

Bachelor Thesis



Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

Application of Machine Learning for the Higgs Boson Mass Reconstruction Using ATLAS Data

Adam Herold

Supervisor: prof Dr. Ing. Jan Kybic
Supervisor–specialist: doc. Dr. André Sopczak
Field of study: Cybernetics and Robotics
January 2022

CERN-THESIS-2022-012
01/02/2022



Acknowledgements

Thank you to my supervisor, Dr. André Sopczak, who has given me a lot of his time and advice for which I am grateful. Thank you to my supervisor, prof. Jan Kybic, who has consulted my work with me. And thank you to my family, especially my mom, whom I love dearly.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, January 4, 2022

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze, 4. ledna 2022

Abstract

This thesis deals with the reconstruction of the Higgs boson mass decaying in the $2lSS + 1\tau_{had}$ channel in the $t\bar{t}H$ production. Based on the reconstructed mass, the goal is to separate the signal from background productions such as the $t\bar{t}Z$.

The data created by the full ATLAS detector simulation are used to develop two neural networks. First, a classification neural network that organizes the data by assigning detected particles to corresponding positions in the channel.

Second, a regression neural network that reconstructs the mass of the Higgs boson. The developed neural network is then tested on different data selections and is shown to outperform the Missing Mass Calculator technique.

Finally, the neural network is tested on real ATLAS data.

Keywords: CERN, ATLAS, Higgs boson, Mass reconstruction, Neural networks

Supervisor: prof Dr. Ing. Jan Kybic

Supervisor–specialist: doc. Dr. André Sopczak

Abstrakt

V této práci se zabýváme rekonstrukcí hmotnosti Higgsova bosonu v rozpadovém kanálu $2lSS + 1\tau_{had}$ v produkci $t\bar{t}H$. Na základě rekonstruované hmotnosti separujeme signál od pozadí, kterým je například produkce $t\bar{t}Z$.

Na datech ze simulace detektoru ATLAS vyvineme dvě neuronové sítě. Nejprve klasifikační neuronovou síť, která data uspořádává přiřazením částic do jednotlivých pozic v kanále.

Poté neuronovou síť, která rekonstruuje hmotnost Higgsova bosonu. Tuto síť testujeme na různých selekcích dat a ukazujeme, že dosahuje lepších výsledků než technika Missing Mass Calculator.

Na závěr je proveden test na skutečných datech z detektoru ATLAS.

Klíčová slova: CERN, ATLAS, Higgsův boson, rekonstrukce hmotnosti, neuronové sítě

Překlad názvu: Aplikace strojového učení pro odhad hmotnosti Higgsova bosonu z dat detektoru ATLAS

Contents

Introduction	1	4 Results	37
1 Theoretical background	3	4.1 Particle assignment	37
1.1 CERN	3	4.2 Mass reconstruction	41
1.1.1 ATLAS	4	4.2.1 Metrics	42
1.2 Standard Model Particles	4	4.2.2 Mass reconstruction results . .	43
1.2.1 Quarks	5	4.2.3 Additional background	44
1.2.2 Leptons	6	4.2.4 Future improvements	47
1.2.3 Neutrinos	6	4.3 Real ATLAS data	47
1.2.4 Representation of Particles . .	6	4.3.1 Particle assignment	48
1.2.5 Invariant mass and angular distance	7	4.3.2 Mass reconstruction	49
1.2.6 Missing transverse energy . . .	7	Conclusion	53
1.3 Higgs Boson	8	A Figures and tables	55
1.3.1 Production and Decay Channels	8	B Real data selection criteria	57
1.3.2 Missing Mass Calculator . . .	10	C Bibliography	59
1.4 Artificial neural networks	11		
1.4.1 Activation functions	12		
1.4.2 Loss function	12		
2 Data	15		
2.1 Analysis levels	15		
2.2 Detector effects	15		
2.3 ROOT	16		
2.4 Provided datasets and selections	16		
2.4.1 Used variables	18		
3 Proposed methods	21		
3.1 Task at hand	21		
3.1.1 Data extraction code	23		
3.2 Data augmentation	23		
3.3 Particle assignment neural network	24		
3.3.1 Lepton and jet permutations	25		
3.3.2 Applicability	26		
3.3.3 Features	26		
3.3.4 Labels	28		
3.3.5 Architecture and hyper-parameters	28		
3.3.6 Particle assignment code . . .	31		
3.4 Mass reconstruction neural network	31		
3.4.1 Mass reconstruction goal . . .	32		
3.4.2 Loss function	32		
3.4.3 Features	34		
3.4.4 Architecture and hyper-parameters	34		
3.4.5 Mass reconstruction code . . .	35		

Figures

1.1 Schematic of the LHC	3
1.2 Model of the ATLAS detector with its distinct layers	4
1.3 Particles of the Standard Model .	5
1.4 Diagram of the $2\ell SS + 1\tau_{had}$ decay channel	9
1.5 Activation functions	13
2.1 Root file diagram	17
3.1 Simplified diagram of the task pipeline	21
3.2 Three data selections used with the NNs	22
3.3 Picture of the detector with a coordinate system	24
3.4 Diagram of the particle assignment process	25
3.5 Delta R separation of correctly and incorrectly paired b jets	29
3.6 The particle assignment neural network diagram	29
3.7 The mass reconstruction neural network diagram	34
4.1 Particle assignment NN confusion matrices	37
4.2 Mass histograms — loss functions and MMC	43
4.3 Mass histograms — additional background	45
4.4 Mass histograms — all data	46
4.5 Mass histograms — scaled to expected number of events	47
4.6 Lepton assignment with reduced features	48
4.7 Simulated data mass with reduced features	49
4.8 Reconstructed mass of real ATLAS events	50
4.9 Mass of the best assigned real events	51
A.1 Mass histograms - data augmentation	56
B.1 Selection used with the real ATLAS data in Python code	57

Tables

1.1 Approximate invariant masses of notable particles	7
2.1 Number of events	17
2.2 Number of expected events of productions in real ATLAS data ..	18
2.3 Number of expected events of productions in real ATLAS data ..	18
3.1 Number of events after selection	23
3.2 Number of event permutations in relation to number of jets	26
3.3 Number of permutations	26
3.4 Specifications of the particle assignment CNN	30
3.5 Weights used in the loss function of the particle assignment CNN	31
3.6 Specifications of the mass reconstruction NN	35
4.1 Particle assignment Matthews correlation	38
4.2 Particle assignment mean of correct assignment	39
4.3 Particle assignment relative mean of correct assignment	39
4.4 Percentage of perfectly assigned events	40
4.5 Percentage of unsolved events by the MMC	41
4.6 Mass reconstruction metrics — loss functions	44
4.7 Mass reconstruction metrics — additional background	45
4.8 Mass reconstruction metrics — additional background	46
4.9 Lepton assignment Matthews correlation with reduced features .	48
4.10 Mass reconstruction with reduced features	50
4.11 The best reconstructed real $t\bar{t}H$ events	51
A.1 Neural network architecture experiments - width	55
A.2 Neural network architecture experiments - depth	55

A.3 Neural network architecture	
experiments - skip connections . . .	55
A.4 Mass reconstruction metrics - data	
augmentation	56



Introduction

In the ATLAS detector, the Higgs boson can be produced alongside a pair of top quarks. As the Higgs boson is short-lived, it decays before it can be detected [1]. The decay products of the Higgs boson include visible particles such as quark jets and leptons but also the undetectable neutrinos, which make the reconstruction of the Higgs boson and its mass a challenging task.

While the Higgs boson mass is known to be 125.18 ± 0.16 GeV [2], the reconstruction of its mass can help us separate events, in which it is created, from background events in which Z boson, W boson or different particles are produced instead of the Higgs boson.

The Higgs boson, alongside the two top quarks, can decay in many different channels, and we will be focusing on a particular one — the $2lSS + 1\tau_{had}$ channel in which two same-charged leptons and one hadronical tau candidate are produced. With the decay narrowed down, we will first assign the detected jets and leptons to the Higgs or one of the top quarks to make the data structured. Then we reconstruct the mass from the organized data.

Both of these tasks will be done using machine learning — in particular, neural networks. Our goal will be to develop such neural networks that will allow us to reconstruct the mass of the Higgs boson and separate it from the background.

Chapter 1

Theoretical background

1.1 CERN

CERN (from French *Conseil Européen pour la Recherche Nucléaire*¹) is an organization focused on research in fundamental physics most notably through the usage of their world-class particle accelerator facilities [3]. It was established in the 1950s and since then has been a great contributor to the world of physics and science [4].

In 2008 the Large Hadron Collider (LHC) started up and to this day it remains the largest and most powerful particle accelerator in the world [5]. It consists of a two-ring hadron accelerator and collider built in a 27 km long tunnel and it is designed for proton beams collisions with a centre-of-mass energy of 14 TeV [6]. The schematic of the LHC is in Figure 1.1.

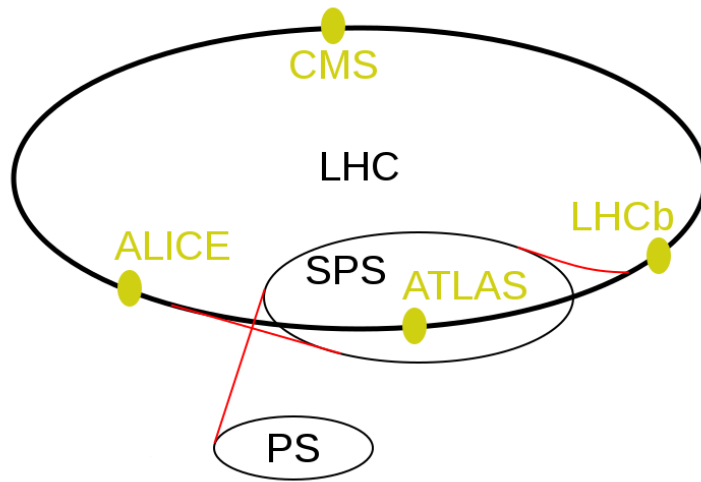


Figure 1.1: Schematic of the LHC

Schematic showing detectors CMS, LHCb, ATLAS and ALICE. Also showing other CERN accelerators — the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). Figure modified from source [7].

¹In English *European Council for Nuclear Research*

1.1.1 ATLAS

There are eight experiments operating at the LHC, focusing on different particles and using different detectors. The two largest experiments are ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid), both being independent general-purpose detectors [8].

The ATLAS detector is 44 meters long and 25 meters in diameter. Around 1 billion proton-proton collisions (events) occurs each second inside of the ATLAS detector. Each event produces multiple particles which are then detected by one of the many sensors of the detector. These measured events are then filtered by a hardware and a software trigger to a rate of around 2000 *interesting* events per second. The ATLAS detector measures properties of the particles such as their direction, momentum, charge, energy and the particle type [9].

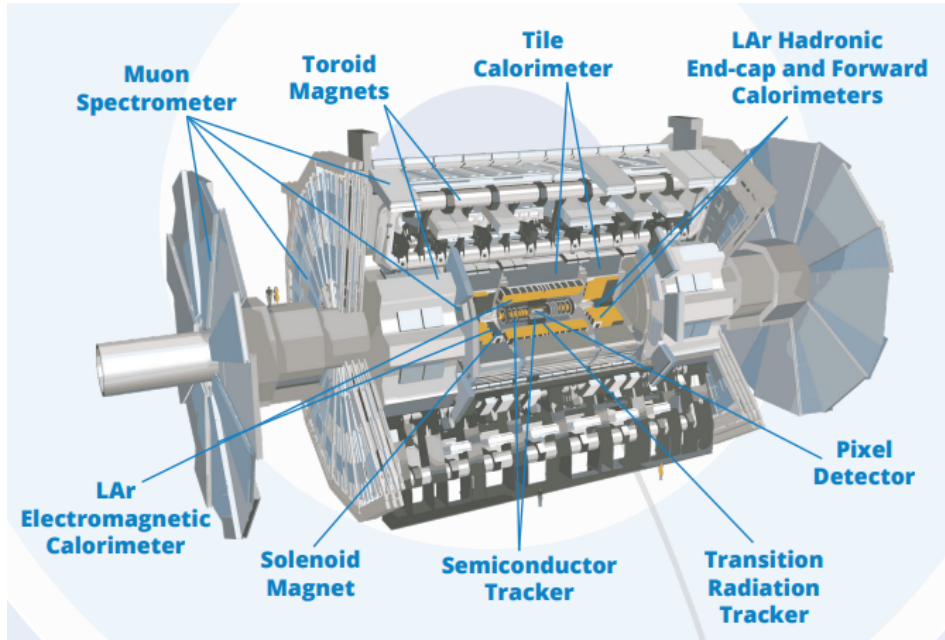


Figure 1.2: Model of the ATLAS detector with its distinct layers
Source [10].

1.2 Standard Model Particles

"The Standard Model of particle physics is the theory used to describe the interactions of fundamental particles (or fermions) and fundamental forces (which are conveyed by particles called bosons) [11]."

The Standard Model further divides fermions into quarks and leptons, each fermion also has an antimatter counterpart with opposite charge but otherwise same properties.

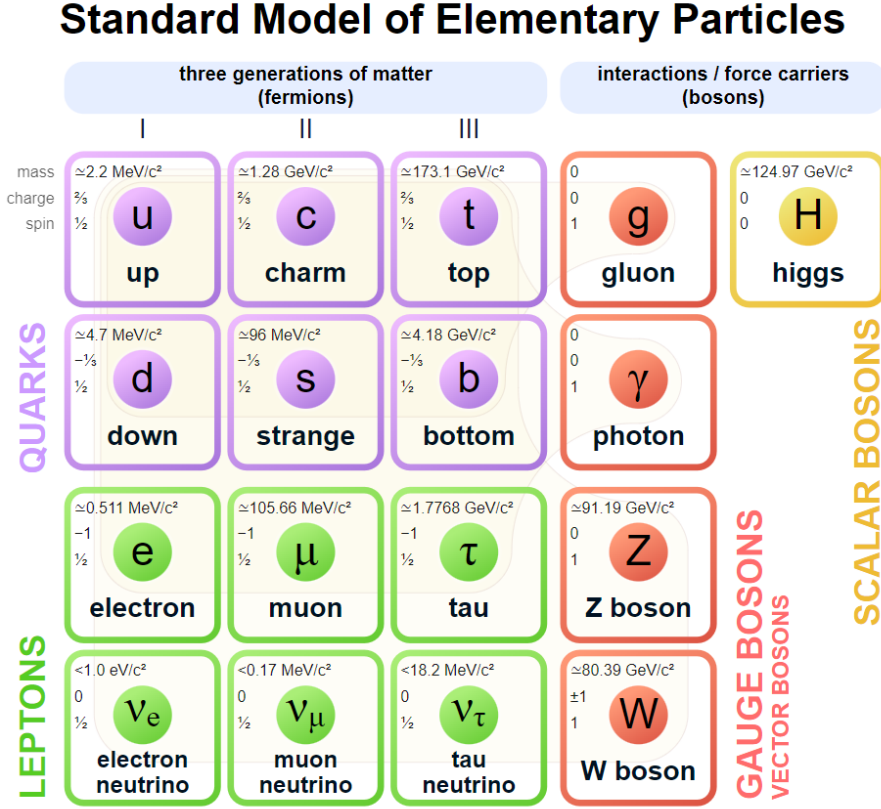


Figure 1.3: Particles of the Standard Model
Source [12].

1.2.1 Quarks

There are six different flavors of quarks: up (u) and down (d), charm (c) and strange (s), top (t) and bottom (b). In this thesis, we will differentiate between top, bottom and other quarks (abbreviated as *non-b* quarks).

