Study of fake tau lepton rates in Higgs and Z Boson decays using advanced machine learning techniques

Chris M. Staude

Masterarbeit in Physik
angefertigt im Physikalischen Institut

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

June 2021

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, 28.06.2021 .....
Date

.....
Signature

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachterin: Prof. Dr. Jochen Dingfelder

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. Dr. Ian C. Brock, not only for giving me the opportunity to write a Master's thesis in his group, but also for his patience and support throughout.

At the beginning of the Master's thesis, the circumstances were very unusual. The spread of the coronavirus led to a nationwide lockdown. While one usually sits in the office and works on solutions together with the other group members, from this point on everything took place online. This was a great challenge for everyone and we tried to master it as best we could. Thanks to everyone for the great cooperation during this time. From time to time, online game evenings took place in the group and there we had the chance to get to know the group members better in private. At this point I would like to thank the organiser Christian Kirfel. A special thanks goes to Tanja Holm, who always knew the answer to every question I had and was able to give me an answer even at untypical times. I would also like to thank Nilima Akolkar, as she constantly motivated and built me up even when she was bombarded by my, sometimes very primitive, comprehension questions about particle physics. Further thanks go to Federico Guillermo Diaz Capriles, who guided me through the thesis and helped me with many comprehension problems. I would also like to thank all the other group members for their support during this time.

Last but not least, I would like to thank my family for their emotional as well as financial support during the whole time. Throughout my studies, I met many great people who made it a time with great memories. At this point, a special thank you to Esra Aslim, Nurida Boddenberg and Sonja Lüdiger. The time with you was truly unforgettable. Also special thanks to Charlotte and Dorothee Perry for the great conversations and cool game evenings. Last but not least, a big thank you to Tiago Suhet Salgado and Sattajit Chandra Bormon, who were able to teach me to laugh again in situations of despair.

Contents

1 Introduction **1**

2 Theoretical Concepts **3**
2.1 The Standard Model of Particle Physics . . . 3
2.2 Feynman Diagrams . . . 7
2.3 Concepts of Accelerator Physics Experiments . . . 8
2.3.1 Basic Concepts in Particle Physics Experiments . . . 9
2.4 Top Quark Physics at Hadron Colliders . . . 12
2.4.1 Properties of the Top Quark . . . 12
2.4.2 Top Quark Production Processes . . . 13
2.4.3 Rare Top Quark Production Processes . . . 16
2.4.4 Single top-quark production associated with a Z boson . . . 17
2.4.5 Single top-quark production associated with a Higgs boson . . . 18
2.4.6 Top Quark Decay . . . 19
2.5 Tau Leptons . . . 19

3 The ATLAS Experiment at the Large Hadron Collider **21**
3.1 The Large Hadron Collider (LHC) . . . 21
3.2 The ATLAS Detector . . . 23
3.2.1 The Inner Detector . . . 24
3.2.2 Calorimetry . . . 27
3.2.3 Muon Spectrometer . . . 28
3.2.4 Magnet System . . . 28
3.2.5 Trigger System . . . 28
3.3 Reconstruction of Physics Objects at ATLAS . . . 29
3.3.1 Electrons and Photons . . . 29
3.3.2 Muons . . . 30
3.3.3 Jets . . . 31
3.3.4 b-jets . . . 31
3.3.5 Hadronically decaying τ leptons . . . 32
3.3.6 Missing Transverse Energy . . . 37

4 Event selection **39**
4.1 tHq channel . . . 39
4.2 $1e/\mu + 1\tau_{\text{had}}$ event selection . . . 40

4.3 Backgrounds . . . 40
4.3.1 Backgrounds containing non-prompt or fake leptons . . . 40
4.3.2 Backgrounds containing prompt leptons . . . 41
4.4 Datasets . . . 42
4.4.1 Data Samples . . . 42
4.4.2 Monte Carlo Simulation Samples . . . 42

5 Neural Networks **45**
5.1 Core Concepts in Machine Learning . . . 45
5.2 Deep Learning . . . 46
5.2.1 The Perceptron Model . . . 47
5.2.2 Multi-Layer Perceptron . . . 48
5.2.3 Activation Functions . . . 48
5.2.4 The Learning Process . . . 51
5.2.5 Dropout . . . 54
5.2.6 The Need for Weak Supervision . . . 54
5.3 Technical Details . . . 57

6 Baseline Classifier Approach for $1e/\mu + 1\tau_{\text{had}}$ **59**
6.1 Selection of Training Variables . . . 59
6.2 Network Architecture and Hyperparameter Optimization . . . 60
6.3 Classification Results . . . 62
6.4 Prediction on unseen MC Samples . . . 64

7 Mixed Sample Classifier (CWoLa) for $1e/\mu + 1\tau_{\text{had}}$ **67**
7.1 CWoLa trained on MC Samples . . . 67
7.1.1 Mixed Samples for Training . . . 67
7.1.2 Selection of Training Variables and CWoLa Network Architecture . . . 70
7.1.3 Classification Results . . . 70
7.1.4 Prediction on unseen MC Samples . . . 73
7.2 CWoLa trained on Data Samples . . . 76
7.2.1 Mixed Samples for Training . . . 76
7.2.2 Selection of Training Variables and CWoLa Network Architecture . . . 76
7.2.3 Classification Results . . . 78
7.2.4 Prediction on unseen MC Samples . . . 79

8 Summary and Conclusion **81**

Bibliography **83**

A Appendix **93**
A.1 List of Data and Monte Carlo (MC) samples . . . 93
A.1.1 Data Samples . . . 93
A.1.2 Monte Carlo Samples . . . 93
A.2 MC Composition Plots . . . 106

A.3 Baseline Classifier training results for 1 and 3-prong tau decays. 106
A.3.1 1-prong 106
A.3.2 3-prong 111
A.4 Prediction on different regions for CWoLa trained on MC 112

List of Figures **117**

List of Tables **121**

CHAPTER 1

Introduction

There is evidence in human history that people have always been driven by curiosity to search for profound questions about nature and the very question of their existence. Observations have allowed humans to draw conclusions and gradually sharpen their understanding of the whole. We recognize that everything that surrounds us is made of matter. It is hard to believe that only hundred years ago the structure of the atom was still completely unknown and it took a few more years to understand that atoms are not fundamental components in nature, but they can be further broken down into sub-components. The search for ever deeper structures has created the need to capture these discoveries in a unified model. This model is called the Standard Model of Particle Physics and it includes all the elementary particles known so far as well as the interactions between them. It is also considered the crown jewel of particle physics, as it was one of the greatest achievements of the past century.

Large accelerator facilities are used to explore the properties of these elementary particles and to test the validity of the Standard Model. An incredible amount of collision data is analyzed in order to extract the actual processes of interest. By increasing the energy of these facilities, rare processes can be studied. Unfortunately, the reconstruction of these is difficult, as background processes have similar signatures.

QCD events represent a large background in this work and can sometimes appear to be tau lepton decays. In this work, rare top quark production processes where a single top quark is produced in association with a Z boson as well as the associated production with a Higgs boson are investigated. Neural Networks are used to discriminate real from fake tau leptons using fully labelled simulated events. However, the simulations are often not perfect or have sufficient statistics. Therefore a new method called CWoLa (Classification Without Labels) is also used to differentiate between real and fake tau leptons using only mixtures of events instead of true labels. The classifier is trained on the enriched with fakes $t\bar{t}$ data and compared to the optimal classifier in the fully-supervised case.

This thesis is structured as follows: In chapter 2, an overview of some theoretical concepts of particle physics is given in order to understand the underlying analysis. Chapter 3 describes the individual parts of the ATLAS detector as well as the reconstruction of physics objects. The event selection is described in chapter 4, followed by an introduction to neural networks in chapter 5. The actual analysis of this thesis is introduced in chapter 6 with the description of the baseline classifier. Subsequently, a weak supervision approach is presented in chapter 7. At the end, a summary with a short conclusion is given in chapter 8.

CHAPTER 2

Theoretical Concepts

"When a stranger, hearing that I am a physicist, asks me in what area of physics I work, I generally reply that I work on the theory of elementary particles. Giving this answer always makes me nervous. Suppose that the stranger should ask, 'What is an elementary particle?' I would have to admit that no one really knows."

- Steven Weinberg (1997),

The first chapter covers the theoretical foundations necessary to understand the analysis of this thesis. At the beginning, a short introduction to the framework of every particle physicist is given, the Standard Model of particle physics. There, the elementary building blocks of matter and their interactions are discussed in detail. Afterwards, the properties and production mechanisms of the top-quark are described. The chapter concludes with a deeper look into tau leptons. In addition to that hadronic tau decays and the key particles of this analysis, the misidentified tau leptons, are discussed.

2.1 The Standard Model of Particle Physics

At the end of the 19th century, particle physics was initiated with the discovery of the first fundamental particle, the electron, by J.J. Thomson. In the following decades, the findings in nuclear and cosmic-ray physics have enabled the study of elementary particles, whose probably most decisive achievement has been the development of the Standard Model of Particle Physics beginning from the last half of the 20th century. According to our present understanding, all elementary particles in the Standard Model are assumed to have no inner substructure and cannot be decomposed into smaller constituents.

The Standard Model is an elegant, though ingenuous far from a simple theory, that is based on gauge field theories that unify three of the four fundamental interactions, starting with quantum electrodynamics (QED), which describes the interaction of electrons with light. Electroweak theory took this further, combining the weak and electromagnetic interaction described by a $SU(2)_L \times U(1)_Y$ gauge symmetry. Finally, the strong nuclear force is the interaction between quarks and is described by quantum chromodynamics (QCD) forming the final non-abelian gauge structure of the Standard

Model as $SU(3)_C \times SU(2)_L \times U(1)_Y$ [1]. Gravity is not included here as the fourth fundamental force, since it is weak compared to the other fundamental forces and is therefore neglected.

Together, the fundamental interactions with the elementary particles form an model which is illustrated in figure 2.1. It seems perfect, but the Standard Model is still an incomplete theory and still does not explain topics like dark energy or the matter-antimatter asymmetry. However, advances in new detector technologies and accelerator structures have also contributed to the existence of a veritable zoo of elementary particles today.

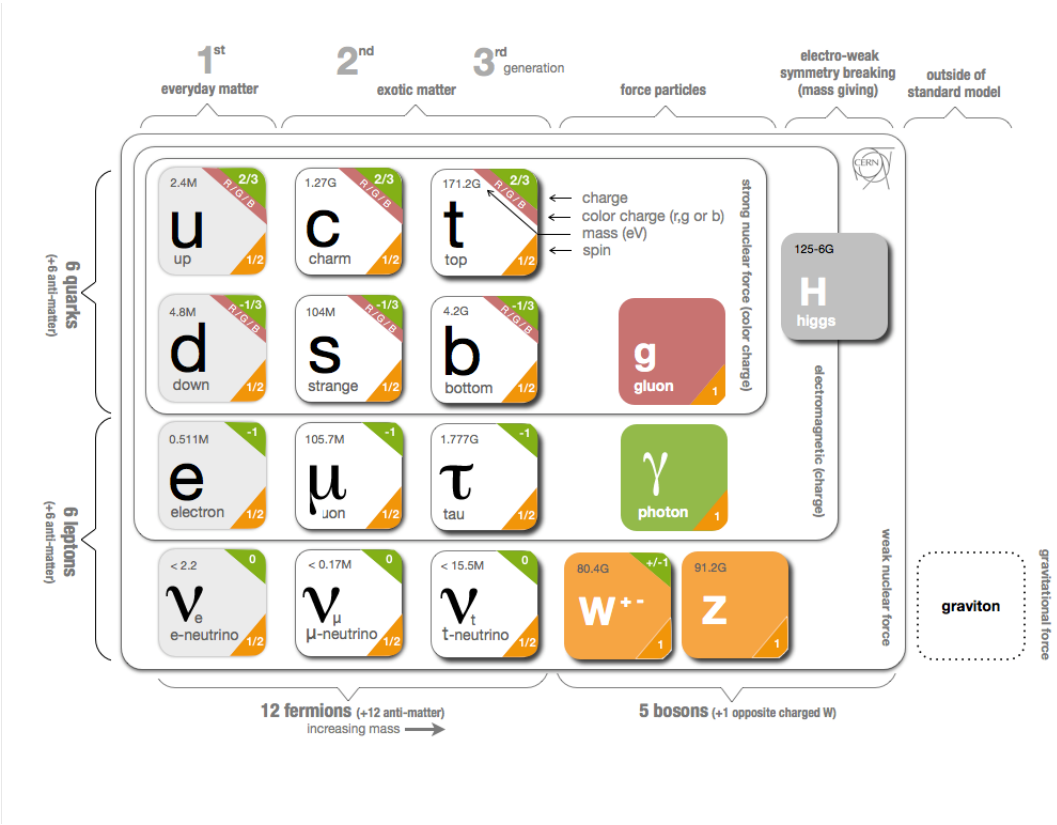


Figure 2.1: An overview of all elementary particles in the Standard Model. They can further be divided into two main groups, fermions (also referred as matter particles) and bosons (also referred as force-carrier particles). Fermions can be further divided into quarks and leptons, and according to their mass into three different generations. The graviton as a hypothetical particle for gravity is not considered in the Standard Model because there is no complete underlying quantum field theoretical description for its existence so far. Furthermore, the charge and spin of the elementary particles is shown in green and orange respectively. [1]

As shown in figure 2.1, all elementary particles known today are divided into two main groups. The first group are fermions (also referred to as *matter particles*), which can further be subdivided into leptons and quarks. The other group are bosons (also being called *force-carrier particles*), which

[1] The subscript C stands for the colour charge of the strong interaction, L denotes that $SU(2)$ is the left-handed weak isospin group and Y stands for the weak hypercharge and connects the electric charge Q with the weak isospin T_3 by $Y = 2\,(Q - T_3)$.

serve as mediators for the interactions. In the following section, these two groups are discussed in more detail, starting with a description of fermions, followed by the description of bosons.

Fermions

Fermions are half-integer spin particles that are subject to Fermi-Dirac statistics; a quantum statistical description for the distribution of identical particles in defined energy states obeying Pauli-Exclusion Principle. In detail, the Pauli-Exclusion Principle states that two or more identical fermions are not able to occupy the same quamtum state. For each fermion, there exist an antiparticle which is a particle with the same mass and spin as the particle, but opposite electric charge. Experimentally this is supported by the fact, when a regular particle and its corresponding antiparticle partner collide, both particles subsequently annihilate and disappear, releasing energy in the form of photons.

There are a total of six leptons in the Standard Model, three of which carry electric charge; the other three leptons are called neutrinos that have no charge and are massless, according to the Standard Model. Each charged lepton with its corresponding uncharged neutrino partner is assigned to a generation (also known as family).

The most accurate determination of the exact number of neutrino generations, was obtained by measurements of the visible interaction cross-section of the e^+e^- annihilation process in the Z-resonance region by measuring the rations of the Z boson partial decay width. There, the number of light neutrino species could be determined to be [2]:

$$N_\nu = 2.9840 \pm 0.0082.$$

The first of the three generations contains the electron (e^-) and the electron neutrino (ν_e). In the other two generations are the muon (μ^-) and the tau (τ^-) with their corresponding neutrino partners, the muon-neutrino (ν_μ) and the tau-neutrino (ν_τ). The main difference between these individual particles is the mass, which increases with generation, making the tau lepton the heaviest lepton in the Standard Model. But stable matter of the universe only comprises particles of the first generation, as the heavier particles of the other generations decay rapidly to lighter particles. Leptons are subject to the weak force, the gravitational force and the electromagnetic force.

Completely analogous to leptons, there are six types of quarks which are also grouped into pairs of different generations. The six quarks are: up (u), down (d), strange (s), charm (c), bottom (b) and top (t).

Because of the strong interaction and in particular the charge of the strong force called colour charge (force that acts between quarks), it is not possible to observe quarks individually. Since only colour-neutral states can occur in nature, quarks are only observed in bound states consisting of two or more quarks. For example, a bound state consisting of one quark (q) and one anti-quark ($\bar{q}$) is called a meson ($q\bar{q}$-state). The proton is an example of a particle consisting of three quarks, two up type quarks and one down type quark. This arrangement is called a baryon (qqq-state).

The weak force is special for quarks because it offers also the possibility to change the flavor of quarks. This is observed in beta decay, where a down quark is transformed into an up-quark.

Bosons

Bosons, on the other hand, are integer spin particles that are subject to Bose-Einstein statistics, which describes a set of identical particles, in contrast to fermions, can occupy the same quantum state.

Bosons act as mediators and are responsible for the fundamental interactions between leptons and quarks or interact among themselves and can further be divided into spin-0 and spin-1 bosons (scalar and vector bosons). There are five vector bosons, that can be seen as the mediators for the four fundamental forces: the gluon (g) for the strong interaction, the W boson (W^+, W^-) and Z boson (Z^0) for the weak interaction and the photon (γ) for the electromagnetic interaction.

The graviton G is proposed to be a spin-2 particle, but according to recent studies, adding the graviton would make the Standard Model inconsistent and requires modification [3]. The recently discovered Higgs boson (H^0) is a spin-0 scalar boson and is understood as a particle that gives mass to the gauge bosons via the higgs-mechanism.

The Higgs-mechanism has its origin in the electroweak symmetry breaking. Since the electroweak interaction describes the combination of the electromagnetic and weak interaction, it requires the electroweak gauge bosons to be massless (if electroweak symmetry would be by nature an exact symmetry). However, only the photon is massless while the $W^\pm$ and Z^0 bosons are massive gauge bosons. Therefore, extra mass terms for $W^\pm$ and Z^0 bosons should arise in the Yukawa terms[2], which would not be gauge-invariant and would lead to a breaking of the electroweak symmetry [4]. The main idea behind the Higgs-mechanism is to overcome this by a adding a complex double of scalar fields, referred as the Higgs field, with which the gauge bosons interact and whose ground state breaks the electroweak symmetry spontaneously giving mass to the gauge bosons. In the Standard Model, after spontaneous symmetry breaking, the lagrangian for the yukawa coupling to the Higgs field is given by

$$\mathcal{L}_{\mathrm{Y}}^{\mathrm{quarks}} = -\left(\bar{d}_L Y_d d_R + \bar{u}_L Y_u u_R\right) + h.c.,$$

where L and R indicate the up (u) and down (d) type quark components with left and right-handed chiralities, the Yukawa coupling matrices are given by Y_u and Y_d respectively [5]. They are three-by-three complex matrices, that need diagonalization using unitary transformations to obtain mass eigenstates in the Standard Model. Mass eigenstates of the quarks, also referred to as *physical states*, can be obtained then by the unitary transformations $\tilde{q}_A = V_{A,q} q_a$ for $q = u, d$ and $A = L, R$, where $V_{A,q} V_{A,q}^\dagger = 1$. After re-definitons of left-handed and right-handed fields separately, the CKM matrix (V_{CKM}) arises from flavour-changing charged current $W^\pm$ interactions of left-handed quarks described by the lagrangian [4]

$$\mathcal{L}_{\mathrm{CC}} = -\frac{g_2}{\sqrt{2}}\left(\bar{\tilde{u}}_L \gamma^\mu W_\mu^+ V_{CKM} \tilde{d}_L + \bar{\tilde{d}}_L \gamma^\mu W_\mu^- V_{CKM}^\dagger \tilde{u}_L\right),$$

where g is the electroweak coupling constant and V_{CKM} is the unitary 3×3 unitary matrix. The Cabibbo-Kobayashi-Maskawa (CKM) matrix is a 3×3 unitary matrix that contains the mixing between the different generation of quarks in the Standard Model and connects mass and weak eigenstates in the following form

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

[2] Yukawa interaction describes the coupling between the massless quark and lepton fields together with the higgs field. It is described by a lagrangian including kinetic terms and interaction terms of the particles.

The matrix itself describes the flavour changing transitions from one up-type quark to another down-type quark inside one family, but also between different families by the exchange of a W boson [6]. Its recent absolute values are giving by [7]

$$V_{CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.9742 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix}. \tag{2.1}$$

Its absolute values $|V_{ij}|$ from the CKM matrix show that the probability for the weak interaction between quarks of the same generation is particularly high, while the probability for the transition of off-diagonal elements between different families is suppressed. The values given in equation 2.1 are approximated values of the CKM elements, values with a higher precision can be found in [7].

2.2 Feynman Diagrams

For elementary processes, such as the scattering of particles or the decay of a particle, the use of Feynman diagrams has proven to be extremely useful [8]. They are widely used as a graphical tool for physicists to visualize particle interactions. Feynman diagrams were first introduced during the theoretical description of quantum electrodynamics.

Particles in a Feynman diagram are represented as lines. The time axis is usually set from left to right. An arrow indicates by its pointing direction, whether it is a particle or an antiparticle while the latter flows backward in time. The point of interaction where particles meet and get either annihilated or new particles are created is called an interaction vertex. The wavy lines generally represent bosons, except that the Higgs boson is indicated by a dashed line and gluons usually by a loop. Real particles are those particles that enter and leave the Feynman diagram, indicated by external lines, while virtual particles are described by intermediate lines.

Feynman diagrams are best explained with a simple example. One example processes is the interaction between an electron and a positron. This reaction is also known as Bhabha scattering and the corresponding Feynman diagram is shown in figure 2.2.

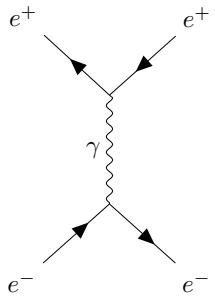


Figure 2.2: Feynman diagram for the electron-positron scattering process (Bhabha scattering).

This Feynman diagram describes the annihilation of an electron-positron pair to create a virtual photon that turns into an electron-positron pair. The diagram shows an example for a lowest order Feynman diagrams which usually has two vertices. At each vertex there is momentum conservation that means that the sum of all incoming momenta is equal to the sum off all outgoing momenta.

2.3 Concepts of Accelerator Physics Experiments

Particle accelerators are used to gain insights into the building blocks of matter. Generally speaking, these matter structures are particularly small and observing them requires high energy from colliding particles. Different types of particle accelerators have been developed but the two basic types are linear and circular accelerators.

As the name suggests, particles in linear accelerators are accelerated in the direction along the longitudinal axis and are shot at a stationary target. In circular accelerators, the particles counter-propagate in a circular path and result in a head-on collision with each other or with a target. Since these particles can be accelerated indefinitely by going around several times, the maximum energy that a circular accelerator can reach is higher than that of a linear accelerator. The most powerful circular accelerator is the Large Hadron Collider (LHC) which is introduced in chapter 3. The two major features that define an accelerator are the center-of-mass energy and luminosity. In this section, these quantities along with some general collider concepts and conventions are explained.

Partons and Parton Distribution Functions (PDFs)

Protons in a typical collision process at the LHC (see chapter 3 for more details) are extremely strongly bound to their constituents, also called *partons*. Responsible for this, is the strong nuclear force, that acts between the quarks and gluons of the proton. This force holds the nucleus together but is, however, not constant. It is described by quantum field theory by introducing a coupling constant α_S which describes the interaction between two particles dependent on the energy of the interacting particles. This dependency is weak in an electromagnetic interaction and becomes strong in the strong interaction [8].

At low energies, the strong coupling constant becomes large, while at high-energy scales, or equivalently at short distances, the coupling becomes asymptotically weak. In this asymptotic limit, quarks can be considered as free particles, which is called asymptotic freedom.

The parton model states, that hadrons and in particular protons are composed of pointlike quasi-free particles, called partons. In a high-energy scattering process of a proton, the reference frame is called an infinite momentum frame, where the mass of the proton and any transverse momentum can be neglected. Each parton from a hadron in an inelastic, hard collision process carries a fraction x of the total momentum of the hadron. This fraction is defined by the Bjorken-x variable as

$$x_i = \frac{p_i}{p_{\text{hadron}}}$$

, where p_i denotes the momentum carried by individual partons and p_{hadron} the momentum of the hadron. For a given parton-type, the momentum-fraction distribution is called parton distribution function (PDF). This distribution gives the probability, that a parton of type i carries a fraction x of the hadron momentum and it can be computed by pertubative methods [9]. At hadron colliders, the determination of the PDFs can be used to provide cross-section predictions. With the known PDF for each parton, the cross-section for a given process $p_1 p_2 \to X$ can be calculated as:

$$\sigma_{p_1 p_2 \to X} = \sum_{i,j=\text{partons}} \int dx_i dx_j \cdot f_1(f_i, \mu^2) f_2(x_j, \mu^2) \sigma_{ij \to X}(s, \mu^2).$$

, where $p1$ and p_2 denote two protons; f_1 and f_2 the PDFs of the protons, s the center-of-mass energy squared and σ (defined in 2.3.1) for the cross-section of the process.

2.3.1 Basic Concepts in Particle Physics Experiments

Natural Units and Conventions

In particle physics, it has been proven to be extremely useful to replace the widely used International System Of Units (SI) by a system of so-called *natural units*. Here, the speed of light c, the Planck constant $\hbar$ and the Boltzmann constant k_B are set to a fixed value of one,

$$c = 1 \quad \hbar = 1 \quad k_B = 1.$$

Since these constants appear relatively often in particle physics equations, setting them to one simplify these equations enormously and are therefore used in this thesis as well. The energy (E), momentum (p) and mass (m) in this system are expressed in units of energy, typically GeV

$$[E] = [p] = [m] = \text{GeV},$$

while the length (x) and time (t) are measure by the inverse of the energy as

$$[x] = [t] = \text{GeV}^{-1},$$

which follows directly from Heisenberg's uncertainty principle. For the interpretation of physical results, however, they must then be converted back into SI units.

Center-of-mass energy

The center-of-mass or center-of-momentum frame, is the reference frame in which the total sum of all incoming and outgoing particle momenta is equal to zero. The total energy of the system is called center-of-mass energy and is defined in equation 2.2. This energy also gives a limit to the particles that can be produced. In a collision of two particles, it can be expressed in a Lorentz-invariant form, given by

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{2} E_i\right)^2 - \left(\sum_{i=1}^{2} p_i\right)^2} \tag{2.2}$$

where E_i denote the energies and p_i the momenta of the two initial particles. When two particles are accelerated and subsequently collide, the center-of-mass energy rises linearly as $\sqrt{s} = 2E_{\text{beam}}$, assuming that the energy of both particles is identical ($E_1 = E_2$).

Luminosity and Cross-section

Over the past decades, the properties of elementary particles have been studied excessively. Many new particles have been discovered, which occur in rare processes where the production cross-section is quite small. Therefore, it is important to collect a lot of data so that enough information is available for further study.

A key indicator is the *luminosity* ($\mathcal{L}$), which provides information on how much the particles are squeezed together in bunches to provide a particularly high collision rate. The luminosity is defined as the number of collisions in a certain time per cross-section. More precisely, when two bunches contain n_1 and n_2 particles respectively in each beam and collide head-on with an average collision frequency f_{coll}, the luminosity reads [7]

$$\mathcal{L} = f_{\text{coll}} \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \tag{2.3}$$

In formula 2.3, σ_x and σ_y denote the transverse beam sizes in horizontal and vertical direction. A high luminosity can therefore be achieved with a high number of particles, many bunches and small particle beam dimensions at the interaction point. The interesting quantity, however, is the integrated luminosity, as it can be used to calculate the total number of interactions N. It can be obtained by the integrated luminosity over the lifetime of the detector operation and the cross-section via

$$N = \sigma \times \int \mathcal{L}(t)dt.$$

The interaction cross-section (σ) defines a measurement of the quantum mechanical probability that an interaction will occur [10]. The unit of the cross-section is barn (b) and is expressed in terms of area, where 1 b = 10^{-28} m^2. It is possible to define the cross-section as a function of variables such as the solid angle, the transverse momentum or the energy.

In this case, the cross-section is called *differential* cross-section (or exclusive). For example, instead of only measuring the event rate one can get additional information about a scattering process by measuring the angular distribution. If the cross-section measurement covers all scattering angles, is called total (or inclusive) cross-seciton.

Rapidity and Pseudorapidity

For the explanation of rapidity and pseudorapidity, the reference system of high energy projectile-target collisions must be considered. At hadron colliders, such as the LHC, the interaction takes place in the center-of-mass system of the colliding particles. Since a particle consists of a composition of sub-constituents, the energies and momenta of these so-called *partons* are not known a priori because the reference system of the partons is not the same as the reference system of the incident colliding particles. For partons, it is more likely that the final-state jets of the collision experience a boost along the beam direction and jet angles being expressed in terms of rapidity [10].

The system is defined by the z-axis being the direction of motion of the incident particle beam, while the interaction point takes place in the origin of the $x - y$ plane. If E denotes the energy and p_z the longitudinal momentum of the jets, then the rapidity is defined as:

$$y = \frac{1}{2} \log\left(\frac{E + p_z}{E - p_z}\right).$$

It is particularly useful that different rapidities are invariant under boosts along the trajectory direction. Thus, measuring the differences in rapidity in any two frames will be the same and they only transform additively [10].

A much easier quantity to determine is the pseudorapidity which only depends on the polar angle θ of a particle's trajectory and is given via:

$$\eta = -\log \tan \frac{\theta}{2}.$$

For relativistic particles, where the mass of the particles becomes negligible at high energies, the precise measurement of energy or longitudinal momentum to determine rapidity can sometimes be difficult. In can be shown that in this case rapidity and psuedorapidity are equal.

Pile-up

Pile-up refers to the occurrence of several collision interactions at the same time that are assigned to one event. In this case, the detector measures objects that do not originate from the primary vertex where the hard scattering interaction occurs. There are two different sources of pile-up: in-time and out-of-time. The first refers to additional proton-proton collisions that occur in the same bunch-crossing as the collision of interest, while the latter one occurs when additional proton-proton collisions occur right before and after the collision of interest [11]. If one increases the luminosity, one increases the pile-up which can then make physics analysis difficult due to the huge amount of unwanted background events in the detector that emerge from these additional interactions. Figure 2.3 shows the different pile-up distributions that were recorded at ATLAS for different years during Run-2.

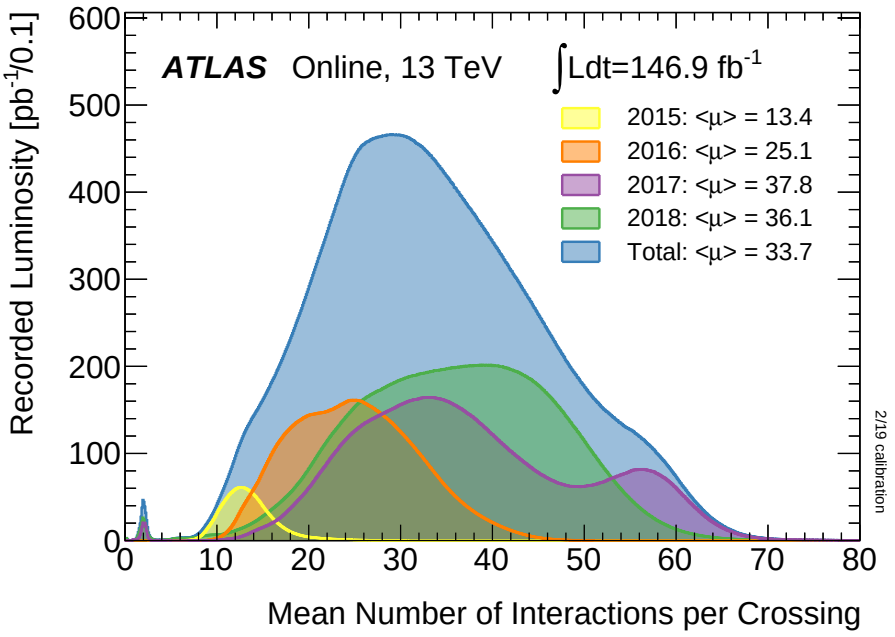


Figure 2.3: Pile-up distributions recorded at ATLAS during Run-2 [12].

2.4 Top Quark Physics at Hadron Colliders

Ever since the discovery of the bottom quark at the Tevatron (Fermilab, Chicago) in the 1970s, the top-quark was predicted to be a member of the weak-isospin doublet of the bottom quark. Its existence would solve the missing puzzle piece in the Standard Model, making it a theory of three quark generations.

The only particle collider at that time that had the necessary energy and was capable to detect the top-quark was the Tevatron. At Tevatron, protons and antiprotons were accelerated to high energies and made to collide, producing top-quark pairs. Exactly this $p\bar{p} \rightarrow t\bar{t}$ reaction at a center-of-mass energy of 1.8 TeV and 1.96 TeV with the subsequent decay into two W bosons and two bottom-quarks from the top-quark decay $t\bar{t} \rightarrow WbW\bar{b}$ ultimately leading to the discovery of the top-quark in 1995 by the CDF and DØ collaborations [13, 14]. For this measurement, approximately 10 fb^{-1} of integrated luminosity data was taken [7].

The top-quark is the heaviest of all fundamental particles, even heavier than a W boson and its properties both in production and decay have been studied in detail. This section gives a brief overview of the properties of the top-quark, followed by the description of top-quark production mechanisms. Furthermore, the decay and rare top-quark production processes are reviewed.

2.4.1 Properties of the Top Quark

The properties of the top-quark have been studied by the CDF and DØ experiments at the Tevatron Collider at Fermilab and by the ATLAS and CMS experiment at the LHC at CERN [15]. All measured properties so far are consistent with the Standard Model of particle physics. The top-quark is the up-type quark partner of the bottom quark and has therefore a weak isospin of $T_3 = +1/2$. Additionally, it has a charge of $Q = +2/3$ and it has like all other fermions a spin of $s = 1/2$, as already shown in figure 2.1. To determine the charge of the top-quarks, one has to look at the decay of the top-quark. The top-quark decays into charged particles whose curvature can be detected in the magnetic fields of the detector. The direction of the curvature can then be used to extract the charge of the decayed charged particles.

The most studied property is the mass that has been measured precisely making the top-quark the heaviest particle of the Standard Model with a mass of $m_t = 173.21 \pm 0.51(\text{sys}) \pm 0.71(\text{stat})$ GeV [7]. Because of the huge mass, it a short lifetime of about $\tau_t \approx 0.5 \times 10^{-24}$ s. A measurement of the top-quark mass can therefore be determined directly from the decay products, because the top-quark decays in the detector before it hadronizes, which then relies on the exact reconstruction of the kinematics and reconstruction efficiency [7, 16]. A summary of the latest results from direct top-quark mass measurements from ATLAS and CMS are given in figure 2.4.

The large mass of the top-quark is directly connected to the Yukawa coupling strength of the Higgs field which is in the order of almost one making this particle an interesting enterprise for physics beyond the Standard Model [17]. Together with the mass of the W boson, the knowledge of the top-quark mass enables then the possibility to measure the mass of the Higgs boson. Since the mass of the top-quark is not predicted by the standard model, this indirect measurement can then be compared with the direct measurement of the Higgs boson mass to check for inconsistencies in the Standard Model [15].

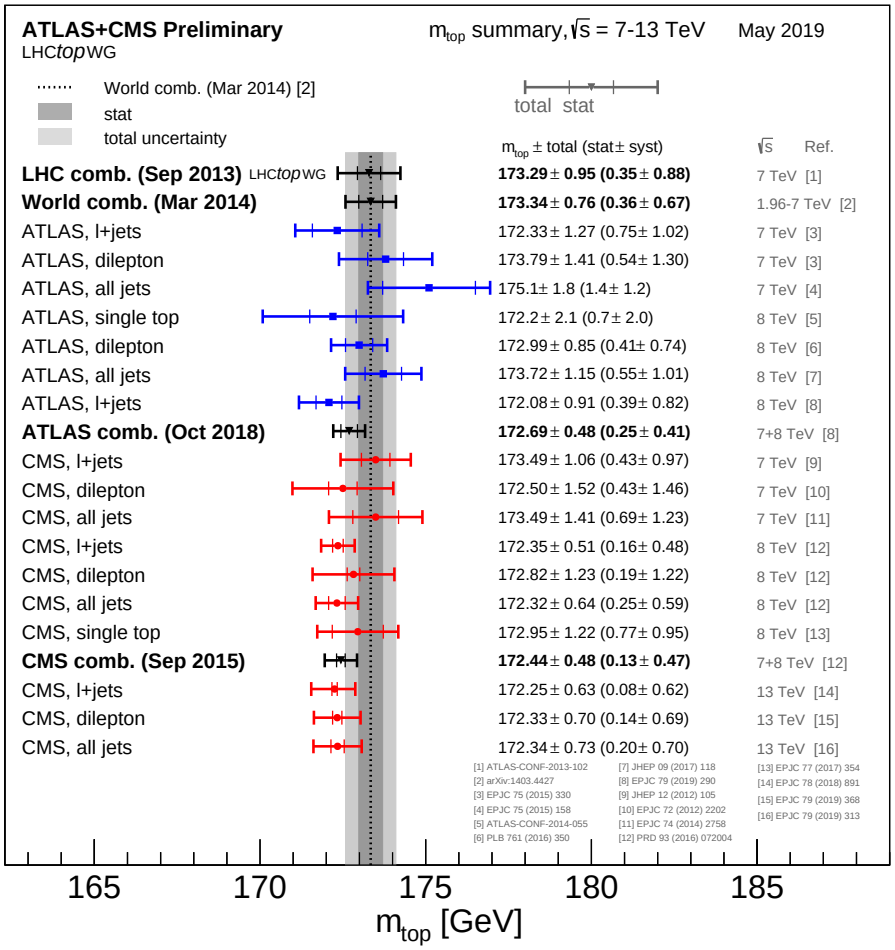


Figure 2.4: Summary of the latest direct top-quark mass m_t, here refered to as m_{top}, measurements from ATLAS [18].

2.4.2 Top Quark Production Processes

The first appearance of the top-quark at the Tevatron in 1995 was produced in pairs in a strong interaction process. This is the dominant production process of top-quarks at hadron colliders. There are several processes that produce top-quarks in the final state, but they all can be divided into two main groups based on their interaction. The first group is the top-quark pair production via the strong interaction and the second group is the production of single top-quarks via the electroweak interaction which was also observed by the CDF and DØ collaborations at Tevatron in 2009 [19]. In the following section the single top-quark production as well as pair production is described individually.

Top Quark Pair Production

The production of top-quarks in pairs is only possible via strong interaction and occurs either via the annihilaton of a quark-antiquark pair ($q\bar{q} \rightarrow t\bar{t}$) or via gluon-gluon fusion ($gg \rightarrow t\bar{t}$). At Tevatron, the dominant production mechanism of $t\bar{t}$ with approximately 85% of the total production cross-section is

the quark-antiquark annihilation while at LHC energies about 90% at $\sqrt{s} = 14$ TeV of the production is from the latter gluon-gluon fusion process (approximately 80% at $\sqrt{s} = 7$ TeV) [7, 20]. The differences in the various interaction cross-sections for these two processes are directly related to the differences in the parton distribution functions (PDFs) for quarks and gluons. Figure 2.5 shows the Feynman diagrams for the top-quark pair production at leading order (LO) for quark-aniquark annihilation as well as for the gluon induced process for both the s-channel and the t-channel.

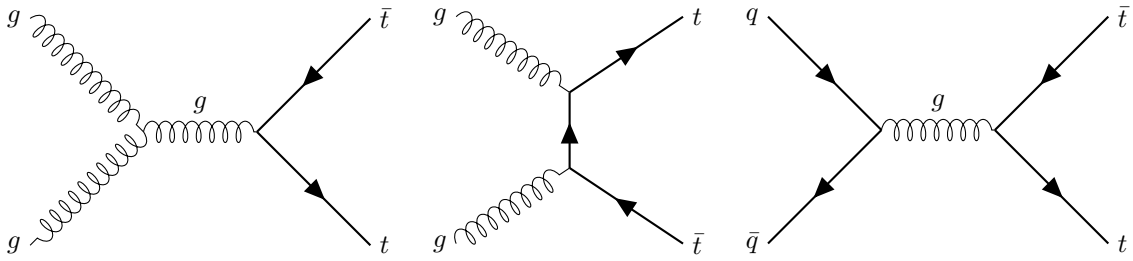


Figure 2.5: Leading Order (LO) Feynman diagrams for the top-quark pair production: gluon induced s-channel (left), t-channel (middle) and quark-antiquark annihilation ($q\bar{q}$) (right).

The latest summary of results for the energy-dependent cross-section at next-to-next-to leading order (NNLO) with next-to-next-to-leading-log (NNLL) soft gluon resummation are shown in figure 2.6. This summary includes the cross-section measurements of LHC and Tevatron as a function of the center-of-mass energy assuming a top-quark mass of 172.5 GeV [21].

Single Top Quark Production

The production of single top-quarks occurs via electroweak interaction and can be distinguished between three different subprocesses based on the virtuality of the exchanged W boson: the s-channel, t-channel and the tW channel[3]. In figure 2.7, the Feynman diagram for each process is shown.

Of all the three processes, the t-channel is the dominant process that involves the exchange of a W boson between a light quark and a bottom quark which results in a jet and a single top-quark ($q\bar{b} \rightarrow q't$) [22]. It contributes around 73% of the total single top-quark production at the LHC. Having a single top-quark in the final state provides a precise measurement of the $|V_{tb}|$ CKM matrix element and may reveal sights for new physics beyond the Standard Model [19].

