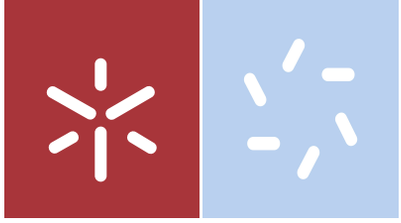
Universidade do Minho
Escola de Ciências

Maura Gabriela Barros Teixeira

Search for Dark Matter in Monotop Events at the Large Hadron Collider

Search for Dark Matter in Monotop Events at the Large Hadron Collider
Maura Gabriela Barros Teixeira
UMinho | 2020

novembro de 2020



Universidade do Minho
Escola de Ciências

Maura Gabriela Barros Texeira

Search For Dark Matter in Monotop Events at the Large Hadron Collider

Dissertação de Mestrado
Mestrado em Física

Trabalho efetuado sob a orientação do
Profesor Doutor Nuno Filipe da Silva Fernandes de Castro
e do
Doutor Miguel Correia dos Santos Crispim Romão

novembro de 2020

Direitos de autor e condições de utilização do trabalho por terceiros

Acknowledgements

This work could not be possible without the support of many people.
I start by ackowledging Professor Nuno Castro and Doctor Miguel Romão for all of their supervision and for allowing me to take part in this project. Thank you for all the invaluable knowledge, experience and for always being supportive and encouraging.
I am grateful to my LIP-Minho colleagues, Ana, Céu, Emanuel, Guilherme, Maria, Rute and Tiago, for their promptness in giving me help every time I needed it.
I also thank the monotop analysis team for the knowledge and the experience of being part of an analysis team.
I would like to thank Igor, Osvaldo and Pedro for always being supportive and helping me during my Master's degree; Alexandra, Catarina, Ice, Joana, Marta and Zé, for always caring and my housemates, Catarina and Helena, for always being by my side.
None of this would be possible without my parents. Thank you for all the support and the effort that allows me to pursue my education in Physics. To my brother and sister for being there. Finally, I want to thank Pedro for always being so proud of me. Thank you for all the motivation and care, for all the things that I can not put into words.

I thank LIP (Laboratório de Instrumentação e Física Experimental de Partículas), FCT/MEC (Fundação Ciência e Tecnologia / Ministério da Educação e da Ciência), FEDER (Fundo Europeu de Desenvolvimento Regional) for funding my activities, through BigDataHEP project POCI/01-0145-FEDER-029147, PTDC/FIS-PAR/29147/2017 financed by funds OE/FCT, Lisboa2020, Compete2020, Portugal 2020, FEDER and ATLAS CERN/FIS-PAR/0002/2019, financed by funds OE/FCT.

Statement of integrity

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Pesquisa por Matéria Escura em Eventos Monotop no Grande Colisionador de Hadrões

Resumo

O Modelo Padrão pode ser considerado uma aproximação a mais baixa energia de uma teoria fundamental, o que encoraja a procura por nova Física. Uma evidência que suporta a existência de novas partículas para além do Modelo Padrão surge de medidas astrofísicas que apontam para a existência de um tipo de matéria que não interage com a força eletromagnética, usualmente designada por Matéria Escura. Apesar de não ser esperado que as partículas associadas com a Matéria Escura interajam significativamente com detetores, as colisões protão-protão no Grande Colisionador de Hadrões podem produzir novas partículas que acoplam às partículas candidatas a Matéria Escura e às partículas do Modelo Padrão permitindo a deteção destes processos. As pesquisas com uma assinatura *monotop* procuram por eventos com um quark *top* e grande energia transversa em falta proveniente dos candidatos a Matéria Escura.

A quantidade de dados produzida no Grande Colisionador de Hadrões é enorme e complexa por isso o uso de técnicas de *Machine Learning* está-se a tornar mais usual neste contexto. O treino de 100 redes neuronais com hiperparâmetros diferentes foi feito e os limites foram calculados. Foram propostas métricas associadas a Física das Altas Energias. A comparação entre os hiperparâmetros, os valores das métricas e os limites é feita com o objetivo de determinar a sua relação.

Outro objetivo para esta tese é apresentar um estudo detalhado na pesquisa por Matéria Escura com uma assinatura *monotop* e também contribuir para a análise feita pela experiência ATLAS no Grande Colisionador de Hadrões, fazendo um estudo *Next-to-Leading Order* versus *Leading Order* na fenomenologia do sinal e produzindo as *JobOptions* para a geração de sinal.

Palavras-chave: aprendizagem automática, matéria escura, métricas, monotop.

Search for Dark Matter in Monotop Events at the Large Hadron Collider

Abstract

The Standard Model (SM) can be considered an approximation at a lower energy of a more fundamental theory which encourages the search for new physics. One evidence supporting that new particles beyond the SM might exist comes from astrophysical measurements that point to the existence of a kind of matter that does not interact with the electromagnetic force, usually referred to as Dark Matter (DM). Although the particles associated with this DM are not expected to interact significantly with detectors, proton-proton collisions at the Large Hadron Collider (LHC) can produce new particles that couple both to DM candidate particles and to SM particles allowing the detection of these processes. The searches with a monotop signature look for events with one top quark and large missing transverse energy from the DM candidates.

The quantity of data produced at the LHC is huge and complex, therefore the use of Machine Learning (ML) techniques is becoming more usual in this context. The training of 100 Neural Networks (NN) with different hyperparameters was done and the limits were calculated. High Energy Physics (HEP) related model performance measures were proposed. A comparison between the hyperparameters, the model performance metrics and the limits is done in order to determine their relation.

Another purpose is to present a detailed study on the search for DM with a monotop signature and also contribute to the analysis being done by the ATLAS experiment at the LHC, by doing a Next-to-Leading Order (NLO) versus Leading Order (LO) study on the signal phenomenology and producing the JobOptions for the signal generation.

Keywords: dark matter, machine learning, model performance metrics, monotop.

Contents

List of Figures

List of Tables

List of Acronyms

SM Standard Model of Particle Physics
DM Dark Matter
LHC Large Hadron Collider
PDG Particle Data Group
QED Quantum Electrodynamics
QCD Quantum Chromodynamics
BSM Beyond the Standard Model
CMB Cosmic Microwave Background
MACHO Massive Compact Halo Object
VLQ Vector-like quarks
FCNC Flavour-Changing Neutral Currents
LO Leading Order
NLO Next-to-Leading Order
CERN *Conseil Européen pour la Recherche Nucléaire*
ATLAS A Toroidal LHC ApparatuS
CMS Compact Muon Solenoid
ALICE A Large Ion Collider Experiment
LHCb Large Hadron Collider beauty
ID Inner Detector
IBL Insertable B-Layer
SCT Semiconductor Tracker
TRT Transition Radiation Tracker
ECAL Electromagnetic Calorimeter

HCAL Hadronic Calorimeter
LAr Liquid Argon
TileCal Tile Calorimeter
HEC Hadronic End-Cap Calorimeter
FCal Forward Calorimeter (FCal)
L1 Level-1
HLT High Level Trigger
RoI Region-of-Interest
CSC Cathode Strip Chambers
RPC Resistive Plate Chambers
HL-LHC High-Luminosity LHC
HEP High Energy Physics
HE-LHC High-Energy LHC
HGTD High-Granularity Timing Detector
DL Deep Learning
NN Neural Networks
ReLU Rectified Linear Unit
SGD Stochastic Gradient Descent
ROC Receiver Operating Characteristic
SIC Significance Improvement Characteristic
AMS Approximate Median Significance
LO Leading-Order
NLO Next-to-Leading Order
CSV Comma-Separated Values
UFO The Universal FeynRules Output
CL Confidence Level

Chapter 1

The Standard Model of Particle Physics and Beyond

In this chapter a brief review of the Standard Model of Particle Physics (SM) as well as its problems is presented. The phenomenology of the monotop events used in the search for Dark Matter (DM) at the Large Hadron Collider (LHC) is described.

1.1 The Standard Model of Particle Physics

The SM is a theory developed in 1960's [1] that describes elementary particles and their interactions, with exception of gravity. The elementary particles in the SM are divided into fermions and bosons.
The fermions are spin$-\frac{1}{2}$ particles that follow the Fermi-Dirac statistics and obey the Pauli exclusion principle. They are divided into leptons and quarks, the quarks interact strongly while the leptons do not, each organised in three generations with the first generation belonging to the lightest and most stable particles and with the second and third generations belonging to the heavier and mostly unstable particles that rapidly decay to other particles. Every generation contains two quarks, one of the up type (up (u), charm (c) and top (t)) and one of the down type (down (d), strange (s) and bottom (b)) and also a charged lepton (electron (e), muon (μ)

and tau (τ)) and a neutrino (electron neutrino (ν_e), muon neutrion (ν_μ) and tau neutrino (ν_τ)), the properties of these particles are described on Table 1.1. Additionally, every fermion has an associated antiparticle that has the same properties but symmetric electric charge.

	Generation	Symbol	Name	Mass	Electric Charge ($\|e\|$)
Quarks	1^{st}	u	Up	2.2 MeV	+2/3
		d	Down	4.7 MeV	-1/3
	2^{nd}	c	Charm	1.275 GeV	+2/3
		s	Strange	95 MeV	-1/3
	3^{rd}	t	Top	173.0 GeV	+2/3
		b	Bottom	4.18 GeV	-1/3
Leptons	1^{st}	e	Electron	0.51 MeV	-1
		ν_e	Electron Neutrino	< 2 eV	0
	2^{nd}	μ	Muon	105.7 MeV	-1
		ν_μ	Muon Neutrino	< 2 eV	0
	3^{rd}	τ	Tau	1.8 GeV	-1
		ν_τ	Tau Neutrino	< 2 eV	0

Table 1.1: Fermions from the SM and their masses and electric charges according to Particle Data Group (PDG) [2].

The bosons are particles that follow the Bose-Einstein statistics. The interactions between the elementary particles, described on the SM, are mediated by the gauge bosons that are spin-1 particles. The photon (γ) is the carrier of the electromagnetic force, it is massless and electrically neutral. The gluon (g) is also a massless and electrically neutral particle, responsible for mediating the strong interaction. Although the gluons are electrically neutral, they carry colour charge with eight combinations, they interact with themselves and only couple to

the strongly charged particles. The $W^{\pm}$ and Z bosons are the mediators of the weak interaction. The W^{+} and W^{-} bosons are electrically charged and massive, while the Z boson is also massive but electrically neutral. Some properties of the gauge bosons and the three interactions are summarised in Table 1.2.

Interaction	Mediator	Mass	Electric Charge ($\|e\|$)
Electromagnetic	Photon (γ)	0	0
Strong	Gluon (g)	0	0
Weak	$W^{\pm}$	80.4 GeV	± 1
	Z	91.2 GeV	0

Table 1.2: Gauge Bosons and correspondent interactions [2].

Beside these gauge bosons, there is a scalar boson on the SM, the Higgs boson, responsible for the masses of the fundamental particles, in the context of a mechanism exposed further in this document.
A scheme of the elementary particles and interactions of the SM is shown in Figure 1.1.

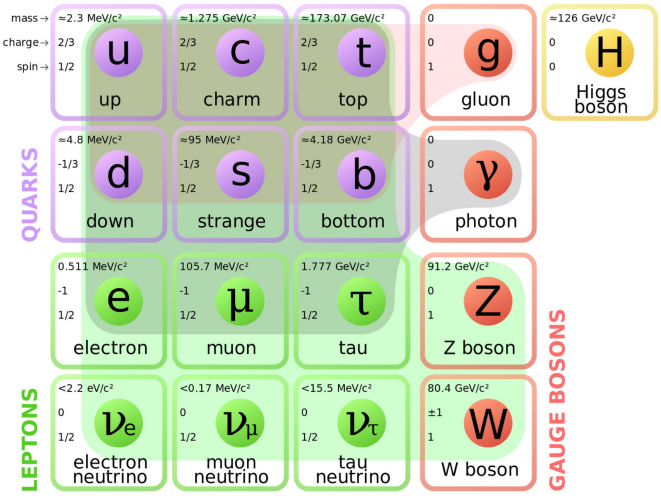


Figure 1.1: Particle content of the SM and their interactions. The shaded areas surrond each gauge boson and the fermions they interact with [3].

1.1.1 Quantum Electrodynamics

Quantum Electrodynamics (QED) is the quantum theory that describes the electromagnetic interaction. It describes not only the interaction of light with matter but also of charged particles with one another. The Dirac Lagrangian describes the free fermion

$$\mathcal{L}_{Dirac} = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi, \tag{1.1.1}$$

with $\bar{\Psi} = \Psi^\dagger\gamma^0$, and where γ^μ ($\mu = 0, 1, 2, 3$) corresponds to the Dirac matrices, Ψ is the Dirac field and m is the fermion mass. The Dirac matrices satisfy the following relations

$$\begin{aligned} \{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \\ \gamma_5 = \gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4, \\ \sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu] = \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu), \end{aligned} \tag{1.1.2}$$

where $g^{\mu\nu}$ is the metric tensor.

This Lagrangian is not invariant under a local $U(1)$ gauge transformation of the form $\Psi \rightarrow e^{-ie\alpha(x)}\Psi$ and $\bar{\Psi} \rightarrow e^{ie\alpha(x)}\bar{\Psi}$, where e is the electric charge and α is a real number. With this transformation the Dirac Lagrangian density is no longer invariant, since it acquires an additional term produced by the derivative $\bar{\Psi}e\gamma_\mu\partial_\mu\alpha\Psi$. The term $\bar{\Psi}e\gamma_\mu A_\mu\alpha\Psi$ is also added to the Lagrangian to mantain local gauge invariance, where a coupling between the A_μ field and the fermion was introduced.

Adding this term and the free field dynamics, described by the Maxwell equation, the QED Lagrangian is obtained

$$\mathcal{L}_{QED} = \bar{\Psi}[i\gamma^\mu(\partial_\mu - eA_\mu) + m]\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{1.1.3}$$

where A_μ is the gauge field and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the gauge field tensor. By imposing a $U(1)$ gauge invariance, the gauge boson has to be a massless and electrically neutral particle. Indeed, the photon is the gauge boson of QED, it has zero mass and no electric charge.

1.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory that describes the interaction between quarks and gluons. QCD and QED have a lot of similarities. In QED, the interaction between electrically charged particles is described, while QCD describes the interaction between particles that carry the QCD charge, known as colour. The colour comes in three varieties: red (R), green (G) and blue (B) instead of only one as the electric charge. The gluon in QCD plays a similar role to the photon in QED. However, while the photon is electrically neutral, the gluon is not colour neutral.

The gauge group of the QCD theory is the $SU(3)$. The quarks are represented as a triplet

$$\Psi = \begin{pmatrix} \Psi_R \\ \Psi_G \\ \Psi_B \end{pmatrix}. \tag{1.1.4}$$

The $SU(3)$ group has eight generators. These generators represent the eight gluons and are written as

$$t^a = \frac{1}{2}\lambda_a, \tag{1.1.5}$$

where λ_a $(a = 1, 2, ..., 8)$ are the Gell-Mann matrices. the Gell-Mann matrices are a set of eight linear independent 3×3 traceless Hermitian matrices and can be written as

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{1.1.6}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

The QCD Lagrangian can be written as

$$\mathcal{L}_{QCD} = \bar{\Psi}(i\gamma^\mu D_\mu - m)\Psi - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a, \tag{1.1.7}$$

with $D_\mu = \partial_\mu + i g_s t_a G^a_\mu$ that corresponds to the covariant derivative. The strength field tensor is defined by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu, \tag{1.1.8}$$

where G^a_μ corresponds to the gluon fields, g_s is the QCD gauge coupling constant and f^{abc} are the structure constants of $SU(3)$ defined by the commutation relation $[t^a, t^b] = i f^{abc} t^c$. The latter term corresponds to the gluon self-interaction term.
The QCD theory exhibits two main properties: asymptotic freedom and colour confinement. When a theory is said to have asymptotic freedom, it means that it is most strongly interacting at low energy scales, but at large scales it is weakly interacting. In this context, at very high energies and short distances, quarks and gluons interact weakly with each other, which makes QCD computable with perturbation theory. For low energies and large distances, the coupling becomes too strong for perturbation theory to be valid, not allowing quarks and gluons from being free, which is known as colour confinement.

1.1.3 Electroweak Theory

The electroweak theory is a unified theory proposed by Weinberg [4], Glashow [1] and Salam [5] which unifies the electromagnetic and weak interactions under the $SU(2)$ group in the use of a $SU(2)_L \bigotimes U(1)_Y$ gauge symmetry, where the L subscript refers to the fact that only left-handed particles interact and the Y subscript corresponds to hypercharge. The hypercharge is given by the Gell-Mann-Nishijima relation $Y = 2Q - 2T_3$ where Q is the electric charge, in terms of $|e|$

and T_3 is the third component of the isospin operator given by $\hat{T} = \frac{\sigma_i}{2}$, with σ_i corresponding to the Pauli matrices in which $i = 1, 2, 3$. The fermions appear as left-handed doublets and right-handed singlets

$$f_L^i = \begin{pmatrix} l_L^i \\ \nu_L^i \end{pmatrix}, \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}, \tag{1.1.9}$$

$$f_R^i = l_R^i, u_R^i, d_R^i,$$

with $i = 1, 2, 3$ being the fermion generation. To maintain gauge invariance, the covariant derivative is written

$$D_\mu \equiv \partial_\mu - ig\overrightarrow{T} \cdot \overrightarrow{W_\mu} - ig'\frac{Y}{2}B_\mu, \tag{1.1.10}$$

where g and g' are, respectively, the $SU(2)_L$ and $U(1)_Y$ coupling constants, $\overrightarrow{T}$ are the $SU(2)$ generators and $\overrightarrow{W}_\mu$ and B_μ are, respectively, the $SU(2)_L$ and $U(1)_Y$ gauge fields. The Lagrangian of the gauge fields is defined as

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \tag{1.1.11}$$

The $SU(2)$ triplet $W_{\mu\nu}^i$ with $i = 1, 2, 3$ and the $SU(2)$ singlet B_μ lead to the field strength tensors

$$\begin{aligned} W_{\mu\nu}^i &\equiv \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk}W_\mu^j W_\nu^k, \\ B_{\mu\nu} &\equiv \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned} \tag{1.1.12}$$

where W_μ^i and B_μ are, respectively, the gauge bosons of $SU(2)_L$ and $U(1)_Y$ and ϵ^{ijk} is the Levi-Civita tensor.
With this, it is possible to write down the Lagrangian of the electroweak theory

$$\mathcal{L}_{EW} = \sum_{f=l_L,q_L} \overline{f}(i\gamma^\mu D_\mu)f + \mathcal{L}_{gauge}. \tag{1.1.13}$$

In order to maintain gauge invariance, the gauge boson fields have to be massless. However, it is known from experimental evidences [6, 7] that the weak

interaction is mediated by heavy bosons ($W^{\pm}$ and Z bosons). To explain this fact, a spontaneous symmetry breaking mechanism which solves this inconsistency was introduced.

1.1.4 The Brout-Englert-Higgs mechanism

This mechanism, which explains the spontaneous break of the electroweak symmetry, was proposed in 1964 by three independent groups (Peter Higgs [8, 9, 10]; Robert Brout and François Englert [11]; Gerald Guralnik, C. R. Hagen and Tom Kibble [12, 13]) and it is known as the Brout-Englert-Higgs mechanism. This mechanism introduces a complex scalar field called the Higgs field.
The Higgs field is a scalar doublet with a hypercharge of $\frac{1}{2}$ and can be written as

$$\Phi \equiv \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} \tag{1.1.14}$$

where ϕ^{+} is the electrically charged field and ϕ^{0} is the electrically neutral field.

The Lagrangian that describes these fields is defined as

$$\mathcal{L}_{\Phi} = (D_{\mu}\Phi)^{\dagger}(D^{\mu}\Phi) - V(\Phi), \tag{1.1.15}$$

where the covariant derivative, D_{μ}, is given by Equation 1.1.10 and the Higgs potential, $V(\Phi)$, is defined as

$$V(\Phi) = -\mu^{2}\Phi^{\dagger}\Phi + \lambda(\Phi^{\dagger}\Phi)^{2}. \tag{1.1.16}$$

where μ^2 and λ are real and positive in order to generate the spontaneous symmetry breaking. This potential is represented in Figure 1.2 and it has a minimum along the line $|\Phi|^2 = \frac{\mu^2}{2\lambda} \equiv v^2$.

Around the vacuum expectation we parametrised Φ as

$$\Phi \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.1.17}$$

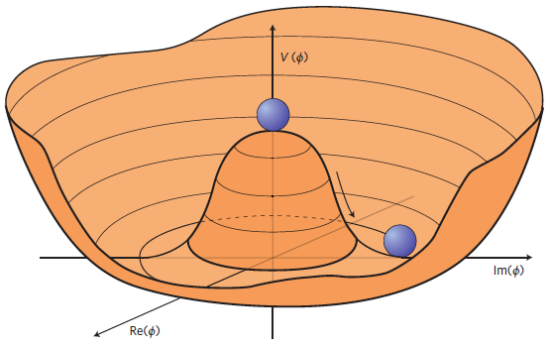


Figure 1.2: Higgs potential, also called "*El Sombrero*" or "Mexican hat" due to its shape [14].

where the term $H(x)$ represents the ground state fluctuations around the vacuum state.

