Maria do Céu Vale Lopes Neiva

Advanced machine learning techniques in rare events research at the Large Hadron Collider

Advanced machine learning techniques in rare events research at the Large Hadron Collider

Céu Neiva

UMinho | 2022

dezembro de 2022

Universidade do Minho
Escola de Ciências

Maria do Céu Vale Lopes Neiva

Advanced machine learning techniques in rare events research at the Large Hadron Collider

Master Thesis
Master in Physics

Work developed under the supervision of:
Professor Doutor Nuno Filipe da Silva Fernandes de Castro
Doutor Miguel Correia dos Santos Crispim Romão

⟨December⟩, ⟨2022⟩

COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.
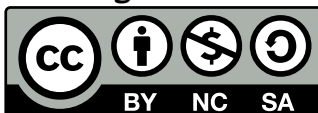
Acknowledgements

Throughout my academic journey, many people have been really supportive of me to whom I owe my most profound and sincere thanks.

First of all, I would like to thank my supervisors Dr. Nuno Castro and Dr. Miguel Romão, that always shared their knowledge and support in an exigent and comprehensive way. I also want to thank them for giving me the best opportunities to grow as a scientist. I also have to thank my LIP colleagues, especially Maura, for their support, advice and help. Thanks to Tiago for being an inspiration to me.

I want to thank my family for all the love and support, especially my mother Carla and father Manuel, that have always believed in me and supported me in every way with everything they got. Without them I would not be able to be in this position today. I want to thank my brother, Rafael - that has always been my role model - for his advice, support and understanding. I want to thank Jorge for having my back through everything and always pushing me to do better and to fight for my goals.

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 Universidade do Minho.

_____, _____
(Place) (Date)

_____
(Maria do Céu Vale Lopes Neiva)

Resumo

Técnicas avançadas de machine learning na pesquisa de eventos raros no Large Hadron Collider

Há ainda várias questões acerca do universo às quais ainda não temos resposta. É fundamental estudarmos e melhorarmos as técnicas já conhecidas, bem como deselvolver novas técnicas, de forma a investigar e descobrir novos fenómenos de física.

Neste trabalho, foram utilizadas técnicas de *machine learning* tais como *boosted decision trees* e *autoencoders*. As *boosted decision trees* foram utilizadas como métodos de *machine learning* supervisionado, enquanto que os *autoencoders* foram utilizados como métodos de deteção de anomalias. Foi feita uma comparação entre a *AUC* dos resultados da *boosted decision tree*, dos *autoencoders* e das *features*, sem estas terem passado por nenhum modelo. Já com os *autoencoders* foi feito outro estudo - a dimensão do espaço latente afeta a deteção do sinal? Foi feito um estudo em várias dimensões do espaço latente e foi calculada a *AUC* para visualizar este efeito.

Os resultados da *boosted decision tree* foram melhores que os resultados dos *autoencoders*, porém utilizar autoencoders foi melhor do que não utilizar nada - a *AUC* dos *autoencoders* foi melhor que a *AUC* das *features*. Além disso, a escolha do espaço latente tem que ser tomada em conta visto que há dimensões mais sensíveis a certos tipos de sinais.

Palavras-chave: aprendizagem automática, árvores de decisão, codificadores automáticos, deteção de anomalias, dimensão do espaço latente

Abstract

Advanced machine learning techniques in rare events research at the Large Hadron Collider

There are still many questions about our universe to which we do not know the answer yet. The development of new techniques and the improvement of techniques that we already developed is fundamental to investigate and discover new physics phenomena.

In this work, machine learning techniques were studied such as boosted decision trees and autoencoders. Boosted decisions trees were used as a supervised technique and the autoencoders as an anomaly detection technique. It was done a comparison between the AUC of the results from the boosted decision tree, the autoencoders and from the features without going through any model. With the autoencoders, another study was done - does the latent space dimension affect the signal detection? For different latent space dimensions, the AUC was performed to visualize this effect.

The results from the boosted decision tree were better than the results coming from the autoencoders but, using autoencoders was better than using nothing - the AUC of the autoencoders were better than the AUC of the features. Furthermore, it was seen that the choice of the latent space dimension must be taken into consideration since some dimensions are sensitive to certain kind of signals.

Keywords: anomaly detection, autoencoders, decision tree, latent space dimension, machine learning

Contents

List of Figures

List of Tables

Standard Model and Beyond the Standard Model Particle Physics

There are many open questions about the constituents of the universe that are yet to be answered. The Standard Model (SM) of particle physics, alone, cannot explain many cosmological observations so there is an unavoidable need to search for physics Beyond the Standard Model (BSM). Despite this, the SM was and still is to this day a very successful model, describing very well experimental data obtained in collider experiments. Such importance needs to be remarked - in section 1.1 the SM of particle physics is described, the particles that it contains and the fundamental interactions between them. The importance of BSM physics is highlighted in section 1.2.

1.1   Standard Model of Particle Physics

The SM was developed in the second half of the 20th century and it accommodates the elementary particles. Their interaction, or in other words the fundamental forces, are also described in SM and they are mediated by elementary particles themselves. They can be split into fermions and bosons, that are represented in table 1 and table 2. Fermions are particles of semi-integer spin that obey the Fermi-Dirac statistics. This group involves quarks and leptons and each one is divided into three families of increasing mass, from the first to the final. The fundamental forces are described in table 2 and their carriers are spin-1 gauge bosons. These bosons obey Bose-Einstein statistics. The photon (γ) is the mediator of the electromagnetic force and has an infinite range and is also responsible for the interaction between charged particles. The photon is massless and neutral. The gluon, an equally massless and neutral particle, is the carrier of the strong force and due to colour charge, it has eight combinations. The strong force is responsible for the confinement of the quarks into protons, neutrons or other hadrons and it also binds

Quarks				
Family	**Symbol**	**Name**	**Mass**	**Electric Charge ($|e|$)**
1st	u	Up	2.2 MeV	+2/3
	d	Down	4.7 MeV	-1/3
2nd	c	Charm	1.3 GeV	+2/3
	s	Strange	96 MeV	-1/3
3rd	t	Top	173 GeV	+2/3
	b	Bottom	4.7 GeV	-1/3
Leptons				
Family	**Symbol**	**Name**	**Mass**	**Electric Charge ($|e|$)**
1st	e	Electron	0.5 MeV	-1
	ν_e	Electron Neutrino	2 eV	0
2nd	μ	Muon	105.7 MeV	-1
	ν_μ	Muon Neutrino	2 eV	0
3rd	τ	Tau	1775.9 MeV	-1
	ν_τ	Tau Neutrino	2 eV	0

Table 1: Elementary particles, their masses and electric charge in units of elementary charge $|e|$ according to Particle Data Group [1]

Fundamental Force	**Boson**	**Mass (GeV)**	**Electric Charge ($|e|$)**
Electromagnetic	Photon (γ)	0	0
Strong	Gluon $\times$ 8 (g)	0	0
Weak	Z	91	0
	W$\pm$	80	$\pm$ 1

Table 2: Fundamental forces, their mediators and their mass and electric charge in units of elementary charge ($|e|$) according to Particle Data Group [1]

the protons and neutrons themselves to create the atomic nuclei. The range of this force is very small, around 10^{-15}m, and the only particles that participate in strong force are quarks and gluons themselves. The weak force, responsible for beta decay, is mediated by Z and W, which both have mass and W is positive or negative charged. The range of this force is also short, around 10^{-18}m.