The second most common process is the associated production of a top-quark with a W-boson in the tW-channel which occurs one fourth of the total single top-quark production at the LHC [19]. In this process, a top-quark is produced along with a W boson ($bg \rightarrow W^{-}t$).

The s-channel is the least dominant process in which a quark-antiquark pair in the initial state produces a $t\bar{b}$ final state ($q\bar{q} \rightarrow t\bar{b}$). Observing single top-quarks is a challenge due to the low expected cross-section of the combined s and t-channel, which is much smaller than relevant background processes. Therefore, a variety of multivariate analysis techniques, such as boosted decision trees or neural networks have been used to separate single top-quark events from background events (see [23] and [24] for more details). Table 2.1 summarizes the recent values for the measured single top-quark production cross-section for different center-of-mass energies at ATLAS and CMS.

[3] s, t and u refer to the Lorentz-invariant mandelstam variables

	Process	ATLAS		CMS	
$\sqrt{s}$ [TeV]	Channel	σ [pb]	Lumi. [fb^{-1}]	σ [pb]	Lumi. [fb^{-1}]
7	t-channel	$68 \pm 2 \pm 8 \pm 1$	4.59	$67.2 \pm 3.7 \pm 4.6 \pm 1.5$	1.17 to 1.59
	tW-channel	$16.8 \pm 2.9 \pm 4.9 \pm 0.6$	2.05	$16 \pm 3^{+4}_{-2} \pm 1$	4.9
	s-channel	< 26.5	0.7	7.1 ± 8.1	5.1
8	t-channel	$89.6 \pm 1.2^{+6.8}_{-5.9} \pm 1.7$	20.2	$83.6 \pm 2.3 \pm 7.1 \pm 2.2$	19.7
	tW-channel	$23.0 \pm 1.3^{+3.2}_{-3.5} \pm 1.1$	20.3	$23.4 \pm 1.9 \pm 5.0 \pm 0.6$	12.2
	s-channel	$4.8 \pm 0.8^{+1.6}_{-1.3}$	20.3	13.4 ± 7.3	19.7
13	t-channel	$247 \pm 6 \pm 45 \pm 5$	3.2	$207 \pm 2 \pm 30 \pm 5$	35.9
	tW-channel	$94 \pm 10^{+28}_{-22} \pm 2$	3.2	$63.1 \pm 1.8 \pm 6.4 \pm 2.1$	35.9
	s-channel	-	-	-	-

Table 2.1: Summary of the measured cross-sections (σ) for the different single top-quark production subprocesses at different center-of-mass energies $\sqrt{s}$ = 7 TeV, $\sqrt{s}$ = 8 TeV and $\sqrt{s}$ = 13 TeV, published by the ATLAS and CMS Collaborations [7, 18].

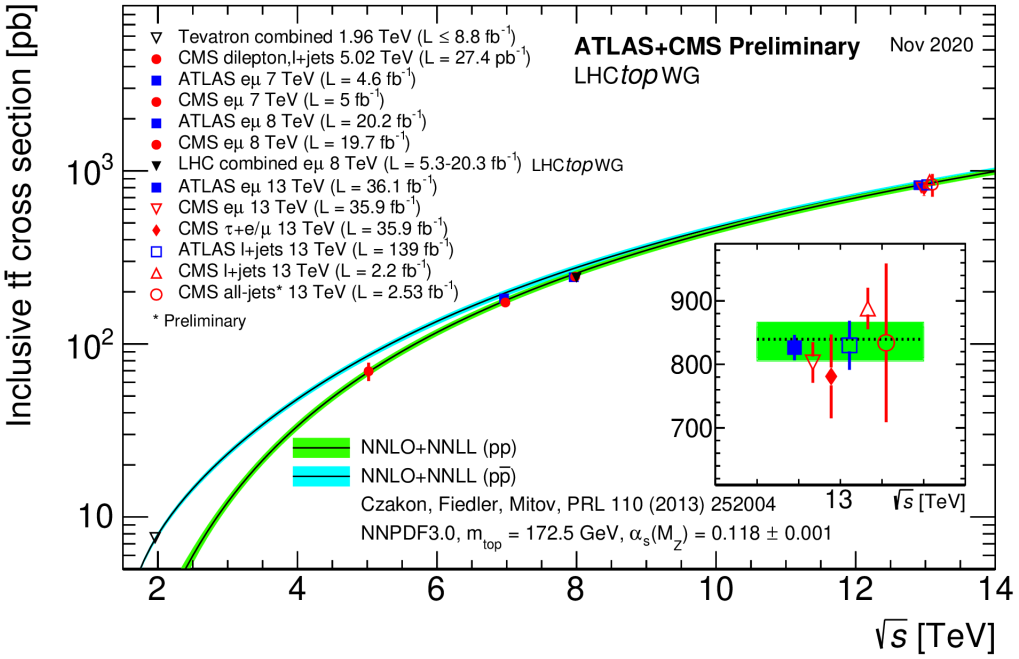


Figure 2.6: Summary of the cross-section measurements by LHC and Tevatron as a function of different center-of-mass energies. The results were performed at next-to-next-to leading order (NNLO) with next-to-next-to-leading-log (NNLL) soft gluon resummation assuming a top-quark mass of 172.5 GeV [21].

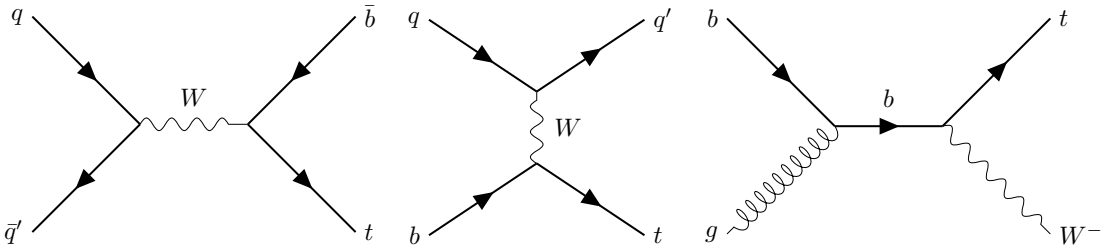


Figure 2.7: Leading Order (LO) Feynman diagrams for single top-quark production: single top-quark production in the s-channel (left), the t-channel (middle) and the tW-channel (right).

2.4.3 Rare Top Quark Production Processes

By increasing the luminosity, the amount of collision data increases and allows processes with relatively low cross-sections to become accessible. Many of such rare processes have top-quarks in the final state, such as tZq, $t\bar{t}H$ or tHq and are referred to rare top-quark production processes [25].

Measurements at the LHC Run-2, for example, have currently reached enough sensitivity with sufficient energy to investigate these processes. Some of these processes include the associated production of a top-quark with a heavy vector boson which provides a direct measurement of the top-quark coupling. In the following, an overview of some of these rare processes is given and in

addition to that a closer look into the associated production of a single top-quark with a Z or a Higgs boson.

- $t\bar{t}V$ $(V = Z, W^{\pm})$: In this process a top-quark pair is created in association with a heavy gauge boson. A measurement of the $t\bar{t}Z$ process for example allows to extract information about the neutral-current coupling of the top-quark. The most recent measurements of the $t\bar{t}Z$ and $t\bar{t}W$ production cross-section have been performed at ATLAS and CMS at $\sqrt{s} = 13\,\text{TeV}$ in multilepton final states with $36.1\,\text{fb}^{-1}$ of events with two, three or four leptons. The results are [7]:

$$\sigma_{t\bar{t}W} = 0.87 \pm 0.13(\text{stat}) \pm 0.14(\text{sys})\,\text{pb}$$

$$\sigma_{t\bar{t}Z} = 0.95 \pm 0.08(\text{stat}) \pm 0.10(\text{sys})\,\text{pb}$$

- $t\bar{t}H$: This process stands for the Higgs boson production in association with a top-quark pair and this production mode allows a direct measurement of the top Yukawa coupling. The production of that process was observed by the ATLAS and CMS experiment at $\sqrt{s} = 13\,\text{TeV}$ with an integrated luminosity of $80\,\text{fb}^{-1}$. The result for the cross-section measurement was [26, 27]:

$$\sigma_{t\bar{t}H} = (294\,^{+182}_{-162})\,\text{fb}$$

- tZq: This process refers to the associated production of a top-quark in association with a Z boson and a quark and will further be described in detail in section 2.4.4.

- tHq: This process refers to the production of a single top-quark in associated with a Higgs boson and a quark, which will also be further explained in section 2.4.5.

The ATLAS collaboration announced recently the evidence for the production of four top-quarks ($t\bar{t}t\bar{t}$) which is extremely rare and occurs only once for every 70 thousand pairs of top-quarks created at the LHC [28]. Due to its large number of top-quarks in the final state, this process is sensitive to the parameters of the Standard Model and can be used to access the Yukawa coupling between the top-quark and the higgs boson [25].

2.4.4 Single top-quark production associated with a Z boson

As described in the previous section, the tZq process is referred to the production of a top-quark in association with a Z boson and a quark. The top-quark is produced via the electroweak interaction through the t-channel process. In the tZq process, the Z boson is either radiated off one of the participating quarks or from the exchanged W boson [29]. The LO Feynman diagrams for this process are given in a 4-flavour scheme in figure 2.8. The figure shows that the Z boson is radiated from any of the quark lines [29].

The Z boson in the tZq process decays into lighter leptons, like electrons or muons and the W boson, that comes from the top-quark decay, decays leptonically into an electron or muon and a corresponding neutrino, and also includes the contribution from a τ lepton that decays into electrons or muons [30]. The tZq process contains WWZ and tZ couplings, therefore this process enables to probe two SM couplings simultaneously [31]. ATLAS reported evidence for this process with a with a measured (expected) significance of $4.2\,\sigma(5.4\,\sigma)$ [32].

The ATLAS collaboration measured the production cross-section for that process in the trilepton channel at a center-of-mass energy of 13 TeV with a corresponding luminosity of 139 fb^{-1} to be [33]:

$$\sigma_{ATLAS}(pp \rightarrow tZq \rightarrow tl^{+}l^{-}q) = (97 \pm 13\,(\text{stat.}) \pm 7\,(\text{sys.}))\,\text{fb}.$$

CMS performed a similar production cross-section measurement is the same trilepton final state using a center-of-mass energy of 13 TeV with a corresponding luminosity of 77.4 fb^{-1} for dilepton invariant masses above 30 GeV to be [34]:

$$\sigma_{CMS}(pp \rightarrow tZq \rightarrow tl^{+}l^{-}q) = (111 \pm 13\,(\text{stat.})\,^{11}_{-9}\,(\text{sys.}))\,\text{fb}$$

Both measured production cross-sections are consistent with the prediction with the Standard Model.

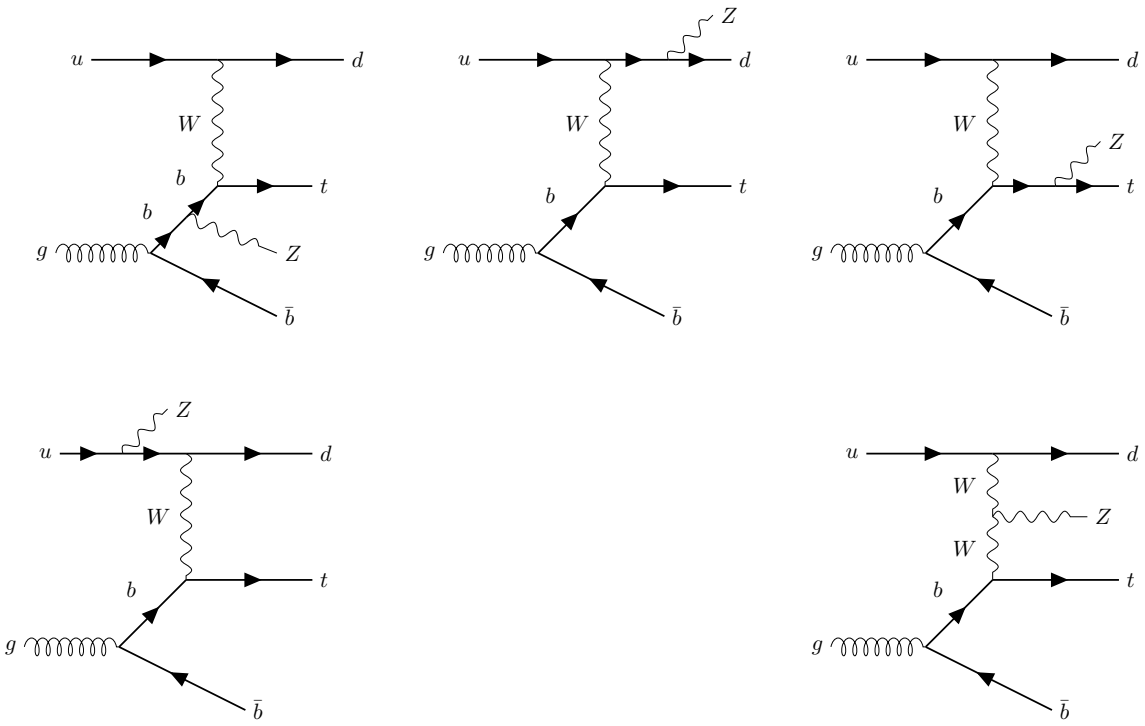


Figure 2.8: LO Feynman diagram for the tZq processes.

2.4.5 Single top-quark production associated with a Higgs boson

The tHq process is already introduced in section 2.4.3 as the process that refers to the production of a single top-quark in association with a Higgs boson and a quark. This process proceeds mainly through t-channel diagrams [35], where the Higgs boson is either emitted from a W boson or a top

quark. The single top-quark production in association with a Higgs boson can also be accompanied by a W boson in the final state (tHW) [35].

Because the Feynman diagrams for both processes interfere destructively, the production cross-section is really small. The cross-sections are 71, 16 and 2.9 fb for the t-channel, tW process and s-channel at a center-of-mass energy of 13 TeV[36].

2.4.6 Top Quark Decay

Because of its large mass, the top-quark decays almost exclusively into a W boson and a bottom quark ($t \to Wb$). This is the favored decay channel of the top-quark while decay into the other down type quarks ($t \to Wd$ and $t \to Ws$), is supressed by the CKM matrix elements, and can be reviewed in formula 2.1 [37]. The W boson then decays leptonically into a lepton and its corresponding neutrino partner ($W \to l\bar{\nu}_l$) or hadronically into a quark-antiquark pair ($W \to q\bar{q}'$).

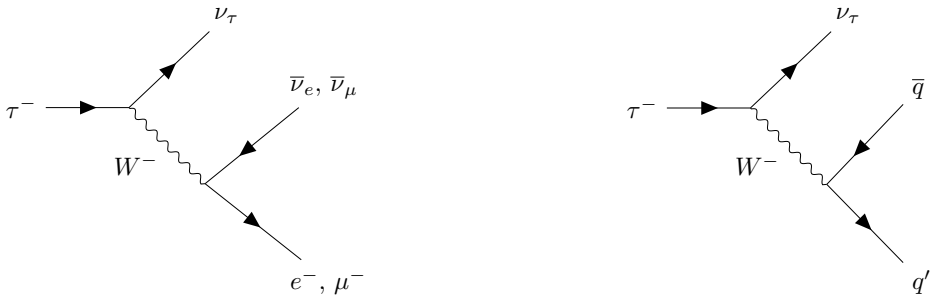
The final state for a leading top pair production $t\bar{t}$ event is therefore most often consisting of six jets, where two hadronic jets arise from the hadronization of the two b quarks (*b-jets*) and the remaining four jets from the hadronic decay of the W bosons into quark-antiquark pairs (further explanation in 4.3.1). This channel is referred to the *all-hadronic* channel [7]. One W boson can also decay leptonically while the other one decays hadronically resulting in a *lepton+jets* (l+jets) channel. In addition, both W bosons can decay leptonically in a so-called *dilepton* channel. The different decay channels with their relative fractions, where l refers to e or μ and are given in the following [7]:

all-hadronic:	$t\bar{t} \to W^+bW^-\bar{b} \to (q\bar{q}')b(q''\bar{q}''')\bar{b}$	(45.7 %),
l+jets:	$t\bar{t} \to W^+bW^-\bar{b} \to (q\bar{q}')bl^-\bar{\nu}_l\bar{b} + l^+\nu_l bq''\bar{q}'''\bar{b}$	(43.8 %),
dilepton:	$t\bar{t} \to W^+bW^-\bar{b} \to l^+\nu_l bl\bar{\nu}'_l\bar{b}$	(10.5 %).

2.5 Tau Leptons

The tau lepton (τ) and its corresponding neutrino partner, the tau neutrino ν_τ, were introduced as leptons of the third fermion generation in the previous section. Since its discovery by the SLAC group in 1975 with the SPEAR accelerator, the tau lepton's properties have been studied during the last decades [38, 39]. Due to its short lifetime of about 2.9×10^{-13} s it can only be observed and reconstructed through its decay products [7].

Like all other leptons in the Standard Model, the tau lepton has a charge of -1 and the third isospin component $-1/2$. Its the heaviest of all leptons with a mass of 1 776.86 MeV and is the only one massive enough to decay both hadronically and leptonically [7]. In 35% of all decays, it decays leptonically in a lighter lepton (electron or muon) plus its neutrino and antineutrinos (leptonic mode τ_{lep}), while the remaining 65% are the hadronic decay in one or more hadrons (hadronic mode τ_{had}). Figure 2.9 shows the Feynman diagrams for both, the leptonic and hadronic decay mode, while figure 2.10 shows a summary of the branching ratios for the most common hadronic decay modes of the tau lepton.



(a) Leptonic tau decay

(b) Hadronic tau decay

Figure 2.9: Feynman diagrams for the tau lepton decay modes. On the left, the leptonic and on the right, the hadronic tau decay.

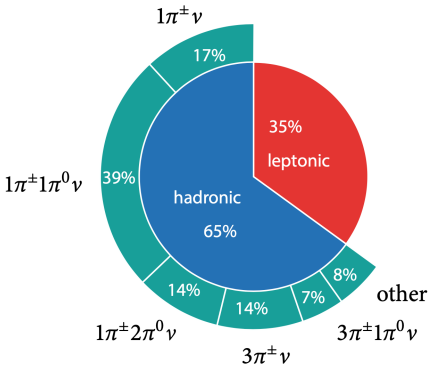


Figure 2.10: Summary of the branching fractions for the most common hadronic tau lepton channels as well as the numbers of charged particles involved [40].

At ATLAS, the reconstruction of electrons and muons from the leptonic decay of the tau lepton are hard to distinguish from electron and muons form the initial proton-proton collision process. The moment the tau decays leptonically, only the electron or muon can be observed in the detector, while both neutrinos pass through the detector without any interaction.

The decay signatures for leptonically decaying tau leptons can therefore not be distinguished experimentally, while in contrast, the emerging quarks from the hadronic mode, leave clear signatures. Thus, the reconstruction only refers to hadronically decaying tau leptons.

The overwhelming background from QCD processes at the LHC are a large unwanted source for the identification of hadronic taus, because the kinematics of QCD jets (jets initiated from the fragmentation of quarks and gluons, section 3.3.3 for more details) look similar to that of hadronically decaying tau leptons. [41]. Another less dominant background is mostly from electrons and muons which can mimic the signature of a tau lepton that is decaying with one charged hadron [42]. With similar kinematics, these sources mentioned above can be misidentified as tau leptons and are referred to as misidentified taus or fake taus. The identification and reconstruction is described in section 3.3.5 of the next chapter.

CHAPTER 3

The ATLAS Experiment at the Large Hadron Collider

"Verlockend ist der äussere Schein. Der Weise dringet tiefer ein."

- Wilhelm Busch,

Various machines for the acceleration of charged particles have been developed since the early 1920s. The desire of particular interest was always to reach higher energies in order to dive deeper into the inner structure of particles and to better understand the interactions between them. Of course, the search for new particles and the proof of the validity of the Standard Model is also one of its crucial challenges.

The LHC is an excellent accelerator for this purpose as the most powerful and largest accelerator ever build. Physics has benefited greatly from its many new discoveries in the recent years. One of the pinnacle of scientific results at the LHC was the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments.

The work in this thesis is based on simulation of events and events taken from the ATLAS experiment at the LHC. In the following sections, an overview of the LHC and the ATLAS experiment is given. Furthermore the sub-componentents of the ATLAS detector are briefly explained. The chapter closes with the reconstruction and identification techniques of physical objects.

3.1 The Large Hadron Collider (LHC)

The Large Hadron Collider was built between 1998 and 2008 and is the world's largest particle accelerator, with a circumference of 27 km. It is located at the European Laboratory for Particle Physics (CERN [1]) near the Swiss-French border close to Geneva, Switzerland. Figure 3.1 shows a map sketch in which the LHC accelerator ring resides at a depth of 100 m under the earth and it shows additionally the four collision points corresponding to the main experiments that are involved: ATLAS, CMS, LHCb and ALICE [43].

When the accelerator was built, the existing tunnel facilities of the previous accelerator, the LEP, were reused as the basis for the conversion to the LHC. The LHC is a proton-proton accelerator that is

[1] CERN: Organisation européenne pour la recherche nucléaire.

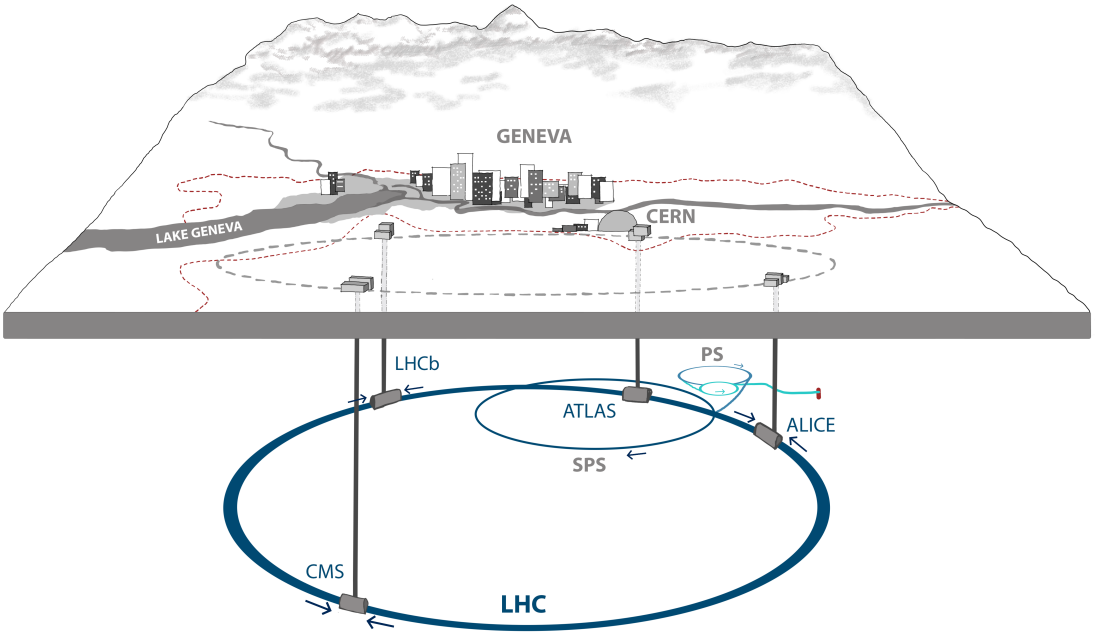


Figure 3.1: Overview sketch of the Large Hadron Collider located near the Swiss-French border and its four main experiments ATLAS, CMS, LHCb and ALICE at CERN. Image drawn based on the original source [44].

designed to deliver a center-of-mass energy of 13 TeV, which is about seven times higher than the energy of the Tevatron at Fermilab. Besides the acceleration of protons, the LHC also accelerates and collides heavier nuclei. In order to keep the particles on stable paths and at particular high energies, more than 6700 magnets and accelerating structures are used. The former LEP accelerator used traditional dipole magnets that provided a 0.3 T magnetic field which have been replaced by superconducting NbTi magnets at LHC due to saturation effects in iron that limit the maximally achievable magnetic field up to 2 T. By the use of superconducting magnets, the magnetic field could be pushed to $B = 8.33$ T at a temperature of 1.9 K [4].

The acceleration complex at CERN consists of accelerator chains that are described in the following. At the beginning of the accelerator chain, protons are extracted from hydrogen gas, where an electric field is used to separate electrons from the hydrogen atoms to yield protons. Single protons are then send to a linear accelerator, the so-called LINAC2, where they get accelerated to a kinetic energy up to 50 MeV. After the beam emerges the Linac, it is injected into the Proton Synchrotron Booster (PSB) which boosts the protons to even higher energies up to 1.4 GeV. The acceleration is continued with

the Proton Synchrotron (PS) up to 25 GeV. They are then send to the Super Proton Synchrotron (SPS) and are eventually injected with an energy of 450 GeV into the two beam pipes of the LHC, where they circulate in opposite direction with each an energy of 7 TeV and brought into collision.

Each proton beam contains 2808 bunches with 10^{11} protons per bunch with a time separation between each bunch of 25 ns. Particles at the LHC reach incredible high velocities near the speed of light at a design luminosity of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which was first reached in June 2016. A future project, the High-Luminosity LHC is planned to increase the luminosity by a factor of 10 beyond the LHC's design value [45].

Experiments at the LHC

As mentioned earlier, the LHC has four main experiments. These are the two general-purpose experiments **ATLAS** and **CMS** (**C**ompact **M**uon **S**olenoid) as well as the two other experiments **ALICE** (**A** **L**arge **I**on-**C**ollider **E**xperiment) and **LHCb** (**LHC** **b**-hadron experiment).

The aim of the first two experiments, the ATLAS experiment and CMS is to study the fundamental particles and their interactions at high precision with the help of an enormous number of particle collisions. The by far most famous achievement to date was the independent discovery of the Higgs boson by the ATLAS and CMS experiment, that opened a new era for further study of new physics that lies beyond the Standard Model. LHCb investigates rare decays of B hadrons and performs precise measurments of the asymmetrie between matter and antimatter which is related to CP violation. At ALICE, the study of the strong interaction sector of the Standard Model is investigated where the study mainly focuses on the search for signatures of the quark-gluon plasma (QGP), which defines the state where quarks and gluons can be regarded as quasi-free particles.

The data of proton-proton collisions used for the analysis in this thesis were recorded by the ATLAS detector, therefore the next section gives an introduction of its major components and how particles are reconstructed.

3.2 The ATLAS Detector

The ATLAS Experiment is a multi-purpose particle detector that sits on one of the four collision points of the LHC, as already seen in figure 3.1. It has a nominal forward-backward symmetric cylindrical geometry with respect to the interaction point. After a collision, the produced particles emerge around the interaction point and their paths and energies are measured with the ATLAS detector. With a length of 46 m, a diameter of 25 m and a total weight of the detector is 7 000 t it is the largest detector ever build [46]. An overview of the ATLAS detector and its main components, the Inner Detector, the Calorimeter, the Muon Spectrometer and Magnet System are shown in figure and introduced in the following figure 3.2.

The Coordinate System of ATLAS

The coordinate system of the ATLAS detector is a right-handed coordinate system with the origin at the interaction point in the centre of the detector [48]. The z-axis is pointing along the beam pipe, the x-axis torwards the centre of the LHC ring and and the y-axis points into the vertical, upwards

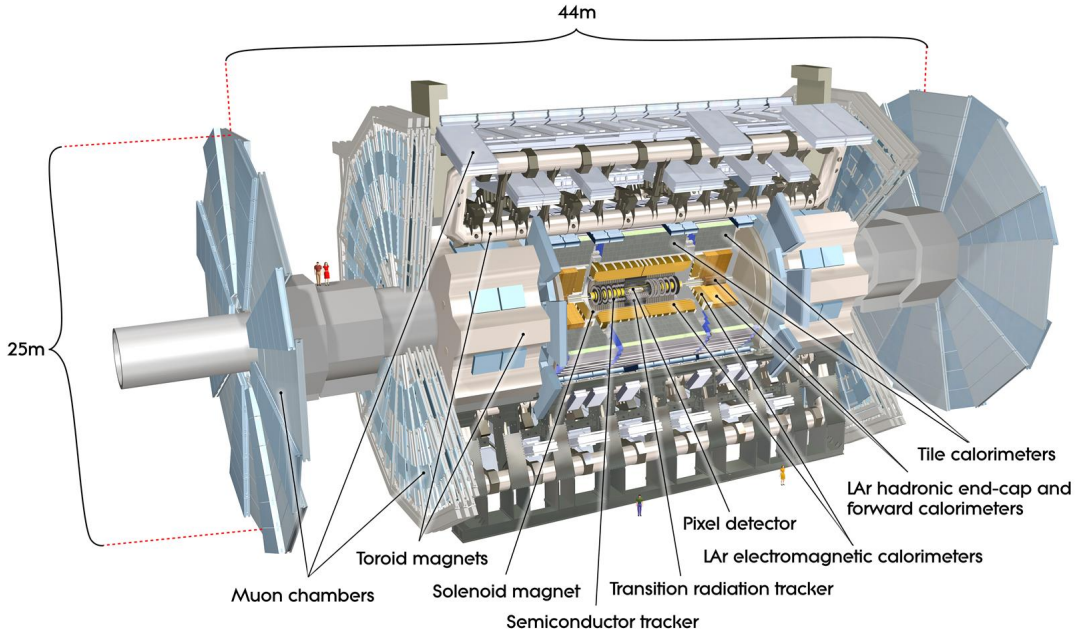


Figure 3.2: Schematic cut-away view of the ATLAS Detector and its four major components: the Inner Detector (ID), calorimeter, muon spectrometer and magnet system [47]

direction [48]. As already mentioned in section 2.3.1, the pseudorapidity is defined in terms of the polar angle θ as $\eta = -\log\tan\frac{\theta}{2}$. The polar angle is measured as the angle between the direction of the particle and the beam axis.

Cylindrical coordinates (r,ϕ) are used in the transverse plane, where ϕ denotes the azimuthal angle around the beam pipe and r being the distance from the beam line [48]. The transverse momentum p_T is referred to the momentum perpendicular to the beam axis. To measure distances, the η-ϕ plane is used and can be calculated via $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ [29].

3.2.1 The Inner Detector

Relatively close to the interaction point is the inner detector (ID) which is responsible for the reconstruction of tracks and vertices of particles that are produced from the proton-proton collision process with high efficiency [49]. This includes the measurement of charge, momentum and direction of these particles with $p_T > 0.5\,\text{GeV}$ and an acceptance of $|\eta| = 2.5$ [50]. Its dimensions are 6.2 m in length and a diameter of 2.1 m. From inside to outside, the inner detector consists of three components: the Pixel Detector covering radial distances between 50.5 mm and 122.5 mm, Semiconductor Tracker (SCT) covering radial distances from 299 mm to 514 mm and the Transition Radiation Tracker (TRT) covering radial distances from 554 mm to 1 082 mm [51]. All three components are situated in a 2 T solenoidal magnetic field and are illustrated in figure 3.4. In addition to that, also their radial distance coverage is shown. For Run-2 a new innermost layer, the IBL was inserted which provides an additional measurement point closer to the interaction point for reconstruction and vertexing [52].

An illustration of the inner detector setup is given in figure 3.3. The three major components are described individually in the following.

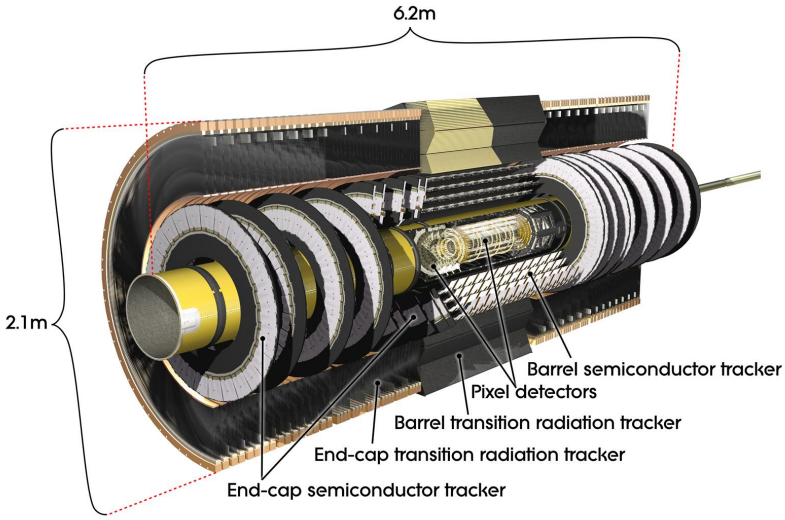


Figure 3.3: Schematic of the ATLAS Inner Detector and its three major components: the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker [53]

The Pixel Detector

The ATLAS Pixel Detector is the innermost component of the inner detector and is subdivided into four concentric barrel layers of silicon pixel sensors and in total six disk layers, with three pixel disks at each end-cap region. Each silicon pixel sensor has a size of $50 \times 400\mu m^2$ and a resolution of $13 \times 115\mu m^2$ [54]. In each barrel are 1736 sensor modules with 46080 readout channels/pixels per module giving a total number of 92 million pixels. Its purpose is the reconstruction of charged particle tracks and in particular the reconstruction of primary and secondary vertices for the detection of long-lived particles and in the search for new physics [55]. The pixel detector has a tracking efficiency in Run-2 of 99 % and a spatial resolution of 8 µm in $R - \phi$ and 75 µm in z [55].

Semiconductor Tracker

The Semiconductor Tracker is based on a large area of $60\,m^2$ silicon microstrip sensors enclosing the pixel detector with an overall coverage of $|\eta| < 2.5$ [57, 58]. It provideds a precise tracking at a lower cost per unit area in comparison to the pixel detector [59]. The SCT includes four concentric barrel layers with a set of nine disks at each end-cap that are populated with 4088 modules. Each module contains 1536 readout strips giving it a total sum of 6.2 million read-out channels with a spatial resolution of 17 µm per layer in the $R - \phi$ plane and 580 µm in z direction [59].

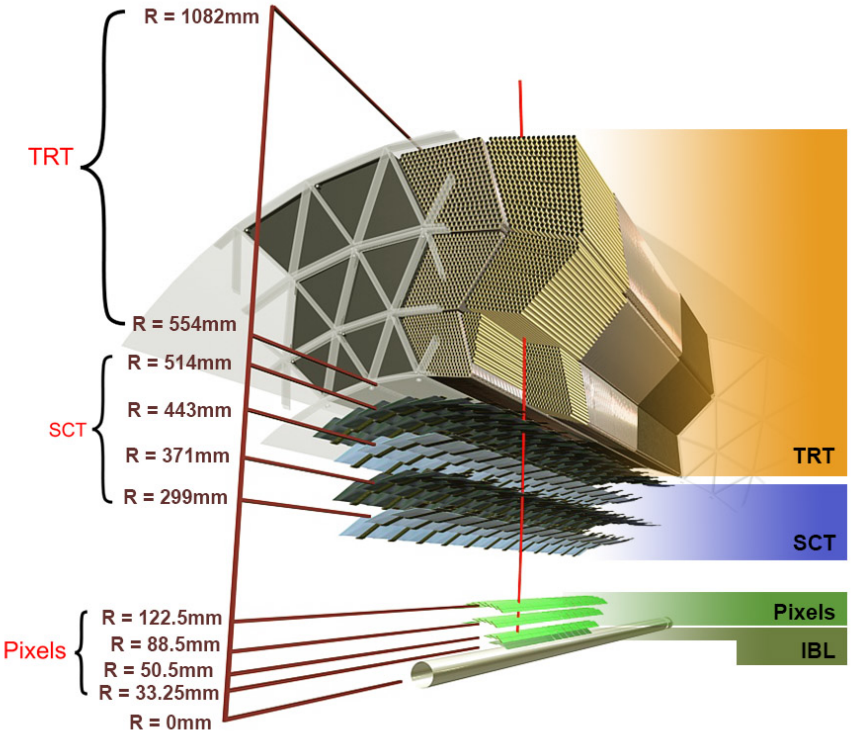


Figure 3.4: A 3D visualisation of the barrel structure and radial distance coverage of the individual parts of the inner detector. The illustration shows the beam pipe, the IBL, the layers of the pixel detector, the four cylindrical layers of the SCT and the tree layers of TRT barrel modules [56].

Transition Radiation Tracker

The outermost and thus final major component of the inner detector is the Transition Radiation Tracker that covers tracks with $|\eta| < 2.0$. The tracker is a gaseous detector (xenon-based and argon-based) that additionally to its tracking capabilities, provides the discrimination of electrons from charged pions based on transiton raditation over the energy range between 1 and 200 GeV [60].

It consists of thin 4 mm diameter gas-filled straw-tubes that are oriented parallel to the beam axis in the barrel region and radially in the end-caps, which are individual drift chambers that offer a high degree of modularity because they can easily be integrated in the detector [61].

Transition radiation occurs, when a relativistic charged particle passes through boundaries between materials of different dielectric constants. In each of the straws, a high-energy particle that passes causes ionization. The ionized electrons get further accelerated torwards the anode, and the remaining ions torwards the straw walls. Fast moving electrons cause an avalanche that implifies a detectable readout signal.

3.2.2 Calorimetry

A collision in ATLAS results in a large number high-energy particles emerging from the interaction point. The purpose of calorimeters in ATLAS is to measure the total energy and momentum of these particles through absorption, more precicely, when a particle passes through the calorimeter material, a particle shower is initiated and can then be measured. They usually consist of layers of passive material interleaved with layers of an active medium. These layers of passive, absorbing materials consist of a high-density material like lead (Pb) while the active medium is mostly lead-glass or liquid argon (LAr) [62]. The passive layers are used to stop and force them to deposit the energy of these particles in the detector and create a signal. Some of them deposit their entire energy already in the beginning, while other particles pass through the material without any interaction like muons or neutrinos.

ATLAS uses an inner electromagnetic calorimeter (ECAL), in particular a high-granularity lead-liquid-argon sampling calorimeter (LAr) for the accurate measurement of electrons, positrons and photons, and an outer tile hadronic calorimeter (HCAL) for the measurement of mostly charged and neutral hadrons. They cover a total rapidity range up to $|\eta| < 4.9$. Calorimeters are important to measure the energy of hadronically decaying tau leptons and are therefore of special interest in the context of this thesis. Both calorimeters are described individually in the following section.

The electromagnetic calorimeter

Electrons, positrons and photons are particles that are subject to the electromagnetic interaction. When high energetic electrons and positrons enter the calorimetry medium, they loose their energy by mainly two processes, ionization or bremsstrahlung while for high-energetic photons the dominant process is the conversion into a pair of e^+e^-. The interaction results then in an electromagnetic shower of particles.

At ATLAS, the ECAL is divided into a barrel ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$), while each end-cap calorimeter is further divided into two coaxial wheels: an outer wheel for the region $1.375 < |\eta| < 2.5$, and an inner wheel for the remaining coverage of $2.5 < |\eta| < 3.2$. The lead absorbers are arranged into an accordion geometry to provide a full azimuthal coverage and are placed into liquid argon. A quantity that defines the energy loss of electrons and photons in matter is called radiation length which is denoted by X_0. It is defined as the mean distance after which an electron has lost its initial energy by a factor of $1/e$ [63]. The design energy resolution in the central region is [64]:

$$\frac{\sigma_E}{E} = \frac{10\,\%}{\sqrt{E}} \oplus 0.7\,\%.$$

The hadron calorimeter

The hadron calorimeter is used to measure the energy of hadrons that interact mainly via the strong interaction but also via electromagnetic interaction. The way charged hadrons loose their energy is mainly through ionization but high energy charged and neutral hadrons can also loose their energy through inelastic hadronic interaction processes with nuclei that result in hadronic showers [40].

Compared to electromagnetic showers, hadronic showers are rather complex because hadrons are produced in jets that have many final states resulting in a large number of tracks. The electromagnetic shower component arises mostly from the decay of neutral pions into two photons $\pi \rightarrow \gamma\gamma$.

In figure 3.2 is shown that the hadron calorimeter is placed directly outside the electromagnetic calorimeter. It consists of three sub-components: the tile calorimeter, the LAr hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCAL). The first one is an iron-scintillating tile calorimeter that uses iron as absorber material and plastic scintillating tiles as active material and it covers a range of $|\eta| < 1.7$. In the second one, the LAr hadronic end-cap calorimeter, copper is used as absorber material embedded into liquid argon which covers a range of $1.5 < |\eta| < 3.2$ [64]. Finally the high-density forward calorimeter consits of copper in the first layers and tungsten for the outter layers to cover a range between $3.1 < |\eta| < 4.9$ [29]. The design energy resolution for the central region of the hadronic calorimeter is [64]:

$$\frac{\sigma_E}{E} = \frac{50\,\%}{\sqrt{E}} \oplus 3.0\,\%. \tag{3.1}$$

3.2.3 Muon Spectrometer

In all the previously described components of the ATLAS detector, muons pass throught it almost unaffected. The amount of absorber material in the calorimeters make sure, that only muons and neutrinos penetrate through the detector layers to reach the outermost component, the muon spectrometer (MS). Its key feature is to measure the trajectories of muons in order to determine their direction, momentum and charge with high accuracy. To measure the muon momenta, the muon spectrometer is equipped with a toroidal magnetic field that forces particles to travel on curved tracks. In terms of size, the muon spectrometer takes up more than half of the Atlas detector volume and surrounds the calorimeters.