The Lagrangian that describes the interaction between the Higgs field and the fermions can be written as

$$\mathcal{L}_{Yukawa} = -y_L \bar{L}\Phi l_R - y_d \bar{Q} d_R - y_u \bar{Q}\tilde{\Phi} u_R + h.c. \tag{1.1.18}$$

where sum is implied, y is the Yukawa coupling between the fermions and quarks with the Higgs field and L and Q are the lepton and quarks doublets, respectively. Through the Yukawa Lagrangian, the fermion masses and the Higgs boson mass can be obtained

$$\begin{aligned} m_f &= y_f \frac{v}{\sqrt{2}}, \\ m_H &= \sqrt{2\lambda} v. \end{aligned} \tag{1.1.19}$$

The mass of the Higgs boson could not be predicted since the λ parameter is unknow by theory. This can only be done experimentally. The electroweak boson masses can also be obtained through the Yukawa Lagrangian

$$m_Z = \frac{vg}{2},$$

$$m_W = v\frac{\sqrt{g^2 g'^2}}{2}, \qquad (1.1.20)$$
$$m_\gamma = 0.$$

1.2 Top Quark Physics

The top quark was discovered in 1995 by the CDF and DØ collaborations at Fermilab [15, 16] and it confirmed experimentally the three-generation structure of the SM. It is the heaviest elementary particle known and its phenomenology is driven by this fact. Some of the top quark's exciting properties are:

- Since it is heavier than a W boson, it is the only quark that decays semi-weakly (into a W boson and a b quark). This process has a branching ratio of approximately 1 [2];
- It has a very short lifetime ($\sim 5 \times 10^{-25}$ s [2]). Therefore, it is the only quark that decays before hadronisation can occur. All the other quarks hadronise on a timescale $\sim \frac{1}{\Lambda_{\mathrm{QCD}}}$, for that reason they cannot be observed free. Λ_{QCD} corresponds to the scale where the perturbatively-defined coupling would diverge;
- Its Yukawa coupling to Higgs boson is of the order of unit and it has the closest mass to the electroweak symmetry breaking scale ($v \sim 246$ GeV [2]). For this reason the top quark has a important role in diverse beyond the Standard Model (BSM) models.

At leading order in QCD and in hadron colliders, top quarks are predominantly produced in pairs through two different processes, shown in Figure 1.3: $q\bar{q}$ annihilation ($q\bar{q} \to t\bar{t}$; corresponding to 85% of the production cross section at Tevatron at $\sqrt{s} = 1.96$ TeV) and gluon-gluon fusion ($gg \to t\bar{t}$ corresponding to 90% of the production cross section at the LHC at $\sqrt{s} = 14$ TeV) [2].

The final states of the top quark pair production can be divided into [2]:

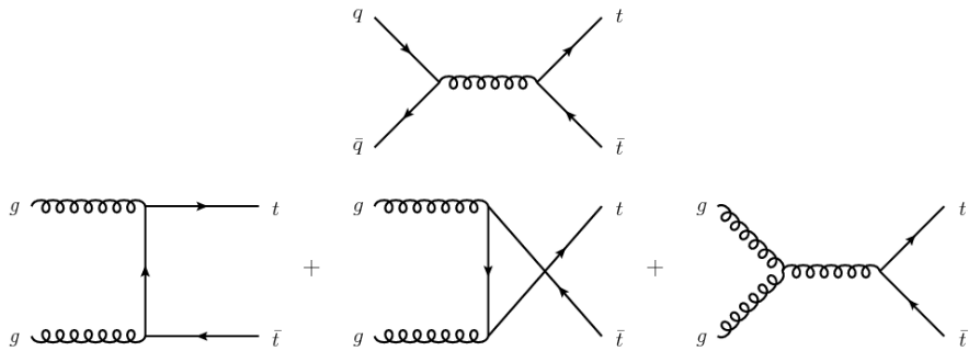


Figure 1.3: Leading Order Feynman diagrams corresponding to pair production of top quarks [17].

- **all-hadronic channel**: $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'\ bq''\ \bar{q'''}\bar{b}$: 45.7%;
- **semi-leptonic channel**: $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'\ b\ell^-\ \bar{\nu}_\ell\bar{b}\ + \ell^+\nu_\ell\ bq''\ \bar{q'''}\bar{b}$: 43.8%;
- **dileptonic channel**: $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu_\ell\ bl'^-\ \bar{\nu}_{\ell'}\bar{b}$: 10.5%.

However, the top quark can also be produced through a process called single top quark production, in which a single top quark in association with other particles is produced, leading to smaller cross-sections. This electroweak single top quark production can be a process mediated by virtual s-channel and t-channel W bosons or it can be a Wt production mechanism.
Considering the single top quark production we can have three different channels, represented in Figure 1.4:

- s-channel: W boson and $q\bar{q}$ annihilation;
- t-channel: W boson and gluon-gluon fusion;
- Wt-channel: production of a top quark in association with a W boson.

Figures 1.5 and 1.6 represent, respectively, the $t\bar{t}$ and single top production cross section measurements from both Tevatron and LHC.

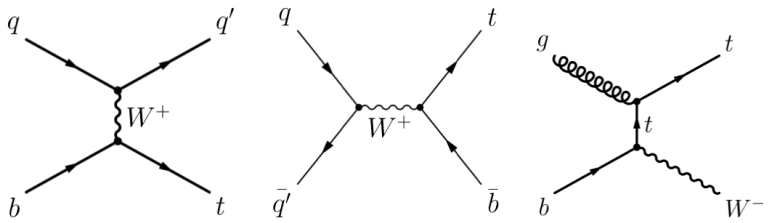


Figure 1.4: Representative Feynman diagrams of the most abundant and most studied single top quark processes at the LHC: t-channel, s-channel and Wt-channel [18].

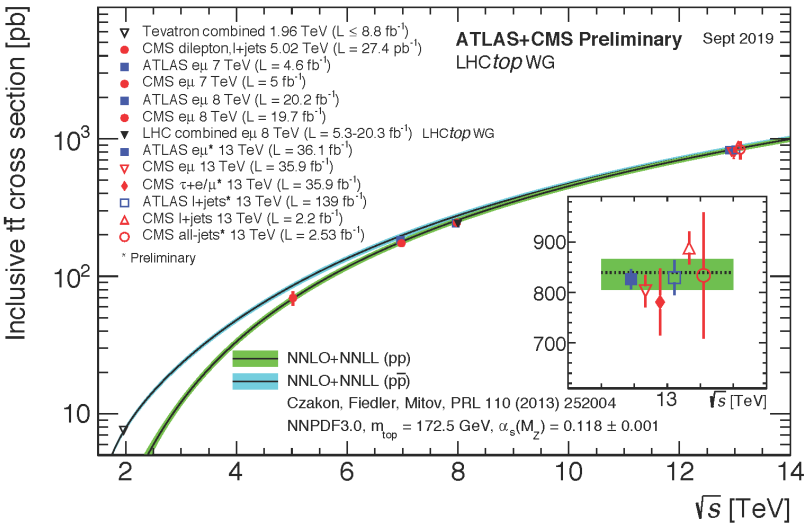


Figure 1.5: $t\bar{t}$ production cross-section measurements and predictions from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp collisions. The value of m_t assumed in the cross-section measurements and in the theory curves and uncertainties is $m_t = 172.5$ GeV/c^2 [19].

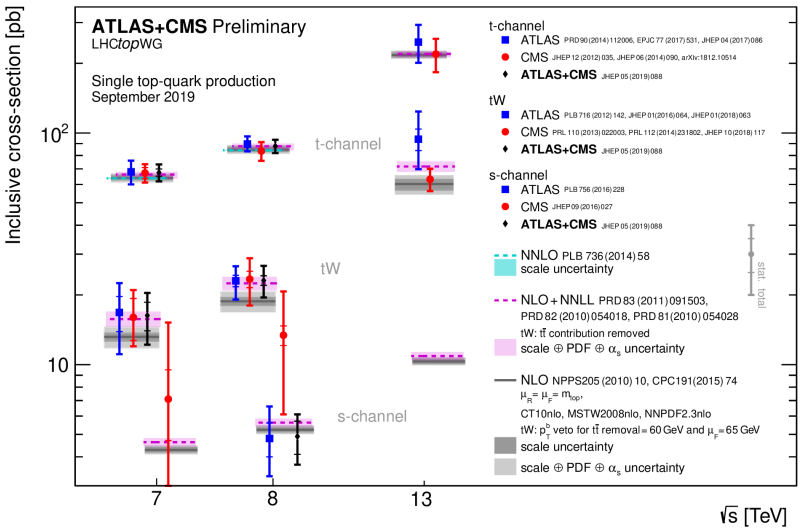


Figure 1.6: Single top production cross-section measurements and predictions from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp collisions [19].

A long quest began in order to better understand the properties of the top quark and its role in the SM.

1.3 Beyond the Standard Model

The SM is a well-tested model that explains successfully most of the present experimental results with high precision and also gives predictions that are in high agreement with the experimental observations. However, it leaves some phenomena unexplained such as gravity, the matter-antimatter asymmetry in the Universe, the neutrino masses, DM and dark energy. In this thesis, the focus will be the DM problem.

Fritz Zwicky, in 1933, measured the radial velocities of the galaxies in the Coma cluster and found out that the velocity dispersions were very large, indicating that the density of the cluster was bigger than the one derived from the luminous matter alone (with a mass-to-light ratio of about 500 [1] [21]).

Vera Rubin *et al.* and Albert Bosma measured the rotation curves of spiral galaxies and also found evidence for a missing matter that was of the non-luminous type [22, 23]. In Figure 1.7 the rotation curve of the NGC 3198 galaxy is shown. It is possible to infer that the velocity maintains constant instead of decreasing as was expected, confirming the theoretical work by Zwicky.

This two discoveries opened up the search for DM and, in the following years, its existence was inferred using cosmological observations from different sources.

In 2004 Clowe *et al.* found the Bullet cluster [25] and this discovery is considered the strongest indication of DM. In the outer region of the Bullet cluster, a greater concentration of mass is observed, while in the central zone, through the emission of X-rays, a higher concentration of gas is verified. The distortion of light by the effect of gravitational lens in the central region did not correspond to the mass observed, implying that there should be more matter in that area.

The cosmic microwave background (CMB) study provides another indications of DM existence since it gives us a precise map of the density mass in the early Universe. By analising the power spectrum, cosmologists were able to determine that our Universe is accelerating, spatially flat and its composition, which is rep-

[1] There were several issues with this estimate, such as the wrong Hubble parameter, the poor statistic and the cluster radius. The current value of the galaxy cluster mass-to-light ratio is 400. [20]

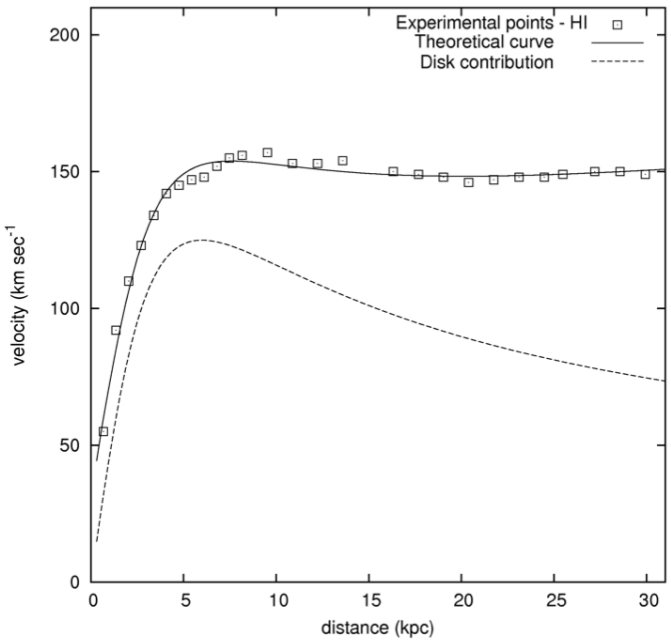


Figure 1.7: Observed rotational velocity versus distance from galactic center of the NGC 3198 galaxy. The dashed line is the contribution from local matter [24].

resented in Figure 1.8; baryonic mass corresponds to ordinary matter; cold dark matter corresponds to DM but it is called cold because it is slower when compared to the speed of light; dark energy is an unknown energy that causes the accelerated expansion of the Universe.

To explain the observations, the massive compact halo objects (MACHOs), *e.g.*, dark holes, neutron stars and brown dwarfs, were considered since they emit little or no radiation. However, Alcock *et al.* concluded that these objects account for only 15% to 20% of DM [26] so they are not enough to explain the observations. With this in mind, the simplest assumption is that DM is, at a microscopic level, a new kind of elementary particles that would not interact with the electromagnetic force and would have a large lifetime. The search for these particles can be done: directly by the observation of the SM particles recoil in collisions with DM particles; indirectly by the cosmological observation of DM particles annihilation products or at the LHC by analysing the SM particles collision products and looking for

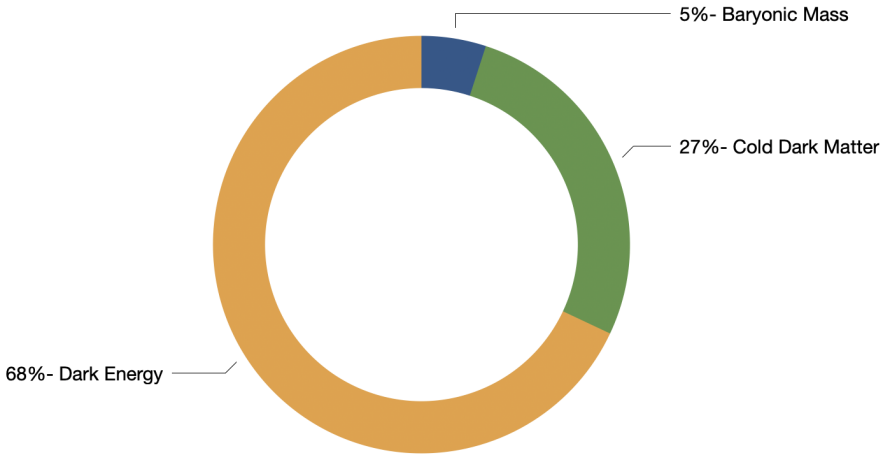


Figure 1.8: Universe composition. Planck 2015 results [20].

missing energy since this DM particles do not interact with matter.
There are several searches for DM at the LHC [27] and it is in this context that the search for DM with a monotop signature appears [28], being one very specific case of these searches.

1.3.1 Dark Matter with a Monotop Signature

The proton-proton collisions at the LHC can produce new particles that couple both to SM particles and to DM candidates. In the events with a monotop signature, the production of invisible particles in association with single-top quark is searched for, which corresponds to a final state with one top quark and large missing transverse momentum.
In such an analysis [29], there is a search for two different signals:

- Single production of vector-like T quark;
- DM candidates produced in association with top quarks.

In this thesis the focus will be the production of DM candidates in association with top quarks.

Dark Matter Production in association with Top Quarks

Two channels, resonant and non-resonant case, for the production of DM candidates in association with top quarks were considered. In both cases there is a top quark- decaying into a W boson and a b quark- and a DM particle, in the final state. The W boson can decay into two leptons (leptonic channel) or into two quarks (hadronic channel).
In the non-resonant case, a top quark is produced via flavour-changing neutral currents (FCNC) originating a new vector particle (V), which decays invisibly into a pair of DM particles. The Feynman diagrams for this process are shown in Figure 1.9.

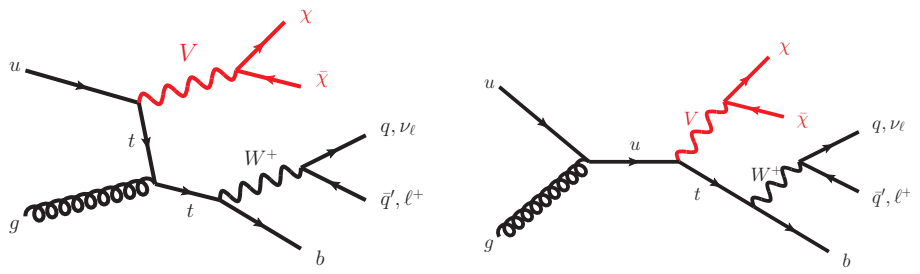


Figure 1.9: Leading order Feynman diagrams corresponding to DM production in association with a top quark for the non-resonant case [29].

The resonant case, takes into account the production of a coloured $\frac{2}{3}$ charged scalar (ϕ) decaying to a top quark and a spin-$\frac{1}{2}$ DM particle (χ). The Feynman diagram for this process is shown in Figure 1.10.

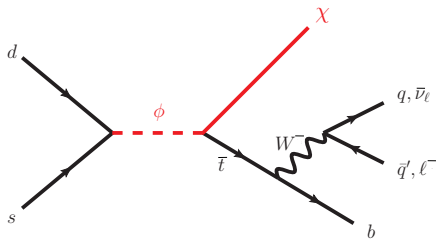


Figure 1.10: Leading order Feynman diagram corresponding to DM production in association with a top quark for resonant case [29].

Since, the resonant case consists in the simplest model featuring monotop production and the hadronic channel has a very clear signature, this was the focus of this thesis. This resonant case is described by the following Lagrangian [29]:

$$\mathcal{L}_{int} = \lambda\phi\bar{d}^{c}P_{R}s + y\phi\bar{\chi}P_{R}t + h.c., \tag{1.3.1}$$

where the λ parameter corresponds to the coupling of the charged scalar to the d and s quarks; the y parameter corresponds to the coupling of the charged scalar to the top quark and the DM particle, χ; the superscript c corresponds to the charge conjugation, $\Psi^{c} = -i\gamma_{2}\Psi*$, and the term P_R is the right-handed chirality projector defined as

$$P_{R} = \frac{1}{2}(1 + \gamma_{5}), \tag{1.3.2}$$

with $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$.

The production of the mentioned final state is very supressed in the SM, so its observation would be a clear sign of BSM physics. The data collected at the LHC experiments is very high in quantity and complexity so there are several tools used and perfected for its study. The use of deep learning techniques, which allow the handling of higher-dimensional and more complex problems, in the study of the LHC collisions is becoming very common. Therefore, having a study on the improvement of these techniques for HEP studies along with the search for DM at the LHC is very relevant.

First, in Chapter 2, an overview of the experimental setup is done. The detectors of ATLAS and CMS collaborations are described, as well as the LHC upgrade. Next, the ML concepts used in this thesis are depicted in Chapter 3. Then, in Chapter 4, the signal and background event generation is described as well as a study between the signal generation at LO and NLO. Finally, the results for the ML techniques associated with HEP are detailed in Chapter 5.

Chapter 2

Experimental Setup

In this chapter, the experimental setup is presented. An overview of the LHC is given, and the two general-purpose detectors are described. The major update that the LHC will undergo is discussed as well as the package used for the fast simulation of a generic general-purpose detector which will be used for the results od this thesis.

2.1 Large Hadron Collider

The LHC [30] is the largest particle accelerator in the world. It is located at the *Conseil Européen pour la Recherche Nucléaire* (CERN), in Geneva, Switzerland. Figure 2.1 shows the accelerators and the detectors currently working at CERN. The LHC consists of a 27 km ring of superconduting magnets and it is located in an underground tunnel with a depth range of 45 m to 170 m. Inside the tunnel, two high-energy proton beams travel in opposite directions in separate beam pipes (two tubes kept at ultrahigh vacuum) which are accelerated until reaching a speed close to the speed of light. They are guided around the accelarator ring by a strong magnetic field maintained by superconducting electromagnets. The magnetic dipoles operate at a temperature of -271.3° C. For this reason much of the accelerator is connected to a distribution system of liquid helium, which cools the magnets.

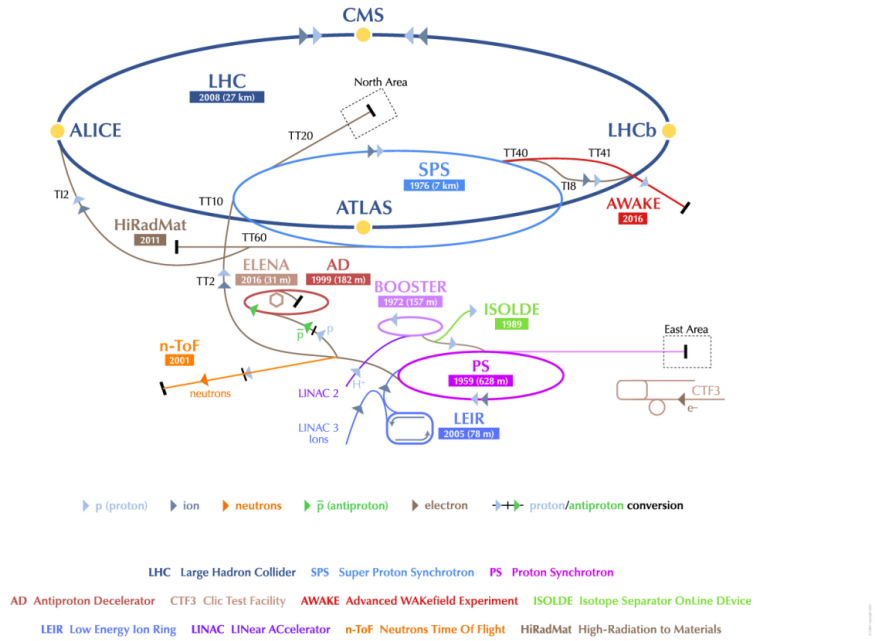


Figure 2.1: Schematic view of the CERN accelerator complex. [31].

The beams inside the LHC are made to collide at four locations, corresponding to the location of four particle detectors: A Toroidal LHC ApparatuS (ATLAS) [32], Compact Muon Solenoid (CMS) [33], A Large Ion Collider Experiment (ALICE) [34] and Large Hadron Collider beauty (LHCb) [35].
The aim of the LHC is to prove the SM and reveal the physics beyond it. The number of events, N_{exp}, generated in LHC collisions for a given process is given by

$$N_{exp} = \sigma_{exp} \times \int L(t)dt \ , \tag{2.1.1}$$

where σ_{exp} is the cross section of the event under study and L is the luminosity, which depends only on the beam parameters. Considering that we have two bunches with n_1 and n_2 particles colliding with a frequency equal to f_b, which

corresponds to the frequency of bunch crossing, the luminosity can be written as

$$L = f_b \frac{n_1 n_2}{4\pi \sigma_x^* \sigma_y^*}, \tag{2.1.2}$$

where σ_x^* and σ_y^* characterise the cross-sectional dimmensions of the beam in the horizontal and vertical directions, respectively, at the point of intersection. A plot of the luminosity at the ATLAS and CMS collaborations is presented in Figures 2.2 and 2.3, respectively. Along the years, a larger delivered luminosity is achieved.