Electrons are well known to be a part of the atom. Orbiting the atom, these electrons are bound to the nucleus by electrostatic attraction between two opposite charged particles. This is a manifestation of Quantum Electrodynamics (QED) (see section 1.1.1). The positively charged nucleus is composed by quarks and their combination creates the proton (uud), which is positively charged, and neutron (udd), which is neutral. Neutrons and protons are kept together by strong force, which is an indication of Quantum Chromodynamics (QCD) (see section 1.1.2). [1] [2]

In addition to these bosons, the SM also includes a scalar boson, the Higgs boson, which is responsible for the mass of the elementary particles, through the Brout-Englert-Higgs mechanism (see section 1.1.4).

1.1.1 Quantum Electrodynamics

The junction of electrodynamics and quantum mechanics gives rise to QED - an electrodynamics quantum field theory invariant under gauge transformations. This theory, quantised by the electromagnetic field, describes the interactions between charged particles and quantises the field. QED is described by the Lagrangian (eq. 1.3):

$$\mathcal{L}_{QED} = \mathcal{L}_{Dirac} + \mathcal{L}_{Maxwell} + \mathcal{L}_{int}, \tag{1.1}$$

with

$$\mathcal{L}_{Dirac} = \bar{\Psi}(i\not{\partial} - m)\Psi$$
$$\mathcal{L}_{Maxwell} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.2}$$
$$\mathcal{L}_{int} = -e\bar{\Psi}\gamma^{\mu}\Psi A_{\mu},$$

where $\mathcal{L}_{Dirac}$ is the Dirac Lagrangian that describes free particles, $\mathcal{L}_{Maxwell}$ is the Maxwell Lagrangian and $\mathcal{L}_{int}$ is a term that considers the couplings between fermions and the four-potential vector A_{μ}. In these expressions, Ψ is the Dirac field, m is the fermion mass, $F^{\mu\nu}$ is the electromagnetic field tensor $F^{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$. The QED Lagrangian becomes:

$$\mathcal{L}_{QED} = \bar{\Psi}(i\not{\partial} - m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - e\bar{\Psi}\gamma^{\mu}\Psi A_{\mu}. \tag{1.3}$$

If one applies a transformation $\Psi \rightarrow e^{i\Lambda(x)}\Psi$, which is a transformation that operates on the field Ψ, and since A_{μ} is invariant under $A_{\mu} \rightarrow A_{\mu} - \partial A$, we get an additional term $\bar{\Psi}e\gamma^{\mu}\partial_{\mu}\Psi$, in order to achieve the invariance. Adding this contribution to the Lagrangian (eq. 1.3), the QED Lagrangian density can be written as:

$$\mathcal{L}_{QED} = \bar{\Psi}(i\not{\partial} - e\gamma^{\mu}A_{\mu} - m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.4}$$

Defining the covariant derivative as $D_{\mu} = \partial_{\mu} + ieA_{\mu}$, the eq. 1.4 becomes:

$$\mathcal{L}_{QED} = \bar{\Psi}(i\not{D} - -m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.5}$$

The photon appears when a local phase invariance (U(1)) is applied to the free Dirac Lagrangian since it introduces a massless field that mediates QED interactions and we know it to be this gauge boson.

1.1.2 Quantum Chromodynamics

QCD appeared as a necessity to explain why bound states of quarks were possible and how it did not violate Pauli's exclusion principle. In QCD the interaction is between colour-charged particles - quarks and gluons. This interaction occurs via gluon exchange and since gluons interact with themselves, they

also have colour. These colours are red (R), green (G) and blue (B). The quarks can be represented as in equation 1.6.

$$\Psi = \begin{pmatrix} \Psi_R \\ \Psi_G \\ \Psi_B \end{pmatrix}. \tag{1.6}$$

The generators of the SU(3) group - the gauge group of the QCD - are written in equation 1.7, where λ_a are the Gell-Mann matrices, with $a = 1, 2, 3, ..., 8$:

$$t^a = \frac{1}{2}\lambda_a \tag{1.7}$$

$$[t^a, t^b] = if^{abc}t^c \tag{1.8}$$

The QCD Lagrangian is:

$$\mathcal{L}_{QCD} = \bar{q}(i\gamma^\mu D_\mu - m)q - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a. \tag{1.9}$$

Defining the covariant derivative

$$D_\mu = \partial_\mu + ig_s t_a G^a_\mu \tag{1.10}$$

and the strength field tensor $G^a_{\mu\nu}$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu, \tag{1.11}$$

where G^a_μ are the gluon fields, g_s corresponds to the gauge coupling constant and f^{abc} are the structure constants of SU(3) defined by the relation 1.8.

There are two properties that one can highlight in this theory. Asymptotic freedom - if quarks and gluons are at high energies, the interaction between quarks and gluons is weak, which means that QCD can be computable by perturbation theory and they can propagate as free particles. On the other hand, if they are at low energies, the coupling is strong and prohibits quarks and gluons from being free which leads to colour confinement - the second property of QCD.

1.1.3   Electroweak Theory

The combination of the electromagnetic and weak theories was achieved by Sheldon Glashow [3], Abdus Salam [4] and Steven Weinberg [5]. The search for this unification started after the Wu experiment [6] - where the discovery of parity violation in the weak interaction took place. This theory was found on a gauge theory with the group $SU(2)_L \otimes U(1)_Y$ - where L refers to the left-handed particles, which are the ones affected by weak interactions, and Y indicates the conserved charge in this theory, the hypercharge, which obeys

$$Y = 2Q - 2T_3 \qquad (1.12)$$

the (Gell-Mann-Nishijima) relation, in which T_3 stands for the weak isospin operator and σ_i are the Pauli matrices:

$$\hat{T} = \frac{\sigma_i}{2}, \text{with i=1,2,3.} \qquad (1.13)$$

The left and right-handed components of the fermions:

$$\Psi_R = \frac{1}{2}(1 + \gamma^5)\Psi$$
$$\Psi_L = \frac{1}{2}(1 - \gamma^5)\Psi \qquad (1.14)$$

These two components transform in different ways under the weak symmetry group - the right-handed fermions transform as singlets while left-handed fermions transform as doublets

$$f_R^i = l_R^i, u_R^i, d_R^i$$
$$f_L^i = \begin{pmatrix} l_L^i \\ v_L^i \end{pmatrix}, \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}, \qquad (1.15)$$

where i=1,2,3 corresponds to the fermion family. The Lagrangian of the gauge field is given by

$$\mathcal{L}_{gauge} = -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \qquad (1.16)$$

The field tensors $W_{\mu\nu}^i$ and $B_{\mu\nu}$ can be written as in

$$W_{\mu\nu}^i \equiv \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk}W_\mu^j W_\nu^k \qquad (1.17)$$

and in

$$B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \qquad (1.18)$$

respectively, where ϵ^{ijk} is the Levi-Civita tensor, W_μ^i is the gauge boson of $SU(2)_L$ and B_μ the gauge bosons of $U(1)_Y$. The $SU(2)_L$ gauge coupling is expressed by g.

One can now define the covariant derivative in order to maintain gauge invariance as

$$D_\mu \equiv \partial\mu - ig\vec{T} \cdot \vec{W}_\mu - ig'\frac{Y}{2}B_\mu, \qquad (1.19)$$

where g' is the coupling of the $U(1)_Y$ and $\vec{T}$ corresponds to the $SU(2)$ generators.

