Department of Physics and Astronomy

Heidelberg University

Master thesis

in Physics

submitted by

Laia Parets Peris

born in Valencia

2021

Study of invisible Higgs boson decays produced via vector-boson fusion

This Master thesis has been carried out by Laia Parets Peris

at the

Kirchhoff-Institute for Physics

under the supervision of

apl. Prof. Dr. Monica Dunford

Untersuchung unsichtbarer Higgs-Bosonzerfälle, die durch Vektorbosonenfusion erzeugt werden

Viele Modelle der Physik, die jenseits des Standardmodells liegen, beinhalten die Kopplung des Higgs-Bosons an Dunkle Materie. Eine genaue Messung des Verzweigungsverhältnisses der unsichtbaren Zerfälle des Higgs-Teilchens könnte uns Einblicke in die Suche nach Neuer Physik gewähren. Das Signalmodell, welches in dieser Masterarbeit untersucht wird, beinhaltet den unsichtbaren Zerfall eines mit Vektorbosonfusion erzeugten Higgs-Bosons. Der Untergrund wird vom Zerfall des Z-Bosons in ein Neutrino und ein Antineutrino in Assoziation mit zwei Jets dominiert. Da der letztere eine größere hadronische Aktivität in der Rapiditätslücke aufgrund des Farbaustauschs aufweist, wird diese Eigenschaft ausgenutzt, um den Untergrund im Vergleich zum Signal zu unterdrücken. Die Strategie dieser Analyse basiert auf der Verwendung von ATLAS Kalorimeter- und Trackerbildern anstatt der traditionellen kinematischen Variablen, wie die Vierervektoren der Jets und die Anzahl der Jets in der Rapiditätslücke, um zu bestimmen, ob die Signal-Hintergrund-Trennung verbessert werden kann. Dazu wurde eine Analyse auf Basis von Autoencodern durchgeführt und die Wirkungsgrade und Ausschussraten untersucht.

Study of invisible Higgs boson decays produced via vector-boson fusion

Various models of physics beyond the Standard Model include the coupling of the Higgs boson to dark matter particles. A precise measurement of the branching ratio of the invisible decay of the Higgs boson could reveal new insights in the search for new physics. The signal model used in this thesis is the invisible decay of the Higgs boson produced via vector-boson fusion. The background is dominated by the Z boson decaying into a neutrino and an anti-neutrino in association with two jets. Since the latter is expected to have a larger hadronic activity in the rapidity gap due to colour exchange, this feature is exploited to supress the background compared to the signal. The strategy of this analysis is to use ATLAS calorimeter and ATLAS tracker images instead of traditional kinematic variables, such as the jet four-vectors and number of jets in the rapidity gap, to determine if the signal to background separation can be improved. For that, an analysis based on autoencoders was implemented and the efficiencies and rejection rates were studied.

Contents

Introduction

The Standard Model of Particle Physics (SM) is the current most accurate description of the known fundamental particles and the strong, weak and electromagnetic forces between them. Despite its large success in describing many experimental measurements, it is not the final theory of fundamental interactions because it leaves some open questions, such as the neutrino masses or the nature of dark matter and dark energy.

Several theories, such as supersymmetry or string theory, have been built to try to give an explaination to these questions. They are called beyond the SM theories. Some of them predict a bigger Higgs sector and different Yukawa couplings, although the coplings could be very similar to the ones predicted by the SM. Therefore, it is crucial to make precision measurements of the properties of the Higgs boson because they could provide hints about the validity of these theories.

The process studied in this thesis is the decay of Higgs boson to particles that are not detected in experiments, classified as invisible particles. The study of this process is motivated by the fact that several beyond the SM theories predict that the Higgs bosson couples to dark matter particles. If this was the case, the branching ratio of invisible Higgs boson decays would be higher than the one predicted by the SM and upper bounds on this ratio could provide information about the properties of dark matter.

A common strategy to separate invisible Higgs decays from its background is to use a boosted decision tree. However, the inputs of this method are traditional kinematic variables and so, it relies on jet algorithms and there is no jet calibration below 17 GeV. A powerful alternative to overcome this limitation and try to enhance Higgs bosons decaying invisibly is to use an autoencoder since its inputs are images of topoclusters and tracks.

This thesis is structured in the following way. The first chapter outlines the theory of the SM and physics beyond the SM. The second one provides an introduction to machine learning and the Toolkit for Multivariate Data Analysis. Next, the ATLAS experiment is described in chapter three. Chapter four presents the reconstruction of the objects used in this thesis. Finally, chapter five and six present the results of the separation between a Higgs boson produced via vector-boson fusion decaying invisibly and its main background, a Z boson decaying into two neutrinos, using boosted decision trees and an autoencoder respectively in each chapter.

1. Theoretical Background

This chapter contains two sections. Section 1.1 gives an overview of the particle content of the SM and the interactions that it describes. This overview is adapted from references [1, 2, 3]. Section 2.2 discusses one of the phenomenons that the SM can not explain: dark matter.

1.1 The Standard Model of Particle Physics

The SM is a gauge theory determined by the local $SU(3)_C$ x $SU(2)_L$ x $U(1)_Y$ symmetry. It combines special relativity and quantum mechanics in a quantum field theory which was developed in the 60s and early 70s. Among its many achievements are the prediction of the weak neutral current and the prediction of particles, like the charm quark or the tau neutrino, before they were experimentally observed.

The elementary particles of the SM are shown in Figure 1.1. The SM contains twelve spin $\frac{1}{2}$ fermions and their corresponding antiparticles. They are classified into quarks and leptons, according to whether they interact via the strong force. Moreover, they are separated into three generations which differ by their mass and their flavour quantum number.

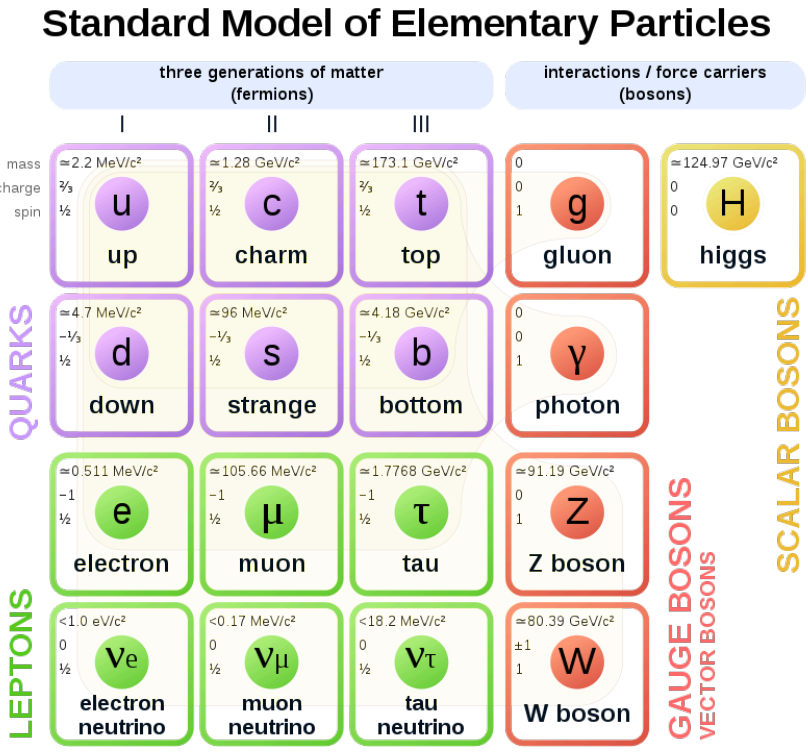


Figure 1.1: Particle content of the SM [4].

The quarks carry a color charge and therefore take part in the strong interaction. There are six different quark flavours that are divided into three up-type quarks and three down-type quarks, according to their electromagnetic charge. The former (the up (u), charm (c) and top (t) quark) have a charge $q = \frac{2}{3}e$ and the latter (down (d),

strange (s) and bottom (b) quark) have $q = -\frac{1}{3}e$. Therefore they interact via the electromagnetic force. In addition, all quarks interact through the weak force.

The leptons are divided into three charged leptons: electron (e), muon (μ) and tau (τ), and three electrically neutral neutrinos: electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). All leptons interact via the weak force and they do not take part in the strong force. Also the charged leptons can interact via electromagnetic force.

The elementary fermions interact with each other via the exchange of spin 1 bosons called gauge bosons. These bosons are one massless photon (γ), three massive bosons (W^+, W^-, Z) and eight massless gluons (g). They mediate the electromagnetic, weak and strong forces, respectively. Furthermore, some gauge bosons are able to self-interact, as depicted in Figure 1.2. One example is the gluons which self-interact because they are also color charged.

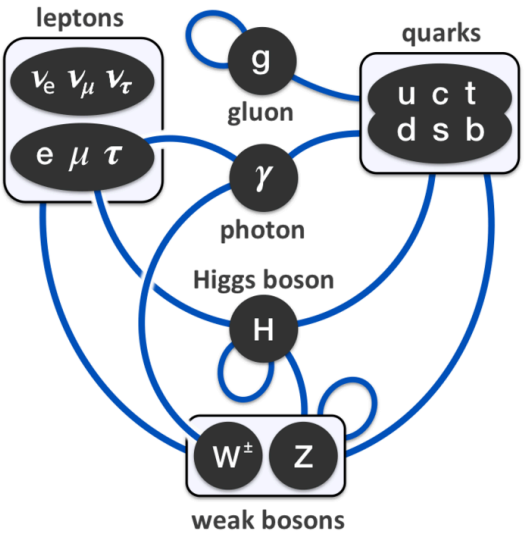


Figure 1.2: Diagram that shows the interactions described by the SM. The blue lines represent not only the interactions between the fermions and bosons, but also the boson self-interaction [4].

The last component of the SM is the Higgs boson. It is a spin 0 boson which has neither electric nor color charge. It was discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) at CERN [5, 6]. A more detailed description of its theoretical prediction and properties is given in Sections 1.1.2 and 1.1.3, correspondingly.

1.1.1 Electroweak Interaction

The electromagnetic and weak interactions were unified in the Glashow-Weinberg-Salam (GWS) theory [7, 8, 9]. The underlying symmetry of this theory is $SU(2)_L$ x $U(1)_Y$ which defines the properties of the interaction. The label L indicates that the gauge fields W^a_ν, with $a = 1, 2, 3$, of $SU(2)_L$ only couple to left-handed particles.

The GWS model replaces the electromagnetic $U(1)_{EM}$ group by $U(1)_Y$ where the label Y is the hypercharge quantum number. For its lagrangian density to be invariant under the electroweak symmetry, the derivative ∂_μ is replaced by the covariant

derivative D_μ

$$D_\mu = \partial_\mu - ig\frac{\sigma^a}{2}W_\mu^a + ig'\frac{\sigma^3}{2}B_\mu, \tag{1.1}$$

where σ represents the pauli matrices that are the generators of $SU(2)_L$, g is the $SU(2)_L$ coupling and g' is the coupling of $U(1)_Y$.

The photon field A_ν and the Z boson field Z_ν are identified as a mixture of the gauge fields B_ν of $U(1)_Y$ and W_ν^3 through

$$A_\nu = cos\,\theta_W\, B_\nu + sin\,\theta_W\, W_\nu^3, \tag{1.2}$$
$$Z_\nu = -sin\,\theta_W\, B_\nu + cos\,\theta_W\, W_\nu^3, \tag{1.3}$$

where θ_W is the weak mixing angle. This mixing arises naturally in the Higgs mechanism explained in the next section.

The physical $W_\mu^\pm$ bosons are linear combinations of W_ν^1 and W_ν^2

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right). \tag{1.4}$$

Since the $W^\pm$ bosons are uniquely formed from the gauge fields of $SU(2)_L$, only left-handed particles take part in the weak charged interactions. Conversely, the Z boson couples to left and right handed particles, but with different strength.

In addition, the GWS model relates the $SU(2)_L$ coupling g', the $U(1)_Y$ coupling g and the electromagnetic coupling e to

$$g'\,cos\,\theta_W = g\,sin\,\theta_W = e. \tag{1.5}$$

1.1.2 Higgs Mechanism

One of the questions that arose when constructing the SM was how fundamental particles obtain their mass without breaking the required gauge symmetry of the theory. Mass terms, such as $-m\bar{\Psi}\Psi$ for fermions or $-\frac{1}{2}m^2A_\nu A^\nu$ for gauge bosons, are not invariant under the electroweak symmetry $SU(2)_L$ x $U(1)_Y$. As a consequence, the unitarity of the SM is lost at the TeV scale if they are included in the lagrangian density.

This difficulty was solved by the Higgs mechanism where the particles acquire mass through the interaction with the Higgs field, a $SU(2)_L$ doublet of complex scalar fields $\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$. The lagrangian density of the Higgs field is

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\Phi)^\dagger\,(D^\mu\Phi) - V(\Phi), \tag{1.6}$$

where $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda\left(\Phi^\dagger\Phi\right)^2$ and D_μ is the covariant derivative defined in Equation 1.1.

For $\mu^2 < 0$ and $\lambda > 0$, the Higgs potential has the well known mexican hat shape. The minima of the potential are at $|\Phi| \neq 0$ and the Higgs field acquires a non zero vacuum expectation value $|\langle\Phi\rangle| = \frac{v}{\sqrt{2}}$, with $v = 246$ GeV. The choice of a vacuum state spontaneously breaks the electroweak symmetry, $SU(2)_L$ x $U(1)_Y \rightarrow U(1)_{EM}$. This process is called electroweak symmetry breaking.

If $\mathcal{L}_{\text{Higgs}}$ is expanded in perturbation theory around the vacuum state, it describes one massive gauge boson and three massless scalar particles known as Goldstone bosons. However, using the gauge freedom to choose the unitary gauge, in which the fields that appear in the lagrangian density correspond to the physical particles, the Higgs field acquires this form

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{1.7}$$

where H is the Higgs boson and the three Goldestone bosons have disappeared.

Regarding the mass of the gauge bosons, $\mathcal{L}_{Higgs}$ contains the following quadratic terms in the gauge fields after its expansion around the vacuum state in the unitary gauge

$$\mathcal{L}_{Higgs} \supset v^2 \frac{g^2 + g'^2}{4} Z^\mu Z_\mu + v^2 \frac{g^2}{4} W^{+\mu} W^-_\mu, \tag{1.8}$$

where A_μ, $W^\pm_\mu$, Z_μ are the mass eigenstates which correspond to the photon, the $W^\pm$ and Z bosons.

Their corresponding masses are:

$$m_W = v\frac{g}{2}, \qquad m_Z = v\frac{\sqrt{g^2 + g'^2}}{2}, \qquad m_A = 0.$$

The three Goldstone bosons have been "eaten" by three of the four electroweak gauge bosons adding a longitudinal polarization which thus makes the electroweak gauge bosons massive. This way of giving mass to the gauge bosons through spontaneous symmetry breaking is the Higgs mechanism.

Similarly, it is possible to construct mass terms for the fermions using the Higgs field. These are called Yukawa terms

$$\mathcal{L}_{Yukawa} = -y_f \left(\overline{\Psi}_L \Phi \Psi_R + \overline{\Psi}_R \Phi^\dagger \Psi_L \right), \tag{1.9}$$

where $\Psi_{L/R}$ are the the left and right handed fermion spinors and y_f is the Yukawa coupling for the fermion f.

After expanding $\mathcal{L}_{Yukawa}$ around the vacuum state, the masses of the fermions are given by

$$m_f = \frac{y_f v}{\sqrt{2}}. \tag{1.10}$$

1.1.3 The Higgs Boson

The Higgs boson can be produced at the LHC by the four main mechanisms depicted in Figure 1.3. The cross section of each of them depends on the mass of the Higgs boson, which is $m_H = 125.38 \pm 0.14$ GeV [10]. For this value of m_H, gluon-gluon fusion is the main production mechanism followed by vector-boson fusion.

Regarding its decay channels, since the coupling of the Higgs boson to a fundamental particle is proportional to the mass of the particle, it decays preferably to the heaviest particles which are kinematically allowed. Therefore, it does not interact directly with the massless photons and gluons. However, it can decay to massless particles through loops of virtual heavy fermions (mostly top quarks) and W bosons, similarly to gluon-guon fusion. The main decay modes are listed in Figure 1.4.

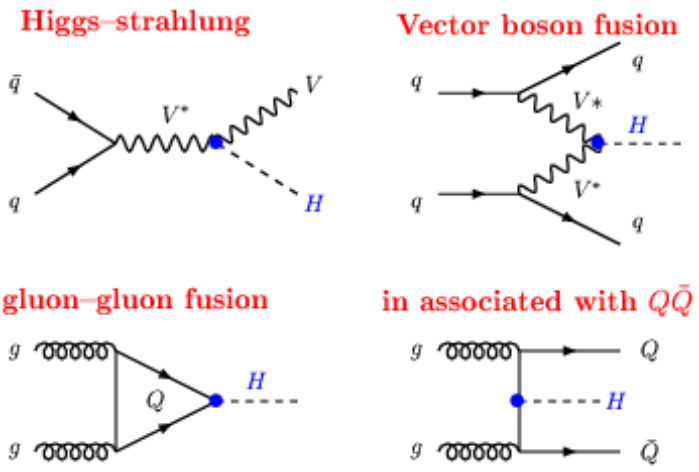


Figure 1.3: Feynman diagrams of the main production mechanisms for SM Higgs bosons at the LHC [11, p.3].