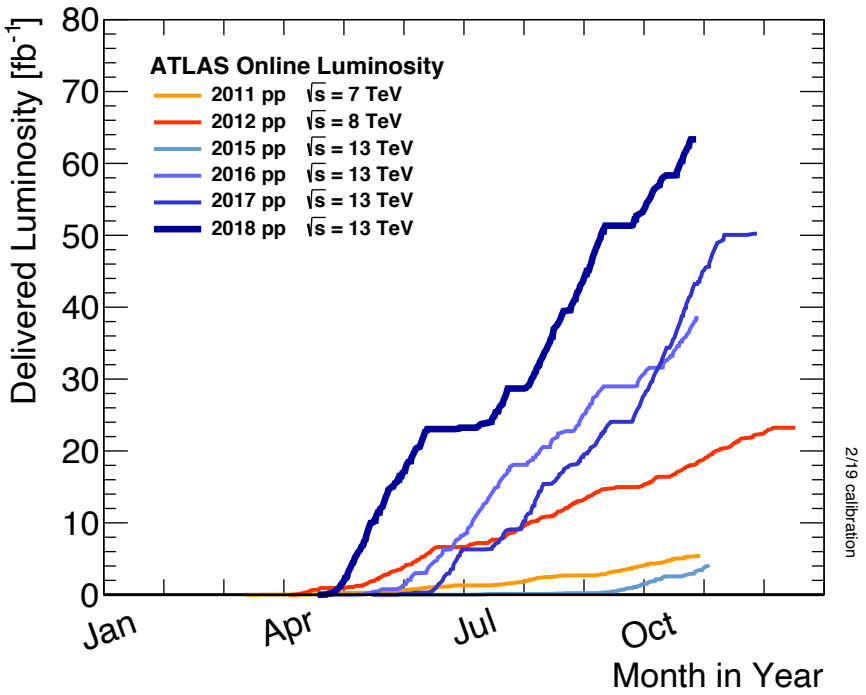


Figure 2.2: ATLAS delivered luminosity versus time for 2011-2018 (p-p data only) [36].

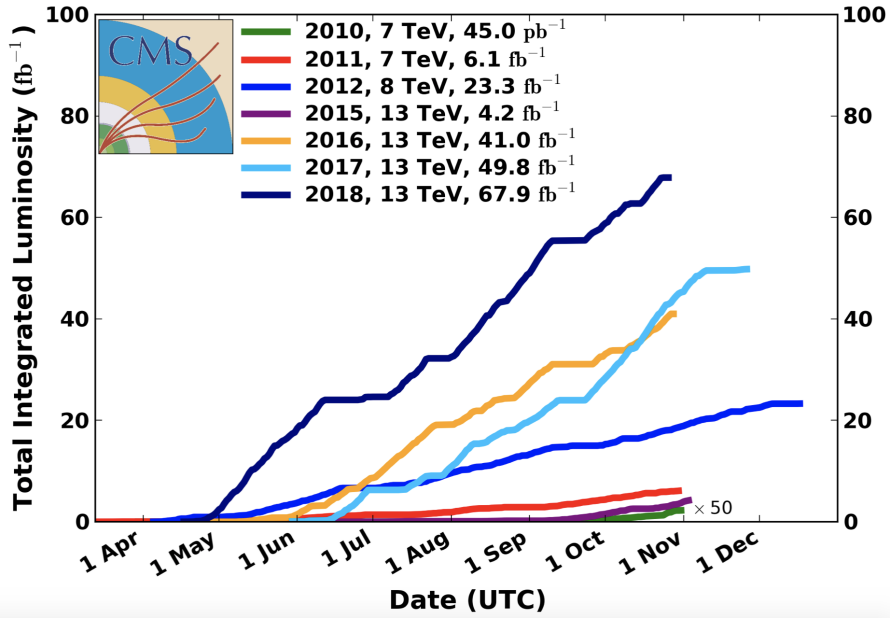


Figure 2.3: CMS integrated luminosity delivered (p-p data only) [37].

2.2 The ATLAS Detector

The ATLAS detector is one of the two general purpose detectors at the LHC, it is 44 meters long, 25 meters wide and weighs 7000 tonnes, as is shown in Figure 2.4.

The coordinate system is a right-handed Cartesian one, with its origin defined as the nominal interaction point. The z-axis is defined along the beam pipe and the xy plane is transverse to the beam direction. The x-axis is defined as pointing from the interaction point to the center of the LHC, the y-axis is defined as pointing upwards. The azimuthal angle ϕ is measured around the beam axis, while the polar angle θ is the angle between the z-axis and the xy plane. The rapidity is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2.2.1}$$

where E is the particle energy and p_z the z-component of its momentum. For

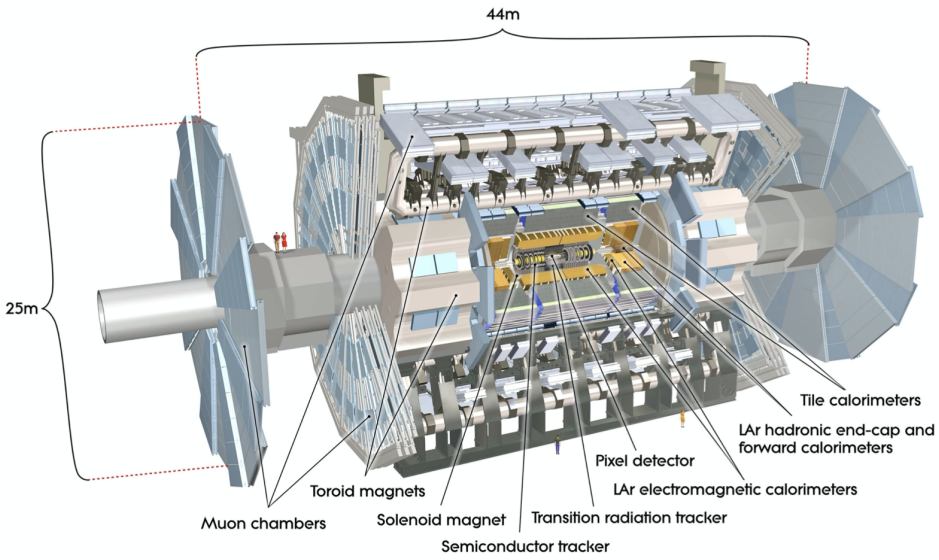


Figure 2.4: Cut-away view of the ATLAS detector [32].

massless particles it can be defined as

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) = \frac{1}{2}\ln\left(\frac{1+\theta}{1-\theta}\right) = \frac{1}{2}\ln\left(\frac{2\cos^2\frac{\theta}{2}}{2\sin^2\frac{\theta}{2}}\right) = \ln\left(\cot\frac{\theta}{2}\right), m = 0 \tag{2.2.2}$$

There is a simple mapping between rapidity and angle for massless particles. This motivates the definition of pseudorapidity η, written as a function of the polar angle

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.2.3}$$

The transverse momentum p_T is defined as

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.2.4}$$

Missing transverse energy, E_T^{miss}, is defined as

$$\vec{E_T^{miss}} = -\sum_i \vec{p_T}(i) \tag{2.2.5}$$

where i corresponds to the particles in the final state. This variable is of crucial importance in the work presented here.

Another variable used is the distance between two particles, ΔR, and it can be written as

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.2.6}$$

The ATLAS detector has several subsystems that are described next, starting from the inner region and their performance goals are represented in Table 2.1.

Detector Component	Resolution
Tracking	$\frac{\sigma_{p_T}}{p_T} = 5 \times 10^{-4} p_T \oplus 0.01$
EM Calorimetry	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$
Hadronic Calorimetry	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$
Muon Spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV

Table 2.1: General performance goals of the ATLAS detector [32]. The units of p_T and E are in GeV.

2.2.1 Inner Detector

The inner detector (ID), presented in Figure 2.5, is the inner subdetector of ATLAS. Since this subdetector is very close to the interaction point where the beams collide, it must deal with a high density of particle tracks. The ID constitutes a tracking system for charged particles which allows momentum measurements and reconstruction of interaction vertices.

It operates embebed in a magnetic field peaking at 2 T provided by a solenoid which bends the particle trajectories in order to measure its momentum. It has full coverage in the azimuthal angle and an acceptance in pseudorapidity of $|\eta| < 2.5$. The ID has three main components: the insertable B-Layer (IBL), the pixel

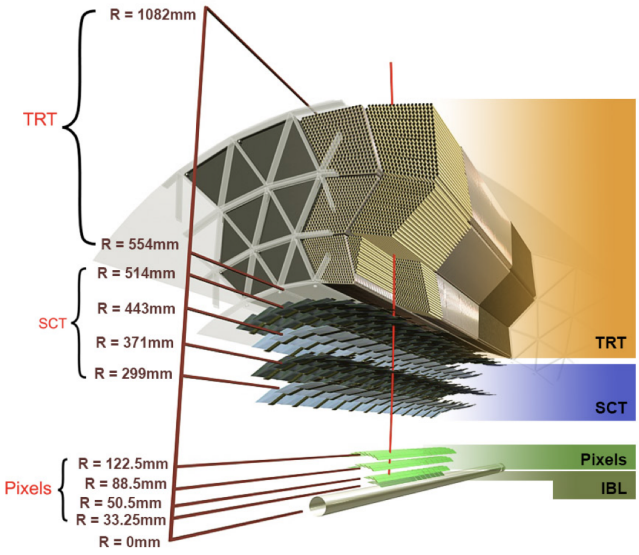


Figure 2.5: Cut-away view of the ATLAS inner detector with all the components labeled [38].

detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT).

The innermost part of the ID consists of a high-granularity silicon pixel detector and the IBL. The pixel detector spans the radial region of $33 - 150$ mm. The IBL is the innermost pixel layer and consists of 14 starves covering the region $|\eta| < 3.03$. The SCT is a silicone microstrip detector, arranged in 4 concentric cylinders around the beam axis and 2 end-caps formed by 9 disk layers each. The TRT consists of straw tube detectors with 4 mm in diameter arranged parallel to the beam axis and provides additional information on the particle type.

2.2.2 Calorimeters

The ATLAS calorimeter system, shown in Figure 2.6, provides the measurement of the particles' energies and their topology. It also determines the missing transverse energy which is produced by particles that escape the detector such as neutrinos.

The calorimeter system is composed of an electromagnetic calorimeter (ECAL)

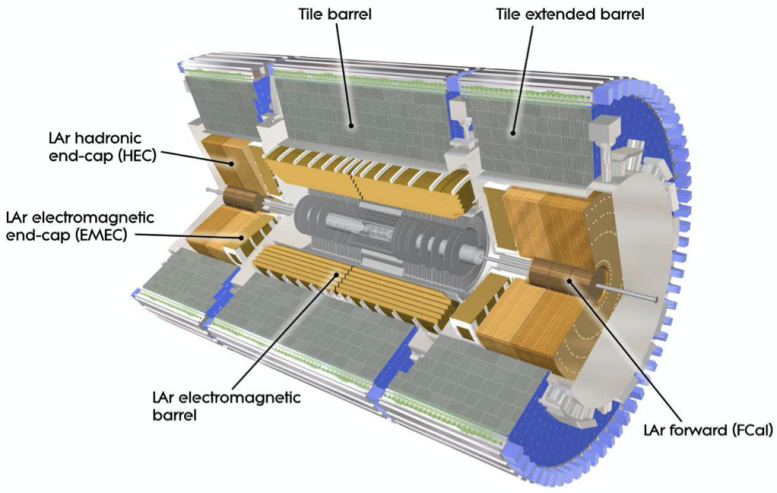


Figure 2.6: Cut-away view of the ATLAS calorimeter system [32].

followed by a hadronic calorimeter (HCAL). The ECAL is the detector layer in which the electromagnetic interacting particles, such as leptons and quarks, leave energy deposits. It is divided into a barrel part ($|\ \eta\ | < 1.475$) and 2 end-cap components ($1.375 < |\ \eta\ | < 3.2$). In the barrel region, the central solenoid and the ECAL share a common vacuum vessel. It consists of two identical half-barrels, separated by a small gap (4 mm) at $z = 0$. The ECAL uses liquid argon (LAr) as an ionisation medium, with lead absorbers arranged in an accordion geometry, which naturally provides a full coverage in ϕ without any gaps.

Surrounding the ECAL is the HCAL, which has the following components: the tile calorimeter (TileCal), the hadronic end-cap calorimeter (HEC) and the forward calorimeter (FCal). The HCAL measures the energy deposits from the hadron showers of high energy hadrons such as protons and neutrons. The TileCal is placed right outside the ECAL envelope. It covers the central region $|\ \eta\ | < 1.7$ and it uses steel as the absorber and scintillating tiles as the active material. The HEC covers the range $1.5 < |\ \eta\ | < 3.2$. It is built with copper acting as the absorber material and with LAr acting as the active material, in which there are electrodes collecting ionisation charges from the hadronic showers. The FCal is installed in the very forward region and it provides a coverage over $3.1 < |\ \eta\ | < 4.9$ in order to

improve the measurement of the missing transverse energy. LAr was again chosen as the active material, while for the absorber material copper and tungsten were chosen.

2.2.3 Muon Spectrometer

Besides the neutrinos, muons are the only particles that are not contained by the calorimeters. The muon spectrometer is the outermost detector, it is represented in Figure 2.7 and it is designed to detect these particles and to measure their momentum in the pseudorapidity range $| \eta | < 2.7$. It is based on the magnetic deflection of muon tracks by a complex arrangement of toroid magnets. The large barrel toroid provides the magnetic bending over the range $| \eta | < 1.4$, for the region $1.6 < | \eta | < 2.7$ two smaller end-cap magnets inserted into both ends of the barrel toroid bend the muon tracks and, in the transition region, $1.4 < | \eta | < 1.6$, a combination of barrel and end-cap fields is responsible for the magnetic deflection. This system is the largest tracking system in ATLAS since it extends from a radius of 4.25 m around the calorimeters out to a radius of 11 m which corresponds to the full radius of the detector.

2.2.4 Trigger System

The ATLAS trigger system consists of: a hardware-based Level-1 (L1) trigger and a software-based high level trigger (HLT). Each level refines the decisions made by the previous level and applies additional selection criteria. The L1 trigger searches for high transverse momentum muons, photons, electrons, jets and τ-leptons decaying into hadrons, and it also searches for large missing and total transverse energy. This level takes up to 2.5 μs and taking as input coarse granularity calorimeter and muon detector information determine Regions-of-Interest (RoIs) in the detector. The RoI information includes the geographical coordinates in η and ϕ, the type of feature identified and the criteria passed. This information is used by the HLT.

The HLT takes the previous RoI as input. It uses full granularity detector infor-

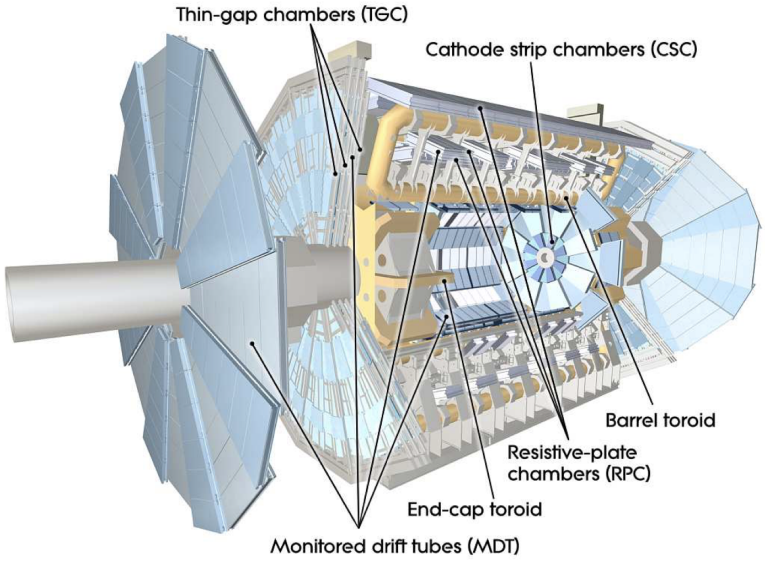


Figure 2.7: Cut-away view of the ATLAS muon spectrometer [32].

mation to run sophisticated selection algorithms. The HLT reduces the rate from the L1 output rate to around 1 kHz in approximately 200 ms.

In order to have a more clear vision, we assume that the LHC is producing around 1 billion events per second. In the ideal scenario, all the events could be recorded to disk in every second, however, due to CPU limitations, this number of events has to be reduced to 100 per second in order to be recorded and analysed. This reduction is done through the triggering system described above. Although we have limitations in the current electronics, the quantity of information produced by the LHC is huge, therefore the use of a tool capable of handling big data is needed. Machine Learning (ML) techniques are able to handle the huge amount and high dimensionality of data from collisions at the LHC. Therefore, the use of this kind of techniques is a good approach in order to study the events produced at the LHC.

2.3 The CMS Detector

The CMS detector has the same purposes as the ATLAS detector. It is 21.6 m long, has 14.6 m of diameter and a total weight of 12500 tonnes. A cutaway view is shown in Figure 2.8. The main distinguishing features of CMS are a homogeneous ECAL based on scintillating crystals and a high-field solenoid.

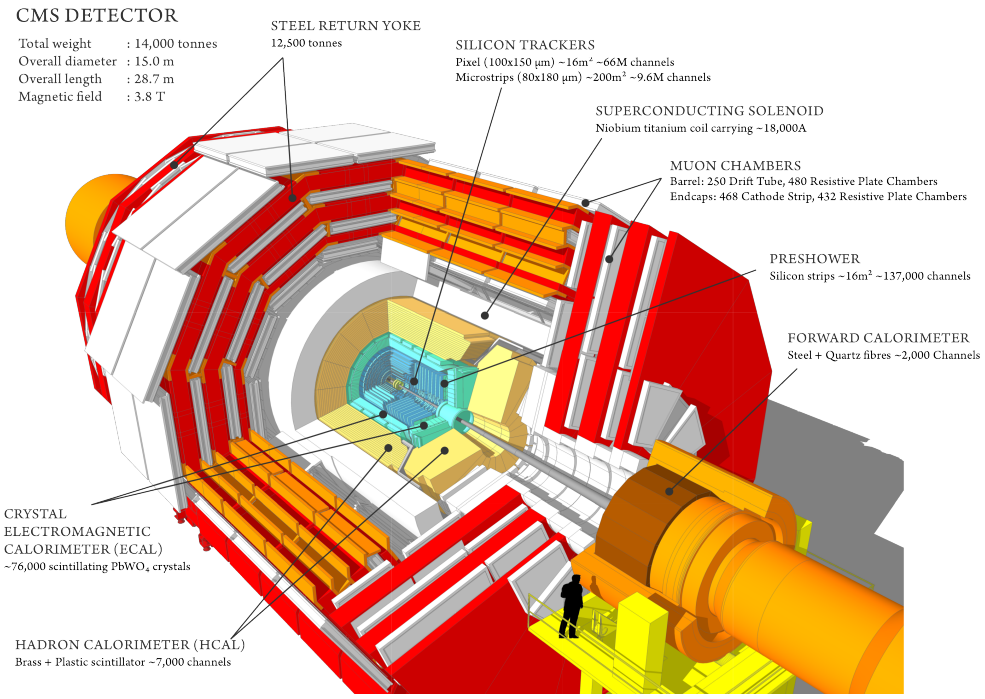


Figure 2.8: Cut-away view of the CMS detector [39].

The CMS detectors performance goals are represented in Table 2.2.

2.3.1 Superconducting Magnet

The main element of the CMS detector is a superconducting solenoid, with 12.5 m of length and 6.3 m of inner diameter. It was designed to reach a 4 T uniform field pointing along the z axis. The coordinate system of the CMS detector is defined in the same way has the ATLAS detector, with a stored energy of 2.6 GJ at full current.

Detector Component	Resolution
Tracking	$\frac{\sigma_{p_T}}{p_T} = 1.5 \times 10^{-4} p_T \oplus 0.005$
EM Calorimetry	$\frac{\sigma_E}{E} = \frac{3\%}{\sqrt{E}} \oplus 0.3\%$
Hadronic Calorimetry	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 5\%$
Muon Spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV

Table 2.2: General performance goals of the CMS detector [33]. The units of p_T and E are in GeV.

Outside of the solenoid, the reflux is returned through a 10000 tonnes iron yoke composed by 6 endcap disks and 5 barrel wheels.

2.3.2 Inner Tracking System

The inner tracking system of CMS is designed to provide a precise and efficient measurement of the trajectories of charged particles and to enable a precise reconstruction of secondary vertices. It has a length of 5.8 m, a diameter of 2.5 m and it surrounds the interaction point. Due to the high flux of particles, a detector technology with high granularity, fast response and radiation hardness is required. These requirements lead to a tracker design entirely based on silicon detector technology. It consists of a three layer silicon pixel detector with radii from 4.4 cm to 10.2 cm, and a silicon strip tracker with 10 barrel detection layers, extending out to a radius of 1.1 m. Each system is completed by end-caps: 2 disks in the pixel detector and 12 disks in the strip tracker, extending the tracker acceptance up to a pseudorapidity of $| \eta | < 2.5$.

2.3.3 Calorimeters

The ECAL of CMS is a hermetic homogeneous calorimeter made of lead tungsten ($PbWO_4$). It is composed by 61200 crystals in the central barrel part and it is closed by 7324 crystals in each of the two end-caps. Avalanche photodiodes and

vaccum phototriodes are used as photodetectors in the barrel and in the end-caps, respectively. The ECAL covers the pseudorapidity range of $| \eta | < 3.0$.
The HCAL is responsible for the measurement of hadron jets and neutrinos or exotic particles resulting in apparent missing transverse energy. The HCAL barrel extends from the ECAL, corresponding to a inner radius of 1.77 m, to the magnetic coil, corresponding to an outer radius of 2.95 m, covering the pseudorapidity range of $| \eta | < 3.0$. This constrains the total amount of material that can be put in order to absorb the hadronic shower. Therefore, a "tail catcher" is placed outside the solenoid. The forward HCALs placed at 11.2 m from the interaction point provide the pseudorapidity coverage to a range of $| \eta | = 5.2$. This calorimeter uses photomultipliers in order to detect the Cherenkov radiation produced by the forward particles travelling inside the quartz fibres.