With this, the Lagrangian of the electroweak theory can be written as:

$$\mathcal{L}_{EW} = \sum_{f=l_L,q_L} \bar{f}(i\gamma^\mu D_\mu)f + \mathcal{L}_{gauge} \qquad (1.20)$$

5

The electroweak theory describes remarkably well the unification of the two fundamental forces - weak and electromagnetic. However, the gauge bosons fields in this theory appear without mass which is inaccurate, since it is known that they have mass [7][8]. This issue was later addressed by Robert Brout, François Englert and Peter Higgs that proposed that the masses arise as a result of the interaction between the particles with a field - the Higgs field - that permeates the universe (see section 1.1.4). [9]

1.1.4   The Brout-Englert-Higgs Mechanism

The mass problem of the gauge bosons was solved by the Brout-Englert-Higgs mechanism [9] [10] [11] that consists of a spontaneous symmetry breaking of the $SU(2)_L \otimes U(1)_Y$ to $U(1)_{EM}$. It was introduced the Higgs field - particles' interaction with this field originates their mass. This field is an isospin doublet of complex scalar fields and can be written as

$$\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix},$$ (1.21)

where ϕ^+ is related to an electrically charged field and ϕ^0 to an electrically neutral field.

The Higgs field can be described by a Lagrangian:

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi),$$ (1.22)

where D_μ corresponds to the covariant derivative given by equation 1.19, and $V(\Phi)$ is the Higgs potential which can be expressed by:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2$$ (1.23)

Considering μ^2 and λ to be real and positive numbers in order to achieve a spontaneous symmetry breaking, the Higgs potential can be represented by figure 1 and has a minimum of $|\Phi|^2 = \Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$ and a vaccum expectation value of $\frac{v}{\sqrt{2}}$.

Equation 1.21 can be rewritten as

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix},$$ (1.24)

where $H(x)$ corresponds to ground state fluctuations around the vacuum state.

The fermions' interaction with the Higgs fields can be described by the Yukawa Lagrangian 1.25, where :

$$\mathcal{L}_{Yukawa} = \sum_{f=l,q} y_f(\bar{f}_L \Phi f_R + \bar{f}_R \bar{\Phi} f_L)$$ (1.25)
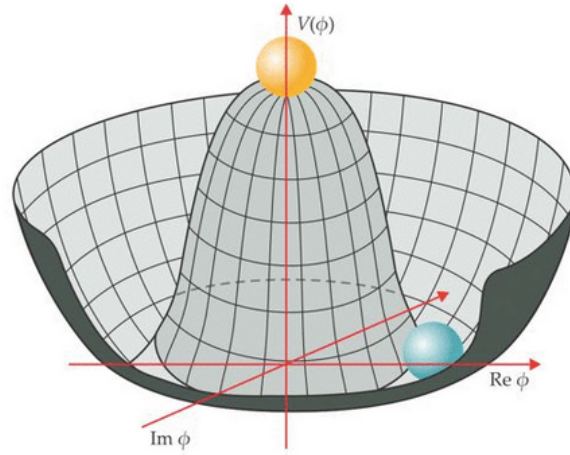
Figure 1: The Higgs Potential [12]

This Lagrangian is gauge invariant and y_f is the coupling between the fermions and the Higgs field. Finally, through Yukawa Lagrangian, Higgs field and Higgs Lagrangian, the mass of the fermions and the Higgs boson can be obtained:

$$
\begin{aligned}
m_f &= y_f \frac{v}{\sqrt{2}} \\
m_H &= \sqrt{2\lambda}v
\end{aligned}
\tag{1.26}
$$

The mass of the Higgs boson can only be obtained experimentally since λ is not known and it is not predicted by theory. In addition, the masses of the electroweak bosons can be obtained from the covariant derivative from the Higgs kinetic term:

$$
\begin{aligned}
m_W &= \frac{vg}{2} \\
m_Z &= v\frac{\sqrt{g^2 + g'^2}}{2} \\
m_\gamma &= 0
\end{aligned}
\tag{1.27}
$$

1.2 Beyond the Standard Model Particle Physics

Some phenomena such as the strong CP violation, matter and antimatter asymmetry and dark matter are not enlightened by the SM. BSM physics goal is to address these flaws and search for solutions for them. These extensions are proposed in an attempt to improve the SM, allowing it to explain nature in a more comprehensive manner.

1.2.1   Vector-like Quarks

Vector-like Quarks (VLQs) are an extension to the SM of semi-integer spin with mass higher than the other generations. This is due to the fact that past experiments must have centre of mass energies that were not able to reach enough energy produce these particles. They have the same righ-hand and left-hand charge components under the gauge group and they can mix with SM quarks - mostly with the 3rd family due to the large Yukawa coupling of the top quark - and have tree-level Flavor-Changing Neutral Currents (FCNC). Since quarks with chiral couplings were excluded by experiments, these VLQs must be non-chiral. [13] They can be singlets, doublets or triplets of SU(2) and can have SM charges of $\frac{2}{3}e$ or $-\frac{1}{3}e$ or exotic charges of $\frac{5}{3}e$ or $-\frac{4}{3}e$.

	Singlets		Doublets			Triplets	
VLQs	T	B	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\begin{pmatrix} X \\ T \end{pmatrix}$	$\begin{pmatrix} B \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$	$\begin{pmatrix} Y \\ B \\ Y \end{pmatrix}$
Charge	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{6}$	$\frac{7}{6}$	$-\frac{5}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$

Table 3: VLQ organized as singlets, doublets and triplets and their charge.

These particles can be single produce by electroweak interactions and pair-produced by QCD. Pair production is dominant for low masses whereas single production is dominant for high masses.

1.2.2   Flavor-Changing Neutral Currents

FCNC are processes where a fermion decays to a different flavor fermion with the emission of a neutral boson. These interactions are not allowed in SM, at tree-level, and beyond tree-level they are highly suppressed. Nevertheless, they can appear in high order correction or in loop-levels. BSM models also predict these interactions and they can occur at tree-level due to their high branching ratios. The branching ratios predicted by SM and BSM models - quark-singlet model (QS), two Higgs doublet model (2HDM), flavour-conserving two Higgs doublet model (FC 2HDM), minimal supersymmetric model (MSSM) and SUSY with R parity violation - are in table 4.

	SM	QS	2HDM	FC 2HDM	MSSM	$\not R$ SUSY
$t \to \mu Z$	8×10^{-17}	1.1×10^{-4}	-	-	2×10^{-6}	3×10^{-5}
$t \to \mu \gamma$	3.7×10^{-16}	7.5×10^{-9}	-	-	2×10^{-6}	1×10^{-6}
$t \to \mu g$	3.7×10^{-14}	1.5×10^{-7}	-	-	8×10^{-5}	2×10^{-4}
$t \to \mu H$	2×10^{-17}	4.1×10^{-5}	5.5×10^{-6}	-	10^{-5}	$\sim 10^{-6}$
$t \to c Z$	1×10^{-14}	1.1×10^{-4}	$\sim 10^{-7}$	$\sim 10^{-10}$	2×10^{-6}	3×10^{-5}
$t \to c \gamma$	4.6×10^{-14}	7.5×10^{-9}	$\sim 10^{-6}$	$\sim 10^{-9}$	2×10^{-6}	1×10^{-6}
$t \to c g$	4.6×10^{-12}	1.5×10^{-7}	$\sim 10^{-4}$	$\sim 10^{-8}$	8×10^{-5}	2×10^{-4}
$t \to c H$	3×10^{-15}	4.1×10^{-5}	1.5×10^{-3}	$\sim 10^{-5}$	10^{-5}	$\sim 10^{-6}$

Table 4: Top FCNC decays and their branching ratios. [13]

Over the years, several searches for FCNC interactions have been done at Tevatron and LHC. [14] [15] Searches performed at ATLAS and CMS gave rise to expected and observed limits of 95% CL for tZ production, $t\bar{t}$ decay and for the tZ production and $t\bar{t}$ decay combined - tZq anomalous couplings. [16] [17] [18] [19] Furthermore, ATLAS had performed searches for tgq and CMS for $t\gamma q$. [20] [21]

Experimental Setup

2.1 CERN

CERN (European Organization for Nuclear Research) is a particle physics research organisation located in Geneva, Switzerland. It is the home of the largest particle collider, the Large Hadron Collider (LHC), with four main detectors that provide the experimental data necessary for physics research. Although the main purpose of CERN is fundamental research, it also has a role in technological development.