The dominant decay channel is a Higgs boson decaying to a pair of b-quarks($b\bar{b}$). Nevertheless, the huge QCD background ($gg \rightarrow b\bar{b}$) at the LHC made it extremely difficult to observe this decay channel. This caused that the discovery channels were those with smaller branching ratios but with clearer signals. They were a Higgs boson decaying to two Z bosons $H \rightarrow ZZ^* \rightarrow l^+l^-l'^+l'^-$, a Higgs boson decaying to two W bosons $H \rightarrow WW^* \rightarrow \bar{l}\nu_l l'\bar{\nu}_{l'}$ and a Higgs boson decaying to two photons $H \rightarrow \gamma\gamma$, where the asterisk (*) indicates that it is a virtual particle since $m_H < 2m_Z$ and $m_H < 2m_W$.

Decay mode	Branching ratio
$H \rightarrow b\bar{b}$	57.8%
$H \rightarrow WW^*$	21.6%
$H \rightarrow \tau^+\tau^-$	6.4%
$H \rightarrow gg$	8.6%
$H \rightarrow c\bar{c}$	2.9%
$H \rightarrow ZZ^*$	2.7%
$H \rightarrow \gamma\gamma$	0.2%

Figure 1.4: The main branching ratios of the Higgs boson decay predicted by the SM, for $m_H = 125$ GeV [1, p.489].

Until now, the production mechanisms shown in Figure 1.3 and the following decays $H \rightarrow$ ZZ* [13], WW* [14], $\gamma\gamma$ [15], $\tau^+\tau^-$ [16] and $b\bar{b}$ [17] have been observed (see Figure 1.5). The last two decays were observed in 2018 and proof that the Higgs boson couples to leptons and to quarks. So far, all these measurements are consistent with the SM predictions within uncertainties.

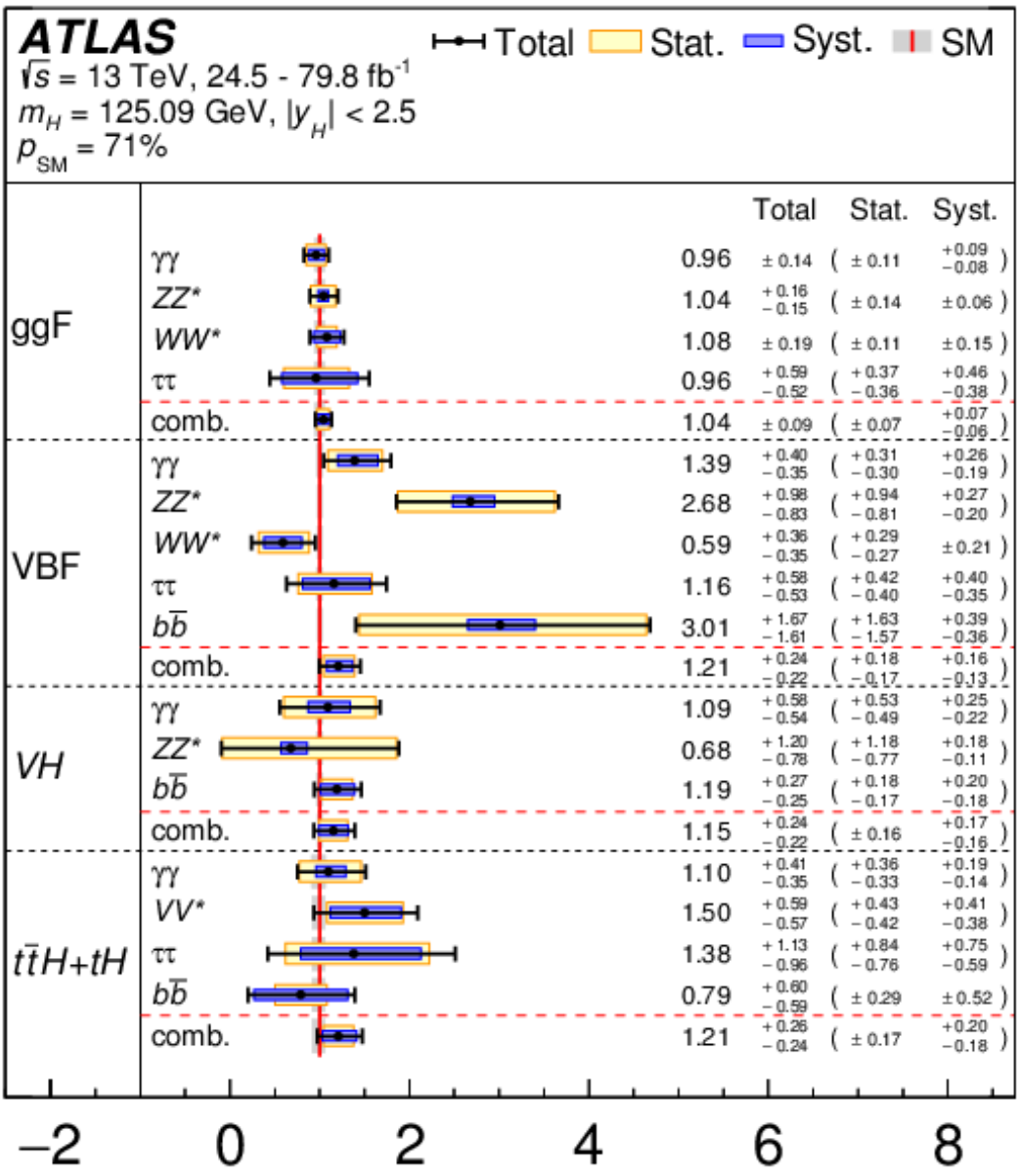


Figure 1.5: Measured cross section times branching ratio for the relevant production mechanisms at the LHC in the main decays channels, normalized to the SM predictions. The black bars and the blue, yellow and grey bands correspond to the total, systematic, statistical and theory uncertainties, correspondingly. [12, p.13].

1.1.4 Quantum Chromodynamics

The strong interaction, described by the Quantum chromodynamics (QCD) theory, is mediated by gluons which couple to colour-charged particles. There are three colour charges and three colour anticharges. Since the gluons carry one colour charge and one anticolour charge each, they are able to self-interact. This feature arises from the non-abelian nature of the underlying $SU(3)$ symmetry of the theory, that is, because the generators of the $SU(3)$ group do not commute.

The $SU(3)$ symmetry requires eight gluon fields G^a_μ and the following covariant

derivative

$$D_\mu = \partial_\mu + i g_S \frac{\lambda^a}{2} G^a_\mu, \tag{1.11}$$

where g_s is the strong coupling constant and λ^a are the Gell-Mann matrices that are the generators of $SU(3)$.

The properties of QCD cause that quarks have not been observed isolated, but in colourless bound states called hadrons. This fact is explained by the colour confinement hypothesis. It states that only colourless objects can propagate freely. This is explained by the fact that the strong force between two particles increases linearly with distance, due to the gluon self-interactions. If two colour charged objects were to be separated by an increasing distance, the energy stored in the strong field between them would also increase. At some point, it is energetically preferable to form new quark-antiquark pairs, thus forming color neutral objects. Hadrons can be further divided into mesons, a quark-antiquark pair, and baryons that are formed by three quarks.

As a consequence of confinement, high-energy quarks and gluons give rise to collimated groups of hadrons known as jets, which are observed at high-energy experiments. The formation of jets occurs through a process called hadronisation. A qualitative example is depicted in Figure 1.6. First, a quark and an antiquark start to separate, interchanging virtual gluons. Since gluons interact with each other, the gluon field is squeezed into a tube. The energy stored in the field increases with distance. Therefore, at some point, the energy is large enough to create a quark and an antiquark pair out of the vacuum, breaking the gluon field. This process is repeated until the energy is low enough that the quarks and antiquaks combine into hadrons. Finally, two jets are formed following the directions of the initial quark and antiquark. Furthermore, the jets contain other particles, like photons or leptons, coming from the decays of some of the hadrons within the jet.

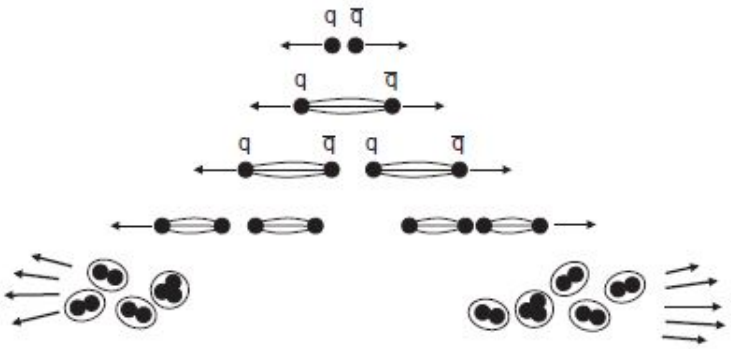


Figure 1.6: Hadronisation of a quark and an antiquark. As a quark and antiquark flight apart, the energy stored in the gluon field increases. At some point, the energy is large enough and a quark and an antiquark pair is created out of the vacuum, breaking the field. The process continues until the energy is low enough that the quarks and antiquaks form hadrons. Eventually, two jets are formed following the directions of the initial quark and antiquark. [1, p. 253].

Another feature of the strong interaction is that its strength decreases as the energy of the interaction increases. This fact is known as asymptotic freedom and occurs due to the gluon self-interactions. At energies around 1 GeV and below, $\alpha_S \, \alpha \, g_s^2$ is of O(1). and perturbation theory can not be used. At larger energies, the value of α_S is low

enough to use perturbation theory and quarks asymptotically behave as free particles. The running of α_S is related to the concept of renormalization which is beyond the scope of this thesis.

1.2 Physics Beyond the Standard Model

A wide range of cosmological and astrophysical observations [18, 19] suggest the existence of vast amounts of massive matter which does not interact via the electromagnetic force. Nevertheless, the SM does not provide an explanation for this special kind of matter. Due to the fact that it does not interact electromagnetically, it is generally referred to as dark matter.

1.2.1 Dark Matter

The most direct evidence for the existence of dark matter on galactic scales is the velocity distributions of stars in spiral galaxies. In these galaxies, the central bulge contains most of the luminous mass. Outside this region, the tangential velocity of a star orbiting the center is

$$v \approx \sqrt{\frac{GM(r)}{r}}, \tag{1.12}$$

where G is the gravitational constant, $M(r)$ is the mass inside a radius r, where it has been assumed that the majority of the mass is located in the bulge. Thus, the velocity of stars should fall as $r^{-1/2}$. However, this theoretical prediction does not agree with the observed velocity distributions, such as the one in Figure 1.7. These observations imply that the mass distribution is given by $M(r) \, \alpha \, r$. This indicates that there is a large non-luminous component which contributes to the mass of the galaxies.

Another evidence comes from the measurements of the fluctuations of the cosmic microwave background (CMB) performed by the PLANCK collaboration. As the universe expanded and cooled, particle species decoupled from the primordial plasma, falling out of thermal equilibrium. The last ones to decouple were electrons, protons and photons. When the electrons and protons combined to form neutral atoms, the photons could travel freely through the universe for the first time. These primordial photons carry the image of the plasma at that moment and constitute the CMB. The CMB is mostly isotropic with small anisotropies which are related to matter density fluctuations. The fluctuations of the barionic matter density alone can not explain these anisotropies. Additional fluctuations of a dark matter density are needed to account for them. According to the PLANCK data, the barionic matter density is $\Omega_b h^2 = 0.0224 \pm 0.0001$ and the dark matter density is $\Omega_c h^2 = 0.120 \pm 0.001$ [21].

During the expansion and cooling of the universe, the light particles, such as neutrinos, remain relativistic (hot) while heavier ones become non-relativistic (cold). They have different effects on the formation of large structures in the universe. The hot particles obstruct the clumping of matter while cold ones clump on small scales. In the case of dark matter, large structure formation indicates that at least the majority of dark matter is cold [22]. For the cold scenario, there are several candidates like axions [23] or weakly interacting massive particles (WIMPs) [24]. The latter emerge naturally in extensions of the SM like supersymmetry. In many supersymetric models,

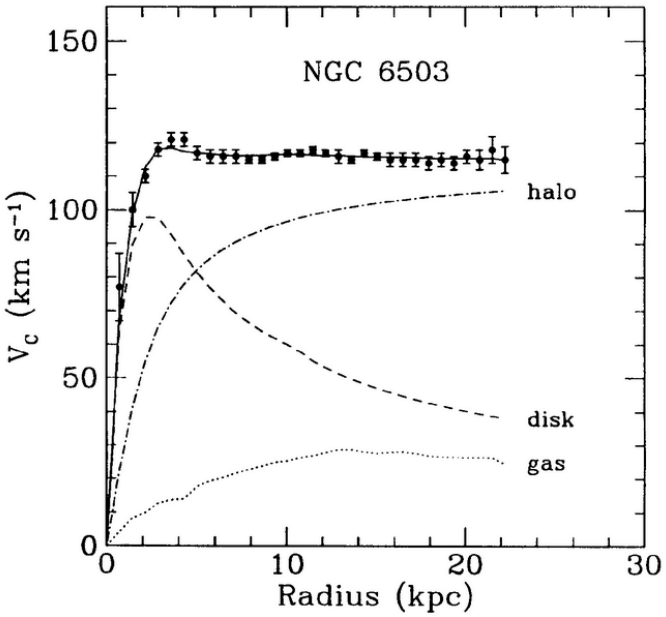


Figure 1.7: Rotation curve of the galaxy NGC 6503. The dashed and dotted curves show the Newtonian rotation curves of the visible disk and the gaseous component. The line depicted by "halo" corresponds to the dark matter component needed to mach the observed data, represented by the points with error bars [20, p. 3].

the lightest particle is a stable, neutral, weakly interacting particle with a mass in the GeV-TeV range.

Currently, there are three main strategies to detect dark matter. These are: searching dark matter scattering off nuclei (direct detection), looking for the decay products of dark matter annihilation in galactic centers (indirect detection) or producing it in SM particle collisions. Here, the third one is discussed.

Since the Higgs boson coupling strength increases with respect to the particle participating in the interaction, the Higgs boson is thought to act as a portal between the SM and dark matter, thus also allowing for Higgs boson decays into dark matter [25]. Dark matter particles traverse the ATLAS detector at the LHC without being directly detected. If the Higgs boson would couple to dark matter, the branching ratio of invisible Higgs boson decays would be higher than the one predicted by the Standard Model $B_{SM}(H \to ZZ^* \to 4\nu) = 1.06 \cdot 10^{-3}$ [26]. Upper bounds on this ratio could give hints about the properties of dark matter. The current most stringent upper limit on this branching ratio is 0.13 at the 95% confidence level (CL) [27].

Although dark matter does not interact with the ATLAS detector, its presence can be inferred from the momentum and energy conservation of the collision products. Before the collision, the initial particles carry no momentum in the plane transverse to the beam direction. Therefore, the sum of the transverse momenta p_T of the final state particles must be zero. If invisible particles are produced, there is a net momentum in the transverse plane. The momentum needed to compensate this net momentum is called the missing transverse momentum. Here, it is referred to missing transverse energy $\vec{E}_T^{miss}$ with absolute value E_T^{miss}. Through this, the missing transverse energy can be related to the existence of dark matter.

2. Machine Learning and Multivariate Data Analysis

Machine learning has become a crucial tool in high energy physics analysis in order to extract the maximum possible information from the data available. In this chapter, an overview to machine learning is presented in Section 2.1. Section 2.2 explains the framework and algorithms used for the multivariate analysis described in Chapter five.

2.1 Machine Learning

In machine learning, a common task is to classify data into two classes, separating out the classes by drawing a decision boundary. For this task, the machine learning algorithm learns how to assign a class label to the data points. To tackle this problem, the data set is usually split into two parts: *training set* and *test set*, which must follow the same distribution. The model is first trained on the training set. The goal here is to find a set of weights that minimize the *loss function* (defined below) using an optimization technique such as *gradient descent* or *stochastic gradient descent*. With the inferred weights, the accuracy of the fitted model is then evaluated using the test set.

If instead of capturing only the underlying pattern, the model also captures the random noise in the training data, this is called *overfitting*. On the other hand, if our trained model fails to capture the relevant relations between the input and the output features, this is named *underfitting*. Neither is satisfactory and a balance between the two is important.

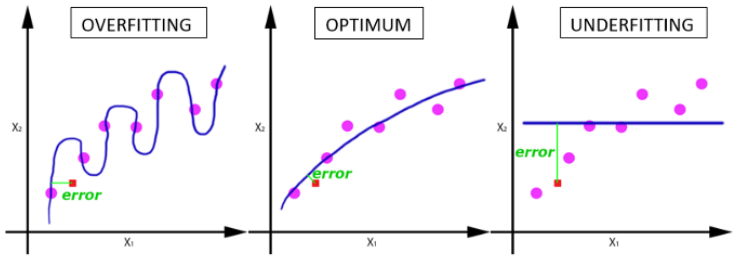


Figure 2.1: Representation of a model fitting too closely to the data points (*overfitting*), fitting correctly (*optimum*) and fitting the data points unsuccessfully (*underfitting*) [28].

Another important aspect of machine learning is that it is usually impossible to pass the entire data set directly, since machine learning algorithms need a lot of data. Therefore, the data set must be divided in a number of *batches*. Then, the *batch size* is defined as the number of training samples that there is in a single batch. Also, every time that the entire data set is passed through the algorithm, it is said that one *epoch* has elapsed.

