

**Implementation and evaluation of a
categorical classifier for signal and
background isolation in the tHq channel at
ATLAS**

Felix Antonio Junjiro Obando Molina

Bachelorarbeit in Physik
angefertigt im Physikalischen Institut

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

August 2023

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

Bonn, ..23.08.2023...
Datum

..........
Unterschrift

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachterin: Priv. Doz. Dr. Tatjana Lenz

Acknowledgements

First and foremost, I would like to thank Prof. Dr. Ian C. Brock for giving me the opportunity to work as part of his research group. I would also like to thank Priv. Doz. Dr. Tatjana Lenz for her willingness to evaluate this thesis.

A special thanks to Christian Kirfel, who supervised and guided my learning, and reviewed my thesis drafts. Furthermore, I would like to express my gratitude to Oliver But and Florian Kirfel for the proofreading of my thesis drafts. I would also like to thank Tanja Holm for her explanations on many concepts related to my thesis. Moreover, I would like to extend my thanks to all group members of the Brock group for their willingness to help, and for providing such a nice atmosphere.

Finally, but most importantly, I want to thank my mother and grandfather for their continuous support.

Contents

1	Introduction	1
2	The Standard Modell	3
2.1	Fermions	3
2.2	Bosons	4
2.3	Limitations of the standard model	5
3	The Measurement of the Higgs boson	7
3.1	Proton-Proton Collisions	7
3.2	Couplings in the Higgs bosons	7
3.3	Production of the Higgs boson	8
3.4	Decay of the Higgs boson	8
3.5	Multi-lepton analysis	10
3.6	tHq signal and backgrounds	10
4	Neural Networks	13
4.1	Artificial Intelligence	13
4.2	Machine Learning	13
4.3	Artificial Neural Networks	14
4.3.1	Important Concepts	16
5	The Large Hadron Collider and the ATLAS experiment	21
5.1	Colliders	21
5.2	The LHC	22
5.3	The ATLAS experiment	24
5.3.1	Coordinates	25
5.3.2	Inner Detector	25
5.3.3	Calorimeter	27
5.3.4	Muon Spectrometer	29
5.3.5	Trigger and Data detection	29
6	Analysis and results	31
6.1	Objective	31
6.2	Software used	31
6.3	Dataset used	32

6.4	Training	32
6.5	Results	33
6.5.1	Loss function and overfitting	33
6.5.2	Performance assessment	34
7	Conclusion	39
	Bibliography	41
	List of Figures	45
	List of Tables	47

Introduction

The standard model is one of the most successful theories of particle physics, describing the 17 fundamental particles and their interactions. One particle, the Higgs boson, is of particular importance as it is associated with the Higgs field, which is a scalar field responsible for the mass of the fundamental particles. One of the largest particle physics experiments, the LHC, can produce the Higgs boson by proton-proton collisions. The detection of the Higgs boson can then be done in the ATLAS experiment. Its detection allows the analysis of its properties and is therefore of great importance for the study of the standard model.

The tHq process is one of the ways in which the Higgs boson can be produced. It is an interesting but rare process, as it allows the study of the coupling of the Higgs boson to other particles, but can easily be overshadowed by backgrounds, which are processes that have similar decay products as those of the tHq signal. The detection of this process is therefore a difficult task, but can be done with the help of neural networks. The use of neural network models to tackle complex problems in particle physics has become increasingly popular in the last years. In this thesis, neural networks were used for the separation or isolation of the tHq signal in the ATLAS detector.

Previous work done by the tHq group involved the implementation of categorical classifiers based on neural network techniques to isolate the tHq signal, a general background, and a tZq background. As part of this thesis, a Python script was implemented to allow the training of categorical classifiers that consider not only the tHq signal and a general background, but also a user-defined number of specific backgrounds. This script also produces different performance plots, including ROC curves for each of the categories.

Several neural network models were trained and their performance was evaluated by means of the loss function and ROC curves. The data used for the training was produced by Monte Carlo simulations and provided by the tHq group. The relationship between the optimal number of training epochs to avoid overfitting and the number of specific backgrounds was also explored. Finally, a trained model will be suggested for its use on the isolation of the tHq signal.

This thesis will begin with a description of the standard model, which will serve as an introduction to the Higgs boson. Afterwards, the Higgs boson will be described along with its importance, production and decay processes. The following chapter will then deal with neural networks and the metrics that can be used to evaluate their performance. Next, an introduction to the ATLAS detector will be given. At the end, the results of the training of the neural network models will be given together with the conclusions.

The Standard Modell

The standard model is the name of one of the most successful theories of particle physics. It comprises 3 of the 4 fundamental forces and describes 17 fundamental particles [1], which are classified into fermions and bosons. A general overview of the fundamental particles of the standard model can be seen in figure 2.1.

In this chapter, the characteristics and classification of fermions and bosons are introduced. The fundamental forces that interact with each particle will also be mentioned. At the end, the limitations of the standard model are briefly discussed.

2.1 Fermions

Fermions are particles that have half-integer spin. Each fermion has its own anti-particle with same quantum numbers but opposite electric charge. Furthermore, they can be classified into quarks and leptons. Fermions are ruled by the exclusion principle, which states that two fermions cannot share the same quantum numbers at the same time [4].

The six existing types of quarks can be differentiated by their flavour charge, being the up u , down d , charm c , strange s , top (or truth) t and bottom (or beauty) b quarks. Moreover, quarks have three different colour charges, being green, blue and red; and they also possess electric charge. As they have colour, flavour and electric charge, they are the only particles that can interact with all 3 fundamental forces: the weak, the strong and the electromagnetic force. As a result, they are the only fundamental particles that can couple to all vector bosons.

On the other hand, leptons can be further classified into charged leptons, which have electric and flavour charge, and neutrinos, which only have flavour charge. As a result, charged leptons can interact with the weak and electromagnetic force, while neutrinos can only interact through the weak force. According to their flavour charge, charged leptons can be classified into electrons e , muons μ and tau τ leptons, while neutrinos can be classified into electron neutrino ν_e , muon neutrino ν_μ and tau neutrino ν_τ .

Neutrinos were initially thought to have no mass; however, the observation of neutrino oscillations showed that they have a very small not yet determined mass [5].

Quarks and leptons can also be further classified according to three generations or families. The first generation quarks encompass the up u and down d quark, while the first generation leptons includes

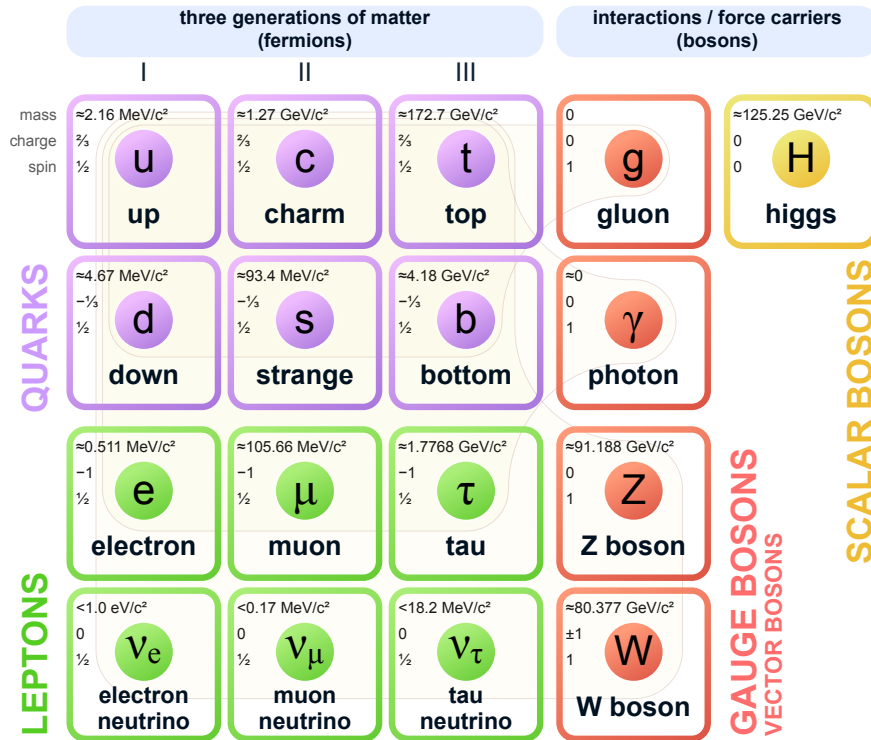


Figure 2.1: Fundamental particles of the standard model of Particle Physics [2]. Values were updated to reflect the actual ones to date 07.19.2023 [3]

the electron e and electron neutrino ν_e ; the second generation quarks incorporates the charm c and strange s quarks, while the second generation leptons consists of the muon μ and muon neutrino ν_μ . Finally, the third generation quarks includes the top t and bottom b quarks, while the third generation leptons encompass the tau τ and tau neutrino ν_τ . Generations also establish a mass hierarchy: higher generation fermions have greater masses than fermions of lower generations. Therefore, the lifetimes of higher generation fermions are shorter than those of lower generation fermions and are consequently less stable.

2.2 Bosons

Bosons are particles with integer spin. Elementary bosons can be classified into vector or gauge bosons, which have spin 1, and scalar bosons, which have spin 0. Vector bosons act as exchange particles in interactions. The photon γ is a massless vector boson that acts as a exchange particle for the electromagnetic interaction and has no electric charge. The gluon g is also a massless vector boson, but it is responsible for the strong interaction. Unlike the photon, the gluon carries the colour charge of the strong interaction. Finally, the W and Z are vector bosons responsible for the weak interaction. The W boson carries negative charge (or positive, in the case of its antiparticle), while the

Z carries neutral charge.

The only fundamental scalar boson is the Higgs boson H and it was the last particle of the standard model whose existence was proven [6]. The Higgs boson is a massive boson associated with the Higgs field, which is a scalar field responsible for the mass of the fundamental particles, being the quarks, leptons and W and Z bosons. The Higgs boson is of particular importance for this thesis, as the main signal treated comes from its decay. It will then be dealt with in more depth in the next chapter.

2.3 Limitations of the standard model

Even though all the particles predicted by the standard model have been experimentally observed, showing the success of this theory, the standard model has some flaws and limitations. For example, it does not provide a description of the fourth fundamental force, namely, gravity. It also does not have an explanation for the matter-antimatter asymmetry in our universe and does not account for the existence of dark energy and dark matter. Furthermore, according to the standard model, neutrinos have no mass; however, neutrino oscillations showed that neutrinos do have mass, yet a very small one.

As a result, several theoretical frameworks have been or are being developed in order to tackle these flaws and limitations (see for example [7]). The frameworks and experimental tests that attempt to explain or test the limitations and flaws of the standard model are commonly known as physics beyond the standard model, or BSM.

The Measurement of the Higgs boson

The Higgs boson was the last fundamental particle of the standard model to be found experimentally. It was first theorised by Higgs in 1964 [8] and found experimentally in 2012 [6] by the ATLAS collaboration. The Higgs boson is a particle associated with the Higgs field, which is a field that couples to fermions and quarks through the Yukawa coupling and thereby contributes to their mass. It also contributes to the mass of the vector bosons.

In this chapter, proton-proton collisions are briefly introduced as a way in which the Higgs boson can be produced in order to be analysed on experimental set-ups, like in the LHC. Then, the coupling mechanism in the Higgs bosons is mentioned in order to emphasise the importance of its study. Later, the production mechanisms of the Higgs boson are discussed to introduce the signal analysed as part of this thesis, namely, the tHq signal. The Higgs boson cannot be directly detected, but by the analysis of its decay products. Therefore, the decay of the Higgs boson and the analysis of its lepton final states are also described. At the end, backgrounds, which are processes that can be confused with the tHq signal, are explained.

3.1 Proton-Proton Collisions

One of the ways to produce the Higgs boson is through proton-proton collisions. A simplistic description of the proton p describes it as being made of 3 valence quarks: 2 u quarks and 1 d quark; however, the properties of these 3 quarks cannot account for some properties (like the mass) of the proton they make. A more complete description of the proton describes it as being made of the mentioned 3 valence quarks, sea quarks and gluons. Sea quarks are virtual quark-antiquark pairs that are created from gluons and can produce gluons through annihilation. During a proton-proton collision, these constituents interact with those of the other proton and can produce new particles, like the Higgs boson. Depending on the interaction, some quantum numbers can be violated, such as parity in the case of the weak interaction.