Before getting to LHC, the beam of particles is accelerated in a chain of older accelerators, with increasingly higher energies. A schematic representation of these accelerators is in figure 2.

Protons are first accelerated at Linear accelerator 2 (Linac2) where they get to an energy of 50 meV and have an increase of 5% in mass. They then get to Proton Synchrotron Booster where they are accelerated to 1.4 GeV and injected into Proton Synchrotron (PS). In this stage, the protons are hastened up until they get an energy of 26 GeV before they get to the next stage, the Super Proton Synchrotron (SPS), and get to an energy of 450 GeV. After these stages, the protons are injected into LHC (see section 2.2).

2.2 LHC

LHC is a 26.7 km perimeter circular particle physics accelerator located underground. In this accelerator, two proton beams that emerge from different beam pipes travel through it, maintaining their direction with the cooperation of strong magnetic fields. These beams collide at specific points throughout their path, being them the four main detectors of LHC: ATLAS, CMS, ALICE and LHCb. Containing different components, these detectors are used for different ends: ATLAS and CMS are both multi-purpose detectors and therefore the ones that are more meaningful to this thesis, LHCb is used to study the b-quark and ALICE is optimized for the context of heavy-ions collisions.
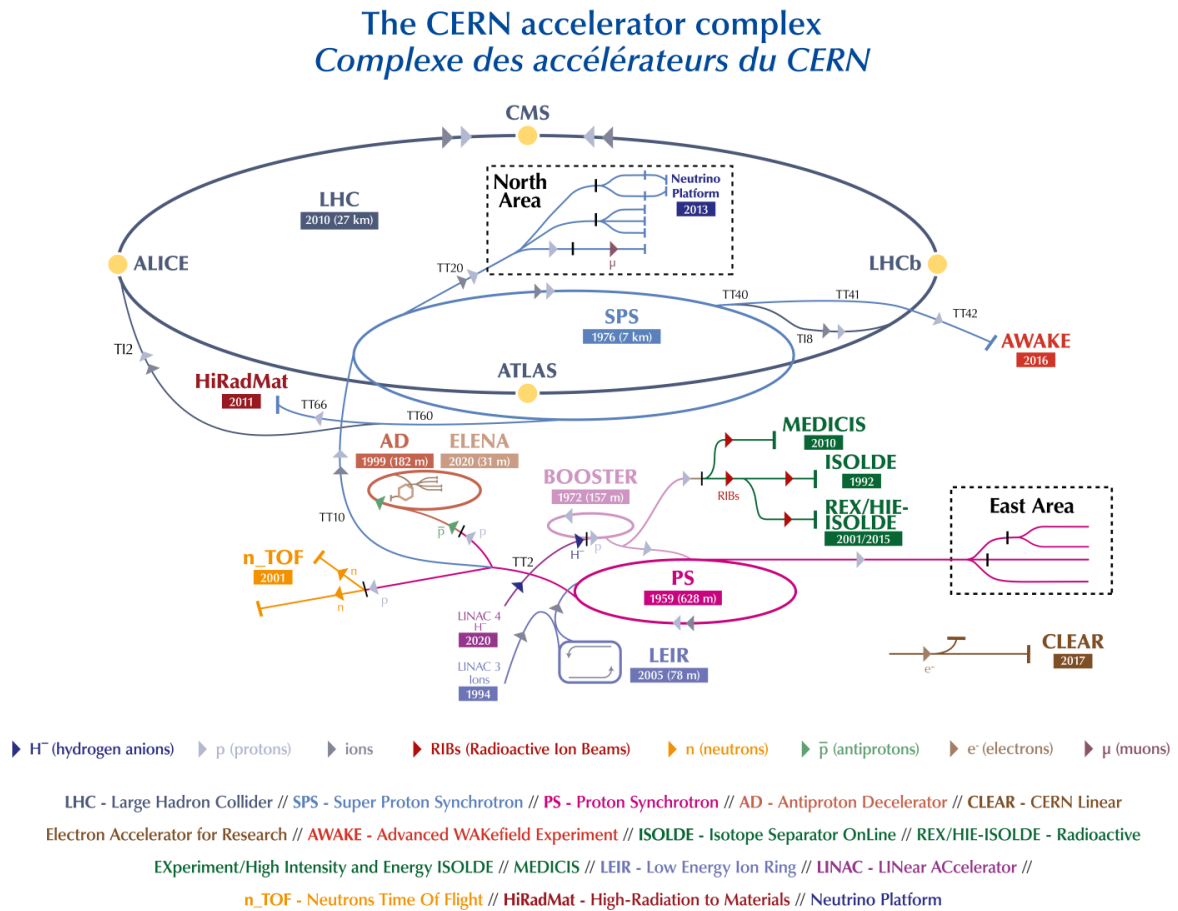
Figure 2: Schematic representation of CERN accelerators. [22]

The first operational run of the LHC occurred from 2009 to 2013 and it peaked at a combined luminosity of 7 TeV. The discovery of the Higgs boson in 2012 was of significant importance, during this run. After a shutdown of around 2 years, the second run started in 2015 and ended in 2018 with an operating energy of 13 TeV of combined energy. This year, 2022, the third run of the LHC started and it already reached a combined energy of 13.6 TeV.

The study of rare events is very ambitious, considering that the proportion that these events occur is very small when compared to SM physics. The number of particles available to collide at a certain time is measured by luminosity, and this quantity appraises the performance of a collider. Writing:

$$\frac{dR}{dt} = L \times \sigma_p, \tag{2.1}$$

luminosity L is the proportional factor of the number of events per second and the production cross-section. Considering the two beams colliding, luminosity can be written as:

$$L = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y}. \tag{2.2}$$

The integral of the instantaneous luminosity over time gives the integrated luminosity:

$$L_{\text{int}} = \int L dt. \tag{2.3}$$
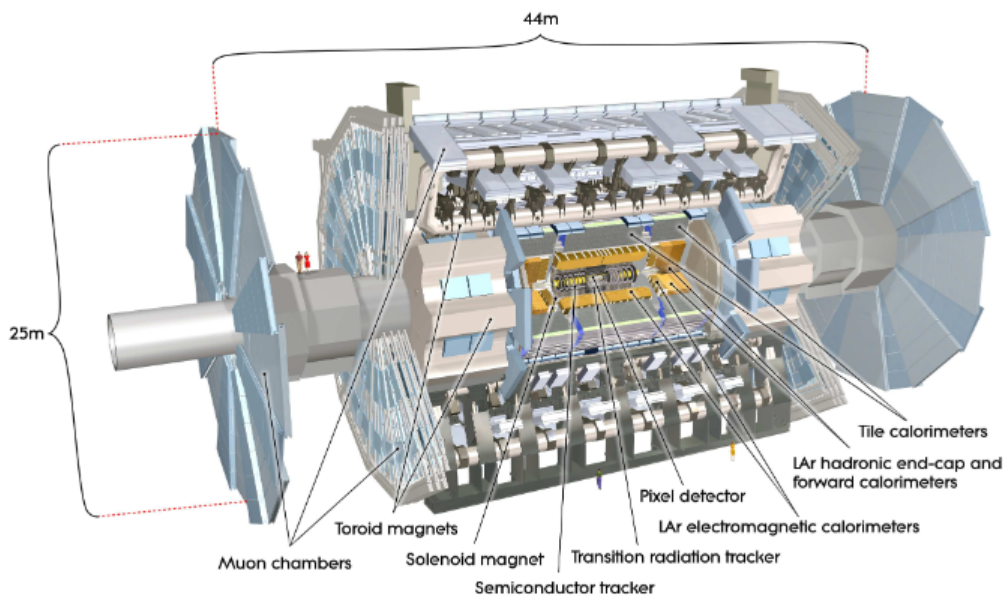
2.2.1 ATLAS