The top quark is heavy enough to decay into a W boson and a b quark, which is the dominant channel. The W boson then decays either into a pair of quarks or a lepton and a neutrino [13, p. 638]:

$$t \rightarrow W^+ b \rightarrow q \bar{q}' b$$

$$t \rightarrow W^+ b \rightarrow \ell^+ \nu_\ell b$$

Quarks are never observed directly in the detector. Instead they are detected as sprays of hadrons called *jets*. Besides quarks, there is another source of jets, which is gluons. Discriminating between quark and gluon jets is a complex task and a focus of studies at ATLAS and CMS [14] [15].

1.2.2 Leptons

In the Standard Model, we differentiate between six leptons: electron (e), muon (μ), tau (τ) and their corresponding neutrinos (ν). In this thesis, we will use a different nomenclature, where when referring to leptons (symbol ℓ), we will only consider the two light leptons, electron and muon.

Much like quarks, taus are also not detected directly, as they have a short lifespan (2.8×10^{-13} seconds) and decay into a tau neutrino and a virtual W boson, which then decays either *leptonically*:

$$\tau^- \rightarrow W^- \nu_\tau \rightarrow \ell^- \bar{\nu}_\ell \nu_\tau \quad (1.1)$$

or *hadronically*:

$$\begin{aligned} \tau^- &\rightarrow W^- \nu_\tau \rightarrow h^- \nu_\tau \\ \tau^- &\rightarrow W^- \nu_\tau \rightarrow h^- h^+ h^- \nu_\tau \end{aligned} \quad (1.2)$$

where h is a hadron [16][17]. In the case of the leptonic decay, it is a non-trivial task to associate the detected lepton to either the tau decay or a different decay process (e.g. leptonic decay of a top quark).

1.2.3 Neutrinos

Neutrinos do not have charge, are nearly massless and very hard to detect, as they only interact weakly [18]. In ATLAS, neutrinos are not detected and, as such, are a source of missing energy in the detected decay process.

Because of this missing energy, it is a challenging task to reconstruct the mass of any particle with a neutrino as one of its decay products.

1.2.4 Representation of Particles

In this thesis, particles will be represented in two interchangeable ways, both of them being a vector of four values, that fully describes the particle kinematics.

First is the $(p_T, \eta, \phi, E)^T$ vector, where $p_T = |\vec{p}_T| = |(p_X, p_Y)^T|$ is the transverse momentum, η is pseudorapidity, ϕ is azimuthal angle and E is energy. This is the representation in which the data of the particles is stored.

Second is the momentum and energy vector $(p_X, p_Y, p_Z, E)^T$, also called the four-vector, where $(p_X, p_Y, p_Z)^T$ is the momentum of the particle in Cartesian coordinates. The variables in this representation will be used in the NN. This representation also has one considerable advantage, which is the fact that momenta and energy are conserved according to the laws of conservation. This allows us to add children particles together to create their parent particle. For example, we can write the following equation for the Higgs boson:

$$H = \tau^+ + \tau^-, \quad (1.3)$$

where H , τ^+ and τ^- are the four-vectors representing those particles.

The relations allowing us to switch between the two mentioned representations without losing information are the equations [19, p.26]

$$\begin{aligned} p_X &= p_T \cdot \cos \phi, \\ p_Y &= p_T \cdot \sin \phi, \\ p_Z &= p_T \cdot \sinh \eta. \end{aligned} \tag{1.4}$$

1.2.5 Invariant mass and angular distance

There are two more particle characteristics we will be using — invariant mass (m_0) and angular distance (ΔR).

The invariant mass is a mass independent of the reference frame of the particle's momenta and energies [20]. It is calculated from the momenta and energy as [19, p.26]

$$m_0 = \sqrt{E^2 - p_x^2 - p_y^2 - p_z^2}. \tag{1.5}$$

Approximate invariant masses of notable particles are in Table 1.1 [2].

Table 1.1: Approximate invariant masses of notable particles

Particle	Invariant mass
Higgs boson	125.18 GeV
Z boson	91.19 GeV
W boson	80.38 GeV
Tau	1.78 GeV
Neutrinos	0.00 GeV

The angular distance between two particles is the angle between their momentum vectors. It is calculated from the difference between their respective η and ϕ as [19, p.22]

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{1.6}$$

1.2.6 Missing transverse energy

The missing transverse energy is a characteristic of a whole event. It stems from the fact that the two protons which collide in the detector come along the beam pipe and thus only have non-zero momentum in the z-axis

$$\vec{p}_p = (0, 0, p_Z)^T \quad p_Z \neq 0.$$

After their collision the total momentum has to be conserved, meaning that the sum of the momentum vectors of all the particles that are created in the collision has to be equal to the sum of the two proton momentum vectors

$$\sum_{created} \vec{p} = \vec{p}_{p_1} + \vec{p}_{p_2}.$$

In reality, when we sum the momentum vectors of all particles detected in an event, the sum will usually have the x and y component non-zero:

$$\sum_{detected} \vec{p} = (p_X, p_Y, p_Z)^T \quad p_X, p_Y \neq 0.$$

This can be attributed to particles that escaped the detector undetected — notably the undetectable neutrinos. We then define the missing transverse energy as [21]

$$\cancel{E}_T = (E_{T_X}, E_{T_Y})^T = - \sum_{detected} \vec{p}_T, \quad (1.7)$$

where *detected* symbolizes the set of all detected particles.

1.3 Higgs Boson

To explain why the carriers of the weak nuclear interaction — the W and Z bosons — have masses, while in theory, they should be massless, in 1964, the Brout-Englert-Higgs (BEH) mechanism was proposed. This mechanism required a new, yet-to-be-discovered field and its associated particle — the Higgs boson [22].

On 4 July 2012, the Higgs boson has been confirmed by the ATLAS and CMS experiments at CERN, when a new particle with a mass around 125 GeV was observed [22].

Since then, more experiments and studies have been carried out to further explore its properties.

1.3.1 Production and Decay Channels

As our primary focus is reconstruction of the Higgs boson mass, we will be distinguishing between *signal* and *background* events. Signal events are the ones in which Higgs boson is produced and background events are the ones where either a Z boson, W boson or other particles are produced instead.

Signal events

The Higgs boson is produced in the LHC mainly through gluon fusion (ggF) or vector boson fusion [23]. Along with the Higgs, there are often other particles produced. In this thesis, we focus on the case, where the Higgs is produced together with a pair of top quarks ($t\bar{t}H$ production):

$$gg \rightarrow t \bar{t} H.$$

This is a production that has been observed recently in 2018 [24].

The Higgs boson has a very short lifetime (1.6×10^{-22} seconds) thus it decays before it can be detected [1]. As the Higgs boson has a great invariant

mass (approx. 125 GeV) it can decay into a pair of bosons or a pair of fermions, for example [23]:

$$\begin{aligned} H &\rightarrow b \bar{b}, \\ H &\rightarrow W W, \\ H &\rightarrow \tau \tau. \end{aligned} \quad (1.8)$$

The last mentioned channel will be the one we will be focusing on.

As the taus also decay (leptonically or hadronically), we will narrow our decay channel even more to a channel with two same charge leptons and a hadronically decaying tau (the $2\ell SS + 1\tau_{had}$ channel) in its final state².

Lastly, we will narrow down the decay of the top quarks to the *lepton+jets* case (Sec. 1.2.1 — one top decays into a pair of quarks and the other one into a lepton and a neutrino, both top decays also include a b quark):

$$t \bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow q \bar{q}' b \ell'^- \bar{\nu}_\ell \bar{b}. \quad (1.9)$$

An example³ Feynman diagram of this channel is shown in Fig. 1.4.

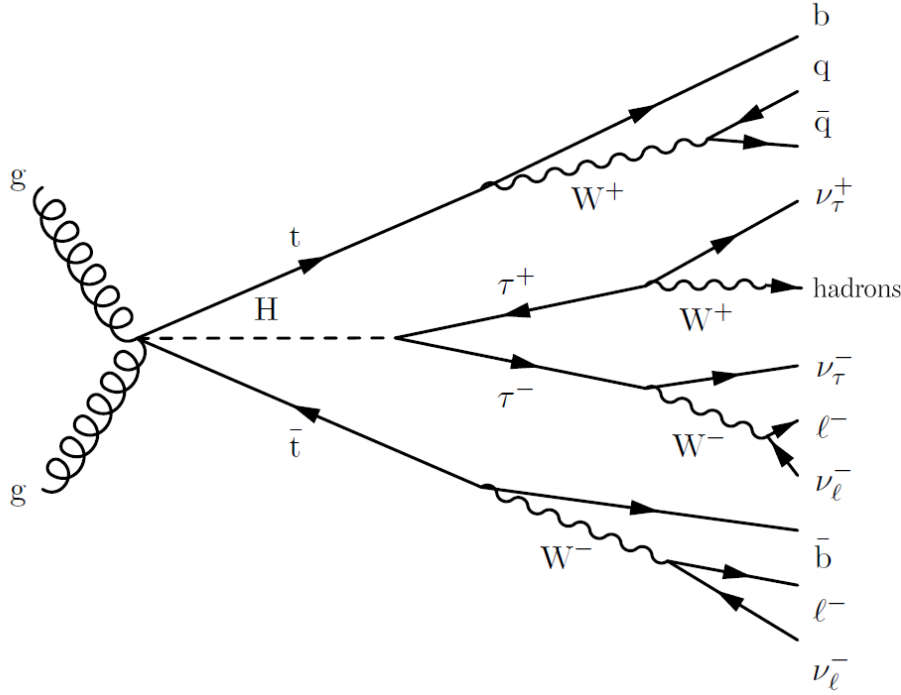


Figure 1.4: Diagram of the $2\ell SS + 1\tau_{had}$ decay channel

The final state particles on the right side of the diagram are the ones detected, except for the undetectable neutrinos. Figure modified from source [19, p.23].

²Final state particles are the ones that are detected by ATLAS.

³There can be slight differences on a case by case basis, such as a permutation of the top and anti-top pair decay or the permutation of the positively and negatively charged tau pair decay.

■ Background events

The Z and W bosons are produced in quark-antiquark annihilations in the LHC [25]. Similarly to the Higgs boson, they are also produced alongside a pair of top quarks ($t\bar{t}H$ production):

$$\begin{aligned} q + \bar{q} &\rightarrow t \bar{t} Z, \\ q + \bar{q} &\rightarrow t \bar{t} W. \end{aligned} \tag{1.10}$$

Another similarity between the Higgs boson and the W and Z bosons is their short lifetime (approx. 3×10^{-25} seconds), which means they also decay before they can be detected.

The Z boson can decay into a lepton-lepton pair, a neutrino-neutrino pair, a tau-tau pair or into hadrons:

$$\begin{aligned} Z &\rightarrow \ell \bar{\ell}, \\ Z &\rightarrow \nu \nu, \\ Z &\rightarrow \tau \tau. \end{aligned} \tag{1.11}$$

The W boson, on the other, hand mainly decays into a lepton-neutrino pair, a tau-neutrino pair or into hadrons [2]:

$$\begin{aligned} W &\rightarrow \ell \nu_{\ell}, \\ W &\rightarrow \tau \nu_{\tau}, \\ W &\rightarrow \text{hadrons}. \end{aligned} \tag{1.12}$$

The important decay mode here is the tau-tau pair, which is available for the Higgs and Z bosons, but not for the W boson. This means that the diagram 1.4 can be also used to describe the decay of the $t\bar{t}Z$ (with the replacement of H for Z in the diagram), but it cannot be used for the $t\bar{t}W$.

Other background productions exist such as $t\bar{t}$. All these background productions can decay in the $2\ell SS + 1\tau_{had}$ channel and as such cannot be easily separated from the signal. The separation of signal and background will be part of our task.

■ 1.3.2 Missing Mass Calculator

Methods for reconstruction of the $\tau\tau$ mass exist, and we will focus on one of them — the Missing Mass Calculator (MMC) — which outperforms other common methods [26, p.18].

The technique first assumes perfect detector resolution and no neutrinos outside of the $\tau\tau$ decay. We then have seven (for the case where one tau decays hadronically and one leptonically) unknowns: p_X , p_Y , p_Z and m^4 for the invisible product of each of the two taus. For the hadronic tau the invisible product is just one neutrino, so we can put its $m = 0$, reducing the

⁴Which is another representation of a particle, with mass instead of energy, similar to the ones we have described in Sec. 1.2.4

eight variables to seven. For these seven unknowns we have the following four momentum and mass conservation equations [26, p.5-6]:

$$\begin{aligned}
\not{E}_{TX} &= p_{mis_1} \sin \theta_{mis_1} \cos \phi_{mis_1} + p_{mis_2} \sin \theta_{mis_2} \cos \phi_{mis_2}, \\
\not{E}_{TY} &= p_{mis_1} \sin \theta_{mis_1} \sin \phi_{mis_1} + p_{mis_2} \sin \theta_{mis_2} \sin \phi_{mis_2}, \\
M_{\tau_1}^2 &= m_{mis_1}^2 + m_{vis_1}^2 + 2\sqrt{p_{vis_1}^2 + m_{vis_1}^2} \sqrt{p_{mis_1}^2 + m_{mis_1}^2} \\
&\quad - 2p_{vis_1} p_{mis_1} \cos \Delta\theta_{vm_1}, \\
M_{\tau_2}^2 &= m_{mis_2}^2 + m_{vis_2}^2 + 2\sqrt{p_{vis_2}^2 + m_{vis_2}^2} \sqrt{p_{mis_2}^2 + m_{mis_2}^2} \\
&\quad - 2p_{vis_2} p_{mis_2} \cos \Delta\theta_{vm_2},
\end{aligned} \tag{1.13}$$

where *mis* and *vis* symbolize the invisible and visible tau products respectively, $M_\tau = 1.777$ GeV as per Table 1.1 and $\Delta\theta_{vm_i}$ is the angular distance between the visible and invisible product of the *i*-th tau [26, p.6].

With seven unknowns and four equations, this is an under-constrained system and as such it does not have one exact solution. From all possible solutions, the MMC chooses the most likely one. It finds it with the help of additional information, such as "*...the expected angular distance between the neutrino(s) and the visible decays products of the τ lepton.*" [26, p.6] The probability density function of such angular distance is obtained from simulated data [26, p.7].

The MMC will serve as a comparison to our mass reconstruction method. A very important thing to note is that the assumption of no neutrinos outside of the $\tau\tau$ decay is not satisfied in our decay channel (a neutrino is coming from the anti-top branch in Fig. 1.4). This results in a drop in efficiency of the MMC, but it does not make it unusable, as the MMC tries to mitigate the effects of resolution in the measurement of $\not{E}_T$ [26, p.10] and the outside neutrino could be viewed as a source of larger resolution.

Lastly, the MMC reconstructs the $\tau\tau$ mass, therefore it is only applicable to events with:

$$\begin{aligned}
Z &\rightarrow \tau\tau, \\
H &\rightarrow \tau\tau,
\end{aligned} \tag{1.14}$$

and not at all to the $t\bar{t}W$ or $t\bar{t}$ production.

1.4 Artificial neural networks

"Artificial neural networks are popular machine learning techniques that simulate the mechanism of learning in biological organisms." [27, p.2] The building stone of a neural network (NN) is a neuron. In the artificial neural network (ANN) the neuron is represented by a computational unit, which takes weighted signals as input [27, p.3], processes them and outputs another signal, which can then serve as an input for other neurons, creating a network. The processing of the inputs will in our case be their addition and subsequent use of an activation function:

$$y = \Phi(\mathbf{w}^T \mathbf{x}). \quad (1.15)$$

To create a NN, the neurons are formed into layers. Besides an input and output layer, there will be other intermediate hidden layers [27, p.5]. In the NNs we will be using, the hidden layers will always be fully connected layers, meaning that each neuron takes as an input the output of each neuron in the preceding layer:

$$y_{i,j} = \Phi(\mathbf{w}_j^T \mathbf{y}_{i-1}), \quad (1.16)$$

where i is the index of the layer and j is the index of the neuron in the layer.

