameters for training

### 3.2.1 Learning rate

The learning rate (LR) is responsible for the amount the training weights are updated with respect to the loss gradient. If the learning rate  $\eta$  tends to be too large, the weights do not converge to a global minimum of the loss but rather leap around it. On the other hand, choosing a small  $\eta$  can lead to the weights not converging in training. This can be visualized with a simple one dimensional example seen in Figure 3.4. Typically, the LR ranges from  $1 \times 10^{-2}$  to  $1 \times 10^{-5}$ . Figure 3.5 shows the impact

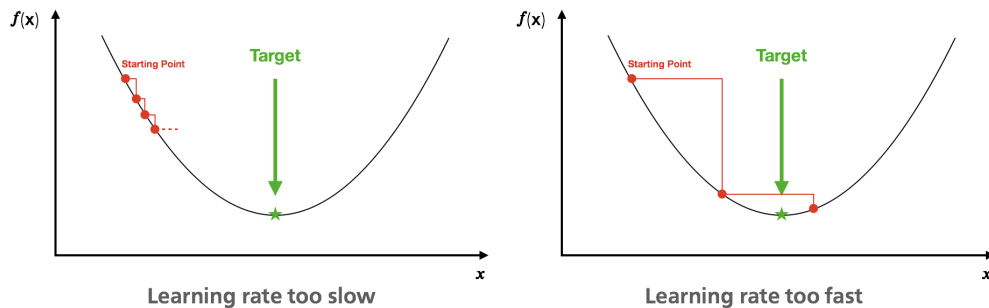


Figure 3.4: Visualization of the impact of the learning rate  $\eta$  in a simplified model. The abscissa represents the one dimensional training weight while the ordinate shows the loss [43]

the LR can have. While with  $\eta = 1 \times 10^{-2}$ , the NN cannot find a global minimum, choosing a learning rate of  $1 \times 10^{-5}$  improves the model significantly. The optimal  $\eta$  is then found by analyzing the AUC value. With  $\eta = 1 \times 10^{-4}$  the NN trains well and improves the AUC slightly. This is seen in Figure 3.6. An AUC value of 87.7% is achieved in Figure 3.6(a) in comparison to 86.9% with  $\eta = 1 \times 10^{-5}$ .

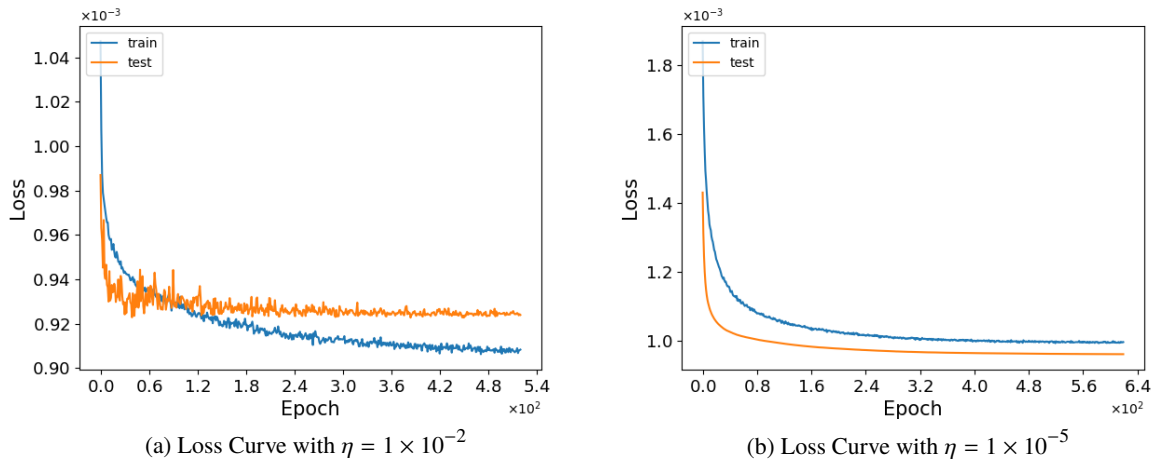


Figure 3.5: Impact of Learning Rate

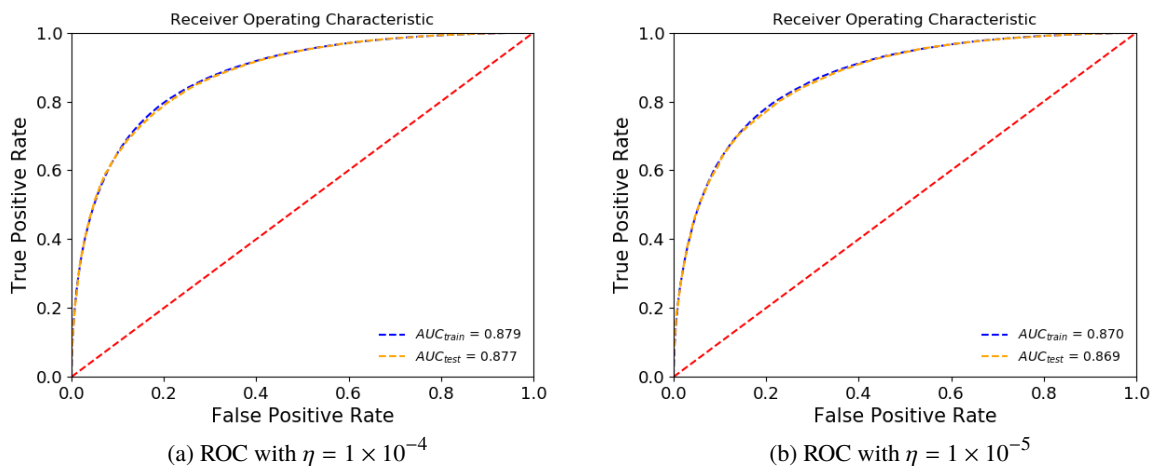


Figure 3.6: Comparison of AUC value with respect to  $\eta$

### 3.2.2 Structure of the neural network

The number of hidden layers and their nodes determines the complexity of the NN. On the one hand, the model may learn insufficiently if there are too few layers or nodes. On the other hand, choosing a very complex structure causes overtraining. A combination of well chosen structure parameters is necessary to realize the best possible outcome.