3.2 Couplings in the Higgs bosons

The study of the properties of the Higgs boson was the next step after discovering it. One of these properties, the strength, also called coupling, of the interaction between the Higgs boson and other

particles depends on the type and mass of the particle [9]. Accordingly, there are three types of couplings to massive particles: the gauge coupling, which couples the Higgs with the W and Z bosons, the self coupling of the Higgs boson with itself, and the Yukawa coupling, which describes the interaction between the Higgs boson and fermions.

The standard model describes the Yukawa coupling as being proportional to the mass of the fermion to which the Higgs couples. The Yukawa coupling can then be calculated from the mass of the coupled particles. When calculating the Yukawa coupling for the t quark it can be observed that it takes a value of $y_t \approx 1$ [3]. This is surprising considering that this value is the only one of this order of magnitude in comparison with the other fundamental particles. The experimental determination of this coupling can then provide relevant information of the standard model.

3.3 Production of the Higgs boson

The Higgs boson can be produced in the following ways [3]:

- via the heavy-quark loop-mediated gluon-gluon process (ggF), where the Higgs production takes place via virtual quark triangle loops (See Figure 3.1(a)).
- via vector-boson fusion (VBF), where a quark from each of the incoming protons radiates off either two W or two Z bosons that later fuse (See Figure 3.1(b)).
- via associated production with a pair of top quarks ($t\bar{t}H$). The Feynman diagram for this process can be seen in Figure 3.1(c).
- via associated production with a weak boson (W or Z), where two initial quarks change their flavour and produce a virtual vector boson that subsequently radiates a Higgs boson (See Figure 3.1(d)).
- via associated production with single top quark (tHq), which is the main interest of this thesis. The Feynman diagrams for this process is shown in Figure 3.2.

The tHq process has a t-channel structure, and its final state consists of two heavy objects, namely the t quark and the Higgs H boson, as well as a light flavour down-quark d , also called the spectator quark.

In the Feynman diagram shown in Figure 3.2(b) a u quark from one proton interacts through the weak interaction with a b quark that comes from the decay of a gluon of the other proton into b and \bar{b} . As a result of the weak interaction, the flavour of both quarks changes, so that the b turns into a t quark, and the u quark turns into a d quark. Then, the t quark radiates the Higgs boson.

In the Feynman Diagram shown in Figure 3.2(a), the Higgs boson is radiated by the W boson. On both cases, the initial quarks that radiated the vector boson are only deflected slightly. These can then be detected as jets by the detector. The tHq process can provide important information about the sign of the top Yukawa coupling. Its study is then relevant for research in the standard model.

3.4 Decay of the Higgs boson

The identification of the Higgs boson in the tHq process is not done directly, but indirectly by identifying its decay products. The most common decay modes and its branching ratios can be seen in

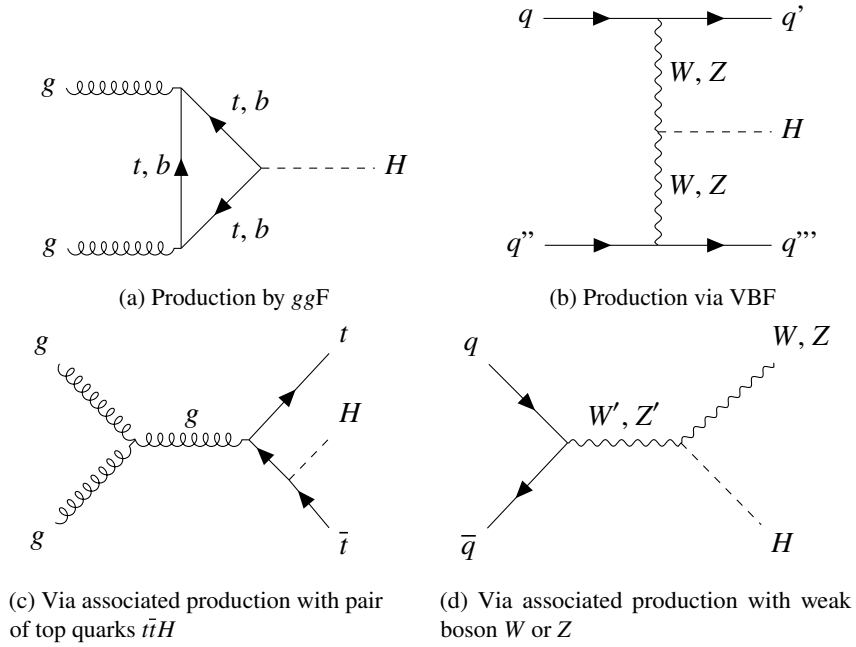


Figure 3.1: Feynman diagrams showing production processes of the Higgs boson.

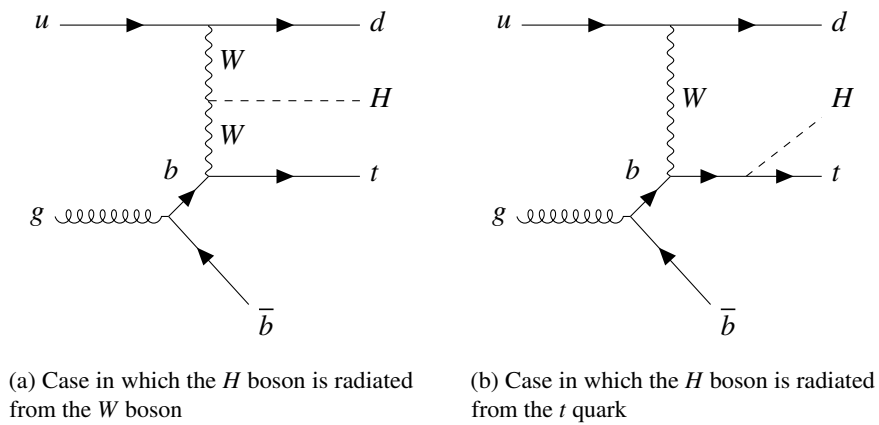


Figure 3.2: Higgs production by associated single top quark tHq .

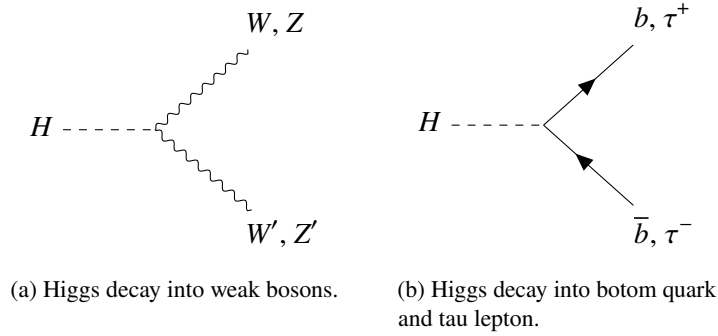


Figure 3.3: Most common Higgs decay modes.

the Table 3.1. In this thesis, we focus on one leptonic decay of the Higgs boson, namely, the decay to $\tau^+\tau^-$ leptons. This can also include the WW' and ZZ' decay, as they can subsequently decay into τ leptons. The most common decay modes are shown in Figure 3.3.

The resulting heavy $\tau^+\tau^-$ leptons can then decay hadronically or into lighter leptons, such as the e^\pm or μ^\pm , which can lead to a multi-lepton final state. The search for the tHq process uses signatures of these final states.

 Table 3.1: Branching ratios for the most common H decay modes [3].

H Decay Modes	Branchingratios %
WW'	25.7 \pm 2.5
ZZ'	2.8 \pm 0.3
$b\bar{b}$	53.0 \pm 8.0
$\tau^+\tau^-$	6.0 \pm 0.7

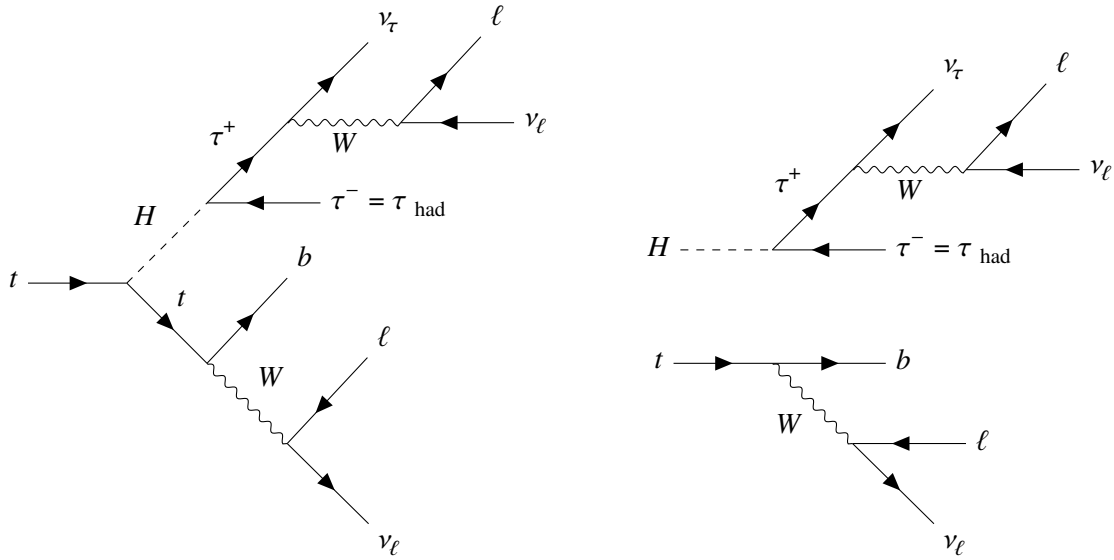
3.5 Multi-lepton analysis

In this analysis, the identification of the tHq process takes place by considering the multi-lepton final state, that is, the decay products of the $H \rightarrow \tau^+\tau^-$ decay mode.

The final state considered here is the decay into two leptons and one hadronic tau: $2\ell + 1\tau_{\text{had}}$. It is produced by the decay of the t lepton into a lighter lepton ℓ , and the decay of the Higgs boson H into two τ tau leptons, one of which further decays into a lighter lepton ℓ , and the other decays hadronically τ_{had} . This signature can be observed in Figures 3.4(a) and 3.4(b).

3.6 tHq signal and backgrounds

It can be observed that the H boson predominantly decays into heavy particles; however, it can be difficult to identify these decays, as there is a considerable amount of signals, called backgrounds, that have similar decay products and can be confused with the signal of the tHq process.

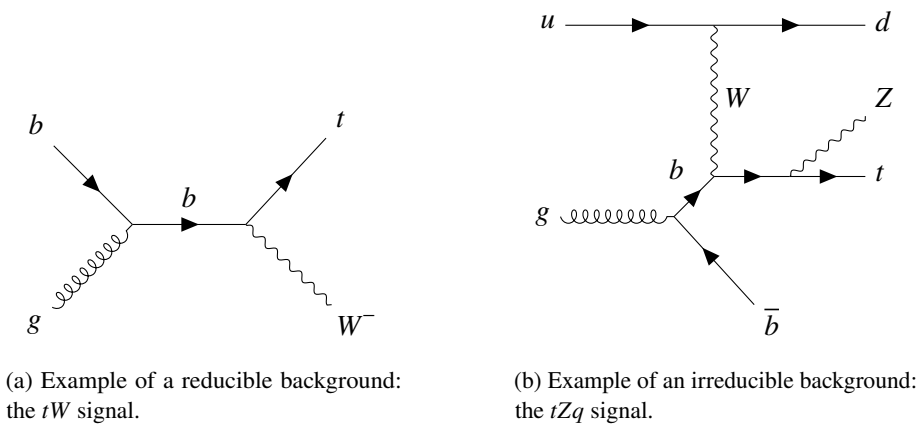


(a) $2\ell + 1\tau_{\text{had}}$ signature for the case in which the H boson was radiated from the t quark (Figure 3.2(b)).

(b) $2\ell + 1\tau_{\text{had}}$ signature for the case in which the H boson was radiated from the W boson (Figure 3.2(a)).