Neural networks can be used for different purposes. The two that will be relevant for us are a regression neural network (rNN⁵), which predicts one or multiple numerical values (e.g. predicting the mass of a particle), and a classification neural network (CNN), which predicts into which of several known categories does the data sample from which the input was generated belong to (e.g. categorization of a particle into quarks and leptons).

■ 1.4.1 Activation functions

Activation functions are used at the output of neurons. There are many different options, the ones we will be using are following (also illustrated in Fig. 1.5).

- Linear function calculated as [27, p.13]

$$\Phi(v) = v. \quad (1.17)$$

- Rectified linear unit function (ReLU) calculated as [27, p.14]

$$\Phi(v) = \max\{v, 0\}. \quad (1.18)$$

- Sigmoid function calculated as [27, p.13]

$$\Phi(v) = \frac{1}{1 + e^{-v}}. \quad (1.19)$$

■ 1.4.2 Loss function

In the training process of a neural network, two phases can be distinguished. First is the forward phase, when the inputs are processed by the neural network to produce outputs.

In the second — backward — phase, the loss function takes in the outputs of the neural network (i.e. the predictions) and the desired outputs (i.e. the

⁵Not to be confused with the recurrent neural network.

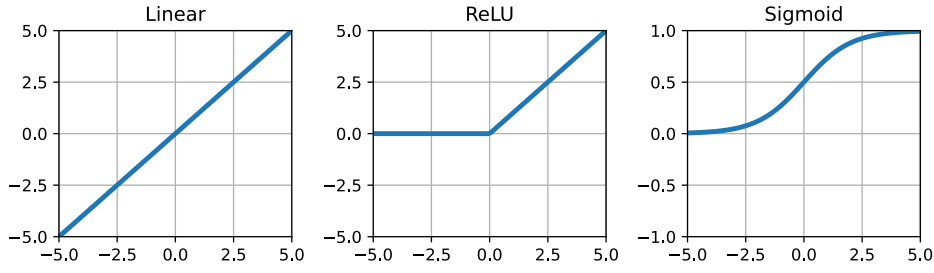


Figure 1.5: Activation functions

truth) and calculates a score that quantifies the quality of performance of the neural network. The neural network then backpropagates by computing the gradient of the loss function with respect to the weights of the layers. Finally, the weights are updated to minimize the loss [27, p.22][28]:

$$\mathbf{w}' = \mathbf{w} - \alpha \cdot \left(\frac{\partial \mathcal{L}}{\partial \mathbf{w}} \right)^T, \quad (1.20)$$

where α is the learning rate.

The choice of the loss function defines to some extent the functionality of the neural network.

For the regression a typical loss function is the mean squared error (MSE) [27, p.176]:

$$\mathcal{L} = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2. \quad (1.21)$$

For the classification into one of two classes (binary classification) the binary cross entropy is used [29].

$$\mathcal{L} = -\frac{1}{n} \sum_{k=1}^n y_k \cdot \log(\hat{y}_k) + (1 - y_k) \cdot \log(1 - \hat{y}_k). \quad (1.22)$$

Chapter 2

Data

2.1 Analysis levels

When it comes to the signal and background productions, there are two different levels of data on which we can study these events:

- **Real data** measured by the ATLAS detector.
- **Generated data** produced by a program (e.g. the Monte Carlo generator PYTHIA [30]). It can be further split into two levels:
 - **Event generator data** (also called *truth* data). It contains full information (p_T , η , ϕ and E) of each particle from the event as well as its *children* and *parent*¹ relations with other particles.
 - **Full ATLAS detector simulation** level of data. At this level, a program takes a generated event and aims to produce data similar to how it would be measured by the real ATLAS detector. The effects this has on the data will be presented in the next section. It is this level of data on which we will study the reconstruction of the Higgs boson mass before applying it to the real data.

2.2 Detector effects

The data produced by the detector simulation contains less information than we have on the event generator level. This is caused by different effects that are taking place in the detector.

The first group of effects stems from the physical nature of the decay process:

- We only detect final state particles (i.e. particles without children in Fig. 1.4).
- Neutrinos are not detected.

¹Children meaning particles it decays into and parent meaning the particle from which it came.

- We do not know the parents of the detected particles. That is, if we detect two leptons in an event, we do not know which lepton is coming from the *Higgs branch*² and which is coming from the *top branch*.
- Quarks are detected as jets.
- More jets than four can be detected, because of gluon jets.
- Less jets than four can be detected, because two or more jets can overlap and be detected as one.

The second group consists of effects associated with the imperfection of the detector:

- There are no sensors in the beam pipe³ and particles produced in the decay can escape the detector's sensors in this direction.
- Particles in the detector can overlap with each other and can be misidentified or one object can be identified as multiple particles. To avoid the latter, there is an overlap-removal procedure, where only certain particles are kept, as part of the detector simulation data processing [31], so in our case this is taken care of, but still it remains a source of more uncertainty and error between the event generator and detector simulation data.
- Detector resolution defined by the ATLAS online glossary as the "*measure of the accuracy of a detector measurement, e.g. of energy or spatial position*" [32].

These effects present obstacles in the process of mass reconstruction. The above is not necessarily a complete list, rather a list of effects which were encountered during the work on this thesis.

2.3 ROOT

ROOT is a framework for working with data created at CERN. Natively, it can be used to write and run C++ programs with the built-in Cling interpreter, but it can also be used with Python through the PyROOT library. It can be used for storing and accessing data in a tree-like structure and plotting or other processing of data [33].

An example of the structure of data stored in a ROOT file is on Figure 2.1.

2.4 Provided datasets and selections

In this thesis two datasets provided by the ATLAS collaboration were used.

The first one contains data of the full ATLAS detector simulation (*detector available*) as well as the event generator (*truth*) level. It consists of multiple

²Higgs branch meaning decay products of the Higgs and their further decay products.

³The pipe through which the beam of protons travels in Fig. 1.2.

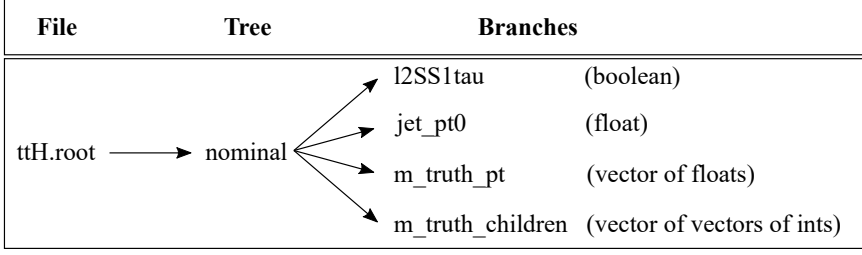


Figure 2.1: Root file diagram

ROOT files each containing events from either $t\bar{t}H$, $t\bar{t}Z$, $t\bar{t}W$ or $t\bar{t}$ production. The total number of available events in each of the named productions is in the *All events* column of Table 2.1. The individual file of these productions are then further separated based on the decay of the top quarks — as mentioned in Section 1.3.1, we will be mostly using the data of the *lepton+jets* decay (Eq. 1.9) of the top quarks.

The events in the dataset were generated with a Monte Carlo generator and the detector simulation is based on Geant4 [34].

As discussed in Section 1.3.1, in this thesis we are focusing on the $2lSS + 1\tau_{had}$ decay channel. Each event contains a detector available boolean variable indicating whether the event is of this specific channel. This variable is computed by requiring the event data to meet some conditions and as it relies on the detected objects, it is not always correct. In addition to that, we will be making a selection of events with at least three detected jets and at least one detected b jet, which is also made based on detector available variables. Making these selections allows for the data that we work with to be more consistent.

The number of events after the application of the selections is in Table 2.1.

Table 2.1: Number of events

Number of all events for each production and number of selected events. The selection requires the event to have the $2lSS + 1\tau_{had}$ decay channel tag and to have at least three detected jets with at least 1 b jet. These selections are based on detector available variables.

Production	All events	Selected events	Percentage selected
$t\bar{t}H$	1 055 628	73 741	6.99%
$t\bar{t}Z$	1 894 217	32 108	1.69%
$t\bar{t}W$	614 984	13 295	2.16%
$t\bar{t}$	252 225	6 027	2.39%
Total	3 819 054	125 171	3.28%

In the next chapter additional selections based on truth information will be introduced on top of the ones mentioned here.

The selected events were further split into three separate datasets for training, validation and testing of the neural networks. The ratio of the split

was 0.8, 0.1 and 0.1 respectively.

The second dataset contains real ATLAS data and as such it cannot be separated into the different productions, but instead it contains all of them. What can still be used, are the detector available selections of $2lSS + 1\tau_{had}$ channel and of the required number of three detected jets with one b jet.

The number of expected events of the different productions with this selection can be approximated and the value are listed in Table 2.3.

Table 2.2: Number of expected events of productions in real ATLAS data
The values were provided by my supervisor.

Production	Events
$t\bar{t}H$	22.8
$t\bar{t}Z$	18.2
$t\bar{t}W$	24.8
$t\bar{t}$	22.2
Other	17.9
Total	105.9

The real data was produced in the years 2015-2018. The number of events obtained each year is in Table A.2.

Table 2.3: Number of expected events of productions in real ATLAS data
The values were provided by my supervisor.

Year of production	Events
2015	4.04×10^6
2016	4.04×10^7
2017	5.28×10^7
2018	6.93×10^7
Total	1.63×10^8

2.4.1 Used variables

Here is an overview of the ROOT variables we will be using in the mass reconstruction and the particle assignment process. The chosen variables are closely related to the $2lSS + 1\tau_{had}$ decay channel, specifically to the final state particles in this decay (see Fig. 1.4).

The detector available variables we will be using are the following:

- The $2lSS + 1\tau_{had}$ decay channel tag and variables indicating the number of detected jets and b jets.
- Four-vectors of up to eight jets (the number differs in each event based on how many have been detected by the detector).

- For each jet a *b-tag* which indicates how likely is the jet coming from a b quark based on characteristics such as "*large mass*" or "*significant lifetime*" [35].
- Four-vectors of two leptons.
- Four-vector of a hadronically decayed tau.
- The decay mode of the hadronically decaying tau (either 1-prong or 3-prong⁴).
- Missing transverse energy characterized by its energy and azimuthal angle.
- The sum of the total visible transverse energy (a scalar variable).

Variables listed above are stored as scalars (e.g. each four-vector is stored in the form of four separate scalar variables).

The truth data on the other hand includes information of each particle occurring in an event. It is stored in the form of eight vectors for each event containing pt, eta, phi, E, ID, particle type, parent and children relations. The last mentioned is a vector of vectors as a particle can decay into multiple particles. Each particle is represented by an index at which the vectors can be accessed to obtain the particle's information.

⁴Decays into 1 or 3 charged particles.

Chapter 3

Proposed methods

3.1 Task at hand

The task at hand is to reconstruct the mass of the Higgs boson from the data produced by the ATLAS detector simulation on a per-event basis. First, the data has to be preprocessed — jets and leptons have to be assigned to their corresponding positions in the decay, so that the data is organized. Figure 3.1 shows a simplified diagram of the described task.

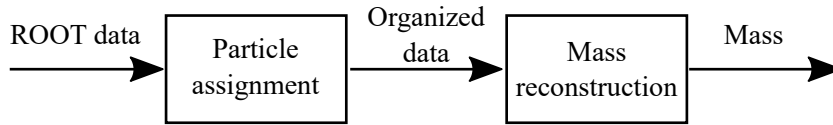


Figure 3.1: Simplified diagram of the task pipeline

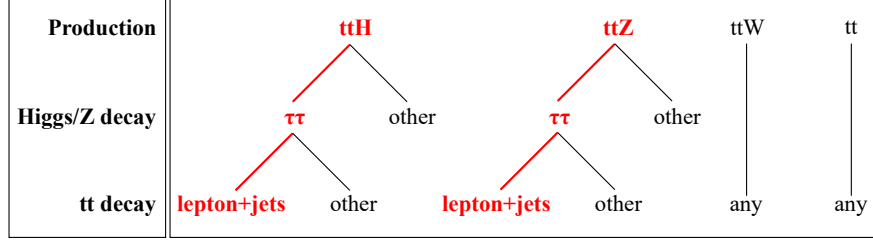
For both the particle assignment and the mass reconstruction a neural network will be used. For the first mentioned, a classification NN based approach inspired by the paper on jet-parton assignment in $t\bar{t}H$ events cited at [36], and for the latter a regression NN. For the rNN, three loss functions will be developed, each representing a different approach to the mass reconstruction.

We will be training and testing the NNs on three selections of the simulated data, which are based on truth information and are illustrated in Figure 3.2. More precisely, the particle assignment NN will be trained only on the *Narrow selection* in (a) in the figure, because it requires a consistent structure of the decay across all events, which is precisely what the Narrow selection ensures. A mass reconstruction NN will then be trained and tested on each of the three selections.

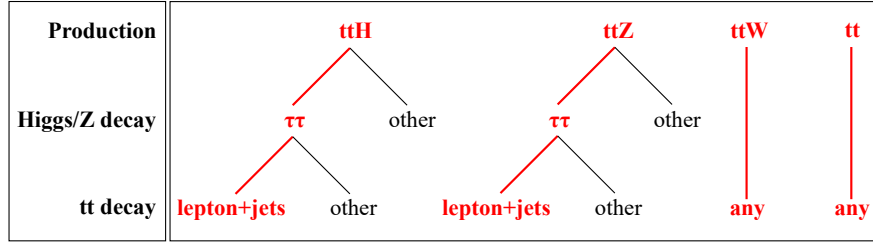
The choice of selections (a) and (b) is based on having the narrowed down decay of the $t\bar{t}H$ and $t\bar{t}Z$, that allows us to rely on the exact structure of the channel (Fig. 1.4), allowing for the assignment of all the jets and leptons. In addition to that, in the *Additional backgrounds selection* in (b) there is also the $t\bar{t}W$ and $t\bar{t}$ background, which cannot have the top pair decay and Higgs/Z boson decay cuts applied to it, since their decay channels are different (e.g. the W never decays into $\tau\tau$), so we opt to take all the data.

The *Real data selection* in (c) simulates the real ATLAS data structure, where the productions cannot be distinguished, therefore we take all of the data, without any truth-based selections.

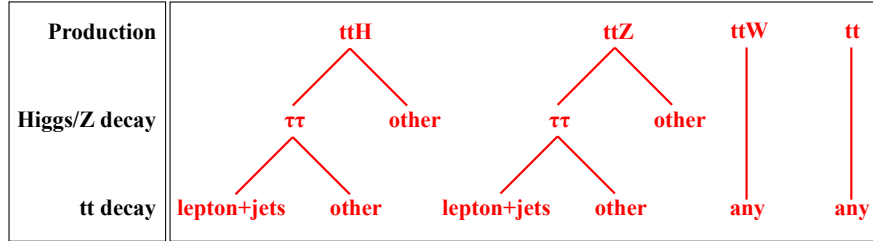
All three of these sets will be using the common selection of requiring the $2lSS + 1\tau_{had}$ channel tag to be true and at least three detected jets with one b jet, which has been discussed in Sec. 2.4. This selection will also be applied to the real ATLAS data, on which we will test the NN trained on dataset (c) from the figure.



(a) Narrow selection



(b) Additional backgrounds selection



(c) Real data selection

Figure 3.2: Three data selections used with NNs

In red is the selected data and in black are the cuts made for each selection. The *Narrow selection* is in (a) as described in Section 2.4. In (b) additional background is added in the form of $t\bar{t}W$ and $t\bar{t}$ productions. Selection simulating the real ATLAS data is in (c), where truth information cannot be used, therefore the data cannot be separated on the conditions in the figure. The selection of $2lSS + 1\tau_{had}$ channel and at least three jets with one b jet detected is used with all three (not shown in the picture).

The number of events for the Narrow selection in Figure 3.2 is in the Table 3.1. The number of events without applying the top pair and Higgs/Z boson decay selection (i.e. Real data selection in the figure) has already been stated in column *Selected events* in Table 2.1, but will be repeated here in Table 3.1

under the column *Real data selection*.