The muon spectrometer consists of four different chambers for tracking and for triggering with accurate time-resolution. For precision tracking, the muon spectrometer uses Drift Tubes (MDT) and Cathode Strip Chambers (CSC) for a region out to $|\eta| = 2.7$, for triggering it uses thin Gap Chambers (TGC) and Resistive Plate Chambers (RPC) in the region $|\eta| < 2.4$ [65, 66].

3.2.4 Magnet System

The ATLAS detector uses a huge magnet system to bend the path of charged particles to allow a good momentum resolution. It is composed of a thin superconducting solenoid in the inner part and three large superconducting toroids, the Barrel Toroid and End-cap Toroids, in the outer region [67].

3.2.5 Trigger System

Due to the enormous number of particle reactions, the need for a trigger system is essential. The ATLAS Trigger and Data Acquisition System (TDAQ) is an important component because it decides in real time whether data from a collision process should be recorded or not [68]. It takes the 40 MHz collision rate and shrinks it to output recorded events at a rate of 1 kHz [69]. For that, selections can be applied to only select the very rare processes of main interest.

The trigger system at ATLAS in Run-2 consists two major components, a hardware-based first level trigger (Level-1) and a software-based high level trigger (HLT) [70]. Level-1 reduces the event rate to

100 kHz by a decision time accept of 2.5 µs [70]. It makes use of the input signals received from the muon (L1Muon) and calorimeter systems (L1Calo) to evaluate threshold and topological selection criteria [69]. The HLT applies software-based decision making to further reduce the event rate from 100 kHz to an output event rate of 1 kHz. Furthermore, it uses a powerful army of CPU-intensive algorithms that are executed on a computing farm known as Processing Units (PUs). During the years, the hardware has been replaced and optimized by newer hardware to increase the computing power [68].

3.3 Reconstruction of Physics Objects at ATLAS

Different particles that emerge from the proton-proton collision process traverse through the detector and leave different signatures. An overview of typical signatures is given in figure 3.5. In this section, these signatures and their reconstruction will be discussed and in addition to that, a detailed look into the reconstruction of hadronically decaying tau leptons is given, as they are the most relevant part for the study of this thesis.

3.3.1 Electrons and Photons

Electrons that pass torwards the detector can lose a significant amount of their energy in form of photons in the electromagnetic calorimeter due to bremsstrahlung when it interacts with the material it traverses [71]. They leave signal in the inner detector and deposit their energy in the electromagnetic calorimeter. Photons resulting from the bremsstrahlung process may convert into electron-positron pairs, known as pair production, that induce further interactions with the detector material. Since electrons are charged particles, their trajectories are bend due to the magnetic field. From the curvature of the trajectory of electrons and positrons, one can determine the momentum and the sign of charge of these particles.

Photons in contrast are particles that are neutral in charge and leave therefore no signal in the detector. However, they can interact with the detector material and produce electron-positron pairs, which is called photon conversion. The electromagnetic calorimeter measures the energy of both, the electrons and photons, by stopping them in a cascade of secondary electrons and photons which results in an electromagnetic shower [72].

The reconstruction of electron candidates is done within the region $|\eta| < 2.47$ and is based on localised clusters of energy deposits in the electromagnetic calorimeter, charged-particle tracks identified in the inner detector and close matching in $\eta \times \phi$ space of the tracks to the clusters [71]. ATLAS uses a sliding window algorithm to find clusters within the $\eta \times \phi$ grid of the electromagnetic calorimeter. For that, the cluster algorithm scans different areas in the electromagnetic calorimeter that contain a local maximum of energy in a search window of fixed size. The search for clusters is performed for every element in the calorimeter [71]. When a cluster is found, one analyses the track that is pointing to it. In case of an electron or positron, a cluster is associated to a reconstructed track in the inner detector, in contrast to photons, that leave no track in the inner detector. For two relatively close cluster candidates that are found in close proximity, the one with the higher transverse energy is retained [71]. The efficiency for the reconstruction is greater than 99 % for $E_T > 15$ GeV [71].

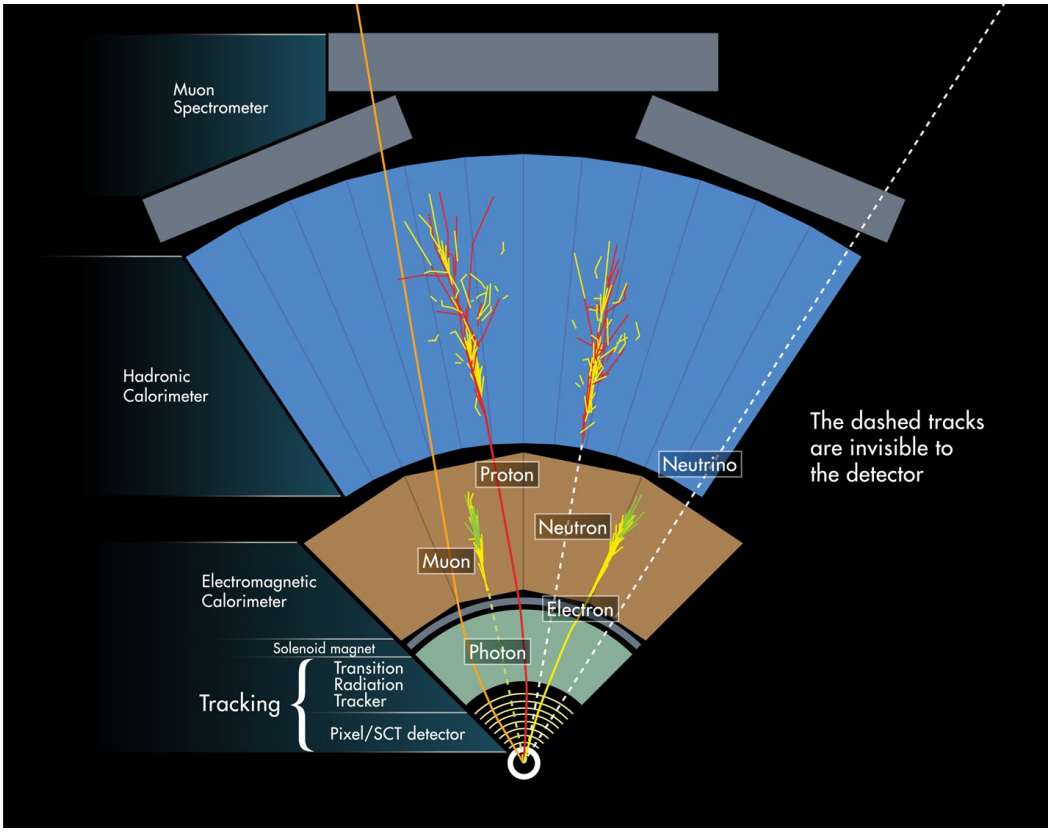


Figure 3.5: Overview of typical signatures from particles that pass through the different components of the ATLAS detector [73].

3.3.2 Muons

Muons that pass through the detector leave signatures in the inner detector, small energy deposits in the calorimeters and tracks in the muon spectrometer. To form muon tracks, the combined information from individual subdetectors is taken into account [74]. The reconstruction of muons is performed individually for the inner detector and the muon spectrometer.

Beginning with the muon spectrometer, hit patterns are searched inside each muon chamber locally to form segments [74]. A global fit is then performed to build a muon track candidate by combining track segments from different layers. Segments form outer chambers are combined with inner layer segments of the spectrometer. A track fit is then applied to combine these segments to build a track candidate. This process is repeated until enough information is available for the final track in the muon spectrometer [75].

For the reconstruction of the complete muon track, the tracks from the inner detector and muon spectrometer are combined. The reconstruction uses different algorithms based on the information that is provided by the inner detector, muon spectrometer and the calorimeters. The measured efficiency for reconstructed muon tracks is measured to be close to 99 % for most of the covered phase space ($|\eta| < 2.5$) for $5 < p_T < 100$ GeV [74].

3.3.3 Jets

Jets are the signatures of quarks and gluons resulting from a high-energy process such as head-on proton-proton collisions [76]. The reason why quarks and gluons cannot be observed as free particles is due to colour-confinement (as already mentioned in section 2.1).

After a collision process, the emerging quarks and gluons combine with other spontaneously created particles from the vacuum to form hadronic bound states. These bound states are observed in the ATLAS detector as collimated particle steams that deposit energy in the calorimeters which are reconstructed as *jets* [77]. For reconstruction, ATLAS used a sequential anti-k_t algorithm [78]. Different distance parameter values R are used for the type of reconstructed jet. Generally speaking, this parameter refers to the size of the jet. Jets originating from quarks and gluons are called small-R jets and are reconstructed with distance parameter value of $R = 0.4$. Hadronic jets, referred as large-R jets are reconstructed with $R = 1.0$ due to the fact that the radius has to be larger for a wider capture of the hadronic decay products. In this thesis, a distance parameter of $R = 0.4$ is chosen.

At the first stage, the anti-k_t algorithm searches for three-dimensional energy clusters of topologically neighbouring calorimeter cells, which are referred to *topo-clusters* [79]. The distance between two objects i and j is computed via:

$$d_{ij} = \min\left(k_{ti}^{2p}, k_{tj}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2},$$

where k_{ti} and k_{tj} are the transverse momenta of the consitutents. The parameter p denotes the relative power of the energy versus geometrical distance [78]. In addition, the distance between an object and the beam is further defined as

$$d_{iB} = k_{ti}^{2p}.$$

For all particles in an event, d_{ij} and d_{iB} are being calculated and stored in a list of distance values $\{d_{ij}, d_{iB}\}$. If the smallest list entry and thus the minimum is d_{ij}, then the two objects i and j are merged into a single object by summing their four-momenta. Otherwise, if d_{iB} is the smallest list entry, the object i is regarded as a complete jet and the entry get removed from the list. This procedure is repeated until there is no element left in the list. Further information related to the jet reconstruction using the anti-k_t algorithm are described in [78].

3.3.4 b-jets

When labelling jets, there is a certain convention regarding the hadron that is participating in the jet. Jets that originate from bottom quarks are called b-jets, those that contain a c-hadron are called c-jets and those that contain neither b nor c hadrons are called light flavour jets.

As already mentioned in section 2.4.6, top-quarks predominantly decay into a W boson and a b quark. Identifying jets originating from b quarks therefore play a crucial role to reduce the background by accurately selecting pure top-quark signal samples but also for a large variety of other processes that include high-p_T b-jets in the final state [80, 81].

The identification of b-jets ("tagging") relies upon the properties of b hadrons: the long lifetime, large mass and decay multiplicity. Due to their long lifetime, B hadrons travel significant distances from the point where they originated, the so-called primary vertex. Subsequently, a large fraction of b

hadrons decay semi-leptonically into lighter hadrons, resulting in a displaced secondary vertex with respect to the primary vertex [81]. At ATLAS the identification of b-jets is based on three algorithms: the impact parameter based algorithms (IP2D and IP3D), an inclusive secondary vertex reconstruction algorithm (SV) and a decay chain multi-vertex reconstruction algorithm (JetFitter) [82]. The results of these different algorithms are combined in a multivariate discriminant (MV2).

3.3.5 Hadronically decaying τ leptons

As described in the pevious chapter, the decay signatures for leptonically decaying tau leptons can not be distinguished experimentally from prompt light electrons or muons, therefore only the identification and reconstruction of hadronically decaying tau leptons is considered and described in the following.

The majority of taus in the hadronic mode decay into one or three charged mesons. These are predominantely pions or less frequently kaons, accompanied by neutral pions. Apart from that, a neutrino is produced along with every decaying tau lepton, which leaves no trace and remains undetected, leaving only the hadrons to be used for reconstruction and identification [83].

Hadronic tau lepton decays can be classified according to their number of reconstructed charged decay particles as 1- or 3-*prong*, denoted by the charged meson in its decay [41]. While the phase space for the decay into one charged particle is much larger than the decay into three particles, the decay into one charged particle is more likely and occurs 72% of the time while the latter process occurs 23% of the time. Higher prong modes, like the 5-*prong* decay, have also been observed but their branching ratios are orders of magnitude smaller, so they are not taken into ATLAS [83].

The final state particles of hadronically decaying tau leptons form a cone of well collimated particle jets in the detector that are used for reconstruction. Figure 3.6 shows an example for a cone where a tau decays into three charged pions, while figure 3.7 shows a typical signature for a QCD event that results in the same cone signature.

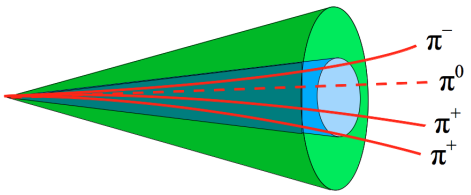


Figure 3.6: Illustration of a hadronically decaying tau lepton resulting in a jet of three charged pions (3-*prong* decay) and a neutral pion [84].

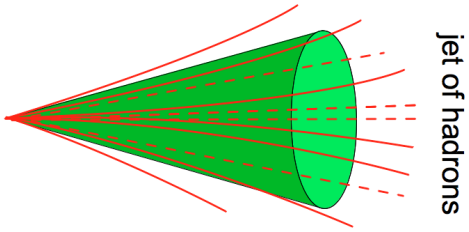


Figure 3.7: Signature of a typical QCD event resulting in a jet of hadrons [84].

The reconstruction takes advantage of the special decay characteristics of the tau leptons and combines the track and calorimeter information obtained from the detector. Possible jet candidates are selected by using the anti-k_t algorithm with a distance parameter of $R = 0.4$ and calibrated TopoClusters as input [85]. The jets are used as seeds for the reconstruction algorithm, if they satisfy $p_T > 10\,\text{GeV}$ and cover the region $|\eta| < 2.4$ [86].

In an event with multiple simultaneous interactions, the primary vertex does not always correspond to the vertex at which the tau lepton is produced. To identify the correct primary vertex, the tau vertex (TV) algorithm takes as input all tau candidate tracks in the core region $\Delta R < 0.2$ around

the direction of the hadronic tau and sums up the p_T of these tracks. The vertex with the highest momentum fraction p_T is selected to be the tau vertex [86]. Additionally, tracks that are associated with the hadronic tau candidate are required to satisfy $p_T > 1\,\text{GeV}$ [86]. This selection maximizes the fraction for 1 and 3-prong tau decays with the correct number of tracks.

Identification of τ_{had} using BDT

For the identification of hadronic taus, trained Boosted Decision Trees (BDT) algorithms, separate for 1 and 3-track hadronic tau decays, are used to discriminate tau lepton decays from hadronic jets. These algorithms were trained using the samples of $Z/\gamma^* \rightarrow \tau\tau$ for signal and di-jet events for background. The discriminating variables for the BDT approach are given in [86] and are listed below.

- **Central energy fraction (f_{cent}):** This fraction refers to the calorimeter transverse energy deposited within a cone of $\Delta R < 0.1$ with respect to the total energy deposited in the region $\Delta R < 0.2$ around the hadronic tau candidate. It is calculated by summing up the energy of all TopoCluster cells that have a barycentre that lies in the respective cone.

- **Leading track momentum fraction ($f^{-1}_{\text{leadtrack}}$):** This variable is defined as the transverse energy sum, deposited in all TopoCluster cells, that lies in the core region of the hadronic tau candidate, divided by the transverse momentum of the hightest p_T charged particle in the core region.

- **Track radius ($R^{0.2}_{\text{track}}$):** The radius refers to the p_T-weighted ΔR distance of the associated tracks to the hadronic tau direction. Only tracks in the core region are used.

- **Leading track IP significance ($|S_{\text{leadtrack}}|$):** This variable defines the absolute value of the transverse impact parameter of the highest p_T track in the core region, calculated with respect to the tau vertex and divided by its estimated uncertainty.

- **Fraction of tracks p_T in the isolation region ($f^{\text{track}}_{\text{iso}}$):** This variable divides the scalar sum of the p_T of tracks associated with the hadronic candidate in the region $0.2 < \Delta R < 0.4$ by the sum of the p_T of all tracks associated with the hadronic tau candidate.

- **Maximum ΔR (ΔR_{Max}):** This radius defines the maximum ΔR between a track associated with the hadronic tau candidate and the direction of the hadronic tau. This variable only uses tracks in the core region.

- **Transverse flight path significance ($S^{\text{flight}}_{\text{T}}$):** This is defined as the decay length of the secondary vertex in the transverse plane, calculated with respect to the tau vertex and divided by its uncertainty.

- **Track mass (m_{track}):** This variable is the invariant mass that is derived from the sum of the four-momentum of all tracks in the core and isolation region.

- **Fraction of EM energy from charged pions ($f_{\text{EM}}^{\text{track-HAD}}$):** This variable refers to the fraction of the electromagnetic energy of tracks associated with the hadronic tau candidate in the core region. The numerator is defined as the difference between the sum of the momentum of tracks in the core region and the sum of the cluster energy deposited in the hadronic part of each TopoCluster, while the denominator states the sum of of cluster energy deposited in the electromagnetic part of the TopoCluster, both associated with the hadronic tau candidate.

- **Ratio of EM energy to track momentum ($f_{\text{track}}^{\text{EM}}$):** This ratio defines the sum of cluster energy deposited in the electromagnetic part of each TopoCluster with the hadronic tau candidate to the sum of the momentum of tracks in the core region.

- **Track-plus-EM-system mass ($m_{\text{EM+track}}$):** Defines the invariant mass of all tracks and up to two most energetic electromagnetic clusters in the core region, where the electromagnetic cluster energy is part of the TopoCluster energy, deposited in the electromagnetic calorimeter and the four-momentum of an electromagnetic cluster is calculated assuming zero mass and using TopoCluster seed direction.

- **Ratio of track-plus-EM-system to p_T ($p_T^{\text{EM+track}}/p_T$):** This variable is the ratio of hadronic tau p_T. It is estimated using the vector sum of track momenta and up to two most energetic electromagnetic clusters in the core region to the calorimeter-only measurement of the hadronic tau p_T.

Figure 3.10 shows an example for the 3-prong distribution for two picked tau lepton identification variables from above, where the red circles represent the signal (tau leptons) and the black squares the background (multi-jet background) [86].

The efficiency of the tau identification algorithm is defined as the product of the reconstruction efficiency and the identification efficiency. Three working points are defined, based on the value of the overall reconstruction and identification efficiency and are labeled as: *Tight*, *Medium* and *Loose*. The values for 1-prong are 0.6, 0.55 and 0.45 for Loose, Medium and Tight and for 3-prong 0.5, 0.4 and 0.3 respectively [86].

Novel approach to identify τ_{had} using RNN

A novel algorithm uses track and calorimeter information in a recurrent neural network (RNN) to distinguish true hadronic taus from fake taus. The previously mentioned algorithm using trained BDTs was used to discriminate jets from hadronic taus [88]. For the RNN approach, a network is trained with MC samples (chapter 4 for more information) with the signal sample consisting of $\gamma^* \rightarrow \tau\tau$ events, while dijet events are used as backgrounds. Compared to the BDT approach, RNN is now superseded and provides a largely improved jet rejection [88]. The discriminating variables for the RNN approach are given in [88] and are listed below:

- **Cluster depth (λ cluster):** This variable refers to the cluster barycentre from the calorimeter front face and it is determined along the cluster axis.

- **Longitudinal cluster extension ($\langle\lambda^2\rangle$):** This variable determines the second moment of the longitudinal distance of cluster cells from the cluster barycentre, along the cluster axis.

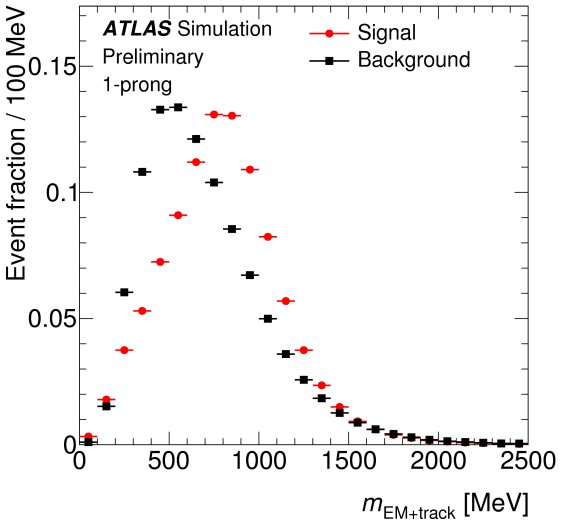


Figure 3.8: (a) 1-prong

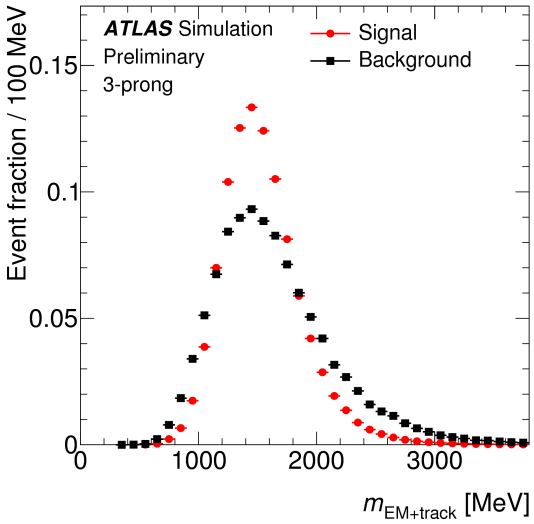


Figure 3.9: (b) 3-prong

Figure 3.10: Example distribution for for one discriminating variable $m_{\text{EM+track}}$ used for the BDT tau identification approach: (a) shows the distribution for 1-prong, (b) the distribution for 3-prong. The red circles indicate the signal from tau leptons and the black squares show the multi-jet background [87].

- **Radial cluster extension ($\langle r^2 \rangle$):** Determines the second moment of the radial distance r of cluster cells from the cluster axis.
- **Transverse momentum of the seed jet ($p_T^{\text{seed jet}}$):** This variable measures the transverse momentum of the orignal seed jet.
- **Transverse momentum at the LC scale (p_T^{uncalib}):** Measured p_T of the hadronic tau candidate at LC scale prior to the tau energy scale calibration. It is calculated from the sum of cluster 4-momenta within $\Delta R < 0.2$ of the hadronic tau candidate. The origin of the cluster is set to the tau production vertex and energies are calibrated at the LC scale.
- **Central energy fraction (f_{cent}):** This fraction refers to the calorimeter transverse energy deposited within a cone of $\Delta R < 0.1$ with respect to all energy deposited in $\Delta R < 0.2$ around the hadronic tau candidate axis. It is calculated by summing up the energy in all cells with a barycentre in these regions. Cells are required to be part of a TopoCluster that belongs to the hadronic tau candidate.
- **Inverse momentum fraction of the leading track ($f_{\text{leadtrack}}^{-1}$):** Defines the transverse energy sum and is calibrated at the EM energy scale, deposited in all cells belonging to TopoClusters int he core region. It is then divided by the transverse momentum of the highest p_T core track of the hadronic tau candidate.
- **Maximum ΔR (ΔR_{Max}):** This radius defines the maximum ΔR between a track associated with the hadronic tau candidate and the direction of the hadronic tau. This variable only uses tracks in the core region.

- **Transverse impact parameter significance of the leading track ($|S_{\text{leadtrack}}|$):** This variable refers to the transverse impact parameter of the highest-p_T core track of the hadronic tau candidate with respect to the tau vertex.
- **Transverse flight path significance ($S_{\text{T}}^{\text{flight}}$):** Measures the decay length of the secondary vertex in the transverse plane with respect to the tau vertex.
- **Momentum fraction of isolation tracks ($f_{\text{iso}}^{\text{track}}$):** Is defined as the scalar sum of the p_T of isolation tracks that are associated with the hadronic tau candidate and its divided by the sum of the p_T of all the associated core and isolation tracks.
- **Ratio of EM energy and track momentum ($f_{\text{track}}^{\text{EM}}$):** Takes the ratio of the sum of cluster energy deposited in the electromagnetic part of the TopoClusters to the sum of the momentum of core tracks. The clusters are calibrated at the LC energy scale.
- **Fraction of track-plus-EM-system ($m_{\text{EM+track}}$):** Measures the ratio of the hadronic tau momentum p_T to the calorimeter-only measurement of the hadronic tau p_T. The hadronic tau momentum p_T is estimated using the vector sum of core track momenta and additionally up to two most energetic EM clusters in the core region.
- **Mass of the track-plus-EM-system ($m_{\text{EM+track}}$):** Provides the invariant mass of the system composed of the core tracks and up to two most energetic EM clusters in the core region. The EM cluster energy is in this case part of the TopoCluster energy deposited in the presampler and first two layers of the LAr calorimeter, and the four-momentum of an EM cluster is calculated assuming zero mass and using TopoCluster seed direction.
- **Mass of the track system (m^{track}):** This variable refers to the invariant mass that is calculated from the sum of the four-momenta of all core and isolation tracks. For each track, a pion mass is assumed.

The RNN Score is defined as the fraction of rejected true hadronic taus [88]. For physics analysis, four working points with increasing background rejection are defined as a function of the hadronic tau p_T and μ: *Very loose*, *Loose*, *Medium* and *Tight* [88]. Here μ refers to the average number of pile-up interactions per bunch crossing. Table 3.1 shows the defined working points with the corresponding signal selection efficiencies and rejection powers [88].

	Signal efficiency		**Background rejection RNN**	
Working point	1-prong	3-prong	1-prong	3-prong
Tight	60%	45%	70	700
Medium	75%	60%	35	240
Loose	85%	75%	21	90
Very loose	95%	95%	9.9	16

Table 3.1: Summary of the defined working points with fixed true hadronic tau selection efficiencies and the corresponding background rejection factors for misidentified taus in dijet events [88].

The RNN approach is used in the analysis part of this thesis, in particular the RNNScore with the *Loose* working point is used to create mixed samples for a weak supervised neural network (CWoLa) that classifies true hadronic taus from fake taus on top of the current recurrent neural network approach.

3.3.6 Missing Transverse Energy

The previously mentioned electromagnetic and hadronic calorimeters provide an excellent energy resolution for most of the particles that were produced during the proton-proton scattering process [89]. However, there are some produced particles that leave the detector system without signatures. Particles that fall under this category are most notably neutrinos and weakly interacting non-SM particles, such as the lightest supersymmetric particle (LSP). Their presence becomes apparent by measuring an imbalance of the total missing transverse momentum.

The missing transverse energy E_T^{miss} quantifies the magnitude of the negative vector sum of the transverse momenta sum of all visible reconstructed particles produced in the collision event and is defined as [90, 91]:

$$\sum E_T^{\text{miss}} = -\sum p_T^e - \sum p_T^\gamma - \sum p_T^{\tau_{\text{had}}} - \sum p_T^\mu - \sum p_T^{\text{jets}} - \sum p_T^{\text{soft}}.$$

For reconstruction, the missing transverse momentum consists of two contributions: from hard-event and soft-event signals. The hard-event signals are characterized by the fully reconstructed calibrated particles and jets. These particles are electrons, muons, photons and hadronic tau-leptons. The latter contribution from soft-event signals consists of reconstructed charged-particle ID tracks, associated with the hard-scatter vertex, that are not identified to any hard object [91].

CHAPTER 4

Event selection

This chapter gives an overview of the tHq process and describes the decay topologies of this process. Furthermore, the chapter describes the event selection for the channel of interest to understand the underlying analysis presented in this thesis. In addition to that, relevant backgrounds containing fake taus are discussed. The chapter closes with a summary of the data and MC samples that are used for the neural networks in the chapter 6 and chapter 7 to differentiate between true and fake taus.

4.1 tHq channel

At tree level, this process consists of three particles: a top quark in association with a Higgs boson and an additional quark jet. In this rare top quark production process, the Higgs boson is in most cases radiated from the exchanged W boson or from the top quark. Based on the decay channel of the Higgs boson, there is a various number of possible final states. In general, tHq channel is a very complex channel with a large multiplicity of objects in the final state.

In particular, the dominant decay modes of the Higgs boson are: the decay into two a pair of W or Z bosons ($H \rightarrow WW/ZZ$), the decay into two b-quarks ($H \rightarrow b\bar{b}$) as well as some multi-lepton processes like the decay into taus ($H \rightarrow \tau\tau$) or two c-quarks ($H \rightarrow c\bar{c}$) [7, 35]. Other less dominant decay modes with a smaller rate are: the Higgs decaying into a pair of photons ($H \rightarrow \gamma\gamma$), the decay into a photon and a Z boson ($H \rightarrow \gamma Z$) and the decay into a pair of muons ($H \rightarrow \mu^{+}\mu^{-}$) [7, 35].

The CMS collaboration has already published searches for the production of a Higgs boson in association with a single top quark in pp collision data at a center-of-mass energy of 8 TeV that corresponds to an integrated luminosity of 19.7 fb^{-1} [35]. Different decay modes have been analyzed, in particular: $H \rightarrow \gamma\gamma$, $H \rightarrow b\bar{b}$, $H \rightarrow WW$ and $H \rightarrow \tau\tau$ [35].

For this analysis, the multilepton channel and in particular the decay $H \rightarrow \tau\tau$ is of special interest. This channel consists of three combinations of possible final states for the Higgs boson. In the first case, both tau leptons can decay leptonically $H \rightarrow \tau_{\text{lep}}\tau_{\text{lep}}$. This channel is also called lep-lep channel and from all three possible decays, this is the one with the lowest branching ratio. Another possible decay mode is the one where both tau leptons decay hadronically $H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$, referred to the had-had channel. In the third case, one of the taus can decay leptonically and the other hadronically $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$. This decay channel is called the lep-had channel and is the one with the highest branching ratio compared to the other two channels.

4.2 $1e/\mu + 1\tau_{\text{had}}$ event selection

For this analysis, the $1e/\mu + 1\tau_{\text{had}}$ channel is of special interest because it has one hadronically decaying tau lepton in the final state. The final state of the $1e/\mu + 1\tau_{\text{had}}$ consists of a total number of two leptons: one light leptons and one hadronically decaying tau lepton, fulfilling the following p_T requirement: $p_T(l) > 10\,\text{GeV}$ and $p_T(\tau_{\text{had}}) > 20\,\text{GeV}$ with $|\eta| < 2.5$. The number of jets is supposed to be $2-6$ jets with $p_T > 25\,\text{GeV}$ and $|\eta| < 4.5$, while events larger or lower values compared to this jet multiplicity range are rejected. In addition, the b-jet requirement is set to have $0-2$ b-jets in the final state with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. Table 4.1 gives an overview of the applied selections.

	Selection
jets	2-6 jets $p_T > 25\,\text{GeV}$, $\lvert\eta\rvert < 4.5$
b-jets	0-2 b-jets $p_T > 25\,\text{GeV}$, $\lvert\eta\rvert < 2.5$

Table 4.1: Overview of the event selection for the $1e/\mu + 1\tau_{\text{had}}$ channel.

4.3 Backgrounds

Background processes are processes that have a similar decay characteristic compared to the signal process of interest. There are several backgrounds that contribute to the channel of interest for this analysis. These backgrounds can be divided into backgrounds containing non-prompt or fake leptons and backgrounds containing prompt leptons.

4.3.1 Backgrounds containing non-prompt or fake leptons

$t\bar{t}$

The $t\bar{t}$ process is one of the largest backgrounds contributing to the study of this thesis. In this decay the top quarks decay into a pair of W-bosons and b-quarks. This background process can contain a fake tau, which is falsely reconstructed as a hadronically decaying tau lepton, since the branching ratio for W-bosons decaying into jets or electrons is higher than that of hadronically decaying tau leptons. Because of the rich proportions of fake taus in this process, it is used as training samples for the neural networks implemented in this thesis. The Feynman diagram for this process is shown in figure 4.1.

Z+jets

In this vector-boson production process, the Z-boson decays into a pair of leptons. The LO Feynman diagram for this process is illustrated in figure 4.2. The W+jets process looks similar to the Z+jets process and is also a large background containing fake leptons.

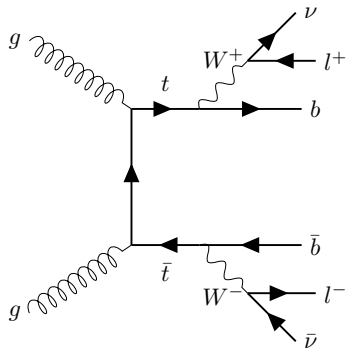


Figure 4.1: LO Feynman diagramm for the $t\bar{t}$ process.

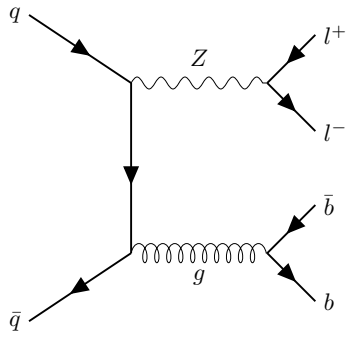


Figure 4.2: LO Feynman diagramm for the Z+jets process with additional b-jets.

4.3.2 Backgrounds containing prompt leptons

Diboson

As the name suggests, diboson refers to the process, where two bosons are produced: WW, WZ or ZZ. The LO Feynman diagrams for the WZ process is given in figure 4.3, the process for the ZZ process in figure 4.4.

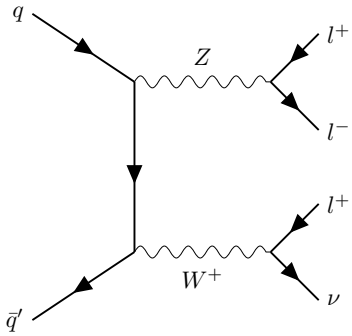


Figure 4.3: LO Feynman diagramm for the diboson WZ process.

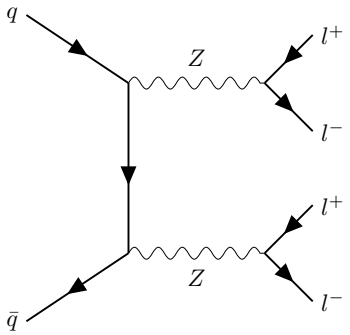


Figure 4.4: LO Feynman diagramm for the diboson ZZ process.

4.4 Datasets

After the discussion of the background processes, a short overview of the samples used for the neural network part of the analysis is given in this section. These samples consists of data samples and Monte Carlo simulated samples (MC).

4.4.1 Data Samples

The data samples that are used in this thesis were recorded from proton-proton collisions at the LHC by the ATLAS detector at $\sqrt{s} = 13$ TeV. The data cover a four-year experimental time range starting from 2015 to the end of 2018. These data were taken at a total integrated luminosity of 139 fb^{-1} and were collected during stable beam LHC operations and with a fully functioning ATLAS detector [30]. Table 4.2 shows an overview of the data samples and their corresponding partial integrated luminosities. The data samples that were used in this thesis are shown in section A.1.1 in the appendix.

Year	Int. Lumi. (fb^{-1})
2015	3.2
2016	33.0
2017	44.3
2018	58.5

Table 4.2: Overview of the integrated luminosities of LHC Run 2 for different years of operation [30].

4.4.2 Monte Carlo Simulation Samples

Monte Carlo simulation events are used to simulate typical collision events in the detector to estimate the expected output. The use of MC generators offers the possibility to simulate the final states of the particles of a collision event by taking into account known physical properties and distributions of the underlying process. These collision processes are generated using stochastic methods.

For the generation of a pp collision event several steps have to be taken into account. A full description of how these MC samples are generated can be found in [29, 92]. After MC generation, the kinematic information for all final state particles is available, which is referred to as the *truth*

information. Up to this point, no detector effects have been taken into account. The use of the software GEANT4 [93] therefore simulates the full detector response. The MC samples that simulate the Run 2 conditions are divided into three campaigns: MC16a, MC16d and MC16e. Each campaigns is associated with a different pile-up condition. MC16a is associated with the pile-up conditions for 2015-2016, MC16d for 2017 and MC16e for 2018. In all generated MC samples, the top mass was fixed to $m_t = 172.5\,\text{GeV}$. Details about the MC samples that were used in this analysis are given in section A.1.2 in the appendix.

Monte Carlo Reweighting

Using simulated data to describe physical processes has the advantage of better understanding the underlying processes and looking at what possible outcomes can be expected. However, for some of these processes, reweighting needs to be done to accurately estimate the correct data samples because these simulations are not always perfect.

There are two procedures for reweighting that will be applied simultaneously: event-by-event rewriting and luminosity weighting. Both will be described in this section. For event-by-event reweighting, each MC simulated event is multiplied by an event-by-event factor ω_{event} (ω denotes the *weight*). The reweighting consists of several terms which can be written as [29]:

$$\omega_{event} = \omega_{MC} \times \omega_{\text{pile-up}} \times \omega_{\text{lepton}} \times \omega_{\text{JVT}} \times \omega_{\text{trigger}} \times \omega_{\text{b-tagging}}.$$

The total weight can therefore be divided in the product of individual weights. These individual weights are described as [29]:

- $\omega_{\boldsymbol{MC}}$: this factor stands for the MC event weight which is applied to estimate the correct number of events in a MC sample.
- $\omega_{\textbf{pile-up}}$: is defined as the MC pile-up weight that has to be taken into account to match the pile-up conditions of data.
- $\omega_{\textbf{lepton}}$: this term is defined as the efficiency of lepton identification, reconstruction and for lepton isolation criteria.
- $\omega_{\textbf{JVT}}$: this factor is taken into account when applying a cut on the output of the JVT discriminant for differences in data and MC efficiencies.
- $\omega_{\textbf{trigger}}$: this factor corrects the trigger selection efficiencies in data and MC.
- $\omega_{\textbf{b-tagging}}$: this factor is applied for differences in data and MC efficiencies due to b-tagged jets and b-tagging efficiency.

The total weight (ω_{total}) has to be matched to the correct luminosity. This is referred to luminosity weighting. Here the event weight is multiplied with a luminosity weight (ω_{lumi}), which is defined as [29]:

$$\omega_{\text{lumi}} = \sigma_{\text{process}} \mathcal{L} / N_0,$$

where $\mathcal{L}$ denotes the luminosity of the data sample, σ_{process} the cross-section for the physics process and N_0 is defined as the number of events from the original MC sample. Then the total weight for each MC event can be written as:

$$\omega_{\text{event}} = \omega_{\text{event}} \cdot \omega_{\text{lumi}}.$$

CHAPTER 5

Neural Networks

All you need is lots and lots of data and lots of information about what the right answer is, and you'll be able to train a big neural net to do what you want.

- Geoffrey Hinton,

In a conventional way in high energy physics, physicists are searching for the evidence of new particles using the bump-hunting-technique. New particles show up as clear signatures called bumps in the invariant mass spectra over a smooth background distribution. A prominent example for the bump-hunting-technique was the historic discovery of the higgs boson in 2012 at the LHC by the ATLAS and CMS experiments.

The higgs boson was observed from the decay into two photons. By measuring their energies and the direction of motion, one can compute their invariant mass. The particle that decays into two photons will show up as a bump in the invariant mass plot sitting on a rather smooth background from other processes that also produce two photons. By taking more data, the bump gets clearer and shows the signature for the higgs boson.

The rise of deep learning, a special field of machine learning, that is based on artificial neural networks, has contributed to bring new levels of performance to a wide spectrum of complex and high dimensional data analysis applications at the LHC. However, all these approaches rely on significant input from simulations because data collected at the LHC is fully unlabeled. The use of simulations from labeled inputs leads to the cost of mis-modelling and has to be adjusted afterwards by data-driven techniques [94]. Instead of training neural networks on simulations, weakly supervised approaches can be used to train directly on data. The method can be used for a variety of collider physics applications like the bump-hunting-technique but also in the scope of this thesis as a powerful tool to distinguish hadronically decaying taus from taus that are mis-identified.

This chapter provides a general overview of neural networks and gives a deeper insight into a weak supervision approach that is used to distinguish true hadronic taus from fake taus.

5.1 Core Concepts in Machine Learning

In contrast to the conventional way of programming, where the computer is feed with defined instructions how to perform a task, the machine learning approach gives computers the ability to

perform tasks without being explicitly programmed for that [95]. A more modern definition of how computers can learn is given by Tom M. Mitchell (1999):

> *'A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.'* (Tom M. Mitchell, 1999)

There are several types of how machine learning systems can learn tasks, which can be differentiated into the following main categories:

- **Supervised Learning**: where the dataset consists of a sequence of input-output data pairs $\{(x_i, y_i)\}_{i=1}^{N}$ to train the model. Here, the notation is used that x_i represents the inputs, also called *features*, and y_i the labeled output corresponding to x_i [96]. For binary classification, each training example comes with a label $y_i \in \{S, B\}$, where S denotes signal and B background, while possible features for $\vec{x}$ might be the measured calorimeter energy or the measured momentum of the particles. The goal of the supervised learning algorithm is to look for a function, that maps the inputs to the output and predicts the label for a set of newly presented input data.