Loss Functions

Loss functions are used to measure how different an estimated output value is from its true value. The most common loss functions are *mean squared error* (*mse*), *mean*

absolute error and *binary crossentropy*. The former is defined as

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{2.1}$$

where n are the number of pixels, y_i are the observed values and $\hat{y}_i$ are the predicted ones.

2.1.1 Neural Networks

A *neural network* is a net of layers of fully-connected nodes that looks like Figure 2.2. It consists of an input layer, one or more hidden layers and an output layer, where the information travels from left to right. The outputs of the nodes of a given layer, which become the input of the nodes of the next layer, are first weighted, summed and passed through a non-linear activation function (Figure 2.3). Therefore, training a neural network means learning the weights associated with each edge.

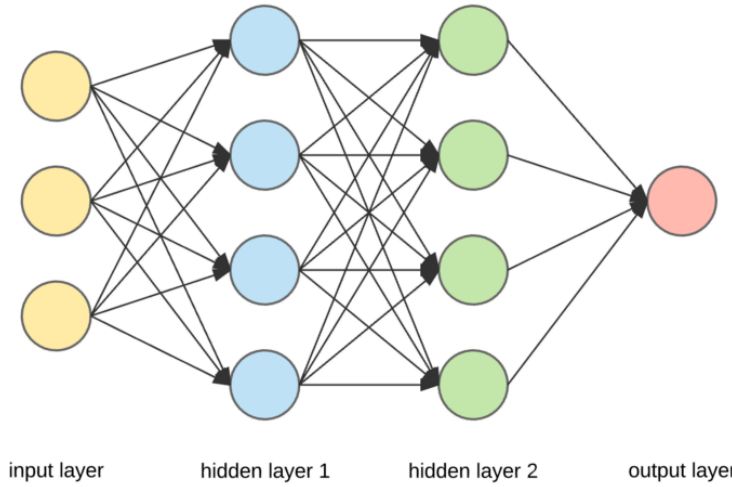


Figure 2.2: Representation of the input of each node [29].

It is also relevant to add that different neural networks are used depending on the type of input data. As an example, convolutional neural networks are used for image data, whereas for test series data, a recurrent neural network is the more common choice.

Activation Functions

As explained in the above neural network section, the outputs from the nodes of the previous layer (x_i) are multiplied by weights (w_i) and then summed. Mathematically, $y = \sum_i w_i x_i$. Note that this is a linear function. If the output of a node is just a linear combination of the outputs of other nodes, then no matter how deep the neural network is, the final output will be just a linear combination of the inputs and, thus, the model will have limited capacity. Because of this limitation, the linear combination of the outputs from nodes of the previous layer is also run through a non-linear *activation function*. The commonly-used activation functions are *ReLU* (*Rectified Linear Unit*), *Sigmoid*, *Linear* and *Tanh*. All these functions are depicted in Figure 2.4.

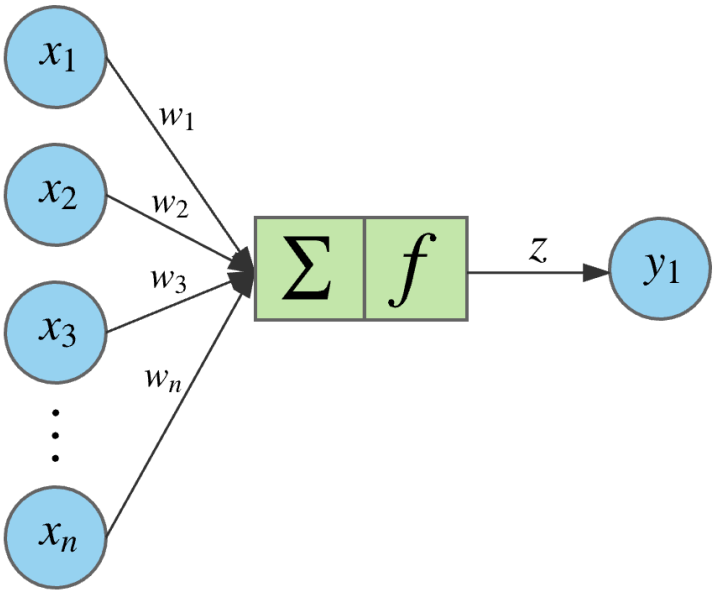


Figure 2.3: Visual representation of how the outputs of a layer are weighted, summed and passed through an activation function before becoming the input of the nodes of the next layer [29].

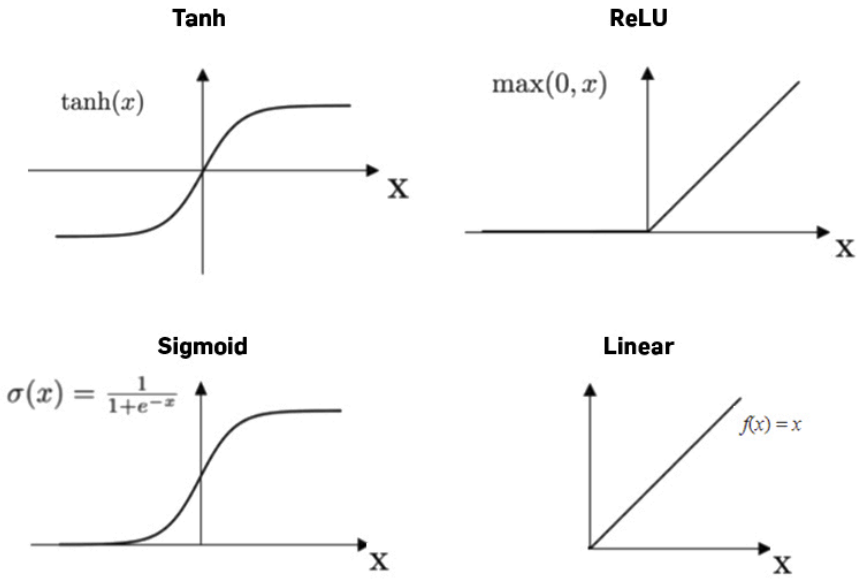


Figure 2.4: Most common activation functions used for neural networks [30].

Convolutional Neural Networks

Convolutional neural networks are used to detect the relevant features of images through the realization of series of convolution and pooling operations. The convolution operation merges two sets of information, using a *convolution filter* and producing a *feature map*. To do the operation the filter is slid over the input image. At each location, an element-wise multiplication is performed and the result is summed. The resulting number is then aggregated to the feature map, as depicted in Figure 2.5.

At each layer of the neural network, multiple convolutions are performed independently on an input, using different filters, resulting in different feature maps. These are stacked along the depth dimension yielding a final output with the same height and width as the input, but deeper (Figure 2.6). This output is then run through a *ReLU* activation function, so that the result is not a linear combination of the input.

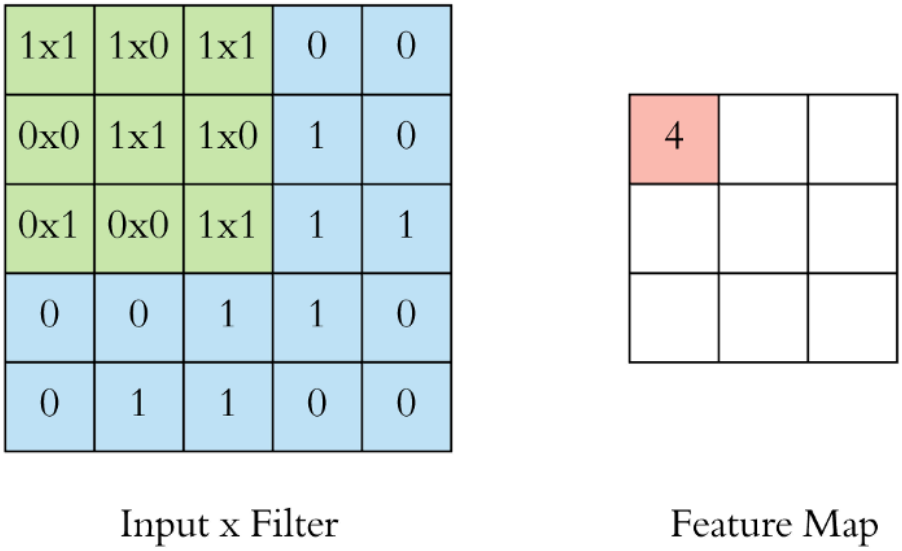


Figure 2.5: On the left, there is the convolution filter (green) and the input image (blue). The filter is placed at every location of the image. At each location, the numbers are multiplied element-wise and the result is summed up. The resulting scalar is then collected in the feature map, shown on the right [32].

After the multiple convolutions, a pooling operation is usually performed. It downsamples the feature maps while keeping the important information. This is achieved by reducing the maps height and width while keeping their depth intact. This reduction results in a shortening of the training time and the chances of overfitting are reduced. The most usual type of pooling is *max pooling* which slides a window over the input and takes the maximum value in the pooling window (Figure 2.7).

2.1.2 Autoencoder

An *autoencoder* is a type of neural network which learns how to compress and then reconstruct the data back from its reduced representation named code. It consists of three parts shown in Figure 2.8. The *encoder*, which is a fully-connected neural network, compresses the input and creates the *code*. The code that is a single layer of a neural network. Finally, the *decoder*, which usually has the reverse architecture than the encoder, gets back the input but using only the code.

Before training the autoencoder, however, there is a series of hyperparameters that must been set. These are: the code size (number of nodes in the code layer), the number of layers of the neural network, the number of nodes per layer and the loss function, which is used to compare the output with the input.

Anomaly Detection

Although autoencoders are mostly used for image recognition, they can be also

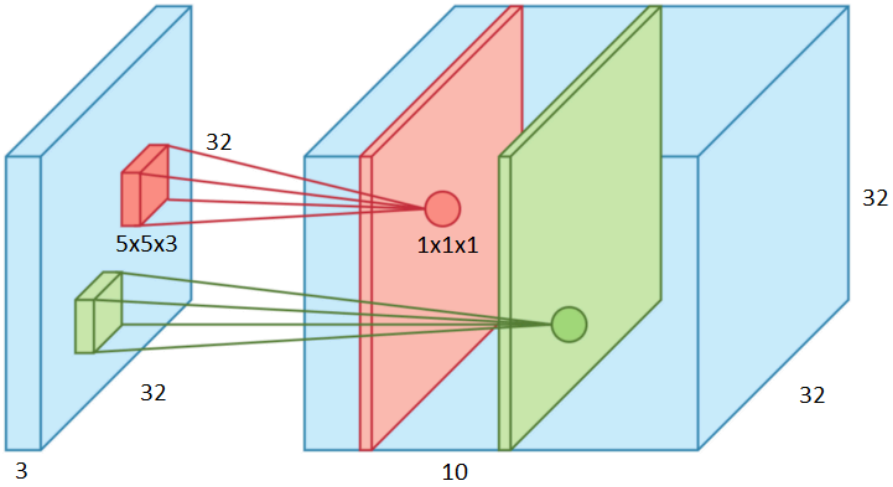


Figure 2.6: Example of 10 convolution operations applied to a 32x32x3 image, using 5x5x3 filters. Each filter leads to a 32x32x1 feature map (red and green slices on the right). These are stacked along the depth dimension resulting in the 32x32x10 output represented by the right blue box [32].

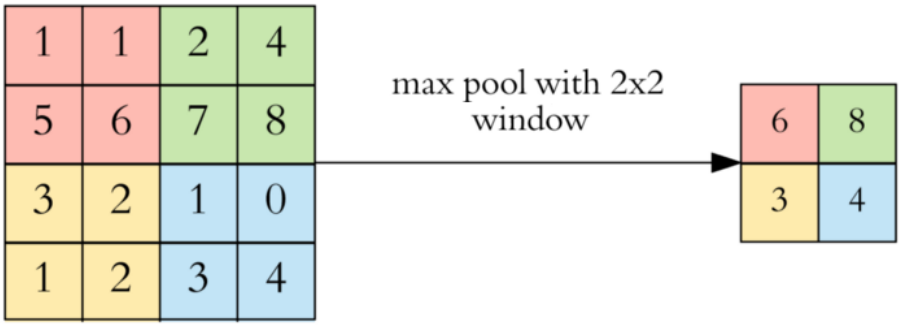


Figure 2.7: Representation of the max pooling operation, where the input image is reduced by sliding a 2x2 window and taking the maximum value of the window [32].

trained to find anomalies in the data set following this procedure. First, the *mse* of each training sample is calculated. Then, the maximum *mse* is set as the threshold for anomaly detection. Eventually, if a test sample has a $mse >$ threshold, it is labeled as an anomaly.

2.2 Toolkit for Multivariate Data Analysis

The Toolkit for Multivariate Data Analysis (TMVA) [33] is a ROOT-integrated analysis framework which allows the processing, performance evaluation and application of several multivariate classification and regression methods, which are based on machine learning techniques. The analysis carried out by the TMVA is separated into a training phase and an application one. In the first part, one or several methods are booked, the input variables are preprocessed and preanalysed, and the chosen methods are trained, tested and evaluated. In the last one, the selected methods are used to solve the classification or regression problem of an unknown data set. The multivariate classification methods used for this analysis are described below.

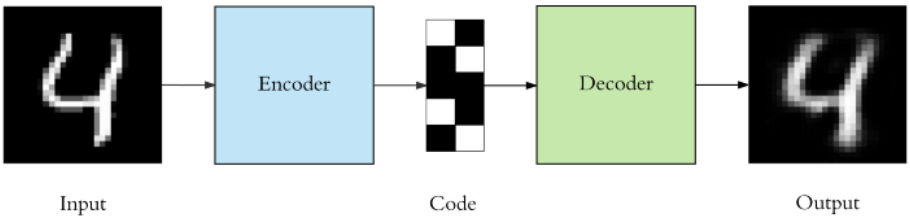


Figure 2.8: Sketch of how an autoencoder functions [31].

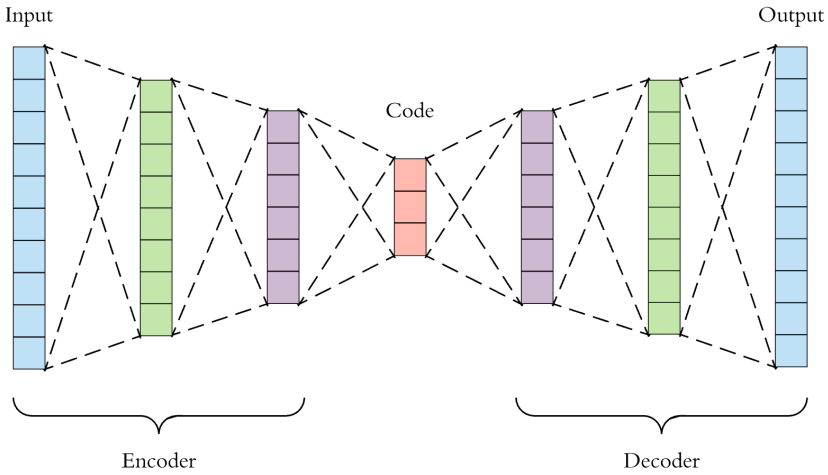


Figure 2.9: Visualization of the structure of an autoencoder [31].

2.2.1 Rectangular Cut Classifier

The rectangular cut classifier is the simplest method that categorizes an event as a signal or a background event. It places a rectangular cut on each variable separately, that is, the cut on one variable does not depend on the cuts on the other variables. The optimisation of these cuts maximises the background rejection at given signal efficiency and it is performed for different signal efficiencies. Furthermore, the performance of the classifier is strongly correlated to the shape of the signal and background distributions since this method relies on the assumption that the signal is clustered in the variable space. If this was not the case, the classifier would underperform.

The fitting method of the cut classifier can be selected using the *FitMethod* option. The primary methods are Monte Carlo sampling [34], Genetic Algorithm [35], and Simulated Annealing [36], being Genetic Algorithm the default one. Moreover, if a simple cut is necessary instead of rectangular cuts, due to the distribution shape being already known, the *Varprop* option should be set to *FMax* or *Fmin*. TMVA can also detect which requirement must be removed by using the option FSmart.

2.2.2 Boosted Decision Trees

A decision tree is a tree-like structured algorithm, like the one depicted in Figure 2.10, which separates the data at each node into two subnodes, establishing a decision requirement. Each requirement depends on one discrimination variable and the data is split over and over until a stop criterion is satisfied. The nodes at the bottom of the tree, called *leaves*, are labeled as signal or background, depending on the majority of events that end up in that node. Finally, the depth of the tree, as well as the minimum

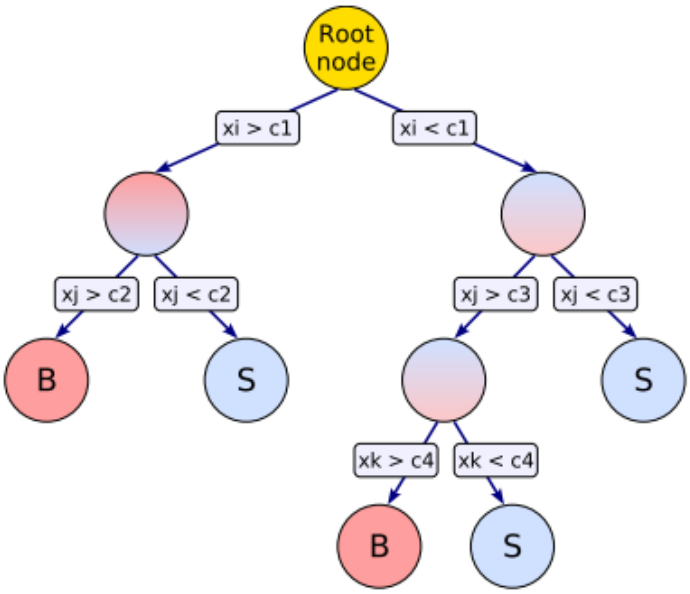


Figure 2.10: Sketch of a decision tree. Beginning in the root node, the data is binary split, using the discriminating variables, until some criteria is met. The leaves (the circles labeled as "B" or "S") are labeled as signal or background depending on the majority of events that end up in that node [33, p.115].

percentage of training events required in each leaf node, can be fixed respectively using the *MaxDepth* and *MinNodeSize* options.