Table 3.1: Number of events after selection

Wide selection comprises of all events with the $2lSS + 1\tau_{had}$ channel and at least three jets with one b jet. *Narrow selection* is obtained by applying the requirement of $\tau\tau$ Higgs (or Z) decay and lepton+jets $t\bar{t}$ decay on the wide selection. Wide selection only uses detector available variables, while the narrow selection requires truth information.

Production	Real data selection	Narrow selection	Percentage narrow
$t\bar{t}H$	73 741	18 124	24.58%
$t\bar{t}Z$	32 108	10 886	33.90%
$t\bar{t}W$	13 295	-	-
$t\bar{t}$	6 027	-	-
Total	125 171	29 010	23.18%

3.1.1 Data extraction code

The Python code used for data extraction from ROOT ntuples to a format fit for the particle assignment NN is in the directory `/source_code/root_data_extraction`. It produces all three of the selections in Fig. 3.2.

In the same directory is also the code for extraction of selected events from the real ATLAS dataset.

3.2 Data augmentation

Data augmentation is a technique used in machine learning to avoid overfitting and to achieve better generalization of a NN on a dataset [27, p.335]. It is commonly used in convolutional neural networks in the way of altering (e.g. rotating, translating or squeezing) an image used as an input for the network. The principle of generating altered data, that could possibly occur in the original dataset (i.e. it follows the original data distribution) can be transferred to our case as well.

We can make use of the rotational symmetry of the detector about the beam pipe — the azimuthal symmetry [37]. While the symmetry is not perfect, as the detector is not fully homogeneous, it is still a valuable tool which brings us the benefits mentioned above.

Before being used as an input for the NN, each event will have its particles rotated about the beam pipe axis. This way the overall number of events stays the same, but the events differ across epochs. In practice this is achieved by changing the azimuthal angle of the particles by the same random value:

$$\phi' = \phi + \Delta\phi, \quad (3.1)$$

where ϕ is a vector of the azimuthal angles of the original particles in an event, ϕ' is the vector of the azimuthal angles of augmented particles and

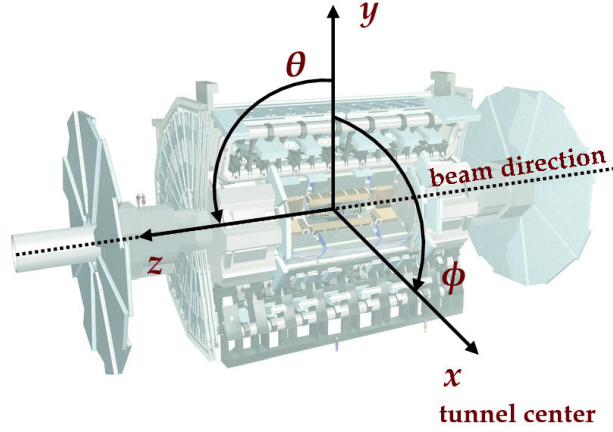


Figure 3.3: Diagram of the detector with a coordinate system
Source of the image is Fig. 4.5 in [38].

$\Delta\phi = (\Delta\phi, \Delta\phi, \dots)^T$ is a vector with the repeated value $\Delta\phi \in [0, 2\pi)$, which is randomly generated for each event.

As we immediately input the ϕ' into a goniometric function to compute the momentum (Eq. 1.4), we do not require $\phi' \in [0, 2\pi)$.

3.3 Particle assignment neural network

For the reconstruction of the Higgs boson mass we will be using a regression neural network. As input for the network, we want to use data organized in a way where we distinguish between two particles of the same type but coming from different parent particles. Another way to put it is, that we want to pair the positions at the decay diagram (Fig. 1.4) with the corresponding detected particles or that we want to assign each jet and lepton to one of the three decay branches (i.e. top, antitop and Higgs (or Z) branch).

The information that would allow us this (i.e. the information of the child-parent relations between particles) is unavailable in the detector as has been discussed in Sec. 2.2. This means that to organize the data of an event, we will have to choose one out of many possible ways the particles can be assigned to their positions. In this thesis, we call this process the particle assignment or particle association.

There are two types of particles that have to be assigned — jets and leptons. For the task of particle assignment, our goal is to create a program which takes the raw ROOT data of an event on input and outputs data with jets and leptons ordered correspondingly to their positions.

The proposed approach is a classification neural network, which for each event takes in each possible permutation of jets and leptons at different positions and outputs the respective probabilities of each of the positions being assigned correctly. The most likely assignment is then chosen from the permutations as the one with the largest product of the respective probabilities.

This process is schematized on Figure 3.4 which expands on Figure 3.1. The inspiration for using permutations as input for the neural network comes from the paper [36] mentioned in Section 3.1.

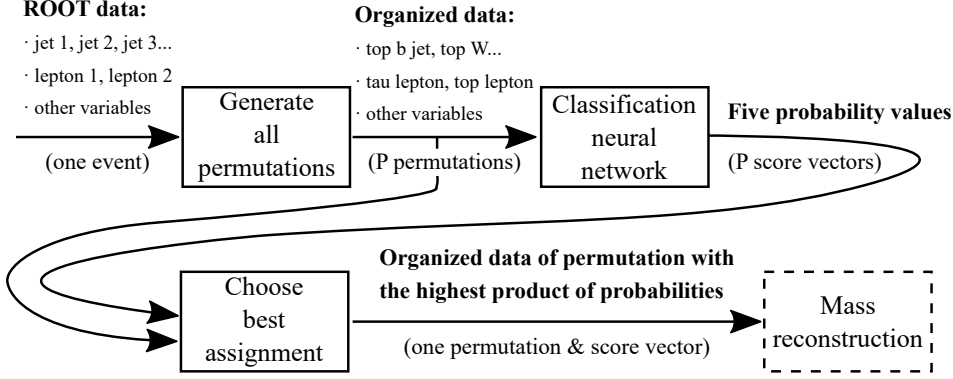


Figure 3.4: Diagram of the particle assignment process

First the ROOT data of an event has all possible permutations of the assignment of leptons and jets generated. The permutations are then processed by the (trained) NN, which assigns a score vector with five values to each permutation. The permutation with the highest product of the individual values is then chosen as the best particle assignment of the event. The event is then ready to have its mass reconstructed.

3.3.1 Lepton and jet permutations

Looking at the decay diagram (Fig. 1.4) we distinguish two lepton positions — one lepton originating from the tau and one lepton originating from the top. Together with the fact that the detector usually detects exactly two leptons, this gives us two possible permutations for each event.

For the jets we distinguish three positions. Two of them are the b jets coming from the top and anti-top respectively. The third one is the sum of the two non-b jets (effectively the W boson that the non-b jets come from, for this reason we will call it the top W), as we do not distinguish between these two, as their positions are interchangeable¹. The number of jet permutations can be quite substantial, depending on the number of detected jets (up to eight).

The number of permutations P of an event with two leptons and n detected jets is

$$P = 2 \cdot \frac{n(n-1)}{2} \cdot (n-2)(n-3), \quad (3.2)$$

where $\frac{n(n-1)}{2}$ is the number of combinations for the W jet pair and $(n-2)(n-3)$ is the number of permutations of the two b jets.

The numbers of permutations of an event for different numbers of jets, obtained by application of the Eq. 3.2, are in Table 3.2.

¹We however still keep their separate four-vectors so as to not lose any information.

Table 3.2: Number of event permutations in relation to number of jets

Jets	4	5	6	7	8
Permutations	24	120	360	840	1680

The total number of permutations, generated from the events of selection (a) from Figure 3.2, which are the events we will be training and testing the particle assignment NN on, is in Table 3.3.

Table 3.3: Number of permutations

The number of permutations generated from the events separated into the two distinct productions. The number of events stated is identical to the numbers in column *Narrow selection* in Table 3.1.

Production	Events	Permutations
$t\bar{t}H$	18 124	4 294 080
$t\bar{t}Z$	10 886	4 000 152
Total	29 010	8 294 232

3.3.2 Applicability

As the particle assignment relates closely to the exact structure of the decay, it is important to note that it can only be applied on the Narrow selection from Fig. 3.2. This selection will be used for the training and testing of the CNN.

Events of the other productions ($t\bar{t}W$ and $t\bar{t}\bar{t}$) and the differently decaying events of $t\bar{t}H$ and $t\bar{t}Z$ (e.g. the top quarks can both decay leptonically; the Higgs can decay into a pair of W bosons instead of taus etc. — these are the events we have removed by using the Narrow selection) will not be used in the evaluation of the CNN. But as we will use them in the mass reconstruction, these events will, in the end, also go through the particle assignment, even though it will be ineffective.

This is an unavoidable issue, because there is too many different productions and ways for them to decay, to have one united way of particle assignment. A possible different approach, would be organizing the particles in an entirely different way than by the positions in the decay. We have decided for the described approach, because our main focus is the narrow selection.

3.3.3 Features

There are 70 features on the input of the particle assignment NN. Most of them are the momentum and energy of final state particles (i.e. their four-vectors) and also their masses and angular distances calculated from these four-vectors (the equations for these calculations are in Section 1.2.5). The full list of variables, grouped by the physical nature of the variables, with

names of the particles based on their origin particle (e.g. anti-top lepton is the lepton originating from the anti-top), is following.

- Four-vectors $(p_X, p_Y, p_Z, E)^T$ of detected particles which are the top b jet, anti-top b jet, both top non-b jets, Higgs boson hadronically decaying tau² (Higgs tau), Higgs boson lepton and anti-top lepton.
- Four-vectors of intermediate particles of the decay, created by addition of selected particles from previous paragraph. Specifically the top W boson (sum of top non-b jets), top quark (sum of top b jet and top W boson), visible part of the Higgs boson (sum of Higgs tau and Higgs lepton) and visible part of the anti-top quark (sum of anti-top b jet and anti-top lepton).
- The mass of each detected particle or intermediate particle mentioned above.
- The angular distance between selected pairs of particles, the focus being primarily on the particles that are being assigned (the jets and the leptons). The pairs are the Higgs tau and Higgs lepton, the Higgs tau and anti-top lepton, the anti-top b jet and Higgs lepton, the anti-top b jet and anti-top lepton, the two top non-b jets, the top b jet and top W boson and, finally, the top quark and the anti-top quark.
- For each of the four jets a *b-tag*, which indicates how likely is the jet coming from a b quark.
- Missing transverse energy characterized by its $\cancel{E}_{T_X}$ and $\cancel{E}_{T_Y}$ components.
- The scalar sum of the total visible transverse energy and the scalar sum of the transverse energy of all detected jets.
- The number of jets with energy over 25 GeV.
- Lastly the decay mode of the hadronically decaying tau.

Inspiration in the choice of the features was in the thesis of Petr Urban [19] and the MMC, as the last three items on the list are variables also used in there. We will be using the same features in the mass reconstruction NN as well.

It should also be emphasized, that the names of the particles in the above list refer to the positions, not to the actual particles assigned to them, because finding the correct particles to assign to the positions is the task at hand.

²By this, we do not mean the tau itself, but rather the hadrons coming from the tau, which are detected.

3.3.4 Labels

Each permutation is labeled with a vector

$$l = (a_{b1}, a_{b2}, a_W, a_{l1}, a_{l2}),$$

$$a_p = \begin{cases} 1, & \text{if particle at position } p \text{ is assigned correctly,} \\ 0, & \text{if particle at position } p \text{ is assigned incorrectly,} \end{cases} \quad (3.3)$$

where $p \in \{b1, b2, W, l1, l2\}$ is one of the five positions we are assigning to.

The label for a permutation is created by comparing the assigned particle to the correct particle for that position, which is known from the truth information of the event, which is available in the dataset. As the particle is not detected perfectly, the detected and true particle have to be compared on some criteria to determine, whether they can be paired and the detected particle assigned to the position.

The criteria on which we compare this pair to decide, whether the assignment is correct, is the angular distance between the two [36, p.6]:

$$a_l = \begin{cases} 1, & \text{dist}(l_{true}, l_{assigned}) \leq 0.12, \\ 0, & \text{dist}(l_{true}, l_{assigned}) > 0.12, \end{cases}$$

$$a_b = \begin{cases} 1, & \text{dist}(b_{true}, b_{assigned}) \leq 0.32, \\ 0, & \text{dist}(b_{true}, b_{assigned}) > 0.32, \end{cases} \quad (3.4)$$

$$a_q = \begin{cases} 1, & \text{dist}(W_{true}, W_{assigned}) \leq 0.32, \\ 0, & \text{dist}(W_{true}, W_{assigned}) > 0.32, \end{cases}$$

where $\text{dist}(p_1, p_2)$ symbolizes the angular distance between two particles.

As the W position is a sum of two non-b jets (symbol q), we compare the two jets and if both meet the distance condition the W is labeled as correctly assigned.

The right-side threshold values of the angular distances were obtained from experimental results, as values that separate the two distributions of correctly and incorrectly assigned particles such as in Fig. 3.5. These distributions become apparent once we plot the distances between all possible b jet combinations and the separation value is then selected.

3.3.5 Architecture and hyper-parameters

The inspiration for the architecture stems from paper on jet-parton assignment [36]. Changes were made to the output and the exact architecture and hyper-parameters were adjusted for our task.

The NN consists of multiple fully connected layers with ReLU activation functions. Dropout layers and data augmentation are used to reduce overfitting on the training data. Furthermore, L2 regularization was used on all weights [27, p.182]:

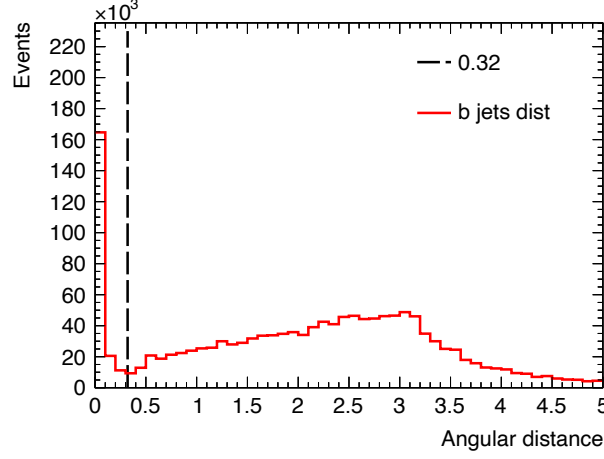


Figure 3.5: Delta R separation of correctly and incorrectly paired b jets
Data used was a combination of $t\bar{t}H$ and $t\bar{t}Z$. All possible pairs were made for each event and the angular distance was calculated. By plotting the distances in a histogram a value is chosen to separate the two distributions that become apparent.

$$\mathcal{L}' = \mathcal{L} + \lambda \cdot \sum_{i=0}^d w_i^2, \quad (3.5)$$

where $\mathcal{L}$ is the old loss function (in this case the binary crossentropy) $\mathcal{L}'$ is the loss function with regularization, λ is the regularization parameter and $\sum w^2$ is the sum of the weights squared.

Skip connections are used to accelerate the learning process [36]. A diagram of the NN architecture is shown in Figure 3.6.