- **Unsupervised Learning**: where the dataset consist of a collection of unlabeled input examples $\{x_i\}_{i=1}^{N}$ without an additional label attached to them. Since there is no label given to the training data, the goal of the unsupervised learning algorithm is to find hidden patterns in the data. One method, that is widely used in solving an unsupervised learning problem is to use *clustering*. Clustering allows to assign similar data, that has not been labeled, into groups.

- **Weak Supervised Learning**: combines the tasks of supervised and unsupervised learning, where the dataset consists of mixtures of both, labeled and usually a smaller set of unlabeled examples. The goal of the algorithm is similar to the goal of the supervised learning algorithm [96].

- **Reinforcement Learning**: where an agent interacts with a dynamic environment and performs actions. The actions of the agent receive feedback, that can either be rewarded or punished, and the system has to find its own strategy to maximize the expected average reward [96].

5.2 Deep Learning

Deep learning is a sub-field of machine learning that uses algorithms inspired by the function of the human brain. The way how humans learn and process information is done by neurons. These neurons are connected by synapses creating a massive mulitilayer network to solve our daily tasks. In analogy to the biological concept of neural networks, deep learning is based on artificial neural networks (ANNs), or just neural networks, that are computer models, used to solve complex mathematical tasks the same way information is processed in our brain. The architecture of a neural network consists of a large number of interconnected artificial neurons, that are typically organized into various layers and pass data through it.

5.2.1 The Perceptron Model

One of the simplest neural network architectures is the perceptron model which can be seen as the building block of a neural network. It describes the simple model of one single neuron, also called perceptron or node, in an artificial neural network. The concept was initially presented by the scientist Franck Rosenblatt in the late 1950s. It is a linear binary classification algorithm that is typically used for supervised learning and only consists of two layers, an input layer and an output layer, as shown in figure 5.1. Binary classification states, that this classifier only distinguishes between two categories by using a linear function of the inputs and can only deal with data that is linearly separable.

The core idea behind the perceptron model is, that in the initial step, the perceptron takes multiple binary input values x_i, each characterized by an assigned weight w_i, that represents the importance of the respective inputs. These weights are initialized most commonly randomly. In the summing node in the next step, a weighted sum of all inputs and weights is computed, as well as an incorporation of an externally bias, denoted by b, is applied [97]. The underlying mathematical operation is a matrix multiplication between the input vector $\vec{x}$ and the weight vector $\vec{\omega}$. A non-linear activation function $h(z)$ maps the weighted sum together with the applied bias and defines the final output of the neuron as:

$$\text{output} = h(z) = h\left(b + \sum_{i=1}^{N} \boldsymbol{\omega}^{\boldsymbol{T}} \boldsymbol{x}\right) = h\left(b + \sum_{i=1}^{N} \omega_i x_i\right). \tag{5.1}$$

The most common activation function for the perceptron is the Heaviside step function that returns *true* when the input passes a certain threshold limit whereas it produces *false*, when the input lies below the threshold value. The perceptron model consists of one layer only, which prevents it from performing non-linear classification. In the next section, a more complex structure consisting of a network of several layers is discussed.

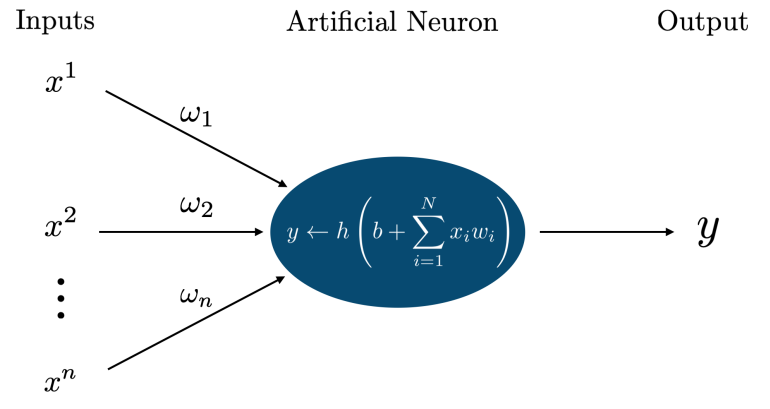


Figure 5.1: Basic architecture of a perceptron. The perceptron receives multiple input values (x_i), each assigned by a weight (w_i) which represents the importance of the respective input. An activated function, denoted as h, maps the weighted sum of all inputs and weights together with an applied bias (b) to produce the output.

5.2.2 Multi-Layer Perceptron

Neural networks that consist of more than one perceptron are called multi-layer perceptrons (MLPs) or a deep artificial neural network. In contrast to the previously mentioned perceptron model, the multi-layer perceptron has an increased complexity that consists, besides the input and output layer, of an arbitrary number of so-called *hidden layers*.

Similar to the perceptron model, the weighted sum of the input values is calculated for each node and an activation function is applied to it. The output of the node is then passed to the next hidden layer and the same process is repeated.

Deep learning is particularly concerned with networks of this kind, where more than one layer exists and networks of this kind are called deep neural networks. Figure 5.2 shows an example for a multi-layer perceptron with n-dimensional input values, three hidden layers and an output layer.

While typically most deep neural networks, and also the network used in this thesis, are so-called *feed-forward neural networks*, where the information is only transferred straight into one direction (from the input layer to the output layer), there is also a second type, called *recurrent neural networks*. Recurrent networks are not feed-forward and are characterised by the fact that they have a feedback loop from neurons from the current layer to neurons from the previous layer. They are particularly useful when temporary aspects or sequences need to be considered [98].

5.2.3 Activation Functions

As already mentioned in the previous section, the activation function is an essential part of a neural network, which is applied to each neuron in the network to determine the output of that neuron. The main purpose of an activation function is to bring non-linearity to that network, allowing the network to learn complex structures in the data. However most patterns, defined by data today, are non-linear, therefore adding non-linearity to the activation function is a crucial condition.

There are a number of activation functions that can be used in a neural network, but not every function is suitable for a particular task. In the following section, the most frequently used activation functions are presented.

Sigmoid

Sigmoid functions are characterized by their S-shaped curve. There are many sigmoid functions that can be used like the hyperbolic-tangent or the arc-tangent function, but the most popular one is the *logistic sigmoid function* as shown in figure 5.3, which is defined as following:

$$h(z) = \left(\frac{1}{1 + \exp^{-z}} \right). \tag{5.2}$$

The output of the sigmoid function lies in between the range 0 and 1 and is therefore particularly useful when probabilities for the output of a neural network have to be determined. One disadvantage of the sigmoid function becomes apparent, when it is used as the activation function within the hidden layers. Figure 5.3 clearly shows, that at very high or very low sigmoid function values, the values of the derivative of the sigmoid function become vanishingly small. As a consequence, this causes the gradients to vanish and learning becomes very poor. One solution to overcome this problem is using the ReLU function that is described in the following section.

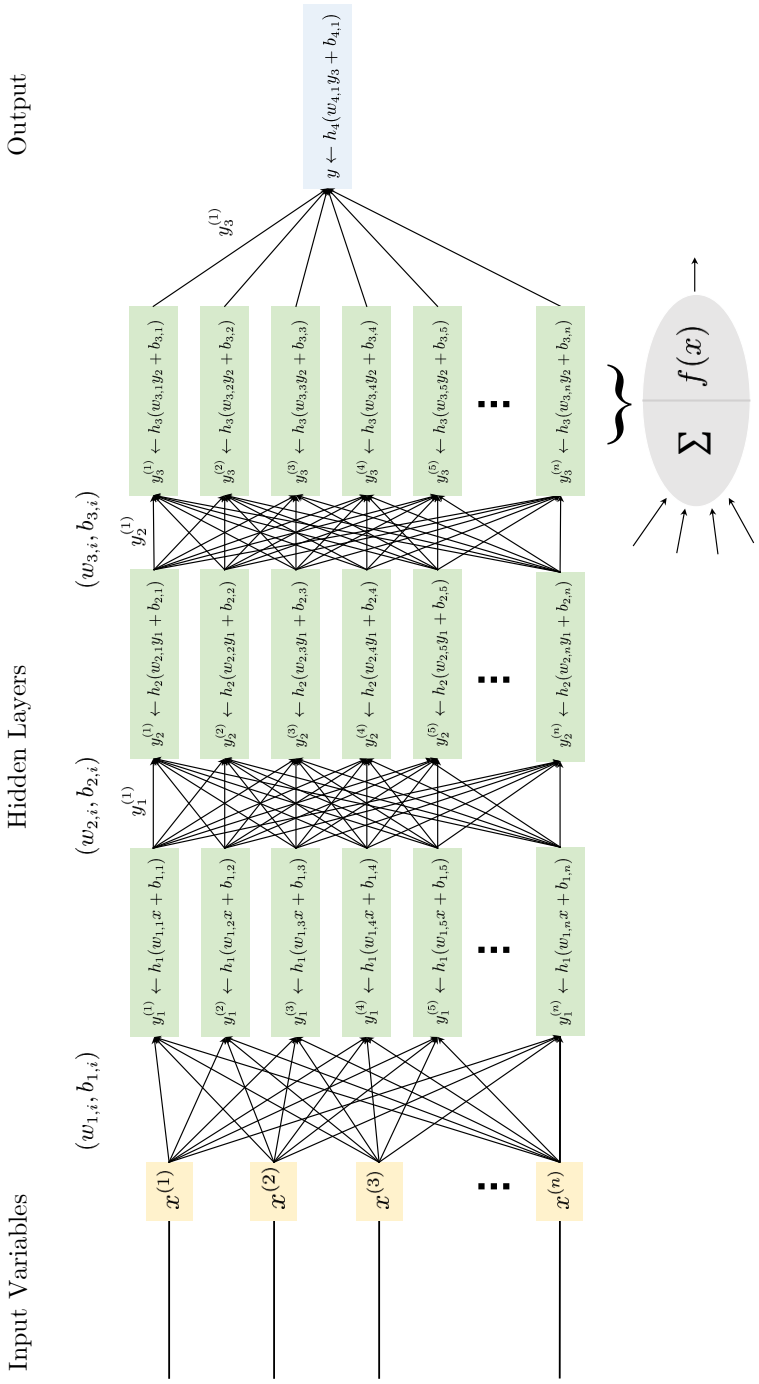


Figure 5.2: A multi-layer perceptron example with an n-dimensional input layer, three hidden layers and an output layer. Modified illustration based on [96]

In the case of the neural networks implemented in this thesis, the sigmoid function is used in the output layer for the binary classification task of distinguishing hadronic taus from misidentified taus.

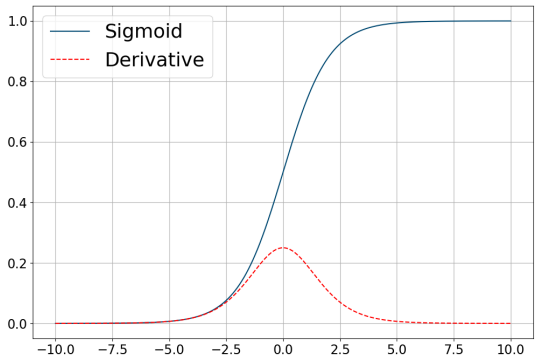


Figure 5.3: S-Shape of the sigmoid function and its derivative.

Tanh

An alternative to the sigmoid function with a similar S-shaped curve look is the hyperbolic-tangent function. Figure 5.4 shows the shape of the tanh function. While the output of the sigmoid function lies in between 0 and 1, the hyperbolic-tangent function has a superior coverage from −1 to 1 and it is defined as:

$$h(z) = \tanh(z).$$

The benefit of using the hyperbolic tangent function over the sigmoid function lies in the extended range. Negative values are correctly mapped to negative outputs while near zero values are mapped to outputs near zero. In addition, the function is symmetrical around zero.

Because of these properties, the hyperbolic tangent function is, in general, preferred over the sigmoid function. The sigmoid function, on the other hand, has the unique property of giving probability values in the output.

Rectified Linear Unit (ReLU)

This type of activation function is widely used and has become the default used activation function for a lot of different neural network applications and is also applied for the hidden layers of the neural networks implemented in this thesis. The ReLU function is a piece-wise non-linear function that checks, whether the input it greater or less than 0 and is defined as:

$$ReLU(z) = \max(0, z) \qquad \text{or as} \qquad ReLU(z) = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } z \geq 0 \end{cases}$$

Depending on the input, it returns 0 weather the input is less than 0, otherwise, if the input is greater than 0, the function returns the value equal to its input. The ReLU function is particular useful because

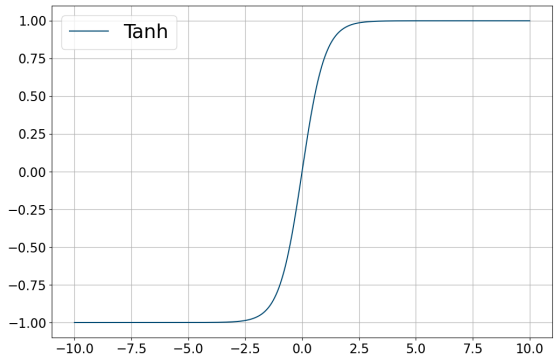


Figure 5.4: S-Shape of the hyperbolic tangent function.

it uses a trivial max function compared to the sigmoid or tangent function, which require exponential calculations, making the overall computation time much faster [99]. Another benefit using the ReLU function is that it overcomes the previously mentioned problem of vanishing gradients because it does not cause small derivatives. Figure 5.5 shows the shape of the ReLU function.

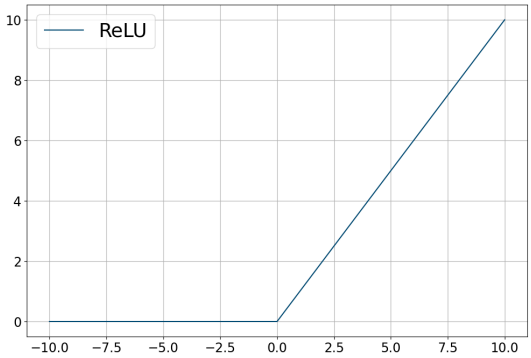


Figure 5.5: Shape of the ReLU function.

5.2.4 The Learning Process

As already seen, the activation function applied to each neuron gives the value of the neuron in the subsequent layer. This calculation continues until the last layer is reached to produce the final output prediction value. The output value, however, has no meaning yet, as there is no information how well or how badly the predicted output value deviates from the actual value (also called *truth label*). Only the direct comparison between these two, gives an indication of how well the network has performed. It is not surprising that the output of the neural network does not match the expected output. Since the weights were initialised randomly, the probabilities of the predicted output value also correspond to random values. In order to achieve a desired output value, the weights have to be adjusted.

The primary step for learning neural networks is to define a *Cost function C* (also called *Loss function*), which is a function to measure the quality of a neural network, in other words, to estimate how well the neural network fits with the given dataset. When the predicted value matches the actual value, the function will return a lower value, otherwise if the predicted value is off the actual value, the function returns a higher value. Depending on the classification task, there are myriad types of cost functions. For a typical binary classification problem, the commonly used cost function is the *binary cross-entropy* which can be calculated as [100]:

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i) \tag{5.3}$$

, where N denotes the total number of possible data, $\hat{y}_i \in \{0, 1\}$ the prediction value for the i-th class and $y_i \in \{0, 1\}$ the corresponding desired target value. Since the output y_i can only be 0 or 1 for a binary classification task, the sum vanishes and equation 5.3 can be rewritten as:

$$\text{BCE} = -(y \log \hat{y} + (1 - y) \log (1 - \hat{y})). \tag{5.4}$$

From equation 5.4 one can see, that the previous mentioned statement is satisfied. When the predicted output gets close to the desired output, the cost function tends towards zero. Figure 5.6 shows the shape of the binary cross-entropy cost function for the target value $y = 0$ and $y = 1$ with respect to the probability output $\hat{y}$. In the case of $y = 0$, the cost function has a high value when the prediction probability is close to 1, while the cost function reaches a minimum value, when the target value is reached with a high probability. The behaviour for $y = 1$ is completely analogous.

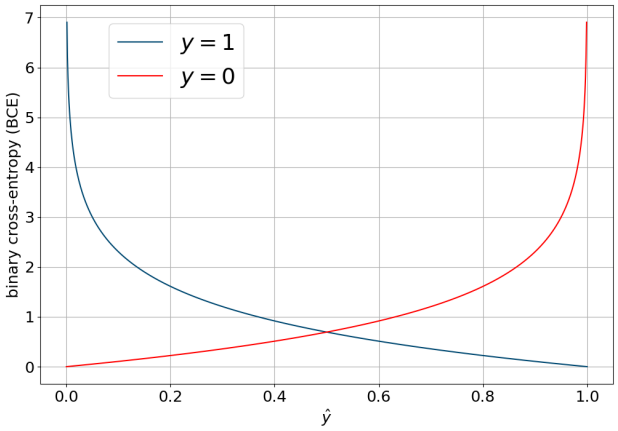


Figure 5.6: Shape of the binary cross-entropy cost function (BCE) for $y = 0$ and $y = 1$.

The goal is to find the overall minimum of the cost function in order to improve the accuracy of the neural network and to achieve the target value. A technique called *gradient descent* is an iterative optimization algorithm that is, in general, used to find the local or global minimum of a function by using partial derivatives. In this case, the method is applied to the cost function, to identify, how the free parameter, weights and biases, have to be shifted to minimize the cost function by calculating its partial derivative.

The derivative points into the direction of the steepest ascent and its sign indicates, if the function increases (positive gradient) or decreases (negative gradient) at this point [101]. For the above given binary-cross entropy cost function, the gradient can be calculated for a single neuron with respect to the weights w_i as:

$$\nabla \mathrm{C} = \left(\frac{\partial \mathrm{C}}{\partial \omega_0}, \frac{\mathrm{C}}{\partial \omega_1}, ..., \frac{\mathrm{C}}{\partial \omega_n} \right)^T$$

By rewriting $\hat{y} = \sigma(z)$, where the definition of z is defined in 5.3, the derivative of the cost function can be calculated by using the chain rule:

$$\frac{\partial C}{\partial \omega_{ij}} = \frac{\partial C}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}}.$$

For all training inputs x, this can be calculated as [102, 103]:

$$\frac{\partial C}{\partial \omega_i} = -\frac{1}{N} \sum_x \left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma(z)} \right) \frac{\partial \sigma}{\partial \omega_i} = -\frac{1}{N} \sum_x \left(\frac{y}{\sigma(z)} - \frac{(1-y)}{1-\sigma(z)} \right) \sigma'(z) x_i$$

, which can be written by using the derivative of the sigmoid function $\sigma'(z) = \sigma(z)(1-\sigma(z))$ in a single denominator expression as:

$$\begin{aligned} \frac{\partial C}{\partial \omega_i} &= \frac{1}{N} \sum_x \left(\frac{y}{\sigma(z)} \sigma(z)(1-\sigma(z)) - \frac{(1-y)}{1-\sigma(z)} \sigma(z)(1-\sigma(z)) \right) \\ &= -\frac{1}{N} \sum_x \left(\frac{y(1-\sigma(z))^2 - (1-y)\sigma(z)(1-\sigma(z))}{(1-\sigma(z))} \right) x_i \\ &= \frac{1}{N} \sum_x x_i \left(\sigma(z) - y \right). \end{aligned}$$

This simple expression tells at which rate the weight learns and its controlled by the error in the output $(\sigma(z) - y)$ [103]. The derivative with respect to the bias can be calculated similarly and the result is:

$$\frac{\partial C}{\partial b} = \frac{1}{N} \sum_x \left(\sigma(z) - y \right).$$

In general after computing the gradients with respect to the parameters, the parameters of the neural network are updated by calculating [100]:

$$\begin{aligned} w &\leftarrow w - \eta \frac{\partial C}{\partial w} \\ b &\leftarrow b - \eta \frac{\partial C}{\partial b} \end{aligned}$$

where η is defined as the learning rate that controls the size of steps of the gradient. In each training epoch, the parameters are updates accordingly by using gradient descent.

However, this calculation can be time-consuming because it has to run through the entire data set after each training epoch. Stochastic Gradient Descent (SGD) is an optimization algorithm used to overcome this problem by only looking at a subset of the training sample to estimate the gradients instead of all of them. Besides SGD, there are a number of other optimizer such as Adam, Nadam or RMSProp (for further information see [104]).

The Backpropagation algorithm is used in neural networks to implement the gradient descent in neural networks with multiple layers and it computes the gradients of the cost function in order to change the weight values in backward direction [105]. Therefore it propagates the errors through the network from the output to the input node. It uses the gradient descent technique to collect the derivatives of the error and to optimizes these parameters. The calculation for the gradient at a particular layer is done by combining all of the following layers via the chain rule. Further information on how the algorithm works in details can be found in [106].

5.2.5 Dropout

The dropout is used in neural networks to drop out units (neurons) in each layer from the network during each training iteration. The values of the weights and biases of this neuron are no longer used during training. This effective method prevents neural networks of performing overfitting. Figure 5.7 shows a typical neural network without dropout on the left and the same network with applied dropout on the right. The neuron is dropped at random by a fixed probability hyperparameter p [107]. A value of $p = 0.5$ indicates, that half of the nodes are dropped.

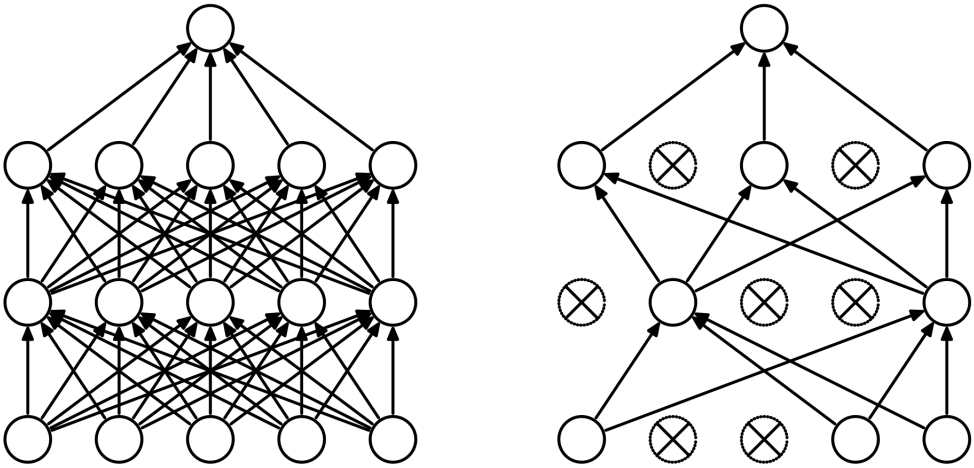


Figure 5.7: Illustration of a standard neural network on the left and a neural network with dropout on the right [108].

5.2.6 The Need for Weak Supervision

In high energy physics, neural networks are usually trained using fully labeled simulated samples (*full supervision*). In the case of binary classification, the training samples consist of two class labels: one

is signal, the other one is background. In the context of this thesis, signal events are referred to real hadronic tau leptons, while background events are fake tau leptons. The network learns to distinguish between signal and background by using the label of the event, which helps to show the network which example it got wrong to update its parameters accordingly and to improve the classification [94].

However, the disadvantage of this method is that it relies on simulated samples, which are not perfect and it may happen that the network is learning artifacts and the classification fails on data. In addition, the label of each event must be known which does not exist for real data.

Therefore, *weak supervision* is also used to differentiate between real and fake tau leptons using mixtures of events instead of true labels. The following section introduces weak supervision and provides deeper insights into two weakly supervised paradigms.

From Supervised to Weakly Supervised Learning

As already seen in section 5.1, the dataset for supervised learning consists of a sequence of input-output data pairs $\{(x_i, y_i)\}_{i=1}^{N}$ to train the model and predict targets on future data by mapping a function that finds a relationship between the inputs and outputs. For this supervised classification task, the binary cross-entropy given in formula 5.3 can be rewritten with the indicator function $\mathbb{I}$ as [94]:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{I}(u_i = S)\log\hat{y}_i + (1 - \mathbb{I}(u_i = S))\log(1-\hat{y}_i)\right).$$

Additionally, for the given probability distributions of $\vec{x}$ defined as $p_S(\vec{x})$ and $p_B(\vec{x})$ for signal and background, the optimal classifier motivated by the Neyman-Pearson lemma (further information given in [109]) is the likelihood ratio $L_{S/B}(\vec{x}) = \hat{y}_{\text{optimal}}(\vec{x}) = p_S(\vec{x})/p_B(\vec{x})$ [94, 110]. Having a large training data set and the right hyperparameters (further explanation in section 6.2) for classification, $\hat{y}$ reaches the optimal prediction output $\hat{y}_{\text{optimal}}$, by minimizing the loss function.

In contrast, the dataset for weak supervision consists of mixtures of both, signal and background examples. The classification goal of predicting the output remains the same as in the fully supervised case. According to [111], there are three different types of weak supervision: incomplete supervision, inexact supervision and inaccurate supervision.

The first type, *incomplete supervision* refers to the case, where a small proportion of training data is labeled, while the rest is unlabeled [111]. Formally the training data set can be written as $\{(x_i, y_i), ..., (x_l, y_l), x_{l+1}, ..., x_m\}$, where l denotes the number of training samples and the index $u = m - l$ the number of unlabeled samples [111]. The second type is *inexact supervision*, where a set of inexact labels is given. Finally, the last type, *inaccurate supervision*, describes the case where some training labels are ground-truth; somehow inexact while the rest has correct labels. In practice, these three types of weak supervision occur often simultaneously.

Two famous paradigms that fall under the umbrella of weak learning that are tailored for physics applications are *Learning from Label Proportions* (LLP) and *Classification Without Labels* (CWoLa). Both are described in the following section separately, while the neural network used in this thesis is based on the latter approach.

Learning from Label Proportions (LLP)

In high energy physics, weak supervision was, regarding to [94], first applied to distinguish jets originating from quarks from those originating from gluons [94]. The classifier for this task only used class proportions during the training by using an approach called *Learning from Label Proportions*. Considering a binary classification task with two mixed datasets that are labeled as M_1 and M_2 containing both signal and background events. Each dataset consists of a known fraction of signal to background of f_1 and f_2 respectively, without knowing the truth label of any event. For both datasets, the distributions read [94]:

$$p_{M_1}(\vec{x}) = f_1 p_s(\vec{x}) + (1 - f_1) p_B(\vec{x})$$
$$p_{M_2}(\vec{x}) = f_2 p_s(\vec{x}) + (1 - f_2) p_B(\vec{x}),$$

where the signal fractions satisfy $0 \leq f_2 < f_1 \leq 1$. The training set is then broken into smaller batches (also called *bags*), annotated with the proportion of instances arising from each class [112]. One possible loss function for training is given in [94]:

$$l_{LLP} = \left| \sum_{i=1}^{N_{M_1}} \frac{\hat{y}_i}{N_{M_1}} - f_1 \right| + \left| \sum_{j=1}^{N_{M_2}} \frac{\hat{y}_j}{N_{M_2}} - f_2 \right|,$$

where N_{M_1} and N_{M_1} denotes the number of M_1 and M_2 samples in each batch.

Classification Without Labels (CWoLa)

Instead of modifying the loss function, the CWoLa approach is a technique to directly discriminate the mixed samples M_1 and M_2 [94]. The optimal classifier to distinguish both mixed samples from each other shows remarkable performance as an optimal classifier for distinguishing signal from background, regarding [94]. The classifier smoothly approaches the fully supervised paradigm, when the fractions approach complete purity ($f_1 \to 0$, $f_2 \to 1$) [113]. And most importantly, CWoLa can be trained directly on the data. An illustration of the CWoLa framework is shown in figure 5.8.

The figure shows, that each mixed sample consists of a different signal fraction, f_1 for the signal fraction for M_1 and f_2 for the signal fraction of M_2. However, the knowledge of the exact class proportions, unlike LLP, is not needed, instead it is sufficient to know that, for example, M_1 contains more signal than M_2. Also no information about the label is used during the training. For classification, data coming from mixed sample M_1 gets labeled as 0 while the data from M_2 by 1, respectively. The classifier is then trained to distinguish data coming from M_1 or M_2 [94], which results then in the same classifier to distinguish signal from background.

To prove that the optimal classifier, to distinguish M_1 from M_2, is also the optimal classifier to distinguish signal from background in terms of pure samples S and B, one can look at the likelihood ratio, considering that $f_1 > f_2$ [94]:

$$L_{M_1/M_2} = \frac{p_{M_1}}{p_{M_1}} = \frac{f_1 p_S + (1 - f_1) p_B}{f_2 p_S + (1 - f_2) p_B} = \frac{f_1 L_{S/B} + (1 - f_1)}{f_2 L_{S/B} + (1 - f_2)}. \tag{5.5}$$

As long as $f_1 > f_2$, the likelihood ratio of the mixed samples given in equation 5.5 is monotonically related to the signal background likelihood $L_{S/B}$ [94]. If otherwise, $f_1 < f_2$, then one gets the reversed classifier. This means, regarding [94], that $L_{S/B}$ and L_{M_1/M_2} are therefore the same classifier.

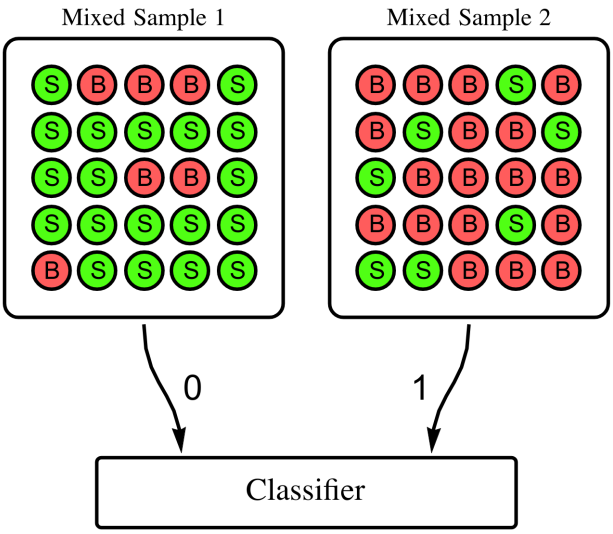


Figure 5.8: Schematic depicting mixed input samples M_1 and M_2 for the CWoLa classifier [94].

Related to the above argument, one has to make sure to have the maximum purity in the samples to have maximal classifier performance. For that, the mixed samples in this thesis are divided in such a way that one of the samples consists of predominantly true taus, while the other one predominantly fake taus. The classifier is then trained to distinguish M_1 from M_2. Further information can be found in chapter 7.

As in all classification problems, one has to make sure to have a large number of training data so that statistical fluctuations are negligible to obtain a great classification performance [94]. Regarding to [94], the performance is poor when the number of training samples is small or the fraction f_1 is close to f_2. Figure 5.9 shows the behaviour of the AUC as a function of the signal fraction f_1, for different training sizes.

As expected, the AUC of the full supervision case is not affected by the signal fraction. More training data results in a better AUC value. For LLP and CWoLa, the AUC drops significantly when $f_1 \to f_2$. While for samples with fewer events, the AUC drops off relatively quickly. With a higher number of samples, the drop is in a narrower range and rises quickly after this range. To summarize, using CWoLa requires to have the maximum purity of the mixed samples and high statistics to achieve maximum performance.

5.3 Technical Details

The neural networks in this thesis are implemented using *CWoLa-Fakes* [114], a sophisticated tool that supports various supervised and unsupervised ML applications. It is based on Keras, which is a powerful open source library that is particularly suitable for numerical computations for machine

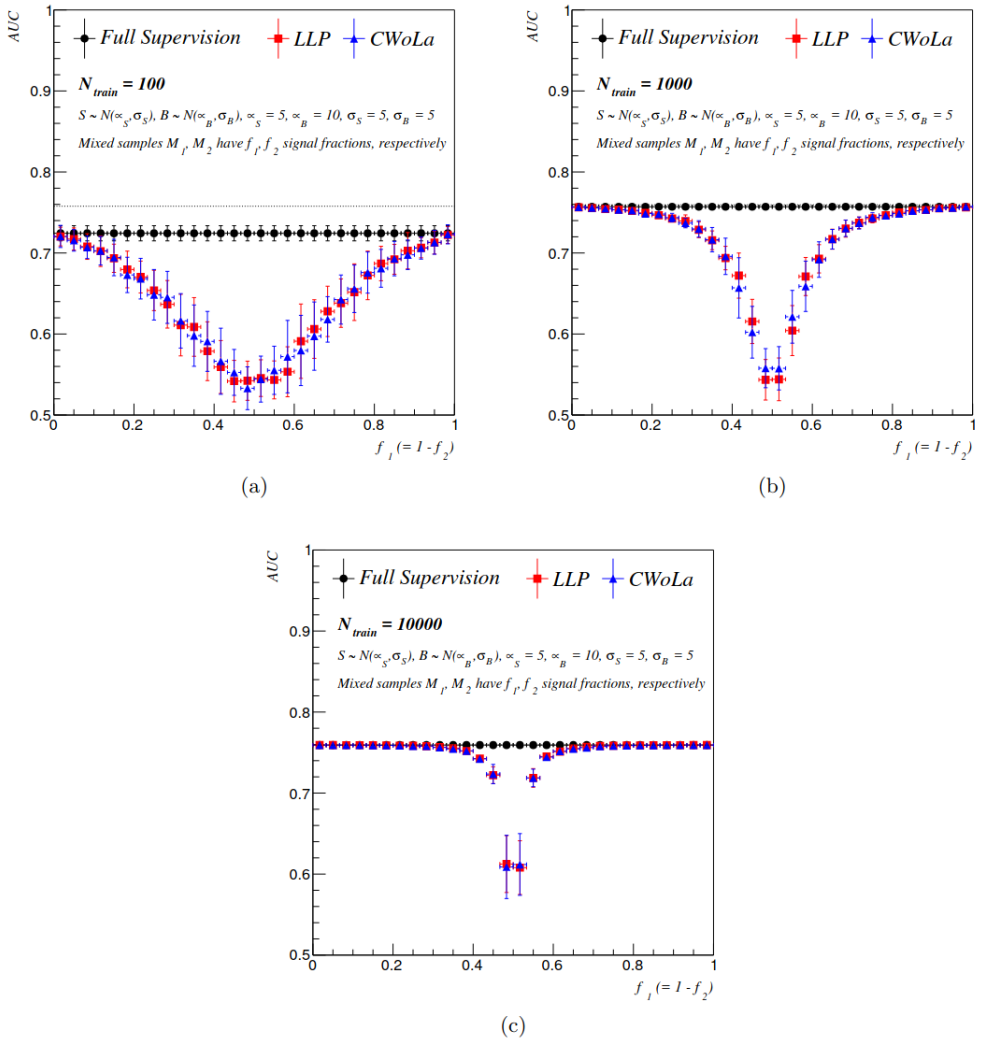


Figure 5.9: A comparison of the AUC value as a function of the signal fraction f_1 for full supervision, LLP and CWoLa. The images show the behaviour for different training sizes: (a) $N_{train} = 100$, (b) $N_{train} = 1000$ and (c) $N_{train} = 10000$ [94].

learning. It has a python interface and offers ready-made functions to train neural networks, load datasets, and to define the architecture of the neural network.

CHAPTER 6

Baseline Classifier Approach for $1e/\mu + 1\tau_{\text{had}}$

This chapter describes the full supervision approach to distinguish true hadronic taus from fake tau leptons. This approach relies on the knowledge of the truth label from every event and requires a training set that is labeled as signal, and a training set labeled as background. Signal events are referred to as those containing only true hadronic taus, while background events store the fake tau information. The classifier is then provided with signal and background training samples from the $1e/\mu + 1\tau_{\text{had}}$ channel to distinguish both classes.

The goal of the full supervision approach is to get a baseline, which classification performance can maximally be achieved by using full supervision to compare it then to a weak supervision approach, which is separately trained on MC and on data that is described in chapter 7. It is expected that due to the complete purity of the training samples in the full supervision approach, the classifier performs better than the weak supervision approach using CWoLa, which is instead trained on mixed samples containing both, signal and background events. For convention, the full supervision approach is from now on always referred to as the baseline approach in the following sections.

This chapter addresses the setup of the baseline classifier, starting with a description of the selection of training variables. The selected variables are ranked according to their significance and the corresponding results are shown. In the last part of this chapter, the results of the baseline classifier approach are discussed and its application is tested on unseen MC samples. The term *unseen* refers to samples that were not used during the training of the classifier. In addition, problems and their possible causes are also mentioned.

6.1 Selection of Training Variables

The baseline classifier is trained on $t\bar{t}$ MC samples (DSID:410470) because as already mentioned in chapter 4.3.1, this process is enriched by fake taus and provides enough statistics for the classifier to find differences between both classes. A tau identification variable differs between true hadronic tau and fake tau events. This variable is used to split the MC samples of the $t\bar{t}$ process into one sample containing only true hadronic tau events, and one sample containing pure fake tau events. In total, there are three $t\bar{t}$ MC samples, that are split into two training sets, one labeled as true, the other one as fake. Table 6.1 shows the separated $t\bar{t}$ MC samples for true and fake, together with the number of events in each sample as well as the total number of events in each training set.

After the splitting, all true samples are merged into one sample as well as all fake samples into a second sample, to create the two final samples for classification. This step is crucial because the higher the statistics, the better the model can adapt weights and biases to increase the performance of the network.

Subsequently, composition plots are created for each available variable from the $t\bar{t}$ MC samples to find variables that fit most suitable for the classifier training. For the composition plots, the signal MC samples from table 6.1(a) are plotted against the background MC samples from table 6.1(b). The variable candidates that show only a small separation between true and fake were neglected for training, while those that show a good separation are taken for the training of the baseline classifier. The selection of variables by looking at signal versus background is a common approach to find variables that are most suitable for training neural networks. Figure 6.1 shows the MC composition plots for true versus fake from the 2015+2016 $t\bar{t}$ MC sample for two example kinematic variables, E_{lep1} and E_{lep2}. Both are an example of two suitable variable candidates that are used for the baseline classifier training.

The complete list of selected variables consists of simple kinematic variables, tau specific variables and some combined variables. These variables are used for training the baseline classifier and are summarized in table 6.2. For clarity's sake, all MC composition plots for these variables are shown in appendix A.2. Since the variables for the network are now defined, the next section describes the setup of the network architecture.

Year	MC Sample	Events
2015+2016	410470	909151
2017	410470	1103749
2018	410470	1485338
Total		3498238

(a) MC True Samples

Year	MC Sample	Events
2015+2016	410470	659685
2017	410470	909887
2018	410470	1217107
Total		2786679

(b) MC Fake Samples

Table 6.1: Available $t\bar{t}$ MC samples (DSID:410470) from v31 single-top ntuples, split by true and fake events, used for training the baseline classifier. The left table 6.1(a) shows the samples containing true hadronic taus, the right table 6.1(b) shows samples with events labeled as fake.

6.2 Network Architecture and Hyperparameter Optimization

This section describes the setup of the baseline classifier neural network. As already described in the previous chapter 5, a neural network consists of a set of nodes, hidden layers and outputs. The network is fed with selected training variables and tries to map the inputs correctly to the outputs.

Before starting the training of the baseline classifier, the network architecture has to be defined. The architecture is defined by a set of hyperparameters. These hyperparameters include, for example, the total number of hidden layers, the number of nodes in each layer or the type of applied activation function. The choice of these hyperparameters is a nontrivial task in deep learning. In general, for this optimization process, many hyperparameters are involved that need to be adjusted to find the overall best classification performance.

When setting up the baseline neural network for the first time, the choice of hyperparameters is arbitrarily and the values of these hyperparameters are adjusted after each training process until the

Kinematic Variables	τ_{had} Variables	Combined Variables
E_{jet1}	$p_{T,\tau_{\text{had}}}$	ΔR_{Min}
E_{jet2}	$\eta_{\tau_{\text{had}}}$	m_{ll}
E_{lep1}	$\phi_{\tau_{\text{had}}}$	ip_{l1j2}
E_{lep2}	$E_{\tau_{\text{had}}}$	ip_{l2j1}
η_{lep1}		ip_{l2j2}
η_{lep2}		ip_{l1j1}
η_{jet1}		$\text{bscore}_{\text{jet1}}$
η_{jet2}		$\text{bscore}_{\text{jet2}}$
m_{met}		
$m_{\phi_{\text{met}}}$		
m_{jet1}		
m_{jet2}		
ϕ_{jet1}		
ϕ_{jet2}		
ϕ_{lep1}		
ϕ_{lep2}		
$p_{T,\text{jet1}}$		
$p_{T,\text{jet2}}$		

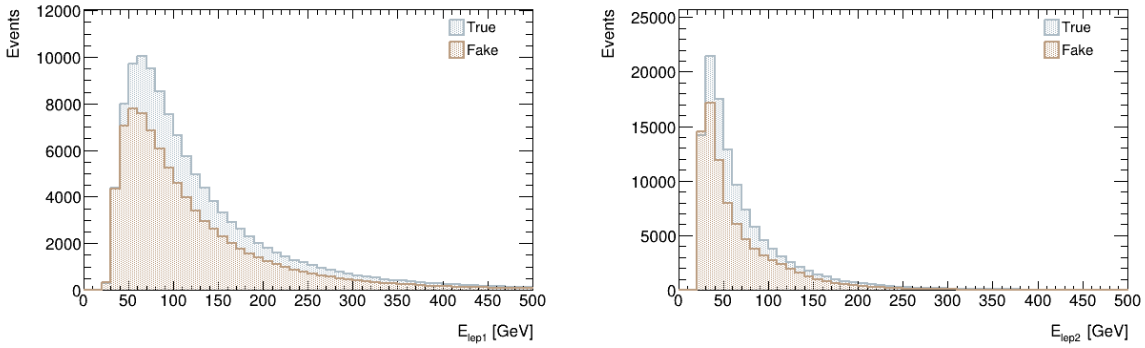
Table 6.2: Complete list of selected training variables for the baseline classifier. The variables can be grouped into kinematic variables, hadronic tau specific variables and combined variables.

best set of hyperparameters is found that minimizes the loss function and give the best classification performance. This process can be time-consuming for powerful networks, as one has to repeat the training with a new set of hyperparameters in the hope of getting a better classification accuracy. A good starting point is to set a minimal configuration of the network, depending on the complexity of the task the neural network has to achieve. Another way and a widely used method for hyperparameter optimization is called grid search, which simply searches for a range of defined hyperparameters, the best combination of values by running over all possible combinations. The side effect of this method is, that it can be really time-consuming because the more dimensions one adds to the search range of hyperparameters, the more the complexity of this task increases.