The main advantage of a decision tree is its interpretability. This helps revealing the interactions between different variables and their relative importance. However, a decision tree is vulnerable to overfitting. In order to avoid this limitation, a technique called *boosting* is included. From that moment on, the tree is referred to as a *boosting decision tree*. It fits many trees to reweighted versions of the training set. The total number of trees is set by *NTrees* and the boosting type by *BoostType*.

3. The ATLAS Experiment

The ATLAS experiment is one of the four experiments at the LHC at CERN. It is designed to make precise measurements of the SM parameters and to search for evidences of physics beyond the SM. In this chapter, the main characteristics of the LHC and the principal components of the ATLAS detector are described in Sections 3.1. and 3.2 correspondingly.

3.1 The Large Hadron Collider

The LHC [37] is the largest and most powerful collider that has been built. It has a circumference of 27 km and is situated at the Large Electron–Positron Collider tunnel at CERN. It is mainly characterized by two parameters: the center-of-mass energy $\sqrt{s}$ and the instantaneous luminosity L. The first parameter determines which particles can be produced, whereas the second one regulates the event rate. The number of events of a specific process is related to the luminosity by

$$N = \sigma \int L(t)\, dt, \tag{3.1}$$

where σ is the production cross section, which is a measurement of the quantum probability for an event to occur. L is integrated over the time that the collider is functioning and it itself depends on the colliding beam parameters

$$L = \frac{N_1 N_2 f}{A}, \tag{3.2}$$

where N_1 (N_2) are the number of particles in the first (second) beam, f is the revolving frequency and A is the effective cross section area.

The LHC collides primarily protons, but also heavy ions. During the first two data taking periods, Run 1 (2009-2013) and Run 2 (2015-2018), proton-proton collisions with center-of-mass energies up to 8 TeV and 13 TeV respectively were achieved. However, it is expected to reach the designed center-of-mass energy of 14 TeV for proton-proton collisions during Run 3. Regarding the instantaneous luminosity, the peak luminosity supplied during Run 2 was $2.0 \cdot 10^{34}$ cm^{-2} s^{-1} [38], exceeding the LHC designed luminosity $L = 10^{34}$ cm^{-2} s^{-1}.

The protons for the LHC are extracted from the ionization of hydrogen gas. Before being injected in the LHC, they are pre-accelerated by the linear accelerator LINAC 2 and the three synchrotron accelerators BOOSTER, Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). Once injected, they are further accelerated by radio frequency cavities with a field strength of 2 MV/m to the final beam energy. The beams are constituted by up to 2808 bunches which contain up to 10^{11} protons. The proton beams are guided along the circular trajectory of the LHC by 1232 superconducting dipole magnets. They produce a magnetic field up to 8.33 T. To maintain their coils' niobium-titanium (NbTi) wires in a superconductive state, 120 tonnes of helium are used to cool the magnets to 1.9 °K [39]. In addition to the dipole magnets, quadrupole magnets focus the beams.

The four main detectors are positioned at the four interacting points. Two of them,

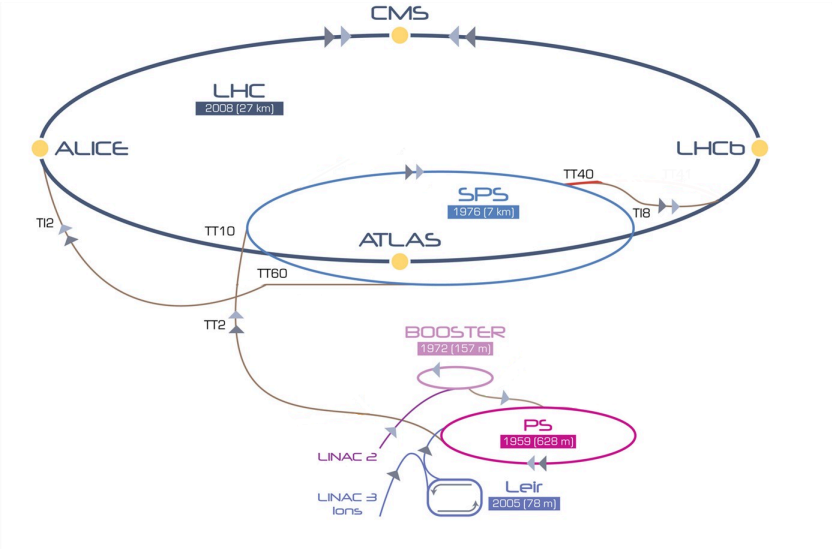


Figure 3.1: Diagram of the acceleration complex at CERN [40].

ATLAS[1] [41] and CMS[2] [42], are general purpose detectors. The LHCb[3] [43] detector was built to make precision measurements of rare decays of B hadrons and CP violation, while the ALICE[4] [44] detector is designed to study heavy ion physics.

One of the challenges that generates the high luminosity of the LHC is that there are multiple interactions per bunch crossing. The additional interactions to the most-energetic hard scatter event are referred as pile-up and constitute a source of background. They can come from the same bunch crossing (in-time pile-up) or from the preceding or following bunch crossing (out-of-time pile-up). The expected pile-up μ is given by

$$\mu = \frac{\mathcal{L}\,\sigma_{inelastic}}{f n_b}, \tag{3.3}$$

where $\sigma_{inelastic}$ is the inelastic cross-section in proton-proton collision, f is the revolving frequency and n_b is the number of bunches. The mean number of interactions per bunch crossing during Run 2 was 33.7 at an integrated luminosity $\mathcal{L} = 147$ fb^{-1} [45]. Higher pile-up is expected as the luminosity is increased.

3.2 The ATLAS Detector

The ATLAS detector is forward-backward symmetric with respect to the interaction point, which is defined as the origin of the coordinate system. The z axis of the system is defined by the beam direction, while the x and y axis point to the center of the LHC and upwards correspondingly. The x-y plane is perpendicular to the beam line and is refereed as the transverse plane. The transverse momentum and energy are defined in this plane as $p_T = |\vec{p}|\ cos\ \theta$ and $E_T = |\vec{E}|\ cos\ \theta$. The azimutal angle ϕ is measured around the z axis and the polar angle θ from the z axis. The later is usually expressed in terms of the pseudorapity $\eta = -ln\ tan\ (\frac{\theta}{2})$. Other important parameters are the

[1] A Toroidal LHC ApparatuS
[2] Compact Muon Solenoid
[3] Large Hadron Collider beauty
[4] A Large Ion Collider Experiment

angular separation between two particles that is quantified by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ and the radial distance from the z axis r.

The structure of the detector is depicted in Figure 3.2. It is formed by several sub-detectors that are concentrically arranged around the interaction point. These are the inner detector, the calorimeters and the muon detector. In addition, a thin superconducting solenoid magnet embeds the inner detector and three large superconducting toroids magnets are placed around the calorimeters.

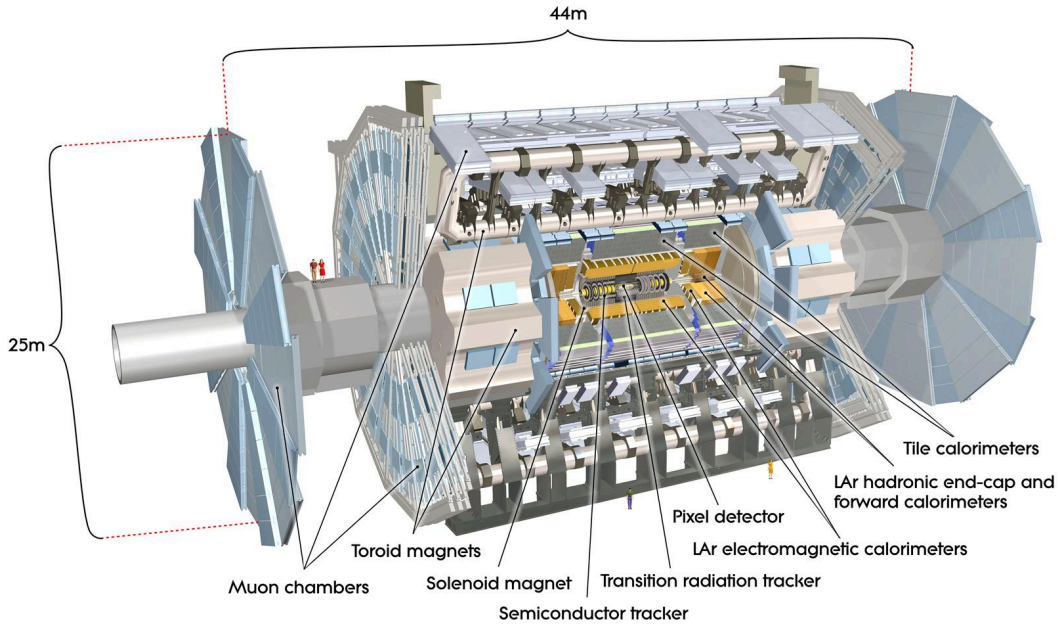


Figure 3.2: Components of the ATLAS detector. Its dimensions are 44 meters in length and 25 meters in diameter [41, p. 4].

3.2.1 Inner Detector

The inner detector (ID) is located the closest to the interaction point. It tracks charged particles in the $|\eta| < 2.5$ region. The tracks, left by the ionization of the ID materials by charged particles, are used for the calculation of their transverse momentum, vertex reconstruction and particle identification. Since the ID is embedded in a 2 T solenoidal magnetic field, the tracks are curved due to the Lorenz force and the transverse momentum can be extracted from the radius of curvature. The accurate measurements of the tracks are possible due to the design of the three subsystems that form the ID: the Pixel Detector (PD), the Silicon Microstrip Tracker (SCT) and the Transition Radiation Tracker (TRT). All three consist of concentric cylindrical layers around the beam axis in the barrel region and disks perpendicular to the beam axis in the end-cap regions.

The PD is situated at the inner part and is made of three layers of silicon pixels in the central region. The majority of the pixels have dimensions of 50 x 400 μm^2 and their intrinsic resolution is 10 μm in r-ϕ[5] and 115 μm in z. The PD is surrounded by the SCT which is composed of four layers of strips in the barrel region with dimensions

[5] r and ϕ are the cylindrical coordinates used in the transverse plane

of 12 cm x 80 μm. The intrinsic resolution of the microstrip sensors is 17 μm in r-ϕ in the central region. For the LHC Run 2, a fourth layer called Insertable B-layer [46] was installed between the beam pipe and the first pixel layer. The size of the IBL's pixels is 50 x 250 μm^2 and they have an intrinsic resolution of 8 μm in r-ϕ and 40 μm in z. This upgrade improved the reconstruction of tracks and vertices, and increased the tracking robustness against higher luminosities.

At the outer part of the ID, there are many layers of drift tubes called straw tubes interleaved with transition radiation material. These 4 mm diameter tubes are filled with a gas mixture of xenon, carbon dioxide and oxygen. They constitute the TRT and are arranged parallel to the beam line in the barrel region and radially in the end-caps. The TRT extends up to $|\eta| < 2.0$ with a resolution of 130 μm in r-ϕ in the barrel region and 130 μm in ϕ-z for the end-caps. Although its resolution is worse than the one achieved in the silicon systems, it improves the momentum resolution in the $|\eta| < 2$ range due to the approximate 36 additional hits per track. The resulting relative momentum resolution of the ID was measured to be $\frac{\sigma_{pT}}{p_T} = (4.83 \pm 0.16) \cdot 10^{-4}$ GeV$^{-1} \cdot p_T$ [47] using data of cosmic rays. Furthermore, the TRT is employed to identify electrons. When a charged particle enters the TRT, transition radiation is realised. Since electrons are much lighter than the other charged particles, their radiation can be distinguished.

3.2.2 Calorimeters

The ATLAS calorimeters cover all ϕ and $|\eta| < 4.9$ range, and are symmetric in ϕ. They are sampling detectors, that is, they are formed by layers of an active medium interspersed with layers of an absorbing high-density material. Incident particles interact with the absorber generating showers of secondary particles which create a signal in the active material. This structure allows to perform energy as well as position measurements of the absorbed particles. The calorimeter system is depicted in Figure 3.3. It consists of the electromagnetic calorimeter (ECAL), that is optimized for measurements of showers produced by electrons and photons, and the hadronic calorimeter (HCAL) for hadronic showers.

The ECAL is located outside the solenoidal magnet. It is constituted by lead plates immersed in liquid argon (LAr) that serves as active medium. The lead plates are interleaved with electrodes arranged in an accordion shape which allows the ECAL to have multiple active layers in depth. Furthermore, the accordion shape provides full coverage in ϕ and fast extraction of the signal.

The ECAL is divided into a LAr electromagnetic barrel, that expands up to $|\eta| < 1.475$, and two electromagnetic end-cap calorimeters (EMEC) in the $1.375 < |\eta| < 3.2$ region. The electromagnetic barrel has three active layers at different radii. The innermost one is finely segmented in η, providing high precision for position measurements. The two EMEC are formed by two coaxial wheels covering the $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$ regions correspondingly. Moreover, a measurement of the energy lost before entering the electromagnetic calorimeters is performed by an instrumented argon layer, called presamplers, that is placed in the $0 < |\eta| < 1.8$ region.

Although it varies with η, the depth of the ECAL is approximately 20 radiation lengths, ensuring that photons and electrons deposit almost all their energy there. Hadrons leave part of their energy in the ECAL, but their hadronic showers extend to the HCAL. The energy resolution of the LAr electromagnetic barrel

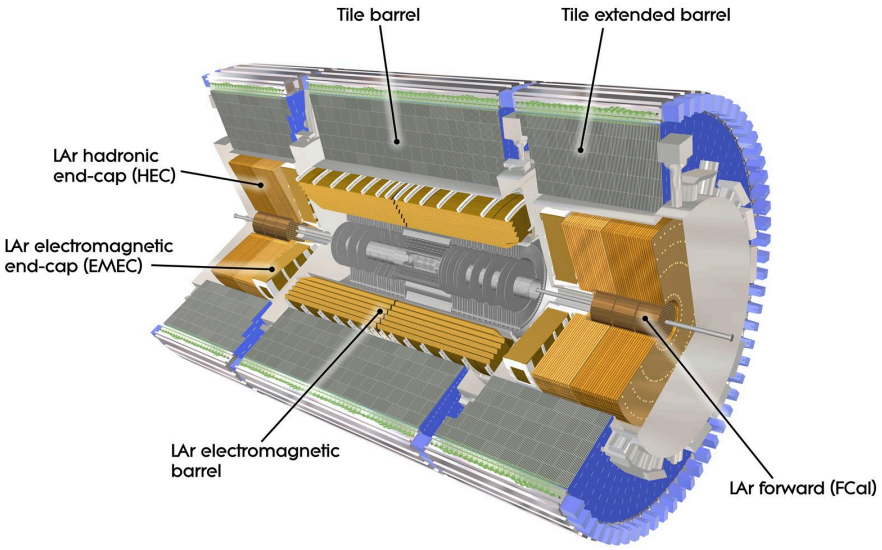


Figure 3.3: The ATLAS calorimeter system [41, p. 8].

$\sigma_E/E = 10.1\%/\sqrt{E(GeV)} \oplus 0.17\%$ [48] was determined using electron beams of known energy.

In the barrel region of the HCAL, there is the tile hadronic calorimeter (TileCal) $0 < |\eta| < 1.7$. It is situated behind the LAr electromagnetic calorimeter and its coarse granularity is used for reconstructing jets and measuring E_T^{miss}. It is segmented into a central barrel and two extended barrels. They rely on scintillator tiles as sampling medium and steel as absorber. The scintillation light from the tiles is read out via wavelength-shifting fibers and transmitted to photo-multiplier tubes.

In each end-cap of the HCAL, there is a Hadronic End-cap Calorimeter (HEC) $1.5 < |\eta| < 3.2$ and a Forward Calorimeter (FCal) $3.1 < |\eta| < 4.9$. The HEC, located behind the EMEC, uses LAr as active medium and copper as absorber. The FCal is placed near the beam line and also relies on LAr as active material. It consists of three modules. The first one is closest to the interaction point. It is designed for electromagnetic showers and is equipped with copper as absorber material. The other two modules use tungsten as absorber and perform measurements of the hadronic showers.

The high radial depth for the TileCal is 7.2 nuclear absorption lengths λ and the HEC and FCal together provide a depth of approximately 10 λ. This configuration ensures that highly energetic jets are contained in the HCAL. The energy resolution for hadrons of the TileCal, HEC and FCal were determined to be $\sigma_E/E = 52.7\%/\sqrt{E(GeV)} \oplus 5.7\%$ [49], $\sigma_E/E = 70.6\%/\sqrt{E(GeV)} \oplus 5.8\%$ [50] and $\sigma_E/E = 94.2\%/\sqrt{E(GeV)} \oplus 7.5\%$ [51] respectively.