As seen in Figure 3.7(a), the highest AUC values are achieved with 2 or 13 hidden layers. 13 hidden layers increase the runtime of the program dramatically and as the trend towards higher numbers of hidden layers indicates worse performance, one cannot rule out that this sudden increase in AUC value is a fluctuation. For that reason, the number of hidden layers is chosen to be two.

With more nodes, the number of calculations in the network increases quadratically. The impact of the nodes can be seen in Figure 3.7(b). The rising difference in training AUC value and validation AUC value indicates overtraining. Figure 3.8 displays this overtraining with an extremely large number of 500 nodes. Choosing too few nodes leads to unsatisfactory results. The optimal number of nodes is therefore chosen if the validation AUC value reaches saturation without having an increasing difference in training and validation AUC values. This is achieved with 100 nodes per hidden layer.

### 3.2.3 Validation fraction

The Monte Carlo  $tZq$  signal sample is mixed with the background sample and afterwards randomly split. One fraction is then used for training while the other part serves for validation. If the training sample tends to be too small, the network does not learn sufficiently. On the other hand, an adequate amount of validation data is necessary to verify the training results. For this thesis, 70% and 30% of the total 332619 MC simulated events in the sample are used for training and validation, respectively.

### 3.2.4 Dropout

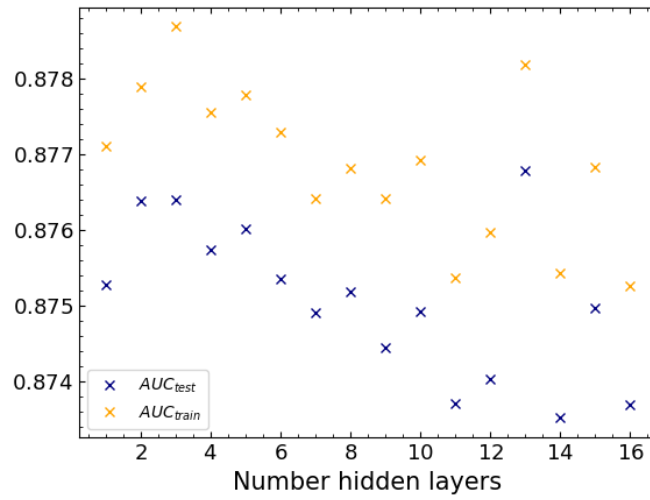
As already mentioned in subsection 2.4.2, the value of the dropout  $d$  determines the probability that a node is ignored in a layer. Finding the optimal dropout value is done in this subsection.

Dropout is only included in the hidden layers. If too few nodes are dropped the network is more likely to overtrain while too many ignored nodes causes the training to fail as the network cannot develop patterns to learn on. For instance, in Figure 3.9(a) a dropout of  $d = 0.01$  is chosen. This is not suitable as one can see significant overtraining, although Figure 3.9(b) indicates a good signal-background separation. In the extreme case of  $d = 0.99$ , only roughly one percent of the randomly chosen nodes are used for training. As a result, signal and background discrimination fails, as seen in Figure 3.10(b). Overtraining is prevented with  $d = 0.4$  while still distinguishing between signal and background. Figure 3.11 shows the evaluation plots.

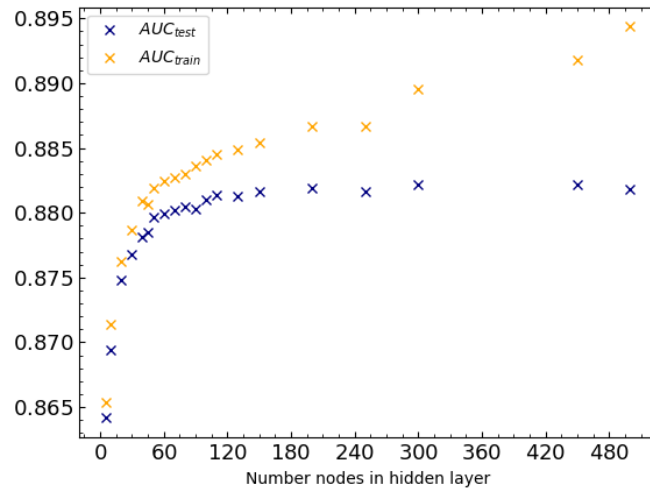
## 3.3 Results of the neural network

For optimal performance, the training parameters are chosen as displayed in Table 3.3.

After training and evaluating the model, the same MC sample is used for predicting signal and background. The total signal and background discrimination can be found in 3.11(b). In order to study the separation power of the backgrounds individually, the sample is split with respect to its



(a) The AUC values for different numbers of hidden layers



(b) The AUC values for different numbers of nodes in two hidden layers

Figure 3.7: Structure impact plots with respect to the AUC values

	Parameter	Value
Learning rate	$\eta$	$1 \times 10^{-4}$
Number hidden layers	$n_{\text{layer}}$	2
Number nodes in hidden layers	$n_{\text{nodes}}$	100
Dropout	$d$	0.4
Validation fraction	$\nu$	0.3