A wide range of hyperparameters are tested for the baseline classifier. Table 6.3 summarizes the search range of hyperparameters that were applied by using grid search. This table also contains the final optimal set of hyperparameters calculated by using grid search. The weights for the training were randomly initialised and get adapted through back-propagation during the training.

Additionally, for each variable in table 6.2 the importance is calculated by using a feature ranking method, which determines the variables that are most significant for this network. Table 6.4 shows the results of the variable ranking. When a variable has a higher score of significance, this variable is expected to have a higher impact on a better classification performance for this network.

It has been shown that by removing features with low significance, no added value in performance could be achieved. Combining significant variables with less significant variables achieved in general a better performance of the baseline classifier, as the network was probably better able to gain information and detect artifacts in the samples that are necessary for the algorithm to discriminate real



(a) E_{lep1} (b) E_{lep2}

Figure 6.1: MC Composition Plots for two examples variables used for training the baseline classifier: (a) E_{lep1}, (b) E_{lep2}.

Network Hyperparameter	Search Range	Optimal Value
Nodes	50-400	350
Hidden layers	1-5	2
Output layers	1	1
Activation function (hidden layers)	relu,elu,tanh	relu
Batch size	100-20000	10000
Epochs	30-400	400
Optimizer	adam,sgd	adam

Table 6.3: Overview of the search range of hyperparameters for the baseline classifier trained on MC, together with the final set of chosen optimal values for the baseline classifier.

hadronic taus from fake taus. In the next section, the training results for the baseline classifier are presented and described.

6.3 Classification Results

After setting the configuration of the network with the variables given in table 6.2 together with an optimized set of hyperparameters (see table 6.3), this section addresses the training results of the baseline classifier approach. The output of the training, is given in three different output curves: the receiver-operating characteristic curve (ROC) shown in figure 6.2(a), the model loss given in figure 6.3(a) and the response curve which shows the separation for signal and background in figure 6.2(b).

The ROC curve shows, for this binary classification task, the performance of the model, as described in the previous chapter 5. The *Area Under the Curve* (AUC) is a measure of the area between the ROC curve and the axes and is a performance measure of the network. It states that a higher AUC value results in a better classification model and is more likely to predict true taus correctly as true taus and fake taus correctly as fake taus. When the ROC curve lies on the line, means the AUC has a value of 0.5, the network was not able to separate true from fake taus. A value of 1.0 indicates a perfect prediction accuracy. In general, the overall goal is to maximize this value. The AUC that could be

Variable	Significance	Variable	Significance	Variable	Significance
ϕ_{jet1}	11.338	$m_{\phi_{\text{met}}}$	9.716	$\eta_{\tau_{\text{had1}}}$	3.092
ϕ_{lep2}	11.334	ΔR_{Min}	9.267	$\phi_{\tau_{\text{had1}}}$	2.591
m_{ll}	11.316	ip_{l2j2}	7.080	$E_{\tau_{\text{had1}}}$	2.388
$p_{T,\text{jet2}}$	11.316	ip_{l1j1}	6.635	E_{lep2}	2.177
ϕ_{lep1}	11.262	$\text{bscore}_{\text{jet1}}$	6.425	η_{lep2}	1.727
$p_{T,\text{jet1}}$	11.245	ip_{l2j1}	6.339	η_{lep1}	1.376
ϕ_{jet2}	11.242	ip_{l1j2}	6.189	E_{jet2}	1.152
m_{jet2}	10.458	E_{lep1}	5.640	$\text{bscore}_{\text{jet2}}$	0.908
m_{jet1}	10.242	$p_{T,\tau_{\text{had1}}}$	5.490	η_{jet2}	0.785
m_{met}	9.846	E_{jet1}	4.365	η_{jet1}	0.693

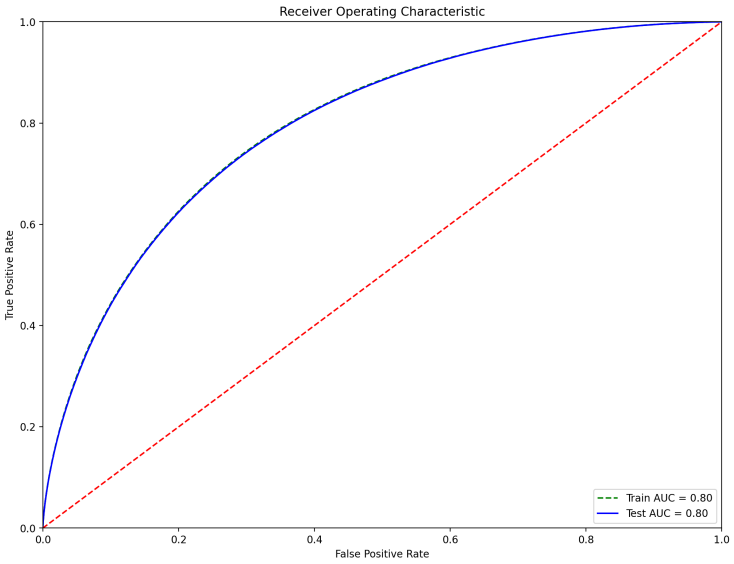
Table 6.4: List of variables (features) ranked by significance for the baseline classifier. Higher values of significance show that this variable is more likely to have a higher impact for a better classification performance.

achieved with the baseline classifier is 0.80, which shows that the possibility is high, that the classifier can predict the desired output.

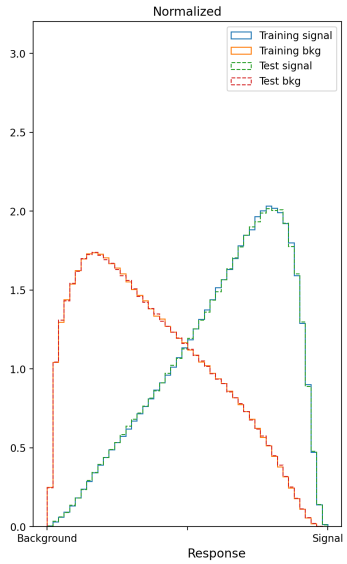
The response of the model in figure 6.2(b) is a measure, how well the model separates signal and background. If the model can not distinguish between these classes, both curves for signal and background would overlap. When both curves are separated and one curve pushed to the signal-like and the other curve pushed to the background-like region, the better the model can distinguish between both classes. Figure 6.2(b) shows a reasonable separation between signal and background. One can see that the trained model is able to distinguish both classes.

Figure 6.3(a) shows the loss-curve. As already discussed in the previous chapter, the goal of a neural network training process is to minimize the loss function. As expected, the loss function decreases with the increasing number of training iterations.

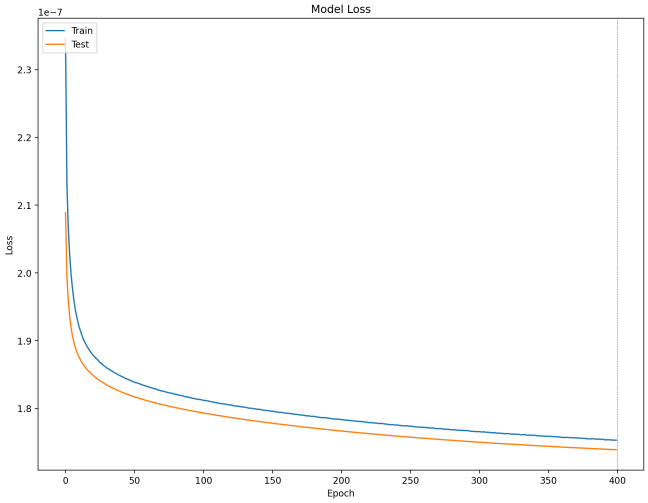
To conclude, all plots shown in figure 6.2(a), 6.3(a) and 6.2(b) behave as one would expect in terms of a good classification performance. However, there is still room for optimization. Taking a larger range of hyperparameters for the grid search could be beneficial for the classifier performance. However, the extra dimensions in the search range of hyperparameters add an increased computing time as a disadvantage. In addition, one could choose a larger number of training epochs to increase the accuracy of the classifier. A significant improvement of classifier performance was also achieved by training a separate classifier for 1-prong and 3-prong taus. The training results for both classifiers are shown in appendix A.3 together with the set of applied hyperparameters.



(a)



(b)



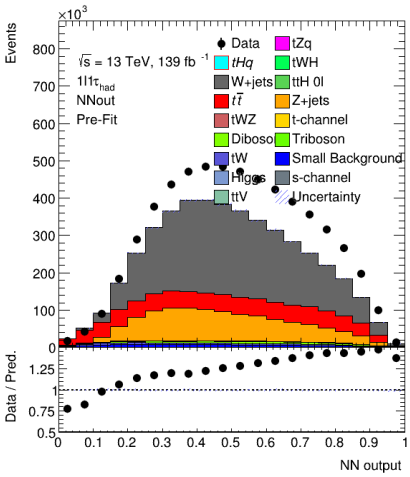
(a)

Figure 6.3: (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve

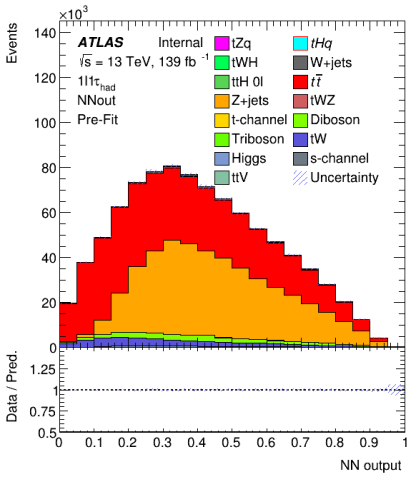
6.4 Prediction on unseen MC Samples

Since the result of this classifier was only trained on MC $t\bar{t}$ samples, it is interesting to check how the model actually performs and predicts on MC samples that were unseen during the training. Therefore, the model has been applied on MC samples from different processes like tZq, W+jets or Z+jets to check, how well the classifier can separate between true and fake taus within these different processes. More information about the complete list of avaiable MC samples is given in appendix A.1.2.

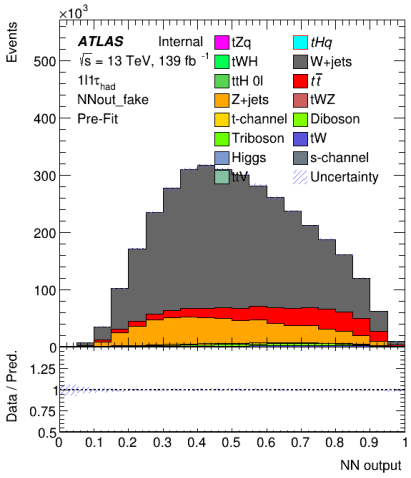
For that, a stacked plot of the neural network output is created that shows the prediction performance of the trained model for the other MC processes. One would expect that the prediction output would look similar to the separation shown in figure 6.2(b). Figure 6.4 shows three neural network output prediction stacked plots. In figure 6.4(a) the complete prediction is shown and additionally shows the MC/data agreement, while figure 6.4(b) and figure 6.4(c) show the output only for true taus and fake taus respectively, by using the truth information.



(a) Complete neural network output including true and fake.



(b) Neural network output for true taus only.

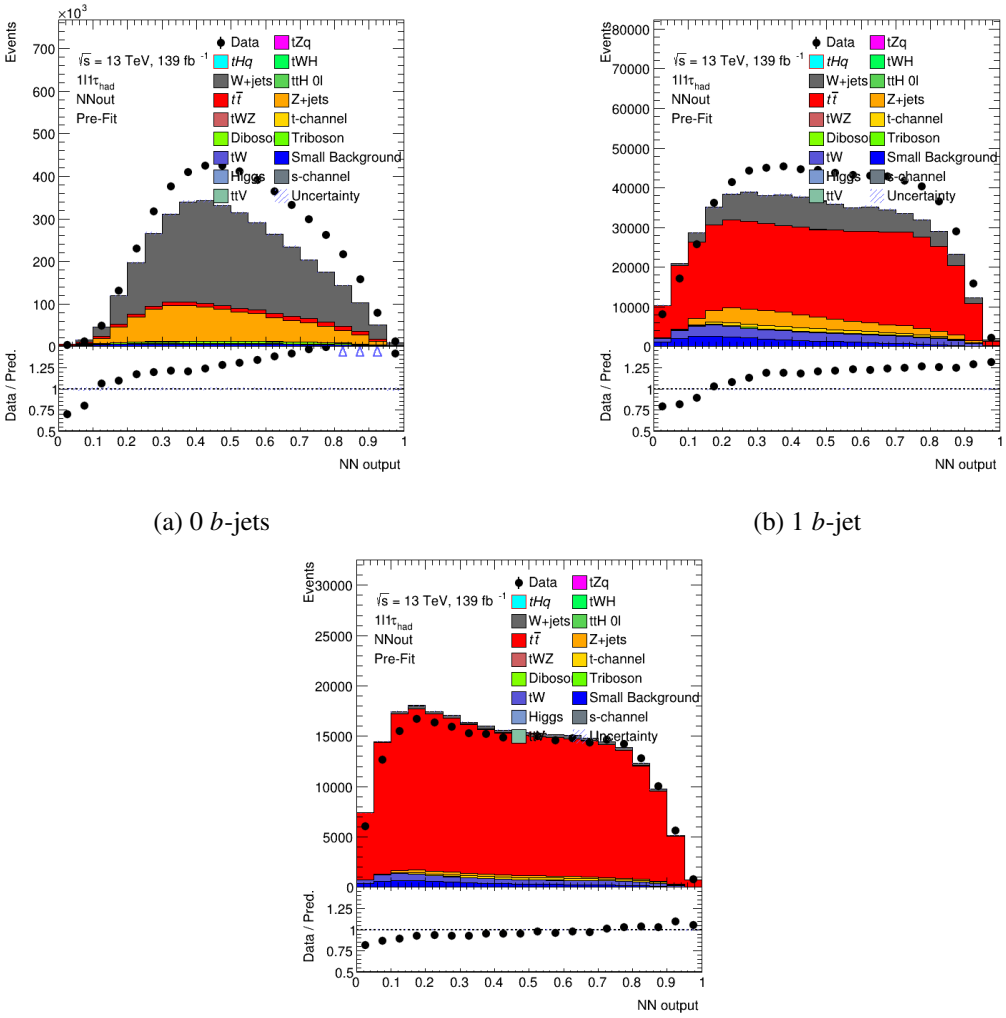


(c) Neural network for fake taus only.

Figure 6.4: Prediction stacked plots for the neural network output on unseen MC samples.

As it can be seen in figure 6.4(a), the neural network output for processes other than $t\bar{t}$ shows a poor separation performance for true and fake. There is also a visible discrepancy between data and MC which is due to the mismodelling of fakes in MC. By using the truth information one would expect figure 6.4(b) to be more signal like, and the output of figure 6.4(c) to be more background like. As it can be seen from these figures, there is almost no such behaviour, suggesting that the baseline classifier does not add any value in other regions than $t\bar{t}$.

Since the classifier was trained inclusive in b-jets, one could check the neural network output for different number of b-jets. Similar to figure 6.4(a), the trained baseline classifier model is applied to the other processes but this for three different b-jet configuration. The neural network output is shown for zero b-jets, for one b-jet and for two b-jets in figure 6.5.



(a) 0 b-jets

(b) 1 b-jet

(c) 2 b-jets

Figure 6.5: Prediction stacked plots for the neural network output for three different b-jet configurations: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets.

One can see an improvement of separation in the 1 b-jet and 2 b-jet configuration compared to 0 b-jets because both b-jet configurations are dominated by $t\bar{t}$. Since the classifier was trained on $t\bar{t}$, the separation shows therefore a significant improvement for this process.

CHAPTER 7

Mixed Sample Classifier (CWoLa) for $1e/\mu + 1\tau_{\text{had}}$

This chapter describes the weakly supervised framework CWoLa as a tool to distinguish real hadronic taus from fake taus. Similar to the baseline classifier described in the previous chapter, the first part of this chapter describes a CWoLa classifier trained on $t\bar{t}$ MC samples from the $1e/\mu + 1\tau_{\text{had}}$ channel.

The chapter starts by explaining how the two mixed samples are prepared. Then, the setup of the classifier is discussed and the classification results are presented. Since this classifier is also trained on MC, a direct comparison can be made with the full supervision approach. One would expect better performance of the baseline classifier since the training set consists of two perfectly separated samples, one for true and one for false, compared to the mixed samples used in the CWoLa approach. CWoLa also offers the possibility to train directly on data, which has the advantage of no longer relying on simulated data.

In the second part of this chapter, this approach is described and the preparation of the mixed samples and the training results are presented. Also, this classifier is trained in the $1e/\mu + 1\tau_{\text{had}}$ channel. In the last part, both classifiers are applied on MC samples that were unseen during training to check their performance.

7.1 CWoLa trained on MC Samples

This section describes the approach of CWoLa trained on $t\bar{t}$ MC samples because this process, as described earlier, has the highest amount of fakes and provides sufficient statistics. Instead of training on two completely separated samples for true and fake, CWoLa uses mixed samples. These mixed samples must be prepared in such a way that they have high purity and high statistics, as already seen in chapter 5. To create these mixed samples, a cut is applied on the RNNScore to have one sample that is enriched by true taus and one sample that is enriched by fakes. In the following section, the preparation of these samples is explained.

7.1.1 Mixed Samples for Training

The $t\bar{t}$ MC samples that are used for training the CWoLa MC classifier are listed in table 7.1. These samples are identical to the $t\bar{t}$ MC samples that were used for the baseline classifier training (see table

6.1), but for CWoLa MC these samples are not divided into one sample containing only true taus and the other sample containing only fake taus because the mixed samples for CWoLa need to have proportions from both classes, true and fake.

Year	MC Sample	Events
2015+2016	410470	1568836
2017	410470	2013636
2018	410470	2702445
Total		6284917

Table 7.1: Available $t\bar{t}$ MC samples (DSID:410470) from v31 single-top ntuples, used to train the CWoLa classifier on MC.

CWoLa requires two mixed samples that are created by applying a cut on the RNNScore variable. Figure 7.1 shows the RNNScore distribution. While this distribution is flat for true hadronic taus, it decreases for fake taus with a higher RNNScore. For each mixed sample a cut is applied on this RNNScore to ensure that one mixed sample consists of mostly true taus, while the other mixed sample is enriched by fake taus. The procedure for the cut selection is described in the following paragraph.

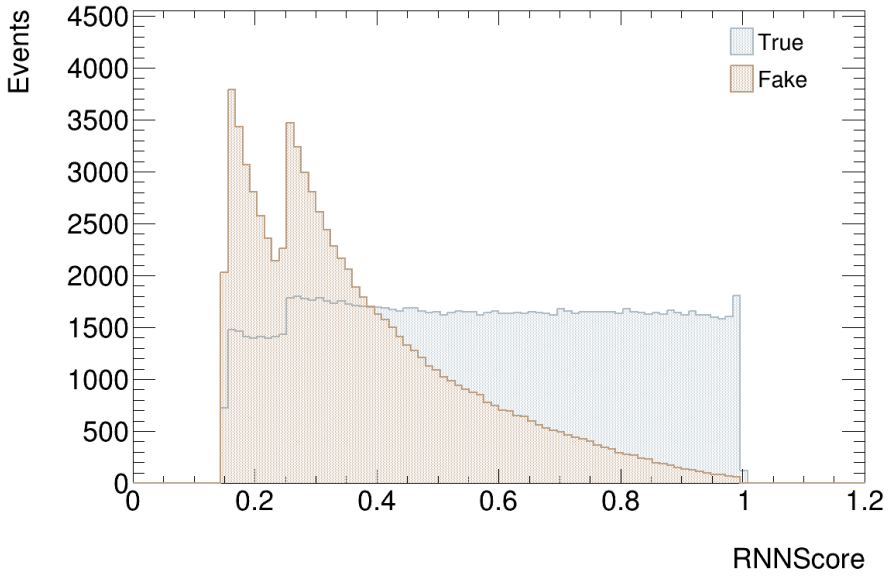


Figure 7.1: Distribution for the RNNScore. The distribution is flat for true hadronic taus, while the distribution decreases for fake taus with a higher RNNScore. Here the loose working point is used with a signal efficiency of 85% for 1-prong tau decays and 75% for 3-prong decays, which explains the appearance of two peaks in the distribution.

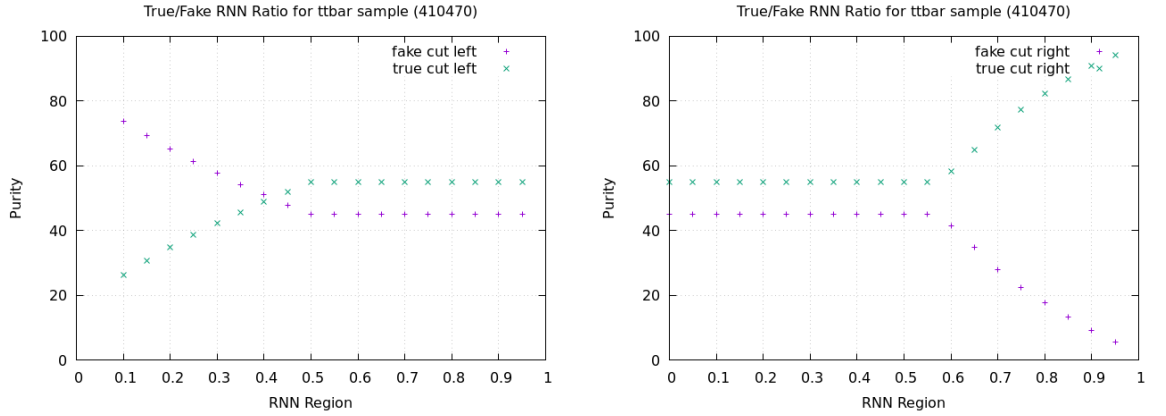
To select the cuts, the ratios for true and fake were calculated for specific regions on the x-axis of the RNNScore to ensure to have samples that have both high purity and high statistics. While the left region of the RNNScore is enriched by fake taus, the right side of the RNNscore clearly shows a dominant number of true taus. For the first cut, the x-axis was scanned starting from the left side in

steps of 0.05 and the ratios for true and fake taus were determined. These ratios were then calculated as:

$$\text{Ratio True} = \frac{\text{Entries True}}{\text{Entries True} + \text{Entries Fake}}, \qquad \text{Ratio Fake} = \frac{\text{Entries Fake}}{\text{Entries True} + \text{Entries Fake}}.$$

In steps of 0.05 the ratios from the current and the previous step were summed up until the complete RNNScore was scanned. The result for the scan from the left to the right side of the RNNScore for the first cut is shown in figure 7.2(a).

Identical to the scan starting from the left side, a scan was also performed from the right side, which is the region that is enriched by mostly true taus. The result of this calculation can be seen in figure 7.2(b).



(a) True and fake ratios for the scan performed from the left to the right side.

(b) True and fake ratios for the scan performed from the right to the left side.

Figure 7.2: Ratios for true and fake: (a) for the scan from the left to the right side for the fake enriched region and (b) for the scan performed from the right to the left side for the true enriched region.

The cut is optimal, if one ratio is at maximum, while the other ratio is at minimum to have samples that are enriched by one class. But this is only correct, if there are enough statistics in this sample available. For figure 7.2(a), the cut for the first mixed sample was set at RNNscore < 0.19. This cut ensures that the first mixed sample (M1) is then enriched by fake taus. The cut that was applied for the second mixed sample (M2) was set at RNNscore > 0.90, which is then enriched by true taus. The exact number of events in each sample along with a summary of the applied cuts and class proportions can be reviewed from table 7.2. These mixed samples are now used as the training input for the CWoLa classifier trained on MC.

Training Sample	Applied Cut	Ratio of true and fake taus	Events
Mixed Sample M1	RNNscore < 0.19	76.2 % fake, 23.8 % true	643020
Mixed Sample M2	RNNscore > 0.90	6.3 % fake, 93.7 % true	435100

Table 7.2: Summary of the cuts applied to the two mixed samples M1 and M2 along with the ratio of true and fake taus in each mixed sample. In addition to that, the total number of events in each sample is shown.

7.1.2 Selection of Training Variables and CWoLa Network Architecture

Initially, the same variables were used as in the baseline classifier approach because these variables showed in general a good visible separation. It has been tested that the performance of the classifier was significantly better after removing some variables from this list. In particular, the removal of the tau specific variables leads to better overall classification performance. Similar to the baseline approach, a feature ranking of the variables has been performed. The result of this ranking is shown in table 7.3.

Variable	Significance
ϕ_{lep2}	14.407
m_{jet2}	14.328
ϕ_{jet2}	14.257
$m_{\phi_{\text{met}}}$	14.220
m_{jet1}	14.210
$d\phi_{l2j_{\text{min}}}$	14.192
η_{lep2}	14.151
ϕ_{lep1}	14.138
$d\phi_{l1j_{\text{min}}}$	14.118
m_{met}	14.087
$p_{T,\text{jet2}}$	14.074

Variable	Significance
ϕ_{jet1}	14.010
$p_{T,\text{jet1}}$	13.992
η_{lep1}	13.890
$\text{bscore}_{\text{jet2}}$	11.254
$\text{bscore}_{\text{jet1}}$	4.357
ΔR_{Min}	1.629
E_{lep2}	1.413
E_{jet1}	1.101
E_{jet2}	0.496
E_{lep1}	0.026

Table 7.3: List of variables (features) ranked by significance for the CWoLa classifier trained on MC. Higher values of significance show that this variable is more likely to have a higher impact for a better classification performance.

Again, the variables are ranked by their significance for the network. For example the removal of the variables E_{lep1} or E_{jet2} showed in general no better separation for true and fake. Instead, the association between variables such as the combination of m_{jet2} together with E_{lep2} showed better classifier performance than considering only the highly significant variable m_{jet2}. In this case the network better adapt differences between both classes.

A grid search is used to find the optimal set of hyperparameters for the training. Table 7.5 shows the applied search range together with the final set of hyperparameters used for the CWoLa MC training. The complete list of variables that were used to train CWoLa on MC is shown in table 7.4. The training is similar to the baseline classifier, inclusive in b-jets. The results of this training are presented in the next section.

7.1.3 Classification Results

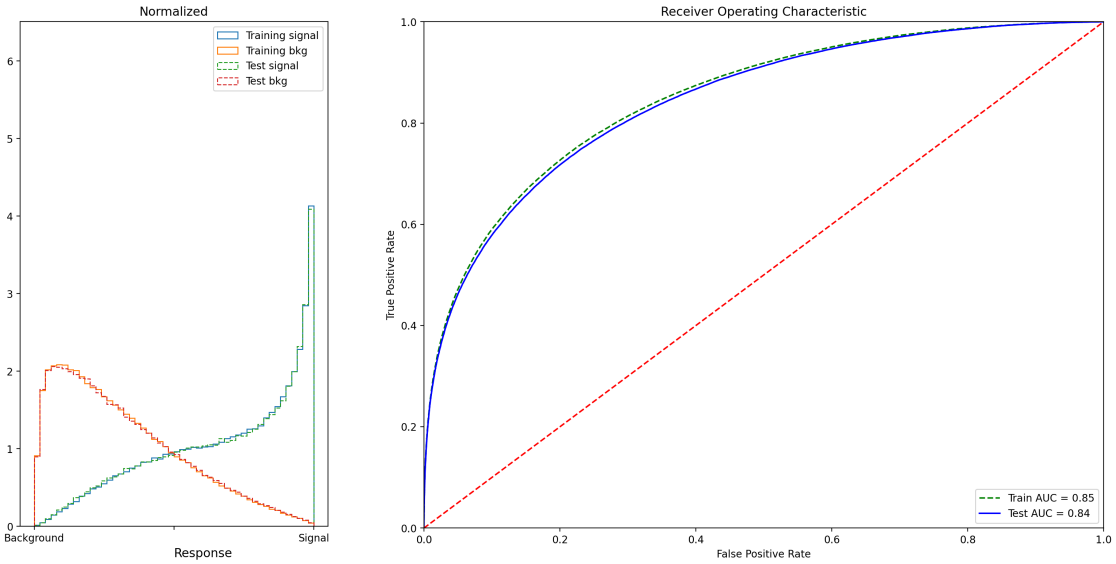
The result of the CWoLa classifier trained on MC shows a great performance compared to the baseline classifer. Compared to the baseline classification results, CWoLa shows a great separation performance for true and fake in the response output in figure 7.3(a). A clear signal-like and background-like behaviour for both curves is visible. The ROC-curve in figure 7.3(b) and the loss-curve in figure 7.3(c) also show a major improvement of classification.

Kinematic Variables	Combined Variables
E_{jet1}	ΔR_{Min}
E_{jet2}	$d\phi_{l1 j_{\text{min}}}$
E_{lep1}	$d\phi_{l2 j_{\text{min}}}$
E_{lep2}	$\text{bscore}_{\text{jet1}}$
η_{lep1}	$\text{bscore}_{\text{jet2}}$
η_{lep2}	
m_{met}	
$m_{\phi_{\text{met}}}$	
m_{jet1}	
m_{jet2}	
ϕ_{jet1}	
ϕ_{jet2}	
ϕ_{lep1}	
ϕ_{lep2}	
$p_{T,\text{jet1}}$	
$p_{T,\text{jet2}}$	

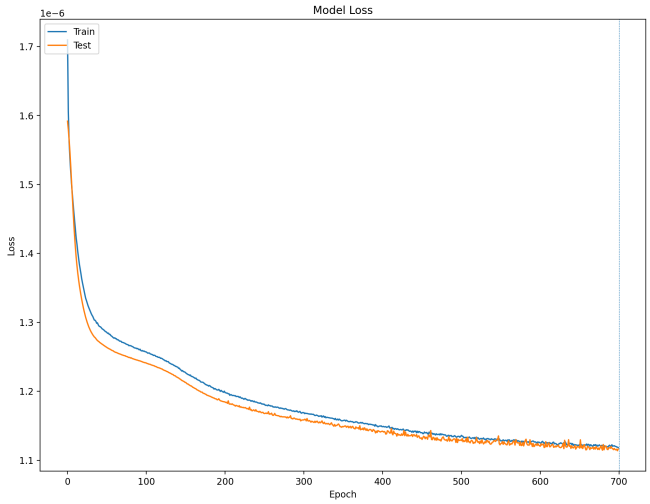
Table 7.4: Selected list of training variables for the CWoLa classifier trained on MC.

Network Hyperparameter	Search Range	Optimal Value
Nodes	50-400	100
Hidden layers	1-5	1
Output layers	1	1
Activation function (hidden layers)	relu,elu,tanh	relu
Batch size	100-20000	10000
Epochs	50-700	700
Optimizer	adam,sgd	adam

Table 7.5: Overview of the search range of hyperparameters for the CWoLa classifier trained on MC, together with the final set of chosen hyperparameters.



(a) (b)



(c)

Figure 7.3: (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve

This result, however, is contrary to expectation, as, in the case of full supervision, the classifier was provided with a complete set of fully separated training samples that contain pure true and fake events, instead of the weak supervision approach where the mixed samples consist of proportions of each class. As described in chapter 5, in case of complete purity of the mixed samples, cwola approaches the full supervision case. From the results of the baseline classifier compared to CWoLa MC, no such behaviour can be seen.

This can be due to some technical issues with the setup of the baseline neural network, the incorrect selection of training variables, since they have the biggest impact on the performance of a neural network or a set of hyperparameters that is not optimal for this classification task. The CWoLa classifier was also tested with the same variables used for the baseline approach, but changing the

variables still could not explain the difference between the two classifiers. In addition, a grid search with an extended search range was performed, which did not produce an improvement.

To test for possible technical issues in the setup of the baseline classifier neural network, a complete new minimal full supervision neural network was set up and trained with the same hyperparameters given in table 6.3 but with a reduced number of training epochs. For this training, only 100 epochs were used. The results show a better performance of classification compared to the previously used baseline network and the results of this training can be seen in figure 7.4. All outputs show a significant improvement over the previously used baseline neural network. The classification performance reaches almost the performance of the CWoLa MC classifier. It can still be improved by using a higher number of training epochs and in general by performing a hyperparameter optimization. A cause of the underperforming of the baseline classifier may therefore indicate a technical issue in setup of the neural network. The next section describes the prediction of the trained CWoLa MC model on unseen MC samples.

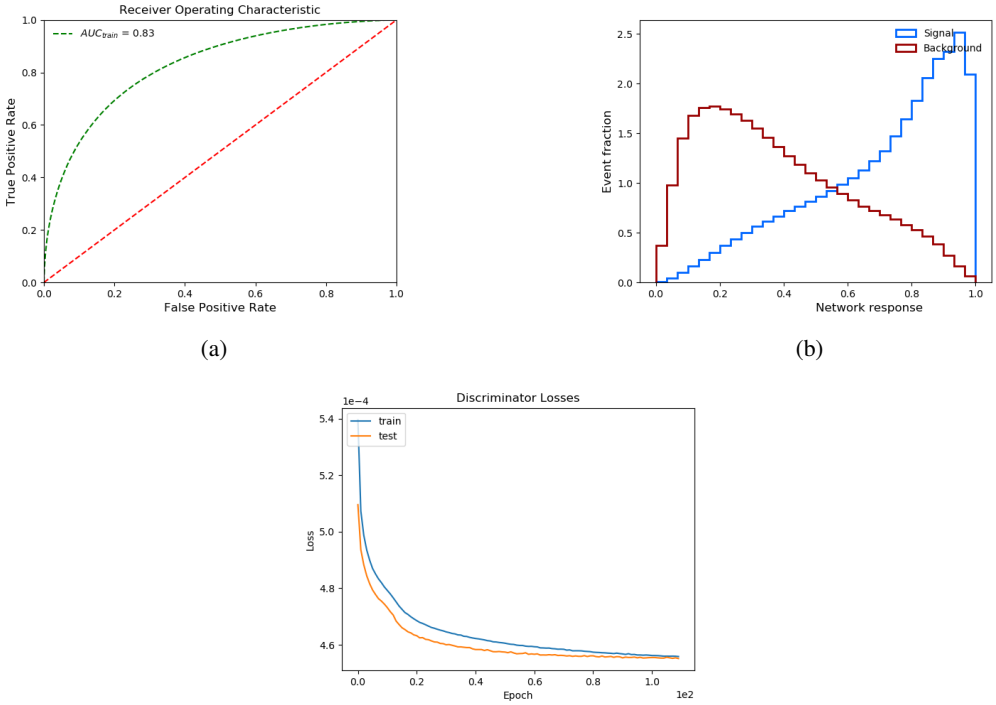
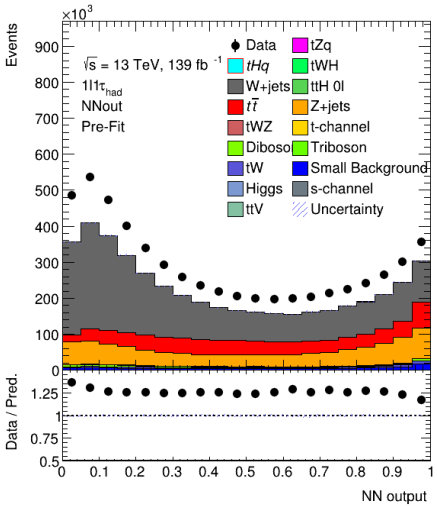


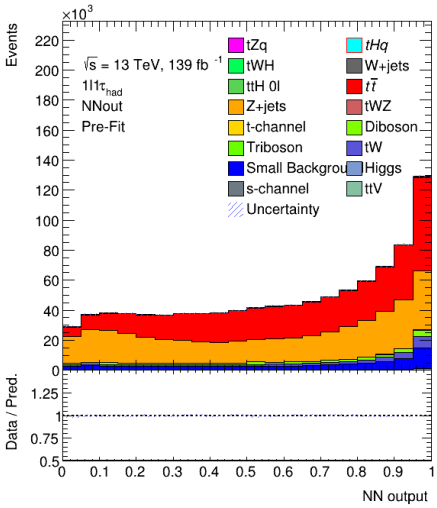
Figure 7.4: (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve

7.1.4 Prediction on unseen MC Samples

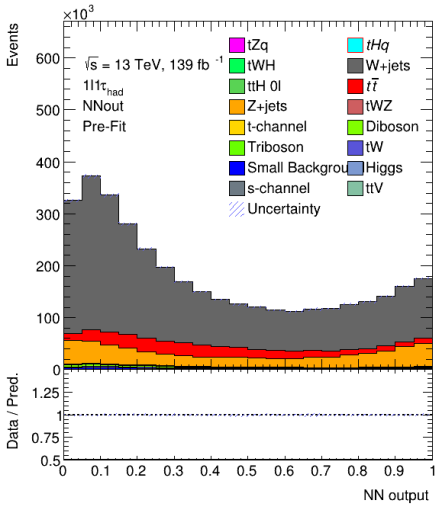
Similar to the prediction of the baseline classifier, described in section 6.4, the CWoLa MC model that is trained on $t\bar{t}$ is now applied to other MC processes to check how the model performs for unseen MC samples. The complete list of used MC samples is shown in section A.1.2 in the appendix. Figure 7.5 shows the neural network output result for the prediction as a stacked plot.



(a) Complete neural network output including true and fake.



(b) CWoLa MC output for true taus only.

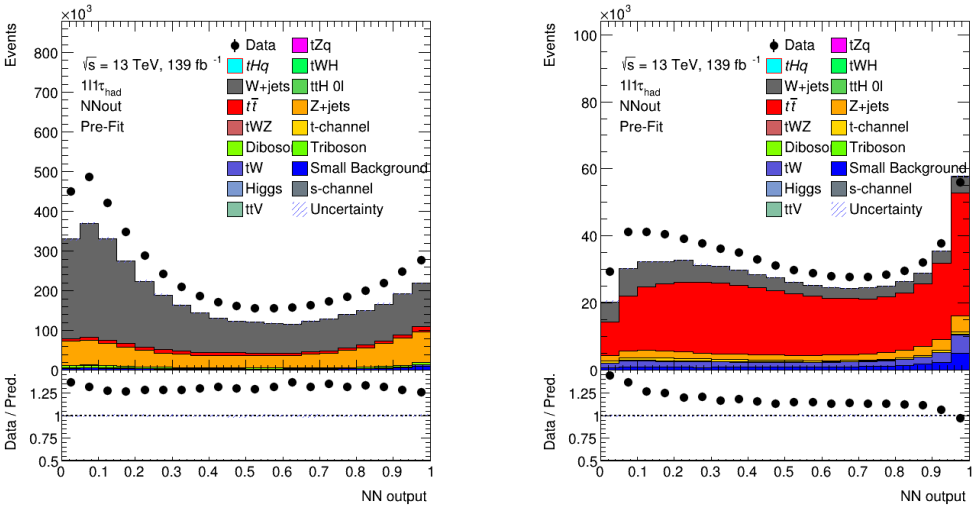


(c) CWoLa MC output for fake taus only.

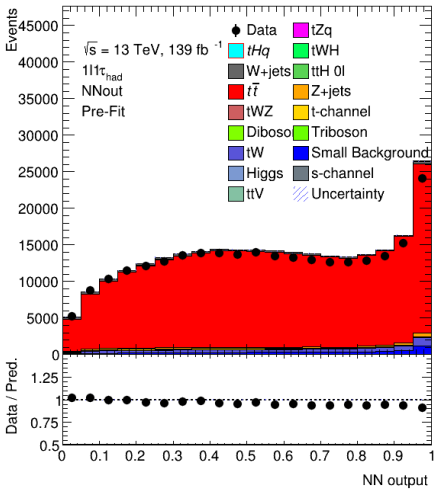
Figure 7.5: Prediction stacked plots for the neural network output on unseen MC samples.