3.2.3 Muon Spectrometer

Although muons deposit a fraction of their energy in the ECAL, they exit the calorimetry system without being absorbed. To identify them, the ATLAS detector contains a muon spectrometer located outside the calorimeters. It makes accurate measurements of their trajectory in the $|\eta| < 2.7$ range to determine their momentum. Moreover, it

triggers on muon tracks for $|\eta| < 2.4$.

The muon spectrometer consists of large superconducting toroid magnets and muon chambers. The magnets provide a magnetic field that is mostly perpendicular to the muon trajectories and bends the muons into the chambers. In the barrel region $|\eta| < 1.4$, the field reaches 0.5 T, whereas in the end-caps region $1.4 < |\eta| < 2.7$ it reaches 1 T. The chambers are arranged in three concentric cylindrical layers in the barrel at radii of approximately 5 m, 7.5 m and 10 m, and four wheels in each end-cap.

For tracking the muons trajectories, two different types of chambers are used: the Monitored Drift Tube Chambers (MDTs) and Cathode Strip Chambers (CSCs). MDTs cover the $|\eta| < 2.7$ region and are formed by drift tubes placed perpendicular to the r-z plane. The tubes are filled with a combination of argon and carbon dioxide mixture, providing a resolution of 35 μm in the r-z plane. CSCs are installed in the innermost layer of the end-caps, covering the $2 < |\eta| < 2.7$ range. They are employed due to their higher rate capability and time resolution. Furthermore, they are built as multi-wire proportional chambers filled with an argon and carbon dioxide mixture. The resulting resolution is 40 μm in the r-z plane and 5 mm in ϕ.

For triggering and bunch-crossing identification, Resistive-Plate Chambers (RPCs) and Thin-Gap Chambers (TGCs) are used. RPCs extend up to $|\eta|$ ¡ 1.05. They are composed of parallel electrode plates separated by 2 mm with a 4.9 kV/mm electric field applied between them. This configuration results in a time resolution of 1.5 ns, and a hit location resolution of 10 μm in the r-z plane and 10 mm in ϕ. TGCs are multi-wire proportional chambers employed in the $1.05 < |\eta| < 2.7$ range. They are filled with a mixture of carbon dioxide and n-pentane, achieving a time resolution of 4 ns. In addition, their hit resolution is 2-6 mm in the r-z plane and 3-7 mm in ϕ.

3.2.4 Trigger System

The LHC delivers a bunch spacing between collisions of 25 ns which corresponds to collision rates up to 40 MHz. However, the ATLAS detector can only record events at approximately 1 kHz due to its bandwidth and data storage limitations. Therefore, the event rate must be reduced. This task is performed by the trigger system which decides which events are interesting. In Run 2, the trigger system [52] was composed of two levels. The hardware-based Level-1 (L1) trigger reduces the collision rate to 100 kHz. It uses reduced-granularity information from the TGC, RPC and the calorimeters to reconstruct high energetic objects such as muons, jets, electrons, photons, τ leptons or E_T^{miss}. The decision to keep or reject an event is mainly based on the energy of the objects in the event and is taken within 2.5 μs. Moreover, the L1 trigger determines the regions where possible trigger objects were identified referred as Regions-of-Interest (RoIs) and send them to the High Level Trigger (HLT).

The HLT runs reconstruction algorithms, mainly over the RoIs selected by the L1 trigger, using the full granularity detector information. As soon as the event is sufficiently reconstructed, sophisticated selection algorithms are used, reducing the acceptance rate of L1 trigger to approximately 1 kHz. This selection is done within 200 ms and the selected events are then stored.

4. Object Reconstruction

The analysis of an event requires that the outputs of the detector must be converted into physics objects. Section 4.1 provides an introduction to Monte Carlo simulations. An overview of the reconstruction and calibration of the physics objects used in this thesis is given in Sections 4.2-4.4. Section 4.5 and 4.6 detail the signal and background processes, and the cuts applied to each event correspondingly.

4.1 Monte Carlo Simulations

Monte Carlo (MC) simulations are a tool used in high energy physics for a wide range of purposes such as event selection, background estimation or detector optimization. They emulate an event from its generation to its corresponding detector output. The simulation software chain is divided into three steps: the generation of the event, the simulation of the interactions between the the long-lived final state particles with the detector, and the digitization of the energy deposited. Furthermore, in order to compare the simulations to the measured data, they are integrated in the ATLAS software framework ATHENA [53].

MC simulations use MC event generators to randomly generate the proton-proton collisions of the LHC, exploiting the MC method [54]. In this step, the incoming and outgoing particles of the hard scatter event, the initial- and final-state radiations and the hadronisation process are simulated. Next, the toolkit GEANT 4 [55] is deployed to simulate the geometry of the ATLAS subdetectors and the interactions of the long-lived final state particles with them. Finally, the detector hits are digitalized and transformed into measurable quantities.

4.2 Jet Reconstruction

As introduced in Section 1.1.4, an energetic quark or gluon gives rise to collimated groups of hadrons (kaons, pions, ...), leptons and photons known as jets. The particles that form the jet deposit their energy in the calorimeters. The corresponding cells that have an energy greater than a noise threshold are clustered into topological clusters known as *topo-clusters* [56]. They are formed from seed cell that contain more than 4σ of energy, where σ is the average amount of noise expected in the cell, defined as the sum of the expected pile-up and electronic noise. The adjacent cells to the seed cell are added to it if they have at least 2σ of energy. This process ends when there are no adjacents cells that satisfied that condition. At the end, all cells adjacent to the topo-cluster are added regardless of their energy [57].

The topo-cluster energy is calibrated either at the electromagnetic (EM) scale, determined using electons at test beams, or at the local cell weighting scale, which adjust the average topo-cluster to the hadronic energy scale. Furthermore, the position of each topo-cluster is rectified so that its four-momentum points to the hard-scatter primary vertex rather than to the center of the ATLAS detector. This is referred to as the *origin correction*.

According to the algorithm used to reconstruct the jet and the object used as input to the algorithm, the definition of jet varies in collider experiments. At the ATLAS

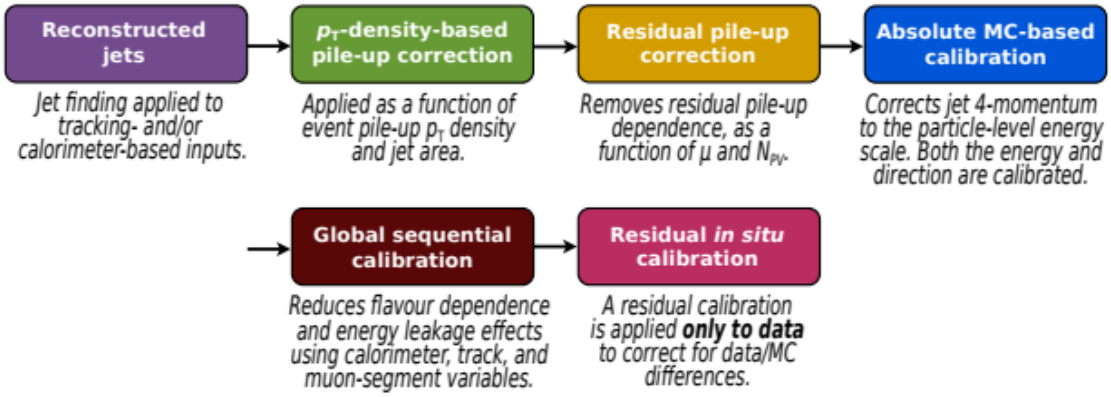


Figure 4.1: Stages of jet calibration [56, p.8].

collaboration, $anti-k_t$ [58] is the most common. The jets considered in this thesis are reconstructed by the $anti-k_t$ algorithm using topo-clusters calibrated at the EM scale, and have a radius parameter R = 0.4.

4.3 Jet Calibration

In order that the measured p_T of the reconstructed jets coincide with those of the particles that form them, they must be calibrated. The calibration consists of five stages which are depicted in Figure 4.1. The first two remove the effect of pile-up on the measured jet transverse momentum. First, the pile-up contribution is subtracted using the average energy density (ρ) and the area of the jet (A). Secondly, the residual dependence of the jet on the expected average number of interactions per bunch crossing (μ) and the number of reconstructed primary vertices (N_{PV}) is removed, yielding

$$p_T = p_T^{reconstructed} - \rho A - \alpha(N_{PV} - 1) - \beta\mu, \tag{4.1}$$

where α and β are the correction coefficients which are determined performing linear fits of the jet p_T as a function of N_{PV} and μ respectively.

Next, the jet energy scale and the jet direction are corrected to the particle-level energy scale by comparing the data to the MC simulations. In this way, the difference of the calorimeter energy response and the various calorimeter granularities are taken into account. Then, the global sequential correction [59] uses properties of the jets, like the portion of the jet energy measured in the first layer of the hadronic calorimeter or the number of tracks of a jet, to account for the jet response dependence on the flavour of the parton which initiated the jet. Finally, a residual in situ calibration measures the jet response in data and MC simulation separately and uses the ratio to correct the data. This final step accounts for differences between the jet response in data and simulation, which are created by the imperfect simulation of the physics processes, the detector materials and their interaction with the particles.

4.4 Missing Transverse Energy

As mention in Section 1.2.1, missing transverse energy is a measure of the transverse momentum imbalance of the detected particles in an event. It is defined as

$$E_{x(y)}^{miss} = -\sum_{particles} p_{x(y)}, \qquad E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}, \tag{4.2}$$

consisting of hard terms and a soft term. The first ones correspond to the missing transverse energy associated to fully calibrated objects such as the photons, electrons, muons, tau-leptons τ and jets, whereas the soft term accounts for the reconstructed momentum not associated to any of the hard terms.

$$E_T^{miss} = E_{T,\gamma}^{miss} + E_{T,e}^{miss} + E_{T,\mu}^{miss} + E_{T,\tau}^{miss} + E_{T,jets}^{miss} + E_{T,soft\,term}^{miss}. \tag{4.3}$$

The soft term is mostly reconstructed at ATLAS using the Track Soft Term algorithm. It is based on tracks of the hard scattering vertex and so, it does not take into account the contribution of neutral particles. Nevertheless, relying on tracks rather than calorimeter depositions makes it very robust against changing the pile-up condition.

4.5 Signal and Background Processes

The signal process of this analysis is the invisible decay of a Higgs boson produced via vector-boson fusion (VBF), which is depicted in Figure 4.2. Its dominant SM background contribution is a Z boson decaying to a neutrino and anti-neutrino pair, $Z \to \nu\bar{\nu}$, in association with two jets. Exemplary Feynman diagrams are shown in Figure 4.3. This process yields a signature in the detector equal to the signal process, turning $Z \to \nu\bar{\nu}$ in an irreducible background.

In this analysis, the signal and the background processes are evaluated by MC samples. For the background, only diagrams where the Z boson is QCD-produced are included.

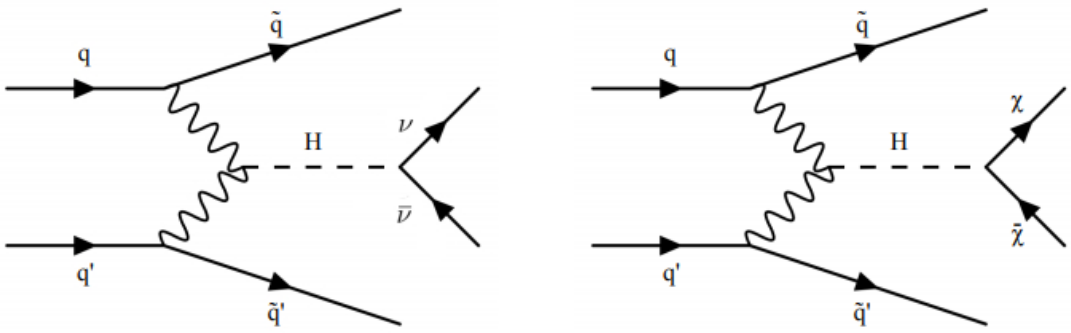


Figure 4.2: Tree-level Feynman diagram of the invisible decay of a Higgs boson produced via VBF according to the SM (left) or decaying into dark matter particles (right) [60, p.2].

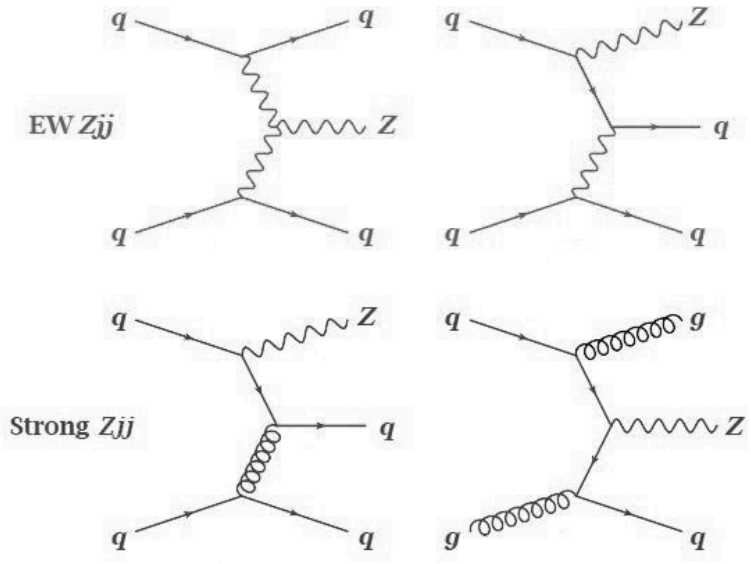


Figure 4.3: Examples of tree-level Feynman diagrams for the dominant SM background in the invisible decay of a Higgs boson produced via VBF. The top left diagram is not included in the MC background sample since only Feynman diagrams where the Z boson is QCD-produced are considered. [61, p.4].

4.6 Event Selection

The selected events of the MC samples used for the analysis were required to fulfill the following cuts. It was imposed that for each event there was

- more than one jet,
- a leading jet with $p_T > 80$ GeV,
- a subleading jet with $p_T > 50$ GeV,
- $|\eta| < 4.5$ for all jets,
- $E_T^{miss} > 180$ GeV.

5. Higgs to Invisible Enhancement Using Kinematic Variables

In order to achieve a good separation between Higgs bosons decaying invisibly and its dominant backgound, $Z \to \nu\bar{\nu}$, this thesis engages in multivariate classification methods based on machine learning techniques. The strategy and workflow of the analysis are described in Section 5.1. Section 5.2. displays the multivariate analysis results, for which the data set was split into a training set (90%) and a test set (10%). Finally, Section 5.3 shows the results of training an autoncoder using kinematic variables.

5.1 Strategy and Workflow of the Analysis

The Higgs boson can be a possible dark matter portal if dark matter particles couple to it. If this is the case, the branching ratio of invisible Higgs boson decays will be higher than the one predicted by the SM. However, it is very challenging to measure Higgs bosons decaying invisibly due to the fact that the signal is around five orders of magnitude lower than the background. To overcome this challenge, the strategy of this analysis is to use detector images instead of kinematic variables to try to better distinguish the invisible decay of a Higgs boson produced via VBF from its dominant background, a Z boson decaying to a neutrino and an anti-neutrino in association with two jets. There are two main differences between these process. The first one is the width of the rapidity gap which is defined as the region between the two jets with the highest p_T (Figure 5.1). The second one is the hadronic activity inside the rapidity gap since the Z to $\nu\bar{\nu}$ process is expected to have more hadronic activity in this region due to color exchange.

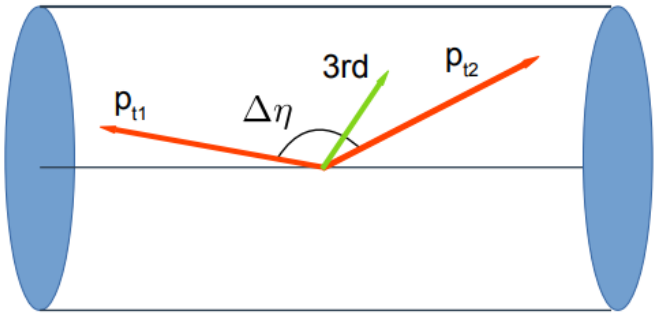


Figure 5.1: Sketch of the rapidity gap definition.

The workflow of the analysis will be the following. First, a multivariate analysis is performed and used as a benchmark to compare to the autoencoder results. Then, as a proof of principle, an autoencoder is trained on event-level variables, such as m_{jj}, in order to see if it can reproduce high-level quantities. Finally, an autoencoder is trained using detector images (Chapter 6).

5.2 Multivariate Analysis Results

The choice of the classification model is decisive for a good signal-background separation. Here, the performance of a BDT and the rectangular cut classifier using the same data set are compared. The configuration of their settings is a bit different from the default one specified in the TMVA users guide [33]. The optimized values are shown in Table 5.1. In order to avoid a very complex model and overtraining, the maximal depth of the BDT is set to 3.