Table 3.3: Optimized parameters for training

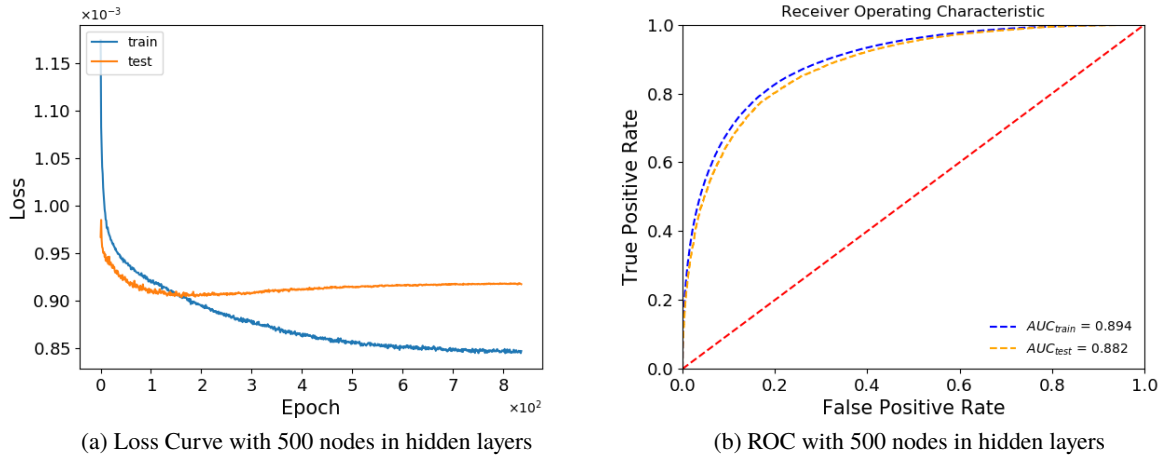


Figure 3.8: Extreme limit of 500 nodes in hidden layers leads to overtraining

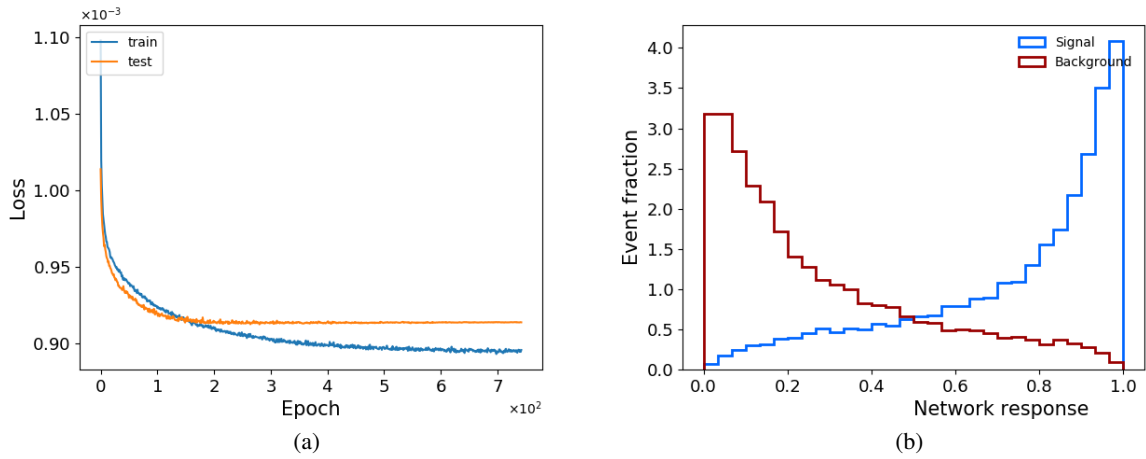


Figure 3.9: Evaluation of model with  $d = 0.01$ . The left plot displays the Loss curve. The right plot shows the signal-background discrimination. The histograms show the network output for signal and background normalized to area unity

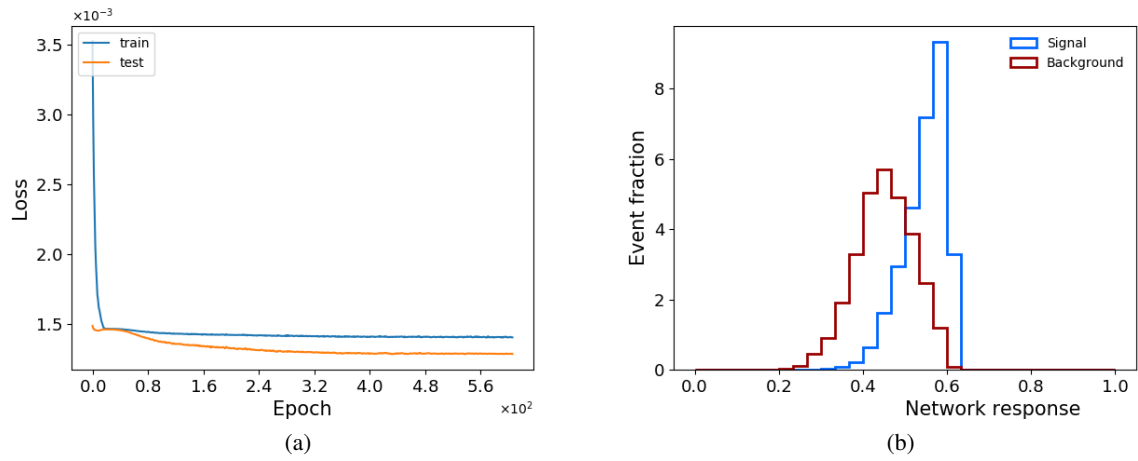


Figure 3.10: Evaluation of model with  $d = 0.99$

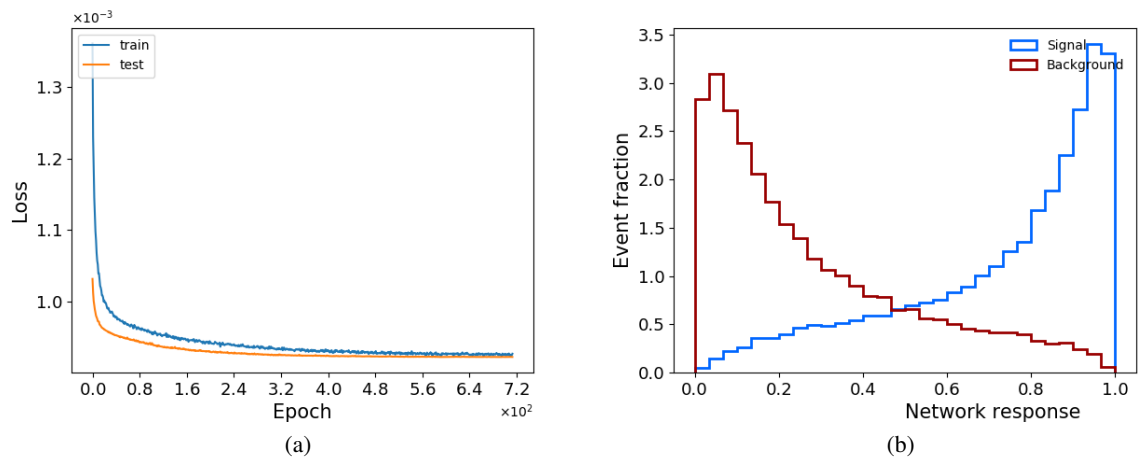


Figure 3.11: Evaluation of model with  $d = 0.4$



background and applied separately to the final model. Figure 3.12 shows the neural network output distributions normalized to unity.