Compared to the prediction output of the baseline classifier, the CWoLa MC classifier shows in figure 7.5(a) a great separation performance for other processes than $t\bar{t}$. CWoLa is therefore able to distinguish true hadronic taus from fake taus. In order to validate this behavior, two separate neural network output prediction stacked plots have been created to show the response for true and fake taus separately, by using the truth information. The response for true taus is shown in figure 7.5(b) and the response for fake taus in figure 7.5(c). These stacked plots show the expected behavior that for true taus the response is more signal-like, while for fake taus the response is background-like. The result shows that the trained CWoLa MC model is able to separate the two classes, true and fake, correctly.

Additionally, one can look at the prediction output for a different number of b-jets, since this CWoLa MC classifier was trained inclusive in b-jets. The stacked plots for 0 b-jets are shown in figure 7.6(a), for 1 b-jet in figure 7.6(b) and for 2 b-jets in figure 7.6(c).



(a) (b)



(c)

Figure 7.6: Prediction stacked plots for the neural network output for three different b-jet configurations for the CWoLa MC classifier: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets.

The results show that the classifier is also able to distinguish true hadronic taus from fake taus for different b-jet configurations. To optimize the results for the CWoLa MC classifier, one could use more statistics in order to improve the separation. An extended grid search can also contribute to a better result. The next section describes a CWoLa classifier that is trained on data and presents, similar to the baseline and CWoLa MC approach, the training results as well as the prediction on MC. Additionally, section A.4 shows the prediction results for CWoLa MC for different regions separated by jet and b-jet multiplicities. These results show the limitations for CWoLa. Regarding [94], when the strategy of CWoLa is employed for training in one event topology and tested in another event topology, there may be systematic uncertainties associated with the extrapolation, which could be seen from the prediction plot of different jet and b-jet multiplicities.

7.2 CWoLa trained on Data Samples

The use of simulated samples in neural network training has the disadvantage that classification on real data can fail because the classifier is too specialized in simulated samples. As described in chapter 5, the great advantage of CWoLa is that one does not have to rely on simulated samples, but can train directly on data. In this section, we describe a CWoLa classifier that is trained directly on data.

The first section shows the available training data on which the classifier was trained. Then the training variables are discussed and the results of the classification are shown. In the last part, the trained model is applied to simulated MC samples.

7.2.1 Mixed Samples for Training

For the CWoLa classifier trained on data, the list of available data samples is given in section A.1.1 in the appendix. Similar to the previous approach, where CWola was trained on MC, the cut on the RNNScore is applied to create two mixed samples. The difference to CWoLa MC is that there is no truth information in data to determine the ratio of true and fake in each sample. The only information that can be extracted from the mixed samples is that one sample must be enriched by true and one sample by fake taus, as this can be seen from the distribution of the RNNScore. Table 7.6(b) summarizes the event yields for each mixed sample after applying the cut on the RNNScore.

Year	Data	Event yields
2015	data15	22787
2016	data16	192682
2017	data17	330555
2017	data18	430859
Total		976883

(a) Mixed Sample M1 with an applied cut of RNN < 0.19.

Year	Data	Event yields
2015	data15	3067
2016	data16	26596
2017	data17	40501
2017	data18	54612
Total		124776

(b) Mixed Sample M2 with an applied cut of RNN > 0.90.

Table 7.6: Event yields for each data sample after RNN selection. Table 7.6(a) shows the event yields for mixed sample M1 with the applied cut RNN < 0.19, table 7.6(b) shows the event yields for mixed sample M2 with the applied cut RNN > 0.90.

7.2.2 Selection of Training Variables and CWoLa Network Architecture

The same variables were used for CWola trained on data as for the CWoLa approach trained on MC, as this allows a direct performance comparison between the two approaches. These variables are summarized in table 7.7.

The hyperparameters for the training have been optimized using a grid search. Both, the search range for each hyperparameter as well as the optimal value that is taken for training is given in table 7.8. It can be seen that the complexity of the network has extremely decreased compared to the previous approach. As in the previous approach, the training has been performed inclusive in b-jets as well. The next section shows the result of the training.

Kinematic Variables	Combined Variables
E_{jet1}	ΔR_{Min}
E_{jet2}	$d\phi_{l1 j_{\text{min}}}$
E_{lep1}	$d\phi_{l2 j_{\text{min}}}$
E_{lep2}	$\text{bscore}_{\text{jet1}}$
η_{lep1}	$\text{bscore}_{\text{jet2}}$
η_{lep2}	
m_{met}	
$m_{\phi_{\text{met}}}$	
m_{jet1}	
m_{jet2}	
ϕ_{jet1}	
ϕ_{jet2}	
ϕ_{lep1}	
ϕ_{lep2}	
$p_{T,\text{jet1}}$	
$p_{T,\text{jet2}}$	

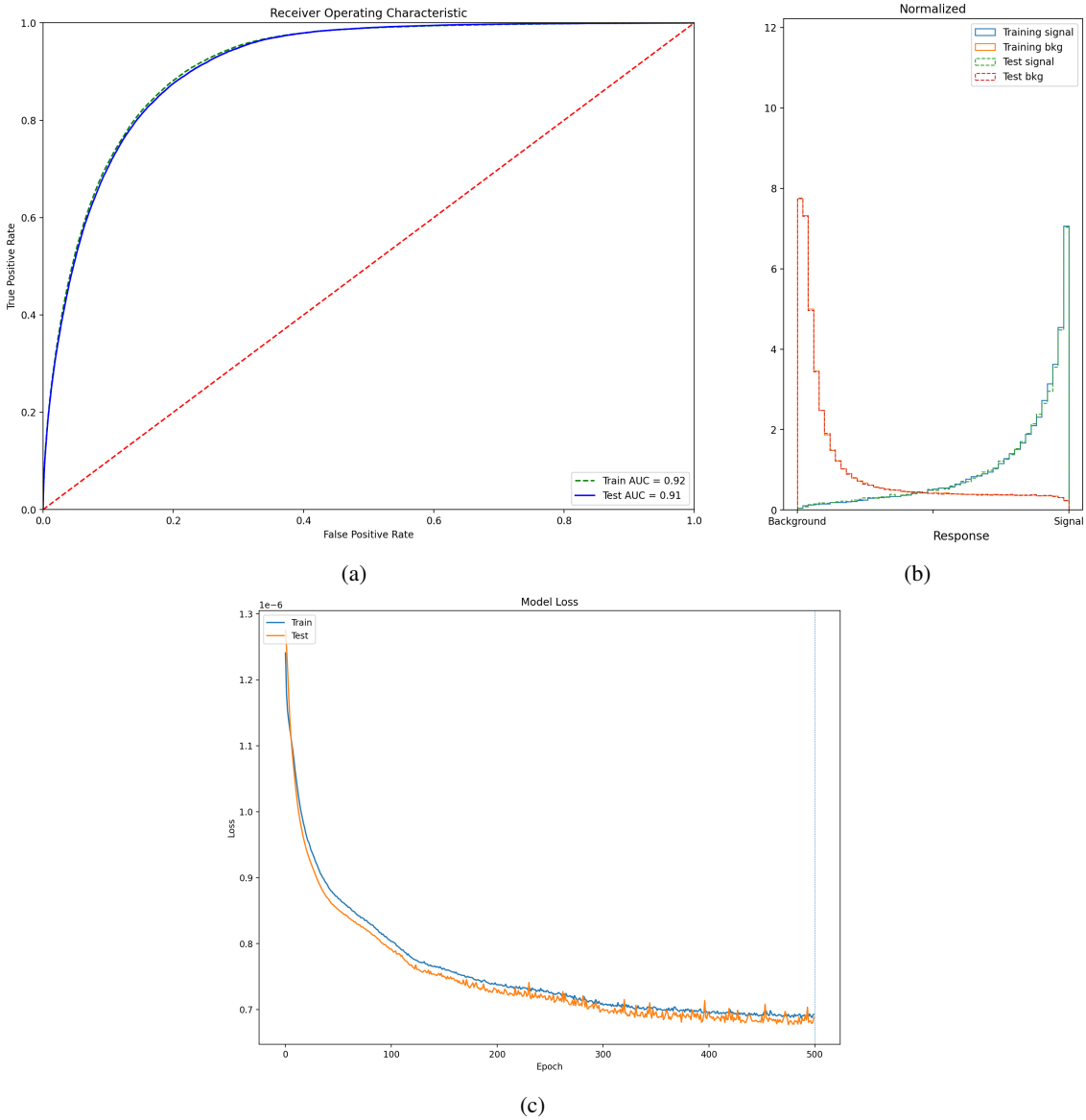
Table 7.7: Selected list of training variables for the CWoLa classifier trained on data.

Network Hyperparameter	Search Range	Optimal Value
Nodes	50-400	40
Hidden layers	1-5	1
Output layers	1	1
Activation function (hidden layers)	relu,elu,tanh	relu
Batch size	100-20000	10000
Epochs	30-500	500
Optimizer	adam,sgd	adam

Table 7.8: Overview of the search range of hyperparameters for the CWoLa classifier trained on data, together with the final set of chosen hyprparameters for training.

7.2.3 Classification Results

This section presents the training results of the CWoLa classifier trained on data. Training on data shows a remarkable performance increase compared to the baseline and the CWoLa MC approach. The corresponding training output curves are given in figure 7.7. Figure 7.7(b) shows a clear separation for true and fake and the high AUC value in figure 7.7(a) indicates that the model correctly predicts the desired output. The model loss, given in figure 7.7(c), also shows that the score is decreasing to a point of stability as one would expect for a good learning curve.



(a) (b)

(c)

Figure 7.7: (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve

The reason why the separation for the CWoLa on data classifier is significantly better is because of the training data itself. In the previous approaches, the classifiers were only trained on $t\bar{t}$ MC samples. The CWoLa on data classifier, however, was trained on data which includes the contribution from all other processes to the training. This added value of information helps the classifier to better distinguish true hadronic taus from fake taus and increase the overall classification performance. In the next section, the model is applied on MC samples.

7.2.4 Prediction on unseen MC Samples

Since the CWoLa classifier has been trained on data, the model can be applied on MC. The MC samples that were used for the prediction are similar to the MC samples used for the baseline and CWoLa MC approach and are listed in section A.1.2 in the appendix. The result for the prediction on unseen MC samples can be seen in figure 7.8. Because of the missing truth label in data, the neural network output prediction can not be separated into predictions for true and fake only. The prediction in figure 7.8 shows a great performance of CWoLa trained on data to distinguish true hadronic taus from fake taus, which can be seen from the signal-like and background-like behavior of the output response.

In addition to the complete prediction output, a selection cut has been applied for different b-jet configurations. The results for a different number of b-jets can be seen in figure 7.9. One can see for different b-jet configurations, that the classifier separates best for 0-bjets, while the separation for 1 and 2-bjets fails. This can be due to effects like mis-modeling of Monte Carlo simulation samples. The CWoLa model trained on data could still be improved by performing a deeper grid search with a larger search range for hyperparameters. For future analysis, it would be also interesting to see how the classifier performs on different channels such as $2l1\tau_{\text{had}}$.

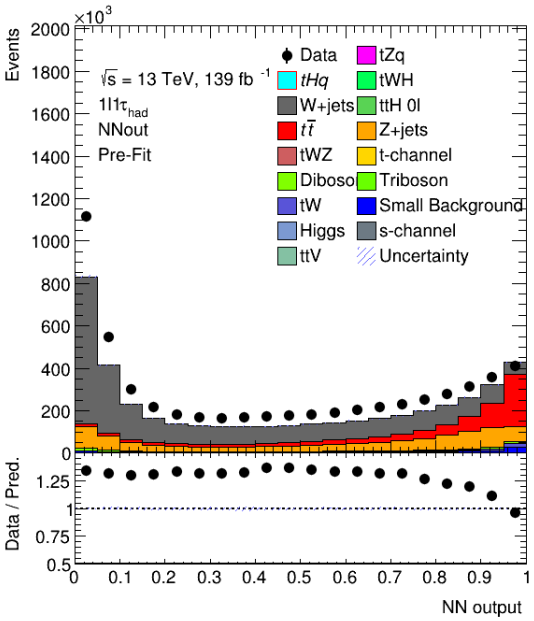
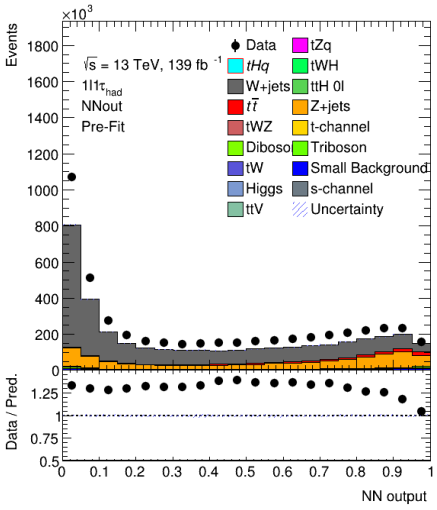
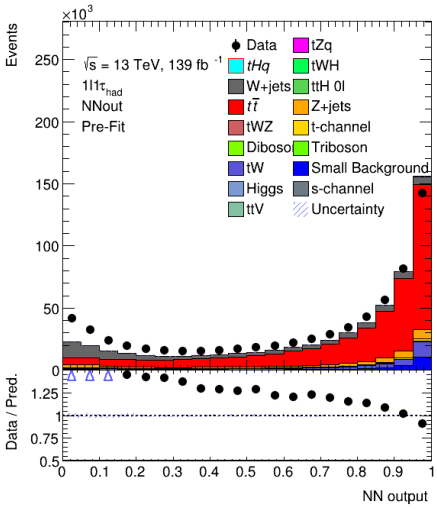


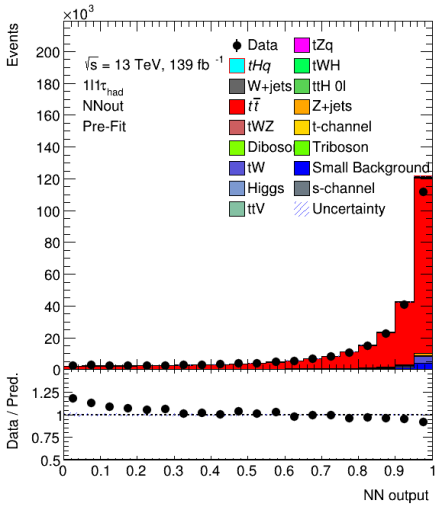
Figure 7.8: Prediction stacked plot for the CWoLa classifier trained on data for unseen MC samples.



(a) 0 b-jets



(b) 1 b-jet



(c) 2 b-jets

Figure 7.9: Prediction stacked plots for the neural network output for three different b-jet configurations for the CWoLa data classifier: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets.

Despite the power of the CWoLa approach, there is a limitation that one has to keep in mind. When one class does not exist in the data, CWoLa does not apply which means that a search for new physics with an exotic signature can not be performed [94]. Further limitations are described in [94].

CHAPTER 8

Summary and Conclusion

The goal of this analysis was to set up a neural network for the 1l1t channel that classifies one hadronically decaying tau lepton. The different approaches have been used that are summarized in the following paragraph.

In the first approach, a baseline neural network classifier was implemented that uses full supervision to distinguish true hadronic taus from fake taus. The idea of the baseline classifier was to set a baseline to see which classification performance can be achieved by using full supervision. This result can then be used as a baseline to compare it to a weak supervision approach.

For the baseline classifier, $t\bar{t}$ MC samples from the $1e/\mu + 1\tau_{\text{had}}$ channel were used. By using the truth information, the $t\bar{t}$ MC samples for the baseline classifier were separated into one sample containing only true taus, and another sample containing only fake taus. These two complete separate samples were used for training and the results were presented. To check the performance of the trained baseline classifier model, it was also applied to MC samples that were unseen during the training. The results showed a poor separation performance due to technical issues in the setup of the neural network which could be solved by an entirely new set up neural network. However this classifier

The second approach is a weak supervision approach which is called *Classification Without Labels* (CWoLa). Instead of two fully separated training samples for true and fake, this technique uses statistical mixtures of events. These mixed samples were created by applying a cut on the RNNScore. Two different training setups were performed. CWoLa was trained on $t\bar{t}$ MC samples from the $1e/\mu + 1\tau_{\text{had}}$ channel. The applied cut separates these $t\bar{t}$ MC samples into two mixed samples with different proportions of true hadronic taus and fake taus: one sample enriched by true taus, the other one enriched by fake taus. The CWoLa classifier was then trained to distinguish these statistical mixtures of classes. The classification showed a great separation performance between both classes which could also be verified with a prediction on a set of unseen MC samples. In addition, the limitations of CWoLa were discussed.

Instead of training on $t\bar{t}$ MC samples, CWoLa was trained directly on data. The results showed a great separation performance between true and fake which is because both, the baseline classifier and the CWoLa MC classifier use $t\bar{t}$ MC samples as training input. Using data includes the contributions not only from $t\bar{t}$ but also from the other processes which lead to better classification performance. The performance of the CWoLa classifier trained on data was verified with a prediction on MC samples.

With the limitations of the CWoLa approach in mind, the analysis in this thesis showed that CWoLa is a powerful tool that can be used to distinguish true hadronic taus from fake taus. Nevertheless,

hyperparameter optimization needs to be performed for every neural network and especially for the new set up baseline classifier neural network, because the result was still poor compared to the weak supervision approach, which is contrary to expectation. Also, the classifier should be applied to other channels like $2l1\tau$ to see how the classifier performs on channels other than $1e/\mu + 1\tau_{\text{had}}$.

Bibliography

[1] C. D. Galbraith, *Standard Infographic of the Standard Model. CERN Webfest.* URL: http://cds.cern.ch/record/1473657/files/ (cit. on p. 4).

[2] S. Mele, *The Measurement of the Number of Light Neutrino Species at LEP*, URL: https://cds.cern.ch/record/2103251/files/9789814644150_0004.pdf?subformat=pdfa&version=1 (cit. on p. 5).

[3] M. P. Hertzberg and M. Sandora, *General relativity from causality*, Journal of High Energy Physics **2017** (2017), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP09(2017)119 (cit. on p. 6).

[4] I. C. Brock and T. Schorner-Sadenius, *Physics at the terascale*, Wiley, 2011, URL: https://cds.cern.ch/record/1354959 (cit. on pp. 6, 22).

[5] S. Descotes-Genon and P. Koppenburg, *The CKM Parameters*, Annual Review of Nuclear and Particle Science **67** (2017) 97, ISSN: 1545-4134, URL: http://dx.doi.org/10.1146/annurev-nucl-101916-123109 (cit. on p. 6).

[6] S. Descotes-Genon and P. Koppenburg, *The CKM Parameters*, Annual Review of Nuclear and Particle Science **67** (2017) 97, ISSN: 1545-4134, URL: http://dx.doi.org/10.1146/annurev-nucl-101916-123109 (cit. on p. 7).

[7] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001, URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001 (cit. on pp. 7, 10, 12, 14, 15, 17, 19, 39).

[8] B. Povh et al., *Povh, B: Teilchen und Kerne.: Eine Einführung in die physikalischen Konzepte.* Physics and astronomy online library, Springer, 2004, ISBN: 9783540210658, URL: https://books.google.de/books?id=PIw0leNupPoC (cit. on pp. 7, 8).

[9] C. Loizides, *Jet physics in ALICE*, (2005) (cit. on p. 8).

[10] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-107-03426-6 (cit. on p. 10).

[11] Z. M. and, *Simulation of Pile-up in the ATLAS Experiment*, Journal of Physics: Conference Series **513** (2014) 022024, URL: https://doi.org/10.1088/1742-6596/513/2/022024 (cit. on p. 11).

[12] URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Pileup_Interactions_and_Data_Tak (cit. on p. 11).

[13] F. Abe et al.,
Observation of Top Quark Production inp¯pCollisions with the Collider Detector at Fermilab,
Physical Review Letters **74** (1995) 2626, ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.74.2626 (cit. on p. 12).

[14] R. Hawkings, *Top quark physics at the LHC*,
Comptes Rendus Physique **16** (2015) 424, Highlights of the LHC run 1 / Résultats marquants de la première période d'exploitation du GCH, ISSN: 1631-0705, URL:
https://www.sciencedirect.com/science/article/pii/S1631070515000766
(cit. on p. 12).

[15] F. Déliot et al., *Properties of the Top Quark*,
Annual Review of Nuclear and Particle Science **64** (2014) 363, ISSN: 1545-4134,
URL: http://dx.doi.org/10.1146/annurev-nucl-102313-025655 (cit. on p. 12).

[16] S.-Q. Wang et al., *A precise determination of the top-quark pole mass*,
The European Physical Journal C **78** (2018), ISSN: 1434-6052,
URL: http://dx.doi.org/10.1140/epjc/s10052-018-5688-1 (cit. on p. 12).

[17] A. M. Sirunyan et al., *Measurement of the top quark Yukawa coupling from tt¯kinematic distributions in the lepton+jets final state in proton-proton collisions at s=13 TeV*,
Physical Review D **100** (2019), ISSN: 2470-0029,
URL: http://dx.doi.org/10.1103/PhysRevD.100.072007 (cit. on p. 12).

[18] A. Collaboration, *Summary of the ATLAS and CMS direct mtop measurements*,
URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/
LHCTopWGSummaryPlots#Top_Quark_Mass (cit. on pp. 13, 15).

[19] Priyanka, K. Ranjan and A. Bhardwaj, *Single-top quark production at CMS*,
Universe **5** (2019) 19 (cit. on pp. 13, 14).

[20] M. Gallinaro, *Top quark physics: A tool for discoveries*,
Journal of Physics Conference Series **447** (2013) 2012 (cit. on p. 14).

[21] URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/
TopPairCrossectionSqrtsHistory (cit. on pp. 14, 16).

[22] T. M. P. Tait and C.-P. Yuan,
Single top quark production as a window to physics beyond the standard model,
Physical Review D **63** (2000), ISSN: 1089-4918,
URL: http://dx.doi.org/10.1103/PhysRevD.63.014018 (cit. on p. 14).

[23] T. Aaltonen et al., *Observation of Electroweak Single Top-Quark Production*,
Phys. Rev. Lett. **103** (9 2009) 092002,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.092002 (cit. on p. 14).

[24] V. M. Abazov et al., *Observation of Single Top-Quark Production*,
Phys. Rev. Lett. **103** (9 2009) 092001,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.092001 (cit. on p. 14).

[25] M. Moreno Llacer,
Studies of rare top quark processes: tZq, $t\gamma q$, $t\bar{t}t\bar{t}$ and flavour changing neutral currents,
PoS **LHCP2019** (2019) 106. 12 p, URL: http://cds.cern.ch/record/2708469
(cit. on pp. 16, 17).

[26] M. Aaboud et al., *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173, ISSN: 0370-2693,
URL: http://dx.doi.org/10.1016/j.physletb.2018.07.035 (cit. on p. 17).

[27] *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*,
tech. rep., All figures including auxiliary figures are available at
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-045: CERN, 2019, URL: https://cds.cern.ch/record/2693930
(cit. on p. 17).

[28] URL: https://atlas.cern/updates/briefing/evidence-four-top-quark-production (cit. on p. 17).

[29] 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*, Presented 24 Jul 2018, 2018,
URL: http://cds.cern.ch/record/2654218 (cit. on pp. 17, 24, 28, 42, 43).

[30] M. Alhroob et al.,
Observation of the associated production of a top quark and a Z boson at 13 TeV with ATLAS,
tech. rep., CERN, 2018, URL: https://cds.cern.ch/record/2304824
(cit. on pp. 17, 42).

[31] C. Collaboration et al., *Search for Associated Production of a Z Boson with a Single Top Quark and for tZ Flavour-Changing Interactions in Pp Collisions at $\sqrt{s} = 8$ TeV*, (2017)
(cit. on p. 17).

[32] M. Aaboud et al., *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://dx.doi.org/10.1016/j.physletb.2018.03.023 (cit. on p. 17).

[33] G. Aad et al., *Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s}13TeV$ with the ATLAS detector*, Journal of High Energy Physics **2020** (2020),
ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP07(2020)124 (cit. on p. 18).

[34] A. M. Sirunyan et al., *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}13TeV$*, Physical Review Letters **122** (2019),
ISSN: 1079-7114, URL: http://dx.doi.org/10.1103/PhysRevLett.122.132003
(cit. on p. 18).

[35] V. Khachatryan et al., *Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ GeV*,
Journal of High Energy Physics **2016** (2016), ISSN: 1029-8479,
URL: http://dx.doi.org/10.1007/JHEP06(2016)177 (cit. on pp. 18, 19, 39).

[36] URL: https://publications.rwth-aachen.de/record/762636/files/762636.pdf
(cit. on p. 19).

[37] D. Chakraborty, J. Konigsberg and D. Rainwater, *TOP-QUARK PHYSICS*, Annual Review of Nuclear and Particle Science **53** (2003) 301, ISSN: 1545-4134, URL: http://dx.doi.org/10.1146/annurev.nucl.53.041002.110601 (cit. on p. 19).

[38] M. Perl, *The discovery of the tau lepton and the changes in elementary particle physics in 40 years*, (2003) (cit. on p. 19).

[39] M. L. Perl, *The Discovery of the tau lepton and the changes in elementary particle physics in forty years*, Phys. Perspect. **6** (2004) 401 (cit. on p. 19).

[40] Martin Johannes Schultens, *Search for Supersymmetry with Hadronically and Leptonically Decaying Tau Leptons at the ATLAS Experiment*, PhD thesis: Rheinische Friedrich-Wilhelms-Universität Bonn, 2017, URL: http://hdl.handle.net/20.500.11811/7101 (cit. on pp. 20, 27).

[41] F. Friedrich, *Tau Lepton Reconstruction and Identification at ATLAS*, EPJ Web of Conferences **28** (2012) 12007, ISSN: 2100-014X, URL: http://dx.doi.org/10.1051/epjconf/20122812007 (cit. on pp. 20, 32).

[42] G. Aad et al., *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ TeV, The European Physical Journal C **75** (2015), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-015-3500-z (cit. on p. 20).

[43] URL: https://home.cern/science/experiments/atlas (cit. on p. 21).

[44] *The Large Hadron Collider, 2020.* URL: https://home.cern/science/accelerators/large-hadron-collider (cit. on p. 22).

[45] URL: https://home.cern/science/accelerators/high-luminosity-lhc (cit. on p. 23).

[46] URL: https://atlas.cern/discover/detector (cit. on p. 23).

[47] URL: https://cds.cern.ch/record/1095924 (cit. on p. 24).

[48] G. Aad et al., *Determination of the top-quark pole mass using* $t\bar{t}$ *+ 1-jet events collected with the ATLAS experiment in 7 TeV pp collisions*, Journal of High Energy Physics **2015** (2015), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP10(2015)121 (cit. on pp. 23, 24).

[49] *ATLAS inner detector: Technical Design Report, 1*, Technical design report. ATLAS, CERN, 1997, URL: https://cds.cern.ch/record/331063 (cit. on p. 24).

[50] A. Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, tech. rep., CERN, 2013, URL: https://cds.cern.ch/record/1537991 (cit. on p. 24).

[51] T. A. collaboration, *Operation and performance of the ATLAS semiconductor tracker*, Journal of Instrumentation **9** (2014) P08009, ISSN: 1748-0221, URL: http://dx.doi.org/10.1088/1748-0221/9/08/P08009 (cit. on p. 24).

[52] A. Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, Eur. Phys. J. C **80** (2020) 1194. 41 p, arXiv: 2007.07624, URL: https://cds.cern.ch/record/2724037 (cit. on p. 24).

[53] URL: https://cds.cern.ch/record/1095926 (cit. on p. 25).

[54] URL: https://atlas.cern/discover/detector/inner-detector (cit. on p. 25).

[55] H. Pernegger, *The Pixel Detector of the ATLAS experiment for LHC Run-2*, Journal of Instrumentation **10** (2015) C06012, URL: https://doi.org/10.1088/1748-0221/10/06/c06012 (cit. on p. 25).

[56] URL: https://cds.cern.ch/record/2724037 (cit. on p. 26).

[57] D. Ferrère, "The Atlas Silicon Microstrip Tracker Construction Status", *Advanced Technology - Particle Physics*, 2002 196 (cit. on p. 25).

[58] URL: https://www.fsp103-atlas.de/e17619/e17620/ (cit. on p. 25).

[59] A. Abdesselam et al., *Engineering for the ATLAS SemiConductor Tracker (SCT) End-cap*, Journal of Instrumentation **3** (2008) P05002, URL: https://doi.org/10.1088/1748-0221/3/05/p05002 (cit. on p. 25).

[60] *Particle Identification Performance of the ATLAS Transition Radiation Tracker*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-128: CERN, 2011, URL: https://cds.cern.ch/record/1383793 (cit. on p. 26).

[61] T. A. T. collaboration et al., *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance*, Journal of Instrumentation **3** (2008) P02013, URL: https://doi.org/10.1088/1748-0221/3/02/p02013 (cit. on p. 26).

[62] URL: https://atlas.cern/discover/detector/calorimeter (cit. on p. 27).

[63] C. W. Fabjan and D. Fournier, "Calorimetry", *Particle Physics Reference Library: Volume 2: Detectors for Particles and Radiation*, ed. by C. W. Fabjan and H. Schopper, Springer International Publishing, 2020 201, ISBN: 978-3-030-35318-6, URL: https://doi.org/10.1007/978-3-030-35318-6_6 (cit. on p. 27).

[64] G. Aad et al., *The ATLAS experiment at the CERN large hadron collider*, Journal of Instrumentation **3** (2008) S08003 (cit. on pp. 27, 28).

[65] S. Aefsky, *Alignment of the Muon Spectrometer in ATLAS*, tech. rep., CERN, 2011, URL: https://cds.cern.ch/record/1380912 (cit. on p. 28).

[66] E_T^{miss} *performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2018-023: CERN, 2018, URL: https://cds.cern.ch/record/2625233 (cit. on p. 28).

[67] URL: https://www.brock.physik.uni-bonn.de/research/atlas-experiment (cit. on p. 28).

[68] T. A. collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** (2020) P10004, URL: https://doi.org/10.1088/1748-0221/15/10/p10004 (cit. on pp. 28, 29).

[69] R. Keyes, *Development, validation and integration of the ATLAS Trigger System software in Run 2*, tech. rep., CERN, 2016, URL: https://cds.cern.ch/record/2239296 (cit. on pp. 28, 29).

[70] A. Ruiz-Martinez and A. Collaboration, *The Run-2 ATLAS Trigger System*, tech. rep., CERN, 2016, URL: https://cds.cern.ch/record/2133909 (cit. on pp. 28, 29).

[71] M. Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV*, The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-019-7140-6 (cit. on p. 29).

[72] URL: https://cms.cern/news/finding-electrons-and-photons-cms-detector (cit. on p. 29).

[73] URL: https://cds.cern.ch/record/1505342 (cit. on p. 30).

[74] G. Aad et al., *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ s =13 TeV*, The European Physical Journal C **76** (2016), ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-016-4120-y (cit. on p. 30).

[75] E. Moyse, *Muon Track Reconstruction in the ATLAS Experiment*, (2009), URL: https://cds.cern.ch/record/1222299 (cit. on p. 30).

[76] URL: https://cms.cern/tags/particle-jet (cit. on p. 31).

[77] S. Schramm, *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2291608 (cit. on p. 31).

[78] M. Cacciari, G. P. Salam and G. Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479, URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063 (cit. on p. 31).

[79] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, The European Physical Journal C **77** (2017) (cit. on p. 31).

[80] *Performance of b-jet identification in the ATLAS experiment*, Journal of Instrumentation **11** (2016) P04008, ISSN: 1748-0221, URL: http://dx.doi.org/10.1088/1748-0221/11/04/P04008 (cit. on p. 31).

[81] M. Lehmacher, *b-Tagging Algorithms and their Performance at ATLAS*, 2008, arXiv: 0809.4896 [hep-ex] (cit. on pp. 31, 32).

[82] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, tech. rep., All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-012: CERN, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on p. 32).

[83] *Identification of the Hadronic Decays of Tau Leptons in 2012 Data with the ATLAS Detector*, tech. rep. ATLAS-CONF-2013-064, CERN, 2013, URL: https://cds.cern.ch/record/1562839 (cit. on p. 32).

[84] URL: http://cds.cern.ch/record/1754701/plots (cit. on p. 32).

[85] A. Pingel, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment*, Nuclear and Particle Physics Proceedings **273-275** (2016) 1141, 37th International Conference on High Energy Physics (ICHEP), ISSN: 2405-6014, URL: https://www.sciencedirect.com/science/article/pii/S2405601415006689 (cit. on p. 32).

[86] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, (2015) (cit. on pp. 32–34).

[87] URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-045/ (cit. on p. 35).

[88] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, (2019) (cit. on pp. 34, 36).

[89] C. A. Lee, *MissingpTReconstruction at ATLAS*, Journal of Physics: Conference Series **645** (2015) 012012, URL: https://doi.org/10.1088/1742-6596/645/1/012012 (cit. on p. 37).

[90] V. Khachatryan et al., *Search for supersymmetry in events with photons and missing transverse energy in pp collisions at 13 TeV*, Physics Letters B **769** (2017) 391, ISSN: 0370-2693, URL: http://dx.doi.org/10.1016/j.physletb.2017.04.005 (cit. on p. 37).

[91] 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. 37).

[92] G. Aad et al., *The ATLAS Simulation Infrastructure*, The European Physical Journal C **70** (2010) 823, ISSN: 1434-6052, URL: http://dx.doi.org/10.1140/epjc/s10052-010-1429-9 (cit. on p. 42).

[93] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688 (cit. on p. 43).

[94] E. M. Metodiev, B. Nachman and J. Thaler, *Classification without labels: learning from mixed samples in high energy physics*, Journal of High Energy Physics **2017** (2017), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP10(2017)174 (cit. on pp. 45, 55–58, 75, 80).

[95] A. L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal of Research and Development **3** (1959) 210 (cit. on p. 46).

[96] V. K. Vemuri, *The Hundred-Page Machine Learning Book*,
Journal of Information Technology Case and Application Research **22** (2020) 136,
eprint: https://doi.org/10.1080/15228053.2020.1766224,
URL: https://doi.org/10.1080/15228053.2020.1766224 (cit. on pp. 46, 49).

[97] S. S. Haykin, *Neural networks and learning machines*, Third, Pearson Education, 2009 (cit. on p. 47).

[98] M. Deru, A. Ndiaye and G. Press, *Deep Learning mit TensorFlow, Keras und TensorFlow.js*, Onleihe. E-Book, Rheinwerk Verlag GmbH, 2019, ISBN: 9783836265119,
URL: https://books.google.de/books?id=MiNPzQEACAAJ (cit. on p. 48).

[99] X. Glorot, A. Bordes and Y. Bengio, "Deep Sparse Rectifier Neural Networks", *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*,
ed. by G. Gordon, D. Dunson and M. Dudík, vol. 15,
Proceedings of Machine Learning Research, PMLR, 2011 315,
URL: http://proceedings.mlr.press/v15/glorot11a.html (cit. on p. 51).

[100] URL: http://neuralnetworksanddeeplearning.com/chap3.html (cit. on pp. 52, 53).

[101] M. J. Kochenderfer and T. A. Wheeler, *Algorithms for Optimization*, The MIT Press, 2019, ISBN: 0262039427 (cit. on p. 53).

[102] R. Yedida, S. Saha and T. Prashanth,
LipschitzLR: Using theoretically computed adaptive learning rates for fast convergence, 2020,
arXiv: 1902.07399 [cs.LG] (cit. on p. 53).

[103] URL: https://machinelearning.tf.fau.de/pdfs/mlisp/LectureNN3.pdf (cit. on p. 53).

[104] URL: https://keras.io/api/optimizers/ (cit. on p. 54).

[105] URL: https://www.nnwj.de/backpropagation.html (cit. on p. 54).

[106] D. E. Rumelhart, G. E. Hinton and R. J. Williams,
Learning Representations by Back-propagating Errors, Nature **323** (1986) 533,
URL: http://www.nature.com/articles/323533a0 (cit. on p. 54).

[107] URL: https://pgaleone.eu/deep-learning/regularization/2017/01/10/anaysis-of-dropout/ (cit. on p. 54).

[108] 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. 54).

[109] J. Neyman and E. S. Pearson,
On the Problem of the Most Efficient Tests of Statistical Hypotheses,
Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933) 289, ISSN: 02643952,
URL: http://www.jstor.org/stable/91247 (cit. on p. 55).

[110] J. Neyman and E. S. Pearson,
"On the Problem of the Most Efficient Tests of Statistical Hypotheses",
Breakthroughs in Statistics: Foundations and Basic Theory, ed. by S. Kotz and N. L. Johnson,
Springer New York, 1992 73, ISBN: 978-1-4612-0919-5,
URL: https://doi.org/10.1007/978-1-4612-0919-5_6 (cit. on p. 55).

[111] Z.-H. Zhou, *A brief introduction to weakly supervised learning*,
National Science Review **5** (2017) 44, ISSN: 2095-5138, eprint:
https://academic.oup.com/nsr/article-pdf/5/1/44/31567770/nwx106.pdf,
URL: https://doi.org/10.1093/nsr/nwx106 (cit. on p. 55).

[112] C. Scott and J. Zhang,
Learning from Label Proportions: A Mutual Contamination Framework, 2020,
arXiv: 2006.07330 [stat.ML] (cit. on p. 56).

[113] P. T. Komiske et al., *Learning to classify from impure samples with high-dimensional data*,
Physical Review D **98** (2018), ISSN: 2470-0029,
URL: http://dx.doi.org/10.1103/PhysRevD.98.011502 (cit. on p. 56).

[114] URL: https://gitlab.cern.ch/fdiazcap/cwola-fakes (cit. on p. 57).

[115] URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SgTopRun2NtuplesSamples
(cit. on p. 105).