Figure 3: The ATLAS detector [23]

ATLAS [23], represented in figure 3, is a multipurpose detector that studies proton-proton collisions. It has the shape of a cylinder with 46 meters long, 25 meters in height, and is located 100 meters below ground. It is formed by different components and each one of them has a specific design and goal - inner detector, calorimeter system, muon spectrometer.

The ATLAS has a right-handed coordinate system in which the z-axis points towards the beam pipe, the x-axis points toward the centre of the LHC and the y-axis points upwards. The x and y axis form the xy plane that in transverse to the beam pipe and therefore it is called the transverse plane. The angle ϕ is the azimuthal angle and corresponds to the angle around the beam pipe while the angle θ is the polar angle and corresponds to the angle between the z-axis and the transverse plane. One can define the rapidity as:

$$y = \frac{1}{2} \ln(\frac{E + p_z}{E - p_z}), \tag{2.4}$$

where E is the energy of the particle and p_z is the component along the z-axis of the momentum of the particle. For a massless particle:

$$y = \frac{1}{2} \ln(\frac{1 + \theta}{1 - \theta}) = \ln(\cot \frac{\theta}{2}). \tag{2.5}$$

It leads to define pseudorapidity that is written as a function of the polar angle θ that does not depend on the energy of the particle and is a purely geometric quantity:

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

(2.6)

and η=0 for particles that travel perpendicular to the beam axis and either positive or negative if the particle travel with a certain angle to the beam axis.

The transverse momentum corresponds to the momentum measured in the transverse plane xy:

$$p_T = \sqrt{p_x^2 + p_y^2},$$

(2.7)

and the missing transverse energy is also the missing energy measured in the transverse plane xy:

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

(2.8)

where i is the final state of the particles.

The distance between two particles can be written using the azimuthal angle and the pseudorapidity:

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

(2.9)

The performance goals for each component of the ATLAS detector are specified in table 5.

ATLAS detector component	Resolution	Coverage
Tracking	$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%$	$|\eta| < 2.5$
Electromagnetic Calorimeter	$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%$	$|\eta| < 3.2$
Hadronic calorimeter barrel	$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%$	$|\eta| < 3.2$
Hadronic calorimeter end-cap forward	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%$	$3.1 < |\eta| < 4.9$
Muon spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV	$|\eta| < 2.7$

Table 5: Performance goals of the ATLAS detector. [23] The energy and transverse momentum are in units of GeV.

2.2.1.1 Inner Detector

The inner detector (ID) objective is to track charged particles, measure their momentum and obtain their primary and secondary vertex above a given threshold for p_T (typically $p_T \approx 0.5$ GeV) and pseudorapidity $|\eta| < 2.9$. A representation of this detector is given in figure 4. The ID is contained in a magnetic field supplied by a solenoid of 2T and the figure 5 shows the sensors and structure of the detector.

The ID is itself composed of other complementary detectors that are listed in figure 5 - insertable B-layer (IBL), pixel detector, semiconductor tracker (STC), and the transition radiation tracker (TRT).
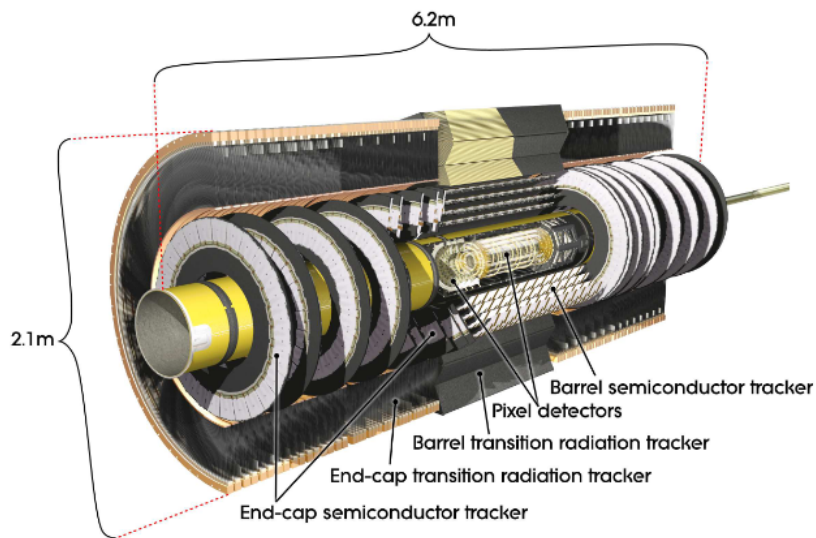
Figure 4: The ATLAS inner detector [23]

The IBL has been operating since 2015 and it improved significantly the ATLAS tracking performance. [24] This detector is constituted by 14 carbon fiber staves, each one with dimensions of 2 cm wide and 64 cm long, covering a pseudo-rapidity of $|\eta| < 3$ and has around 26880 pixels. The pixel detector has 4 layers of sensors in the barrel region and three pixels disks in the end-cap region. The PD has a high contribution to the detection of charged particles.[25] The STC consists of 4 cylindrical barrel layers of silicon detectors and 18 disks distributed over the 2 endcaps while the TRT has 4 mm diameter straw tubes filled with gas that provide information about the particles due to the transition radiation property.

2.2.1.2 Calorimeter System

The calorimeter system, represented in figure 6, gives an accurate measure of the energy of the particles, taking into consideration the deposits that particles leave as they advance throughout the detector. It also provides information about the particles themselves, as they interact differently with the material. This property is the reason behind the existence of two main components of the calorimeter system: the electromagnetic calorimeter which provides information about the particles that have electromagnetic interactions with the detector and the hadronic calorimeter which gives information about the particles that have strong interactions with the detector.

The electromagnetic calorimeter has liquid argon as its medium and it is composed of a barrel and two end-caps. The barrel itself is an aggregation of two half-barrels separated by a small gap of 4 mm at z=0. The organization of the electrodes and lead absorber plates grants it full coverage and it achieves a complete ϕ symmetry.