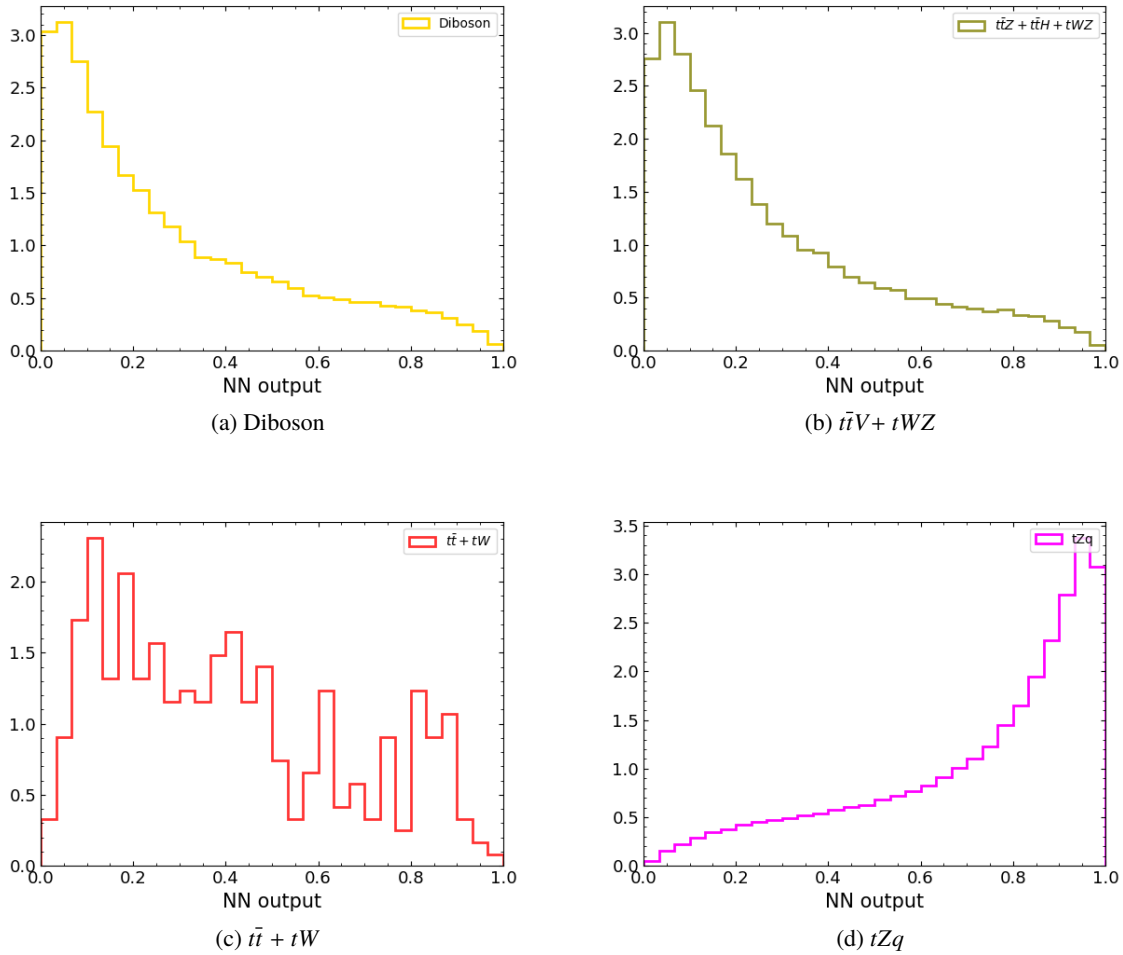


Figure 3.12: Keras NN output distributions

While the diboson and  $t\bar{t}Z$  samples peak at roughly zero (background), indicating a good separation, the  $t\bar{t}$  sample network outputs are distributed more uniformly and due to its low statistics, more fluctuating, but a tendency to the background implies that a discrimination has taken place. The signal output in Figure 3.12(d) also yields that the analysis was successful in signal identification.

Hereby, an accuracy and AUC value of 79.8% and 88.2% were achieved, respectively.



---

## Comparison with NeuroBayes

---

The main goal of the thesis is to implement the NN with Keras and to compare its performance to the trilepton  $tZq$  production multivariate analysis done before in Ref. [22]. It is based on the NeuroBayes package [12]. This chapter focuses on a brief introduction to the network structure of NeuroBayes and proceeds to study its performance in comparison to the results of Keras.

### 4.1 Network structure of NeuroBayes

The main difference to Keras is the predefined number of hidden layers, namely one. The number of nodes in this hidden layer is user-defined and set to 25.

Preprocessing is done similarly to this thesis. First, the sample is normalized to unity, transformed to a gaussian distribution and rescaled to compensate the large background sample. After ranking and choosing the variables for training, five separate networks are trained by using 80% and the remaining 20% of the sample for training and validation, respectively. Having five networks, the entire dataset can then be used for training and validation. This so called 5-fold cross validation technique is implemented to avoid overtraining [44]. Table 3.1 shows the most important variables found by the NeuroBayes analysis and also used in the Keras analysis.

### 4.2 Comparison of Results

#### Training quality

In Ref. [29], the training quality is checked by plotting the signal purity against the network output. Following Ref. [45], a well-trained network's signal purity should have a linear dependence on the output. The signal purity is defined as the ratio of signal events and total events in a bin. A comparison of purity plots is seen in Figure 4.1. This evaluation shows that both networks follow the linear distribution and that Keras provides somewhat better results by eye. If one wants to analyze the training quality in detail, a linear fit with uncertainty calculation is necessary.

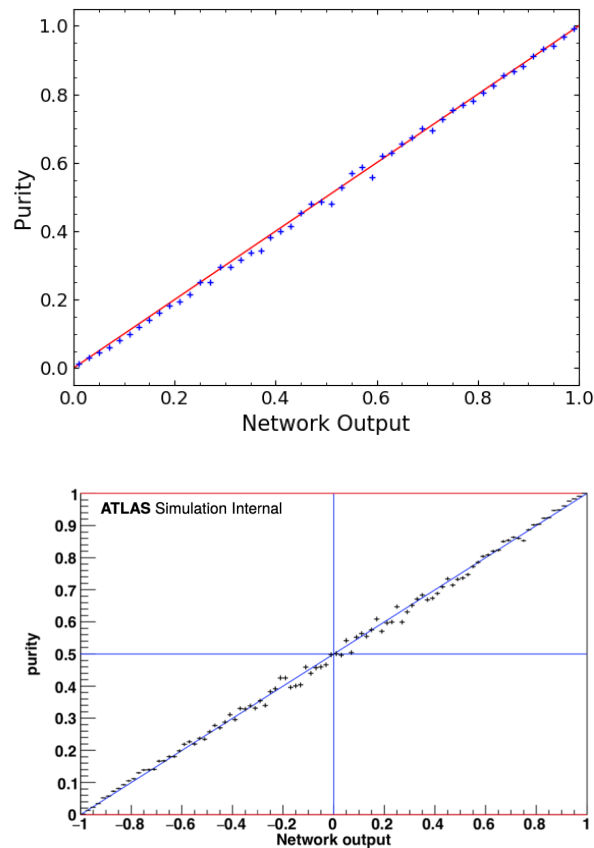


Figure 4.1: Signal purity as a function of the network output for Keras (top) and NeuroBayes (bottom) [29]