2.3.4 Muon System

The muon identification is of central importance to CMS, and the muon system has 3 functions: muon identification, momentum measurement and triggering. The high field solenoidal magnet and its flux return yoke enable good muon momentum resolution and trigger capability. The flux return yoke also enables muon identification since it serves as a hadron absorber. The muon system has a cylindrical barrel section and 2 planar end-cap regions.
In the barrel section the muon rate is low and there is a 4 T magnetic field which is uniform and mostly contained in the steel yoke. The barrel drift tube chambers cover the pseudorapidity of $| \eta | < 1.2$ and are organised in 4 stations. The first 3 stations, each contain 8 chambers, in 2 groups of 4 which are separated as much as possible to provide the best angular resolution, measuring the muon coordinate in the $r\phi$ bending plane, and 4 chambers that provide a z measurement. In the fourth station there is no z measurement. In this station, some segments contain 2 chambers, resulting in a total of 70 chambers for this station, and 250 chambers for the whole muon system.
In the 2 end-cap regions of CMS the muon rate is high and the magnetic field is large and non-uniform. Here, the muon system uses cathode strip chambers (CSC).

The CSCs identify muons with a pseudorapidity coverage of $0.9 <| \eta |< 2.4$ and it has a fast response time, a fine segmentation, and a radiation resistance. The cathode strips of each chamber run radially outward and provide a precision measurement of ϕ.

Because the muon detector has a pseudorapidity coverage of $| \eta |< 2.4$ with no acceptance gaps, the offline reconstruction efficiency of simulated single-muon samples is above 95% for almost the entire range of pseudorapidity coverage, represented in Figure 2.9. The muon transverse momentum is measured independently by the inner tracker and the muon system, which enhances fault finding and permits cross-checking between the systems.

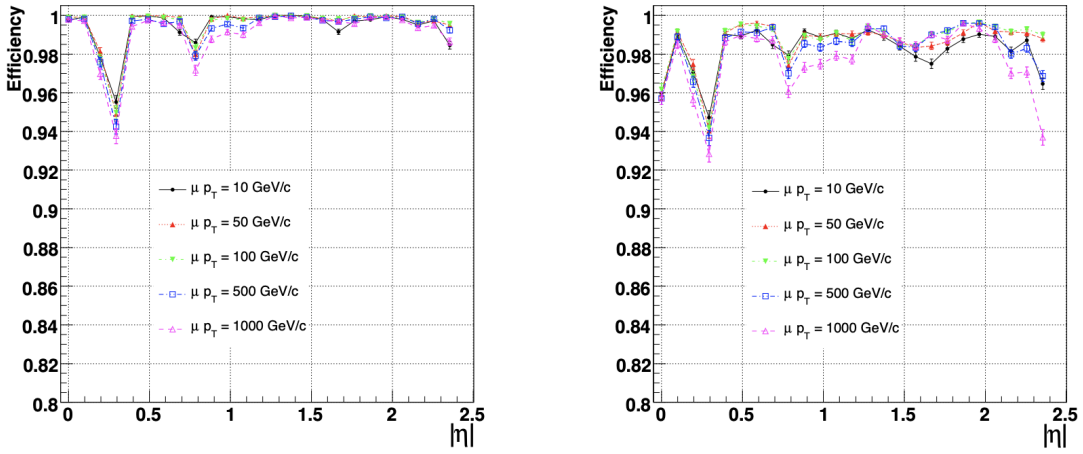


Figure 2.9: Muon reconstruction efficiency in function of the pseudorapidity range for selected values of p_T. In the left pannel the reconstruction using only hits from the muon system with a vertex constraint is shown. In the right pannel the reconstruction using hits from both the muon system and the tracker is shown [33].

A complementary trigger system was added in both the barrel and endcap regions and it consists in resistive plate chambers (RPC). They produce an independent fast response with good time resolution, but coarser position resolution than the barrel drift tube chambers or CSCs. A total of 6 layers of RPCs were placed in the barrel muon system: 2 in each of the first 2 stations, and 1 in each of

the last 2 stations. A plane of RCPs in each of the first 3 stations was introduced in the end-cap region. The placement of these trigger chambers improve the time resolution for bunch crossing identification and it uses the coincidences between stations to reduce background. Finally, in order to optimise the muon momentum resolution, a system measures the positions of the muon detectors with respect to the inner tracker and to each other.

2.3.5 Trigger System

The trigger system is comprised of an L1 hardware trigger and a HLT. The L1 trigger consists in field-programmable gate arrays and it uses coarsely segmented data from the calorimeters and the muon system, whereas the HLT is a software system which has access to the complete read-out data and therefore can perform complex calculations similar to the ones done in the analysis offline software.
The L1 trigger has local, regional and global components. The local triggers are based in energy deposits in the calorimeter trigger towers and track segments in muon chambers. Regional triggers combine their information and use pattern logic to determine electron or muon candidates in limited spatial regions. The global calorimeter and global muon triggers determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the global trigger. The L1 latency is 3.2 μs.
The rate reduction capability is designed to be at least a factor of 10^6 for the combined L1 trigger and HLT due to CPU limitations.

2.4 High Luminosity LHC

To extend the LHC discovery potential, it will undergo a major update, whose main objective is to increase its luminosity, to become the High-Luminosity LHC (HL-LHC) [40]. The HL-LHC will collide protons at 14 TeV center of mass energy and with an integrated luminosity of 3 ab^{-1}. Luminosity is an important indicator of the LHC performance and, as can be seen in equation 2.1.1, it is proportional to the number of collisions that occur in a given amount of time. Therefore, and in

addition to the upgraded detectors and expected improvements in the theoretical understanding, it will extend the sensitivity to rare processes. The quantity of data produced in the HL-LHC will be much larger than the quantity produced nowadays at the LHC. This upgrade will enhance the necessity for a tool capable of handling large datasets and solve complex problems. The increase on the particle rate will demand new tools for data processing, particle identification, offline analysis and other tasks. It is in this context that the use of ML is becoming essential in High Energy Physics (HEP) studies.

In order to delivery a more realistic result, a mix of approaches was used to assess the sensitivity in searching for new physics at the HL-LHC. Detailed-simulations were used to assess the reconstruction of objects performance. To estimate the sensitivity for the HL-LHC dataset, extrapolations of existing results were used. The reconstruction efficiency for the upgraded detectors was estimated employing fast-simulation in order to produce a large number of Monte Carlo events. To allow a full re-optimisation of the analysis selections, fast simulation which parameterises the detector effects is used. The expected performance of the upgrade ATLAS and CMS detectors has been studied in order to exploit the physical potential of this upgrade to the LHC [41, 42].

The harsher radiation environment and higher detector occupancies at the HL-LHC imply changes to the detectors. The detectors will also operate at a very high pile-up, which will be a major challenge for the experiments. The pile-up μ is the average number of events per bunch crossing and it can be written as

$$\mu(t) = \frac{\sigma L(t)}{f n_b}, \tag{2.4.1}$$

where σ is the cross-section, $L(t)$ is the instantaneous luminosity at time t, n_b is the number of colliding bunches and f the revolution frequency. Looking at the expression, the increase of luminosity will lead to an increase of the pile-up. The current pile-up is approximately of 36 number of interactions per crossing, the expected pile-up for the HL-LHC is around 140 number of interactions per crossing [43]. Pile-up produces more hits in the tracking detectors, leading to mismeasured or misidentified tracks and adds extra energy to the calorimeter measurements. It

can also influence the trigger system and the offline reconstruction and interpretation of events. Therefore, the detectors will have to be upgraded in order to be able to mantain their high performance.

2.4.1 The ATLAS Upgrade

Inner Tracker

Due to the expected high-pile environment at the HL-LHC, the inner tracker of the ATLAS detector will be completely replaced. It will improve the momentum resolution and also extend the pseudorapidity coverage from $\mid \eta \mid < 2.5$ to $\mid \eta \mid < 4.0$ and it will have a new silicon-only design. A silicon pixel detector composed of 5 barrel layers will be placed closer to the beamline [44] and a silicon strip detector with 4 barrel layers will extend allowing the tracking to higher radii [45].

Calorimeters

The ATLAS LAr calorimeter will have entirely new electronics optimised in order to withstand the radiation conditions and it is designed to output digitised signals of full granularity at 40 MHz [46].
The ATLAS TileCal will also have new electronics, power supplies and optical link interface boards to deal with the radiation conditions [47].

Muon Spectrometer

In order to maintain muon reconstruction perfomance, suppress the rate of random coincidences and increase trigger acceptance, additional muon chambers will be installed. Most of the muon spectrometer frontend and on- and off-readout and trigger electronics will also be replaced allowing higher trigger rates and longer latencies [48].

Trigger and Data Acquisition

ATLAS will use a single-level based trigger and data acquisition system as a baseline: a Level-0 trigger leading to a detector readout of 1 MHz and a maximum

latency of 10 μs that uses the calorimeter and muon spectrometer information and, in addition, the event filter reduces the output data rate to 10 kHz. Besides this, the processors will be complemented by additional ones in order to provide more background rejection.
If the pile-up conditions at the HL-LHC become a challenge, the system is designed to be capable of evolving to a dual-level hardware trigger: a Level-0 with a trigger rate up to 4 (2) MHz and a maximum latency also of 10 μs and a Level-1 with a trigger rate up to 600 (800) kHz and a maximum latency of 35 μs. In Level-1 hardware-based track reconstruction is implemented; these tracks are combined with calorimeter and muon spectrometer information to forme the Level-1 trigger decision [49].

High-Granularity Timing Detector

In order to make a precise measurement of the charged particles' timing, the ATLAS high-granularity timing detector (HGTD) will be installed having a pseudorapidity coverage of $2.4 <\mid \eta \mid< 4.0$. This installation will be made in front of the LAr calorimeter to reduce background from pile-up jets [50].

2.4.2 The CMS Upgrade

Inner Tracking System

Such as the ATLAS inner tracker, the inner tracking system of CMS will also be completely replaced. It will have an extended geometrical coverage with a pseudorapidity coverage up to $\mid \eta \mid= 4$, an improved radiation hardness, higher granularity and it will provide information on tracks (above a transverse momentum threshold) to the L1 trigger [51].

Calorimeters

The lead-tungsten crystals that are present in the ECAL will be cooled to a lower temperature. This provides the mitigation of noise in the avalanche photodiodes due to radiation damage. The front-end electronics will also be improved allowing

the exploitation of the information from single crystals in the L1 trigger.
In the HCAL, the hybrid photodiodes will be replaced by silicon photomultipliers. The scintillator tiles close to the beam line will also be replaced [52, 53].

Muon System

In order to increase radiation tolerance, readout speed and performance, the front-end electronics for the drift tube chambers and cathode strip chambers (CSCs) will be replaced with improved versions. New chambers based on gas electron multiplier technique and improved resistive plate chambers will enhance the muon system in the forward region, increasing the pseudorapidity coverage up to about $\mid \eta \mid = 2.8$ and improving the triggering and reconstruction performance [54].

Triggering System

The L1 trigger will be upgraded leading to an increase of L1 rate to about 750 kHz and to a latency of 12.5 μs. This level will also feature inputs from the silicon tracker, allowing real-time track fitting and particle-flow reconstruction objects.

2.5 Delphes Framework

Complex and sophisticated detectors are built in order to detect and precisely measure particles originated at the high energy collisions that occur at the LHC. Experimental colaborations rely on Monte-Carlo event generation for the definition of a specific analysis strategy and on tools that fully simulate a detector response, which reproduces the particle interactions with the matter content of the detector. Such studies that require a high level of accuracy are fully simulated using a GEANT4 package [55] which provides a simulation of the passage of particles through matter, namely through the particle detectors' material.
This complete framework requires expertise as well as computational resources that can only be handled by large collaborations. In phenomenological studies, such as the ones presented here, such level of complexity is not needed and a fast

detector simulation is often used. This is a simple approach that parameterises the response of the detector. The DELPHES framework [56] is one example of such a tool and it is used for the work presented here.
This framework simulates the response of a detector organised concentrically with a cylindrical symmetry around the beam axis and composed by: an inner tracker, electromagnetic and hadronic calorimeters and a muon system. There are several parameters that can be set by the user: the detector active volume, the calorimeter segmentation and the strenght of the uniform magnetic field. Each subdetector has a response described below.

2.5.1 Particle Propagation

The first step is the propagation of the particles resulting from the physics processes in the inner tracker. Neutral particles have a straight trajectory, while charged particles have a helicoidal trajectory until they reach the calorimeter. Tracks are assumed to have a perfect angular resolution. Its efficiency, energy and momentum resolution can be specified by the user and depends on the particle type, p_T and η. Particles originated outside the tracker volume are ignored.

2.5.2 Calorimeters

In DELPHES, the calorimetry is also performed by two calorimeters: the ECAL and the HCAL. The calorimeters are segmented in a rectangular grid in (η, ϕ) and the size of these rectangles can be defined by the user and can depend on η. The segmentation is uniform in ϕ and the same granularity is assumed for both ECAL and HCAL. Each long-lived particle that reaches the calorimeters has parameters f_{ECAL} and f_{HCAL}, corresponding to the fraction of its total energy which is deposited in the ECAL and HCAL, respectively. In DELPHES, by default, photons and electrons have $f_{ECAL} = 1$ since they leave all of their energy at the ECAL; hadrons have $f_{HCAL} = 1$ since it is assumed that they leave all of their energy at the HCAL, except for Kaons and Λ particles for which $f_{ECAL} = 0.3$ and $f_{HCAL} = 0.7$. Although it is assumed for hadrons to have $f_{HCAL} = 1$, in a real

detector, stable hadrons deposit a significant fraction of its energy at the ECAL. Muons and neutrinos do not deposit energy in the calorimeters. The default setup can be changed by the user .
A straight line coming from the point of interaction crosses one ECAL cell and one HCAL cell in the same (η, ϕ) region. These pairs called calorimeter towers and, along with the tracks, are used for the object reconstruction, such as isolated electrons and photons, as well as jets and missing transverse energy.

2.5.3 Particle-Flow Reconstruction

The particle-flow approach aims to use the maximum amount of information from all the subdetectors in order to reconstruct the event. In DELPHES, an approach based on the tracking system and the calorimeters for the particle-flow even reconstruction is used. Below a certain energy threshold, the tracker resolution is better and above this threshold the calorimeter energy deposit is more reliable to estimate the momentum. In the particle-flow context, the information from the tracker is always preferred in order to estimate the charged particle momenta.
The particle-flow algorithm produces two sets of 4-vectors, particle-flow tracks and particle-flow towers, used as input for jets and missing transverse energy reconstruction. Each reconstructed track results in a particle-flow track. The total energy deposited in the ECAL and HCAL is defined by the parameters E_{ECAL} and E_{HCAL}, respectively, and the energy deposited in the ECAL and HCAL originating from charged particles for which the track has been reconstructed is defined by the parameters $E_{ECAL,trk}$ and $E_{HCAL,trk}$, respectively. Then, for each calorimeter tower, the energy deposits originating from particles with reconstructed tracks are subtracted. If $E_{tower} > 0$, a particle-flow tower is created with energy E_{tower} and with the direction of the tower (η, ϕ) coordinates.
This definition implies that the particle-flow tracks contain charged particles measured with good resolution, while the particle-flow towers contain a combination of neutral particles, charged particles with no reconstructed track, and additional excess deposits induced by the positive smearing of the calorimeters, measured with lower resolution. While this algorithm is very simple when compared to what

is actually required in real experiments, it is shown that it accurately reproduces the performance achieved at the LHC.

2.5.4 Object Reconstruction

Charged Leptons and Photons

Muons: For muons, the momentum resolution, which is a function of the muon p_T and η, and the probability of reconstruction can be defined by the user. This probability of reconstruction vanishes outside tracker acceptance and for muon momenta below a certain threshold in order to reject looping particles. The final muon momentum is obtained by a Gaussian smearing of the original 4-momentum, according to a defined resolution.
Isolation criteria for the muons is determined according the content in a cone with maximum ΔR of 0.3 and maximum p_T ratio between the cone and the isolated object of 0.2.

Electrons: Electrons are usually reconstructed using the combined information from the tracking system and the electromagnetic calorimeter. DELPHES avoids this necessity by parameterising the combined reconstruction efficiency as a function of the energy and η. The electron reconstruction efficiency also vanishes outside the tracker acceptance and below a certain energy threshold. For the electron energy resolution, a combination of the ECAL and tracker resolution is used, such that at low energies the tracker resolution dominates, while at high energies the ECAL energy resolution dominates.
Isolation criteria for the electrons is determined according the content in a cone with maximum ΔR of 0.3 and maximum p_T ratio between the cone and the isolated object of 0.1.

Photons: The photon reconstruction relies only on the ECAL. Photon conversions to e^+e^- pairs are neglected and electrons with no reconstructed track that reach ECAL are reconstruced as photons in DELPHES. Its energy is obtained by applying the ECAL resolution to the original photon. In this thesis photons were

not used.

Currently, DELPHES does not simulate fake rate for electrons, muons and photons. The fake rate parameterises the possibility of a certain object being misidentified as a lepton or a photon. An electron, muon or photon is said to be isolated if the activity in its vicinity is small enough. The reconstruction of a lepton or photon has to pass an isolation criterion. The isolation variable can be defined as:

$$I(P) = \frac{\sum_{i \neq P}^{\Delta R < R,\ p_T(i) > p_T^{min}} p_T(i)}{p_T(P)}, \tag{2.5.1}$$

where the denominator is the transverse momentum of the particle of interest, P. The numerator is the sum of transverse momenta above p_T^{min} and within a cone of $\Delta R < R$ around the particle of interest, except for P. The particle is said to be isolated if $I(P) < I_{min}$ is verified. The default values are $p_T^{min} = 0.1$ GeV, $R = 0.5$ and $I_{min} = 0.1$, these values can be set by the user.

Jets

The final states, in a hadron collider, are often dominated by jets. For this reason, an accurate jet reconstruction is crucial. This reconstruction can be performed using three different collections of objects: the generated jets corresponding to the long-lived particles obtained after parton-shower and hadronisation, in which no detector simulation nor reconstruction is taken into account; the calorimeter jets that use calorimeter towers as input; and the particle-flow jets that are the result of clustering the particle-flow tracks and particle-flow towers. DELPHES integrates the FASTJET package [57] which allows jet reconstruction with the most popular jet clustering algorithm and to set the corresponding parameter values. To avoid double-counting of particles, DELPHES automatically removes jets that have already been reconstructed as isolated electrons, muons or photons.
Jets were reconstructed using the anti-k_t algorithm [58], with $p_T > 25$ GeV and

with a radius parameter of $\Delta R = 0.4$. The large-R jets taken into account were reconstruced with a radius parameter of $\Delta R = 0.8$.

b and τ jets: The identification of jets that result from τ decays or the hadronisation of heavy flavour quarks is important in hadron collider experiments. Their identification is based in a parametric approach. A jet can become a potential b or τ jet candidate if its direction is within a certain ΔR cone relatively to a generated b or τ, respectively. The user can define the tagging efficiency, which is parameterised and gives the probability for the jet to be identified as a b or a τ. Jets coming from b quarks with $2.5 \leqslant \mid \eta \mid < 4$ and 10 GeV $< p_T <$ 500 GeV are tagged as b-jets with efficiency of 60%.

Missing Transverse Energy and Scalar Transverse Energy

Missing transverse energy, E_T^{miss}, is defined as the negative of the vectorial sum of the p_T of all detected particles and the scalar transverse energy, H_T, is defined as the scalar sum of the p_T of all detected particles.
DELPHES is able to estimate both E_T^{miss} and H_T from generator level information, pure calorimetric information or particle-flow objects.

In this work DELPHES with the HL-LHC card was used [59].

Chapter 3

Machine Learning

In this chapter, an overview of relevant ML concepts is done. ML is the science of computer algorithms that improve automatically through experience by building a mathematical model based on sample data in order to "learn from data". These algorithms are applied in a large variety of fields.

3.1 Types of Learning

ML approaches are divided into: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning, amongst others.
In supervised learning the dataset is the collection of labeled examples. The goal of this kind of algorithm is to produce a model that takes as input a feature vector $\mathbf{x}$ and outputs information that allows to deduce the label of this feature vector. An example of the use of this algorithm is the spam detection in email messages. Supervised learning was the approach used in this thesis for the task of separating signal from background.
Supervised learning can be grouped into classification and regression problems. The main difference between these two tasks is that the dependent attribute is categorical for classification and numerical for regression. Regression is used, for example, when provided with a dataset of houses and the goal is to determine their prices. The regression algorithm is going to attempt to estimate the map-

ping function from the input variables to numerical values that are continous and correspond to the price. Using the same dataset, if the goal is to find if the price of the house is above or below the retail price, we want a discrete output with only two possibilities. This problem is a classification problem in which the algorithm attempts to estimate the mapping function from the input variables to categorical variables that are discrete.

In unsupervised learning the dataset is a collection of unlabeled examples. The goal of this kind of learning algorithm is to create a model that takes as input a feature vector $\mathbf{x}$ transforming it into another vector or into a value that can be used to solve a practical problem. Among several applications, this algorithm is often used in clustering in order to, for example, detect groups of similar visitors in a blog. The algorithm does not receive information about which group a visitor belongs to, it finds those connections by itself.

In semi-supervised learning the dataset contains both labeled and unlabeled examples. The goal of this kind of algorithm is the same as the supervised one: use the dataset to produce a model that takes as input a feature vector $\mathbf{x}$ and outputs information that allows to deduce the label of this feature vector. Many unlabeled examples are used in order to find a better model. Photo-hosting services are a good example of the use of this algorithm. Several photos are uploaded and the algorithm recognises that the same person appears in several ones. The user only has to provide one label per person and the system is able to label every person in every photo.

In reinforcement learning the system is capable of perceiving the state of an environment as a vector of features. The machine performs actions that bring different rewards, or penalties, in order to learn by itself what is the best policy, with the goal of getting the most reward over time. A policy is a function that takes as input the feature vector of a state and outputs an optimal action to execute in that state. An example of the application of this kind of algorithm is when robots "learns" how to walk.