Table 5.1: Optimized setting values for the cut classifier and the BDT.

cut classifier	BDT
VarTransform = None	VarTransform = None
VarProp = FSmart	MaxDepth = 3
	NTrees = 1000
	nCuts = 10
	MinNodeSize = 3

The kinematic variables passed as input to the models are: the p_T, η and Φ of the leading and subleading jets, m_{jj}, E_T^{miss}, $|\Delta\eta|$ and the number of jets in the rapidity gap. The input distributions are displayed in Figure 5.2 and Figure 5.3. The variables with highest discrimination power are the η of the jets and so, $|\Delta\eta|$. The number of jets in the rapidity gap also provides a good signal-background separation because more jets are expected in the rapidity gap in the $Z \to \nu\bar{\nu}$ process due to color exchange. Furthermore, since the jets of dijets events are uniformly distributed in Φ, the Φ distributions is flat for both events and Φ has no discrimination power.

Figure 5.4 represents the result of the training processes as ROC curves for the trained models. The closer the curve is to the right corner, the better is the signal-background separation because the (x,y) = (1,1) point of the graph corresponds to the case where all background is rejected, while all signal is kept. As it can be seen, the cut-based classifier performs worse than the BDT since it places a cut on each variable separately. It is therefore not able to provide a good separation since it does not take into account correlations between the variables. The BDT yields a very good separation of signal and background events.

5.3 Input Regeneration Using An Autoencoder

As explained in Section 2.1.2, an autoencoder is able to regenerate the input. The goal of this test is to exploit this feature in order to reconstruct the MC distributions of the following kinematic variables: the p_T and η of the leading and subleading jets, m_{jj}, E_T^{miss}, $|\Delta\eta|$ and the number of jets in the rapidity gap. The input is structured in a "image", like the one depicted in Figure 5.5, which contains eight pixels, each for one of the input kinematic variables.

Before the autoencoder is trained, each pixel value is normalized in the following

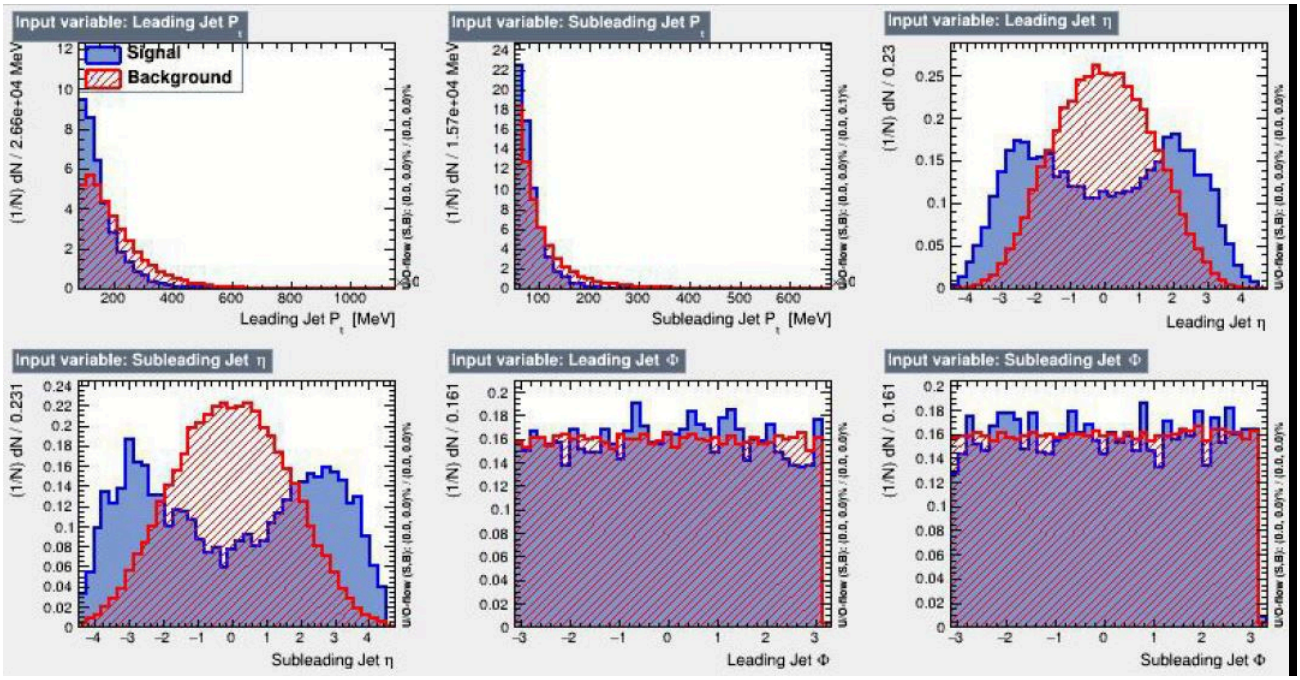


Figure 5.2: p_T, η and Φ distributions for the leading and subleading jets used as input for the MVA. The blue distribution corresponds to H $\rightarrow$ invisible and the red one to Z $\rightarrow \nu\bar{\nu}$.

way

$$normalized\,pixel\,value = \frac{x - min(x)}{max(x) - min(x)}. \tag{5.1}$$

The data is separated into a training set (90%) and a test set (10%). The training set is further split into data the model is trained on (80%) and a validation set (20%). This helps in reducing the chances of overfitting, as the model is being validated on data that it has not seen before.

The architecture and hyperparameters of the convolutional autoencoder are based on the ones described in the Datacamp tutorial [62]. The latter are shown in Table 5.2. Its architecture consists of six convolutional layers interspersed with four max-pooling layers, converting the input from wide (1 x 8 image) and thin (a single channel) to small (1 x 2 image at the code layer) and thick (128 channels). The encoder architecture is made up of three convolutional layers with 32, 64 and 128 filters correspondingly of size 3x3. The first two layers are followed each one by a downsampling max-pooling layer which halves the length of the "images". The decoder has a similar structure as the encoder. Its convolutional layers are followed by upsampling layers and the convolutional layers have 128, 64 and 1 filters respectively.

Figure 5.6 and Figure 5.7 show the $|\Delta\eta|$ and m_{jj} input MC distributions (blue) as well as the ones regenerated by the autoencoder (red). The autoencoder is able to reconstruct the MC distributions of the input kinematic variables in such a way that no significant deviation is visible.

In order to validate the output of the autoencoder, distributions of variables which are not part of the training process are investigated. Therefore, the autoencoder network is trained with a different set of eight variables. Instead of using the m_{jj} and $|\Delta\eta|$ variables, the masses of the jets (m_1 and m_2) are part of the input "image".

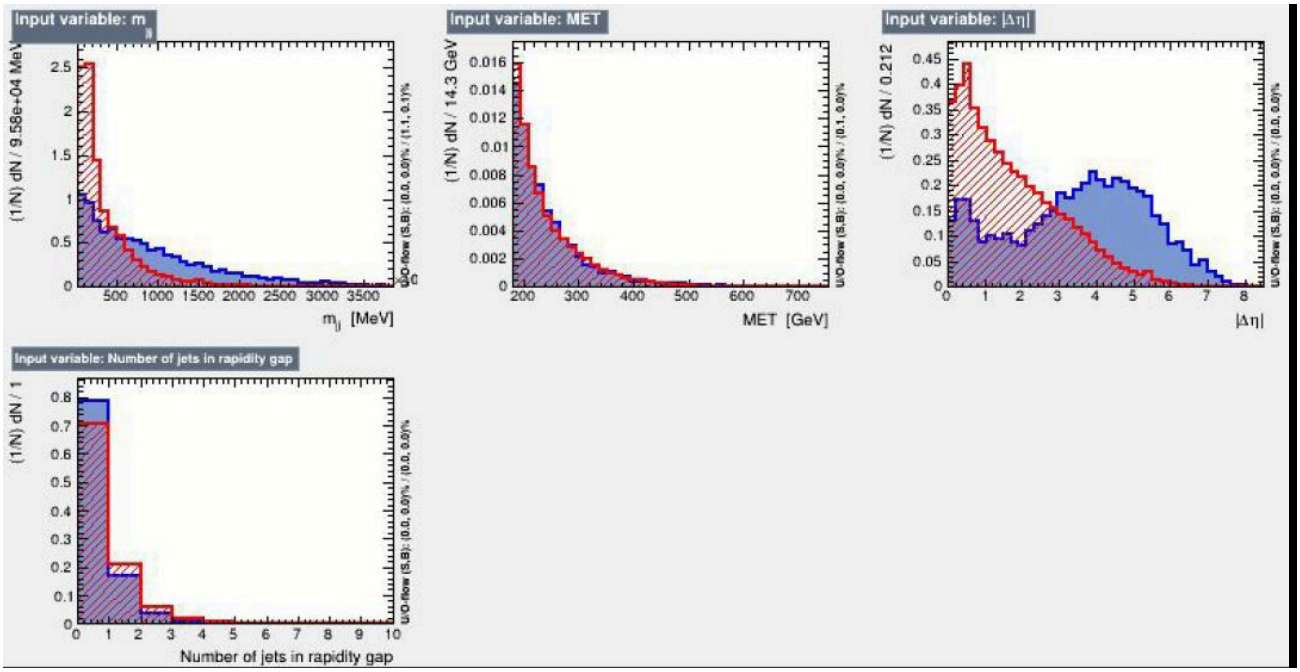


Figure 5.3: m_{jj}, E_T^{miss}, $|\Delta\eta|$ and number of jets in the rapidity gap distributions used as input for the MVA. The blue distribution corresponds to H $\rightarrow$ invisible and the red one to Z $\rightarrow \nu\bar{\nu}$.

Table 5.2: Parameters of the convolutional autoencoder.

Parameters
Batch size = 128
Epochs = 50
loss function = mean squared error
optimizer = RMSprop

The autoencoder architecture and hyperparameters are chosen just as in the autocoder training described above.

The input invariant mass distribution is compared to the reconstructed distribution by the autoencoder in Figure 5.8. Although the invariant mass itself is not part of the input "image", the autoencoder is able to reproduce the m_{jj} MC distribution. This implies that the network is learning the underlying properties for the signal events.

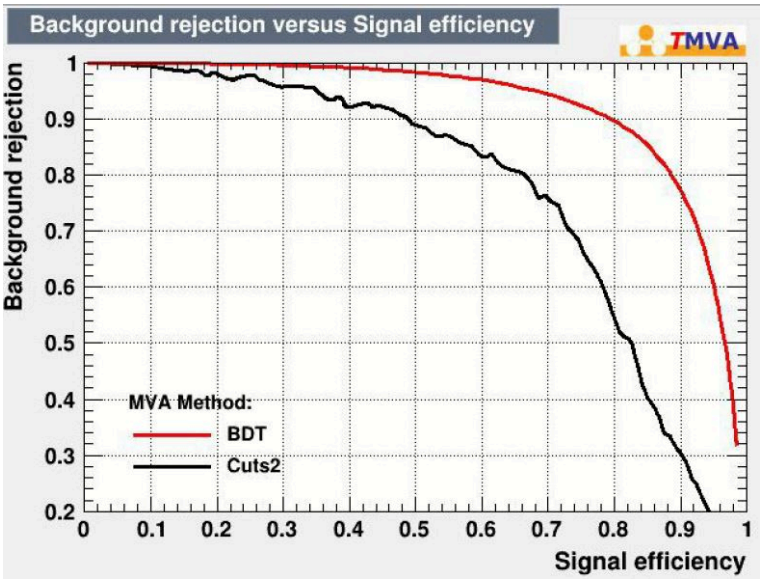


Figure 5.4: Plot showing the background rejection as a function of the signal efficiency for the BDT (red) and the cut classifier (black).

p_{t1}	p_{t2}	η_1	η_2	m_{jj}	MET	$\lvert\Delta\eta\rvert$	N_{jets}
pixel 1	pixel 2	pixel 3	pixel 4	pixel 5	pixel 6	pixel 7	pixel 8

Figure 5.5: Example of an input ”image” where each pixel contains the value of one of the eight kinematic variables of an event.

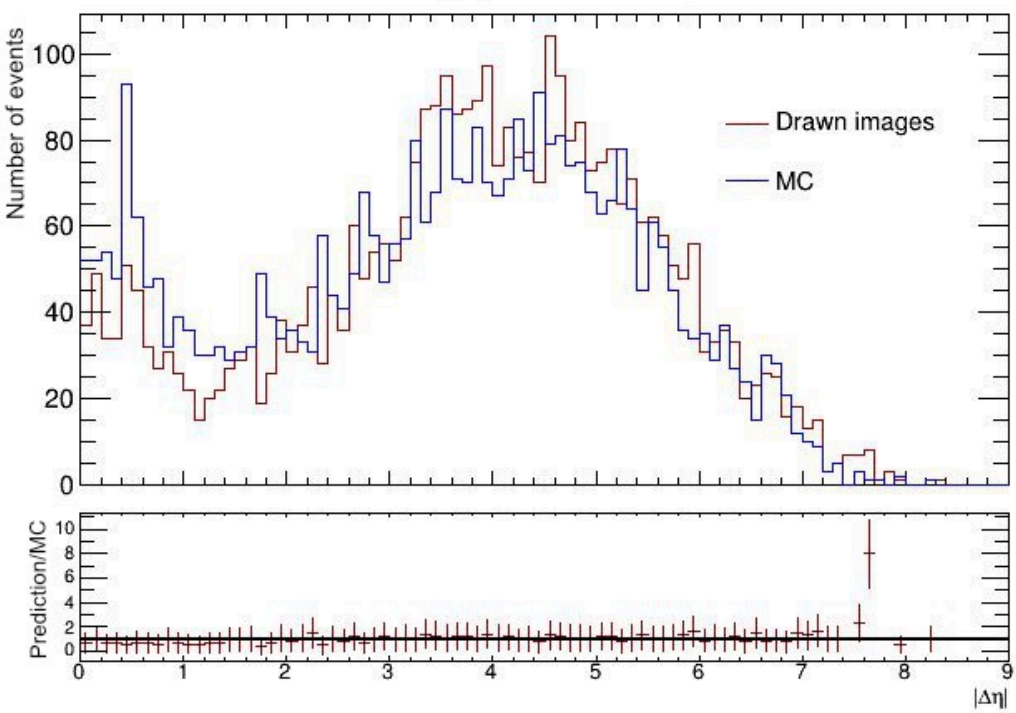


Figure 5.6: Comparison of the $|\Delta\eta|$ MC distribution (blue) with the $|\Delta\eta|$ distribution predicted by the autoencoder (red), for H→invisible using $|\Delta\eta|$ as the input variable. The lower plot represents the ratio of the input and predicted distributions.

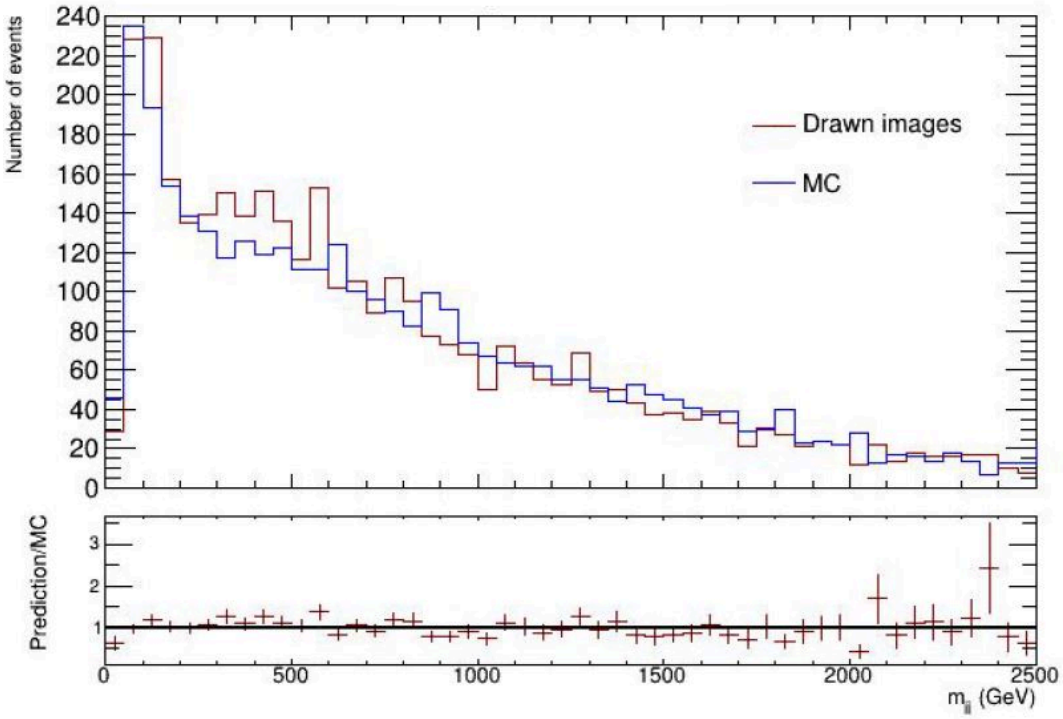


Figure 5.7: Comparison of the m_{jj} MC distribution (blue) with the m_{jj} distribution predicted by the autoencoder (red), for H→invisible using m_{jj} as the input variable. The lower plot represents the ratio of the input and predicted distributions.