### Network Output Distributions

As mentioned in Section 3.3, the  $t\bar{t}$  discrimination has a flatter distribution compared to the other backgrounds. Comparing the Keras  $t\bar{t}$  predictions in Figure 3.12(c) with Figure 4.2(d), a similar behaviour can be seen in the NeuroBayes analysis.

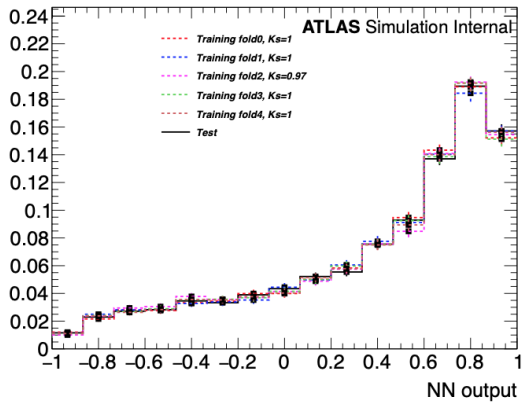
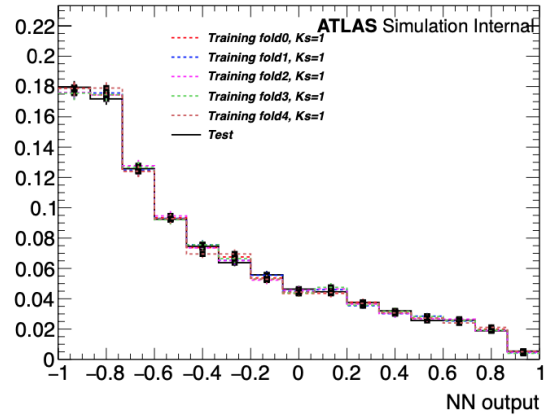
The other NeuroBayes network outputs in Figure 4.2 show a clear signal-background discrimination. Compared to Figure 3.12, which displays the network’s predictions of this thesis, the distributions look very alike.

### Receiver Operating Characteristic Curves

The ROC-curve and the corresponding AUC value gives a good estimation of the network’s performance. Ultimately, a higher AUC value represents a better model <sup>1</sup>.

In Ref. [29], the ROC-Curve was defined by plotting 1-TPR against 1-FPR. Although this an unusual definition, it gives the same insights into the model as it still scans FPR-TPR pairs for different signal

<sup>1</sup> With NeuroBayes, the so called Gini Index is calculated. Converting to the AUC value yields  $AUC = Gini + 0.5$

(a)  $tZq$ 

(b) Diboson

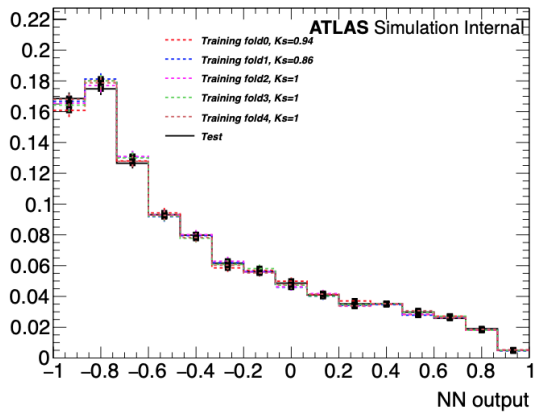
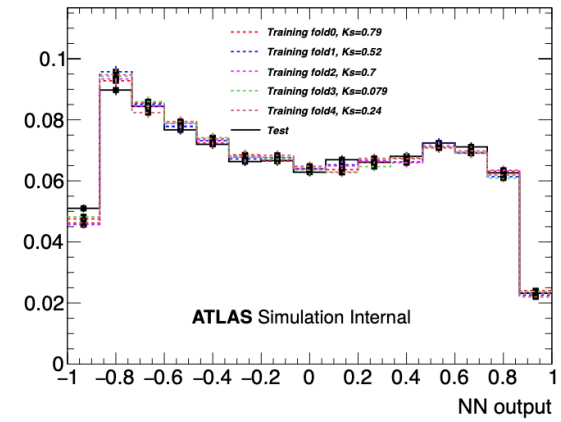
(c)  $t\bar{\nu} + t\bar{t}H + tWZ$ (d)  $t\bar{t} + tH$ 

Figure 4.2: NN outputs of the NeuroBayes  $tZq$  discrimination. The network outputs are distributed between  $-1$  and  $1$  for background and signal, respectively. The histograms are also normalized to unity [29]