3.1.1 Logistic Regression

Logistic regression is a classification learning algorithm. In this case we want to model the predicted label y_i as a function of our feature x_i. Since y_i will only two possible values, it is neccessary to find a function whose codomain is $[0, 1]$. The standard logistic function (also known as sigmoid function) has such a property, so the logistic regression model is given by:

$$f_{\mathbf{w},b}(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{1 + e^{-(\mathbf{w}\mathbf{x}+b)}}, \tag{3.1.1}$$

where $\mathbf{w}$ is a vector with the same dimension as the input vector $\mathbf{x}$ and b is a real number. The goal is to find the optimal values $\mathbf{w}*$ and $b*$. This set of parameters defines the model that makes the most accurate predictions. The output $f(\mathbf{x})$ can be interpreted as the probability of y_i being positive within a choice of a threshold. To find the optimal $\mathbf{w}*$ and $b*$, the negative log-likelihood of our training set is minimised. The log-likelihood can be written as

$$\text{Log}L_{\mathbf{w},b} \stackrel{\text{def}}{=} \ln(L_{\mathbf{w},b}(\mathbf{x})) = \sum_{i=1}^{N} y_i \ln f_{\mathbf{w},b}(\mathbf{x}) + (1 - y_i) \ln(1 - f_{\mathbf{w},b}), \tag{3.1.2}$$

where N correponds to the number of labels.
This is a linear learner, which means it will only learn an optimal linear decision function.

3.2 Neural Networks and Deep Learning

Neural Networks (NN) are a ML model inspired by the biological neurons of the human brain [60] and it consists in stacked layers, each with a certain number of units or neurons. Deep Learning (DL) is a field of ML that uses multiple layers. To build and train an NN the essential ingredients are:

- The input data and respective targets;
- Layers that are combined into a network;

- The loss function which measures the performance of the NN;
- The optimiser to update the weights of the model.

The interaction between these objects is represented in Figure 3.1.

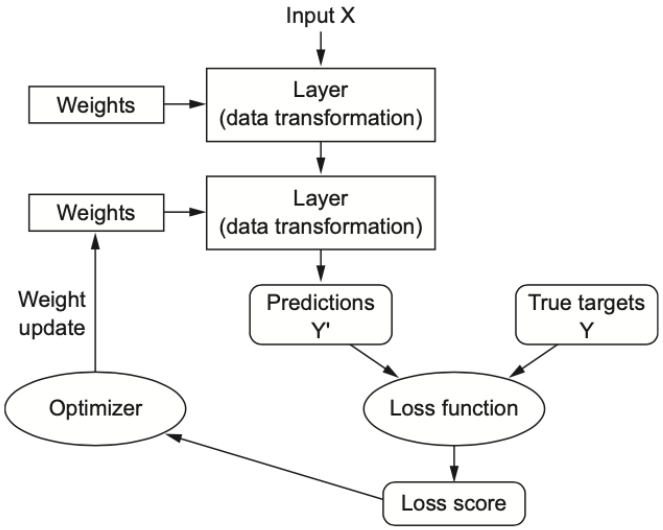


Figure 3.1: Pictogram of the interaction between network, layers, loss function, and optimiser [60].

Being able to apply a learning algorithm to a problem requires building the dataset first. A dataset is a collection of labeled examples $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ in which each element $\mathbf{x}_i$ among $\mathbf{N}$ corresponds to the feature vector. The feature vector contains values called features that are denoted as x_i. In this thesis, y_i has two possible values: 1 for the signal events and 0 for the background events since the goal of the NN created is to distinguish events of these two types.
After the transformation of the data in order for it to be readable, a partition of the dataset in the following three subsets should be done:

- training set;
- validation set;
- test set.

The training set is the one used to build the model. The weights of the different layers are initialised and the model generates predictions for the training set. The training loss score is computed and the optimiser updates the weights of the NN in order to minimise the loss score. This process is iterated a given number of times and an epoch corresponds an iteration through the whole dataset.
The validation and test sets are called the holdout sets and cannot be used to build the model. The validation set is used, at the end of each epoch, to compute a validation loss score which is used to perform early-stop and model selection. After the training of the NN is completed, the model obtained can be tested in a third independent set: the test set which purpose is to simulate the application of the model. This test set will be the one used to ascertain what we can expect from the actual data.
A layer is a data processing module that takes as input one or more tensors and outputs also one or more tensors. Different layers are used for different types of data processing. A dense layer is a frequently used layer for when the data is already a dense vector and it makes the following operation:

```
output = activation(dot(input, kernel) + bias),
```

which applies the activation function to the sum of the dot product between the input data and the weight data (kernel) with the bias that represents a biased value used in ML to optimise the model. The output of this kind of layer will be affected by the number of units specified. A linear model is too restricted to solve most problems. The deeper the NN, the more complicated will be the function that it can learn. In order to give a layer the ability of performing a non-linear transformation, an activation function is associated with it. In this thesis, dense layers were used with rectified linear unit (ReLU) as the activation function of the intermediate layers, and a sigmoid as the activation function of the final layer. The sigmoid function is given by Equation 3.1.1. The ReLU is defined as:

$$f(z) = \begin{cases} 0, & \text{if } z < 0. \\ z, & \text{otherwise.} \end{cases} \tag{3.2.1}$$

Both functions are represented in Figures 3.2 and 3.3. The ReLU is a function that returns zero for negative values, while the sigmoid function returns a value in the $[0, 1]$ interval for any arbitary value.

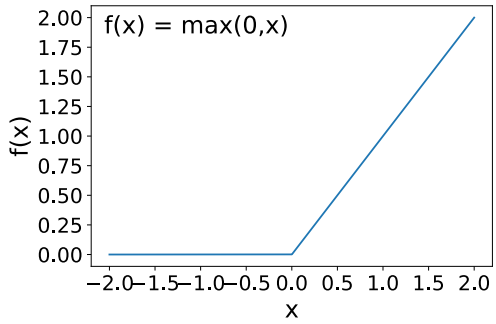


Figure 3.2: The ReLU function.

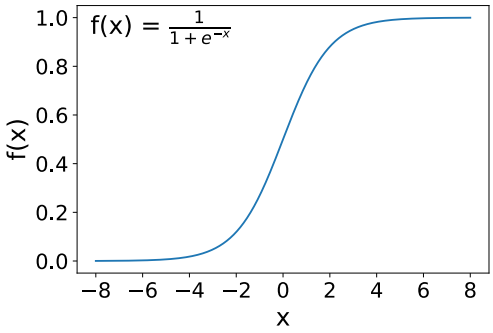


Figure 3.3: The sigmoid function.

When training an NN, two problems can arise: overfitting and underfitting. A model overfits when it predicts very well the traning data but poorly the new data. The dropout layer is a layer frequently used for regularisation techniques. This kind of technique is a widely used approach to prevent overfitting and consists in randomly dropping out a number of output features of the layer during training. The probability p of the neuron or unit being temporarily dropped out is called the dropout rate. Opposite to overfitting is underfitting which consists in the inability of the model to predict well the labels of the training set. To contour this problem a more complex model can be created. Additionally to a dropout layer, early-stopping methods can be applied to prevent overfitting. These methods consist in monitoring a quantity, in this case the validation loss, and stopping training if the quantity does not improve over a defined number of epochs.

A technique commonly used for training DNN is the batch normalisation which stabilises the learning process and reduces the number of training epochs required to train DNN. This technique consists of standardising the outputs of each layer before the units of the next layer receive them as input. Standardisation is the procedure that rescales the feature values in order to give them standard normal distribution properties with mean μ equal to 0 and standard deviation from the

mean σ equal to 1. The standardised values are calculates as

$$\hat{x}^{(j)} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}}, \tag{3.2.2}$$

where $\mu^{(j)}$ corresponds to the mean value and $\sigma^{(j)}$ to the standard deviaton.

Once the network architecture is defined, the loss function and the optimiser have to be chosen. The loss function is minimised during training and it represents a measure of success. The binary cross-entropy is used in a binary classification problem as the one presented here. The binary cross-entropy is defined as

$$\text{BCE} = -(y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)), \tag{3.2.3}$$

where $\hat{y}_i$ corresponds to the predicted probability of an example $\mathbf{x}_i$ belonging to class 1 and y_i is the label.

The minimisation of this function is the average of all binary cross-entropy terms across all training examples and all labels of those examples.
The optimiser determines how the network will be updated based on the loss function. To find the minimum of a differentiable function analitically, the derivative has to be performed. However, applied to an NN, a gradient descending algorithm can be used instead. Gradient descending is a generic algorithm used to find optimal solutions, finding the combination of weight values that yields the smallest possible loss function. This algorithm computes the local gradient of the loss function and moves the parameters in the opposite direction from the gradient. The optimiser used on this thesis was one that implements the Adam algorithm [61]. Adam optimisation is a stochastic gradient descent (SGD) method. This kind of method computes the gradient picking a random mini-batch in the training set at every step and using the mean of the instead of using the whole training set.
After the model is built, tuning the hyperparameters is an important step. A hyperparameter is a property of the learning algorithm that influences its performance. These hyperparameters are not learned by the algorithm, they have to be defined by the user. In this thesis, 100 random sets of the following hyperparameters were used:

- Number of layers;
- Number of units;
- Dropout rate;
- Batch normalisation.

All of them, with the exception of batch normalisation, are numerical hyperparameters. Batch normalisation when set to `True` performs the regularisation method described before.

3.3 Model Performance Measures

A metric function is used to assess the performance of the model on the validation set. The most widely used metrics to assess the performance are accuracy, precision, recall and area under the receiver operating characteristic (ROC) curve.

Accuracy

This metric is given by the ratio of of correctly classified examples divided by the total number of classified examples:

$$\text{accuracy} \stackrel{\text{def}}{=} \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{3.3.1}$$

where TP are the true positives, TN are the true negatives, FP are the false positives and FN are the false negatives.

Precision and Recall

Precision is the positive predictive value, it corresponds to the ability of the classifier not to label as positive a sample that is negative and it is given by:

$$\text{P} \stackrel{\text{def}}{=} \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3.3.2}$$

Recall is the true positive rate or sensitivity, it corresponds to the ability of the classifier to find all the positive samples and is given by:

$$\mathrm{R} \stackrel{\text{def}}{=} \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{3.3.3}$$

It is usual to define, with these two variables, the precision-recall curve which shows the tradeoff between precision and recall for every possible cut-off also referred to as threshold.
Average precision summarises a precision-recall curve as the weighted mean of precisions achieved at each threshold, with the increase in recall from the previous threshold used as the weight, and it is given by:

$$\mathrm{AP} = \sum_n (\mathrm{R}_n - \mathrm{R}_{n-1})\,\mathrm{P}_n, \tag{3.3.4}$$

where P_n is the precision at the nth threshold and R_n the recall at the nth threshold.

ROC

The ROC curve is a commonly used method to assess the performance of the classification models. It uses a combination of the true positive rate and the false postive rate defined as

$$\mathrm{TPR} \stackrel{\text{def}}{=} \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \qquad \mathrm{FPR} \stackrel{\text{def}}{=} \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}. \tag{3.3.5}$$

The better classifier is the one with the higher area under the ROC curve.

Besides these metrics commonly used in ML studies, metrics correlated with high energy physics (HEP) were implemented: the significance improvement characteristic (SIC) and the approximate median significance (AMS).

SIC

SIC can be written as

$$\text{SIC} = \frac{s}{\sqrt{b}}, \tag{3.3.6}$$

where s corresponds to the number of signal events and b to the number of background events. The area under the SIC curve was taken into account. SIC can be compared with ROC since it is computed per threshold.

AMS

The AMS [62] can be defined as

$$\text{AMS} = \sqrt{\sum_i \left\{ 2\left[(s_i + b_i)\ln\frac{s_i + b_i}{b_{0i}} - s_i - b_i + b_{0i} \right] + \frac{(b_i - b_{0i})^2}{\sigma_{bi}^2} \right\}}, \tag{3.3.7}$$

where i corresponds to the bin number, s_i and b_i to the number of signal and background events, respectively, in that specific bin, σ_{bi} corresponds to the background absolute error and b_{0i} is defined as

$$b_{0i} = \frac{1}{2}\left(b_i - \sigma_{bi}^2 + \sqrt{(b_i - \sigma_{bi}^2)^2 + 4(s_i + b_i)\sigma_{bi}^2} \right). \tag{3.3.8}$$

The AMS is an objective function used to determine a region in the feature space where an enhanced number of signal events is expected. The goal is to maximise this function since it corresponds to a significance in which the background-only hypothesis is rejected. If we take the limit in which $b \gg s$ we obtain Equation 3.3.6.

One of the goals of this thesis is to compare these metrics with the values obtained for the limits that will be introduced in the following chapters.

Chapter 4

Signal and Background Generation

In this chapter, the generation and simulation of both signal and background events is described. A study on the signal model is also presented.

4.1 Signal Generation

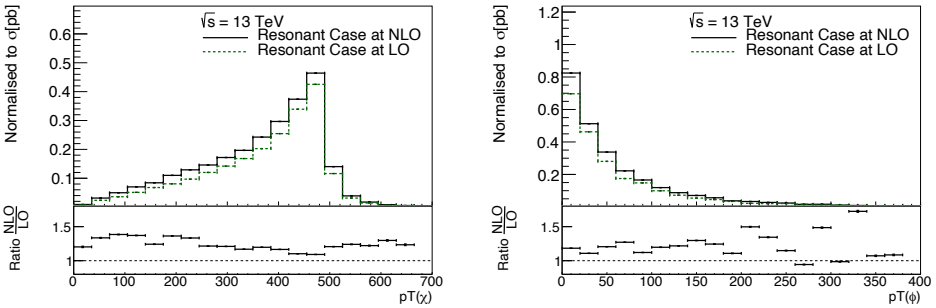
The generation of both signal and background simulated events was done using MadGraph5_aMC@NLO [63]. This tool is based on the Monte Carlo method and it is used for the generation of events at the LHC. The MadGraph5 software can be extended to incorporate: Pythia8 [64], used for simulation of parton showers and hadronisation, and Delphes [56], used for the detector simulation described in Section 2.5.
A leading-order (LO) versus next-to-leading order (NLO) study on the signal generation was performed. MadGraph5 can perform generation of simulated events both at LO and NLO. However, in this particular case, where the signal corresponds to the monotop resonant case, there is a further challenge in the NLO generation with the hadronisation phase at Pythia8 due to colour junction of the signal under study, which is beyond the normal capability of Monte Carlo generators. Therefore, obtaining NLO samples at particle level for this process cannot be

achieved without further steps. To contour this issue, NLO and LO samples with parton shower and without hadronisation were generated to be able to compare both samples. For a preliminary study, 10000 events of pp collisions at centre-of-mass energy of 13 TeV were generated with mediator mass, m_ϕ, set to 1000 GeV, dark matter particle mass, m_χ, set to 10 GeV, the couplings $\lambda = 0.2$ and $y = 0.4$ and with the fully hadronic top quark decay, $t(\bar{t}) \to W^+b$ ($W^-\bar{b}$). The following cross-sections were obtained: $\sigma_{\text{LO}} = 2.191 \pm 0.005$ pb and $\sigma_{\text{NLO}} = 2.591 \pm 0.014$ pb. The cross-section for the LO process is not within the uncertainty of the cross-section for the NLO process and it represents a significant increase in the predicted cross-section. The kinematics of both samples were also compared and are shown in Figure 4.1.

As referred, the generation of this process at NLO is not possible without further steps. Some of the information from an NLO calculation can be compacted in the K-factor [65]. The K-factor is used to normalise the LO calculations, and the theoretical one is usually defined as the ratio between the NLO and LO. At first order, the K-factor can be defined as

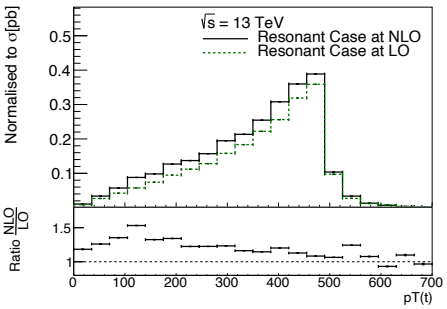
$$K = \frac{\text{NLO}}{\text{LO}} \tag{4.1.1}$$

Usually the K-factor can also be defined as the ratio between NLO and LO cross sections. However, the NLO corrections can result in a shape change, so that one K-factor is not sufficient to describe the impact of the NLO corrections on the LO cross-section. For this reason, since both distributions are normalised to their respective cross-sections, and there is a difference between the two samples in the cross-section values as well in the kinematics, the ratio plot can be used as the K-factor to rescale the LO generation.

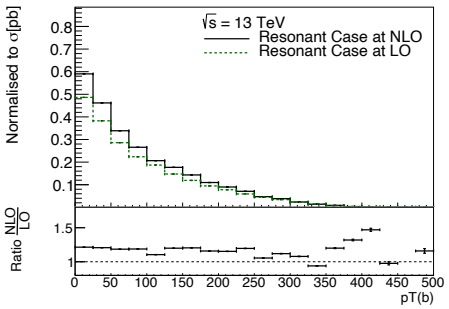


(a)

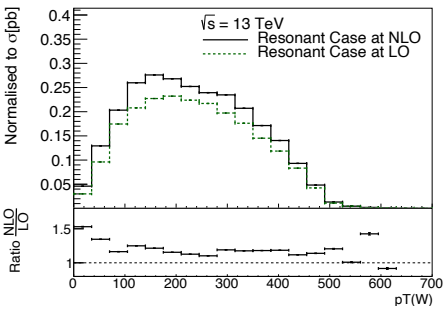
(b)



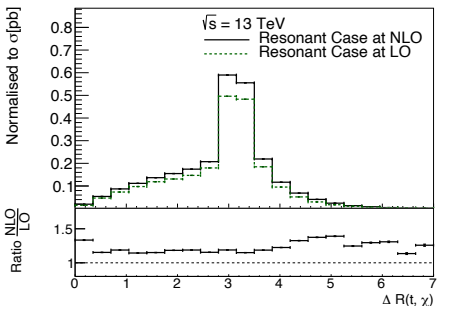
(c)



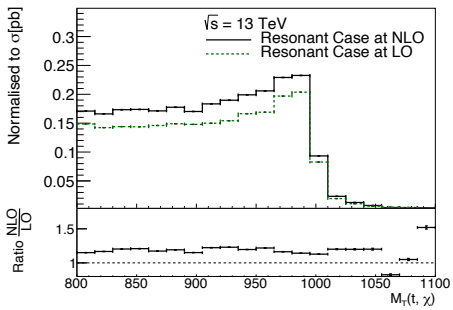
(d)



(e)



(f)



(g)

Figure 4.1: NLO and LO kinematic distributions of (a) DM particle, (b) mediator particle, (c) top quark, (d) bottom quark, (e) W boson, (f) ΔR between the top quark and χ particle and (g) transverse mass between the top quark and χ particle.

In order to obtain the K-factor values, a total of 500000 events were generated to increase statistics. The cross-section values, with more statistics, are the following: $\sigma_{\mathrm{LO}} = 2.194 \pm 0.011$ pb and $\sigma_{\mathrm{NLO}} = 2.597 \pm 0.030$ pb. Again, we notice that the cross-section of the LO process is also not within the uncertainty of the cross-section for the NLO process. The first variable used in order to obtain the K-factor values was the transverse momentum of the χ particle. The values for the K_{χ}-factor can be obtained from the ratio in the distribution represented in the Figure 4.2. These values were used to rescale the LO distributions.

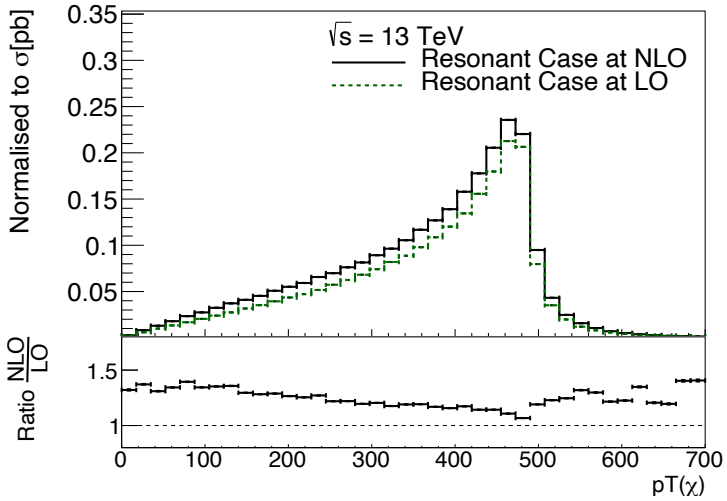


Figure 4.2: Transverse momentum of the χ particle. The values represented in the ratio correspond to the K-factors used to rescale LO distributions.

The distributions for the LO samples rescaled with these values are represented in Figure 4.3.