Figure 3.4: tHq multi-lepton final states.

In order to detect the tHq signal, we need to be able to identify the mentioned signatures and distinguish them from those of other backgrounds. Backgrounds are processes that are different from tHq , but that can have a multi-lepton final state that is identified as the same as that of the tHq process. Background processes can be classified into irreducible and reducible backgrounds. An irreducible background is a process whose final state is identical to that of tHq , while a reducible background is a process whose final state is different to that of the tHq , but that can be misreconstructed to end up looking like the final state of the tHq . An example of a reducible background is the tW process (see figure 3.5(a)). On the other hand, an example of an irreducible background, the tZq process, can be seen in figure 3.5(b).



(a) Example of a reducible background: the tW signal.

(b) Example of an irreducible background: the tZq signal.

Figure 3.5: Examples of tHq backgrounds.

Neural Networks

This chapter will start with the definition of artificial intelligence and machine learning. Then, artificial neural networks are described as a type of machine learning algorithm. As part of this description, important concepts related to neural networks are mentioned, including the metrics that can be used to evaluate the performance of a neural network model.

4.1 Artificial Intelligence

Artificial intelligence is a general term used to describe a technology, be it hardware or software, that is capable of mimicking a human behaviour or action [10].

Machine learning is a branch of artificial intelligence (AI) that imitates the way humans learn, gradually improving its accuracy [11].

4.2 Machine Learning

To better understand what machine learning is and its impact on engineering designs, we first have to mention what a conventional engineering design is in order to be able to understand the advantages of a machine learning based design.

The conventional engineering design requires a series of steps:

1. First, identifying the rules that govern the phenomenon that is going to be addressed by the design.
2. Second, based on these rules, the design of a mathematical model that captures the way the phenomenon works.
3. Third, the production of an algorithm and model optimised to the mentioned phenomenon under the assumption that that model is an idealised representation of reality.

A conventional engineering design is not suitable for real life phenomena, as real life phenomena are not ideal and can involve a huge amount of unexpected events that increase the complexity of the system and make it unpredictable for an ideal model. The Machine Learning based engineering design

does not require the first step, namely, the identification of the ideal rules that describe a phenomenon. It instead involves the acquisition of a big amount of samples that represent the phenomenon, and the subsequent training of a model to make it able to give the same outcomes as those of the samples [12]. The large amount of samples provides the design with the flexibility to learn from real life situations, making the machine learning model able to model a real life phenomenon without the need to know the rules that govern it.

There are three categories of machine learning [12]:

- **Supervised learning:** In this category, the data set used for the training of the model consists of labelled samples, which consist of input data and their expected outcomes or corresponding outputs. Supervised Machine Learning algorithms can be of two types: regression and classification. Regression algorithms are able to predict continuous values, while classification algorithms or models are able to predict discrete values, that is, whether an input corresponds to a category or not. Classification models can be used for classification between 2 categories, in which case it is called binary classifier, or for classification between more categories, in which case it is called categorical classifier.
- **Unsupervised learning:** the data set used consists of unlabelled samples, which consist of input data without their outcomes. This kind of model aims to find common properties of the samples in the system.
- **Reinforcement learning:** A reinforcement model receives feedback from the environment. This feedback indicates to what extent the output of the model satisfies a general intended outcome. For example, in a video game, the main character can obtain a greater score according to the greater amount of coins that it collects. In this example, the general intended outcome is to get a greater score, and the feedback received is the number of coins that the character obtains.

In this project, supervised machine learning will be used to train a categorical classifier model.

4.3 Artificial Neural Networks

An artificial neural network is a type of machine learning algorithm or model inspired by the way the human brain works.

The basic unit in a neural network is called the perceptron. It was first proposed in the year 1958 [13] and is a mathematical linear model that aims to emulate how a biological neuron works. It consists of a series of inputs x_n that are multiplied with corresponding weights w_n connected to a summing node, which is sometimes called neuron, to which a bias b is also added. The output is given by $y = \sum_{i=1}^n w_n \cdot x_n + b$, which is a linear function. A block diagram of a perceptron can be seen in figure 4.1.

An artificial neural network uses a series of perceptrons connected in parallel and in many layers. The first layer is called the input layer. The last layer is called the output layer, and its number of nodes corresponds to the total number of categories to which the neural network is tasked to classify. The layers between the input and output layers are called hidden layers. The total number of layers of the neural network define its depth. A block diagram of a neural network can be seen in figure 4.2.

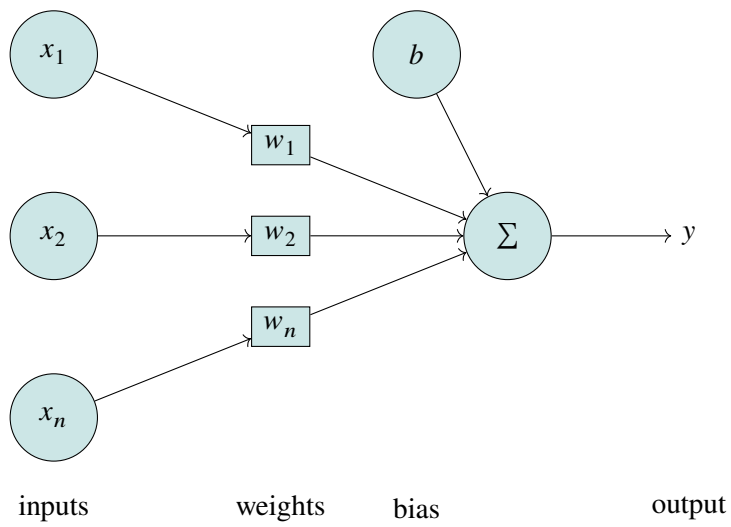


Figure 4.1: Block Diagram of a perceptron.

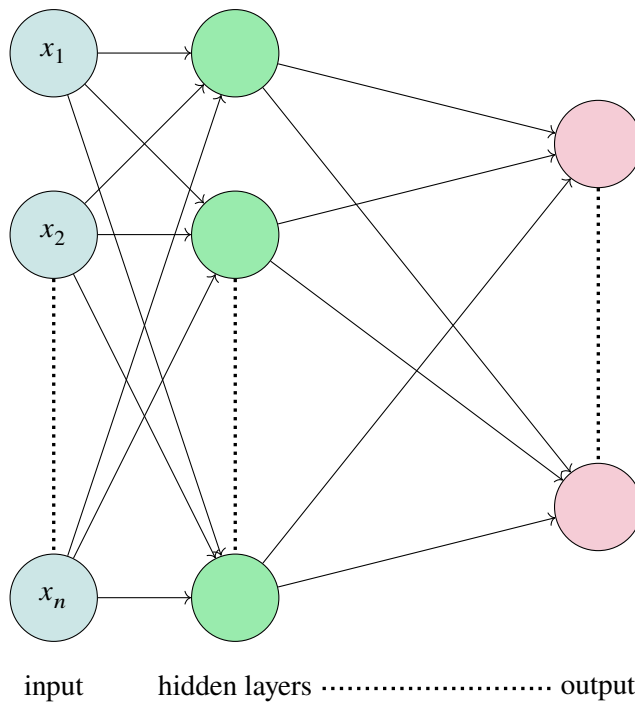


Figure 4.2: Block Diagram of a neural network with input nodes in blue, hidden layers nodes in green, and output nodes in red.

4.3.1 Important Concepts

The following concepts are important to understand how the neural network is used for this project.

Activation functions

Activation functions are used in the nodes and provide non-linearity to the neural network. According to the universal theorem [14], a neural network of one or more hidden layers that implements non-linearity is capable of approximating any convex continuous function. Activation functions can be either linear or non-linear, but it is only non-linear activation functions that are useful for neural networks, as they allow to perform back-propagation [15]. Common activation functions are the sigmoid function, the Exponential Linear Unit (ELU) function and the rectified linear activation function. This project uses multi-class classification, as it aims to classify a signal between many categories. There are two main types of multi class situations: non-exclusive classes, where a sample can be assigned to multiple categories; and exclusive classes, where a sample can be assigned to only one category. Accordingly, there is two types of activator functions that can be used for multi class situations:

- The sigmoid function, which is used for non-exclusive classes and gives a value between 0 and 1 depending on how probable it is for a sample to be classified in a category. The sum of the probabilities for all categories can be greater than 1.
- The softmax function, which is used for exclusive classes. Just as the sigmoid function, it gives a value between 0 and 1 for each category, but in contrast to the sigmoid function, the sum of the probabilities for all categories will be equal to 1.

In this project, the softmax function was used for the last layer, as a signal has to be exclusively classified in a class. On the other hand, the ELU function was used for the hidden layers.

Backpropagation

Backpropagation is an algorithm followed by the neural network in order to improve its performance. We describe its steps as follows:

After each training session or epoch, the model's performance is evaluated. This is done through the cost function, which quantifies the difference between the expected outcome (already known by the labelled data) and the calculated or predicted outcome.

In this project, the cost function used was the cross entropy loss function. The cross-entropy loss function $L(y, t)$ for a calculated output vector y and a labelled output vector t is given by the negative of the sum of each sample of the labelled output vector t_i times the natural logarithm of each sample of the calculated output vector y_i . Accordingly, the cross-entropy loss function can be seen in Eq. 4.1.

$$L(y, t) = - \sum_i t_i \ln y_i \quad (4.1)$$

The neural network tries to reduce the value of the cost function. For its reduction, the neural network has different algorithms, also called optimisation algorithms, at its disposal. The most basic optimisation algorithm is called Gradient Descent. It calculates the derivative or slope of the cost

function and attempts to reduce it so that it could get closer to a minimum. This reduction is done by updating the weights and biases of the neural network for the next epoch. The pace at which this update is done is given by the hyperparameter called learning rate. Extensions of the Gradient Descent algorithm can make the optimisation process more efficient and effective. The Adam algorithm [16], an extension of the gradient descent, constantly adapts the learning rate for each weight and is used in this project.

By means of the optimisation algorithm, the neural network is able to update its weights and biases for the next epoch, yielding better results for the cost function.

Evaluating performance

In order to evaluate the performance of the neural network, some metrics have to be calculated. There are 4 prediction types used to calculate the metrics. The prediction types are dependent on whether a prediction is positive or negative.

A prediction is said to be positive if it is classified as part of a category. On the other hand, a prediction is said to be negative if it is classified as not being part of that category. For example, if we want to identify a signal as a *tHq* signal, and not as a background signal (that is, *tHq* vs. background), then, the prediction will be positive if the signal is predicted to be a *tHq* signal, or it will be negative if the signal is predicted to be a background signal.

From the definition of positive and negative prediction, we can then define the prediction types, which are given as follows:

- True positive (TP): Describes the number of predictions rightly classified as positive.
- False positive (FP): Describes the number of prediction that are classified as positive but are not right, as they should have not been classified as positive. They are also known as error type 1.
- True negative (TN): Describes the number of predictions that are rightly classified as negative.
- False negative (FN): Describes the number of predictions that are classified as negative, but should have been classified as positive. They are also known as error type 2.

The metrics of the model are calculated based on these 4 prediction types. The three most important metrics are:

1. Accuracy: The accuracy measures the amount of right predictions, both negative and positive, with respect to the total amount of predictions. This metric is useful if the data set is well balanced. The accuracy is given by the formula:

$$\text{acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

2. Precision: It describes the ability of the model to find the most relevant cases from the point of view of the model, that is, the amount of retrieved and relevant cases with respect to the amount of all retrieved cases. It is measured with respect to the positive or negative value. For example, the precision with respect to the positive value measures the amount of positive predictions that

were right (retrieved and relevant cases) with respect to all positive predictions (retrieved cases). The precision with respect to the positive value is given by:

$$\text{prec} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Analogously, the precision with respect to the negative value is given by:

$$\text{prec} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

3. **Recall:** Describes the ability of the model to find the amount of retrieved and relevant cases with respect to all relevant cases. This metric is also measured with respect to the positive or negative value. If it is measured with respect to the positive value, then it will measure the amount of positive predictions that were right (retrieved and relevant cases) with respect to all samples that are labelled as positive (relevant cases). In this case, the value is given by

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The recall with respect to the negative value is also called specificity, and is given by

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

4. **F1-Score:** This metric takes the precision and recall metrics into account. It is defined as

$$F_1 = 2 \cdot \frac{\text{prec} \cdot \text{recall}}{\text{prec} + \text{recall}}$$

During the training and testing of the neural network, a plot of accuracy with respect to the number of epoch will be shown. This will be useful to evaluate the overall performance of the neural network during training.