The hadronic calorimeter is composed of the tile calorimeter, LAr hadronic end-cap calorimeter and LAr forward calorimeter. The tile calorimeter surrounds the electromagnetic calorimeter and uses steel as
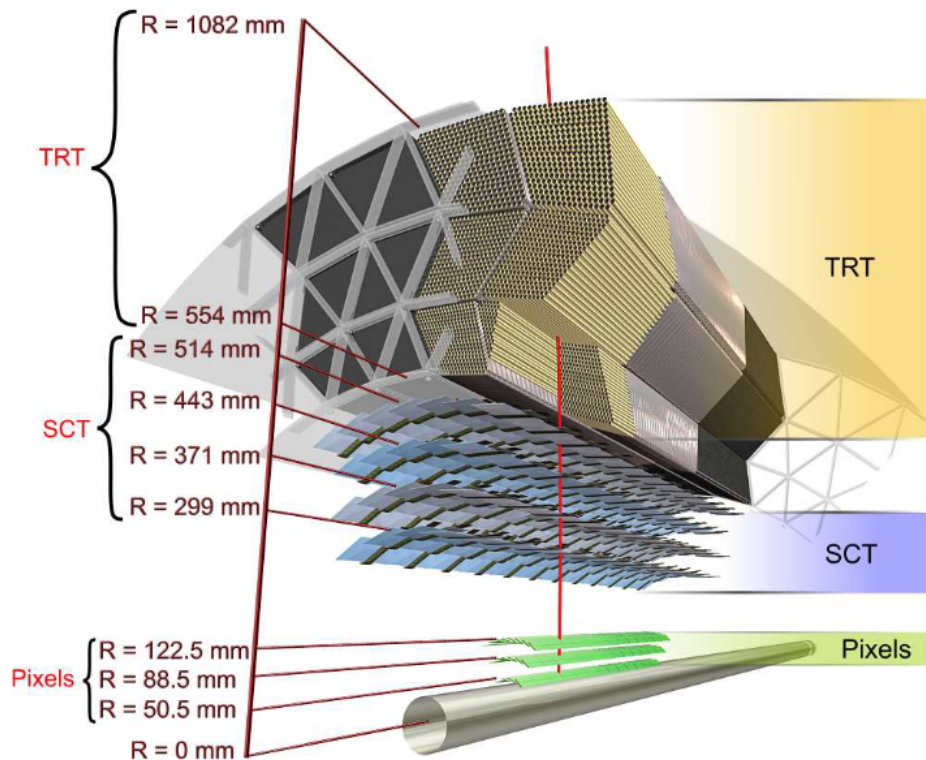
Figure 5: The ATLAS inner detector sensors and structure [23]

Calorimeter System		Coverage
Electromagnetic calorimeter	Barrel region	$\lvert\eta\rvert < 1.475$
	End-cap region	$1.375 < \lvert\eta\rvert < 3.2$
Hadronic calorimeter	Tile calorimeter	$\lvert\eta\rvert < 1.7$
	Hadronic end-cap calorimeter	$1.5 < \lvert\eta\rvert < 3.2$
	Forward calorimeter	$3.1 < \lvert\eta\rvert < 4.9$

Table 6: η coverage of the calorimeter system components

the absorber and scintillating tiles as the active material. The LAr hadronic end-cap calorimeter is located behind the end-cap electromagnetic calorimeter and consists of two wheels, one per end-cap. The material close to the interaction point is built of copper, interleaved with LAr gaps. The LAr forward calorimeter consists of modules made in copper, optimized for electromagnetic measurements, and tungsten, optimized for hadronic measurements.

The information about the coverage of the components of the calorimeter system is mentioned in table 6. This system was designed to stop most of the particles that go through it. Despite this, muons and neutrinos go through it without stopping. Muons are later detected in the muon spectrometer, however, neutrinos do not leave signatures nor interact with any component of the ATLAS detector - their presence is induced by the measure of the missing transverse momentum E_T^{miss}, due to the conservation of the
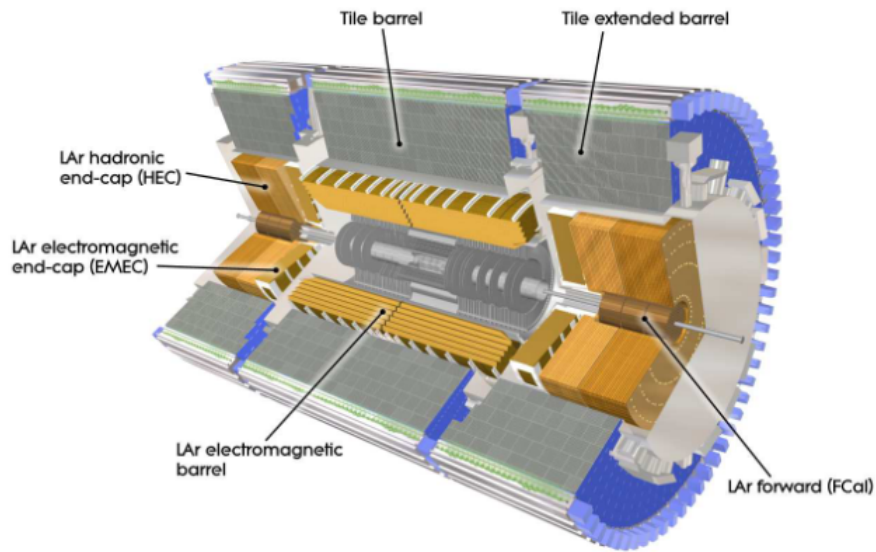
Figure 6: The ATLAS calorimeter system [23]

total momentum of the event.

2.2.1.3   Muon Spectrometer

The Muon Spectrometer, represented in figure 7, is the outer detector and its goal is to detect muons and measure their momentum. It relies on the magnetic deflection of muon tracks surrounded by a toroid magnet system. At a pseudorapidity of $|\eta| < 1.4$, the bending is done by the large barrel toroid while for a pseudorapidity between $1.6 < |\eta| < 2.7$ the bending is provided by two small magnets located at each end of the barrel toroid. In the transition region, $1.4 < |\eta| < 1.6$, the muon tracks are bent by the combination of the toroid barrel and the small magnets. The tracks are measured in chambers that are distributed over the barrel, transition and end-cap regions - Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC).

2.2.1.4   Trigger System

The trigger system selects the events according to defined criteria. It is composed of three levels: level 1 (L1), level 2 (L2) and the event filter.

L1 searches for events such as muons, electrons, photons, jets and τ leptons with high transverse-momentum decaying into hadrons, events with large missing energy and large total energy. Trigger chambers in the barrel and end-cap regions of the spectrometer are used to select high transverse-momentum muons. For calorimeters, it is used reduced-granularity selection method collecting information on all calorimeters. In each event, this level defines one or more regions of interest (RoI), that include information about the type of feature selected and the criteria it passed. This information is passed to the
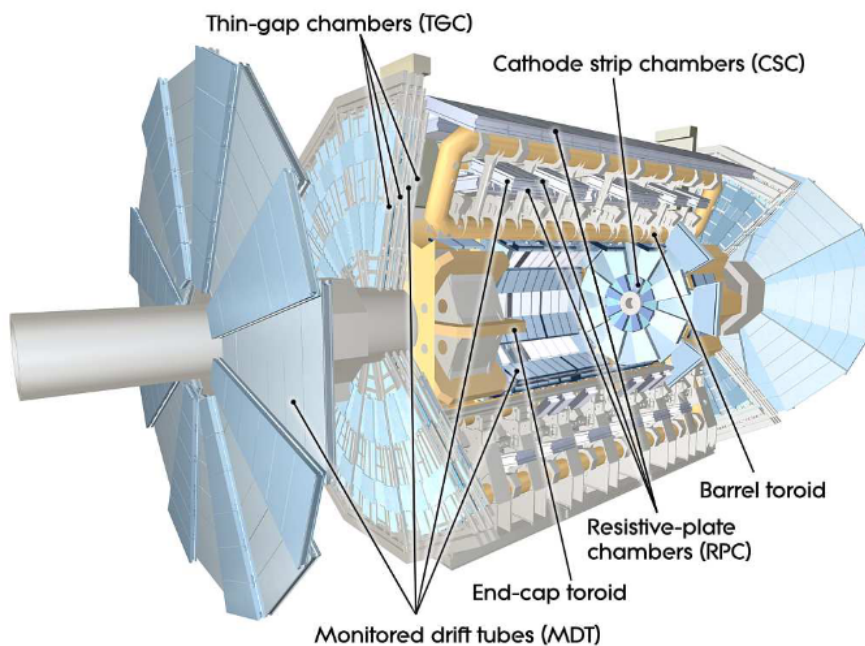
Figure 7: The ATLAS muon system [23]

high-level trigger (HLT).