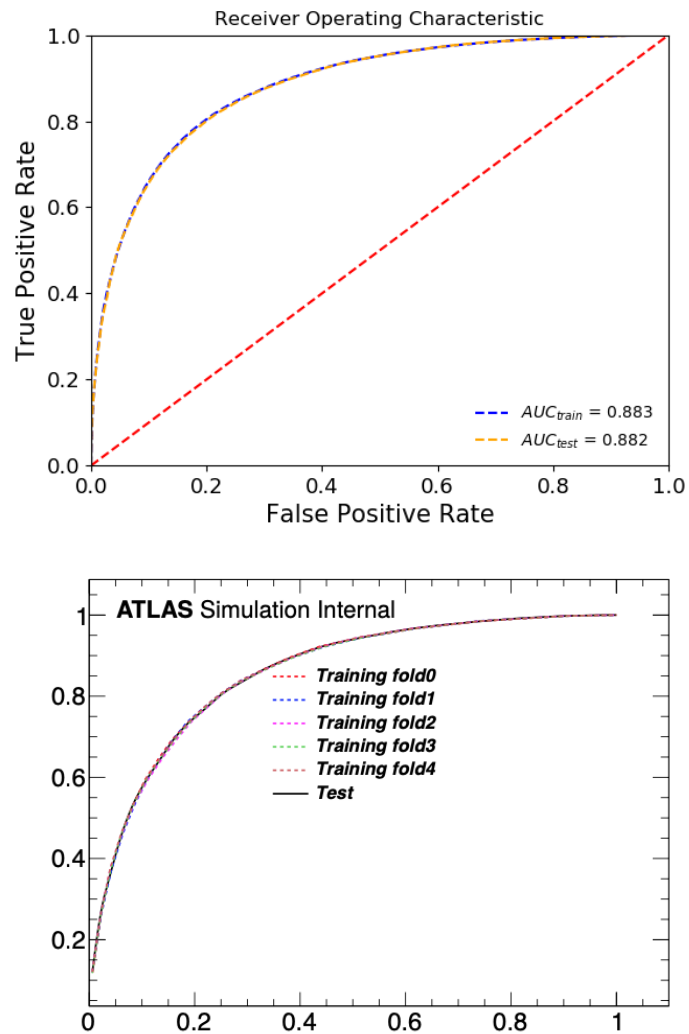


Figure 4.3: ROC-Curves of the analysis done in this thesis (top) and NeuroBayes (bottom) [29]. On the bottom plot, the x-axis is  $1 - \text{TPR}$  and the ordinate is given by  $1 - \text{FPR}$

thresholds. Furthermore, this definition does not change the AUC value.

In this thesis, the Keras model achieved an AUC value of 88.2% being slightly higher than the AUC value in the NeuroBayes analysis, yielding 87.5%. Figure 4.3 illustrates the ROC-curves of both analyses.





---

## Conclusion and Outlook

---

Even though the associated  $tZq$  production has already been studied intensively, there is still room for improvement. This thesis presents the multivariate analysis of a leptonically decaying associated  $tZq$  production using the Keras package.

After giving a general introduction and the theoretical background in Chapters 1 and 2, respectively, the multivariate analysis of MC simulated data generated for the ATLAS detector (corresponding to 2015 to 2018 ATLAS data from  $pp$  collisions at  $\sqrt{s} = 13$  TeV) is done in Chapter 3. At first, the kinematic variables were preprocessed to then optimize the hyperparameters of the NN. For this, the MC data was split into and training and validation part. The aim was to get consistent results for the loss curves for both fractions of the data and to maximize the AUC values of the calculated ROC-Curve. This allowed determination of the best values for the learning rate, number of hidden layers, nodes within the hidden layers, the validation fraction and the dropout.

A previous analysis has been performed with the non open-source package NeuroBayes. Therefore, in Chapter 4, the performance of the Keras results for the optimal hyperparameters and of NeuroBayes is compared.

While NeuroBayes already shows remarkable results in signal-background discrimination, they have even been further improved by the open-source package Keras. Both, training quality and predictions of the Keras network, can compete with NeuroBayes. In fact, they are quite indistinguishable by eye.

In order to evaluate the outcome in detail, the ROC-Curve and corresponding AUC value were analyzed. NeuroBayes reached an Area under Curve value of 87.5%. The Keras model achieved an AUC value of 88.2% having a slightly better performance than the previous model in  $tZq$  analysis. Furthermore, being an open-source Python module, Keras grants the user more freedom in terms of programming.

This leads to the conclusion that Keras is very suitable as a future replacement for NeuroBayes in not only  $tZq$  analysis, but also for particle physics in general.

Before Keras can become a standard tool, further improvements should be considered which could not be done in this thesis due to time constraints. First, the kinematic variables used for the network training are still chosen by the preprocessing of NeuroBayes. To be completely independent, an implementation of a Python based preprocessing is necessary. Here, the variable ranking can be done similarly to NeuroBayes, as shown in Appendix A. For future analyses, this ranking should be taken into account for network training, so that a comparison of the ranking performance can be used to identify the optimal set of kinematic variables.

After these improvements, it will be interesting to use actual ATLAS data rather than MC simulations. Based on the performance shown for the MC data, an extraction of cross-sections for associated  $tZq$  is possible [18]. The analysis strategy could then lay the foundation of other rare top quark related processes.

## Bibliography

---

- [1] S. Abachi et al., *Observation of the Top Quark*, *Physical Review Letters* **74** (1995) 2632, ISSN: 1079-7114, URL: <http://dx.doi.org/10.1103/PhysRevLett.74.2632> (cit. on pp. 1, 3).
- [2] G. A. et al, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012) 1, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/S037026931200857X> (cit. on pp. 1, 3).
- [3] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008) S08001, URL: <https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001> (cit. on p. 1).
- [4] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08003, URL: <https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003> (cit. on pp. 1, 5, 6).
- [5] V. Gribov, *The theory of quark confinement*, *The European Physical Journal C* **10** (1999) 91, ISSN: 1434-6052, URL: <http://dx.doi.org/10.1007/s100529900052> (cit. on p. 1).
- [6] A. Ali and G. Kramer, *JETS and QCD: a historical review of the discovery of the quark and gluon jets and its impact on QCD*, *The European Physical Journal H* **36** (2011) 245, ISSN: 2102-6467, URL: <http://dx.doi.org/10.1140/epjh/e2011-10047-1> (cit. on p. 1).
- [7] R. Kogler et al., *Jet substructure at the Large Hadron Collider*, *Rev. Mod. Phys.* **91** (4 2019) 045003, URL: <https://link.aps.org/doi/10.1103/RevModPhys.91.045003> (cit. on pp. 1, 10).
- [8] T. Golling, *LHC searches for physics beyond the Standard Model with top quarks*, *Journal of Physics: Conference Series* **452** (2013) 012010, ISSN: 1742-6596, URL: <http://dx.doi.org/10.1088/1742-6596/452/1/012010> (cit. on p. 2).
- [9] R. KEHOE, M. NARAIN, and A. KUMAR, *REVIEW OF TOP QUARK PHYSICS RESULTS*, *International Journal of Modern Physics A* **23** (2008) 353, ISSN: 1793-656X, URL: <http://dx.doi.org/10.1142/S0217751X08039293> (cit. on p. 2).
- [10] ATLAS collaboration, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, *Physics Letters B* **780** (2018) 557, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/S0370269318302120> (cit. on pp. 2, 10).