ROC Curve and AUC

After the last epoch is done, the final model performance will be evaluated by means of the ROC curve [17], which stands for receiver operating characteristics curve. The ROC curve plots the True Positive Rate (TPR), which is nothing more than the recall with respect to the positive value, against the False Positive Rate (FPR), which is defined as the complement of the specificity: $\text{FPR} = 1 - \text{specificity}$. The definitions of the TPR and FPR can be seen in equation 4.2.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (4.2)$$

After getting the ROC curve, the area under the ROC curve (AUC) is calculated. The AUC, and by extension the ROC curve, provides a measurement of how well the neural network model is at class separation. In our project, the AUC should show us how well our model is at classifying the *tHq* samples as *tHq*, and the background samples as background. The values of AUC range between 0 and

1, and a value of 1 means that the trained model has a good separability capability. On the other hand, a value of 0.5 means that the model has, like a tossed coin, no ability to separate classes. An example of the ROC curve with its AUC can be seen in figure 4.3.

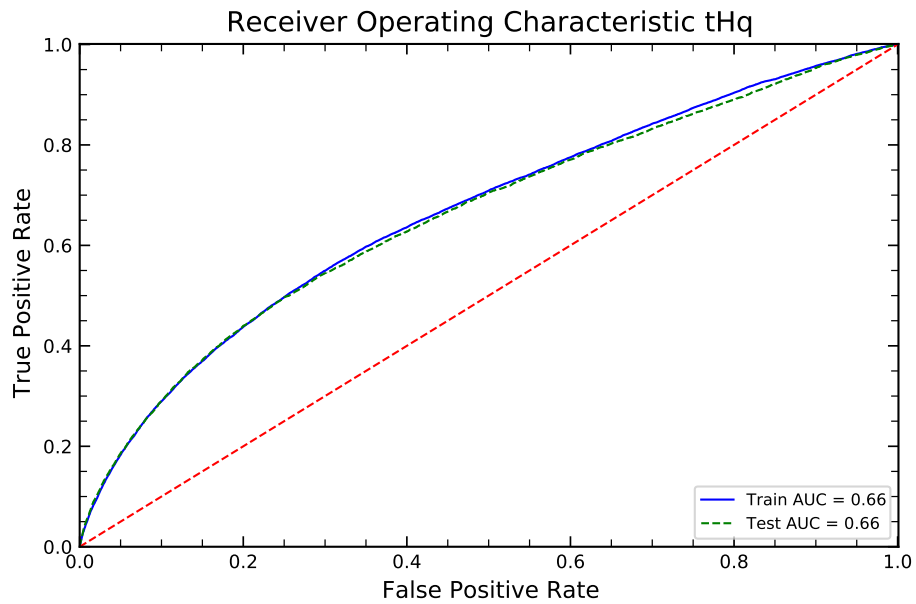


Figure 4.3: Example of ROC curve and AUC for a tHq signal.

The Large Hadron Collider and the ATLAS experiment

The ATLAS experiment, located in the Large Hadron Collider, or LHC, is the experimental set-up that can be used to produce and analyse the Higgs boson. In this chapter, a brief introduction to colliders and how they can be used to produce particles, including the Higgs boson, is given. Then, the LHC and its detectors are briefly introduced. Further, the ATLAS experiment, one of the LHC detectors, will be discussed. This discussion includes the coordinate system used by the experiment, together with the 3 subsystems used for particle detection.

5.1 Colliders

Colliders can accelerate particles through the use of the electromagnetic force and bring them to collide. From the collision of two particles, new particles, for example, the Higgs boson, can be created. Let $p_1 = (E_1, \vec{p}_1)$ and $p_2 = (E_2, \vec{p}_2)$ be the four-momentum of the first and second colliding particles with relativistic energies E_1 and E_2 and momenta \vec{p}_1 and \vec{p}_2 respectively. By assuming that both particles have the same rest mass $m_1 = m_2 = m$ and come from opposite directions, that is, they have opposing momentum $\vec{p}_1 = -\vec{p}_2 = \vec{p}$, the Lorentz-invariant energy variable \sqrt{s} , which is called the centre-of-mass energy, can be calculated:

$$s = (p_1 + p_2)^2 = m_1^2 + m_2^2 - 2\vec{p}_1\vec{p}_2 = 2m^2 + 2|\vec{p}|^2 + 2E_1E_2 \quad (5.1)$$

In a circular collider, both particles have the same relativistic energy $E = E_1 = E_2$, achieving a centre-of-mass energy of $\sqrt{s} = 2E$. New particles with energies smaller than \sqrt{s} can be created from the collision of two initial particles [5]. As a result, massive particles, like the Higgs boson or the top quark can be produced by the collision.

Circular colliders are particularly useful, as they allow particles to circulate multiple times and, by being subjected to an accelerating force for more time, achieve greater energies. However, a downside of circular colliders is that circulating charged particles lose energy by emitting synchrotron radiation. Accordingly, there is an energy loss U_0 that is anti proportional to the fourth power of the mass m of

the particle and the radius ρ of the circular path of the collider [18].

$$U_0 \propto \frac{1}{m^4 \rho} \quad (5.2)$$

When an electron e^- and a positron e^+ are used as colliding particles, synchrotron radiation makes it impossible in practice to achieve energies needed for the production of the Higgs boson. However, the bigger mass of the proton p makes the synchrotron radiation negligible. This is one of the reasons why protons are used in the LHC to study the production of the Higgs boson.

5.2 The LHC

The Large Hadron Collider, or LHC, is the world's largest particle collider and is located in the CERN's accelerator complex (see figure 5.1), in Geneva, Switzerland.

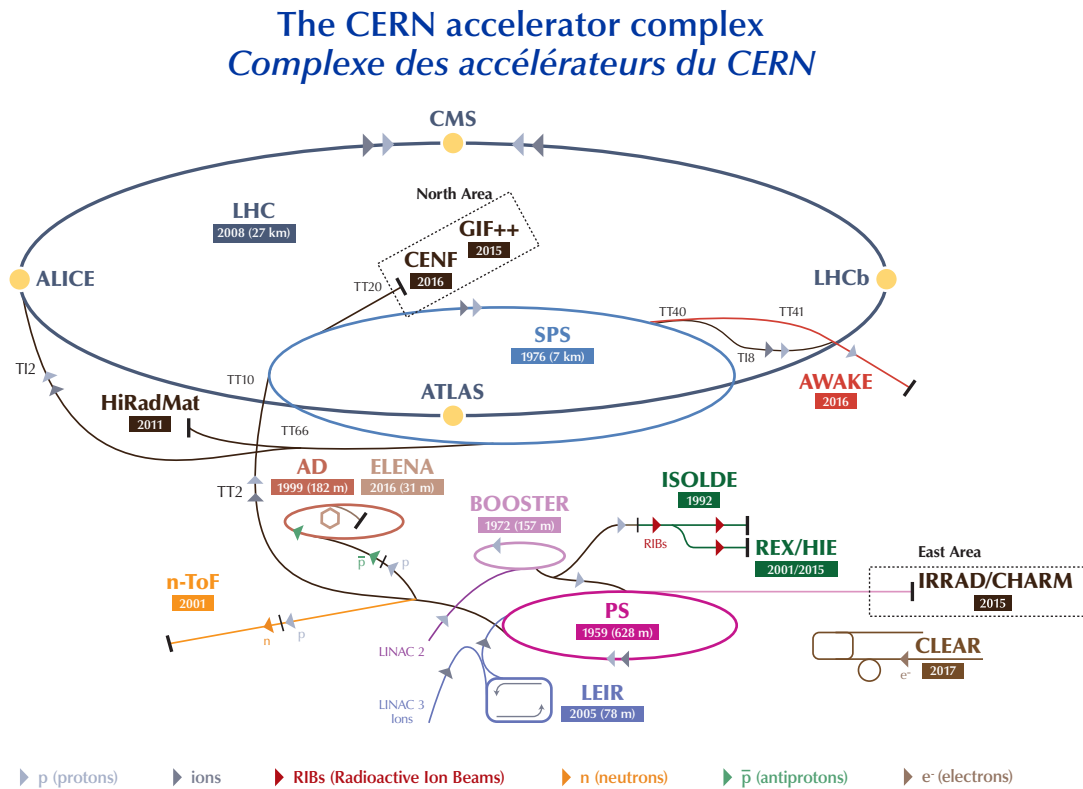
The Large Hadron Collider is a 27 km long circular collider in which protons or ions, which are accelerated to nearly the speed of light, are brought to collision. It consists of a series of accelerators that successively accelerate particles, delivering two high energy particle beams that travel in opposite directions and along two tubes. Both particle beams are then brought to collision in 4 locations in the LHC.

The LHC has three vacuum systems [20]: one for the insulation of the beam pipes and two for the insulation of the helium distribution line and the magnets. The beam pipes are kept in vacuum in order to avoid unwanted collisions of the circulating particles with gas particles. On the other hand, the helium distribution line and the magnets are kept in vacuum in order to ensure thermal isolation.

The LHC has a series of electromagnets along its path in order to bend the particles beam trajectory. These electromagnets have to be cooled to temperatures of -271.3°C so that they can be operated as superconductors and consequently be more efficient and reduce their resistance. For this reason, there is a liquid helium supply that allows the electromagnets to be kept at very low temperatures [21].

The collision of the particles takes place in 4 locations where the two tubes intersect and where the main particle detectors are located:

- ATLAS: A general purpose detector that consists of four major subsystems wrapped concentrically in layers around the collision point to measure different variables of the debris after the collision [22]. Since this is the most important detector for this thesis, it will be discussed further in the next sessions.
- CMS: The Compact Muon Solenoid, or CMS, is also a general purpose detector [23].
- ALICE: A Large Ion Collider Experiment, or ALICE, is a heavy-ion detector at the LHC designed to allow the study of quark-gluon plasma, which is a state of matter in which confinement is no longer required [24].
- LHCb: The Large Hadron Collider beauty experiment, or LHCb, is a detector specialised in the study of the beauty, more commonly called bottom, quark [25].



LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKEfield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n-ToF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials // CHARM - Cern High energy AcceleRator Mixed field facility // IRRAD - proton IRRADIation facility // GIF++ - Gamma Irradiation Facility // CENF - CERN Neutrino platFORM

Figure 5.1: CERN accelerator complex [19].

5.3 The ATLAS experiment

ATLAS stands for A Toroidal LHC Apparatus and it is a detector built around the collision point that has a cylindrical shape with dimensions of 46 m length, 25 m diameter and 7 000 t weight. It aims to measure the debris of the collision of two protons that have energies of 6.5 TeV each, measuring variables like the mass, momentum, lifetime and nuclear spin of the resulting particles after the collision. The detector system consists of a series of concentric cylinders that comprise 4 major subsystems which can be seen in figure 5.2:

- Inner detector: It is the first detector that detects the particles produced after the collision. It is used to measure the path and position of electrically charged particles, with the exception of the muon. It can also be used to determine the momentum and energy of the particles by measuring the curvature of their path.
- Calorimeters: Used to measure the energy of particles that interact through the strong and electromagnetic force, like hadrons, electrons and photons.
- Muon spectrometer: Used to measure the energy and path of muons.
- Magnet Systems: 2 magnetic systems are used to bend charged particles in the inner detector and in the muon spectrometer. They allow for the measurement of the electric charges and the momenta of the particles by bending their path through the Lorentz force.