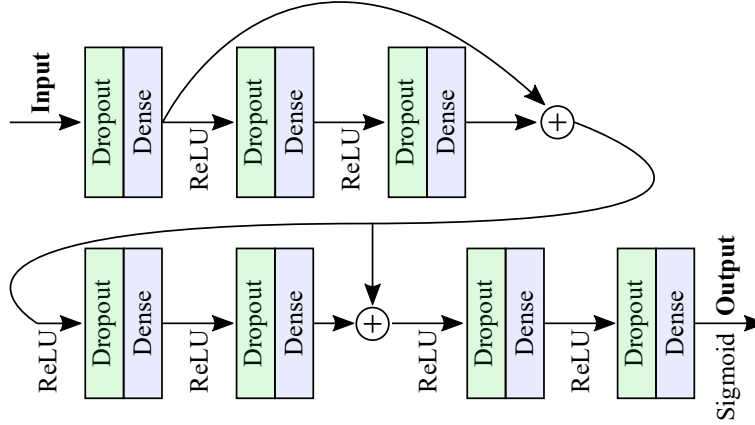


Figure 3.6: The particle assignment neural network diagram

The exact specifications of the NN are in Table 3.4

Learning rate step decay was used to make the training process smoother [27, p.136]:

Table 3.4: Specifications of the particle assignment CNN

Number of inputs	71
Number of outputs	5
Learning rate	0.0003
Optimizer	Adam
Dropout rate	0.2
Dense layer neurons	500
Hidden layers activations	ReLU
Output layer activation	Sigmoid

$$\alpha_t = \alpha_0 \cdot 0.99^t, \quad (3.6)$$

where t is the number of the epoch and α_0 is the initial learning rate.

The loss function is the binary cross entropy (Eq. 1.22). Since we are outputting five values we calculate it for each output and the loss is then their weighted mean. The mean has to be weighted, because the five outputs each have a different ratio between the total number of zeros and ones in their respective labels across all samples (imbalanced classification). For the leptons, where in the vast majority³ of cases there is one correct and one incorrect permutation the classification can be considered balanced. For the jets, on the other hand, there are usually multiple contenders but only one of them is correct, this leads to more zero labels than ones.

This unbalance of the label classes (zero corresponding to incorrect assignment and one corresponding to correct assignment) has to be offset, for which weights in the binary cross entropy are used. These are calculated from the exact ratio of the classes for each output separately (Table 3.5):

$$\begin{aligned}
b_p^0 &= \sum_{k=1}^n a_{p_k}, \\
b_p^1 &= \sum_{k=1}^n 1 - a_{p_k}, \\
w_p^0 &= \frac{2b_p^0}{b_p^0 + b_p^1}, \\
w_p^1 &= \frac{2b_p^1}{b_p^0 + b_p^1},
\end{aligned} \quad (3.7)$$

where n is the number of events, a_{p_k} is the label of position p of k -th event, b_p^0 is a helper variable equal to the number of labels with 0 and w_p^0 is the weight for class 0 for the position p being assigned to. The equations are designed for the sum of the two class labels to be two:

$$w_p^0 + w_p^1 = 2, \quad \forall p. \quad (3.8)$$

³Not always because of the detector effects which can (very rarely) cause for example one of the leptons to be detected poorly.

Table 3.5: Weights used in the loss function of the particle assignment CNN

Particle	Class 0	Class 1
Top b jet	0.28	1.72
Anti-top b jet	0.29	1.71
Top W	0.07	1.93
Tau lepton	0.98	1.02
Anti-top lepton	0.99	1.01

3.3.6 Particle assignment code

The Jupyter notebook for the training of the particle assignment NN is at `/source_code/particle_assignment/particle_assignment_training.ipynb`. The code trains on the data of the Narrow selection (Fig. 3.2) extracted from the ROOT ntuples.

To choose the best permutation for each event and thus process the data for the mass reconstruction the Jupyter notebook at `/source_code/particle_assignment/particle_assignment_training.ipynb` can be used.

A trained NN is also included in directory `/source_code/trained_NN_models`.

3.4 Mass reconstruction neural network

As has been stated in Sec. 3.1, for the mass reconstruction a regression neural network will be used. The data on which the NN will be trained and tested will be the three selections of Fig. 3.2, with the data processed by the particle assignment NN. The particle assignment is only effective for the Narrow selection from the figure. For the other two wider selections from the figure, we will use the same particle assignment NN, although it will be mostly ineffective, as the decay channels of the events added by these wider selections are different. In the approach we have chosen, we do not have a better method to assign the particles of these added events.

The MMC (Sec. 1.3.2) will serve as a comparison for the NN. An MMC library from *Athena* (described as "the *ATLAS Experiment's main offline software repository*" [39]) has been provided by the supervisor. The library has been used in our script with the *2015 calibration set* and the data used with the MMC was the same as for the NN, which means, it was also processed by our particle assignment NN (the MMC requires only the assignment of leptons).

The C++ code for our implementation of a script that uses the MMC library is in the directory `/source_code/MMC`, where are also the Jupyter notebooks for evaluating the reconstructed mass data of the MMC.

3.4.1 Mass reconstruction goal

The goal of the mass reconstruction is to predict the mass of the desired particle. For the signal ($t\bar{t}H$) events this particle is the Higgs boson and for the background events this particle is the Z boson for $t\bar{t}Z$, the W boson for $t\bar{t}W$ and for the $t\bar{t}$ we will ideally predict a zero mass, as there is no particle to be reconstructed⁴.

The values to which we will be comparing the predicted masses will be calculated from the truth information that is present in the simulated dataset. The masses calculated this way are not exactly equal to the known constant masses but have a negligible variance, which is a property of the dataset we are working with. Another almost equivalent approach would be to take the constant invariant masses from Table 1.1.

3.4.2 Loss function

We propose multiple possible approaches each requiring a different loss function to be used with the NN. When choosing an approach (and its loss function) we have to consider multiple things.

We expect the NN to distinguish between the signal and the background and for the reconstructed masses to be close to their actual values.

There is also the question of what background signals do we want the NN to be able to process. The MMC only works well with the $t\bar{t}H$ and $t\bar{t}Z$ productions, but in the real ATLAS data there are also other productions such as the $t\bar{t}W$ and $t\bar{t}$.

The loss functions will use different labels, but each will be able to produce the reconstructed mass from its output.

MMC inspired loss

Inspired by the MMC, a loss function incorporating equations of the invariant mass of reconstructed particles and the MET. The NN outputs the predicted four-vectors of the four neutrinos. Then the masses of the neutrinos and also of particles, reconstructed by adding together the neutrinos and the visible particles (e.g. the neutrino and the lepton coming from the anti-top branch when added together make the anti-top W boson), are calculated by the formula Eq. 1.5. With four neutrinos, two W bosons, two taus, one Higgs and one top this gives us ten predicted masses.

The predicted MET is calculated from the addition of all four neutrinos and application of the formula Eq. 1.7⁵.

To calculate the loss itself, we subtract each of the predicted masses from the known constant of the given particle (e.g. $\hat{m}_W - 80.38$), square it and

⁴We could reconstruct the leptonically decaying top quark, but that would be inconsistent with the other productions, where we also could reconstruct the top quark, but we do not, as we reconstruct a different particle.

⁵Just without the negation, since in the formula we are summing the visible particles and here we are summing the undetected ones.

multiply it by a weight. We do the same with the MET. We then sum all of these terms to obtain the loss.

The described loss can be written as

$$\mathcal{L}_1 = \sum_{k=1}^{10} \alpha_k (m_k - \hat{m}_k)^2 + \alpha_{11} (\mathcal{E}_{T_X} - \hat{\mathcal{E}}_{T_X})^2 + \alpha_{11} (\mathcal{E}_{T_Y} - \hat{\mathcal{E}}_{T_Y})^2. \quad (3.9)$$

The weights are chosen in a way, that each part of the loss is approximately equal after the training has gone through some initial epochs:

$$\alpha_1 (m_1 - \hat{m}_1)^2 \approx \alpha_2 (m_2 - \hat{m}_2)^2 \approx \dots \quad (3.10)$$

The optimization of the weights has not been necessary, as even large perturbations, such as multiplying some weights by four and dividing others by four had negligible effect on the results of the trained network.

We obtain the predicted Higgs boson four-vector by adding the neutrinos coming from the Higgs boson branch to the visible part of the Higgs decay (a hadronically decaying tau and a lepton coming from the leptonically decaying tau):

$$\hat{H} = \hat{\nu}_{\tau_{had}} + \hat{\nu}_{\tau_{lep}} + \hat{\nu}_\ell + \tau_{had} + \ell. \quad (3.11)$$

The mass is calculated from the predicted Higgs boson four-vector with the use of Eq. [1.5](#).

The applicability of this loss function is even stricter than of the MMC as it requires precisely the Narrow selection in Fig. [3.2](#) (MMC does not require the specific $t\bar{t}$ decay).

■ Four-vector MSE loss

The output is the four-vector of the reconstructed particle (Higgs boson in $t\bar{t}H$, Z boson in $t\bar{t}Z$, W boson in $t\bar{t}W$, four-vector of zeros in $t\bar{t}$). From a good prediction of the four-vector, we would ideally obtain a good prediction of the mass. The used loss is MSE:

$$\mathcal{L}_2 = \frac{1}{4} [(p_x - \hat{p}_x)^2 + (p_y - \hat{p}_y)^2 + (p_z - \hat{p}_z)^2 + (E - \hat{E})^2]. \quad (3.12)$$

The advantage of this loss function compared to the MMC is the option to use it on $t\bar{t}H$ and $t\bar{t}Z$ as well as $t\bar{t}W$ and $t\bar{t}$. This makes it useful even for the real ATLAS data, where these background productions occur.

■ Single output mass loss

The neural network outputs a single value equivalent to the reconstructed particle's mass. The difference between the predicted and truth value is then squared, which serves as the loss function:

$$\mathcal{L}_3 = (m_0 - \hat{m}_0)^2. \quad (3.13)$$

Also this loss function can be used on $t\bar{t}H$ and $t\bar{t}Z$ as well as $t\bar{t}W$ and $t\bar{t}$.

This loss is similar to what Petr Urban has used in his thesis [19]. His work is mentioned in the points one and two in the assignment of my thesis. His NN code suffers from the issue that it has been trained and tested on $t\bar{t}H$ only. When training, the NN only encountered labels with values around the Higgs mass constant and when used on anything besides the $t\bar{t}H$, it would predict the same values no matter what would be on input (bias for the constant Higgs boson invariant mass). By including other productions in the training data we avoid this issue.

3.4.3 Features

The input features will stay the same as in Sec. 3.3.3 with the addition of the output of the particle assignment NN added as five additional features for each event. The five additional features represent the probabilities of correct assignment and we use them to give the mass reconstruction NN the information of how well are the particles (which four-vectors are also on the input) assigned.

3.4.4 Architecture and hyper-parameters

The neural network will consist of a number of fully connected layers. The architecture optimization was done on the Narrow selection ((a) in Fig. 3.2) and with the last mentioned loss function (mass loss). This selection and loss function were selected, as they were considered to be the most important, ideally we would adjust the architecture for each loss and each selection.

Experiments with the specifics of the architecture were conducted and some of the results are in the tables A.1 through A.3 in Appendix A. The tables serve as more of an illustration of the differences between the architectures being insignificant, than as a rigid optimization process, as each record in the tables comes from only one iteration of the NN being trained and evaluated. The chosen architecture is schematized on Fig. 3.7 with parameters per Table 3.6.

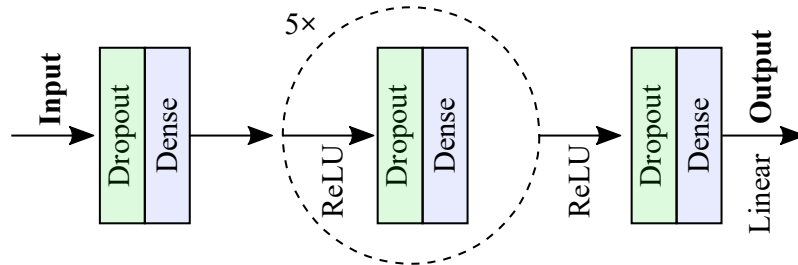


Figure 3.7: The mass reconstruction neural network diagram

The NN contains an input layer, five sequential fully-connected hidden layers and an output layer. Each fully-connected layer also includes a dropout layer.

Table 3.6: Specifications of the mass reconstruction NN

Number of inputs	76
Initial learning rate	0.0005
Batch size	4096
Optimizer	Adam
Dropout rate	0.2
L2 regularization	0.001
Dense layer neurons	720
Hidden layers activations	ReLU
Output layer activation	Linear

As with the particle assignment NN (Sec. 3.3.5), the same learning rate step decay was used.

To reduce overfitting the same data augmentation (Sec. 3.2) was used as with the particle assignment. The effect of data augmentation on mass reconstruction is demonstrated in Table A.4 and Figure A.1 in Appendix A. L2 regularization was used on all weights to prevent overfitting on the training data.

3.4.5 Mass reconstruction code

The Jupyter notebooks for the mass reconstruction NN, each containing the NN with one of the three loss functions can be found in the directory `/source_code/mass_reconstruction`. Each notebook contains both the training and evaluation process.

Trained NN models are in the directory `/source_code/trained_NN_models`.

Chapter 4

Results

4.1 Particle assignment

A visualisation of the performance of the particle assignment NN on a test dataset is in Figure 4.1. The data in the test data set had the same properties as the data the NN was trained on, it was of the Narrow selection from Fig. 3.2 and each event was represented by all possible permutations of jets and leptons.

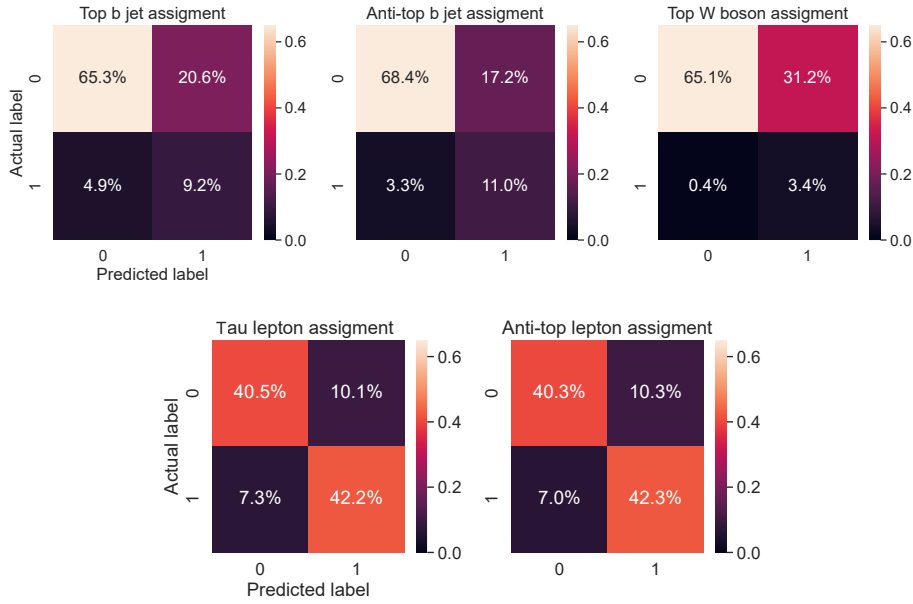


Figure 4.1: Particle assignment NN confusion matrices

Confusion matrices of the assignment of particles of the NN on a test dataset. Each event from the dataset is represented by all possible jets and leptons permutations.

The main metric measured in the particle assignment is the Matthews correlation (MCC) (Eq. 4.1), which is a suitable metric for imbalanced classification problems [40].

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}, \quad (4.1)$$

where TP stands for the number of true positives, TN for true negatives, FP for false positives and FN for false negatives. The MCC expected value is 0 for random classification and 1 for perfect classification [40]. The MCC obtained on the test dataset is in Table 4.1.

Table 4.1: Particle assignment Matthews correlation

Assigned particle	Matthews correlation
Top b jet	0.314
Anti-top b jet	0.442
Top W	0.232
Tau lepton	0.654
Anti-top lepton	0.655

From the confusion matrices and the MCC we see, that the assignment of the leptons is more accurate than the assignment of the jets. This can be attributed to the fact, that leptons are detected better by the detector than the quarks which evolve into jets before being detected and also are less likely to be misidentified. Jets can also come from gluons, which further increases the uncertainty. Also, with the leptons we only choose one permutation out of two, while with the jets there are usually many more permutations to choose from.