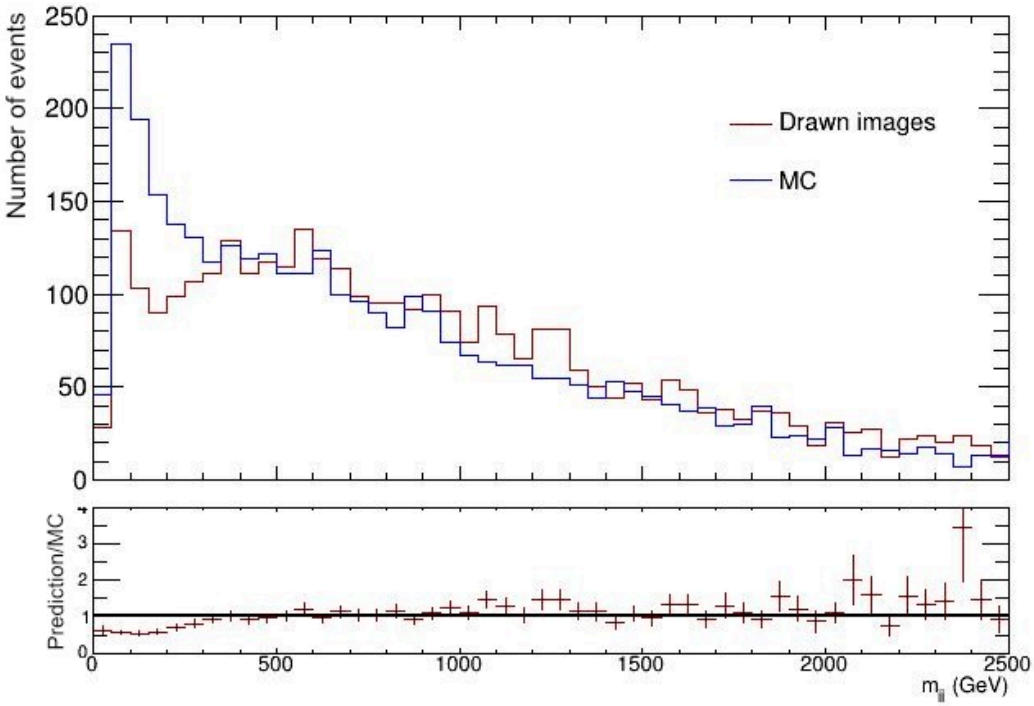


Figure 5.8: Comparison of the m_{jj} MC distribution (blue) with the m_{jj} distribution predicted by the autoencoder (red), for H→invisible where m_{jj} is not an input variable. The lower plot represents the ratio of the input and predicted distributions.

6. Higgs to Invisible Separation Using Images

To determine if the signal to background separation can be improved, an autoencoder is trained using detector images. Section 6.1 describes the detector images, while Section 6.2 presents the results obtained.

6.1 Detector Images

Instead of kinematic event variables, topocluster energy η x Φ maps (calorimeter images) or tracks p_T η x Φ maps (tracker images), like the ones of Figure 6.1 and 6.2, can be used as input. One advantage of it is that these images do not rely on a jet algorithm. Therefore, it is possible to go below 17 GeV where there is no jet calibration. Another benefit is that a description of the new physics that are being looking for is not needed. Furthermore, images contain more information. However, jet algorithms have pile-up subtraction. Another disadvantage is that there could be differences between the data and the MC simulation because the fact that they agree at jet level does not mean that they agree at cluster level.

It is difficult for neural networks to reconstruct images with many empty pixels. To make it easier for the autoencoder, so that it does not have to learn the rotational symmetry of the detector, the detector images were rotated so that the ϕ of the highest p_T jet in the rapidity gap lies at $\phi = 0$. Also, the granularity of the tracker images, $\Delta\Phi = \Delta\eta = 0.3$, was chosen to be bigger than the ATLAS tracker one so that the images did not contain that many empty pixels. The granularity of the calorimeter images is the same as the Level-1 trigger one,

- $\Delta\Phi = \Delta\eta = 0.1$ for $|\eta| \leq 2.5$
- $\Delta\Phi = \Delta\eta = 0.2$ for $2.5 < |\eta| \leq 3.1$
- $\Delta\Phi = 0.2$ $\Delta\eta = 0.1$ for $3.1 < |\eta| \leq 3.2$
- $\Delta\Phi = \Delta\eta = 0.4$ for $|\eta| > 3.2$

6.2 Autoencoder Results

Regarding its architecture, the autoencoder used for the calorimeter images contains seven layers with 512, 64, 64, 16, 64, 64, 512 nodes respectively in each layer. These numbers were chosen trying different configurations, taking into account the size of the input images. The architecture of the autoencoder used for the tracker images has the same structure as the one used for calorimeter images, with the difference that the number of nodes in each layer is different. This is because the tracker and calorimeter images have different number of pixels. The former are 20 x 20 images, while the latter are 66 x 64 images. The number of nodes in each layer is 512, 64, 64, 16, 64, 64, 512 respectively. Unlike the autoencoder used for the results display in Section 5.3, these

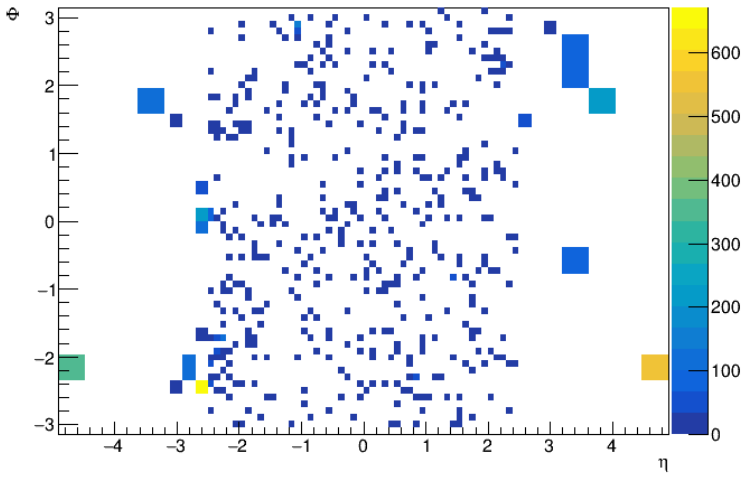


Figure 6.1: Example of the calorimeter image of an event where the topo cluster energy (GeV) is represented.

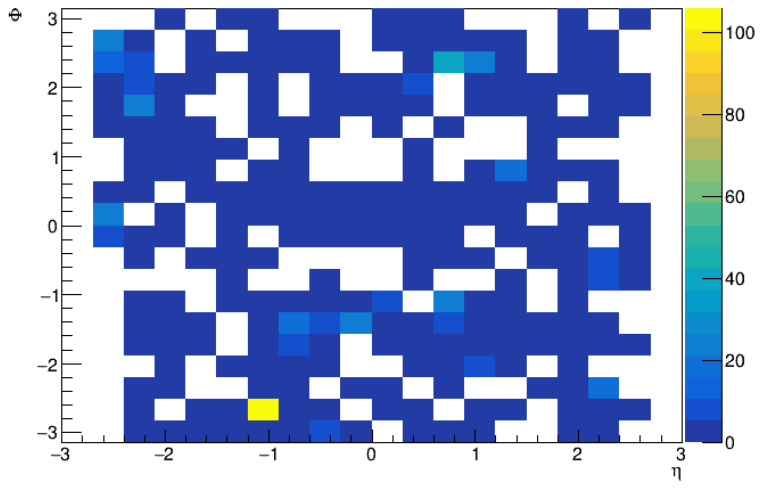


Figure 6.2: Example of the tracker image of an event where the p_T is represented.

are not convolutional autoencoders. Hence, the images are flattened (compressed into 1 x number of pixels images) before entering the autoencoder.

As explained in Section 2.1.2, an autoencoder is able to regenerate the input. However, in this case, the autoencoder does not reproduce the pixel energy and p_T distributions (Figure 6.3). This is due to the fact that there are many empty pixels in the image. The autoencoder tends to not produce totally empty pixels, but rather fills very low energies into each. Therefore, all the pixels are triggered ($E > 0$ or $p_T > 0$) for the regenerated images (Figure 6.4). Nevertheless, when the detector images are averaged, the images regenerated by the autoencoder are very similar to the input ones (Figure 6.5). So, the autoencoder regenerates the total energy and p_T per image (Figure 6.6).

In order to determine if the autoencoder can yield a better signal-background separation, the anomaly detection technique is used (Section 2.1.2). First, the data is separated into a training and test set following the same procedure as in Section 5.3. Then, the autoencoder is trained with detector images of either H→invisible or Z→ $\nu\bar{\nu}$

events. In the end, it is tested with H→invisible and Z→ $\nu\bar{\nu}$ events.

Figure 6.7 represents the results obtained as ROC curves. Compared to Figure 5.4, the autoencoders yields a worse signal-background separation than the boosted decision trees, since there is lower background rejection for same signal efficiency. This may be because for the training of the boosted decision trees, both processes are used as input, whereas here only one is employed at a time. Also, the autoencoder performs worse when it uses tracker images.

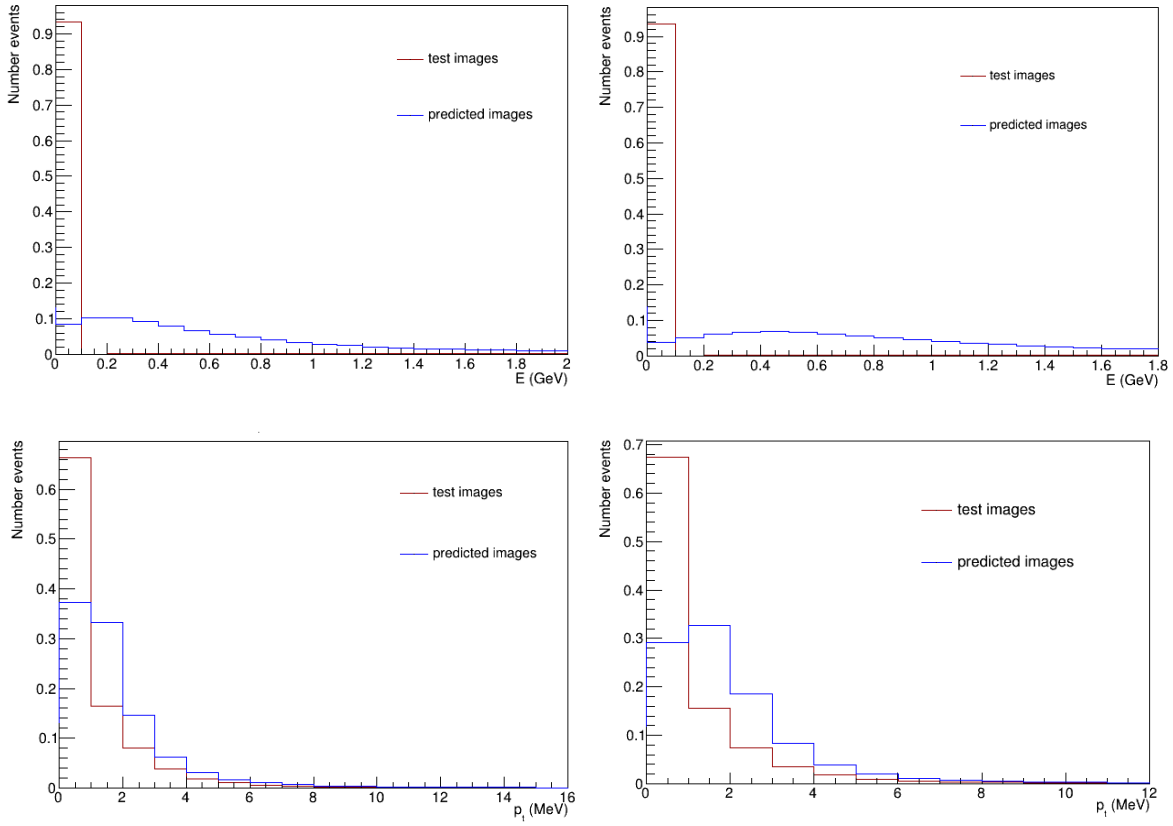


Figure 6.3: Energy distributions of the calorimeter (top) and transverse momentum distributions of the tracker (bottom), in the cases where the autoencoder is trained on H→invisible data (left) or on Z→ $\nu\bar{\nu}$ data (right).

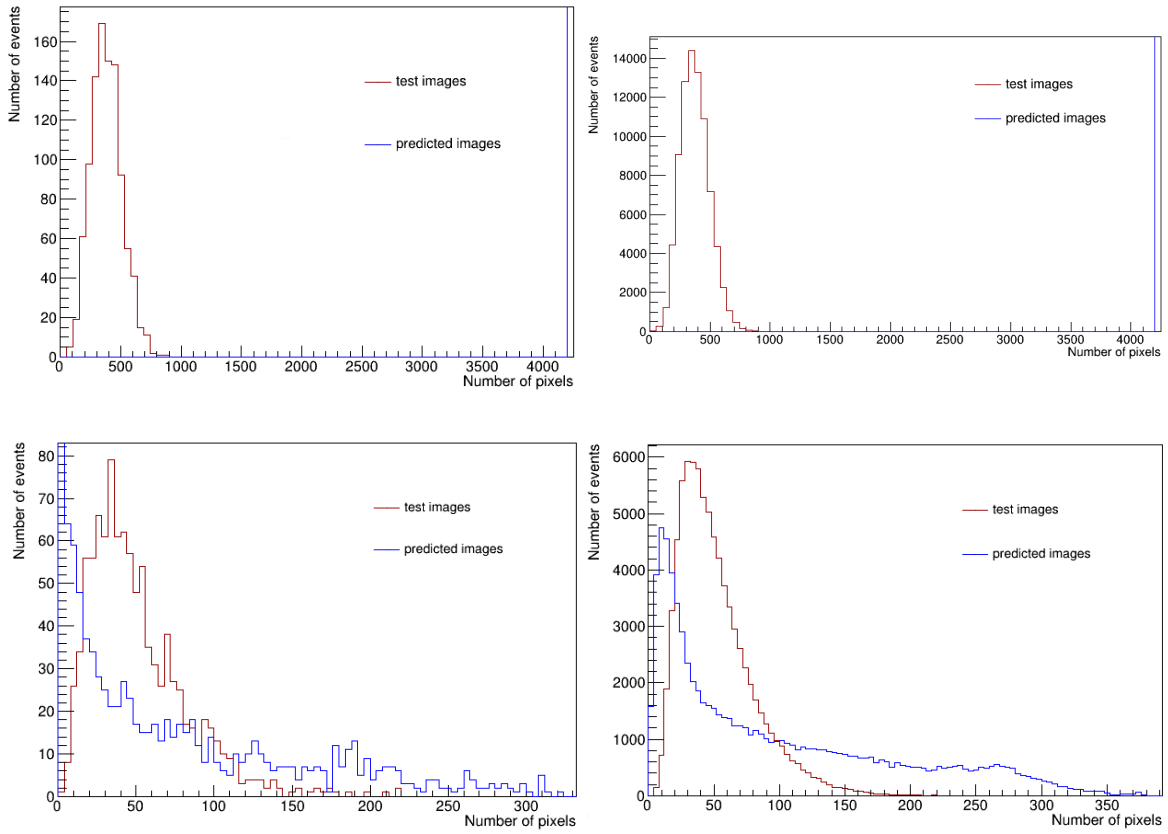


Figure 6.4: Number of pixels triggered per image for the calorimeter (top) and the tracker (bottom) in the cases where the autoencoder is trained on H→invisible data (left) or on Z→ $\nu\bar{\nu}$ data (right). In the top graphs, all the pixels are triggered for the regenerated images. In the bottom graphs, that is not the case because a lower limit has been placed to consider that a pixel is triggered.

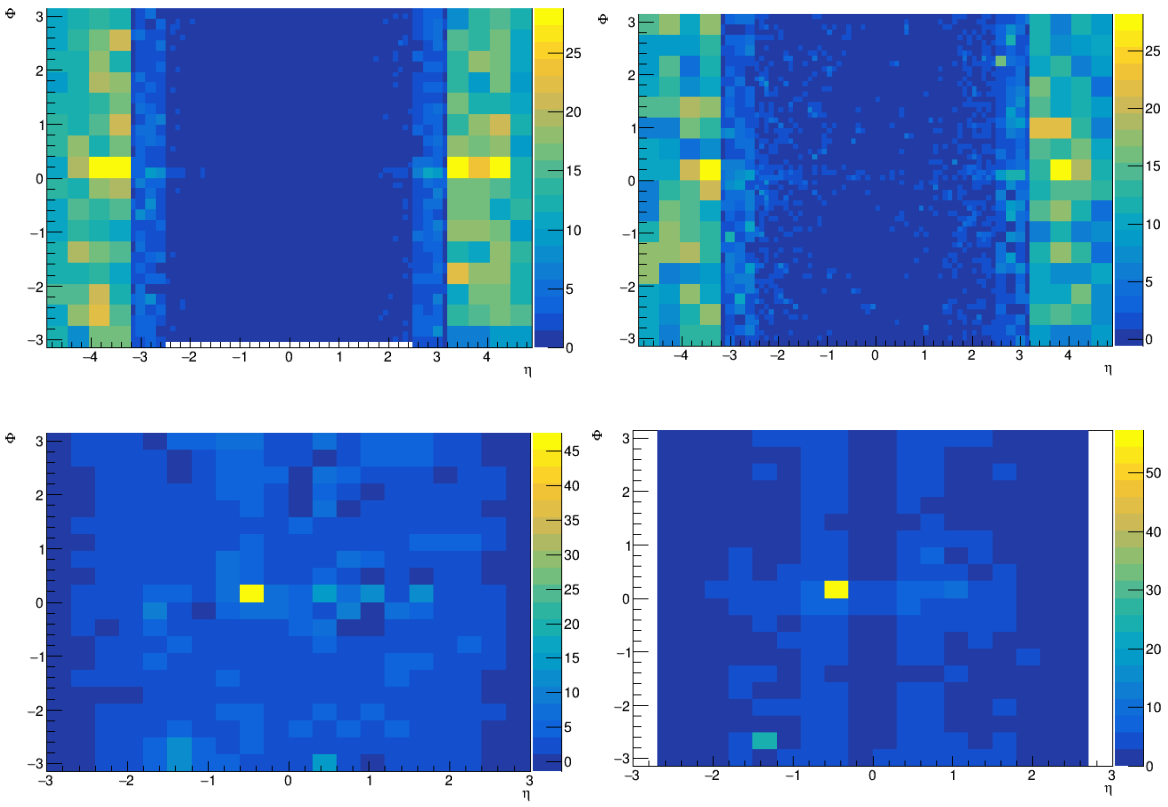


Figure 6.5: H→invisible averaged images for the calorimeter (top) and Z→ $\nu\bar{\nu}$ averaged images for the tracker (bottom), in the cases where the input images are averaged (left) and where the reconstructed images are averaged.