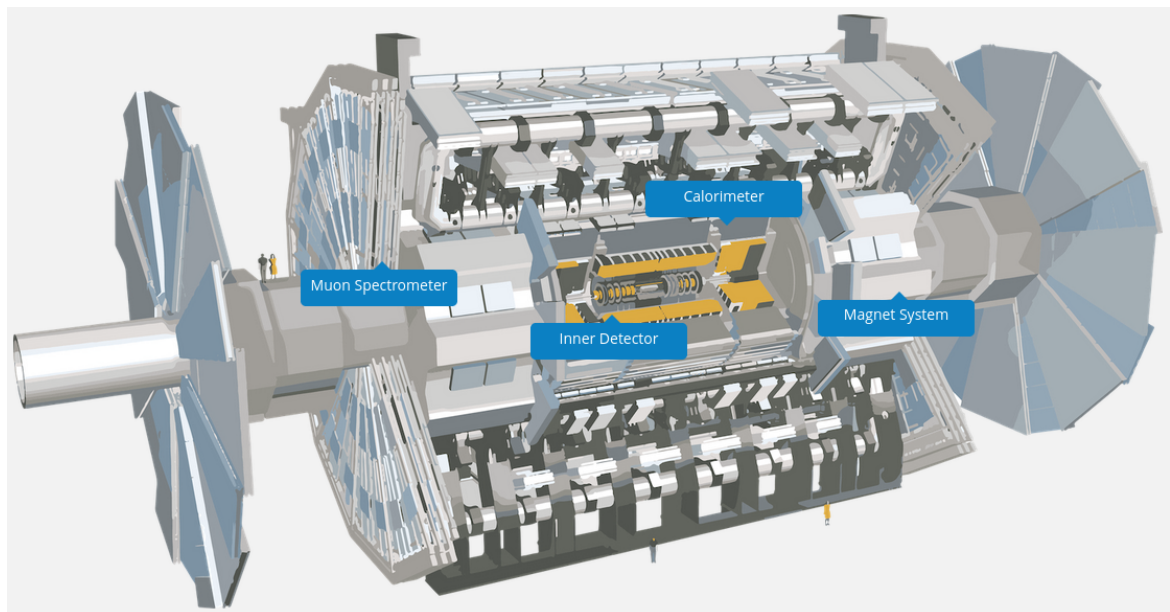


Figure 5.2: Sketch of the ATLAS detector [26].

Neutrinos are a special case. Neutrinos have a small mass and cannot interact with other particles neither through the electromagnetic force nor through the strong force. As the detectors used in the ATLAS experiment make use of the electromagnetic force and the strong force to detect and measure particles, the neutrinos cannot be directly detected. Their presence can however be indirectly inferred

by other variables, like the missing momentum. In order to satisfy the momentum conservation, the total momentum of the particles after the collision should be equal to the momentum before the collision; however, the measured momentum after the collision is smaller than the momentum before the collision. The difference between the measured momentum after the collision and the momentum before the collision is the missing momentum.

This section will begin with an introduction to the coordinate system used by the ATLAS experiment. Then, the subsystems of the LHC will be introduced, including the inner detector, the calorimeters and the muon spectrometer. The magnets systems will be introduced as part of the subsections of the calorimeter and the muon spectrometer. At the end, we will include a subsection about the data acquisition and triggering.

5.3.1 Coordinates

Many of the variables measured by the detector are expressed in terms of its coordinates. Let IP be the interaction point. According to the Cartesian coordinates, x is the vector pointing from the interaction point to the centre of the LHC, y is the vector perpendicular to the IP and to the proton beam and z is the vector pointing in the direction of the particle beam. The coordinate system used by the LHC uses the vector of the momentum of the particle $\vec{p} = (p_T, \eta, \phi)$ which is expressed in terms of the transversal momentum p_T , the pseudo-rapidity η and the azimuthal angle ϕ . The azimuthal angle ϕ takes values between π and $-\pi$ and is measured in the xy -plane from the positive x axis towards the positive y axis. The pseudo-rapidity η is calculated from the polar angle θ of the spherical coordinates and is given by Eq. 5.3. It can take values between $-\infty$ and ∞ .

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad (5.3)$$

The pseudo-rapidity is motivated from the rapidity, which is a spatial coordinate that is Lorentz invariant under longitudinal boosts. The pseudo-rapidity η is, as opposed to the rapidity, independent of the particle energy and momentum. Furthermore, the pseudo-rapidity and the rapidity are equal for high energies. Therefore, the use of the pseudo-rapidity is convenient for particle collisions in the LHC.

The transversal momentum p_T can be calculated by any of the spatial components of the particle momentum or its module, and the azimuthal angle or the pseudo-rapidity (see Eq. 5.7).

$$p_x = p_T \cdot \cos\phi \quad (5.4)$$

$$p_y = p_T \cdot \sin\phi \quad (5.5)$$

$$p_z = p_T \cdot \sinh\phi \quad (5.6)$$

$$|\vec{p}| = p_T \cdot \cosh\eta \quad (5.7)$$

The coordinate system used by the ATLAS detector can be seen in figure 5.3.

5.3.2 Inner Detector

The inner detector is located 3.3 cm from the LHC proton beam axis, extends to a radius of 1.2 m and has 6.2 m in length along the beam pipe [28].

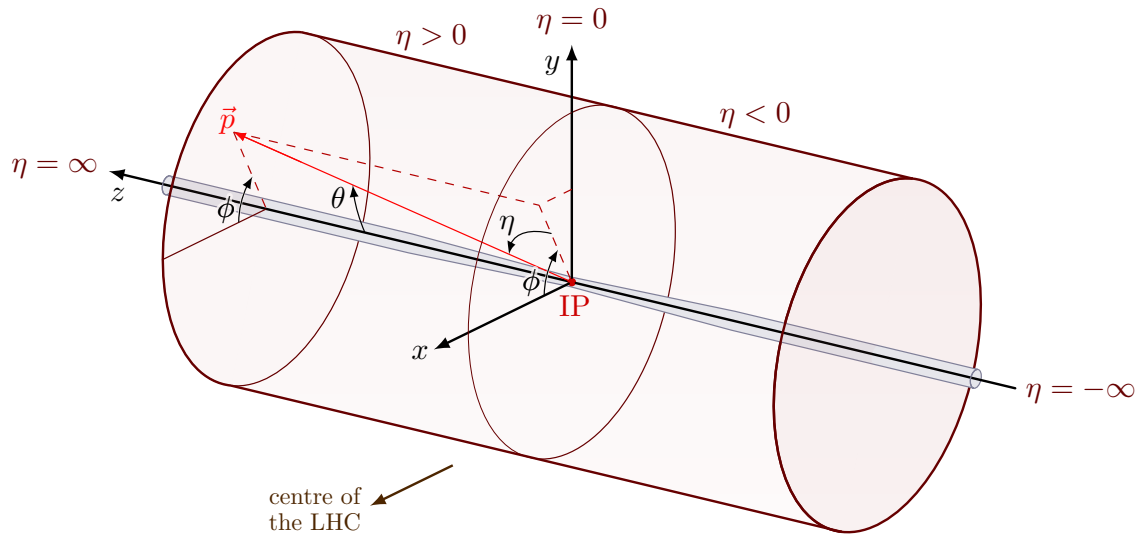


Figure 5.3: Coordinates used by the ATLAS detector [27].

It is immersed in a magnetic field that causes the particles, which result from the collision, to curve, delivering information about the particle electric charge, momentum and direction. It has tracking capabilities for values of $\eta < 2.5$ and consists of 3 main components: The pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT) [29].

An sketch of the ATLAS inner detector can be seen in figure 5.4.

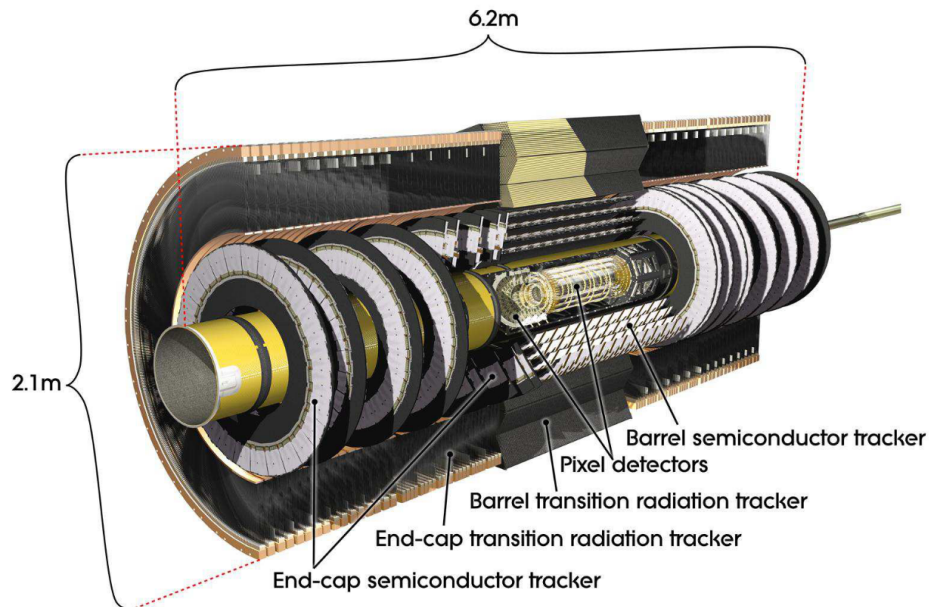


Figure 5.4: Sketch of the ATLAS inner detector [29].

Pixel Detector

The pixel detector is the layer of the ATLAS detector closest to the interaction point and consists of 4 concentric barrel layers with 1 736 sensor modules and 3 discs on either side of the beam tube with 288 modules. The innermost barrel layer was the last layer added as part of the LHC upgrade Run-2 and is called Insertable B-Layer or IBL. The 4 layers provide 3-dimensional space points that can be used for the precise tracking of the particles.

Semiconductor Tracker (SCT)

The SCT is a silicon microstrip detector located around the pixel detector. It measures particles over a larger area than the Pixel Detector with roughly the same accuracy, but only in 1 dimension. The SCT consists of 4 088 two sided modules and over 6 million readout strips and covers around 60 m².

Transition Radiation Tracker (TRT)

The TRT is the final layer of the inner detector. It consists of 300 000 drift tubes filled with gas with a 30 μ m gold-plated tungsten wire in the middle. Particles passing through the straws ionise the gas creating a detectable electric signal that is then used to trace the path of the particles and, thanks to transition radiation, provide information about the particle type. It is not as precise as the other detectors, but allows for the covering of a larger detection volume while reducing costs.

5.3.3 Calorimeter

The calorimeters surround the solenoidal magnet outside of the inner detector. They measure the energy of the particles by absorbing their energy and forcing them to stop.

There are two types of calorimeters that are implemented as part of the two ATLAS calorimeter systems: the electromagnetic calorimeter and the hadronic calorimeter.

1. **Electromagnetic Calorimeter (ECAL):** An ECAL absorbs energy from particles that interact through the electromagnetic force. In the case of photons, they lose their energy to the ECAL mainly through pair production. In the case of charged particles like electrons (including those coming from the pair production), they lose their energy through the emission of Bremsstrahlung, emitting photons of lower energies during this process. The simultaneous pair production and Bremsstrahlung emission creates an electromagnetic shower. The interaction of this shower with an active material can produce an electric current that can be amplified and measured.
2. **Hadronic Calorimeter:** A hadronic calorimeter absorbs the energy from particles that interact through the strong force, mainly hadrons. Hadrons interact with heavy nuclei in the passive material and produce other hadronic particles which can again interact with other heavy nuclei. This succession of interactions creates hadronic showers that are normally larger than electromagnetic showers. The hadronic showers can produce hadrons, like pions, that decay electromagnetically. These electromagnetically decaying hadrons produce then an electromagnetic shower that can interact with an active material and produce a measurable current.