In L2, selection criteria are applied at full precision and all detector data in the RoI available is considered.

The last stage, the event filter, consists of an offline analysis operation.

Considering that the bunch-crossing rate at ATLAS is about 40 MHz per second, the first level L1 aims to reduce the rate of events to about 75 kHz and the HLT to a rate of about 200 Hz.

2.2.2 CMS

As ATLAS, CMS is a multipurpose detector and it is represented in figure 8. This detector is 21.6 meters long and 14.6 meters in diameter. Each one of the components of this detector - the inner detector, calorimeters and muon system - has a different scope and layout. [26] The general performance of the components of the CMS detector are in table 7.

2.2.2.1 Superconducting Magnet

The CMS superconducting magnet is arranged to provide a 4T field in a 12.5 m of length and 6.3 m of inner diameter solenoid. The coordinate system is defined the same way as ATLAS, so the magnetic field is pointing towards the beam axis or z-axis. It has a stored energy of 2.6 GJ at full current. Outside of it, the flux returns through a 10 000 t yoke that has 5 barrel wheels and 2 end-caps with 3 end-cap disks each.
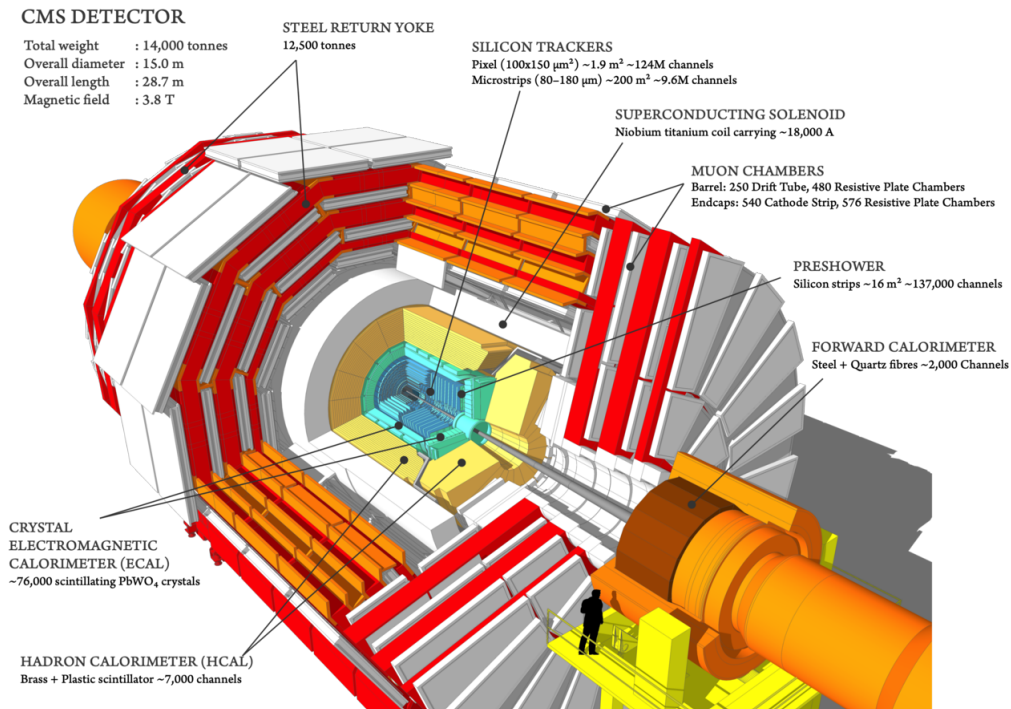
CMS DETECTOR
Total weight   : 14,000 tonnes
Overall diameter : 15.0 m
Overall length   : 28.7 m
Magnetic field   : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm²) ~1.9 m² ~124M channels
Microstrips (80–180 μm) ~200 m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000 A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16 m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 8: The CMS detector [27]

CMS detector component	Resolution	Coverage
Tracking	$\frac{\sigma_{pT}}{pT} = 1.5 \times 10^{-4} pT \oplus 0.005$	$|\eta| < 2.5$
Electromagnetic calorimeter	$\frac{\sigma_E}{E} = \frac{3\%}{\sqrt{E}} \oplus 0.3\%$	$|\eta| < 3$
Hadronic calorimeter	$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 5\%$	$|\eta| < 5.2$
Muon spectrometer	$\frac{\sigma_{pT}}{pT} = 10\%$ at $p_T = 1$ TeV	$|\eta| < 2.4$

Table 7: Performance goals of the CMS detector [26]. The energy and transverse momentum are in units of GeV.

2.2.2.2   Inner Detector

The main purpose of the inner detector is to track charged particles with high precision, trace their trajectories and reconstruct their secondary vertex.

The ID has 2 parts - pixel detector and strips detector. Their construction is based on silicon technology since it is required that the detector can resist harsh environments originated from the intense flux of particles that cause radiation damage to the system. The pixel detector is composed of 3 barrel layers that cover a radius from 4.4 cm to 10.2 cm and 2 end-cap disks, one on each side of the barrel. The silicon strip tracker is formed by 10 layers in the barrel region, 4 of them being inner barrels and the other 6 outer barrels, and 3 plus 9 disks on each side of the barrel, covering $|\eta| < 2.5$. The silicon tracker has an active region of 200 m².

2.2.2.3 Calorimeters

CMS has 2 calorimeters: an electromagnetic calorimeter and a hadron calorimeter.

The electromagnetic calorimeter is made of tungstate ($PbWO_4$) crystals located in the central part of the barrel and in the 2 end-caps. The photodetectors in the barrel are made of Avalanche photodiodes (APDs) while in the end-caps they are made of vacuum phototriodes (VPTs). This detector has a particularity of being a homogeneous crystal calorimeter which provides a good energy resolution. It covers the pseudorapidity region of $|\eta| < 3$. [28]

The hadron calorimeter's target is to measure the energy of hadron jets and the missing transverse energy that results from neutrinos or exotic particles. This calorimeter's barrel is radially restricted outerly by the electromagnetic calorimeter and innerly by the magnet coil. This has a direct impact on the amount of material that can be used to absorb the hadronic shower and therefore, another hadron calorimeter is situated outside the solenoid in order to extend the barrel. The forward hadron calorimeter provides a coverage of $|\eta| < 5.2$ using Cherenkov radiation method within an active medium of quartz fibres, where the charged shower particles above the Cherenkov threshold that travel through this calorimeter generate Cherenkov light and produce a signal. [28]

2.2.2.4 Muon System

The CMS has a cylindrical barrel and 2 planar end-caps and it is composed of 3 different types of gaseous detectors to reconstruct muons' momentum and charge. Another function of this detector is triggering, and the success of these measurements and trigger are provided by the high-field solenoidal magnet and its flux-return yoke.

In the barrel region the flux of muons is low and therefore the 4T magnetic field is uniform. The barrel drift-tube (DT) chambers covers $|\eta| < 1.2$ and the arrangements of the chambers is able to, with good precision, measure the muon time.