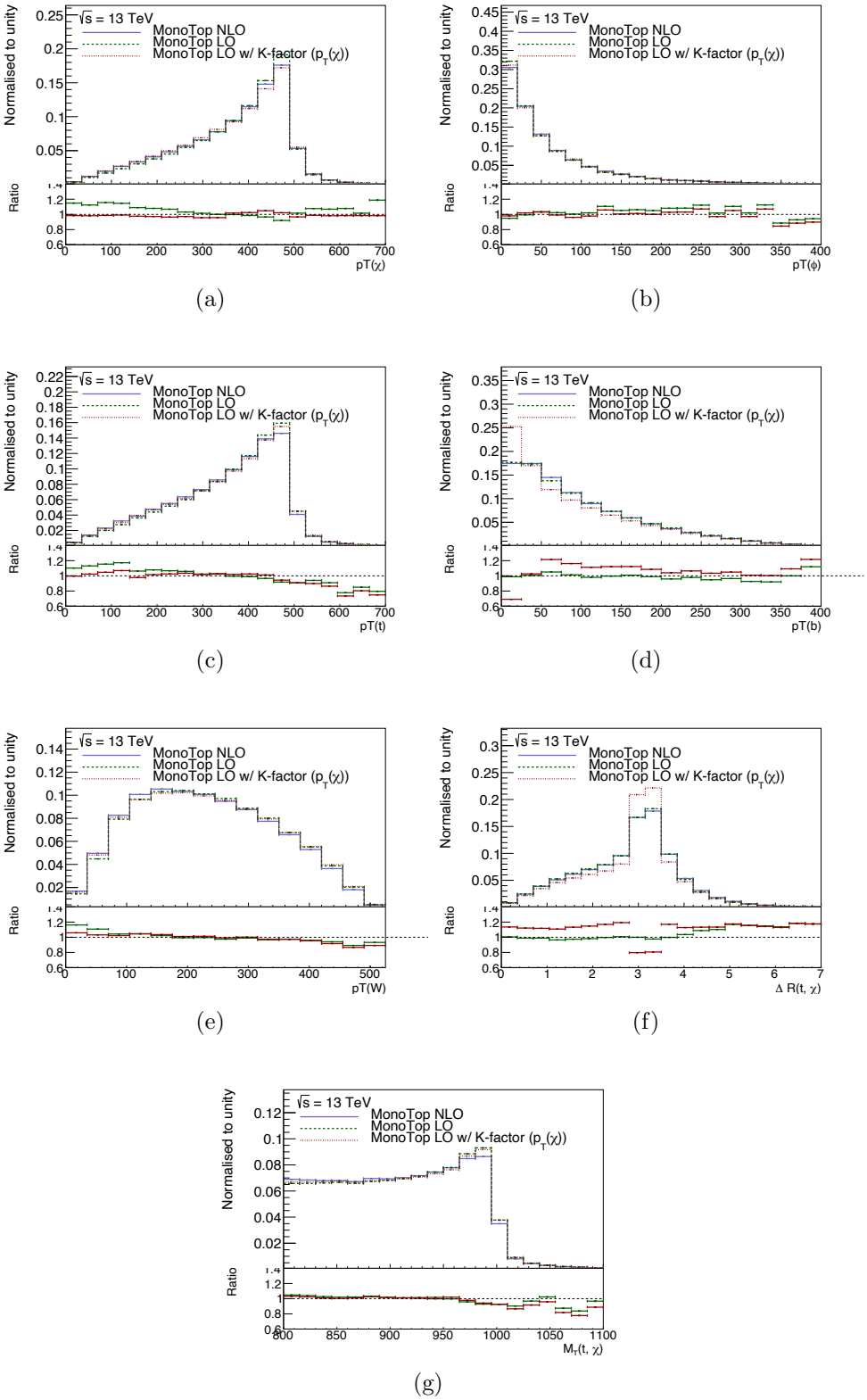


Figure 4.3: NLO and LO kinematic distributions, with LO samples rescaled to K-factor obtained from the $p_T(\chi)$ of (a) DM particle, (b) mediator particle, (c) top quark (d) bottom quark, (e) W boson, (f) ΔR between the top quark and χ particle and (g) transverse mass between the top quark and χ particle. All ratio plots correspond to $\frac{\text{NLO}}{\text{line colour}}$.

The ratio between NLO and LO samples rescaled with the K_χ-factors is atenuated in some distributions, however for others, mainly from the top quark decay products, the difference is not atenuated using the K_χ-factor to rescale the LO samples. In order to obtain new K-factor values, other variables were tested. The distribution used for the new K_ϕ-factor values is represented in Figure 4.4 and corresponds to the values in the ratio plot in the transverse momentum of the ϕ particle distribution.

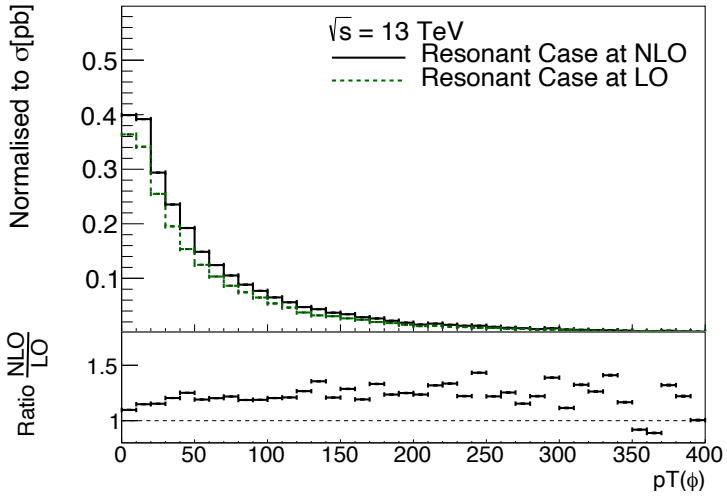
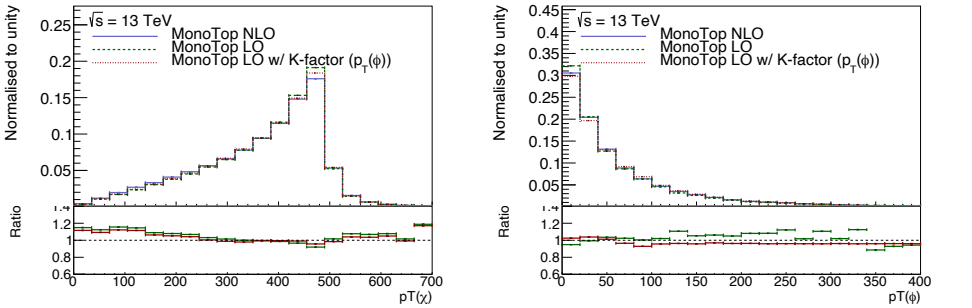


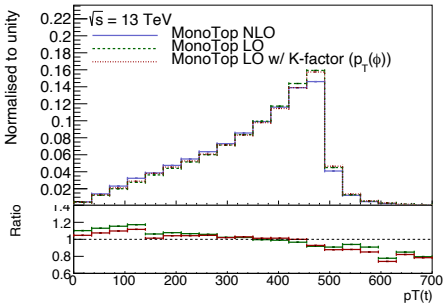
Figure 4.4: Transverse momentum of the χ particle. The values represented in the ratio correspond to the K-factors used to rescale LO distributions.

The distributions with the LO sample rescaled with these values for the K_ϕ-factors are represented in Figure 4.5.

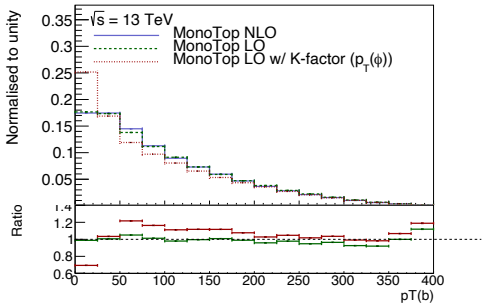


(a)

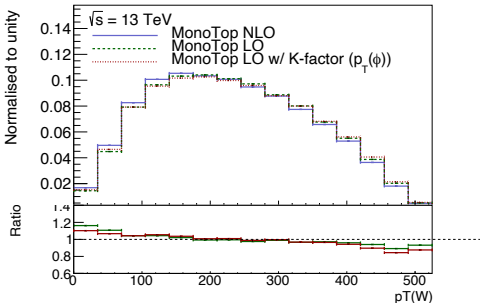
(b)



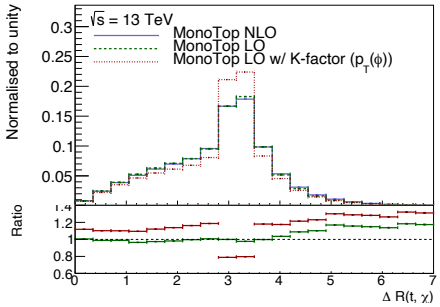
(c)



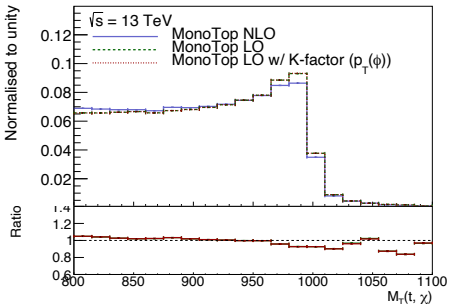
(d)



(e)



(f)



(g)

Figure 4.5: NLO and LO kinematic distributions, with LO samples rescaled to K-factor obtained from the $p_T(\phi)$ of (a) DM particle, (b) mediator particle, (c) top quark, (d) bottom quark, (e) W boson, (f) ΔR between the top quark and χ particle and (g) transverse mass between the top quark and χ particle. All ratio plots correspond to $\frac{\text{NLO}}{\text{line colour}}$.

The results for these values of the K_ϕ-factors are similiar to the ones obtained for the K_χ-factors values. Therefore, yet another variable was tested, the distribution of the top quark transverse momentum, which is represented in Figure 4.6.

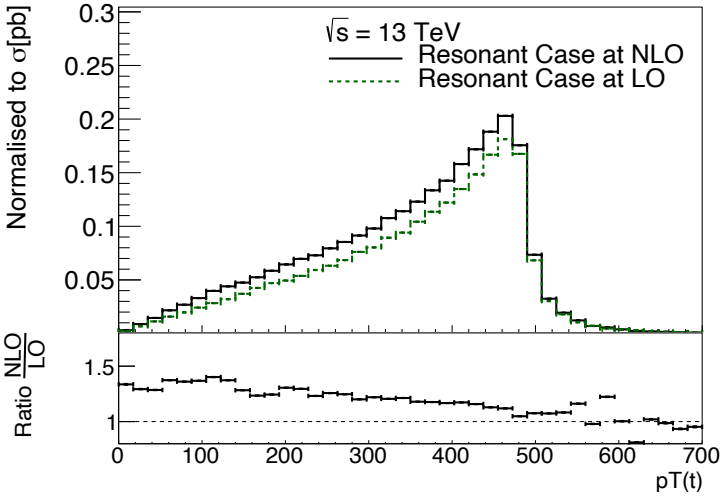
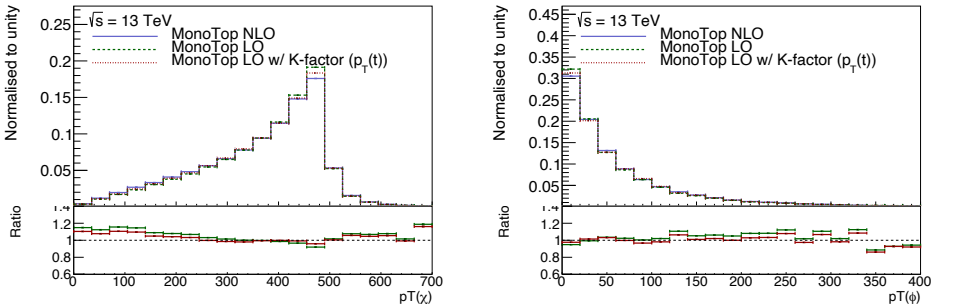


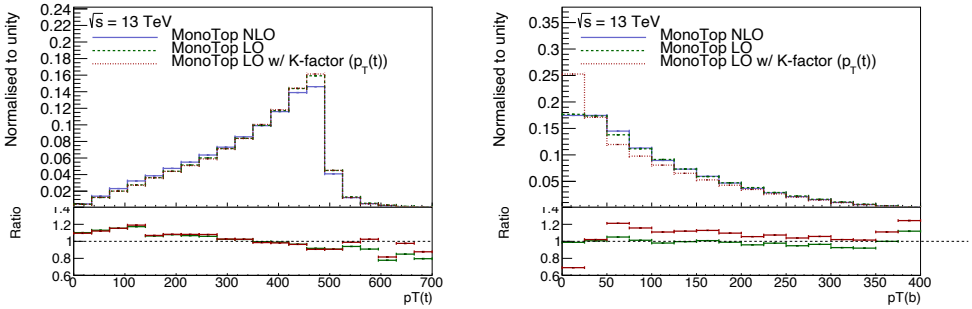
Figure 4.6: Transverse momentum of the top quark. The values represented in the ratio correspond to the K-factors used to rescale LO distributions.

The distributions with the rescaling using these values of the K_t-factors are represented in Figure 4.7.



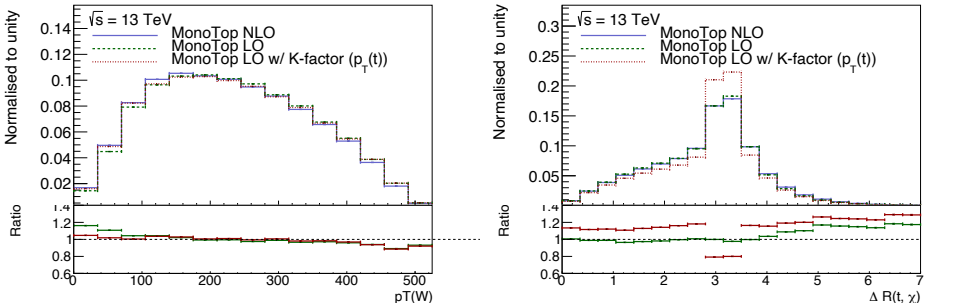
(a)

(b)



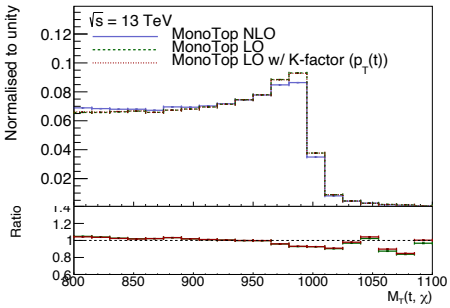
(c)

(d)



(e)

(f)



(g)

Figure 4.7: NLO and LO kinematic distributions, with LO samples rescaled to K-factor obtained from the $p_T(t)$ of (a) DM particle, (b) mediator particle, (c) top quark, (d) bottom quark, (e) W boson, (f) ΔR between the top quark and χ particle and (g) transverse mass between the top quark and χ particle. All ratio plots correspond to $\frac{\text{NLO}}{\text{line colour}}$.

The results are very similar to the other two K-factors, therefore K-factors using 2D distributions were obtained and used to rescale the LO samples. The $K_{\phi,t}$-factor values are represented in Figure 4.8 and were obtained from the distribution where the y-axis corresponds to the $p_T(\phi)$ and the x-axis corresponds to the $p_T(t)$.

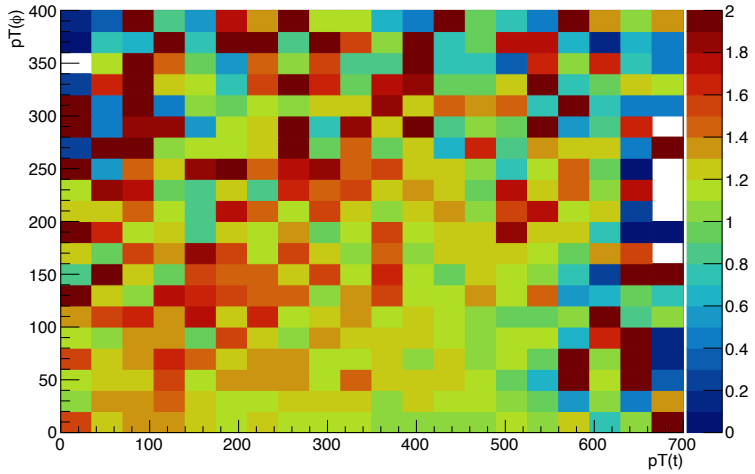
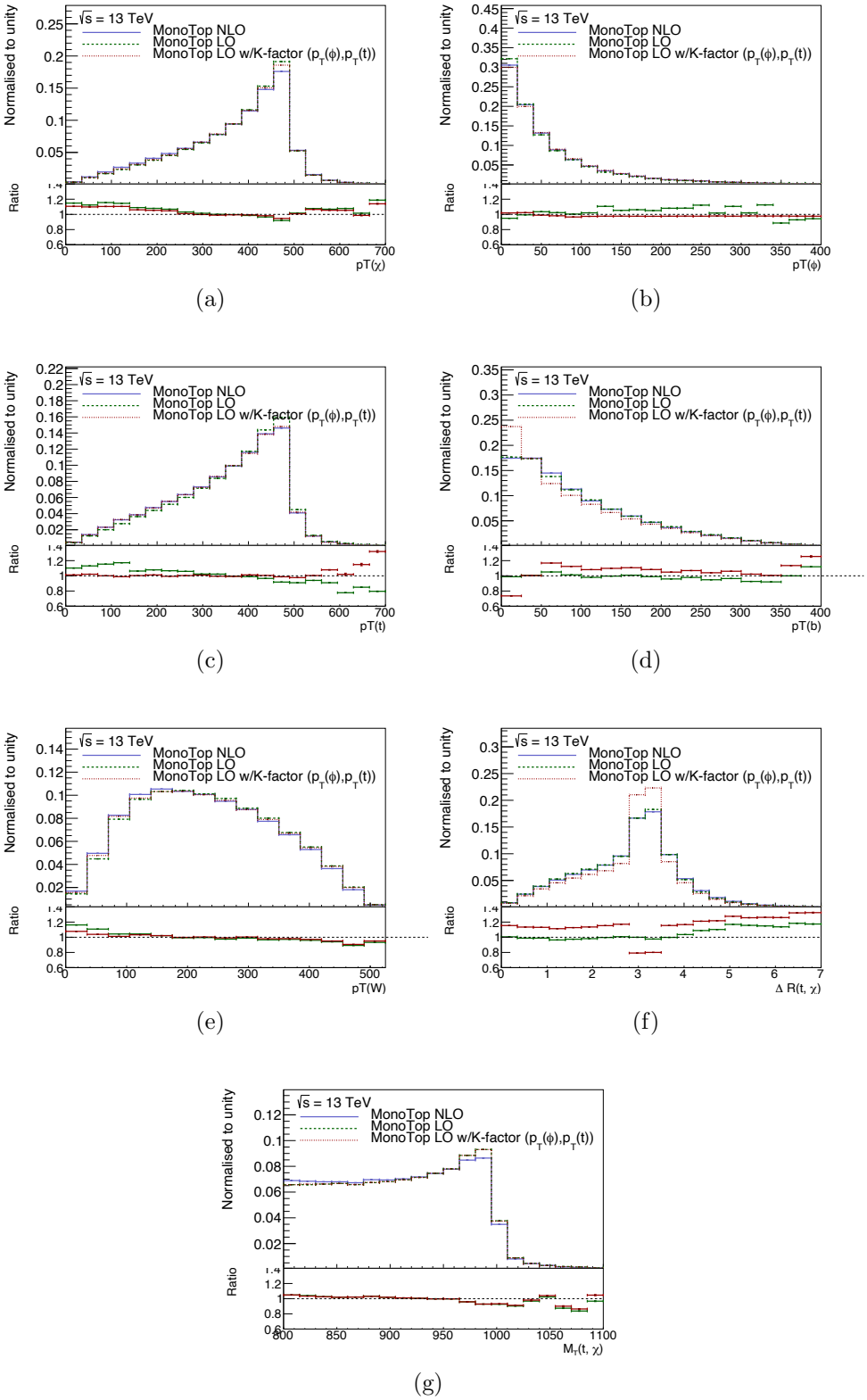


Figure 4.8: K-factor values obtained from a 2D distribution: $p_T(\phi)$ vs $p_T(t)$.

As previously, these values were used to rescale de LO sample in order to compare it with the NLO sample. The distributions are represented in Figure 4.9.



(a) (b) (c) (d) (e) (f) (g)

Figure 4.9: NLO and LO kinematic distributions, with LO samples rescaled to K-factor obtained from the $p_T(\phi)$ vs $p_T(t)$ of (a) DM particle, (b) mediator particle, (c) top quark, (d) bottom quark, (e) W boson, (f) ΔR between the top quark and χ particle and (g) transverse mass between the top quark and χ particle. All ratio plots correspond to $\frac{\text{NLO}}{\text{line colour}}$.

The results using a 2D distribution to rescale the LO samples are slightly better but not yet fully satisfactory. Since the use of K-factors did not solve the difference between the LO and NLO generated events, the work presented in this thesis was done using LO samples both for the signal and background. These studies will be obtained in the future, given the relevance for the ATLAS collaboration. The JobOptions for the resonant and non-resonant case were also developed for the analysis being done for the ATLAS collaboration.
The signal was generated using MadGraph5 2.7.3 at a centre-of-mass energy $\sqrt{s} = 13$ TeV using the monotop DMF universal FeynRules output (UFO) model [66, 67]. 2500000 events of the monotop resonant case were generated with $m_\phi = 3$ TeV, $m_\chi = 10$ GeV, $y = 0.6$ and $\lambda = 0.4$ obtaining the cross-section $\sigma = 0.0662$ pb.

4.2 Background Generation

The background samples were also generated with MadGraph5 2.7.3 at a centre-of-mass energy $\sqrt{s} = 13$ TeV but using the SM UFO model [68]. 1000000 events for each background were generated using Pythia8 for showering and Delphes high-luminosity card for the detector simulation. At parton level, the minimum p_T for the jets was required to be 20 GeV, while for the charged leptons it was required to be 10 GeV; the maximum $\mid \eta \mid$ was required to be 5 and 2.5 for jets and charged leptons, respectively. The backgrounds generated for this thesis were:

- $t\bar{t}$ di- and semileptonic;
- WW inclusive;
- ZW inclusive;
- ZZ inclusive.

This choice was made since these are the known processes, from the SM that contributes for the same final state that our signal has: 3 jets, being 1 b-tagged

and large E^T_{miss}.

The cross-sections for these processes are presented in Table 4.1.

Background	Cross-Section (pb)
$t\bar{t}$ SL	204.2850
$t\bar{t}$ DL	51.0645
WW	62.4545
ZW	23.7390
ZZ	8.9201

Table 4.1: Background samples and respective cross-section.

All of these samples, both background and signal, were generated with limited resources in a shared machine with other institutions that allows the use of only one node per user and with limited space. These processes take about 3 to 4 days to complete the generation.

After the MadGraph5 generation, it is necessary to build the dataset to be readable by the NN, as referred in Section 3.2. In order to do that, the ouput files were converted into a comma-separated values (CSV) file. For this processes more resources, including space, are needed.

Chapter 5

Results

In this chapter the kinematic cuts, as well as the NN training details will be described. The results of this thesis will be presented.