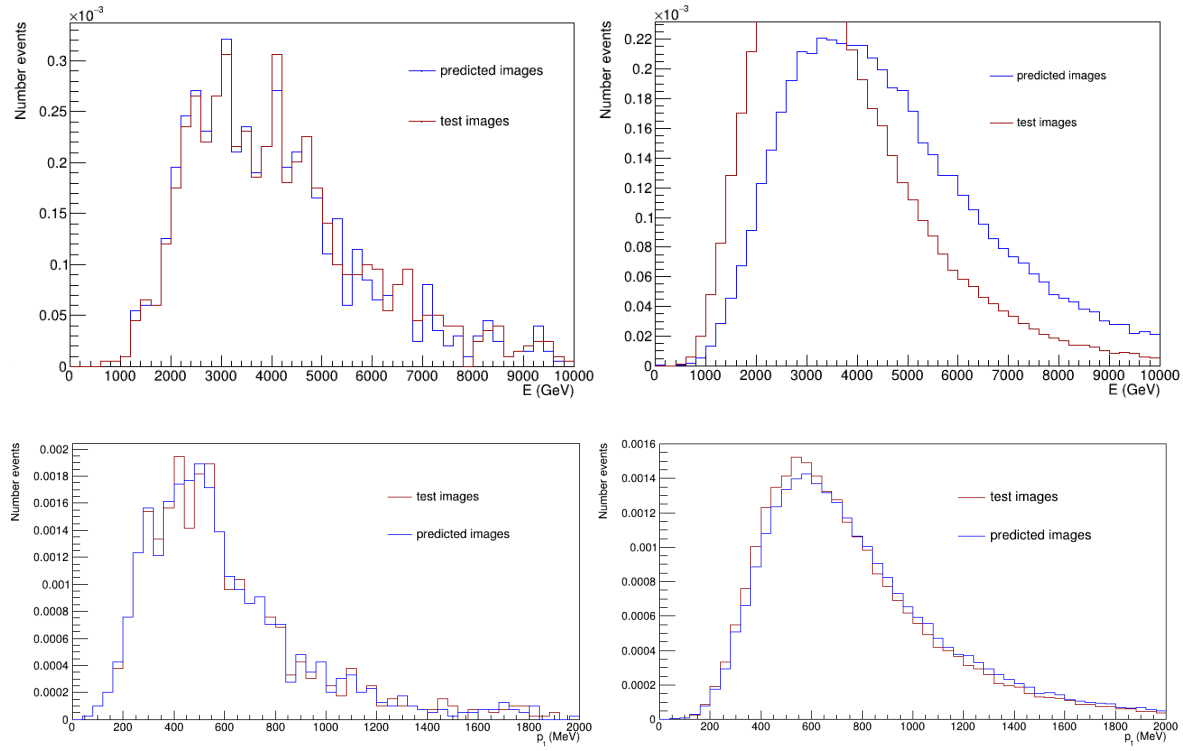


Figure 6.6: Energy per image distributions of the calorimeter (top) and transverse momentum per image distributions of the tracker (bottom) in the cases where the autoencoder is trained on H→invisible data (left) or on Z→ $\nu\bar{\nu}$ data (right).

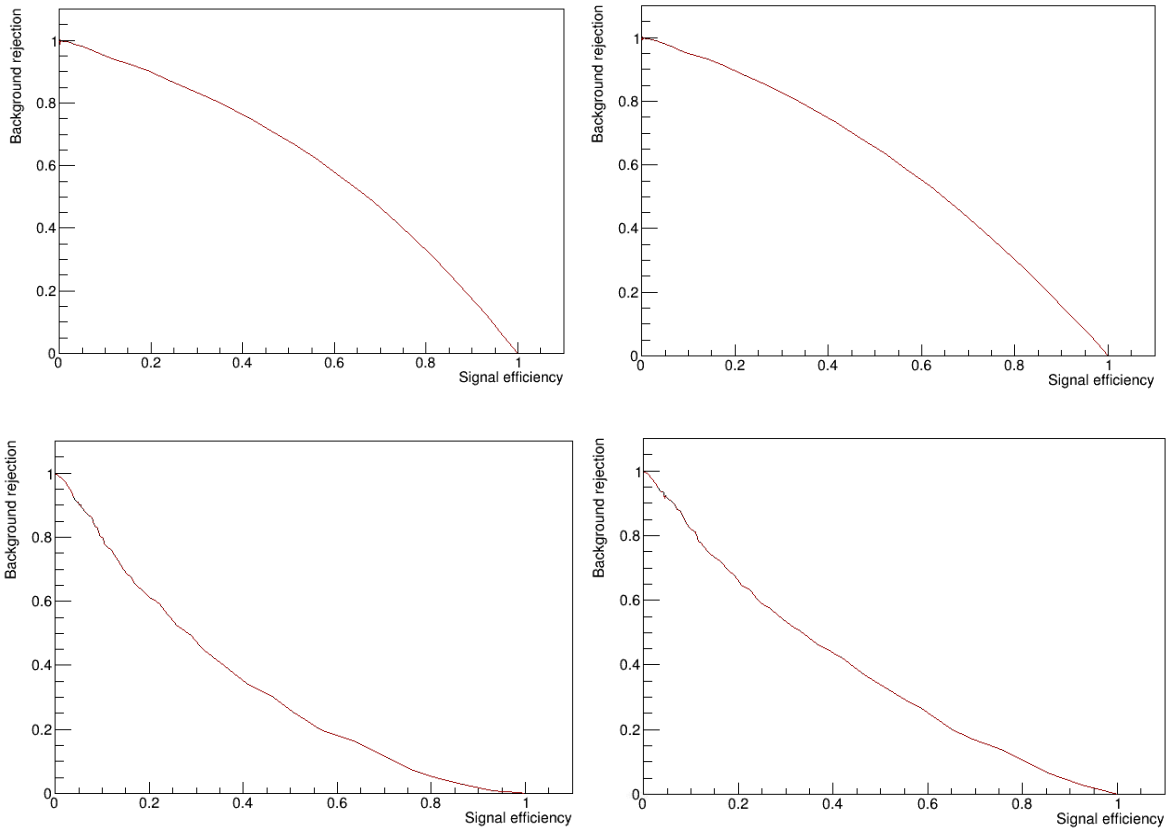


Figure 6.7: ROC curves for the calorimeter (top) and the tracker (bottom) in the cases where the autoencoder is trained on H→invisible data (left) or on Z→ $\nu\bar{\nu}$ data (right).

Conclusion

Autoencoders provide the opportunity to classify images by learning their features. In this study, an autoencoder for anomaly detection was implemented to investigate if it was able to better distinguish invisible Higgs boson decays where the Higgs boson is produced via vector-boson fusion, from its dominant background, Z bosons decaying into a neutrino and an anti-neutrino in association with two jets. Although the autoencoder could regenerate the total energy and p_T of the events, it yielded a worst separation between the two processes than boosted decision trees. These results were obtained by studying the efficiencies and rejection rates of both algorithms.

In order to extract features and recognize patterns, machine learning algorithms need a lot of data. If more events where used for the autoencoder during the training phase, the potential of the autoencoder in the Higgs to invisible enhancement would be bigger. This is important because, unlike for the boosted decision trees, autoencoders can only use Higgs or Z events, but not both at the same time, in the training process. Furthermore, the signal-background separation obtained using the ATLAS tracker images, which is worse than the one obtained using the ATLAS calorimeter images, could be improved with a further optimization of the autoencoder architecture, since that would allow to use a smaller granularity for these images.

List of Figures

List of Tables

Bibliography

[1] Thomson, Mark. *Modern particle physics*. Cambridge University Press, (2013).

[2] Pich, Antonio. "The Standard model of electroweak interactions." *arXiv preprint arXiv:1201.0537* (2012).

[3] Dawson, S. "Introduction to electroweak symmetry breaking." *arXiv preprint hep-ph/9901280* (1999).

[4] Wikipedia (2020). *Standard Model*. Available from: https://en.wikipedia.org/wiki/Standard_Model. [Accessed 6th October 2020].

[5] Aad, Georges, et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." *Physics Letters B* 716.1 (2012): 1-29.

[6] Chatrchyan, Serguei, et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." *Physics Letters B* 716.1 (2012): 30-61.

[7] Glashow, Sheldon L. "The renormalizability of vector meson interactions." *Nuclear Physics* 10 (1959): 107-117.

[8] Weinberg, Steven. "A model of leptons." *Physical review letters* 19.21 (1967): 1264.

[9] Salam, Abdus. "Weak and electromagnetic interactions." *Selected Papers Of Abdus Salam: (With Commentary)*. (1994). 244-254.

[10] Sirunyan, Albert M., et al. "A measurement of the Higgs boson mass in the diphoton decay channel." *Physics Letters B* (2020): 135425.

[11] Djouadi, Abdelhak. "Higgs physics: theory." *Pramana* 79.4 (2012): 513-539.

[12] Aad, Georges, et al. "Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment." *Physical Review D 101.1* (2020): 012002.

[13] ATLAS Collaboration, "Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4l$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector", *J. High Energy Phys. 03* (2018) 095.

[14] ATLAS Collaboration, "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B 789, 508* (2019).

[15] ATLAS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D 98, 052005* (2018).

[16] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D 99, 072001* (2019).

[17] ATLAS Collaboration, "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector", *Phys. Lett. B 786, 59* (2018).

[18] Bertone, Gianfranco, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints." *Physics reports 405.5-6* (2005): 279-390.

[19] Clowe, Douglas, et al."A direct empirical proof of the existence of dark matter." *The Astrophysical Journal Letters 648.2* (2006): L109.

[20] Freese, Katherine. "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars." *EAS Publications Series 36* (2009): 113-126.

[21] Aghanim, Nabila, et al. "Planck 2018 results-VI. Cosmological parameters." *Astronomy Astrophysics 641* (2020): A6.

[22] Taoso, Marco, Gianfranco Bertone, and Antonio Masiero. "Dark matter candidates: a ten-point test." *Journal of Cosmology and Astroparticle Physics 2008.03* (2008): 022.

[23] Preskill, John, Mark B. Wise, and Frank Wilczek. "Cosmology of the invisible axion." *Physics Letters B 120.1-3* (1983): 127-132.

[24] Roszkowski, Leszek, Enrico Maria Sessolo, and Sebastian Trojanowski. "WIMP dark matter candidates and searches—current status and future prospects." *Reports on Progress in Physics 81.6* (2018): 066201.

[25] Burgess, C. P., Maxim Pospelov, and Tonnis Ter Veldhuis. "The Minimal model of nonbaryonic dark matter: A Singlet scalar." *Nuclear Physics B 619.1-3* (2001): 709-728.

[26] de Florian, Daniel, et al. "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector." *arXiv. org* (2016).

[27] ATLAS Collaboration. "Search for invisible Higgs boson decays with vector boson fusion signatures with the ATLAS detector using an integrated luminosity of 139 fb^{-1}." *ATLAS-CONF-2020-008*, (2020).

[28] Sagar Sharma. (2017) *Epoch vs Batch Size vs Iterations* Available from: https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9 [Accessed 16th June 2021].

[29] Arden Dertat. (2017) *Applied Deep Learning - Part 1: Artificial Neural Networks* Available from: https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6 [Accessed 16th June 2021].

[30] AI Wiki. (2018) *Activation Function* Available from: https://docs.paperspace.com/machine-learning/wiki/activation-function [Accessed 5th June 2021].

[31] Arden Dertat. (2017) *Applied Deep Learning - Part 3: Autoencoders.* Available from: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798 [Accessed 16th September 2020].

[32] Arden Dertat. (2017) *Applied Deep Learning - Part 4: Convolutional Neural Networks*. Available from: https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2 [Accessed 19th July 2021].

[33] Hoecker, Andreas, et al. "TMVA-toolkit for multivariate data analysis." *arXiv preprint physics/0703039* (2007).

[34] Shapiro, Alexander. "Monte Carlo sampling methods." *Handbooks in operations research and management science 10* (2003): 353-425.

[35] Mirjalili, Seyedali. "Genetic algorithm." *Evolutionary algorithms and neural networks*. Springer, Cham, (2019). 43-55.

[36] Van Laarhoven, Peter JM, and Emile HL Aarts. "Simulated annealing." *Simulated annealing: Theory and applications*. Springer, Dordrecht, (1987). 7-15.

[37] Evans, Lyndon, and Philip Bryant. "LHC machine." *Journal of instrumentation 3.08* (2008): S08001.

[38] Boyd, J. T. "LHC Run-2 and Future Prospects." *arXiv preprint arXiv:2001.04370* (2020).

[39] Hoecker, Andreas. "Physics at the LHC Run-2 and Beyond." *arXiv preprint arXiv:1611.07864* (2016).

[40] Christiane Lefèvre (2008). *The CERN accelerator complex*. Available from: https://cds.cern.ch/record/1260465. [Accessed 24th October 2020].

[41] Aad, Georges, et al. "The ATLAS experiment at the CERN large hadron collider." *Jinst* 3 (2008): S08003.

[42] Chatrchyan, Serguei, et al. "The CMS experiment at the CERN LHC." *Journal of instrumentation 3.8* (2008): S08004.

[43] Alves Jr, A. Augusto, et al. "The LHCb detector at the LHC." *Journal of instrumentation 3.08* (2008): S08005.

[44] Aamodt, Kenneth, et al. "The ALICE experiment at the CERN LHC." *Journal of Instrumentation 3.08* (2008): S08002.

[45] Aad, Georges, et al. "ATLAS data quality operations and performance for 2015-2018 data-taking". *No. CERN-EP-2019-207. ATLAS-DAPR-2018-01-003*, (2019).

[46] Capeans, M., et al. "ATLAS insertable B-layer technical design report". *No. CERN-LHCC-2010-013. ATLAS-TDR-019*, (2010).

[47] ATLAS collaboration. "The ATLAS Inner Detector commissioning and calibration." *European Physical Journal C 70.3* (2010): 787-821.

[48] Aharrouche, M., et al. "Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 568.2* (2006): 601-623.

[49] Davidek, T. "ATLAS Tile Calorimeter performance for single particles in beam tests." *J. Phys.: Conf. Ser. Vol. 160.* (2009).

[50] Dowler, B., et al. "Performance of the ATLAS hadronic end-cap calorimeter in beam tests." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 482.1-2* (2002): 94-124.

[51] Archambault, J. P., et al. "Energy calibration of the ATLAS liquid argon forward calorimeter." *Journal of Instrumentation 3.02* (2008): P02002.

[52] Collaboration, A.T.L.A.S. "Performance of the ATLAS trigger system in 2015." *Eur. Phys. J. C 77* (2017): 317.

[53] ATLAS Collaboration. "ATLAS Computing Technical Design Report". *ATLAS-TDR-017, CERN-LHCC-2005-022* (2005).

[54] Metropolis, Nicholas, and Stanislaw Ulam. "The monte carlo method." *Journal of the American statistical association 44.247* (1949): 335-341.

[55] Agostinelli, Sea, et al. "GEANT4—a simulation toolkit." *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506.3* (2003): 250-303.

[56] Aad, Georges, et al. "Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". *No. arXiv: 2007.02645. ATLAS-JETM-2018-05-003*, (2020).

[57] Schramm, Steven. "ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentzboosted Objects." *EPJ Web of Conferences. Vol. 182. EDP Sciences*, (2018).

[58] Cacciari, Matteo, Gavin P. Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm." *Journal of High Energy Physics* 2008.04 (2008): 063.

[59] ATLAS collaboration. "Jet global sequential corrections with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 8$ TeV." *ATLAS-CONF-2015-002*, (2015).

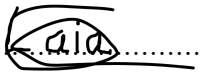
[60] Sirunyan, Albert M., et al. "Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV." *Physics Letters B 793* (2019): 520-551.

[61] Aad, Georges, et al. "Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton - proton collisions at ATLAS." *No. CERN-EP-2020-045. ATLAS-STDM-2017-27-003*, (2020).

[62] Aditya Sharma (2018), "Implementing Autoencoders in Keras: Tutorial" : https://www.datacamp.com/community/tutorials/autoencoder-keras-tutorial [Accessed 12th May 2021]

Erklärung:

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

Heidelberg, den 28.09.2021 .....  .....