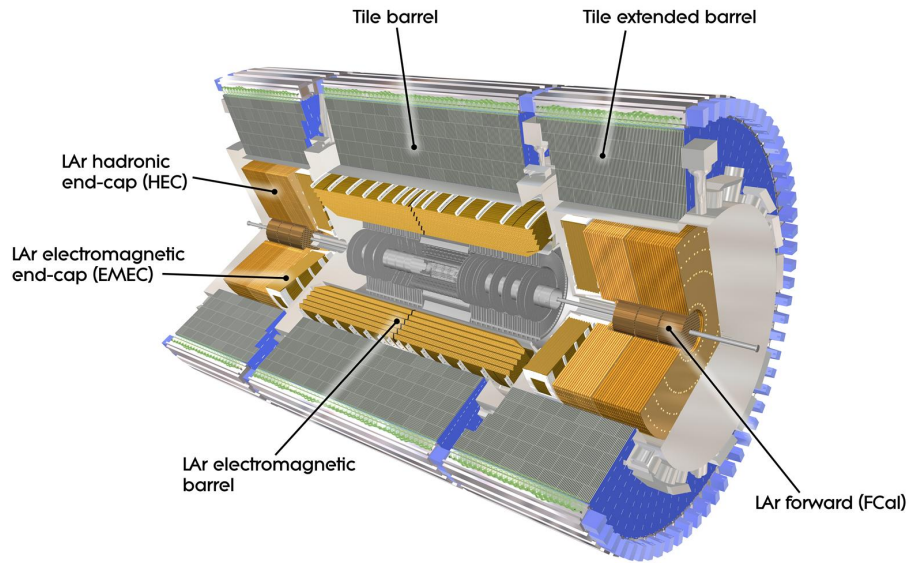


Figure 5.5: Sketch of the ATLAS calorimeter systems [30].

The ATLAS detector consists of two calorimeter systems: the Liquid Argon (LAr) calorimeter and the Tile Hadronic calorimeter [31]. Both systems make use of electromagnetic and hadronic calorimeters. A sketch of the calorimeter systems of the ATLAS detector can be seen in figure 5.5.

Liquid Argon Calorimeter (LAr)

The LAr calorimeter surrounds the ATLAS inner detector and consists of 4 parts [32]:

- An electromagnetic barrel (EMB) of 6.4 m length and 53 cm thickness. Here, lead is used as the passive material, which is used for the creation of the particle shower. It covers a pseudo-rapidity value of $|\eta| < 1.5$.
- The LAr electromagnetic end-cap (EMEC) has a thickness of 0.632 m and a radius of 2.077 m. Lead is also used as passive material. The EMEC spans values between $1.4 < |\eta| < 3.2$
- The hadronic endcap (HEC) consists of two wheels of thickness 0.8 m and 1.0 m and radius of 2.09 m. Plates of copper are used as passive material. The HEC covers a range of $1.5 < |\eta| < 3.2$.
- A forward calorimeter (FCal) which has three modules of radius 0.455 m and thickness 0.450 m each. It covers the pseudo-rapidity range between $3.1 < |\eta| < 4.9$. In the FCal, the first layer has as passive element copper, and for the second and third layer the passive element is tungsten.

After particles interact with the passive layers, electromagnetic showers are created. These showers ionise the active material, which in this case is liquid Argon at -184°C , creating ions and electrons that produce a current. The induced current can then be amplified and measured, so that the energy of the initial particles can be read.

Tile Hadronic Calorimeter

The Tile Calorimeter surrounds the LAr calorimeter and consists of an array of interleaved plates of steel (passive element) and plastic scintillators (active element). As particles hit steel, they generate a shower of new particles which generate photons in the scintillators. By means of a photomultiplier, photons create an electric current that can be measured to get the energy of the initial particles. It consists of a fixed central barrier (LB) and two moveable extended barrels (EB) and spans a pseudo-rapidity range of $|\eta| < 4.9$ [33].

5.3.4 Muon Spectrometer

Muons and electrons interact with matter mainly through Bremstrahlung; however, muons are less affected by Bremstrahlung as they have a mass that is a considerably bigger than the electron's mass. As a result, muons can cross the inner detector and calorimeters without leaving their energy and therefore without being measured.

A muon spectrometer is used to identify and measure the muon's momentum. The muon spectrometer of the ATLAS detector consists of a barrel covering values of $|\eta| \leq 1.2$, and two end-caps spanning the ranges of $1 < |\eta| < 2.7$ [34].

In the barrel, a toroidal magnet consisting of 8 superconducting coils creates a magnetic field that curves the path of the muon. The muon momentum can then be measured by tracking its curved trajectory, which is done by means of three stations equipped with Monitored Drift Tubes or MDT.

MDT are detectors composed of aluminium tubes of a diameter of 3 cm filled with a gas mixture of Ar and CO₂ and with a W-Rn 50 μm cable in the middle [34]. Muons crossing the MDT detectors ionize the gas whose ions drift to the wire and produce a measurable electric signal.

The MDT detectors are precision detectors used to determine the position of the particles. Additionally, the barrel also uses RPCs (Resistive Plate Chambers), which are fast response detectors used for triggering and selecting the relevant events.

On the other hand, each end-cap region consists of three tracking stations of chambers mounted to form three big discs and of an air core toroid placed between the first and second tracking station. MDT detectors are used in all stations with the exception of the inner station in the region $2 < |\eta| < 2.7$, where a Cathode Strip Chamber (CSC) is used. Thin Gap Chambers (TGC) are used as fast response detectors to select relevant events.

5.3.5 Trigger and Data detection

The ATLAS detector observes around 1.7 billion proton-proton collisions per second; however, only a small portion of these events contains relevant information, which amounts to approximately 1 000 events s^{-1} . The Trigger and Data Detection system consists of two stages and is used to reduce the event frequency [35].

- The first level trigger L1 is a hardware based trigger that reduces the amount of events to around 100 000 events s^{-1} . It takes its decision of whether to keep an event or not in less than 2.5 μs .
- The High Level Trigger or HLT is a software based trigger that reduces the number of events from 100 000 events s^{-1} to around 1 000 events s^{-1} .

Analysis and results

6.1 Objective

The main objective of this thesis is to implement a python script that can be used for the training of a categorical classifier capable of distinguishing between several categories:

1. The tHq signal.
2. A user defined amount of specific backgrounds, such as the $t\bar{t}$ process.
3. A general background that does not contain any of the specific backgrounds.

The implemented python script will be capable of distinguishing between all categories and deliver the corresponding performance plots, such as the loss function plots or the ROC curves of each category.

It will also be explored whether an increasing amount of specific backgrounds can improve the isolation of the tHq signal. Additionally, several classifiers will be trained and a suggestion for the choice of the best one will be given.

6.2 Software used

In this section, the python libraries used for the processing of data and the training of the neural network model are introduced.

The initial dataset that contained the tHq signal and the backgrounds was available in the .root format. This format can be processed and read by the ROOT [36] framework. ROOT was developed at CERN and it is a framework used for the handling and processing of large amounts of data. By using the ROOT API, the .root files were read and represented as Pandas dataframes, which are data structures commonly used for data science applications. Pandas [37] is a Python library that was used for the preprocessing of the data. NumPy [38], a library for scientific computing, was used for the reshaping of the data to satisfy the shape requirements for the training of the categorical classifier.

The categorical classifier was trained and implemented by making use of Keras [39], which is a library that provides an interface for TensorFlow [40]. TensorFlow is a Python library used for the implementation and training of neural network models. For the evaluation of the performance of the

models, both TensorFlow and Scikit-learn [41], a machine learning library, are used. Scikit-learn is also used for the data handling. For data visualisation and plotting, including the plotting of the ROC curves and the loss functions, matplotlib [42] was used.

6.3 Dataset used

The dataset used for the training of the neural network models consisted of events generated by Monte Carlo simulations [43]. This dataset contained samples from the tHq signal and several backgrounds in the 2ℓ OS + $1\tau_{\text{had}}$ channel. The 2ℓ OS + $1\tau_{\text{had}}$ is the channel in which the final lepton state of the channel $2\ell + 1\tau_{\text{had}}$ has two light leptons of opposite sign. Here, the lepton that has an opposite charge with respect to the τ_{had} is identified as originating from the Higgs-boson decay; while the lepton that has the same charge is associated with the top quark decay. Previous related research [44] pre-selected the events to only those events whose final states fulfilled several criteria, which required the energy of one light lepton to have at least 27 GeV, the presence of at least 2 jets with $|\eta| < 4.5$, the presence of one or two b -tagged jets, and the missing energy of each event to satisfy $5 \text{ GeV} < E_{\text{T}}^{\text{miss}} < 800 \text{ GeV}$. The event yields satisfying these criteria and used in this project are shown in table 6.1.

Table 6.1: Event yields in the 2ℓ OS + $1\tau_{\text{had}}$ channel [44].

Process	Event yields	
tHq	1.259 ± 0.032	
tZq	34.8	± 0.5
$t\bar{t}$	4 631	± 15
tW	223	± 6
W + jets	2.0	± 0.7
Z + jets	2 840	± 70
Diboson	137.8	± 2.0
$t\bar{t}W$	37.5	± 0.6
$t\bar{t}Z$	97.1	± 0.9
$t\bar{t}H$	36.49	± 0.18
tWZ/tWH	17.56	± 0.11
Other	8.9	± 1.8
Total background	8 070	± 70

6.4 Training

For the training of the neural network, the following steps were taken:

1. Reading of the .root files that contained samples from the tHq signal and backgrounds, and conversion of data to Pandas dataframes.
2. Preprocessing of the data to be able to use it for the training of the neural network. This included the filtering of opposite sign events (2ℓ OS + $1\tau_{\text{had}}$), the removal of unused features,

the conversion from Pandas dataframes to Numpy arrays, and the reshaping of Numpy arrays to satisfy the shape format required by TensorFlow.

3. Splitting of the dataset into testing and training datasets.
4. Training of the neural network model. The model is retrained many times by means of a series of training sessions or epochs in order to reduce the loss function and improve the performance.
5. Evaluation of the final neural network model performance by means of the loss function and the ROC curve for the tHq signal. It is worth mentioning that the implemented python script also delivers ROC curves for all categories. For example, the ROC curve for a specific background will consider that specific background as a positive value, and all other categories as negative values.

The hyperparameters are parameters that are defined by the user before the training of the neural network model and whose values manage and control the learning process. The hyperparameters used for the training of the models can be seen in table 6.2. They were chosen based on previous research on tHq signal separation [44], where an optimisation of the hyperparameters was carried out for similar situations. These hyperparameters have been shown to provide models with better prediction capabilities in other related research [44].

Additionally, the features used for the training are described in table 6.3. The chosen features were kinematic variables of the final state products of the tHq signal together with missing energy values. Other variables were not considered, as they did not deliver better results in other related research [44].

Table 6.2: Hyperparameters used in the training of the neural network models.

Hyperparameter	Value
Model	Categorical classifier
Number of nodes for input and hidden layers	120
Number of nodes for output layer	Same as number of categories
Number of hidden layers	6
Dropout	0.65
Batch Size	1000
Activation function for input and hidden nodes	Exponential Linear Unit (ELU)
Activation function for output node	softmax
Optimization algorithm	adam
Learning rate	0.01

6.5 Results

6.5.1 Loss function and overfitting

Initially, five models were trained for 200 epochs. These models considered as categories the signal tHq , a general background, and a certain number of specific backgrounds, which were:

1. $Z+$ jets
2. $Z+$ jets and $t\bar{t}$
3. $Z+$ jets, $t\bar{t}$ and tW
4. $Z+$ jets, $t\bar{t}$, tW and $t\bar{t}V$
5. $Z+$ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson

These specific backgrounds were chosen as they are the dominant backgrounds, that is, they possess the greatest number of event yields (see table 6.1).

The corresponding loss functions of the models are shown in figure 6.1. As expected, a greater number of epochs can introduce overfitting to our models. Overfitting is a big issue in machine learning models that occurs when a model performs well for the training dataset, but not so well for the testing dataset. As a result, overfitted models are not expected to reach their best performance when doing predictions with unseen data. Signs of overfitting can be observed in loss functions when the loss function of the testing dataset increases and has higher values than those of the training dataset. Accordingly, an optimal number of epochs is that for which the loss function of the training dataset and the testing dataset reach approximately the same value. The approximate optimal number of epochs (obtained from figure 6.1) can be read in table 6.4.

No proportionality relation can be observed between the optimal number of epochs and the number of categories, that is, an increase in the number of categories does not necessarily involve an increase in the optimal number of epochs. Other hyperparameters, such as the learning rate, might still play a role in how fast the optimal number of epochs can be reached; however, the tuning of other hyperparameters was not explored as part of this thesis.

In the next section, the performance of the new trained models with the optimal number of epochs and same hyperparameters will be assessed.