5.1 Kinematic Cuts

After transforming the MadGraph5 output files into a CSV file, the features were analysed. The dataset was divided equally into the three sets and the weights were calculated for each sample and each set. The weights are the yield contribution from each sample and it is calculated as

$$\omega = \frac{\sigma L}{N}, \tag{5.1.1}$$

where σ is the cross-section, L is the luminosity defined in Section 2.1 and it was set as 300 fb^{-1} and N is the number of events of the sample in question.
The expected yield for a given luminosity L is given by

$$N_{\text{expected}} = \sum_i \omega_i \tag{5.1.2}$$

where i corresponds to the different processes.

Basic cuts were applied to the data taking into account the signal phenomonology and decay products. The cuts applied to all data were:

- 0 leptons;
- 1 large-R jet;
- at least 1 b-tagged jet.

These cuts have been chosen since the final state of the signal corresponds to a fully hadronic one. Therefore, the first cut to be applied was to set the number of leptons to 0, then since the 2 jets provenient from the W boson are very close the number of large-R jets was set to 1 and lastly 1 b-tagged jet was required since it was expected from the top quark decay.
Some important features before and after the cuts are shown in Figure 5.1.

A lower number of weighted events can be observed in the distributions both for background and signal after the cuts were applied. The number of events before and after the cuts are shown in Table 5.1.

	Number of Events							
	Total Set		Training Set		Test Set		Validation Set	
	Before Cuts	After Cuts	Before Cuts	After Cuts	Before Cuts	After Cuts	Before Cuts	After Cuts
Background	5 000 000	16 865	1 667 205	5 700	1 666 280	5 637	1 666 515	5 528
Signal	2 500 000	283 525	832 794	94 125	833 721	94 353	833 485	95 047
Total	7 500 000	300 390	2 499 999	99 825	2 500 001	99 990	2 500 000	100 575

Table 5.1: Number of events for the background, signal and total before and after cuts.

The events that pass the cuts are in the majority signal events as shown. Although the number of signal events after the cuts is larger than the number of background events, if the weighted events are taken into account, the weighted background events are significantly higher than the weighted signal events. This can be inferred from equation 5.1.1 since the cross section for the background events, represented in Table 4.1, is much bigger than the signal cross section. This leads to statistic problems, therefore the statistical uncertainty has to be taken into account.
In Table 5.2, the weighted events for signal and background are shown. It is

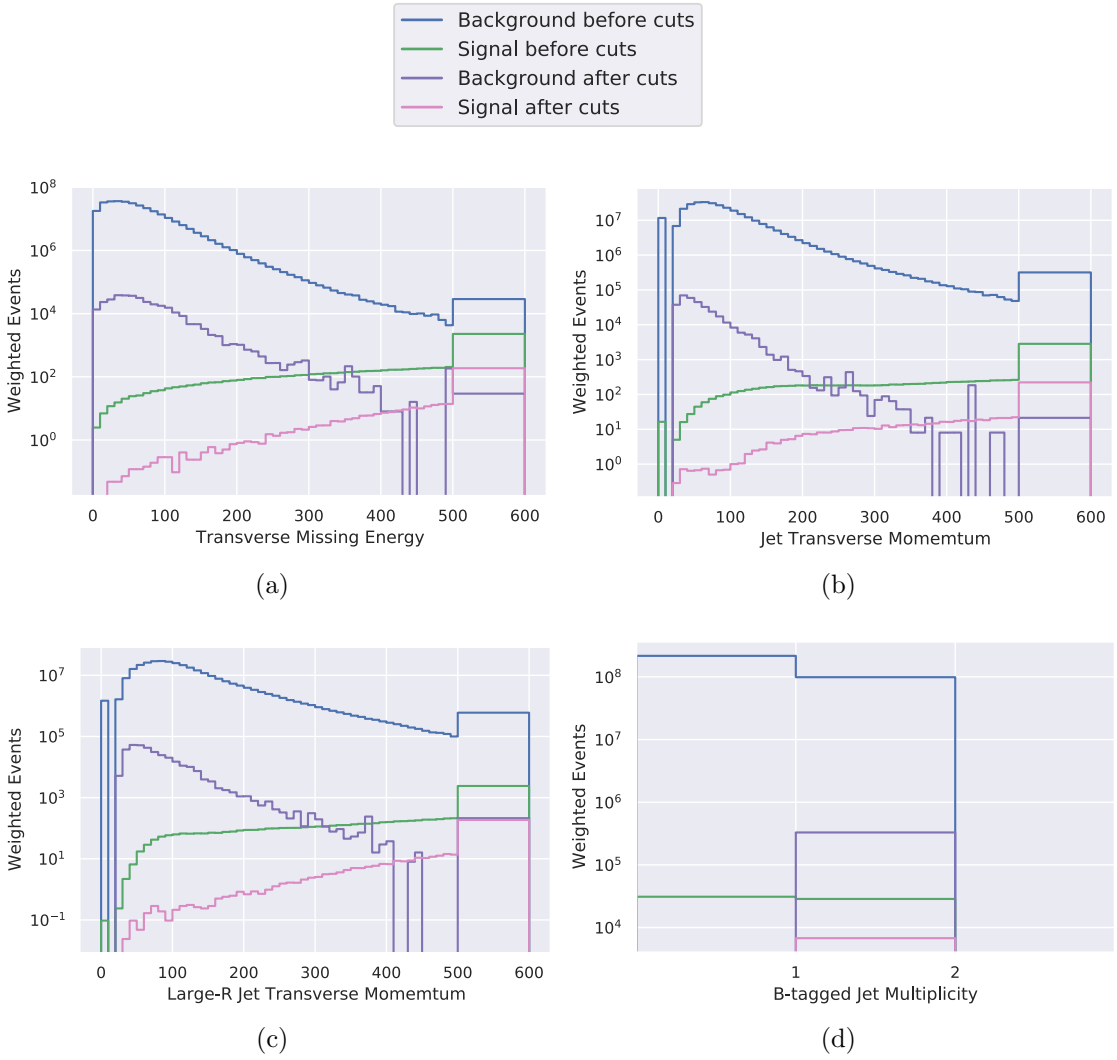


Figure 5.1: Distributions before and after the cuts with weighted events of (a) transverse missing energy, (b) jet transverse momentum, (c) large-R jet transverse momentum and (d) b-tagged jet multiplicity.

possible to verify that the weighted events for the signal are at least 2 orders of magnitude smaller.

		$\sigma L/N$		
		Training Set	Test Set	Validation Set
Signal	Resonant Fully-Hadronic	0.0239	0.0238	0.0238
Background	$t\bar{t}$ SL	183.6	183.9	183.9
	$t\bar{t}$ DL	45.9	45.9	45.9
	WW	56.1	56.1	56.4
	ZW	21.3	21.3	21.3
	ZZ	8.1	8.1	8.1

Table 5.2: Signal and Background $\frac{\sigma L}{N}$ values.

5.2 Neural Network Training

The events of the training set are used to train the model as referred in Chapter 3. This training was done using 100 random combinations of the following hyperparameters:

- layers: random integer between 1 and 10;
- units: random integer between 16 and 128;
- droupout rate: log uniform between 0.001 and 0.5;
- batch normalisation: true or false.

The training was done for a total of 200 epochs using the early stop method with a patience of 10 epochs. The predictions of the NN for the 100 combinations as well as the hyperparameters used and the performance assessment measures were saved in a CSV file and then used to calculate the limits. Since the statistic is limited, the AMS was calculated for the threshold where it takes its maximum value and this threshold value was also saved.

Before the limit calculation, the results were analysed and it was determined for which hyperparameters set the model performance metrics were better. These

performance metrics are described in Section 3.3 and the maximum value of each one was searched for in order to ascertain which NN had the better performance. These results are presented in Tables 5.3.

Trial Number	Batch Normalisation	Dropout Rate	Layers	Units	**ROC**	Average Precision	SIC	AMS
41	True	0.0463	5	29	**0.9999**	0.9988	1.2754×10^6	52.1599

(a)

Trial Number	Batch Normalisation	Dropout Rate	Layers	Units	ROC	**Average Precision**	SIC	AMS
72	True	0.1865	3	66	0.9999	**0.9989**	1.3346×10^6	48.5417

(b)

Trial Number	Batch Normalisation	Dropout Rate	Layers	Units	ROC	Average Precision	**SIC**	AMS
96	False	0.3250	9	27	0.9987	0.9194	3.4364×10^6	10.0404

(c)

Trial Number	Batch Normalisation	Dropout Rate	Layers	Units	ROC	Average Precision	SIC	**AMS**
59	True	0.1451	6	49	0.9999	0.9987	1.3714×10^6	**79.5489**

(d)

Table 5.3: Maximum value for (a) ROC, (b) Average Precision, (c) SIC and (d) AMS.

The combination of hyperparameters that provides the best results for both ROC is the 41^{st} combination which corresponds to 5 layers, 29 units, a dropout rate of 0.0463 and batch normalisation set to True, while for the average precision is the 72^{nd} combination which corresponds to 3 layers, 66 units, a dropout rate of 0.1865 and batch normalisation set to True. The 96^{th} combination of hyperparameters is the one that maximises the SIC value and it corresponds to 9 layers, 27 units, a dropout rate of 0.3250 and batch normalisation set to False. While for the AMS, it

is the 59^{th} hyperparameter combination that maximises its value and corresponds to 6 layers, 49 units, a dropout rate 0.1451 and batch normalisation set to True. Since the set of hyperparameters that provides the best model performance metric is different depending on which metric is chosen, the harmonic mean between the four metrics was done. The harmonic mean can be written as

$$H = \frac{n}{\sum_{i=n}^{n} \frac{1}{x_i}}, \tag{5.2.1}$$

where in this case x_i corresponds to the metrics values in the ith position. The maximum value for the harmonic mean was found and the correspondent values of the performance assessment measures are represented in Table 5.4.

Trial Number	Batch Normalisation	Dropout Rate	Layers	Units	ROC	Average Precision	SIC	AMS
59	True	0.1451	6	49	0.9999	0.9987	1.3715×10^6	79.5489

Table 5.4: Line corresponding to maximum harmonic mean.

Although the NN that provides the maximum harmonic is the same as the one that provides the maximum value for the AMS, conclusions can not be drawn right away. In order to achieve a more detailed study, the limits for each combination of hyperparameters were calculated to obtain a correlation between these values, the metric values and also the hyperparameters values.

5.3 Limits

The upper limit at 95 % confidence level (CL) was calculated using pyhf [69] wich implements the CL_s method [70]. This method consists in considering two hypothesis: one that includes the SM known processes and it is often referred to as the background-only (b) hypothesis and another that corresponds to the SM known processes with the addition of a new signal process and it is often referred to as signal-plus-background ($s+b$) hypothesis. In this thesis the SM without the monotop resonant case is considered the background-only scenario, while the SM with the monotop resonant case is considered the signal-plus-background scenario. The ratio of likelihoods for the two hypothesis of interest is given by

$$Q = \frac{P(\text{data}|\text{signal + background})}{P(\text{data}|\text{background})}, \tag{5.3.1}$$

with Q being the test statistical and where the probabilities are defined as

$$P(\text{data}|\text{background}) = \frac{b^n e^{-b}}{n!}, \tag{5.3.2}$$

$$P(\text{data}|\text{signal + background}) = \frac{(s+b)^n e^{-(s+b)}}{n!},$$

with s being the signal and b the background expected number events and n the observed number of events.

The confidence level for excluding the possibility of $s+b$ hypothesis is given by CL_{s+b} and can be defined as the probability that Q would be less than or equal to the observed in the data, assuming the presence of both signal and background. The confidence level for excluding the b hypothesis is given by CL_b and can be defined as the probability that Q would be less than or equal to the observed in the data, assuming the presence of only background. These two confidence levels can be written as

$$CL_{s+b} = P(Q \leq Q_{obs}|\text{signal + background}), \tag{5.3.3}$$

$$\mathrm{CL}_b = P(Q \leq Q_{obs}|\text{background}).$$

With these two confidence levels defined, the signal-only (s) hypothesis can be obtained,

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b}. \tag{5.3.4}$$

If $\mathrm{CL}_s < 0.05$, the signal-plus-background hypothesis with a signal strength μ is excluded at 95 % CL. The signal strenght can be defined as

$$\mu = \frac{\sigma^{95\%}}{\sigma^{\text{theoretical}}}. \tag{5.3.5}$$

If $\mu < 1$ the signal is excluded at 95 % CL.

Using the predictions of the 100 NN, the limits were calculated using pyhf and the results were saved in a CSV file. A pre-processing of the bins was done, if the weighted number of events of the background and the signal was less than 0.01 the correspondent bin was eliminated, since pyhf implements the assymptotic formulae.
After the limit calculation, the NN output of three minimum and maximum μ were analysed. These outputs are represented in Figures 5.2 and 5.3 , respectively.

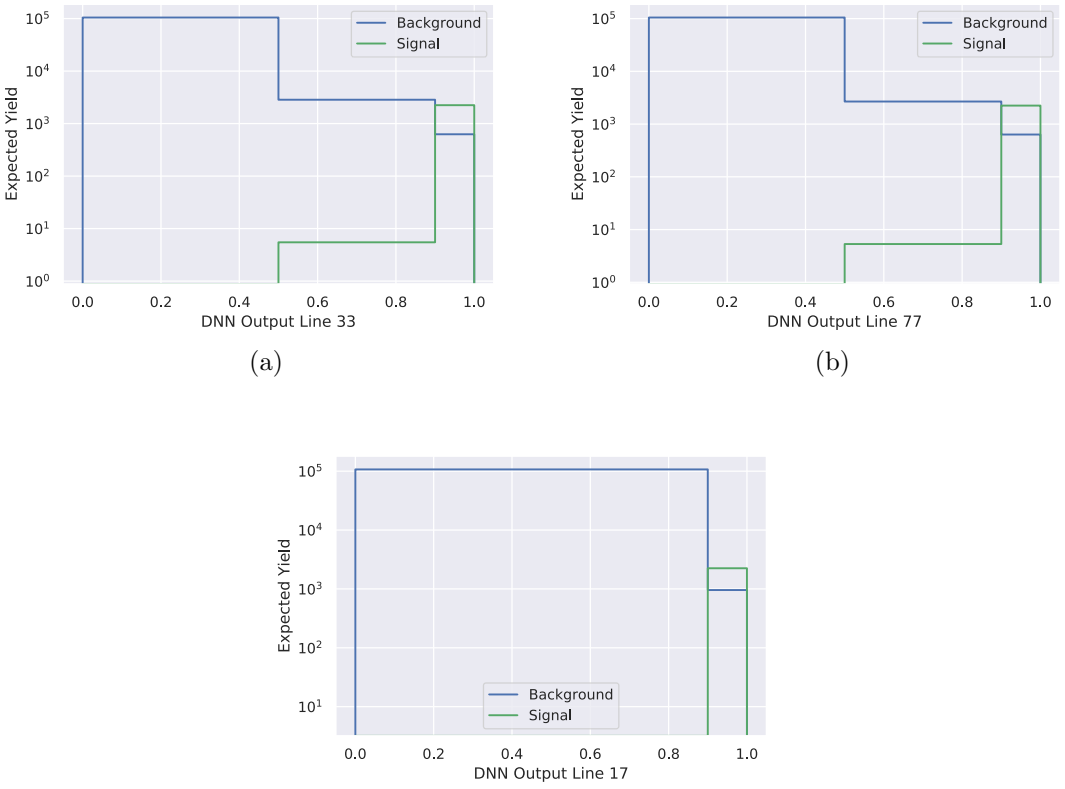


Figure 5.2: Output of the NN, with background and signal expected yield, corresponding to (a) the minimum value for the μ, (b) the second minimum μ and (c) the third minimum μ.

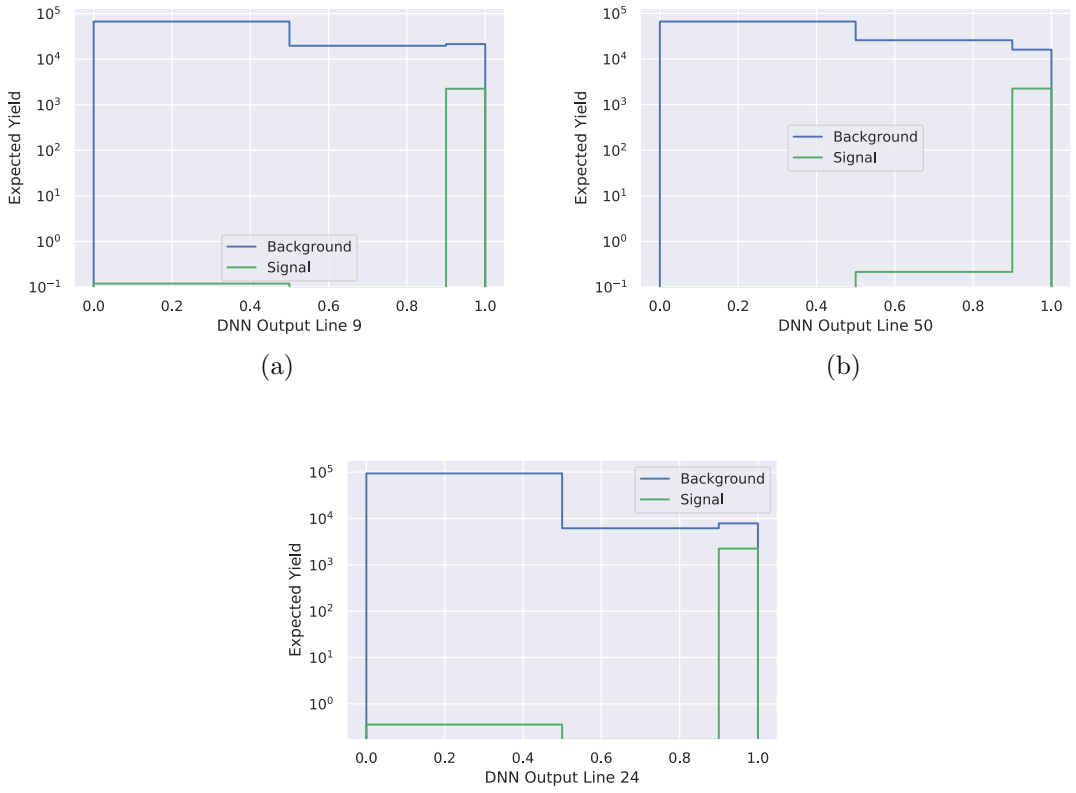


Figure 5.3: Output of the NN, with background and signal expected yield, corresponding to (a) the maximum μ, (b) the second maximum μ and (c) the third maximum μ.

For the output of the NN which corresponds to the minimum values for μ, the signal weighted events in the last bin is higher than the background weighted events while, for the NN outputs corresponding to the maximum value for μ, the opposite happens. The minimum and maximum values of μ are represented in Table 5.5, as well as the corresponding hyperparameters and metric values.

	$\mu = 0.2023^{+0.028}_{-0.0002}$	$\mu = 1.0053^{+0.3675}_{-0.2617}$
Number of Layers	1	9
Number of Units	61	101
Dropout Rate	0.0881	0.3449
Batch Normalisation	False	True
ROC	0.9999	0.9951
Average Precision	0.9981	0.6755
SIC	1.2139×10^6	5.6655×10^5
AMS	79.1910	4.9672

Table 5.5: Values for the hyperparameters and metrics for the minimum μ and the maximum μ.

In order to have a better correlation between the metrics and the μ values, a correlation plot as well as a scattering plot was done. The correlation plot is represented in Figure 5.4.

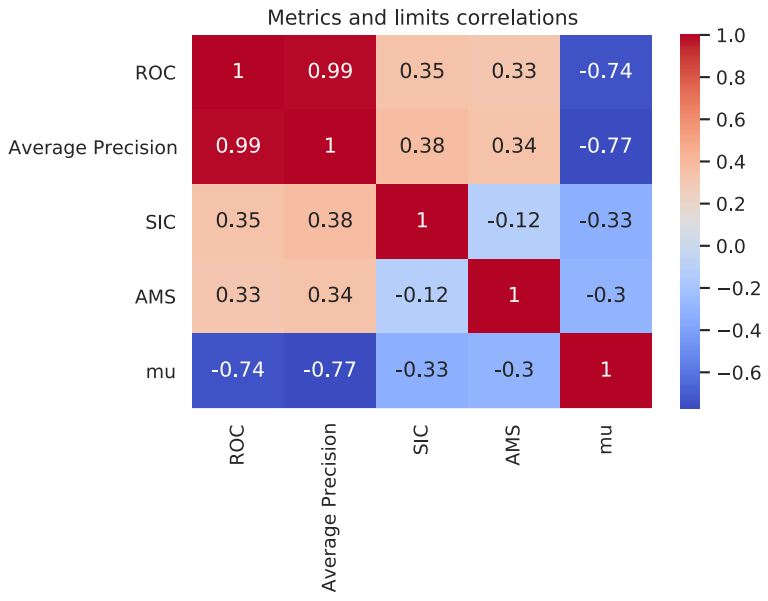


Figure 5.4: Correlation plot between metrics and μ values.

The scattering plot between these variables was done and it is represented in Figure 5.5.