Another point to notice is that while the assignment of the two leptons has very similar performance, the jet assignment performances differ between one another. The anti-top b jet seems to be the easiest of the three to assign, which can be explained by it being in the same branch of the decay as one of the leptons, while the top b jet and the top W boson are both coming from the same parent particle (Fig. 1.4). Since the lepton is assigned correctly in a large number of cases, it is then easier for the corresponding b jet to be chosen.

The top W, which is in reality the two non-b jets which the W boson decays into, is the most difficult to assign, because if one of the two jets is poorly detected or misidentified, which is two times more likely than for the single b jet, the top W is then more difficult to reconstruct.

What the confusion matrices and MCC do not show, is how well we are able to choose the best possible assignment out of all the permutations for each event. Out of all the permutations we choose the one with the largest product of the five output values:

$$\arg \max_i \prod_p \hat{a}_p^i, \quad (4.2)$$

where i symbolizes the current permutation, $p \in \{b1, b2, W, l1, l2\}$ is the position being assigned to and $\hat{a}$ is the output of the neural network — that is the predicted probability of the particle being correctly assigned.

We can then look at the rate with which the chosen assignment has the respective particles assigned correctly.

It needs to be taken into account, that it is not always the case, that the best assignment has each particle assigned correctly (e.g. when a detected jet has been misidentified). To account for this, we calculate the mean of correct assignment of each position (MCA; symbol A) calculated as

$$A_p = \frac{1}{n} \sum_{k=1}^n a_p^k, \quad (4.3)$$

where n is the number of events and a^k is one if assigned correctly and zero otherwise (as in Eq. 3.3).

In Table 4.2 is the MCA per particle in the chosen permutation and in the best possible permutation (i.e. the one with the most correctly assigned particles) for $t\bar{t}H$ and $t\bar{t}Z$.

Table 4.2: Particle assignment mean of correct assignment
Separately for the chosen assignment and the best possible assignment in each event. Separated into $t\bar{t}H$ and $t\bar{t}Z$.

Assigned particle	MCA chosen		MCA best possible	
	$t\bar{t}H$	$t\bar{t}Z$	$t\bar{t}H$	$t\bar{t}Z$
Top b jet	0.43	0.42	0.84	0.84
Anti-top b jet	0.57	0.53	0.85	0.86
Top W	0.20	0.18	0.50	0.48
Tau lepton	0.80	0.85	0.99	0.98
Anti-top lepton	0.80	0.86	0.99	0.99

Additionally, by dividing the value of the chosen and best possible MCA,

$$A_p^r = \frac{A_p^c}{A_p^b}, \quad (4.4)$$

we can obtain the relative MCA (Table 4.3).

Table 4.3: Particle assignment relative mean of correct assignment

Assigned particle	relative MCA	
	$t\bar{t}H$	$t\bar{t}Z$
Top b jet	0.52	0.50
Anti-top b jet	0.67	0.62
Top W	0.40	0.38
Tau lepton	0.81	0.87
Anti-top lepton	0.81	0.87

The above tables confirm that the leptons are assigned more accurately and that the anti-top b jet is easier to assign than the other jets. We can

now also compare and see that the $t\bar{t}H$ has slightly worse lepton assignment than the $t\bar{t}Z$, which does not seem to have an obvious explanation, since the decay channels are the same. The only obvious difference between the two production is, that the Z boson is less massive than the Higgs boson, which could have an positive effect on the lepton assignment, as one of the leptons is coming from the Higgs/Z boson branch.

A comparison was made between the above results and the paper [36], which inspired the NN architecture and the use of permutations for the assignment task. A direct comparison is not possible, because the two differ in some very important aspects. Some of the major differences from the side of the paper are the $H \rightarrow b\bar{b}$ decay channel being analysed, not including any other production besides $t\bar{t}H$, having 700 000 events, the events selection being stricter and six jet positions being assigned to.

The cited paper states, that in 52% of cases it reconstructed the event perfectly [36]. We can calculate a similar metric, where we compare the chosen assignment to the best possible assignment and see if they have the same number of correctly assigned particles:

$$r_k = \begin{cases} 1, & \sum_p a_p^{c_k} \geq \sum_p a_p^{b_k}, \\ 0, & \sum_p a_p^{c_k} < \sum_p a_p^{b_k}, \end{cases} \quad (4.5)$$

where r_k marks, if for the k -th event we have chosen the best possible assignment, and $\sum_p$ is sum of the label of either the chosen ($a_p^{c_k}$) or the best possible ($a_p^{b_k}$) assignment for event k . Then we calculate the mean across all events and finally get the percentage of events in which the perfect assignment has been chosen as:

$$R = 100\% \cdot \frac{1}{n} \sum_{k=1}^n r_k. \quad (4.6)$$

On the test set this metric gives us:

Table 4.4: Percentage of perfectly assigned events

Perfectly assigned event is an event in which a permutation that was chosen has the maximum possible number of correctly assigned particles.

Perfectly assigned events	
$t\bar{t}H$	$t\bar{t}Z$
32%	31%

Compared to the paper's 52%, our performance is worse, but as has been said a direct comparison does not make sense, since there are too many differences, it is rather to give a different perspective on the achieved performance. Some major differences have been already mentioned above, some of them cannot be dealt with (different decay channel, different productions, different particles being assigned), but what could be improved is the number of events

the neural network is trained on, by obtaining a larger dataset, which could boost the performance.

4.2 Mass reconstruction

In Section 3.1 we have proposed three data selections and in Section 3.4.2 we have proposed three loss functions. We will first compare the loss functions on the Narrow selection from Fig. 3.2, that is the selection comprising of only $t\bar{t}H$ and $t\bar{t}Z$ with the specified decay channel. In the comparison we will also include the MMC used on the same test dataset as the NNs. It is important to note, that using the same dataset implies, that not only the NN but also the MMC will be affected by the particle assignment, by which all the data was processed.

In some of the events, the MMC does not produce a result as it does not find a solution for the equations 1.13. It might be in part because of the broken assumption of no neutrinos outside of the Higgs that does not apply to our channel, as has been discussed in 1.3.2. Even though they are not shown in Fig. 4.2, the events which lead to the zero mass prediction are accounted for in Table 4.6. The exact numbers for the unsolved events are in Table 4.5, which includes not only the $t\bar{t}H$ and $t\bar{t}Z$ but also the other productions, which will be relevant later on.

Table 4.5: Percentage of unsolved events by the MMC

The first two rows in the table correspond to the currently discussed Narrow data selection ((a) in Fig. 3.2), narrow meaning the selection of the exact structure of the decay channel of the productions made using the truth information. The remaining rows correspond to the selections without the truth-based requirement on the Higgs/Z or the top pair decay from the same figure.

Dataset	Unsolved events
$t\bar{t}H$ narrow	14.1%
$t\bar{t}Z$ narrow	17.6%
$t\bar{t}W$ all	30.5%
$t\bar{t}$ all	18.2%
$t\bar{t}H$ all	23.2%
$t\bar{t}Z$ all	19.9%

The higher numbers in Table 4.5 for the $t\bar{t}W$ and $t\bar{t}$ can be attributed to the fact, that the MMC is not suited to reconstruct mass for these productions, as has been discussed in 1.3.2. For the $t\bar{t}H$ and $t\bar{t}Z$ the particle assignment can play a role as well. As we can see, the events of the *narrow* selection, which the particle assignment has been developed on, are solved more often than the ones of the *all* selection, which includes many events with poorly assigned particles.

4.2.1 Metrics

The main metric on which we compare the different approaches is the weighted maximum separation of the signal from the background calculated as:

$$S = \max_d \left\{ \frac{|z| + |h|}{2 \cdot |h|} \cdot |h_c(d)| + \frac{|z| + |h|}{2 \cdot |z|} \cdot |z_c(d)| \right\}, \quad (4.7)$$

where h and z are the test sets of Higgs (or Z) events, d is a value between the minimum and maximum of all predicted masses, which serves as a separator between predicted signal and background, where we assume all predicted masses larger than the separator to be from signal events and others to be from background events:

$$\text{event } e \text{ is } \begin{cases} \text{signal,} & \hat{m}(e) > d, \\ \text{background,} & \hat{m}(e) \leq d, \end{cases} \quad (4.8)$$

and $h_c(d)$ and $z_c(d)$ are the sets of events correctly assigned from the original sets by Eq. 4.8 for the given separator value d . The separation is calculated for one thousand evenly distributed separator values. The values in Eq. 4.7 are set in a way so that for both extreme separations¹ $|h_c| = |h| \wedge |z_c| = 0$ and $|h_c| = 0 \wedge |z_c| = |z|$ the separation is 50%.

For the separation, the larger the value, the better, and 50% is the baseline of random separation.

Other metrics are the mean μ and standard deviation σ of the Gaussian function (Eq. 4.9, 41) fit on the histogram of the predicted masses.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (4.9)$$

We also calculate the scaled σ' of the Gaussian fit scaled to the Higgs boson mass constant 125.18 as:

$$\sigma' = \sigma \cdot \frac{125.18}{\mu}. \quad (4.10)$$

For the Higgs boson mass mean, the value should be as close to the constant 125.18 GeV as possible and have a small σ .

The separation is considered to be the most important metric as it is not trivial to achieve. For the mean and the σ of the mass distribution, we could always predict the constant 125.18 GeV for each event and we would achieve a perfect mean and σ , but no separation.

The results that will be presented, will have the original predicted mass presented, but the mass distributions could also be scaled to have the mean equal to 125.18 GeV, while the separation would stay the same, but the σ would change. To represent such scaling we introduced the scaled σ' (Eq. 4.10).

¹Predicting all events to be signal or all events to be background.

4.2.2 Mass reconstruction results

From Figure 4.2 and Table 4.6 we can analyze the performance of the different loss functions.

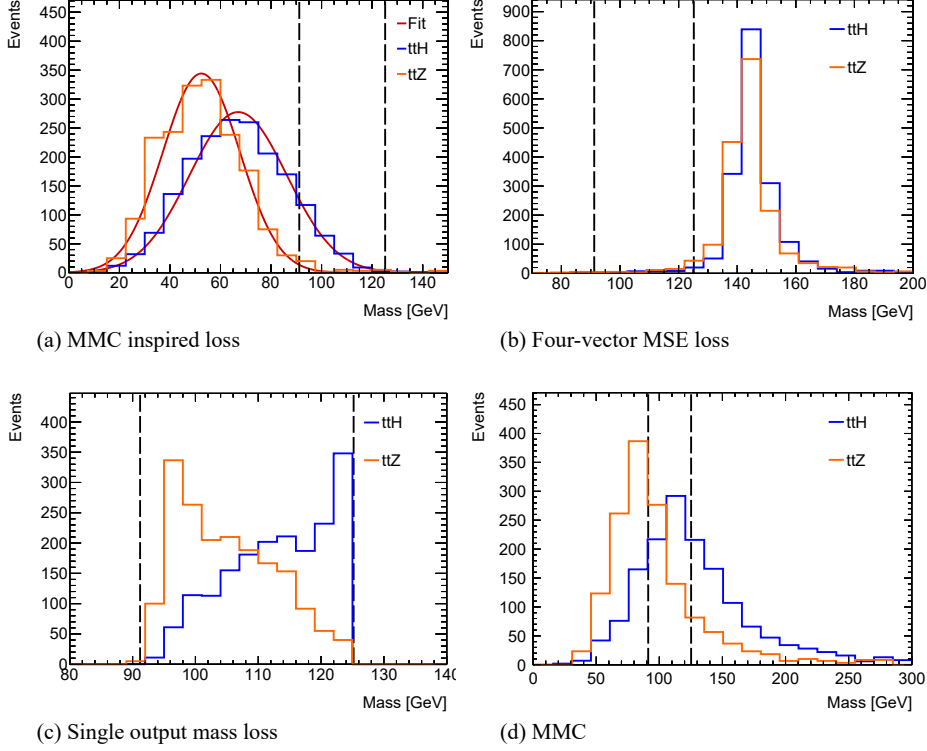


Figure 4.2: Mass histograms — loss functions and MMC

Mass reconstructed by neural networks with different loss functions in (a), (b) and (c). In (d) is the reconstructed mass from the MMC for comparison with unsolved events omitted from the plot. Data of Narrow selection from Fig. 3.2. Number of events scaled by weighting $t\bar{t}Z$ to match the $t\bar{t}H$, which had a larger dataset. Gaussian function fit in (a) for illustration of mean and STD measurement — left out of (c), (b) and (d) for better readability. Dashed lines mark the Higgs invariant mass 125.18 GeV and Z invariant mass 91.19 GeV.

The poor performance of the four-vector MSE loss in (b) of the figure can be accounted to the loss being focused on the values of the four-vector (momentum and energy) and not directly taking into account the mass, which is the goal value. The thought behind the loss — which was that if we predict the four-vector we can calculate the mass — would require for the four-vector prediction to perform better.

The MMC inspired loss in (a) performed slightly worse than the MMC in terms of separation. The idea of having the NN learn to conserve masses and MET equations translated to the NN quite well and the distributions have a similar shape to the MMC in (d). However, the mean of the distribution is almost two times smaller and even changing the weight of the Higgs (or Z) mass in 3.9 had no effect on this, as the NN always had about the same

Table 4.6: Mass reconstruction metrics — loss functions

Mean μ and STD σ of Gaussian fit of the mass distribution (as in Fig. 4.2). Highlighted in bold the best performing of the given metric. For Mass loss we also note a $\mu = 113.14$ GeV, $\sigma = 8.11$ GeV and scaled $\sigma' = 8.97$ of the distribution itself (not the Gaussian fit), as the Gaussian does not fit the distribution well.

Metric	MMC inspired loss	Four-vector MSE loss	Mass loss	MMC
Separation	65.84%	56.25%	69.33%	67.63%
Higgs mass μ (GeV)	67.00	144.85	120.77	118.98
Higgs mass σ (GeV)	19.28	5.77	12.95	33.55
Higgs mass σ' scaled (GeV)	36.02	4.99	13.42	35.30

performance.

The best separation is achieved by the Single output mass loss in (c) even outperforming the MMC. With the loss fully focused on the goal variable, the NN achieves the best results. The mass distributions are between the Z (91.19 GeV) and Higgs (125.18 GeV) boson constant mass and the NN never predicts outside of the range, as it has not been trained on any values outside of the mentioned range. This additional information (i.e. that the masses lie in this range) that is given to the NN lets it achieve a very good mean and width of the Higgs mass distribution.

The peaks of the MMC mass distributions in (d) being shifted slightly towards lower values than the mass constants, can be explained by the missing energy in the form of an outside² neutrino coming from the anti-top branch, which, as has been discussed in Sec. 1.3.2, is not accounted for by the MMC.

4.2.3 Additional background

The Additional backgrounds selection as illustrated in (b) of Fig. 3.2 expands the previous selection to include the $t\bar{t}W$ and $t\bar{t}$ background productions.

The separation metric will still be used as defined by Eq. 4.7, but with all three backgrounds instead of just the $t\bar{t}Z$.

On this broader dataset we will train and test our NN with the best performing loss function — the Single output mass loss. To have a comparison for the NN, we will once again use the MMC, although as has been discussed in Sec. 1.3.2, the MMC is not suited to deal with these added backgrounds, so we will not make any conclusions in regards to the following comparisons.

Figure 4.3 shows the distributions of the predicted masses for the different productions. The histograms of the background had to be scaled to match the $t\bar{t}H$, which had the most test events.

²Outside of the Higgs/Z branch of decay.