APPENDIX A

Appendix

A.1 List of Data and Monte Carlo (MC) samples

A.1.1 Data Samples

Process	Data DSID	Generator
Data	data15	grp15_v01_p4173
	data16	grp16_v01_p4173
	data17	grp17_v01_p4173
	data18	grp18_v01_p4173

Table A.1: $t\bar{t}$ Data Samples (Run 2)

A.1.2 Monte Carlo Samples

Process	MC DSID	Generator	$\sigma[pb]$	k-factor
t-channel	410658	e6671_s3126_r9364_p4174	36.99	1.19
	410658	e6671_s3126_r10201_p4174	36.99	1.19
	410658	e6671_s3126_r10724_p4174	36.99	1.19
	410659	e6671_s3126_r9364_p4174	22.18	1.18
	410659	e6671_s3126_r10201_p4174	22.18	1.18
	410659	e6671_s3126_r10724_p4174	22.18	1.18
tW	410646	e6552_s3126_r9364_p4174	37.94	0.94
	410646	e6552_s3126_r10201_p4174	37.94	0.94
	410646	e6552_s3126_r10724_p4174	37.94	0.94
	410647	e6552_s3126_r9364_p4174	37.91	0.95
	410647	e6552_s3126_r10201_p4174	37.91	0.95
	410647	e6552_s3126_r10724_p4174	37.91	0.95
tWZ	412118	e7518_s3126_r9364_p4174	0.02	1.00
	412118	e7518_s3126_r10201_p4174	0.02	1.00

	412118	e7518_s3126_r10724_p4174	0.02	1.00
ttV	410155	e5070_s3126_r9364_p4174	0.55	1.10
	410155	e5070_s3126_r10201_p4174	0.55	1.10
	410155	e5070_s3126_r10724_p4174	0.55	1.10
	410156	e5070_s3126_r9364_p4174	0.15	1.11
	410156	e5070_s3126_r10201_p4174	0.15	1.11
	410156	e5070_s3126_r10724_p4174	0.15	1.11
	410157	e5070_s3126_r9364_p4174	0.53	1.11
	410157	e5070_s3126_r10201_p4174	0.53	1.11
	410157	e5070_s3126_r10724_p4174	0.53	1.11
	410218	e5070_s3126_r9364_p4174	0.04	1.12
	410218	e5070_s3126_r10201_p4174	0.04	1.12
	410218	e5070_s3126_r10724_p4174	0.04	1.12
	410219	e5070_s3126_r9364_p4174	0.04	1.12
	410219	e5070_s3126_r10201_p4174	0.04	1.12
	410219	e5070_s3126_r10724_p4174	0.04	1.12
	410220	e5070_s3126_r9364_p4174	0.04	1.12
	410220	e5070_s3126_r10201_p4174	0.04	1.12
	410220	e5070_s3126_r10724_p4174	0.04	1.12
	410276	e6087_s3126_r9364_p4174	0.02	1.00
	410276	e6087_s3126_r10201_p4174	0.02	1.00
	410276	e6087_s3126_r10724_p4174	0.02	1.00
	410277	e6087_s3126_r9364_p4174	0.02	1.00
	410277	e6087_s3126_r10201_p4174	0.02	1.00
	410277	e6087_s3126_r10724_p4174	0.02	1.00
	410278	e6087_s3126_r9364_p4174	0.00	1.00
	410278	e6087_s3126_r10201_p4174	0.00	1.00
	410278	e6087_s3126_r10724_p4174	0.00	1.00
tZq	412063	e7054_s3126_r9364_p4174	0.03	1.00
	412063	e7054_s3126_r10201_p4174	0.03	1.00
	412063	e7054_s3126_r10724_p4174	0.03	1.00
$t\bar{t}$	410470	e6337_s3126_r9364_p4174	396.87	1.14
	410470	e6337_s3126_r10201_p4174	396.87	1.14
	410470	e6337_s3126_r10724_p4174	396.87	1.14
	410471	e6337_s3126_r9364_p4174	332.97	1.19
	410471	e6337_s3126_r10201_p4174	332.97	1.19
	410471	e6337_a875_r10724_p4174	332.97	1.19
tWH	346678	e7816_a875_r9364_p4174	0.02	1.00
	346678	e7816_a875_r10201_p4174	0.02	1.00
	346678	e7816_a875_r10724_p4174	0.02	1.00

tHq	346676	e7815_a875_r9364_p4174	0.06	1.00
	346676	e7815_a875_r10201_p4174	0.06	1.00
	346676	e7815_a875_r10724_p4174	0.06	1.00
Z+jets	364100	e5271_s3126_r9364_p4172	1 630.22	0.98
	364100	e5271_s3126_r10201_p4172	1 630.22	0.98
	364100	e5271_s3126_r10724_p4172	1 630.22	0.98
	364101	e5271_s3126_r9364_p4172	223.72	0.98
	364101	e5271_s3126_r10201_p4172	223.72	0.98
	364101	e5271_s3126_r10724_p4172	223.72	0.98
	364102	e5271_s3126_r9364_p4172	127.18	0.98
	364102	e5271_s3126_r10201_p4172	127.18	0.98
	364102	e5271_s3126_r10724_p4172	127.18	0.98
	364103	e5271_s3126_r9364_p4172	75.02	0.98
	364103	e5271_s3126_r10201_p4172	75.02	0.98
	364103	e5271_s3126_r10724_p4172	75.02	0.98
	364104	e5271_s3126_r9364_p4172	20.35	0.98
	364104	e5271_s3126_r10201_p4172	20.35	0.98
	364104	e5271_s3126_r10724_p4172	20.35	0.98
	364105	e5271_s3126_r9364_p4172	12.39	0.98
	364105	e5271_s3126_r10201_p4172	12.39	0.98
	364105	e5271_s3126_r10724_p4172	12.39	0.98
	364106	e5271_s3126_r9364_p4172	24.29	0.98
	364106	e5271_s3126_r10201_p4172	24.29	0.98
	364106	e5271_s3126_r10724_p4172	24.29	0.98
	364107	e5271_s3126_r9364_p4172	9.28	0.98
	364107	e5271_s3126_r10201_p4172	9.28	0.98
	364107	e5271_s3126_r10724_p4172	9.28	0.98
	364108	e5271_s3126_r9364_p4172	6.01	0.98
	364108	e5271_s3126_r10201_p4172	6.01	0.98
	364108	e5271_s3126_r10724_p4172	6.01	0.98
	364109	e5271_s3126_r9364_p4172	4.77	0.98
	364109	e5271_s3126_r10201_p4172	4.77	0.98
	364109	e5271_s3126_r10724_p4172	4.77	0.98
	364110	e5271_s3126_r9364_p4172	2.27	0.98
	364110	e5271_s3126_r10201_p4172	2.27	0.98
	364110	e5271_s3126_r10724_p4172	2.27	0.98
	364111	e5271_s3126_r9364_p4172	1.49	0.98
	364111	e5271_s3126_r10201_p4172	1.49	0.98
	364111	e5271_s3126_r10724_p4172	1.49	0.98
	364112	e5271_s3126_r9364_p4172	1.79	0.98
	364112	e5271_s3126_r10201_p4172	1.79	0.98
	364112	e5271_s3126_r10724_p4172	1.79	0.98
	364113	e5271_s3126_r9364_p4172	0.15	0.98
	364113	e5271_s3126_r10201_p4172	0.15	0.98

364113	e5271_s3126_r10724_p4172	0.15	0.98
364114	e5299_s3126_r9364_p4172	1 627.18	0.98
364114	e5299_s3126_r10201_p4172	1 627.18	0.98
364114	e5299_s3126_r10724_p4172	1 627.18	0.98
364115	e5299_s3126_r9364_p4172	223.73	0.98
364115	e5299_s3126_r10201_p4172	223.73	0.98
364115	e5299_s3126_r10724_p4172	223.73	0.98
364116	e5299_s3126_r9364_p4172	126.45	0.98
364116	e5299_s3126_r10201_p4172	126.45	0.98
364116	e5299_s3126_r10724_p4172	126.45	0.98
364117	e5299_s3126_r9364_p4172	76.29	0.98
364117	e5299_s3126_r10201_p4172	76.29	0.98
364117	e5299_s3126_r10724_p4172	76.29	0.98
364118	e5299_s3126_r9364_p4172	20.34	0.98
364118	e5299_s3126_r10201_p4172	20.34	0.98
364118	e5299_s3126_r10724_p4172	20.34	0.98
364119	e5299_s3126_r9364_p4172	12.62	0.98
364119	e5299_s3126_r10201_p4172	12.62	0.98
364119	e5299_s3126_r10724_p4172	12.62	0.98
364120	e5299_s3126_r9364_p4172	25.03	0.98
364120	e5299_s3126_r10201_p4172	25.03	0.98
364120	e5299_s3126_r10724_p4172	25.03	0.98
364121	e5299_s3126_r9364_p4172	9.37	0.98
364121	e5299_s3126_r10201_p4172	9.37	0.98
364121	e5299_s3126_r10724_p4172	9.37	0.98
364122	e5299_s3126_r9364_p4172	6.08	0.98
364122	e5299_s3126_r10201_p4172	6.08	0.98
364122	e5299_s3126_r10724_p4172	6.08	0.98
364123	e5299_s3126_r9364_p4172	4.87	0.98
364123	e5299_s3126_r10201_p4172	4.87	0.98
364123	e5299_s3126_r10724_p4172	4.87	0.98
364124	e5299_s3126_r9364_p4172	2.28	0.98
364124	e5299_s3126_r10201_p4172	2.28	0.98
364124	e5299_s3126_r10724_p4172	2.28	0.98
364125	e5299_s3126_r9364_p4172	1.49	0.98
364125	e5299_s3126_r10201_p4172	1.49	0.98
364125	e5299_s3126_r10724_p4172	1.49	0.98
364126	e5299_s3126_r9364_p4172	1.81	0.98
364126	e5299_s3126_r10201_p4172	1.81	0.98
364126	e5299_s3126_r10724_p4172	1.81	0.98
364127	e5299_s3126_r9364_p4172	0.15	0.98
364127	e5299_s3126_r10201_p4172	0.15	0.98
364127	e5299_s3126_r10724_p4172	0.15	0.98
364128	e5307_s3126_r9364_p4172	1 627.73	0.98
364128	e5307_s3126_r10201_p4172	1 627.73	0.98

364128	e5307_s3126_r10724_p4172	1 627.73	0.98
364129	e5307_s3126_r9364_p4172	223.88	0.98
364129	e5307_s3126_r10201_p4172	223.88	0.98
364129	e5307_s3126_r10724_p4172	223.88	0.98
364130	e5307_s3126_r9364_p4172	127.73	0.98
364130	e5307_s3126_r10201_p4172	127.73	0.98
364130	e5307_s3126_r10724_p4172	127.73	0.98
364131	e5307_s3126_r9364_p4172	76.03	0.98
364131	e5307_s3126_r10201_p4172	76.03	0.98
364131	e5307_s3126_r10724_p4172	76.03	0.98
364132	e5307_s3126_r9364_p4172	20.21	0.98
364132	e5307_s3126_r10201_p4172	20.21	0.98
364132	e5307_s3126_r10724_p4172	20.21	0.98
364133	e5307_s3126_r9364_p4172	12.29	0.98
364133	e5307_s3126_r10201_p4172	12.29	0.98
364133	e5307_s3126_r10724_p4172	12.29	0.98
364134	e5307_s3126_r9364_p4172	24.80	0.98
364134	e5307_s3126_r10201_p4172	24.80	0.98
364134	e5307_s3126_r10724_p4172	24.80	0.98
364135	e5307_s3126_r9364_p4172	9.33	0.98
364135	e5307_s3126_r10201_p4172	9.33	0.98
364135	e5307_s3126_r10724_p4172	9.33	0.98
364136	e5307_s3126_r9364_p4172	5.48	0.98
364136	e5307_s3126_r10201_p4172	5.48	0.98
364136	e5307_s3126_r10724_p4172	5.48	0.98
364137	e5307_s3126_r9364_p4172	4.79	0.98
364137	e5307_s3126_r10201_p4172	4.79	0.98
364137	e5307_s3126_r10724_p4172	4.79	0.98
364138	e5313_s3126_r9364_p4172	2.28	0.98
364138	e5313_s3126_r10201_p4172	2.28	0.98
364138	e5313_s3126_r10724_p4172	2.28	0.98
364139	e5313_s3126_r9364_p4172	1.50	0.98
364139	e5313_s3126_r10201_p4172	1.50	0.98
364139	e5313_s3126_r10724_p4172	1.50	0.98
364140	e5307_s3126_r9364_p4172	1.81	0.98
364140	e5307_s3126_r10201_p4172	1.81	0.98
364140	e5307_s3126_r10724_p4172	1.81	0.98
364141	e5307_s3126_r9364_p4172	0.15	0.98
364141	e5307_s3126_r10201_p4172	0.15	0.98
364141	e5307_s3126_r10724_p4172	0.15	0.98
364198	e5421_s3126_r9364_p4172	2 330.19	0.98
364198	e5421_s3126_r10201_p4172	2 330.19	0.98
364198	e5421_s3126_r10724_p4172	2 330.19	0.98
364199	e5421_s3126_r9364_p4172	82.26	0.98
364199	e5421_s3126_r10201_p4172	82.26	0.98

364199	e5421_s3126_r10724_p4172	82.26	0.98
364200	e5421_s3126_r9364_p4172	44.88	0.98
364200	e5421_s3126_r10201_p4172	44.88	0.98
364200	e5421_s3126_r10724_p4172	44.88	0.98
364201	e5421_s3126_r9364_p4172	5.11	0.98
364201	e5421_s3126_r10201_p4172	5.11	0.98
364201	e5421_s3126_r10724_p4172	5.11	0.98
364202	e5421_s3126_r9364_p4172	2.76	0.98
364202	e5421_s3126_r10201_p4172	2.76	0.98
364202	e5421_s3126_r10724_p4172	2.76	0.98
364203	e5421_s3126_r9364_p4172	0.47	0.98
364203	e5421_s3126_r10201_p4172	0.47	0.98
364203	e5421_s3126_r10724_p4172	0.47	0.98
364204	e5421_s3126_r9364_p4172	2 331.22	0.98
364204	e5421_s3126_r10201_p4172	2 331.22	0.98
364204	e5421_s3126_r10724_p4172	2 331.22	0.98
364205	e5421_s3126_r9364_p4172	81.36	0.98
364205	e5421_s3126_r10201_p4172	81.36	0.98
364205	e5421_s3126_r10724_p4172	81.36	0.98
364206	e5421_s3126_r9364_p4172	44.97	0.98
364206	e5421_s3126_r10201_p4172	44.97	0.98
364206	e5421_s3126_r10724_p4172	44.97	0.98
364207	e5421_s3126_r9364_p4172	5.48	0.98
364207	e5421_s3126_r10201_p4172	5.48	0.98
364207	e5421_s3126_r10724_p4172	5.48	0.98
364208	e5421_s3126_r9364_p4172	2.78	0.98
364208	e5421_s3126_r10201_p4172	2.78	0.98
364208	e5421_s3126_r10724_p4172	2.78	0.98
364209	e5421_s3126_r9364_p4172	0.47	0.98
364209	e5421_s3126_r10201_p4172	0.47	0.98
364209	e5421_s3126_r10724_p4172	0.47	0.98
364210	e5421_s3126_r9364_p4172	2 333.93	0.98
364210	e5421_s3126_r10201_p4172	2 333.93	0.98
364210	e5421_s3126_r10724_p4172	2 333.93	0.98
364211	e5421_s3126_r9364_p4172	81.10	0.98
364211	e5421_s3126_r10201_p4172	81.10	0.98
364211	e5421_s3126_r10724_p4172	81.10	0.98
364212	e5421_s3126_r9364_p4172	44.84	0.98
364212	e5421_s3126_r10201_p4172	44.84	0.98
364212	e5421_s3126_r10724_p4172	44.84	0.98
364213	e5421_s3126_r9364_p4172	5.54	0.98
364213	e5421_s3126_r10201_p4172	5.54	0.98
364213	e5421_s3126_r10724_p4172	5.54	0.98
364214	e5421_s3126_r9364_p4172	2.79	0.98
364214	e5421_s3126_r10201_p4172	2.79	0.98

	364214	e5421_s3126_r10724_p4172	2.79	0.98
	364215	e5421_s3126_r9364_p4172	0.47	0.98
	364215	e5421_s3126_r10201_p4172	0.47	0.98
	364215	e5421_s3126_r10724_p4172	0.47	0.98
ttH	346343	e7148_s3126v_r9364_p4174	0.23	1.00
	346343	e7148_s3126v_r10201_p4174	0.23	1.00
	346343	e7148_s3126_r10724_p4174	0.23	1.00
	346344	e7148_s3126_r9364_p4174	0.22	1.00
	346344	e7148_s3126_r10201_p4174	0.22	1.00
	346344	e7148_s3126_r10724_p4174	0.22	1.00
	346345	e7148_s3126_r9364_p4174	0.05	1.00
	346345	e7148_s3126_r10201_p4174	0.05	1.00
	346345	e7148_s3126_r10724_p4174	0.05	1.00
Triboson	364242	e5887_s3126_r9364_p4172	0.01	1.00
	364242	e5887_s3126_r10201_p4172	0.01	1.00
	364242	e5887_s3126_r10724_p4172	0.01	1.00
	364243	e5887_s3126_r9364_p4172	0.00	1.00
	364243	e5887_s3126_r10201_p4172	0.00	1.00
	364243	e5887_s3126_r10724_p4172	0.00	1.00
	364244	e5887_s3126_r9364_p4172	0.00	1.00
	364244	e5887_s3126_r10201_p4172	0.00	1.00
	364244	e5887_s3126_r10724_p4172	0.00	1.00
	364245	e5887_s3126_r9364_p4172	0.00	1.00
	364245	e5887_s3126_r10201_p4172	0.00	1.00
	364245	e5887_s3126_r10724_p4172	0.00	1.00
	364246	e5887_s3126_r9364_p4172	0.00	1.00
	364246	e5887_s3126_r10201_p4172	0.00	1.00
	364246	e5887_s3126_r10724_p4172	0.00	1.00
	364247	e5887_s3126_r9364_p4172	0.00	1.00
	364247	e5887_s3126_r10201_p4172	0.00	1.00
	364247	e5887_s3126_r10724_p4172	0.00	1.00
	364248	e5887_s3126_r9364_p4172	0.00	1.00
	364248	e5887_s3126_r10201_p4172	0.00	1.00
	364248	e5887_s3126_r10724_p4172	0.00	1.00
	364249	e5887_s3126_r9364_p4172	0.00	1.00
	364249	e5887_s3126_r10201_p4172	0.00	1.00
	364249	e5887_s3126_r10724_p4172	0.00	1.00
s-channel	410644	e6527_s3126_r9364_p4174	2.03	1.01
	410644	e6527_s3126_r10201_p4174	2.03	1.01
	410644	e6527_s3126_r10724_p4174	2.03	1.01
	410645	e6527_s3126_r9364_p4174	1.27	1.01
	410645	e6527_s3126_r10201_p4174	1.27	1.01

	410645	e6527_s3126_r10724_p4174	1.27	1.01
Higgs	342282	e4850_s3126_r9364_p4172	30.20	1.45
	342282	e4850_s3126_r10201_p4172	30.20	1.45
	342282	e4850_s3126_r10724_p4172	30.20	1.45
	342283	e4246_s3126_r9364_p4172	3.83	0.98
	342283	e4246_s3126_r10201_p4172	3.83	0.98
	342283	e4246_s3126_r10724_p4172	3.83	0.98
	342284	e4246_s3126_r9364_p4172	1.10	1.25
	342284	e4246_s3126_r10201_p4172	1.10	1.25
	342284	e4246_s3126_r10724_p4172	1.10	1.25
	342285	e4246_s3126_r9364_p4172	0.60	1.45
	342285	e4246_s3126_r10201_p4172	0.60	1.45
	342285	e4246_s3126_r10724_p4172	0.60	1.45
Diboson	364250	e5894_s3126_r9364_p4172	1.25	1.00
	364250	e5894_s3126_r10201_p4172	1.25	1.00
	364250	e5894_s3126_r10724_p4172	1.25	1.00
	364253	e5916_s3126_r9364_p4172	4.58	1.00
	364253	e5916_s3126_r10201_p4172	4.58	1.00
	364253	e5916_s3126_r10724_p4172	4.58	1.00
	364254	e5916_s3126_r9364_p4172	12.50	1.00
	364254	e5916_s3126_r10201_p4172	12.50	1.00
	364254	e5916_s3126_r10724_p4172	12.50	1.00
	364255	e5916_s3126_r9364_p4172	3.23	1.00
	364255	e5916_s3126_r10201_p4172	3.23	1.00
	364255	e5916_s3126_r10724_p4172	3.23	1.00
	364283	e6055_s3126_r9364_p4172	0.01	1.00
	364283	e6055_s3126_r10201_p4172	0.01	1.00
	364283	e6055_s3126_r10724_p4172	0.01	1.00
	364284	e6055_s3126_r9364_p4172	0.05	1.00
	364284	e6055_s3126_r10201_p4172	0.05	1.00
	364284	e6055_s3126_r10724_p4172	0.05	1.00
	364285	e6055_s3126_r9364_p4172	0.12	1.00
	364285	e6055_s3126_r10201_p4172	0.12	1.00
	364285	e6055_s3126_r10724_p4172	0.12	1.00
	364286	e6055_s3126_r9364_p4172	0.03	1.00
	364286	e6055_s3126_r10201_p4172	0.03	1.00
	364286	e6055_s3126_r10724_p4172	0.03	1.00
	364287	e6055_s3126_r9364_p4172	0.04	1.00
	364287	e6055_s3126_r10201_p4172	0.04	1.00
	364287	e6055_s3126_r10724_p4172	0.04	1.00
	364288	e6096_s3126_r9364_p4172	1.43	1.00
	364288	e6096_s3126_r10201_p4172	1.43	1.00
	364288	e6096_s3126_r10724_p4172	1.43	1.00

	364289	e6133_s3126_r9364_p4172	2.92	1.00
	364289	e6133_s3126_r10201_p4172	2.92	1.00
	364289	e6133_s3126_r10724_p4172	2.92	1.00
	364290	e6096_s3126_r9364_p4172	0.17	1.00
	364290	e6096_s3126_r10201_p4172	0.17	1.00
	364290	e6096_s3126_r10724_p4172	0.17	1.00
	363355	e5525_s3126_r9364_p4172	15.56	0.28
	363355	e5525_s3126_r10201_p4172	15.56	0.28
	363355	e5525_s3126_r10724_p4172	15.56	0.28
	363356	e5525_s3126_r9364_p4172	15.56	0.14
	363356	e5525_s3126_r10201_p4172	15.56	0.14
	363356	e5525_s3126_r10724_p4172	15.56	0.14
	363357	e5525_s3126_r9364_p4172	6.80	1.00
	363357	e5525_s3126_r10201_p4172	6.80	1.00
	363357	e5525_s3126_r10724_p4172	6.80	1.00
	363358	e5525_s3126_r9364_p4172	3.44	1.00
	363358	e5525_s3126_r10201_p4172	3.44	1.00
	363358	e5525_s3126_r10724_p4172	3.44	1.00
	363359	e5583_s3126_r9364_p4172	24.72	1.00
	363359	e5583_s3126_r10201_p4172	24.72	1.00
	363359	e5583_s3126_r10724_p4172	24.72	1.00
	363360	e5983_s3126_r9364_p4172	24.73	1.00
	363360	e5983_s3126_r10201_p4172	24.73	1.00
	363360	e5983_s3126_r10724_p4172	24.73	1.00
	363489	e5525_s3126_r9364_p4172	11.41	1.00
	363489	e5525_s3126_r10201_p4172	11.41	1.00
	363489	e5525_s3126_r10724_p4172	11.41	1.00
W+jets	364156	e5340_s3126_r9364_p4172	15 770.00	0.97
	364156	e5340_s3126_r10201_p4172	15 770.00	0.97
	364156	e5340_s3126_r10724_p4172	15 770.00	0.97
	364157	e5340_s3126_r9364_p4172	2 493.38	0.97
	364157	e5340_s3126_r10201_p4172	2 493.38	0.97
	364157	e5340_s3126_r10724_p4172	2 493.38	0.97
	364158	e5340_s3126_r9364_p4172	844.20	0.97
	364158	e5340_s3126_r10201_p4172	844.20	0.97
	364158	e5340_s3126_r10724_p4172	844.20	0.97
	364159	e5340_s3126_r9364_p4172	637.42	0.97
	364159	e5340_s3126_r10201_p4172	637.42	0.97
	364159	e5340_s3126_r10724_p4172	637.42	0.97
	364160	e5340_s3126_r9364_p4172	219.97	0.97
	364160	e5340_s3126_r10201_p4172	219.97	0.97
	364160	e5340_s3126_r10724_p4172	219.97	0.97
	364161	e5340_s3126_r9364_p4172	71.46	0.97
	364161	e5340_s3126_r10201_p4172	71.46	0.97

364161	e5340_s3126_r10724_p4172	71.46	0.97
364162	e5340_s3126_r9364_p4172	212.56	0.97
364162	e5340_s3126_r10201_p4172	212.56	0.97
364162	e5340_s3126_r10724_p4172	212.56	0.97
364163	e5340_s3126_r9364_p4172	98.44	0.97
364163	e5340_s3126_r10201_p4172	98.44	0.97
364163	e5340_s3126_r10724_p4172	98.44	0.97
364164	e5340_s3126_r9364_p4172	36.91	0.97
364164	e5340_s3126_r10201_p4172	36.91	0.97
364164	e5340_s3126_r10724_p4172	36.91	0.97
364165	e5340_s3126_r9364_p4172	39.38	0.97
364165	e5340_s3126_r10201_p4172	39.38	0.97
364165	e5340_s3126_r10724_p4172	39.38	0.97
364166	e5340_s3126_r9364_p4172	22.92	0.97
364166	e5340_s3126_r10201_p4172	22.92	0.97
364166	e5340_s3126_r10724_p4172	22.92	0.97
364167	e5340_s3126_r9364_p4172	9.61	0.97
364167	e5340_s3126_r10201_p4172	9.61	0.97
364167	e5340_s3126_r10724_p4172	9.61	0.97
364168	e5340_s3126_r9364_p4172	15.01	0.97
364168	e5340_s3126_r10201_p4172	15.01	0.97
364168	e5340_s3126_r10724_p4172	15.01	0.97
364169	e5340_s3126_r9364_p4172	1.23	0.97
364169	e5340_s3126_r10201_p4172	1.23	0.97
364169	e5340_s3126_r10724_p4172	1.23	0.97
364170	e5340_s3126_r9364_p4172	15 769.64	0.97
364170	e5340_s3126_r10201_p4172	15 769.64	0.97
364170	e5340_s3126_r10724_p4172	15 769.64	0.97
364171	e5340_s3126_r9364_p4172	2 492.64	0.97
364171	e5340_s3126_r10201_p4172	2 492.64	0.97
364171	e5340_s3126_r10724_p4172	2 492.64	0.97
364172	e5340_s3126_r9364_p4172	844.64	0.97
364172	e5340_s3126_r10201_p4172	844.64	0.97
364172	e5340_s3126_r10724_p4172	844.64	0.97
364173	e5340_s3126_r9364_p4172	630.32	0.97
364173	e5340_s3126_r10201_p4172	630.32	0.97
364173	e5340_s3126_r10724_p4172	630.32	0.97
364174	e5340_s3126_r9364_p4172	215.49	0.97
364174	e5340_s3126_r10201_p4172	215.49	0.97
364174	e5340_s3126_r10724_p4172	215.49	0.97
364175	e5340_s3126_r9364_p4172	97.74	0.97
364175	e5340_s3126_r10201_p4172	97.74	0.97
364175	e5340_s3126_r10724_p4172	97.74	0.97
364176	e5340_s3126_r9364_p4172	202.84	0.97
364176	e5340_s3126_r10201_p4172	202.84	0.97

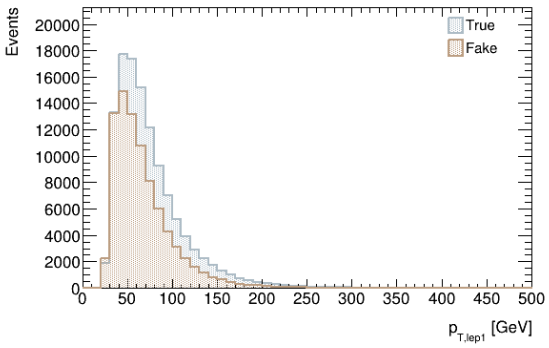
364176	e5340_s3126_r10724_p4172	202.84	0.97
364177	e5340_s3126_r9364_p4172	98.44	0.97
364177	e5340_s3126_r10201_p4172	98.44	0.97
364177	e5340_s3126_r10724_p4172	98.44	0.97
364178	e5340_s3126_r9364_p4172	37.00	0.97
364178	e5340_s3126_r10201_p4172	37.00	0.97
364178	e5340_s3126_r10724_p4172	37.00	0.97
364179	e5340_s3126_r9364_p4172	39.24	0.97
364179	e5340_s3126_r10201_p4172	39.24	0.97
364179	e5340_s3126_r10724_p4172	39.24	0.97
364180	e5340_s3126_r9364_p4172	22.85	0.97
364180	e5340_s3126_r10201_p4172	22.85	0.97
364180	e5340_s3126_r10724_p4172	22.85	0.97
364181	e5340_s3126_r9364_p4172	9.66	0.97
364181	e5340_s3126_r10201_p4172	9.66	0.97
364181	e5340_s3126_r10724_p4172	9.66	0.97
364182	e5340_s3126_r9364_p4172	15.22	0.97
364182	e5340_s3126_r10201_p4172	15.22	0.97
364182	e5340_s3126_r10724_p4172	15.22	0.97
364183	e5340_s3126_r9364_p4172	1.23	0.97
364183	e5340_s3126_r10201_p4172	1.23	0.97
364183	e5340_s3126_r10724_p4172	1.23	0.97
364184	e5340_s3126_r9364_p4172	15 799.44	0.97
364184	e5340_s3126_r10201_p4172	15 799.44	0.97
364184	e5340_s3126_r10724_p4172	15 799.44	0.97
364185	e5340_s3126_r9364_p4172	2 477.25	0.97
364185	e5340_s3126_r10201_p4172	2 477.25	0.97
364185	e5340_s3126_r10724_p4172	2 477.25	0.97
364186	e5340_s3126_r9364_p4172	854.55	0.97
364186	e5340_s3126_r10201_p4172	854.55	0.97
364186	e5340_s3126_r10724_p4172	854.55	0.97
364187	e5340_s3126_r9364_p4172	638.55	0.97
364187	e5340_s3126_r10201_p4172	638.55	0.97
364187	e5340_s3126_r10724_p4172	638.55	0.97
364188	e5340_s3126_r9364_p4172	210.38	0.97
364188	e5340_s3126_r10201_p4172	210.38	0.97
364188	e5340_s3126_r10724_p4172	210.38	0.97
364189	e5340_s3126_r9364_p4172	98.02	0.97
364189	e5340_s3126_r10201_p4172	98.02	0.97
364189	e5340_s3126_r10724_p4172	98.02	0.97
364190	e5340_s3126_r9364_p4172	202.33	0.97
364190	e5340_s3126_r10201_p4172	202.33	0.97
364190	e5340_s3126_r10724_p4172	202.33	0.97
364191	e5340_s3126_r9364_p4172	98.58	0.97
364191	e5340_s3126_r10201_p4172	98.58	0.97

	364191	e5340_s3126_r10724_p4172	98.58	0.97
	364192	e5340_s3126_r9364_p4172	40.06	0.97
	364192	e5340_s3126_r10201_p4172	40.06	0.97
	364192	e5340_s3126_r10724_p4172	40.06	0.97
	364193	e5340_s3126_r9364_p4172	39.33	0.97
	364193	e5340_s3126_r10201_p4172	39.33	0.97
	364193	e5340_s3126_r10724_p4172	39.33	0.97
	364194	e5340_s3126_r9364_p4172	22.78	0.97
	364194	e5340_s3126_r10201_p4172	22.78	0.97
	364194	e5340_s3126_r10724_p4172	22.78	0.97
	364195	e5340_s3126_r9364_p4172	9.67	0.97
	364195	e5340_s3126_r10201_p4172	9.67	0.97
	364195	e5340_s3126_r10724_p4172	9.67	0.97
	364196	e5340_s3126_r9364_p4172	15.05	0.97
	364196	e5340_s3126_r10201_p4172	15.05	0.97
	364196	e5340_s3126_r10724_p4172	15.05	0.97
	364197	e5340_s3126_r9364_p4172	1.23	0.97
	364197	e5340_s3126_r10201_p4172	1.23	0.97
	364197	e5340_s3126_r10724_p4172	1.23	0.97
Small Background	412043	e7101_a875_r9364_p4174	0.01	1.13
	412043	e7101_a875_r10201_p4174	0.01	1.13
	412043	e7101_a875_r10724_p4174	0.01	1.13
	412119	e7518_s3126_r9364_p4174	0.01	1.00
	412119	e7518_s3126_r10201_p4174	0.01	1.00
	412119	e7518_s3126_r10724_p4174	0.01	1.00
	361600	e4616_s3126_r9364_p4172	10.63	1.00
	361600	e4616_s3126_r10201_p4172	10.63	1.00
	361600	e4616_s3126_r10724_p4172	10.63	1.00
	412043	e7101_a875_r9364_p4174	0.01	1.13
	412043	e7101_a875_r10201_p4174	0.01	1.13
	412043	e7101_a875_r10724_p4174	0.01	1.13
	410649	e6615_s3126_r9364_p4174	3.99	0.95
	410649	e6615_s3126_r10201_p4174	3.99	0.95
	410649	e6615_s3126_r10724_p4174	3.99	0.95
	410648	e6615_s3126_r9364_p4174	4.00	0.94
	410648	e6615_s3126_r10201_p4174	4.00	0.94
	410648	e6615_s3126_r10724_p4174	4.00	0.94
	304014	e4324_s3126_r9364_p4174	0.00	1.00
	304014	e4324_s3126_r10201_p4174	0.00	1.00
	304014	e4324_s3126_r10724_p4174	0.00	1.00
	361610	e4711_s3126_r9364_p4172	2.27	1.00
	361610	e4711_s3126_r10201_p4172	2.27	1.00
	361610	e4711_s3126_r10724_p4172	2.27	1.00
	361609	e4711_s3126_r9364_p4172	10.10	1.00

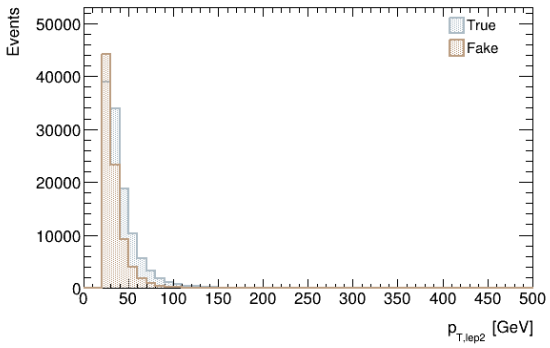
361609	e4711_s3126_r10201_p4172	10.10	1.00
361609	e4711_s3126_r10724_p4172	10.10	1.00
361608	e4711_s3126_r9364_p4172	5.77	1.00
361608	e4711_s3126_r10201_p4172	5.77	1.00
361608	e4711_s3126_r10724_p4172	5.77	1.00
361607	e4711_s3126_r9364_p4172	3.28	1.00
361607	e4711_s3126_r10201_p4172	3.28	1.00
361607	e4711_s3126_r10724_p4172	3.28	1.00
361606	e4711_s3126_r9364_p4172	44.18	1.00
361606	e4711_s3126_r10201_p4172	44.18	1.00
361606	e4711_s3126_r10724_p4172	44.18	1.00
361604	e4475_s3126_r9364_p4172	0.92	1.00
361604	e4475_s3126_r10201_p4172	0.92	1.00
361604	e4475_s3126_r10724_p4172	0.92	1.00
361603	e4475_s3126_r9364_p4172	1.26	1.00
361603	e4475_s3126_r10201_p4172	1.26	1.00
361603	e4475_s3126_r10724_p4172	1.26	1.00
361602	e4054_s3126_r9364_p4172	2.78	1.00
361602	e4054_s3126_r10201_p4172	2.78	1.00
361602	e4054_s3126_r10724_p4172	2.78	1.00
361601	e4475_s3126_r9364_p4172	4.46	1.00
361601	e4475_s3126_r10201_p4172	4.46	1.00
361601	e4475_s3126_r10724_p4172	4.46	1.00
361600	e4616_s3126_r9364_p4172	10.63	1.00
361600	e4616_s3126_r10201_p4172	10.63	1.00
361600	e4616_s3126_r10724_p4172	10.63	1.00
346799	e8018_a875_r9364_p4174	0.01	1.00
346799	e8018_a875_r10201_p4174	0.01	1.00
346799	e8018_a875_r10724_p4174	0.01	1.00

Table A.2: Monte Carlo Simulation v31 Samples (Run 2) [115]

A.2 MC Composition Plots



(a) $p_{T,\text{lep1}}$



(b) $p_{T,\text{lep2}}$

A.3 Baseline Classifier training results for 1 and 3-prong tau decays.

A.3.1 1-prong

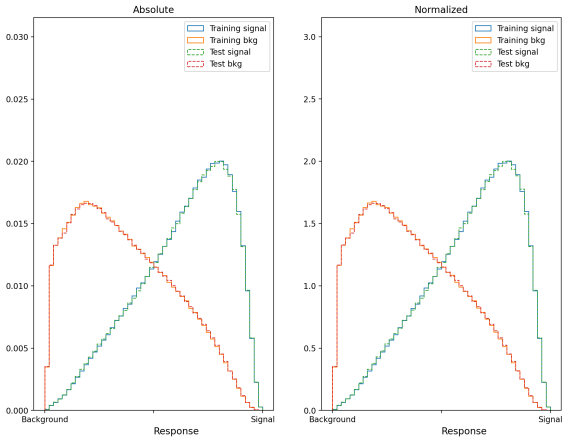


Figure A.3: ROC-curve for the baseline classifier trained on 1-prong tau decays.

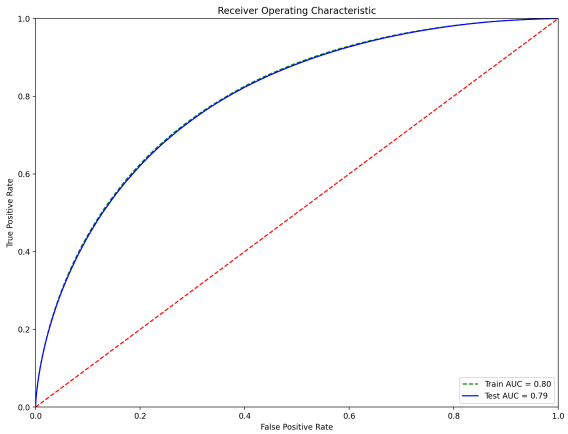
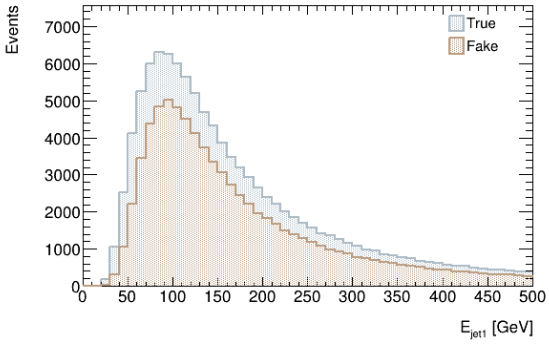
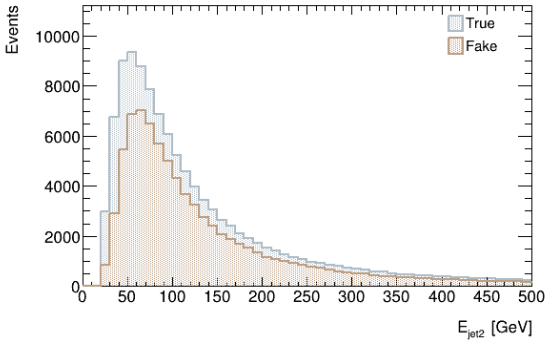


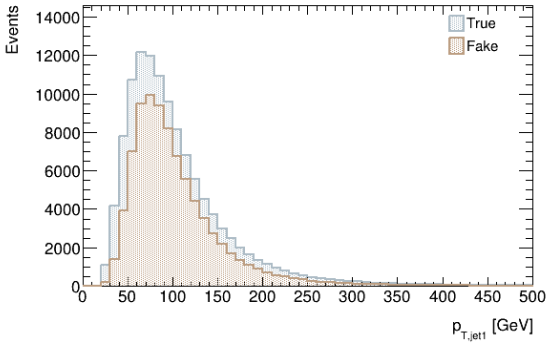
Figure A.4: Response-curve for the baseline classifier trained on 1-prong tau decays.