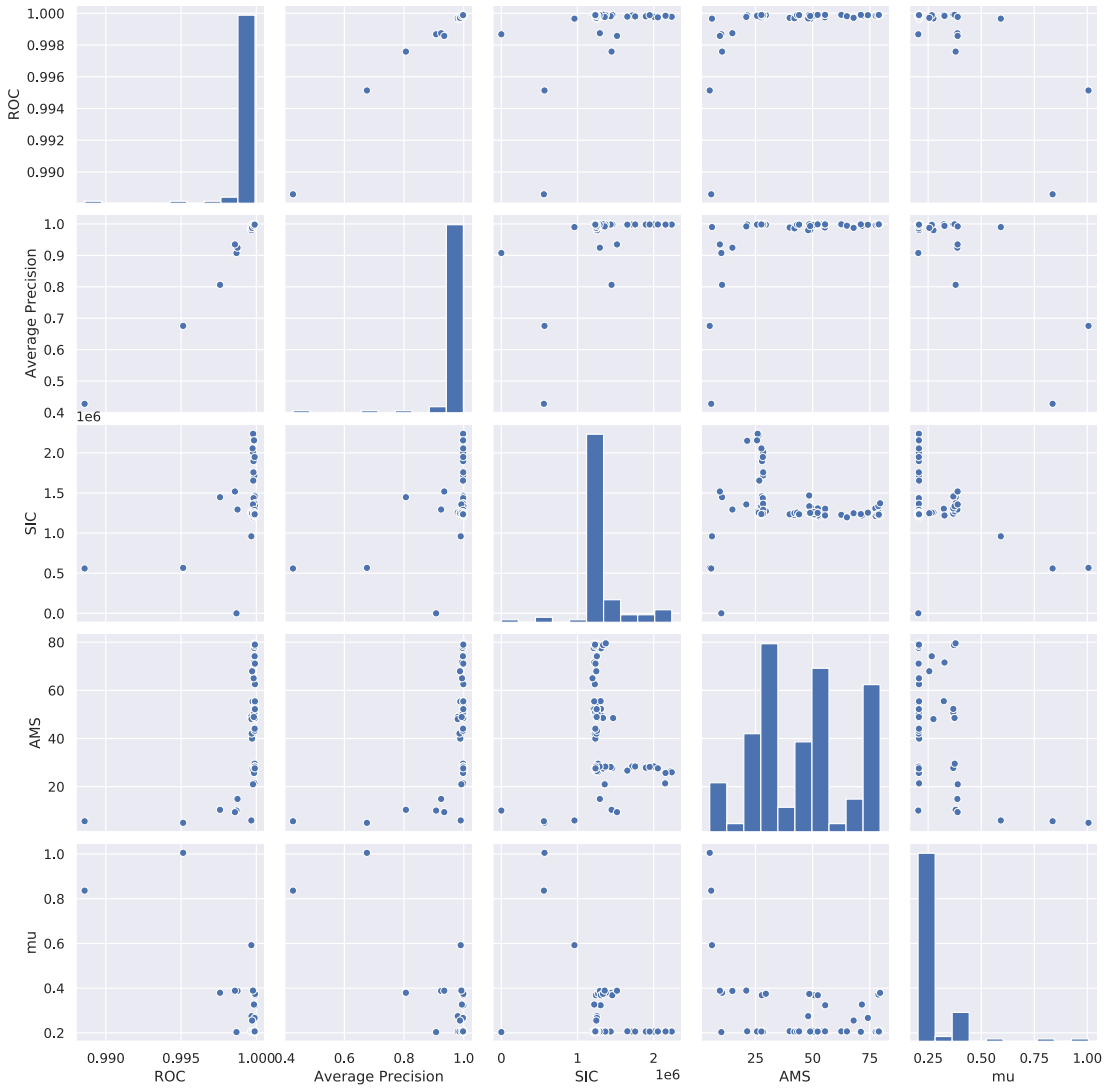


Figure 5.5: Scattering plot of metrics and μ values.

Looking at the correlation plot values and the scattering plot, the value of μ looks more correlated with the ROC and average precision values. However, ROC and average precision are saturated at 1. For this reason, the same procedure was done but, instead of taking the metric values x for ROC and average precision, $\log(1-x)$ was taken into consideration. The correlation plot for this case is

represented in Figure 5.6.

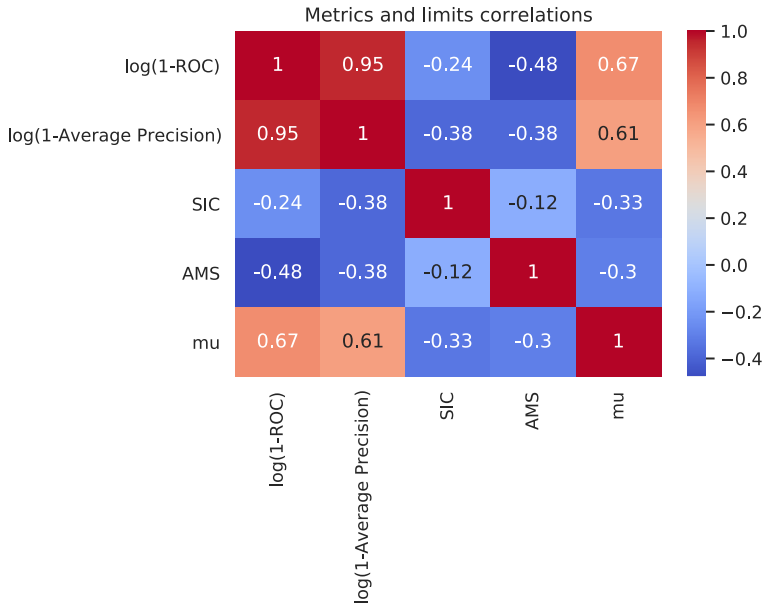


Figure 5.6: Correlation plot between metrics and μ values, taking $\log(1-x)$ with x being ROC and average precision.

Looking at the correlation plot of Figure 5.6, it is possible to observe that the values of μ are very correlated with the values of $\log(1 - \text{ROC})$ and $\log(1 - \text{Average Precision})$. The scattering plot for these values was also done and it is represented in Figure 5.7.

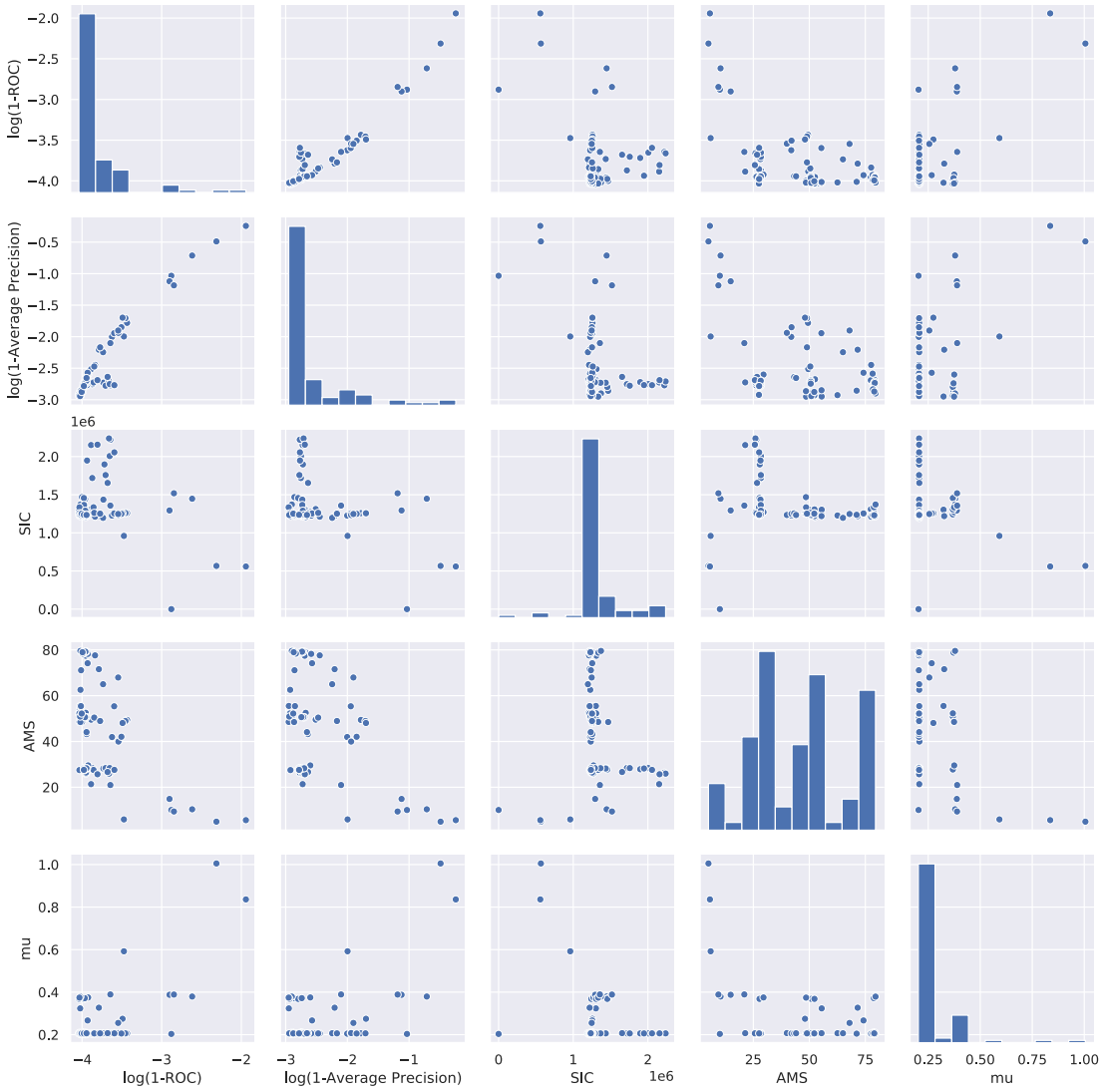


Figure 5.7: Scattering plot of metrics and μ values, taking $\log(1-x)$ with x being ROC and average precision.

The results of ROC and average precision values are very close and the drawn conclusions are the same. For the $\log(1-\text{ROC})$ values, the μ values are lower for lower $\log(1-\text{ROC})$ values and higher for higher $\log(1-\text{ROC})$ values. The $\log(1-\text{Average Precision})$ values have the same behavior as the $\log(1-\text{ROC})$ values. Looking at the scatter plot where the y axis corresponds to the μ values and the x axis corresponds to AMS values, AMS has a more disperse distribution.

However, it is possible to observe that the higher values of μ are related to the lower values for the AMS. The values of the SIC have a similar behaviour as the AMS ones.

The same procedure was done with the hyperparameters values. The correlation plot is represented in Figure 5.8.

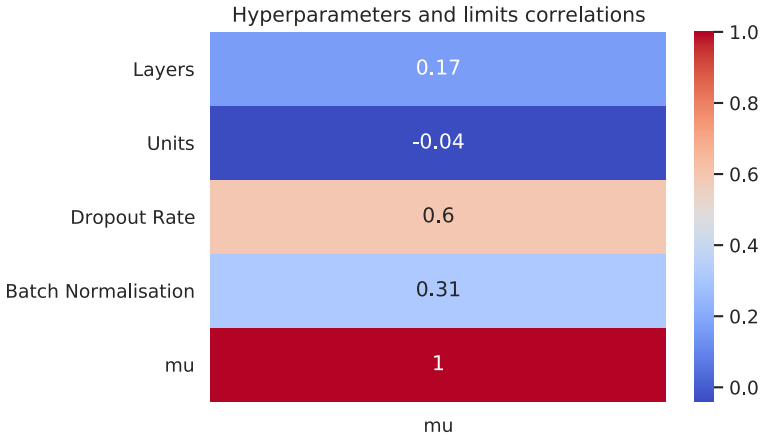


Figure 5.8: Correlation plot between hyperparameters and μ values.

Looking at the correlation coefficients, the hyperparameter more correlated with the μ value is the dropout rate. However, it is important to look at the scattering plot represented in Figure 5.9.

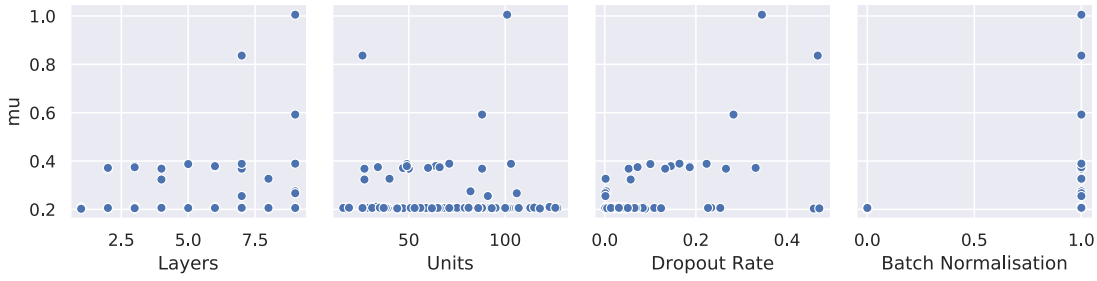


Figure 5.9: Scattering plot of hyperparameters and μ values.

The number of layers has a small influence in the μ values. In the μ vs layers plot the points are a little disperse, however the lower number of layers provides

the lower values for μ. The number of units is not correlated with the μ values, since it has a more disperse set of points. For the batch normalisation, where 0 corresponds to the batch normalisation set to False and 1 to the batch normalisation set to True, the lower values of μ seem to be obtained from NN with batch normalisation set to False. The dropout rate seems to have more influence in the μ values, with the lower values of μ corresponding to the lower values of the dropout rate.

Therefore, we see that simpler networks with little regularisation seem to be preferred for this classification task.

Looking at the results from the metrics, ROC and average precision are a better proxy for the μ value than SIC and AMS. Looking at the scattering and correlation plots, Figures 5.6 and 5.7, the value of μ looks more correlated with the values of $\log(1 - \text{ROC})$.

Chapter 6

Conclusion

As previously stated, the SM is able to describe the elementary particles and their interactions. However, it does not provide answers to some questions. Astrophysical measurements point to the existence of DM. In fact, only $\sim 5\%$, corresponding to baryonic mass, of the Universe is known. This kind of matter can be searched for at the LHC with the production of particles that couple both to the DM candidate particles and to the SM particles.

In the present thesis, a search for DM at the LHC in association with a top quark was done. The fully hadronic resonant case was considered, where the final state is characterised with high E^T_{miss}. An LO versus NLO study on the signal generation was performed using different variables to obtain the K-factors in order to rescale the LO sample. This study was presented to the ATLAS collaboration team contributing to the analysis being performed. However, the results were not yet fully satisfactory so the LO generation was used for the ML studies.

Simply applying the selection cuts led to low statistic problems. The generation of samples was conditioned by computation limitations such as time and available space. Several NNs with random hyperparameters were trained and evaluated with ML and HEP model performance metrics. For the limit calculation using the CL_s method, the evaluation of the relations between the hyperparameter, model performance metrics and limit value was done.

From the scattering plots presented, simpler networks with little regularisation

seem to be preferred for this classification task. The results for the metrics are very similar for the ROC and average precision but it seems that the value of μ is more correlated with the values of the ROC.
For future work, the same study could be redone with more statistic in order to validate the results obtained in this thesis. A comparison between the performance of a NN and a boosted decision tree can also be performed, as well as the use of the AMS as a loss function.

Bibliography

[1] S. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, pp. 579–588, 1961.

[2] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev. D*, vol. 98, no. 3, p. 030001, 2018.

[3] https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg.

[4] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[5] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.

[6] G. Arnison *et al.*, "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV," *Phys. Lett. B*, vol. 122, pp. 103–116, 1983.

[7] M. Banner *et al.*, "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider," *Phys. Lett. B*, vol. 122, pp. 476–485, 1983.

[8] P. W. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Lett.*, vol. 12, pp. 132–133, 1964.

[9] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[10] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons," *Phys. Rev.*, vol. 145, pp. 1156–1163, May 1966.

[11] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.

[12] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964.

[13] T. W. B. Kibble, "Symmetry Breaking in Non-Abelian Gauge Theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, Mar 1967.

[14] J. Ellis, "Higgs Physics," pp. 117–168. 52 p, Dec 2013. 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report.

[15] F. Abe and *et al.*, "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab," *Phys. Rev. Lett.*, vol. 74, pp. 2626–2631, Apr 1995.

[16] S. Abachi and *et al.*, "Observation of the Top Quark," *Phys. Rev. Lett.*, vol. 74, pp. 2632–2637, Apr 1995.

[17] K. Kröninger, A. B. Meyer, and P. Uwer, "Top-Quark Physics at the LHC," *The Large Hadron Collider*, p. 259–300, 2015.

[18] A. Giammanco, "Single top quark production at the LHC," *Reviews in Physics*, vol. 1, p. 1–12, Nov 2016.

[19] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWG.

[20] P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al., "Planck2015 results," *Astronomy Astrophysics*, vol. 594, p. A13, Sep 2016.

[21] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," *Helvetica Physica Acta*, vol. 6, pp. 110–127, Jan. 1933.

[22] V. C. Rubin, D. Burstein, J. Ford, W. K., and N. Thonnard, "Rotation velocities of 16 SA galaxies and a comparison of Sa, SB and SC rotation properties.,"

[23] A. Bosma, *HI Velocity Fields and Rotation Curves*, pp. 11–22. Dordrecht: Springer Netherlands, 1983.

[24] A. Carati, "Gravitational effects of the faraway matter on the rotation curves of spiral galaxies," 2011.

[25] D. Clowe, A. Gonzalez, and M. Markevitch, "Weak [U+2010] Lensing Mass Reconstruction of the Interacting Cluster 1E 0657−558: Direct Evidence for the Existence of Dark Matter," *The Astrophysical Journal*, vol. 604, p. 596–603, Apr 2004.

[26] C. Alcock, R. A. Allsman, D. R. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, N. Dalal, A. J. Drake, K. C. Freeman, M. Geha, K. Griest, M. J. Lehner, S. L. Marshall, D. Minniti, C. A. Nelson, B. A. Peterson, P. Popowski, M. R. Pratt, P. J. Quinn, C. W. Stubbs, W. Sutherland, A. B. Tomaney, T. Vandehei, and D. Welch, "The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations," *The Astrophysical Journal*, vol. 542, pp. 281–307, oct 2000.

[27] D. Abercrombie, N. Akchurin, E. Akilli, J. A. Maestre, B. Allen, B. A. Gonzalez, J. Andrea, A. Arbey, G. Azuelos, P. Azzi, and et al., "Dark Matter benchmark models for early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum," *Physics of the Dark Universe*, vol. 27, p. 100371, Jan 2020.

[28] J.-L. Agram, J. Andrea, M. Buttignol, E. Conte, and B. Fuks, "Monotop phenomenology at the Large Hadron Collider," *Physical Review D*, vol. 89, Jan 2014.

[29] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al., "Search for large missing transverse momentum in association with one top-quark in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Journal of High Energy Physics*, vol. 2019, May 2019.

[30] L. Evans and P. Bryant, "LHC Machine," *Journal of Instrumentation*, vol. 3, pp. S08001–S08001, aug 2008.

[31] C. De Melis, "The CERN accelerator complex," Jan 2016. https://cds.cern.ch/record/2119882.

[32] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[33] S. Chatrchyan *et al.*, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[34] K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[35] J. Alves, A.Augusto *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[36] "LuminosityPublicResultsRun2." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_201.

[37] "Public CMS Luminosity Information." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[38] A. Collaboration, "Alignment of the ATLAS Inner Detector in Run-2," 2020.

[39] "CMS detector design: CMS Experiment," *CMS detector design — CMS Experiment*.

[40] P. Azzi and *et al.*, "Standard Model Physics at the HL-LHC and HE-LHC," 2019.

[41] "Expected performance of the ATLAS detector at the High-Luminosity LHC," Tech. Rep. ATL-PHYS-PUB-2019-005, CERN, Geneva, Jan 2019.

[42] T. C. Collaboration, "Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC," Tech. Rep. CMS-NOTE-2018-006. CERN-CMS-NOTE-2018-006, CERN, Geneva, Dec 2018.

[43] "Expected pileup values at the HL-LHC," Tech. Rep. ATL-UPGRADE-PUB-2013-014, CERN, Geneva, Sep 2013.

[44] "Technical Design Report for the ATLAS Inner Tracker Pixel Detector," Tech. Rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, Geneva, Sep 2017.

[45] "Technical Design Report for the ATLAS Inner Tracker Strip Detector," Tech. Rep. CERN-LHCC-2017-005. ATLAS-TDR-025, CERN, Geneva, Apr 2017.

[46] "Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter," Tech. Rep. CERN-LHCC-2017-018. ATLAS-TDR-027, CERN, Geneva, Sep 2017.

[47] "Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter," Tech. Rep. CERN-LHCC-2017-019. ATLAS-TDR-028, CERN, Geneva, Sep 2017.

[48] "Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer," Tech. Rep. CERN-LHCC-2017-017. ATLAS-TDR-026, CERN, Geneva, Sep 2017.

[49] "Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System," Tech. Rep. CERN-LHCC-2017-020. ATLAS-TDR-029, CERN, Geneva, Sep 2017.

[50] "Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade," Tech. Rep. CERN-LHCC-2018-023. LHCC-P-012, CERN, Geneva, Jun 2018.

[51] "The Phase-2 Upgrade of the CMS Tracker," Tech. Rep. CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun 2017.

[52] "The Phase-2 Upgrade of the CMS Endcap Calorimeter," Tech. Rep. CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov 2017.

[53] "The Phase-2 Upgrade of the CMS Barrel Calorimeters," Tech. Rep. CERN-LHCC-2017-011. CMS-TDR-015, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC.

[54] "The Phase-2 Upgrade of the CMS Muon Detectors," Tech. Rep. CERN-LHCC-2017-012. CMS-TDR-016, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC.

[55] J. Allison and *et al.*, "Geant4 developments and applications," *IEEE Transactions on Nuclear Science*, vol. 53, no. 1, pp. 270–278, 2006.

[56] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3: a modular framework for fast simulation of a generic collider experiment," *Journal of High Energy Physics*, vol. 2014, Feb 2014.

[57] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.

[58] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, pp. 063–063, apr 2008.

[59] "Hl/he-lhc physics workshop – simulation tools." https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HLHEWG_MC.

[60] F. Chollet, *Deep Learning with Python*. USA: Manning Publications Co., 1st ed., 2017.

[61] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2017.

[62] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, Feb 2011.

[63] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *Journal of High Energy Physics*, vol. 2014, Jul 2014.

[64] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1," *Computer Physics Communications*, vol. 178, p. 852–867, Jun 2008.

[65] R. Vogt, "The Usage of the K Factor in Heavy Ion Physics," *Acta Physica Hungarica A) Heavy Ion Physics*, vol. 17, p. 75–92, Feb 2003.

[66] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter, "UFO – The Universal FeynRules Output," *Computer Physics Communications*, vol. 183, p. 1201–1214, Jun 2012.

[67] http://feynrules.irmp.ucl.ac.be/attachment/wiki/Monotops.

[68] https://feynrules.irmp.ucl.ac.be/wiki/StandardModel.

[69] Heinrich, Lukas and Feickert, Matthew and Stark, Giordon, "pyhf: v0.5.2."

[70] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys. G*, vol. 28, pp. 2693–2704, 2002.