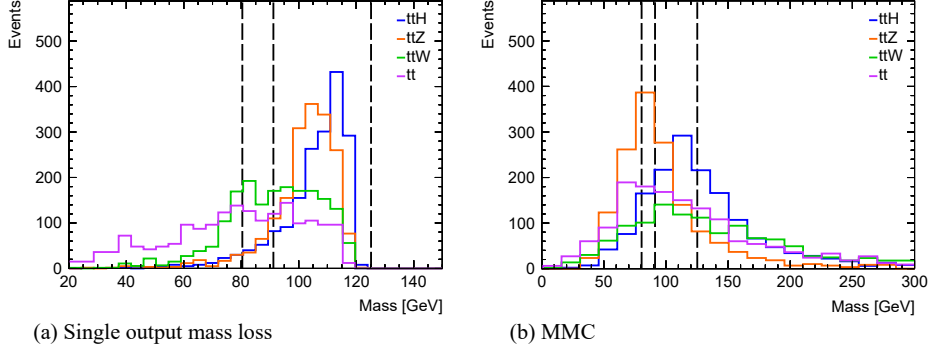


Figure 4.3: Mass histograms — additional background

Histograms of reconstructed mass by the NN with the Single output mass loss function in (a) and by the MMC in (b). Data of the Additional backgrounds selection from Fig. 3.2. Number of events scaled by weighting background productions to match the signal, which had the largest test dataset. Dashed lines mark the Higgs invariant mass 125.18 GeV, Z invariant mass 91.19 GeV and W invariant mass 80.38. For $t\bar{t}$ a 0.00 GeV mass is assumed as the goal value.

From the figure it is visible, that while the $t\bar{t}H$ and $t\bar{t}Z$ data stayed the same as in the previous data selection in Fig. 4.2, the separation has worsened, as the NN had to account for the new backgrounds as well. The $t\bar{t}W$ and $t\bar{t}$ both have a similar wide distribution, with W having a sharper peak at its corresponding constant mass. The Higgs boson separation from these two backgrounds is much better than from the $t\bar{t}Z$, which can be explained by the Higgs and Z boson productions being very similar, while the other two differ more from the Higgs boson.

The MMC performs exactly the same on the $t\bar{t}H$ and $t\bar{t}Z$, as in Fig. 4.2, because it does not train on the data as the NN does, instead it is deterministic in its calculations. For the $t\bar{t}W$ and $t\bar{t}$ it performs poorly, with a wide distribution with similar peaks to the Higgs, as was expected.

Values supporting the above figure are in Table 4.7

Table 4.7: Mass reconstruction metrics — additional background

Mean μ and STD σ of Gaussian fit of the mass distribution (in Fig. 4.3). For Mass loss we also note a $\mu = 105.09$ GeV, $\sigma = 12.68$ GeV and scaled $\sigma' = 15.10$ GeV of the distribution itself (not the Gaussian fit), as the Gaussian does not fit the distribution well.

Metric	NN (Mass loss)	MMC
Separation	67.77%	61.59%
Higgs mass μ (GeV)	107.78	118.98
Higgs mass σ (GeV)	5.71	33.55
Higgs mass σ' (GeV)	6.63	35.30

The last detector simulation dataset selection — the Real data selection in Fig. 3.2 — includes all data tagged with the $2lSS + 1\tau_{had}$ decay channel

variable and with at least three detected jets and one b jet. Compared to the previous selection above, we no longer require the exact decay of the $t\bar{t}H$ and $t\bar{t}Z$ and as such, this selection is done without the use of truth information. These additional $t\bar{t}H$ and $t\bar{t}Z$ events go through the particle assignment NN first, but as their decay structures differ from the one the NN has been developed on, the particle assignment is not effective on these events.

Such events are then more poorly reconstructed by both the NN and the MMC and further worsen the achieved separation as can be seen in Figure 4.4 and Table 4.8.

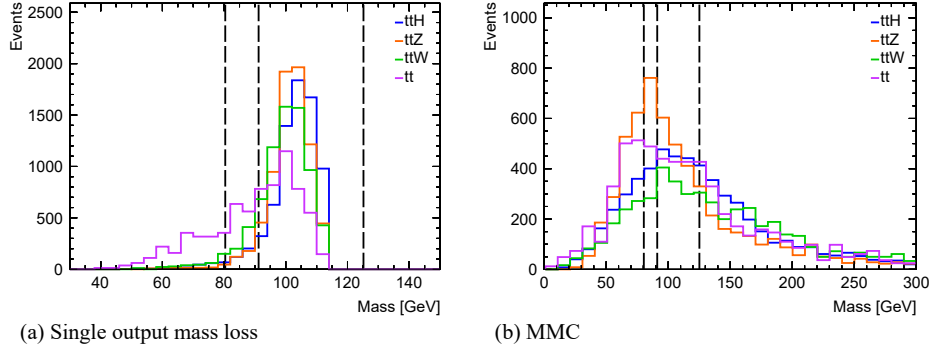


Figure 4.4: Mass histograms — all data

Histograms of reconstructed mass by the NN with the single output mass loss function in (a) and by the MMC in (b). Data of selection (c) from Fig. 3.2, which simulates the real ATLAS data. Number of events scaled by weighting background productions to match the signal, which had the largest test dataset. Dashed lines mark the Higgs invariant mass 125.18 GeV, Z invariant mass 91.19 GeV and W invariant mass 80.38.

Table 4.8: Mass reconstruction metrics — additional background

Mean μ and STD σ of Gaussian fit of the mass distribution (as in Fig. 4.4). For Mass loss we also note a $\mu = 102.25$ GeV, $\sigma = 8.19$ GeV and scaled $\sigma' = 10.03$ GeV of the distribution itself, although the Gaussian function fits quite well.

Metric	NN (Mass loss)	MMC
Separation	59.21%	53.39%
Higgs mass μ (GeV)	105.08	116.44
Higgs mass σ (GeV)	7.41	37.71
Higgs mass σ' (GeV)	8.83	40.54

The NN is still able to achieve some separation, but the Higgs and Z further overlap and the $t\bar{t}W$ and $t\bar{t}$ peaks also shift to higher values and overlap with the Higgs more than previously. Still some separation is achieved with the 59.21%.

The MMC also achieves poorer reconstruction, than on the previous selection, of both the $t\bar{t}H$ and $t\bar{t}Z$ with wider and more over-lapped distributions.

As this data selection is equal to the data in the real ATLAS data, we also scale the mass distribution by the expected number of events from Table 2.3. Mass distribution scaled to the expected number of events is shown in Figure 4.5.

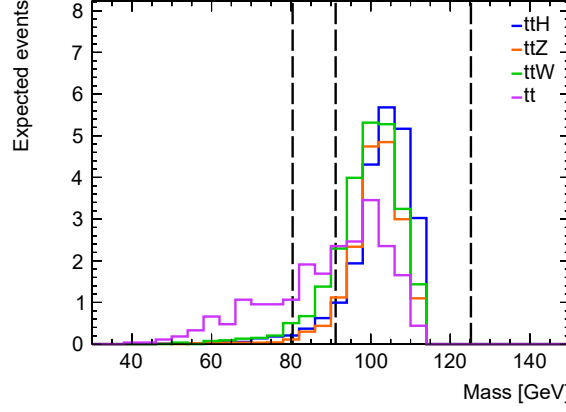


Figure 4.5: Mass histograms — all data scaled
Mass reconstructed by the NN with Mass loss. Data of selection (c) from Fig. 3.2, which simulates the real ATLAS data. Number of events scaled to be equal for each production to the number of expected events as per Table 2.3.

4.2.4 Future improvements

The performance of the NN could be further improved by a more careful feature selection, as the features we have selected were based on the thesis of Petr Urban [19] and the inputs of the MMC.

The architecture and choice of the parameter values of the NN also has some space for improvement as it has not been thoroughly optimized (e.g. by grid search [27, p.125]).

4.3 Real ATLAS data

The neural networks trained on the simulated data will now be used on the real ATLAS data.

For a test on the real data, ideally, we would use the NN trained on the Real data selection from Fig. 3.2, which uses no truth information and therefore a neural network trained on such selection is suited to be used on the real ATLAS data, but there are some differences in what variables are in the simulated and the provided real ATLAS datasets.

In the simulated dataset we had the p_T , η , ϕ and E information of up to eight detected jets.

The provided real ATLAS dataset has the information of the p_T and η of the three most energetic jets detected.

We have therefore decided to study the particle assignment and mass reconstruction of the real ATLAS data without jets and the inclusion remains to be done.

From the features listed in Sec. 3.3.3 we will only use the ones independent of the jets.

4.3.1 Particle assignment

We retrain the particle assignment NN to accommodate for the new reduced features.

The particle assignment will only assign the two detected leptons in one of two ways. It is again trained and evaluated on the simulated dataset, specifically on the Narrow selection from Fig. 3.2, but with only the reduced features.

The results of the evaluation of the trained NN on permutations generated from the simulated test dataset are shown in Figure 4.6 in the form of confusion matrices.

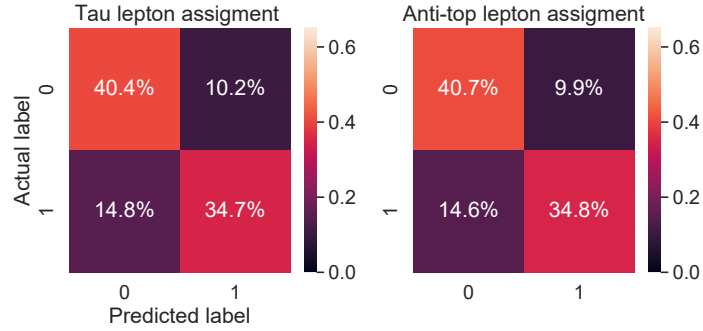


Figure 4.6: Leptons assignment with reduced features

Confusion matrices of the assignment of the leptons of the NN on a simulated test dataset. Results of assignment over all permutations of all events in the dataset. Reduced features without jet information used.

The Matthews correlation is listed in Table 4.9.

Table 4.9: Lepton assignment Matthews correlation with reduced features

Assigned particle	Matthews correlation
Tau lepton	0.502
Anti-top lepton	0.512

If we compare the results of the particle assignment that are in Figure 4.6 and Table 4.9 with reduced variables with the particle assignment with all the variables (Sec. 4.1), we see that the performance has dropped, as the Matthews correlation went from 0.654 and 0.655 to 0.502 and 0.512 and the percentage of correctly assigned leptons went from 42.2% to 34.7% as per the

confusion matrices. This was expected because we are training the NN on less information. The drop in performance is not too significant for the real data, as the particle assignment is only effective for the Narrow selection of the events, as has been discussed in Sec. 3.3.2.

4.3.2 Mass reconstruction

The mass reconstruction NN will be the same as presented in previous sections. We will be using the Single output mass loss, which has been evaluated on the Real data selection (Fig. 4.4).

We retrain the mass reconstruction on the simulated dataset with the reduced features and then use the trained NN on the real ATLAS data.

The mass distribution plot with results on the simulated dataset scaled to the expected number of events for each production (similar to the one in Fig. 4.5) is shown in Fig. 4.7.

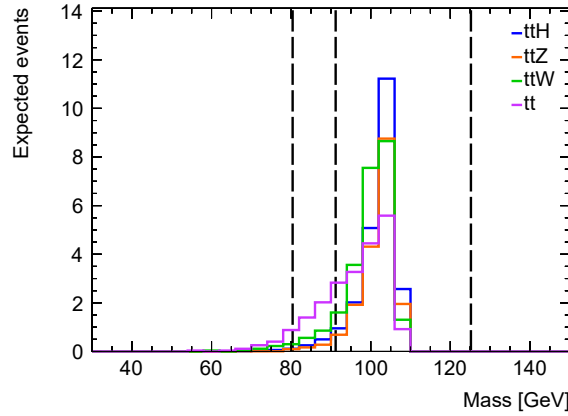


Figure 4.7: Simulated data mass with reduced features

Mass reconstructed by the NN with Mass loss. Data of the Real data selection from Fig. 3.2. Only available features used without jet information. Number of events scaled to be equal for each production to the number of expected events as per Table 2.3.

The metrics derived from the reconstructed Higgs boson distribution produced by the NN with reduced features are in Table 4.10.

A comparison with the results obtained using all the variables (Tab. 4.8) is in place. The separation dropped from 59.21% to 55.65%, which is a significant difference and is caused by the reduced features containing less information. Still there is some separation achieved.

With the trained neural networks with reduced features ready, we move on to their application on the real data.

First the real data has been extracted from the ROOT ntuples and 101 events were obtained compared to the expected 105.9 events (as was stated in Table 2.3). The selection used for the extraction of the events was similar to the one we used with the simulated data which was the requirement of the $2lSS + 1\tau_{had}$ tag to be true in an event and at least three detected jets

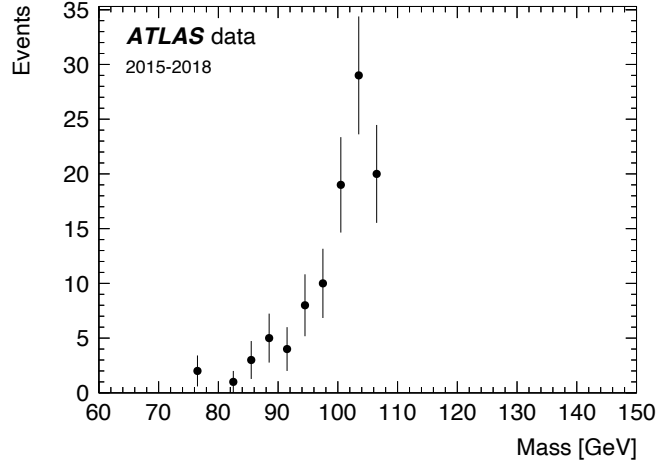
Table 4.10: Mass reconstruction with reduced features

Separation of the signal from the background. Mean μ and STD σ of Gaussian fit of the Higgs boson mass distribution (as in Fig. 4.7). We also note a $\mu = 101.49$ GeV, $\sigma = 5.12$ GeV and scaled $\sigma' = 6.31$ GeV of the distribution without a Gaussian fit.

Metric	NN (Mass loss)
Separation	55.65%
Higgs mass μ (GeV)	102.91
Higgs mass σ (GeV)	3.45
Higgs mass σ' (GeV)	4.20

with at least one of them being a b jet. By using the $2lSS + 1\tau_{had}$ tag and number of jets requirements, we would select 953 events, so such selection would be less strict. The exact selection we used for the real data is in Fig. B.1 in Appendix B.

The 101 events were processed by the trained particle assignment NN and then the mass was reconstructed for each event by the trained mass reconstruction NN. The reconstructed mass distribution is shown in Fig. 4.8.

**Figure 4.8:** Reconstructed mass of real ATLAS data

Mass reconstructed by the NN with Mass loss trained on simulated data. Reconstructed from real ATLAS data with the selection according to Appendix B, which is similar to the selection of $2lSS + 1\tau_{had}$ decay with at least three jets and one b jet. The error bars show the statistical uncertainties. Graphical representation of the distribution different from the previous figures chosen for the real data.

The distribution in Fig 4.8 is similar to the one in Fig. 4.7, which confirms that the events from the simulated dataset are similar to the real ATLAS data.

Lastly, we would like to select three events as the best reconstructed events in which a Higgs boson is produced. We will first reduce the 101 events to 10

events with the best assigned leptons. We do that by choosing the events for which the particle assignment NN returns the highest probabilities of correct assignment (we compute the product of the probabilities similar to Eq. 4.2, but across events not permutations).

The predicted masses of the 10 best assigned events are in Fig. 4.9.

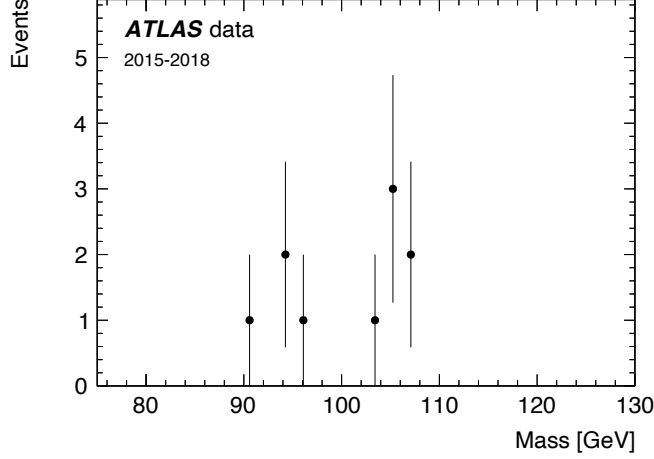


Figure 4.9: Mass of the best assigned real events

The 10 best assigned events were chosen from the 101 selected real events. The mass was reconstructed by the NN with the Mass loss trained on simulated data with reduced features. The error bars show the statistical uncertainties.