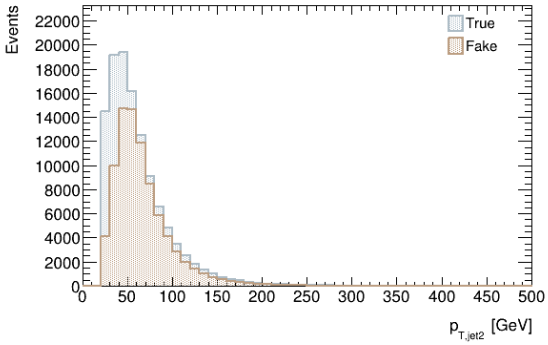
(c) E_{jet1}



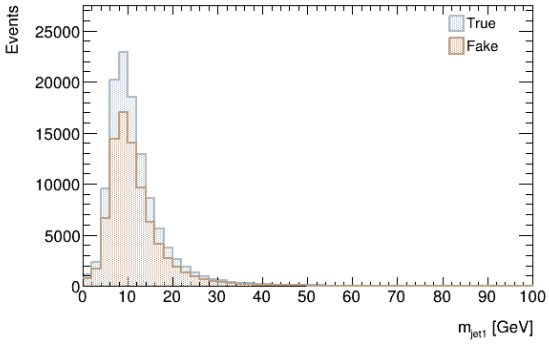
(d) E_{jet2}



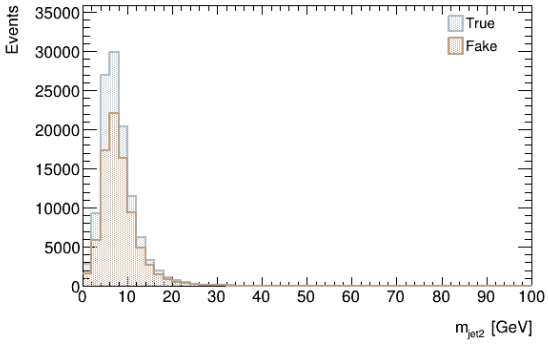
(e) $p_{T\,,\text{jet1}}$



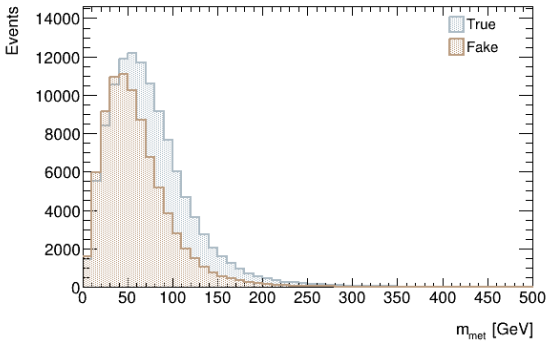
(f) $p_{T\,,\text{jet2}}$



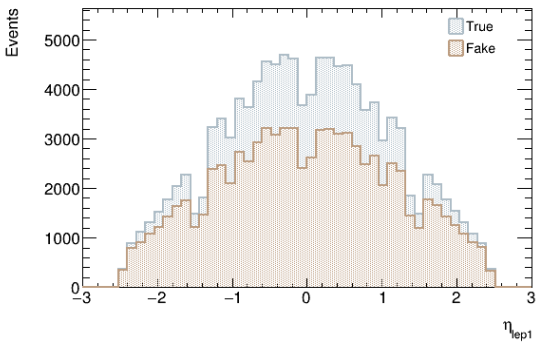
(g) m_{jet1}



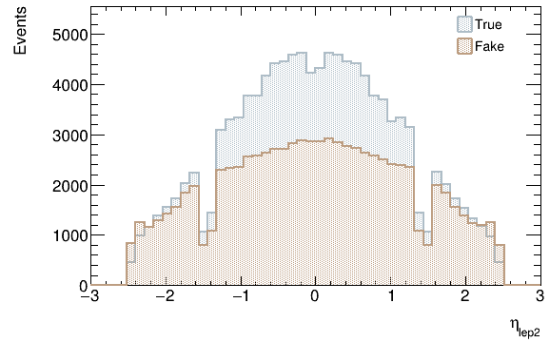
(h) m_{jet2}



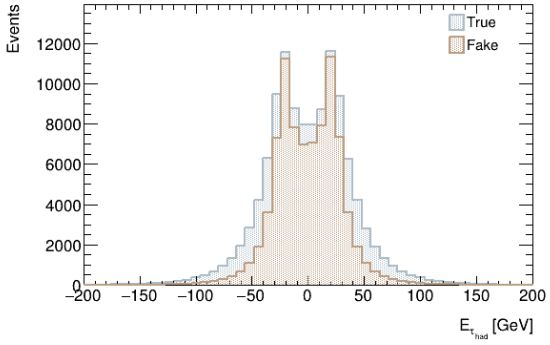
(i) m_{met}



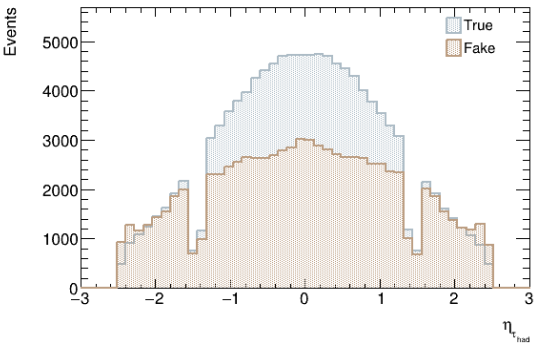
(j) η_{lep1}



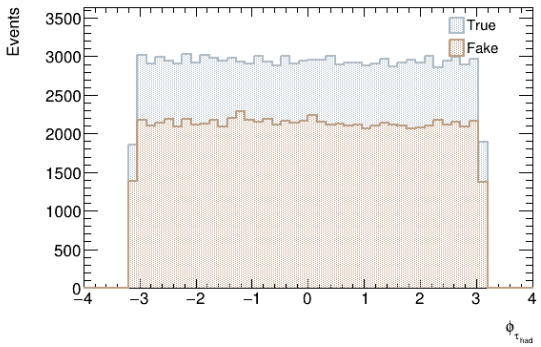
(k) η_{lep2}



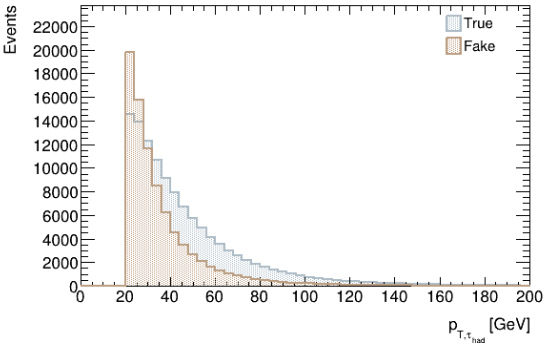
(l) $E_{\tau,\text{had}}$



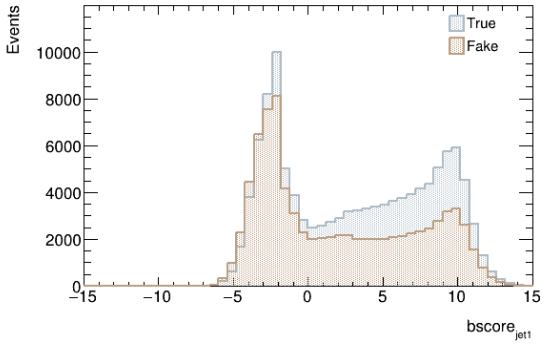
(m) $\eta_{\tau_{\text{had}}}$



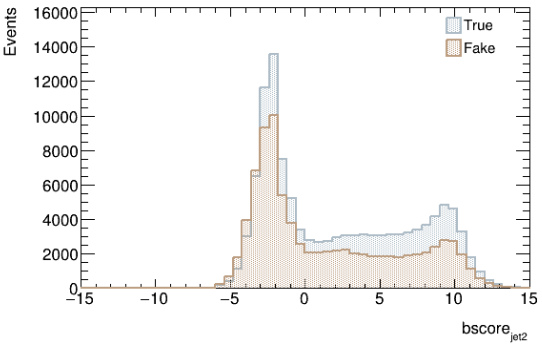
(n) $\phi_{\tau_{\text{had}}}$



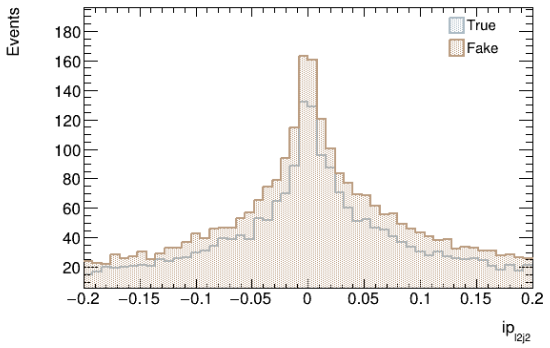
(o) $p_{T,\tau_{\text{had}}}$



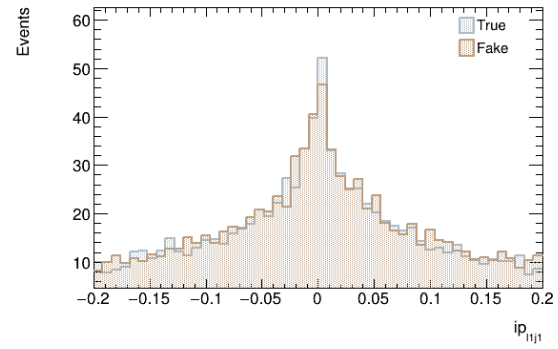
(p) bscore$_{\text{jet1}}$



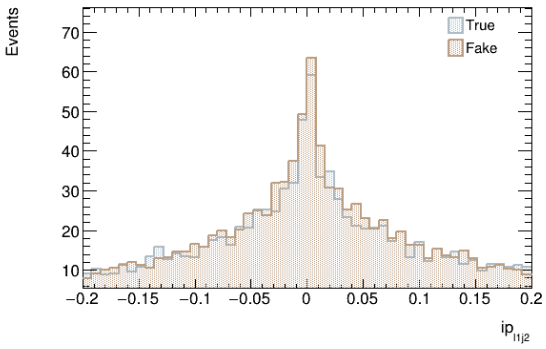
(q) bscore$_{\text{jet2}}$



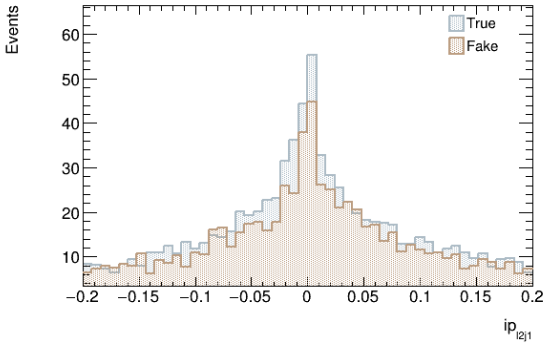
(r) ip_{l2j2}



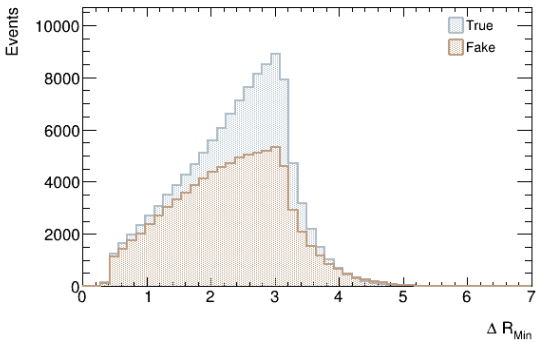
(s) ip_{l1j1}



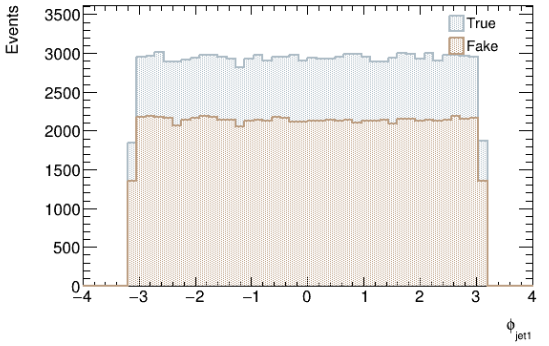
(t) ip_{l1j2}



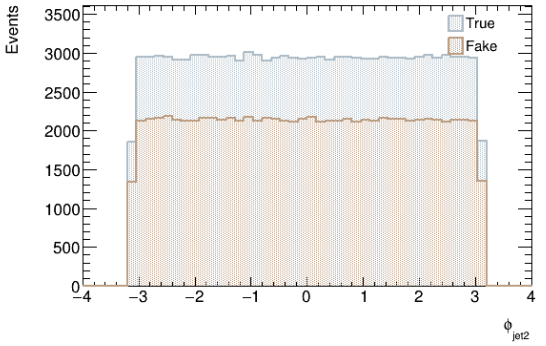
(u) ip_{l2j1}



(v) ΔR_{Min}

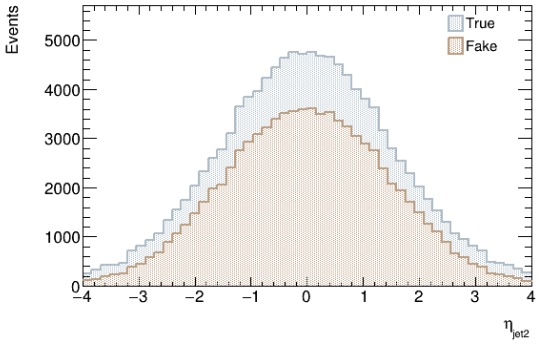


(w) ϕ_{jet1}

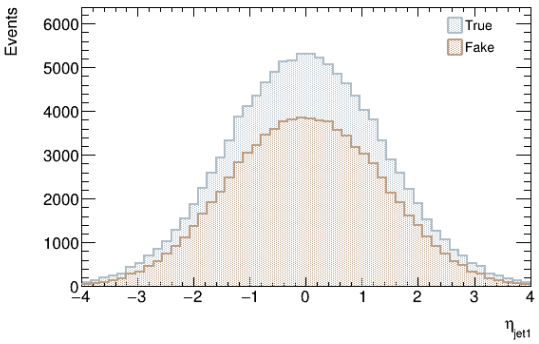


(x) ϕ_{jet2}

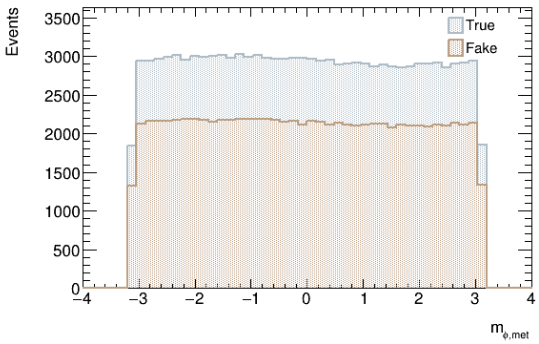
Figure A.1: Monte Carlo Composition Plots 1.



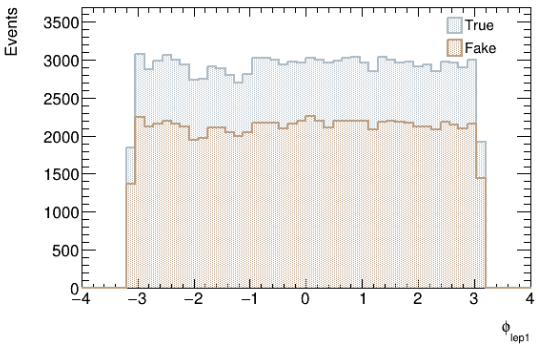
(a) η_{jet2}



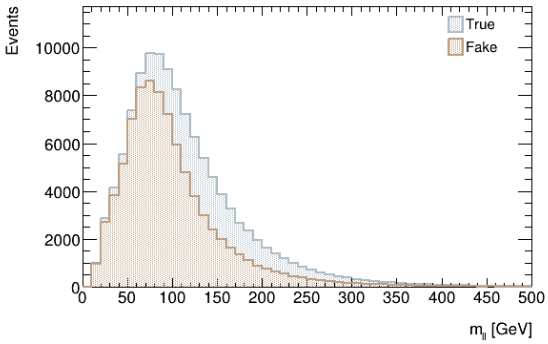
(b) η_{jet1}



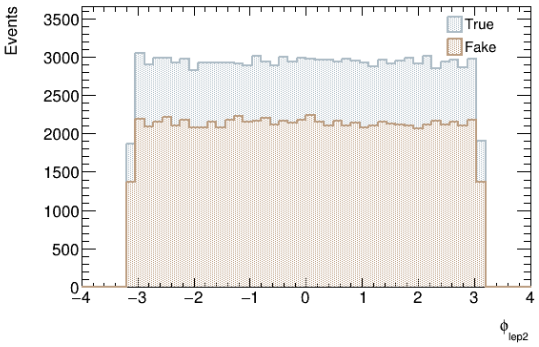
(c) $m_{\phi_{\text{met}}}$



(d) ϕ_{lep}



(e) m_{ll}



(f) ϕ_{lep2}

Figure A.2: Monte Carlo Composition Plots 2.

Network Hyperparameter	Value
Nodes	350
Hidden layers	2
Output layers	1
Activation function (hidden layers)	relu
Batch size	10000
Epochs	400
Optimizer	adam

Table A.3: Overview of the hyperparameters used for the baseline classifier for 1-prong tau decays trained on MC.

A.3.2 3-prong

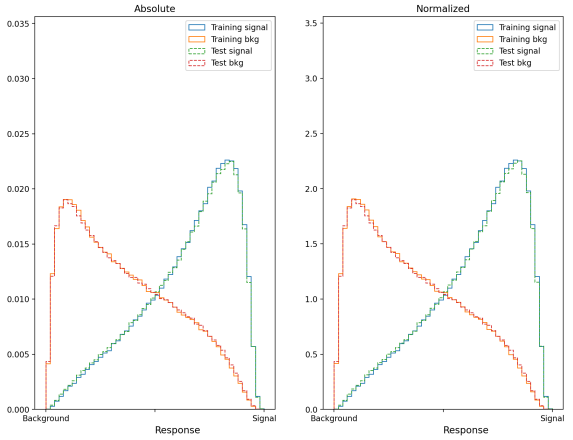


Figure A.5: ROC-curve for the baseline classifier trained on 3-prong tau decays.

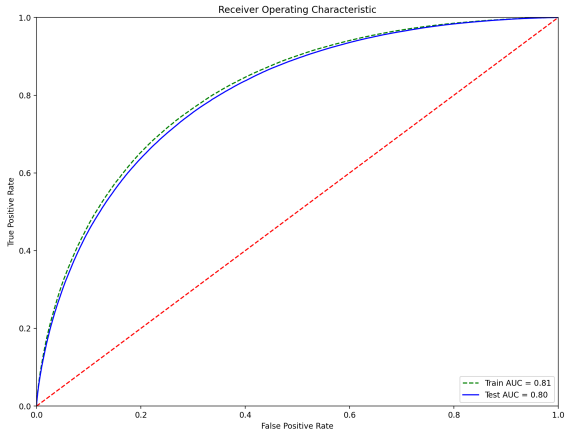
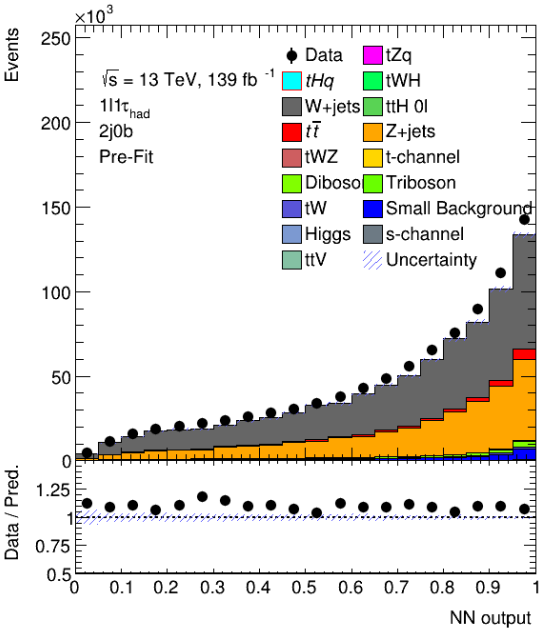


Figure A.6: Response-curve for the baseline classifier trained on 3-prong tau decays.

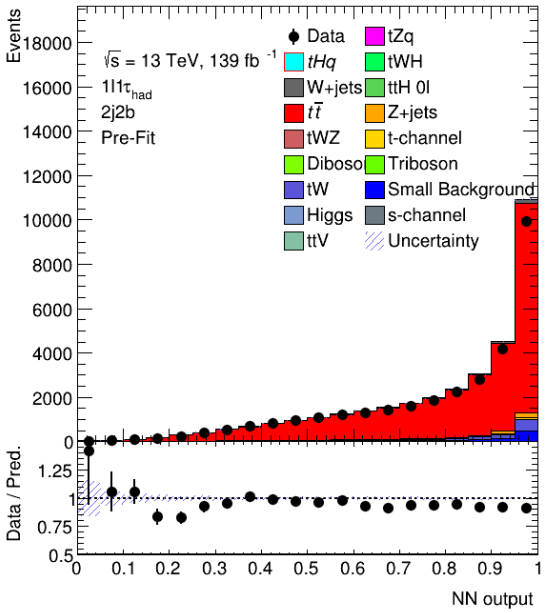
Network Hyperparameter	Value
Nodes	350
Hidden layers	2
Output layers	1
Activation function (hidden layers)	relu
Batch size	10000
Epochs	400
Optimizer	adam

Table A.4: Overview of the hyperparameters used for the baseline classifier for 3-prong tau decays trained on MC.

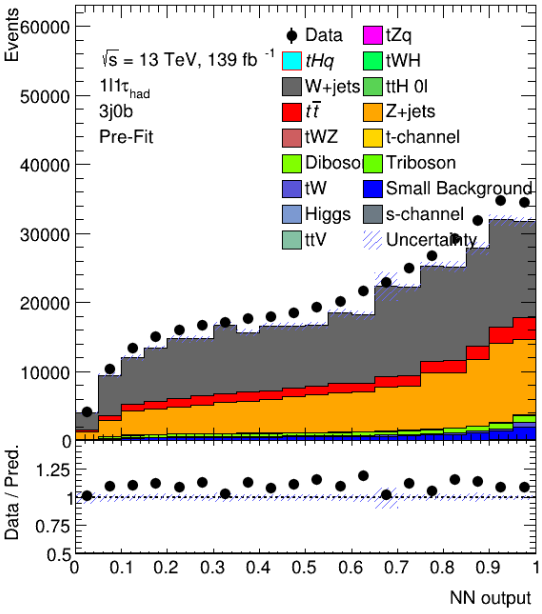
A.4 Prediction on different regions for CWoLa trained on MC



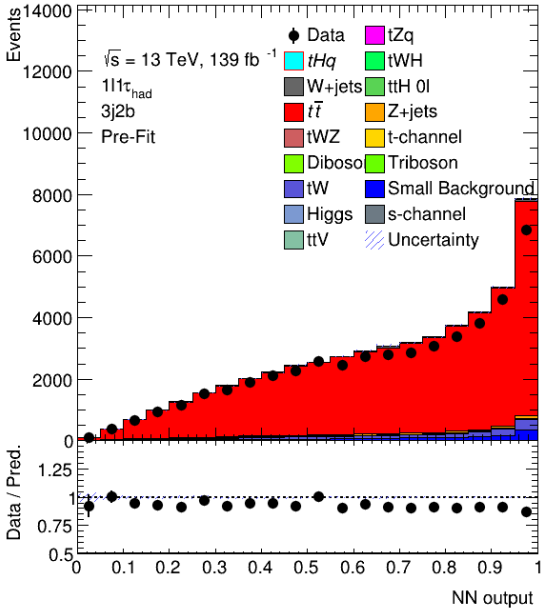
$2j0b$



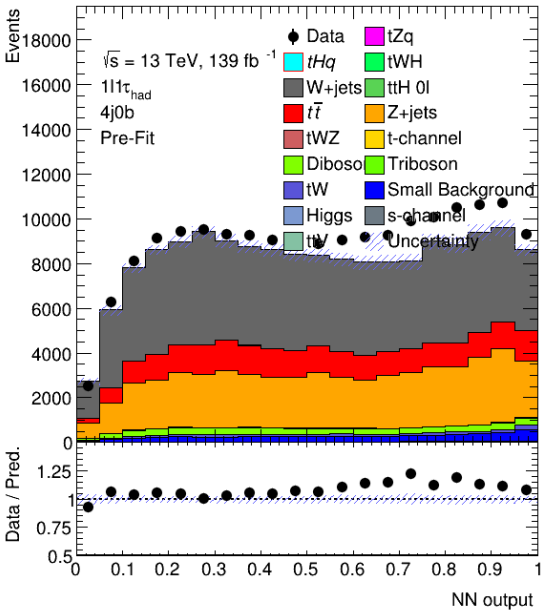
$2j2b$



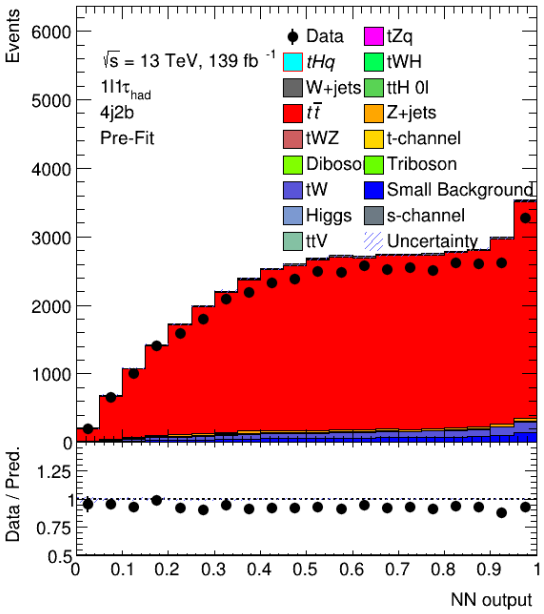
$3j0b$



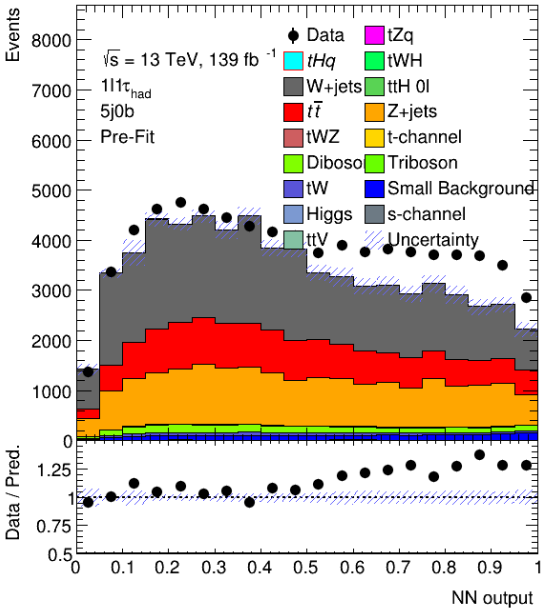
$3j2b$



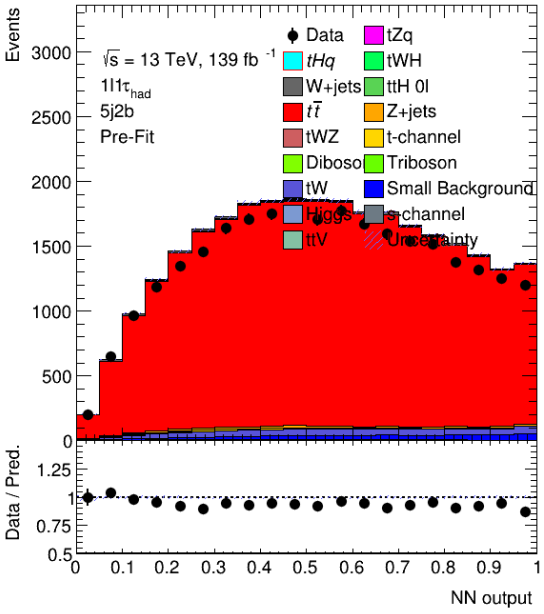
$4j0b$



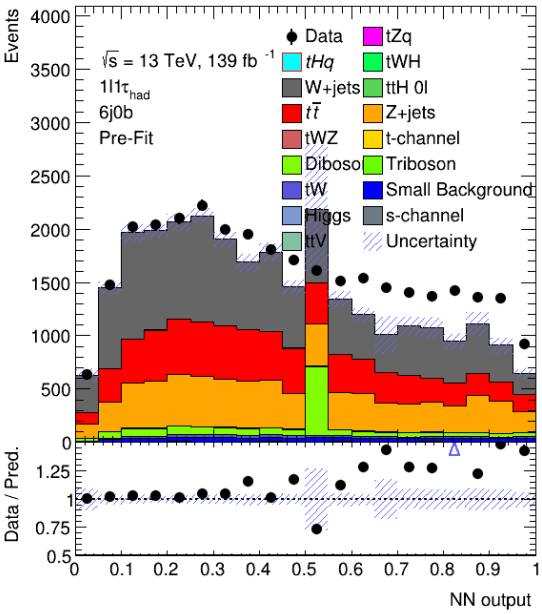
$4j2b$



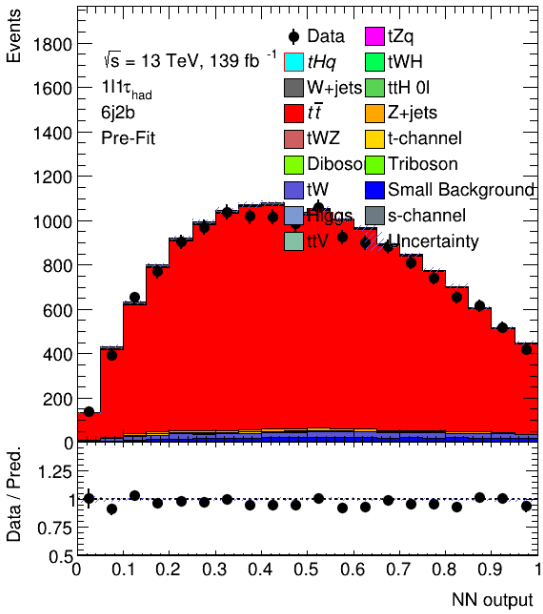
$5j0b$



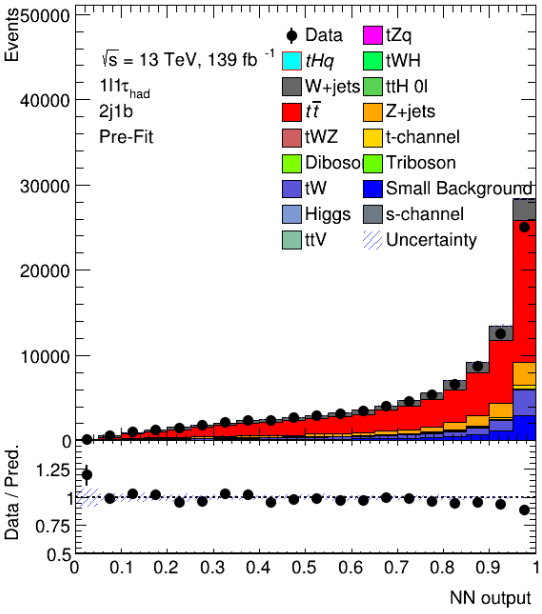
$5j2b$



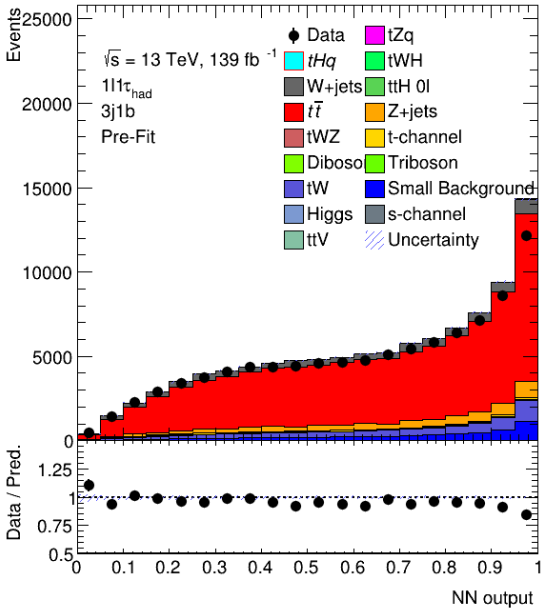
$6j0b$



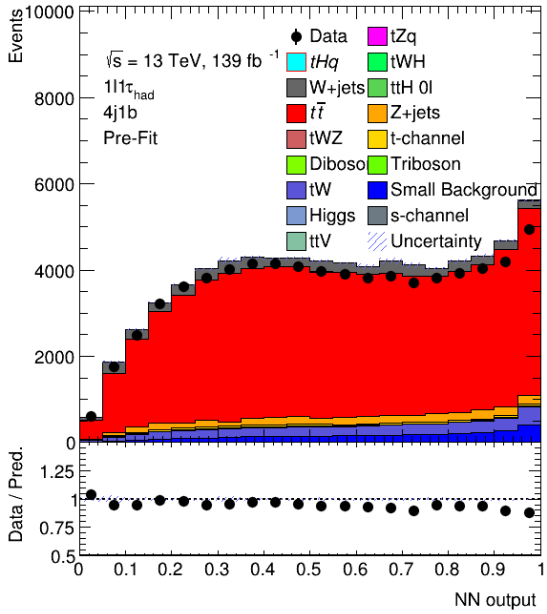
$6j2b$



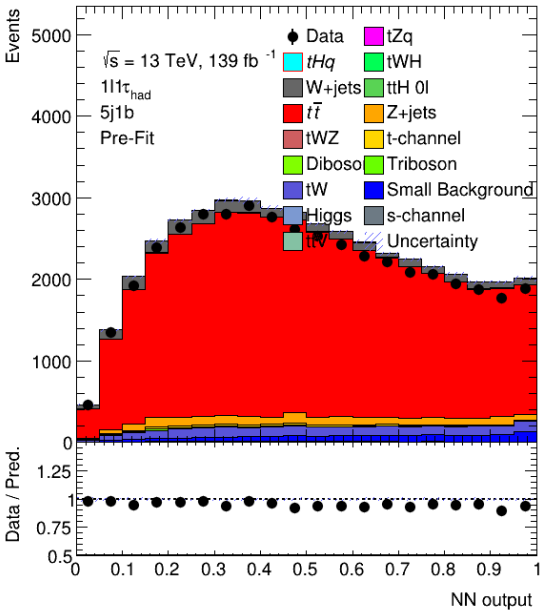
$2j1b$



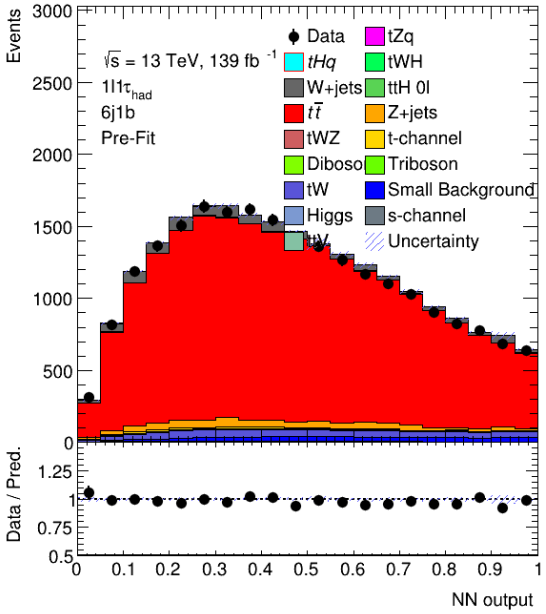
$3j1b$



$4j1b$



$5j1b$



$6j1b$

List of Figures

2.1 An overview of the Standard Model of Particle Physics . . . 4
2.2 Feynman diagram for the electron-positron scattering process (Bhabha scattering). . . 7
2.3 Pile-up distributions recorded at ATLAS during Run-2 [12]. . . . 11
2.4 Summary of the latest direct top-quark mass m_t, here refered to as m_{top}, measurements from ATLAS [18]. . . . 13
2.5 Leading Order (LO) Feynman diagrams for the top-quark pair production: gluon induced s-channel (left), t-channel (middle) and quark-antiquark annihilation ($q\bar{q}$) (right). . . . 14
2.6 Summary of the cross-section measurements by LHC and Tevatron as a function of different center-of-mass energies. The results were performed at next-to-next-to leading order (NNLO) with next-to-next-to-leading-log (NNLL) soft gluon resummation assuming a top-quark mass of 172.5 GeV [21]. . . . 16
2.7 Leading Order (LO) Feynman diagrams for single top-quark production: single top-quark production in the s-channel (left), the t-channel (middle) and the tW-channel (right). . . . 16
2.8 LO Feynman diagram for the tZq processes. . . . 18
2.9 Feynman diagrams for the tau lepton decay modes. On the left, the leptonic and on the right, the hadronic tau decay. . . . 20
2.10 Summary of the branching fractions for the most common hadronic tau lepton channels as well as the numbers of charged particles involved [40]. . . . 20

3.2 Schematic cut-away view of the ATLAS Detector and its four major components: the Inner Detector (ID), calorimeter, muon spectrometer and magnet system [47] 24
3.3 Schematic of the ATLAS Inner Detector and its three major components: the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker [53] . . . 25
3.4 A 3D visualisation of the barrel structure and radial distance coverage of the individual parts of the inner detector. The illustration shows the beam pipe, the IBL, the layers of the pixel detector, the four cylindrical layers of the SCT and the tree layers of TRT barrel modules [56]. . . . 26
3.5 Overview of typical signatures from particles that pass through the different components of the ATLAS detector [73]. . . . 30
3.6 Illustration of a hadronically decaying tau lepton resulting in a jet of three charged pions (3-*prong* decay) and a neutral pion [84]. . . . 32
3.7 Signature of a typical QCD event resulting in a jet of hadrons [84]. . . . 32
3.8 (a) 1-prong . . . 35
3.9 (b) 3-prong . . . 35

3.10 Example distribution for for one discriminating variable $m_{\text{EM+track}}$ used for the BDT tau identification approach: (a) shows the distribution for 1-prong, (b) the distribution for 3-prong. The red circles indicate the signal from tau leptons and the black squares show the multi-jet background [87]. 35

4.1 LO Feynman diagramm for the $t\bar{t}$ process. 41
4.2 LO Feynman diagramm for the Z+jets process with additional b-jets. 41
4.3 LO Feynman diagramm for the diboson WZ process. 41
4.4 LO Feynman diagramm for the diboson ZZ process. 42

5.1 Basic architecture of a perceptron. The perceptron receives multiple input values (x_i), each assigned by a weight (w_i) which represents the importance of the respective input. An activated function, denoted as h, maps the weighted sum of all inputs and weights together with an applied bias (b) to produce the output. 47
5.2 A multi-layer perceptron example with an n-dimensional input layer, three hidden layers and an output layer. Modified illustration based on [96] 49
5.3 S-Shape of the sigmoid function and its derivative. 50
5.4 S-Shape of the hyperbolic tangent function. 51
5.5 Shape of the ReLU function. 51
5.6 Shape of the binary cross-entropy cost function (BCE) for $y = 0$ and $y = 1$. 52
5.7 Illustration of a standard neural network on the left and a neural network with dropout on the right [108]. 54
5.8 Schematic depicting mixed input samples M_1 and M_2 for the CWoLa classifier [94]. . 57
5.9 A comparison of the AUC value as a function of the signal fraction f_1 for full supervision, LLP and CWoLa. The images show the behaviour for different training sizes: (a) $N_{train} = 100$, (b) $N_{train} = 1000$ and (c) $N_{train} = 10000$ [94]. 58

6.1 MC Composition Plots for two examples variables used for training the baseline classifier: (a) E_{lep1}, (b) E_{lep2}. 62
6.3 (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve 64
6.4 Prediction stacked plots for the neural network output on unseen MC samples. 65
6.5 Prediction stacked plots for the neural network output for three different b-jet configurations: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets. 66

7.1 Distribution for the RNNScore. The distribution is flat for true hadronic taus, while the distribution decreases for fake taus with a higher RNNScore. Here the loose working point is used with a signal efficiency of 85% for 1-prong tau decays and 75% for 3-prong decays, which explains the appearance of two peaks in the distribution. . 68
7.2 Ratios for true and fake: (a) for the scan from the left to the right side for the fake enriched region and (b) for the scan performed from the right to the left side for the true enriched region. 69
7.3 (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve 72
7.4 (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve 73
7.5 Prediction stacked plots for the neural network output on unseen MC samples. 74

7.6 Prediction stacked plots for the neural network output for three different b-jet configurations for the CWoLa MC classifier: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets. 75
7.7 (a) ROC-Curve, (b) Response-Curve and (c) Loss-Curve 78
7.8 Prediction stacked plot for the CWoLa classifier trained on data for unseen MC samples. 79
7.9 Prediction stacked plots for the neural network output for three different b-jet configurations for the CWoLa data classifier: (a) for 0 b-jets, (b) for 1 b-jets and (c) for 2 b-jets. 80

A.3 ROC-curve for the baseline classifier trained on 1-prong tau decays. 106
A.4 Response-curve for the baseline classifier trained on 1-prong tau decays. 106
A.1 Monte Carlo Composition Plots 1. 109
A.2 Monte Carlo Composition Plots 2. 110
A.5 ROC-curve for the baseline classifier trained on 3-prong tau decays. 111
A.6 Response-curve for the baseline classifier trained on 3-prong tau decays. 111

List of Tables

2.1 Summary of the measured cross-sections (σ) for the different single top-quark production subprocesses at different center-of-mass energies $\sqrt{s} = 7$ TeV, $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV, published by the ATLAS and CMS Collaborations [7, 18]. 15

3.1 Summary of the defined working points with fixed true hadronic tau selection efficiencies and the corresponding background rejection factors for misidentified taus in dijet events [88]. 36

4.1 Overview of the event selection for the $1e/\mu + 1\tau_{\text{had}}$ channel. 40
4.2 Overview of the integrated luminosities of LHC Run 2 for different years of operation [30]. 42

6.1 Available $t\bar{t}$ MC samples (DSID:410470) from v31 single-top ntuples, split by true and fake events, used for training the baseline classifier. The left table 6.1(a) shows the samples containing true hadronic taus, the right table 6.1(b) shows samples with events labeled as fake. 60
6.2 Complete list of selected training variables for the baseline classifier. The variables can be grouped into kinematic variables, hadronic tau specific variables and combined variables. 61
6.3 Overview of the search range of hyperparameters for the baseline classifier trained on MC, together with the final set of chosen optimal values for the baseline classifier. . . 62
6.4 List of variables (features) ranked by significance for the baseline classifier. Higher values of significance show that this variable is more likely to have a higher impact for a better classification performance. 63

7.1 Available $t\bar{t}$ MC samples (DSID:410470) from v31 single-top ntuples, used to train the CWoLa classifier on MC. 68
7.2 Summary of the cuts applied to the two mixed samples M1 and M2 along with the ratio of true and fake taus in each mixed sample. In addition to that, the total number of events in each sample is shown. 69
7.3 List of variables (features) ranked by significance for the CWoLa classifier trained on MC. Higher values of significance show that this variable is more likely to have a higher impact for a better classification performance. 70
7.4 Selected list of training variables for the CWoLa classifier trained on MC. 71
7.5 Overview of the search range of hyperparameters for the CWoLa classifier trained on MC, together with the final set of chosen hyperparameters. 71

7.6 Event yields for each data sample after RNN selection. Table 7.6(a) shows the event yields for mixed sample M1 with the applied cut RNN < 0.19, table 7.6(b) shows the event yields for mixed sample M2 with the applied cut RNN > 0.90. 76
7.7 Selected list of training variables for the CWoLa classifier trained on data. 77
7.8 Overview of the search range of hyperparameters for the CWoLa classifier trained on data, together with the final set of chosen hyprparameters for training. 77

A.1 $t\bar{t}$ Data Samples (Run 2) 93
A.2 Monte Carlo Simulation v31 Samples (Run 2) [115] 105
A.3 Overview of the hyperparameters used for the baseline classifier for 1-prong tau decays trained on MC. 111
A.4 Overview of the hyperparameters used for the baseline classifier for 3-prong tau decays trained on MC. 111