- [11] R. H. R. Hahnloser et al., *Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit*, *Nature* **405** (2000) 947, URL: <https://doi.org/10.1038/35016072> (cit. on pp. 2, 13).
- [12] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **559** (2006) 190, Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900205022679> (cit. on pp. 2, 31).
- [13] F. Chollet et al., *Keras*, <https://github.com/fchollet/keras>, 2015 (cit. on pp. 2, 19).
- [14] K. Kodama et al., *Observation of tau neutrino interactions*, *Physics Letters B* **504** (2001) 218, ISSN: 0370-2693, URL: [http://dx.doi.org/10.1016/S0370-2693\(01\)00307-0](http://dx.doi.org/10.1016/S0370-2693(01)00307-0) (cit. on p. 3).
- [15] M. e. a. Tanabashi, *Review of Particle Physics*, *Phys. Rev. D* **98** (3 2018) 030001, URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001> (cit. on p. 3).
- [16] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on p. 4).
- [17] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 4).
- [18] I. A. Cioara, *Associated Production of a Top Quark and a Z Boson in pp Collisions at  $\sqrt{s} = 13$  TeV Using the ATLAS Detector*, BONN-IR-2018-07, PhD Thesis: University of Bonn, 2018, URL: [http://hss.ulb.uni-bonn.de/diss\\_online](http://hss.ulb.uni-bonn.de/diss_online) (cit. on pp. 4, 10, 11, 38).
- [19] S. Profumo, L. Giani, and O. F. Piattella, *An Introduction to Particle Dark Matter*, 2019, arXiv: [1910.05610](https://arxiv.org/abs/1910.05610) [hep-ph] (cit. on p. 4).
- [20] G. C. Nayak, *Matter-Antimatter Asymmetry Of The Universe and Baryon Formation From Non-Equilibrium Quarks and Gluons*, 2019, arXiv: [1909.05640](https://arxiv.org/abs/1909.05640) [physics.gen-ph] (cit. on p. 4).
- [21] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, (2016), General Photo, URL: <https://cds.cern.ch/record/2197559> (cit. on p. 5).
- [22] *Observation of the associated production of a top quark and a Z boson in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, tech. rep. ATLAS-CONF-2019-043, CERN, 2019, URL: <https://cds.cern.ch/record/2690716> (cit. on pp. 5, 10, 19, 31).
- [23] N. Ilić, “Observation of the Higgs Boson decaying to  $WW^* \rightarrow l\nu\nu l$ ,” 2015 (cit. on p. 6).
- [24] J. Pequeno, “Computer generated image of the ATLAS inner detector,” 2008, URL: <https://cds.cern.ch/record/1095926> (cit. on p. 7).
- [25] J. Pequeno, “Computer Generated image of the ATLAS calorimeter,” 2008, URL: <https://cds.cern.ch/record/1095927> (cit. on p. 8).

- 
- [26] J. Pequeno, “Computer generated image of the ATLAS Muons subsystem,” 2008,  
URL: <https://cds.cern.ch/record/1095929> (cit. on p. 9).
- [27] M. Aaboud et al., *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13$  TeV*,  
*The European Physical Journal C* **78** (2018), ISSN: 1434-6052,  
URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9> (cit. on p. 9).
- [28] P. Sinervo, *Highlights of top-quark production measurements at ATLAS*, (2019),  
URL: <https://cds.cern.ch/record/2686246> (cit. on p. 10).
- [29] M. Alhroob et al.,  
*Observation of the associated production of a top quark and a Z boson at 13 TeV with ATLAS*,  
tech. rep. ATL-COM-PHYS-2018-140, CERN, 2018,  
URL: <https://cds.cern.ch/record/2304824> (cit. on pp. 10, 11, 19, 22, 31–34).
- [30] O. B. Bylund, *Modelling  $Wt$  and  $tWZ$  production at NLO for ATLAS analyses*, 2016,  
arXiv: [1612.00440](https://arxiv.org/abs/1612.00440) [hep-ph] (cit. on p. 11).
- [31] C. Boever, *Differential cross-section measurement of the  $tZq$  process with the ATLAS detector*,  
MA thesis: Universität Bonn, 2019 (cit. on p. 12).
- [32] C. Nwankpa et al.,  
*Activation Functions: Comparison of trends in Practice and Research for Deep Learning*, 2018,  
arXiv: [1811.03378](https://arxiv.org/abs/1811.03378) [cs.LG] (cit. on p. 11).
- [33] J. Han and C. Moraga,  
“The influence of the sigmoid function parameters on the speed of backpropagation learning,”  
*From Natural to Artificial Neural Computation*, ed. by J. Mira and F. Sandoval,  
Springer Berlin Heidelberg, 1995 195, ISBN: 978-3-540-49288-7 (cit. on p. 13).
- [34] Y. S, *Build up a Neural Network with python*, 2019,  
URL: <https://towardsdatascience.com/build-up-a-neural-network-with-python-7faea4561b31> (cit. on p. 13).
- [35] M. A. Nielsen, *Neural Networks and Deep Learning*, misc, 2018,  
URL: <http://neuralnetworksanddeeplearning.com/> (cit. on p. 13).
- [36] T. Zhang, “Solving Large Scale Linear Prediction Problems Using Stochastic Gradient Descent Algorithms,”  
*Proceedings of the Twenty-First International Conference on Machine Learning*,  
ICML ’04, Association for Computing Machinery, 2004 116, ISBN: 1581138385,  
URL: <https://doi.org/10.1145/1015330.1015332> (cit. on p. 13).
- [37] Y. L. Cun, *A Theoretical Framework for Back-Propagation*, 1988 (cit. on p. 13).
- [38] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014,  
arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG] (cit. on pp. 13, 14).
- [39] J. Sjöberg and L. Ljung,  
*Overtraining, Regularization, and Searching for Minimum in Neural Networks*,  
*IFAC Proceedings Volumes* **25** (1992) 73, 4th IFAC Symposium on Adaptive Systems in  
Control and Signal Processing 1992, Grenoble, France, 1-3 July, ISSN: 1474-6670,  
URL: <http://www.sciencedirect.com/science/article/pii/S1474667017507156>  
(cit. on p. 14).