The $t\bar{t}H$ achieves the best separation in the highest mass values, as the Higgs boson is the most massive of the reconstructed particles. From the 10 events, we select the 3 with the highest mass as the best reconstructed Higgs boson events. The event numbers of these three events are listed in Table 4.11.

Table 4.11: The best reconstructed real $t\bar{t}H$ events

The event number is an identifier of an event produced in ATLAS. The year of production is the year in which the event was detected in ATLAS. The run number is the identifier of the run of ATLAS.

Event number	Year of production	Run number
777623619	2016	307716
1112036381	2018	351455
1139537422	2018	358577



Conclusion

A method for the assignment of detected leptons and jets in the $2lSS + 1\tau_{had}$ decay of the $t\bar{t}H$ and $t\bar{t}Z$ productions has been developed in the form of a classification neural network trained and tested on the ATLAS detector simulation data. The neural network achieved the rate of 32% perfectly assigned $t\bar{t}H$ events and 31% perfectly assigned $t\bar{t}Z$ events.

On the data processed by the particle assignment neural network, a mass reconstruction neural network has been developed. The best out of the three proposed loss functions has achieved a separation of signal from background of 69.33% on a test set of the aforementioned decay, outperforming the Missing Mass Calculator which achieved a separation of 67.63%.

The NN was further trained and tested on a dataset with the addition of the $t\bar{t}W$ and $t\bar{t}$ background productions. On this data the NN achieved a separation of 67.77%.

The final selection of the simulated data included the data from all the mentioned productions without any cuts based on truth information. This selection was equivalent to the data produced by the real ATLAS detector and the achieved separation of the NN was 59.21%.

The developed methods were applied on the real ATLAS data. We have decided for a study without the jets information. The adjusted networks were used and three best reconstructed Higgs boson events were identified.

In conclusion, the developed neural networks fulfilled the assigned tasks.

Appendix A

Figures and tables

In this appendix are additional figures and tables.

Table A.1: Neural network architecture experiments - width

Width (neurons in layer)	Separation	Higgs mass mean (GeV)
360	70.22	112.22
720	70.74	112.73
1440	70.48	113.19

Table A.2: Neural network architecture experiments - depth

Depth (layers)	Separation	Higgs mass mean (GeV)
3	70.13	112.23
4	70.63	112.57
5	70.29	112.24
7	70.66	113.16

Table A.3: Neural network architecture experiments - skip connections

Skip connections	Separation	Higgs mass mean (GeV)
yes	70.31	112.77
no	70.48	113.27

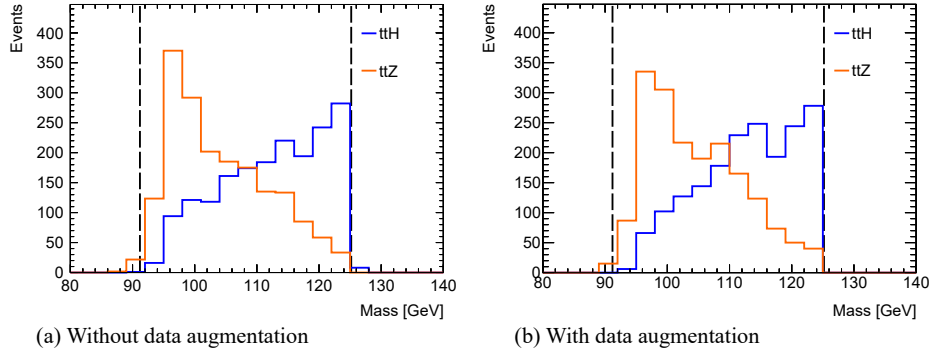


Figure A.1: Demonstration of the effect of data augmentation on the mass reconstruction NN. The values derived from the histograms are in Table [A.4](#).

Table A.4: Mass reconstruction metrics - data augmentation
Neural network trained with data augmentation achieved better separation and mean of the mass distribution closer to the 125 GeV constant.

	Metric	Augmentation	No augmentation
	Separation	71.01%	69.38%
	Higgs mass Gaussian fit μ	117.27	111.30
	Higgs mass Gaussian fit σ	10.42	7.39
	Higgs mass μ	112.87	112.48
	Higgs mass σ	7.85	8.41

Appendix B

Real data selection criteria

In this appendix is the selection used on the real ATLAS data to extract the wanted events.

```
if ((abs(lep_ID_0)==13 and (lep_isMedium_0) and (lep_isolationFCLoose_0)
and event.passPLIVVeryTight_0) or (abs(lep_ID_0)==11 and (lep_isolationFCLoose_0)
and (lep_isTightLH_0) and event.lep_chargeIDBDTResult_recalc_rel207_tight_0>0.7
and event.passPLIVVeryTight_0)) and ((abs(lep_ID_1)==13 and (lep_isMedium_1)
and (lep_isolationFCLoose_1) and event.passPLIVVeryTight_1) or (abs(lep_ID_1)==11
and (lep_isolationFCLoose_1) and (lep_isTightLH_1) and
event.lep_chargeIDBDTResult_recalc_rel207_tight_1>0.7 and event.passPLIVVeryTight_1))
and (((abs(lep_ID_0) == 13) or ( abs( event.lep_ID_0 ) == 11 and
(lep_ambiguityType_0) == 0 and ( not ((lep_Mtrkrk_atPV_CO_0<0.1
and event.lep_Mtrkrk_atPV_CO_0>0) and not (lep_RadiusCO_0>20 and
(lep_Mtrkrk_atConvV_CO_0<0.1 and event.lep_Mtrkrk_atConvV_CO_0>0))))
and not (lep_RadiusCO_0>20 and (lep_Mtrkrk_atConvV_CO_0<0.1 and
event.lep_Mtrkrk_atConvV_CO_0>0)))))) and ((abs( event.lep_ID_1 ) ==
11 and (lep_ambiguityType_1) == 0 and not ((lep_Mtrkrk_atPV_CO_1<0.1
and event.lep_Mtrkrk_atPV_CO_1>0) and not (lep_RadiusCO_1>20 and
(lep_Mtrkrk_atConvV_CO_1<0.1 and event.lep_Mtrkrk_atConvV_CO_1>0))))
and not (lep_RadiusCO_1>20 and (lep_Mtrkrk_atConvV_CO_1<0.1 and
event.lep_Mtrkrk_atConvV_CO_1>0)))) or (abs(lep_ID_1) == 13))) and
(nTaus_OR_Pt25)>=1 and ((nJets_OR_TauOR)>2 and (nJets_OR_DL1r_70)>0)
and (dilep_type and event.lep_ID_0*event.lep_ID_1>0)
```

Figure B.1: Selection used with the real ATLAS data in Python code

Appendix C

Bibliography

- [1] ATLAS Collaboration. *The Higgs boson's shadow*. July 2014. URL: <https://atlas.cern/updates/briefing/higgs-boson-s-shadow>.
- [2] M. Tanabashi et al. (Particle Data Group). *Review of Particle Physics*. An excerpt of data from the original article. 2018. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001). URL: <https://pdg.lbl.gov/2018/tables/rpp2018-sum-gauge-higgs-bosons.pdf>.
- [3] CERN. *Our Mission*. URL: <https://home.cern/about/who-we-are/our-mission>.
- [4] CERN. *Where did it all begin?* URL: <https://home.cern/about/who-we-are/our-history>.
- [5] CERN. *The Large Hadron Collider*. URL: <https://home.cern/science/accelerators/large-hadron-collider>.
- [6] Lyndon R Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008). This report is an abridged version of the LHC Design Report (CERN-2004-003), pp. i, 1. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001). URL: <https://cds.cern.ch/record/1129806>.
- [7] Arpad Horvath. *The LHC experiments and the preaccelerators*. Indications of used particles removed. 2006. URL: <https://commons.wikimedia.org/wiki/File:LHC.svg>.
- [8] CERN. *Experiments*. URL: <https://home.cern/science/experiments>.
- [9] ATLAS Outreach. *ATLAS Fact Sheet : To raise awareness of the ATLAS detector and collaboration on the LHC*. 2010. DOI: [10.17181/CERN.1LN2.J772](https://doi.org/10.17181/CERN.1LN2.J772). URL: <https://cds.cern.ch/record/1457044>.
- [10] Ana Maria Rodriguez Vera, Katarina Anthony, and Sascha Mehlhase. *Detector Overview - ATLAS Fact Sheet*. Image cut from fact sheet. Mar. 2021. URL: <https://cds.cern.ch/record/2759657>.
- [11] Tara Shears. “The Standard Model”. In: *Phil. Trans. R. Soc. A* 370 (2012), p. 805. DOI: [10.1098/rsta.2011.0314](https://doi.org/10.1098/rsta.2011.0314). URL: <https://royalsocietypublishing.org/doi/10.1098/rsta.2011.0314>.

- [12] MissMJ and Cush. *Standard Model of Elementary Particles*. Sept. 2019. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [13] M. Tanabashi et al. (Particle Data Group). “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [14] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. “An operational definition of quark and gluon jets”. In: *Journal of High Energy Physics* 2018.11 (Nov. 2018). ISSN: 1029-8479. DOI: [10.1007/JHEP11\(2018\)059](https://doi.org/10.1007/JHEP11(2018)059). URL: [http://dx.doi.org/10.1007/JHEP11\(2018\)059](http://dx.doi.org/10.1007/JHEP11(2018)059).
- [15] Giorgia Rauco. *Distinguishing quark and gluon jets at the LHC. Quark-gluon separation at the LHC*. Tech. rep. Geneva: CERN, Jan. 2017. URL: <https://cds.cern.ch/record/2254861>.
- [16] Patricia R. Burchat. “Decays of the tau lepton”. PhD thesis. Stanford Linear Accelerator Center, Stanford University, Feb. 1986, pp. 1–3. URL: <https://www.slac.stanford.edu/pubs/slacreports/reports14/slac-r-292.pdf>.
- [17] Camilla Galloni. *Hadronic tau reconstruction and identification performance in ATLAS and CMS*. Presentation slides. June 2018. URL: https://indico.cern.ch/event/681549/contributions/2930926/attachments/1661163/2680398/LHCP_2018_tau.pdf.
- [18] Dr Susan Cartwright. *Neutrinos and Neutrino Oscillations*. URL: https://www.hep.shef.ac.uk/cartwright/pubweb/neutrinos_advanced.html.
- [19] Bc. Petr Urban. “Large-Scale Data Analysis for Higgs Boson Mass Reconstruction in ttH Production”. PhD thesis. CTU, Faculty of Information Technology, Department of Theoretical Computer Science, May 2019. URL: <https://cds.cern.ch/record/2699450>.
- [20] ATLAS Collaboration. *Mass / Invariant mass*. ATLAS online glossary definition. URL: <https://atlas.cern/glossary/mass>.
- [21] J.D. Jackson. *Kinematics*. 2008. URL: <https://pdg.lbl.gov/2008/reviews/kinemarpp.pdf>.
- [22] CERN. *The Higgs boson: a landmark discovery*. URL: <https://atlas.cern/discover/physics/higgs>.
- [23] André Sopczak. *Precision Measurements in the Higgs Sector at ATLAS and CMS*. Presentation slides. June 2019. URL: <https://sopczak.web.cern.ch/sopczak/lhc/ATL-PHYS-SLIDE-2019-258.pdf>.

- [24] M. Aaboud et al. “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”. In: *Physics Letters B* 784 (Sept. 2018), pp. 173–191. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2018.07.035](https://doi.org/10.1016/j.physletb.2018.07.035). URL: <http://dx.doi.org/10.1016/j.physletb.2018.07.035>.
- [25] Vincent Hedberg and Oxana Smirnova. *W and Z bosons, Neutral current reactions, The discovery of the W and Z, W and Z physics at LEP, Charged current reactions*. Slides of a lecture. 2005. URL: http://hedberg.web.cern.ch/hedberg/lectures/ch8_2005_lec3.pdf.
- [26] A. Elagin et al. “A new mass reconstruction technique for resonances decaying to”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (Oct. 2011), pp. 481–489. ISSN: 0168-9002. DOI: [10.1016/j.nima.2011.07.009](https://doi.org/10.1016/j.nima.2011.07.009). URL: <http://dx.doi.org/10.1016/j.nima.2011.07.009>.
- [27] Charu C Aggarwal. *Neural networks and deep learning: A textbook*. en. 1st ed. Basel, Switzerland: Springer International Publishing, 2018. ISBN: 978-3-319-94463-0. ebook.
- [28] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning Representations by Back-propagating Errors”. In: *Nature* 323.6088 (1986), pp. 533–536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). URL: <http://www.nature.com/articles/323533a0>.
- [29] Daniel Godoy. *Understanding binary cross-entropy / log loss: a visual explanation*. 2018. URL: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a>.
- [30] PYTHIA authors. *PYTHIA 8*. Authors listed at the URL. URL: <https://pythia.org/latest-manual/Welcome.html>.
- [31] ATLAS Collaboration. “Search for squarks and gluinos in events with hadronically decaying tau leptons, jets and missing transverse momentum in proton–proton collisions at $\sqrt{s} = 13$ s = 13 TeV recorded with the ATLAS detector”. In: *The European Physical Journal C* 76.12 (Dec. 2016), p. 8. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4481-2](https://doi.org/10.1140/epjc/s10052-016-4481-2). URL: <http://dx.doi.org/10.1140/epjc/s10052-016-4481-2>.
- [32] ATLAS Collaboration. *Resolution*. ATLAS online glossary definition. URL: <https://atlas.cern/glossary/resolution>.
- [33] ROOT Team. *About ROOT*. URL: <https://root.cern/about/>.
- [34] ATLAS collaboration. *Measurements of Higgs boson production cross-sections in the $H \rightarrow \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, Aug. 2021. URL: <https://cds.cern.ch/record/2779179>.
- [35] Bingxuan Liu. “B-tagging in ATLAS”. In: (Jan. 2020). URL: <https://cds.cern.ch/record/2707367>.

- [36] M. Erdmann, B. Fischer, and M. Rieger. “Jet-parton assignment in $t\bar{t}H$ events using deep learning”. In: *Journal of Instrumentation* 12.08 (Aug. 2017). ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/08/p08020](https://doi.org/10.1088/1748-0221/12/08/p08020). URL: <http://dx.doi.org/10.1088/1748-0221/12/08/P08020>.
- [37] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008). Also published by CERN Geneva in 2010. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003). URL: <https://cds.cern.ch/record/1129811>.
- [38] Ahmed Tarek Abouelfadl Mohamed. “The LHC and the ATLAS Experiment”. In: *Measurement of Higgs Boson Production Cross Sections in the Diphoton Channel: with the full ATLAS Run-2 Data and Constraints on Anomalous Higgs Boson Interactions*. Cham: Springer International Publishing, 2020, pp. 61–101. ISBN: 978-3-030-59516-6. DOI: [10.1007/978-3-030-59516-6_4](https://doi.org/10.1007/978-3-030-59516-6_4). URL: https://doi.org/10.1007/978-3-030-59516-6_4.
- [39] ATLAS. *Athena*. 2021. URL: <https://gitlab.cern.ch/atlas/athena>.
- [40] Davide Chicco and Giuseppe Jurman. “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.” In: *BMC Genomics* 21.6 (2020). ISSN: 1471-2164. DOI: [10.1186/s12864-019-6413-7](https://doi.org/10.1186/s12864-019-6413-7). URL: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7>.
- [41] T. Mrkvička. *Základy teorie pravděpodobnosti a matematické statistiky*. Jihočeská univerzita, 2016.