6.5.2 Performance assessment

The newly trained models with the approximate optimal number of epochs, as shown in table 6.4, were assessed. The ROC curves of these new trained models can be observed in figure 6.2. The ROC curves of the tHq signal were calculated by considering the tHq signal as a positive value, while considering the general background together with all specific backgrounds as negative values. The AUC values for each model can be read in table 6.5.

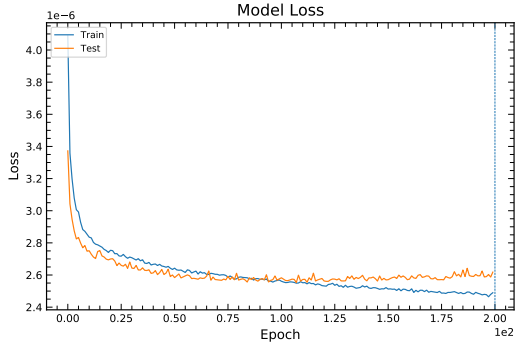
Additionally, a new model was trained (see figure 6.2(f)). This new model was included as its AUC results were considerably better as those of the other models. This model considered all previously used categories without the tW background, that is: $Z+$ jets, $t\bar{t}$, $t\bar{t}V$ and diboson. The table 6.5 also includes the results of this new trained model.

It can be seen that the models with either one or all categories delivered the second best AUC values for both the training and testing dataset; however, the AUC difference with the other models with lower AUC values is less than 0.03, and can therefore be considered as negligible. On the other hand, certain background combinations can achieve better results, as it is the case of the last model trained, which included all previously used categories without the tW background. This model achieved the best AUC results.

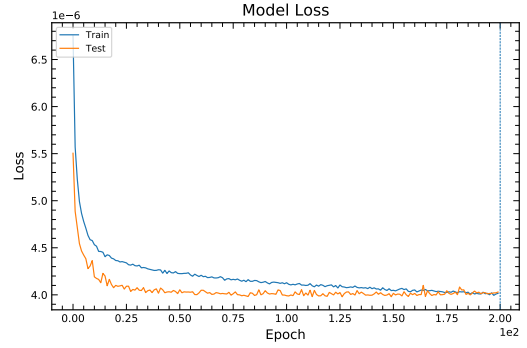
There are then four options that can be considered:

Table 6.3: Features considered for the training of the neural network models.

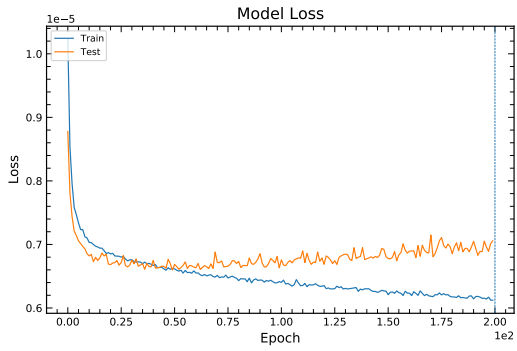
Feature	Meaning
eta_jf	η coordinate of the forward jet
pt_jf	p_T coordinate of the forward jet
phi_jf	ϕ coordinate of the forward jet
eta_b	η coordinate of the b-jet
pt_b	p_T coordinate of the b-jet
phi_b	ϕ coordinate of the b-jet
MMC_out_1	Reconstructed mass of the Higgs boson.
HvisPt	p_T of four-vector sum of hadronic taus
HvisEta	η of the four-vector sum of hadronic taus
TvisMass	Mass of the visible top quark decay products
TvisPt	p_T value of the visible top quark decay products
TvisEta	η value of the visible top quark decay products
m_met	Missing transverse energy
m_sumet	Sum of transverse energy without the missing energy
lep_Top_pt	p_T value of the lepton that decayed from the top quark
lep_Top_eta	η value of the lepton that decayed from the top quark
lep_Top_phi	ϕ value of the lepton that decayed from the top quark
m_phi_met	ϕ value of the missing transverse energy
pt_jet1	p_T value of the leading jet
eta_jet1	η value of the leading jet
phi_jet1	ϕ value of the leading jet
HT_all	Sum of the module of the transverse momentum
deltaPhiTau	Difference in ϕ of the hadronic tau and associated lepton
deltaRTau	Difference in distance R of the hadronic tau and associated lepton
had_tau_pt	p_T value of the hadronic tau
had_tau_eta	η value of the hadronic tau
had_tau_m	Mass of the hadronic tau
lep_Higgs_pt	p_T value of the Higgs associated light lepton
lep_Higgs_eta	η value of the Higgs associated light lepton
lep_Higgs_phi	ϕ value of the Higgs associated light lepton
M_b_jf	Mass of sum of the four-vector of the b-jet and forward jet
lep_top_higgs_mass	Sum of the mass of the Higgs boson and top quark associated leptons



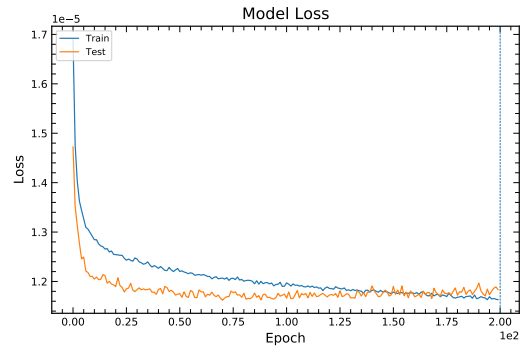
(a) Loss curve of model with specific background of $Z+$ jets.



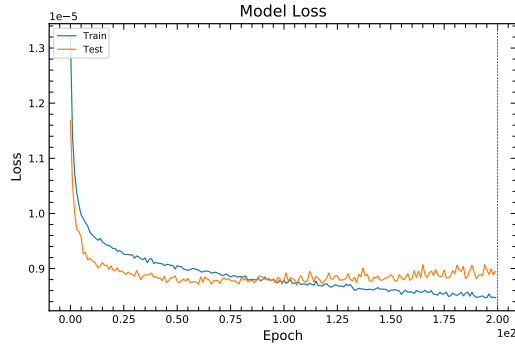
(b) Loss curve of model with specific backgrounds of $Z+$ jets and $t\bar{t}$.



(c) Loss curve of model with specific backgrounds of $Z+$ jets, $t\bar{t}$ and tW .

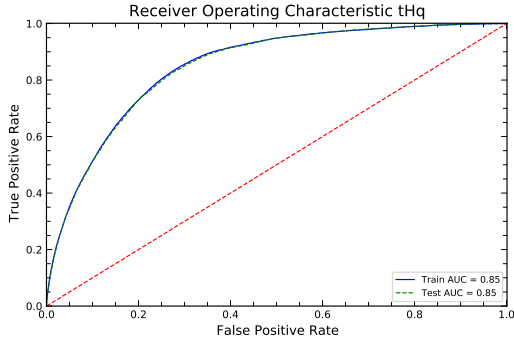


(d) Loss curve of model with specific backgrounds of $Z+$ jets, $t\bar{t}$, tW and $t\bar{t}V$.

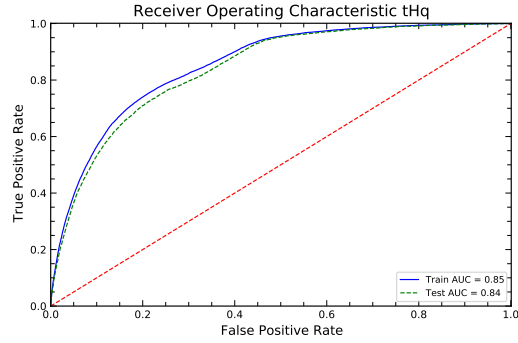


(e) Loss curve of model with specific backgrounds of $Z+$ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson.

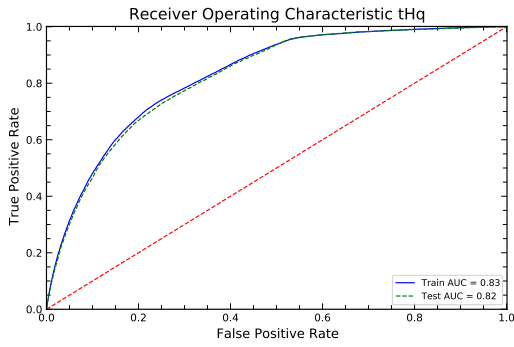
Figure 6.1: Loss function of different models trained for 200 epochs. The loss function for the testing dataset is shown in orange and for the training dataset in blue. All models consider the tHq signal, a general background and additional specific backgrounds: $Z+$ jets (a); $Z+$ jets and $t\bar{t}$ (b); $Z+$ jets, $t\bar{t}$ and tW (c); $Z+$ jets, $t\bar{t}$, tW and $t\bar{t}V$ (d); $Z+$ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson (e).



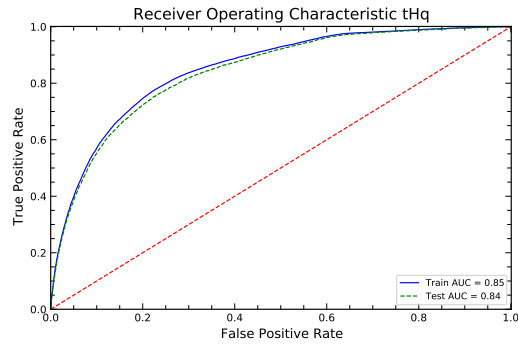
(a) ROC curve of model with categories tHq , background and Z + jets.



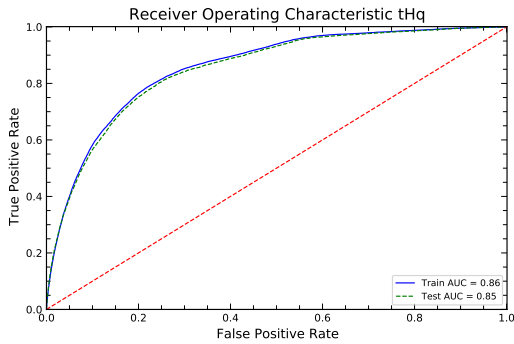
(b) ROC curve of model with categories tHq , background, Z + jets and $t\bar{t}$.



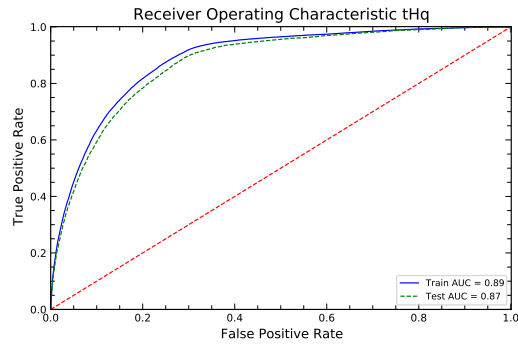
(c) ROC curve of model with categories tHq , background, Z + jets, $t\bar{t}$ and tW .



(d) ROC curve of model with categories tHq , background, Z + jets, $t\bar{t}$, tW and $t\bar{t}V$.



(e) ROC curve of model with categories tHq , background, Z + jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson.



(f) ROC curve of model with categories tHq , background, Z + jets, $t\bar{t}$, $t\bar{t}V$ and diboson.

Figure 6.2: ROC curves with AUC values for different models trained with the optimal number of epochs. All models consider the tHq signal, a general background and additional specific backgrounds: Z + jets (a); Z + jets and $t\bar{t}$ (b); Z + jets, $t\bar{t}$ and tW (c); Z + jets, $t\bar{t}$, tW and $t\bar{t}V$ (d); Z + jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson (e); Z + jets, $t\bar{t}$, $t\bar{t}V$ and diboson (f).

Table 6.4: Specific backgrounds for 5 different neural network models with their corresponding optimal number of epochs.

Specific backgrounds	Approximate optimal number of epochs	Time needed for training of model with optimal number of epochs / hh:mm:ss
Z+ jets	90	3:22:55
Z+ jets and $t\bar{t}$	200	5:41:26
Z+ jets, $t\bar{t}$ and tW	40	2:18:28
Z+ jets, $t\bar{t}$, tW and $t\bar{t}V$	155	2:47:50
Z+ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson	95	4:09:23

Table 6.5: AUC values for trained models with optimal number of epochs that considered different specific backgrounds.