- [40] N. Srivastava et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, *Journal of Machine Learning Research* **15** (2014) 1929, URL: <http://jmlr.org/papers/v15/srivastava14a.html> (cit. on p. 15).
- [41] A. P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms*, *Pattern Recognition* **30** (1997) 1145, ISSN: 0031-3203, URL: <http://www.sciencedirect.com/science/article/pii/S0031320396001422> (cit. on p. 16).
- [42] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825 (cit. on p. 19).
- [43] *Build a Neural Network*, 2019, URL: <https://www.wandb.com/tutorial/build-a-neural-network> (cit. on p. 23).
- [44] J. Brownlee, *A Gentle Introduction to k-fold Cross-Validation*, 2019, URL: <https://machinelearningmastery.com/k-fold-cross-validation/> (cit. on p. 31).
- [45] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*, 2004, arXiv: [physics/0402093](https://arxiv.org/abs/physics/0402093) [[physics.data-an](https://arxiv.org/abs/physics/0402093)] (cit. on p. 31).
- [46] S. van der Walt, S. C. Colbert, and G. Varoquaux, *The NumPy Array: A Structure for Efficient Numerical Computation*, *Computing in Science Engineering* **13** (2011) 22, ISSN: 1558-366X (cit. on p. 43).

---

## Variable Ranking with Keras

---

As already explained in Section 3.1, the variable ranking is done by omitting the variables step by step and calculating the deviation in correlation loss caused by its removal.

The implementation in Keras can be done similarly. Here, one by one, the variables are transformed to a flat distribution using the Numpy `random uniform` module [46]. The change of the accuracy metric is then calculated and the variables are sorted by their importance, i.e. the largest deviation of the accuracy compared to all variables.

This is not sufficient. Having about 95 kinematic variables in total, they might be strongly correlated among themselves. For a meaningful ranking, the strongly correlated variables have to be removed first, otherwise the ranking might decide to keep correlated variables which will lead to poor performances in training.

Unfortunately, the necessary decorrelation could not be performed due to time constraints. The general idea is to compute the correlation matrix using Panda dataframes. A new matrix can then be created with Cholesky decomposition. Variables above a certain correlation threshold could be dropped to repeat the variable ranking.





### Keras network code

---

For the interested reader, the code used for the NN can be found and downloaded from: [https://github.com/pietnogga/tZq\\_discrimination](https://github.com/pietnogga/tZq_discrimination)

In the repository, there is also the code to sort the input MC samples into directories with respect to the background and signal.

Both codes run on Python 3.7 and some additional packages have to be downloaded for the code to run. If you want to run the code on University Bonn computers, run

```
module load anaconda/2019.10-py37
```

to load Python 3.7. Then, activate the Anaconda environment to ensure all used packages are available:

```
source activate /cephfs/user/s6pinogg/PietBachelor/my_pythonized_analysis/
```

A list of the Python packages downloaded to the environment can also be found on the git repository.



# List of Figures

---

2.1	Overview over the current Standard Model of modern particle physics . . . . .	4
2.2	Overview of the accelerators of the Large Hadron Collider . . . . .	5
2.3	Overview of the structure of the ATLAS detector . . . . .	6
2.4	The ATLAS coordinate system . . . . .	6
2.5	Schematic view of the subcomponents of the Inner Detector . . . . .	7
2.6	Overview of the ATLAS calorimeter system . . . . .	8
2.7	Structure of the Muon Spectrometer . . . . .	9
2.8	Example Feynman graphs for the associated production of a top quark and a Z boson	10
2.9	Feynman graphs of the background sources . . . . .	12
2.10	The Structure of a simple neural network . . . . .	13
2.11	Loss curves of different models . . . . .	15
2.12	Further evaluation methods of neural networks . . . . .	17
3.1	Kinematic distributions in order of their significance in the $3j1b$ signal region, normalized area to unity . . . . .	20
3.2	Kinematic distributions in order of their significance in the $3j1b$ signal region, normalized area to unity . . . . .	21
3.3	Kinematic distributions in order of their significance in the $3j1b$ signal region, normalized area to unity . . . . .	22
3.4	Visualization of the impact of the learning rate $\eta$ in a simplified model . . . . .	23
3.5	Impact of Learning Rate . . . . .	24
3.6	Comparison of AUC value with respect to $\eta$ . . . . .	24
3.7	Structure impact plots with respect to the AUC values . . . . .	26
3.8	Extreme limit of 500 nodes in hidden layers leads to overtraining . . . . .	27
3.9	Evaluation of model with $d = 0.01$ . . . . .	27
3.10	Evaluation of model with $d = 0.99$ . . . . .	28
3.11	Evaluation of model with $d = 0.4$ . . . . .	28
3.12	Keras NN output distributions . . . . .	29
4.1	Signal purity as a function of the network output . . . . .	32
4.2	NN outputs of the NeuroBayes $tZq$ discrimination . . . . .	33
4.3	ROC-Curves of the analysis done in this thesis and NeuroBayes . . . . .	34



## List of Tables

---

3.1	Kinematic variables used for training in the NeuroBayes multivariate analysis ordered by their importance . . . . .	22
3.2	Initial set of parameters for training . . . . .	23
3.3	Optimized parameters for training . . . . .	26



## Acknowledgements

---

I would like to take this opportunity to thank several people who helped me with this thesis in so many ways.

Firstly, to Prof. Ian C. Brock, who gave me the chance to work in his lovely team. Your continuous advice, discussions and of course your thesis guide really helped me a lot during these four very informative months. Then, to Dr. Philip Bechtle, for kindly taking the time for reading this thesis as second reviser.

I would also like to thank all members of the Brock research Group. Special thanks to Christian, who agreed to supervise me during this thesis and always had an open ear for my problems and questions. Thanks Tanja, who always took the time to discuss the physics topics I had not understood yet. I am very grateful for all the explanations you provided.

Of course, I'd like to thank Elli for all the coffee breaks we had during this time. The resulting off topic talks always helped me when I had so many things to work on. I would also like to acknowledge the rest of the group. Thank you Maxx, who always put a smile on my face, Anji, Chris, Nilima, Oleh, Rico, Nicolas and Richard for the support you all gave me!

Prof. Brock and Chris Boever also saved much of my time by kindly providing me with the Feynman Graphs I used in this thesis. Many thanks for that!

Luckily, I had Christian, my father and my brother who all agreed to proof read my thesis. Your time you all took to advise me concerning the topic, structure and english of my thesis is dearly appreciated. Without you, the thesis would have had more logic, grammar and comma errors.

All in all, I can say that the support of all mentioned people really meant a lot. These fun four months wouldn't have been the same without you. I am looking forward to working with all of you in the future.