Specific backgrounds	Epochs	AUC for training dataset	AUC for testing dataset
Z+ jets	90	0.85	0.85
Z+ jets and $t\bar{t}$	200	0.85	0.84
Z+ jets, $t\bar{t}$ and tW	40	0.83	0.82
Z+ jets, $t\bar{t}$, tW and $t\bar{t}V$	155	0.85	0.84
Z+ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson	95	0.86	0.85
Z+ jets, $t\bar{t}$, $t\bar{t}V$ and diboson	400	0.89	0.87

1. Use a model that considers only one specific background, such as the model with the Z+ jets category.
2. Use the model that considers all five dominant backgrounds as specific backgrounds: Z+ jets, $t\bar{t}$, tW , $t\bar{t}V$ and diboson. The advantages of this model is that it can also identify and separate each of the five specific backgrounds, while its performance is almost the same as that of the previous case.
3. Train models that consider all possible combinations of dominant backgrounds. This is not a realistic option, as it would involve the training of a big number of models and unrealistic computing times.
4. Use the model that considers the Z+ jets, $t\bar{t}$, $t\bar{t}V$ and diboson backgrounds as specific backgrounds. This would be the best option in terms of performance, as it delivered the best AUC results. It also has the second best ability to identify and separate the specific backgrounds.

Finally, it is suggested to use the model that considers the specific backgrounds Z+ jets, $t\bar{t}$, $t\bar{t}V$ and diboson, as this model achieved the best AUC values and it allows the user to identify and separate four specific backgrounds, a general background, and the tHq signal.

Conclusion

Previously trained models [44] only had the ability to isolate the tHq signal from a general background and a single specific background. As part of this thesis, a python script was implemented for the training of neural network models that can consider a higher number of specific backgrounds. By using this script, several models or categorical classifiers were trained that can be used for the isolation of the tHq signal and any of the specific backgrounds chosen. It is worth mentioning that the implemented script also delivers ROC curves for all chosen categories, including the tHq signal, the general background, and any of the specific backgrounds.

With respect to the optimal number of epochs, it was observed that an increasing number of specific backgrounds does not necessarily require an increasing number of epochs. In the same way, the models trained with the optimal number of epochs did deliver AUC values that only differ slightly; however, the training of these models used the same hyperparameters. The tuning of these hyperparameters depending on the categories could improve the models performance and is therefore suggested for future implementations.

To conclude, an additional model that considered the categories Z + jets, $t\bar{t}$, $t\bar{t}V$ and diboson could achieve the highest AUC results. It is suggested to use this model as it achieved the highest AUC results, and is also able to isolate the tHq signal, a general background, and four specific backgrounds.

Bibliography

- [1] Institute of Physics, *The Standard Model*, en, URL: <https://www.iop.org/explore-physics/big-ideas-physics/standard-model> (visited on 22/08/2023) (cit. on p. 3).
- [2] MissMJ, Cush, *The Standard Model*, en, 2023, URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 08/08/2023) (cit. on p. 4).
- [3] Particle Data Group et al., *Review of Particle Physics*, *Progress of Theoretical and Experimental Physics* **2022** (2022) 083C01, ISSN: 2050-3911, eprint: <https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf>, URL: <https://doi.org/10.1093/ptep/ptac097> (cit. on pp. 4, 8, 10).
- [4] G. Elert, “The Standard Model”, en, *The Physics Hypertextbook*, hypertextbook, 2021, URL: <https://physics.info/standard/> (visited on 22/08/2023) (cit. on p. 3).
- [5] B. Povh, K. Rith, C. Scholz, F. Zetsche and W. Rodejohann, *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte*, de, Springer-Lehrbuch, Berlin, Heidelberg: Springer, 2014, ISBN: 9783642378218, URL: <https://link.springer.com/10.1007/978-3-642-37822-5> (cit. on pp. 3, 21).
- [6] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012) 1, arXiv: 1207.7214, URL: <http://arxiv.org/abs/1207.7214> (cit. on pp. 5, 7).
- [7] J. D. Lykken, *Beyond the Standard Model*, (2011), arXiv: 1005.1676, URL: <http://arxiv.org/abs/1005.1676> (cit. on p. 5).
- [8] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Physical Review Letters* **13** (1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on p. 7).
- [9] ATLAS Collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022), 2207.00092: arXiv, URL: <http://arxiv.org/abs/2207.00092> (cit. on p. 8).
- [10] Collins English Dictionary, *Artificial intelligence definition and meaning*, en, 2023, URL: <https://www.collinsdictionary.com/dictionary/english/artificial-intelligence> (visited on 22/08/2023) (cit. on p. 13).

- [11] IBM, *What is Machine Learning?*, en-us, URL: <https://www.ibm.com/topics/machine-learning> (visited on 23/07/2023) (cit. on p. 13).
- [12] O. Simeone, *A Very Brief Introduction to Machine Learning With Applications to Communication Systems*, *IEEE Transactions on Cognitive Communications and Networking* **4** (2018) 648 (cit. on p. 14).
- [13] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, *Psychological Review* **65** (1958) 386, ISSN: 1939-1471 (cit. on p. 14).
- [14] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are universal approximators*, en, *Neural Networks* **2** (1989) 359, ISSN: 0893-6080 (cit. on p. 16).
- [15] R. M. Tharsanee, R. S. Soundariya, A. S. Kumar, M. Karthiga and S. Sountharajan, *7 - Deep convolutional neural network-based image classification for COVID-19 diagnosis*, en, (2021) 117, ed. by U. Kose, D. Gupta, V. H. C. de Albuquerque and A. Khanna, URL: <https://www.sciencedirect.com/science/article/pii/B9780128245361000125> (cit. on p. 16).
- [16] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, (2017), arXiv: 1412.6980 [cs.LG] (cit. on p. 17).
- [17] Google for Developers, *The ROC Curve and the AUC*, en, URL: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc> (visited on 22/08/2023) (cit. on p. 18).
- [18] Walker, R P, *Synchrotron radiation*, (1994), URL: <https://cds.cern.ch/record/398429> (cit. on p. 22).
- [19] E. Lopienska, *The CERN accelerator complex, layout in 2022*, URL: <https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1> (visited on 22/08/2023) (cit. on p. 23).
- [20] CERN, *A vacuum as empty as interstellar space*, en, 2023, URL: <https://home.cern/science/engineering/vacuum-empty-interstellar-space> (visited on 22/08/2023) (cit. on p. 22).
- [21] CERN, *Cryogenics: Low temperatures, high performance*, en, 2023, URL: <https://home.cern/science/engineering/cryogenics-low-temperatures-high-performance> (visited on 22/08/2023) (cit. on p. 22).
- [22] CERN, *The ATLAS Detector*, en, 2023, URL: <https://home.cern/science/experiments/atlas> (visited on 22/08/2023) (cit. on p. 22).
- [23] CERN, *The CMS Detector*, en, 2023, URL: <https://home.cern/science/experiments/cms> (visited on 22/08/2023) (cit. on p. 22).

-
- [24] CERN, *The ALICE Detector*, en, 2023,
URL: <https://home.cern/science/experiments/alice> (visited on 22/08/2023)
(cit. on p. 22).
- [25] CERN, *The LHCb Detector*, en, 2023,
URL: <https://home.cern/science/experiments/lhcb> (visited on 22/08/2023)
(cit. on p. 22).
- [26] ATLAS Experiment at CERN, *Detector and Technology*, en, 2023,
URL: <https://atlas.cern/Discover/Detector> (visited on 22/08/2023) (cit. on p. 24).
- [27] I. Neutelings, *CMS coordinate system*, en-US, 2021,
URL: https://tikz.net/axis3d_cms/ (visited on 22/08/2023) (cit. on p. 26).
- [28] L. Skinnari, *Alignment of the ATLAS Inner Detector tracking system*, (2011),
URL: <https://cds.cern.ch/record/1401179> (cit. on p. 25).
- [29] ATLAS Experiment at CERN, *The ATLAS Inner Detector*, en, URL:
<https://atlas.cern/Discover/Detector/Inner-Detector> (visited on 22/08/2023)
(cit. on p. 26).
- [30] J. Pequeno, *Computer Generated image of the ATLAS calorimeter*, 2008,
URL: <https://cds.cern.ch/images/CERN-GE-0803015-01> (visited on 22/08/2023)
(cit. on p. 28).
- [31] ATLAS Experiment at CERN, *The ATLAS Calorimeter*, en,
URL: <https://atlas.cern/Discover/Detector/Calorimeter> (visited on 22/08/2023)
(cit. on p. 28).
- [32] D. J. Mahon, *ATLAS LAr calorimeter performance in LHC Run 2*, en,
Journal of Instrumentation **15** (2020) C06045, ISSN: 1748-0221 (cit. on p. 28).
- [33] A. Henriques, “The ATLAS tile calorimeter”,
*2015 4th International Conference on Advancements in Nuclear Instrumentation Measurement
Methods and their Applications (ANIMMA)*, 2015 1 (cit. on p. 29).
- [34] L. Pontecorvo, *The ATLAS muon spectrometer*, en,
The European Physical Journal C - Particles and Fields **34** (2004) s117,
URL: <https://doi.org/10.1140/epjcd/s2004-04-013-y> (cit. on p. 29).
- [35] ATLAS Experiment at CERN, *ATLAS Trigger and Data Acquisition System*, en,
URL: <https://atlas.cern/Discover/Detector/Trigger-DAQ> (visited on 22/08/2023)
(cit. on p. 29).
- [36] R. Brun et al., *root-project/root: v6.18/02*, version v6-18-02, 2019,
URL: <https://doi.org/10.5281/zenodo.3895860> (cit. on p. 31).
- [37] The pandas development team, *pandas-dev/pandas: Pandas*, version v2.1.0rc0, 2023,
URL: <https://doi.org/10.5281/zenodo.8239932> (cit. on p. 31).
- [38] C. R. Harris et al., *Array programming with NumPy*, *Nature* **585** (2020) 357,
URL: <https://doi.org/10.1038/s41586-020-2649-2> (cit. on p. 31).
- [39] F. Chollet et al., *Keras*, <https://keras.io>, 2015, (visited on 22/08/2023) (cit. on p. 31).

Bibliography

- [40] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: <https://www.tensorflow.org/> (visited on 22/08/2023) (cit. on p. 31).
- [41] L. Buitinck et al., “API design for machine learning software: experiences from the scikit-learn project”, *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013 108 (cit. on p. 32).
- [42] J. D. Hunter, *Matplotlib: A 2D graphics environment*, *Computing in Science & Engineering* **9** (2007) 90 (cit. on p. 32).
- [43] ATLAS Collaboration, *Atlas Production Group MC16a/d/e*, 2021, URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SingleTopProductionV3411> (visited on 10/08/2023) (cit. on p. 32).
- [44] T. Holm, *Towards a measurement of the tHq process in channels with hadronic tau lepton decays*, PhD thesis: University of Bonn, 2023 (cit. on pp. 32, 33, 39).

List of Figures

2.1	Fundamental particles of the standard model of Particle Physics	4
3.1	Feynman diagrams showing production processes of the Higgs boson	9
3.2	Higgs production by associated single top quark tHq	9
3.3	Most common Higgs decay modes	10
3.4	tHq multi-lepton final states	11
3.5	Examples of tHq backgrounds	11
4.1	Block Diagram of a perceptron	15
4.2	Block Diagram of a neural network	15
4.3	Example of ROC curve and AUC for a tHq signal	19
5.1	CERN accelerator complex	23
5.2	Sketch of the ATLAS detector	24
5.3	Coordinates used by the ATLAS detector	26
5.4	Sketch of the ATLAS inner detector	26
5.5	Sketch of the ATLAS calorimeter systems	28
6.1	Loss function of different models trained for 200 epochs	36
6.2	ROC curves with AUC values for different models trained with the optimal number of epochs	37

List of Tables

3.1	Branching ratios for the most common H decay modes	10
6.1	Event yields in the $2\ell OS + 1\tau_{\text{had}}$ channel	32
6.2	Hyperparameters used in the training of the neural network models	33
6.3	Features considered for the training of the neural network models	35
6.4	Specific backgrounds for 5 different neural network models with their corresponding optimal number of epochs	38
6.5	AUC values for trained models with optimal number of epochs that considered different specific backgrounds	38