In the 2 end-caps, the flux of muons is high and this system uses cathode strip chambers (CSC). It covers $0.9 < |\eta| < 2.4$. The layers of CSC gives enough statistics that allows to reject non-muon backgrounds.

Resistive plate chambers (RPC) were added as a complementary trigger system, in both barrel and end-cap regions in order to produce a fast response with good resolution. There was a total of 6 layers of RPCs settled - 2 in each one of the first 2 stations and 1 in each one of the last 2 stations of the barrel muon system - and a plane of RPCs in each of the first 3 stations of the end-cap muon system. This allows to improve time resolution for bunch crossing identification and it reduces background using concurrences between stations.

The muons system has a coverage of $|\eta| < 2.4$ and the overall offline reconstruction efficiency of simulated single-muon samples is between 95 and 99% for almost the entirety of this pseudorapidity coverage except in the regions surrounding $|\eta| = 0.25$ and $|\eta| = 0.8$.

2.2.2.5   Trigger system

The trigger system is composed of two steps - Level-1 (L1) Trigger and High-Level Trigger (HLT). The L1 is the first level of data selection that uses custom eletronic systems and has a latency of 3.2 μs. HLT is a software trigger that reduces the data output of L1 trigger to a more pleasant size for storage and further physics analysis. The combined L1 and HLT rate reduction capability is designed to be around 10^{-6}.

L1 uses coarsely data from calorimeter and muon system whereas HLT uses completely read-out data hence it can execute more complex calculations.

L1 has local, regional and global components. Trigger Primitive Generators (TPG) receive information on energy deposits in calorimeter trigger towers and track segments on muon chambers. Regional triggers combined the information and use pattern logic to determine object candidates in limited regions. The Global Calorimeter and Global Muon Triggers determine the highest rank calorimeter and muon object and transfers them to the Global Trigger, that has the job to either select or reject them. If selected, it passes to the HLT for further evaluation.

Machine Learning

Machine learning is a subfield of artificial intelligence that uses computer algorithms to learn and study data and retrieve knowledge from it. These algorithms build a model based on a sample data that aims to make predictions. There are many types of learning: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning, among others.

3.1 Supervised Learning

Supervised learning uses labeled inputs to train algorithms that improve over time and create a model that is used to make predictions. As it is being trained, a parametric model has its weights updated until it fits properly the data. This fitting is monitored so that it does not fall into an over-fitting or under-fitting problem.

One can deal with two types of problems when using supervised learning: classification problem or regression problem. Classification problems aim to categorize the data into their specific category. It tries to learn patterns or other entities in the data that could point them into a specific label. Some examples of algorithms used in classification methods are linear classifiers, decision trees, random forests. Regression problems aim to find a relation between an independent and a dependent variable and it is commonly used for forecasting. Some algorithms that are used in these types of problems are linear regression and decision trees.

3.1.1 Decision Tree

Decision Tree - a non-parametric supervised learning method - is an algorithm that splits data recursively due to a certain condition until it gets to a terminal node. A representation of a decision tree is shown
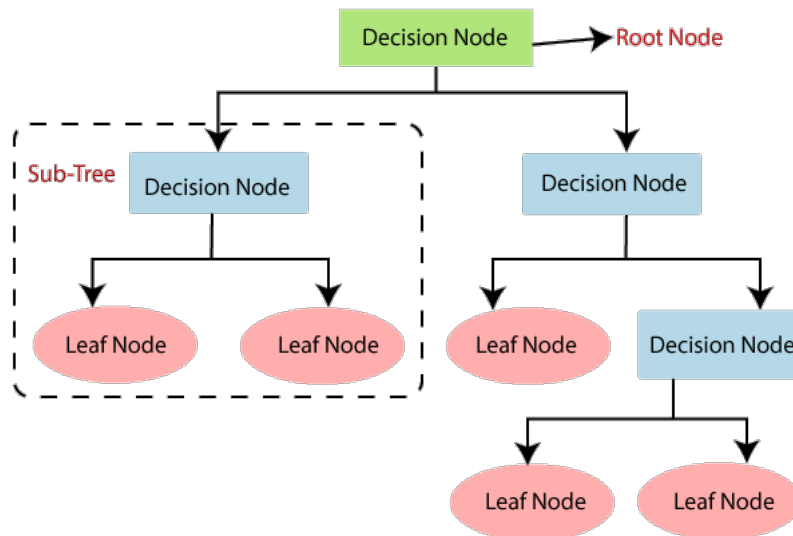
Figure 9: Representation of a decision tree and its elements.[29]

in figure 9.

Nodes and leaves are a part of this algorithm - representing a tree-like structure - and they have different meanings: a node corresponds to a condition that splits the data based on whether they meet or do not meet the condition until it gets to a leaf that correspond to terminal nodes and they display the probability or label of the class, for classification problems.

The trees can be created recursively to form an ensemble of trees. Its output is the weighted sum:

$$\hat{y}(x) = \sum_t w_t h_t(x), \tag{3.1}$$

being w the weight and h the tree's output. The goal of the ensemble of trees is to minimize their objective function:

$$O(x) = \sum_i l(\hat{y}_i, y_i) + \sum_t \Omega(f_t), \tag{3.2}$$

where $l(\hat{y}_i, y_i)$ is the loss function which is the distance between the value predicted and the true value for the ith sample and $\Omega(f_t)$ is the regularization function that measures the penalty of the model based on the complexty of the tree.

A decision tree can be boosted in order to help with the minimization of the loss function. There are some types of boosting such as Adaboost that stands for "Adaptive boosting", Gradient Boosting that uses gradient descent and XGBoost that means "eXtreme Gradient Boosting"that uses a gradient boosting.

The tree has some adjustable parameters that aim to increase accuracy and prevent overfitting. One can define:

- The maximum depth of the tree.

- The maximum number of features that can be used by the tree.

- The minimum number of samples per leaf, that creates a threshold that needs to be achieved in order to make a new leaf.

- The loss function.

- The learning rate, that defines the amount of adjustment that the tree can apply to the data weights after each interaction.

- The number of trees.

Using boosted decision trees can have many benefits such as the increase of the train speed, it is easy to tune, it has good performance, among others, however, it can easily fall into overfitting problems, so it is very important to cross-validate while training.

3.2 Anomaly Detection

Anomaly detection aims at find data points that do not follow the same pattern as the rest of the data. There are two main methods used in anomaly detection - outlier detection and novelty detection. In outlier detection, the dataset that is given to the model has outliers as observations and the model tries to fit the data that is concentrated, ignoring deviant data points. In novelty detection, the dataset does not contain outliers and the model tries to find if a new observation is an outlier (commonly called novelty) or not. The work done in this thesis fits in novelty detection and aims to identify novelties. This identification can be done with machine learning tools, such as autoencoders, for instance.

Autoencoders are [30], in literature, an unsupervised machine learning technique. It consists of two neural networks that are connected by a bottleneck, called latent space or code layer, that has variable dimension. Before getting in the details of autoencoders, lets define a neural network.

3.2.1 Neural Network

A neural network (NN) is a structure inspired by the biological brain, composed by an input layer that is fed with the data, hidden layers and the output layer. These layers are itself composed by nodes or neurons that connect to another nodes from the neighbor layers. In figure 10 there is a representation of a deep neural network.

Each node has associated a weight and a bias and, when it receives an input, it assigns the weights that establish the importance of the variable and then sums the bias. One can describe this as:

$$y = \sum_i w_i x_i + b, \tag{3.3}$$

where x is the input data, w is the weights and b is the bias.
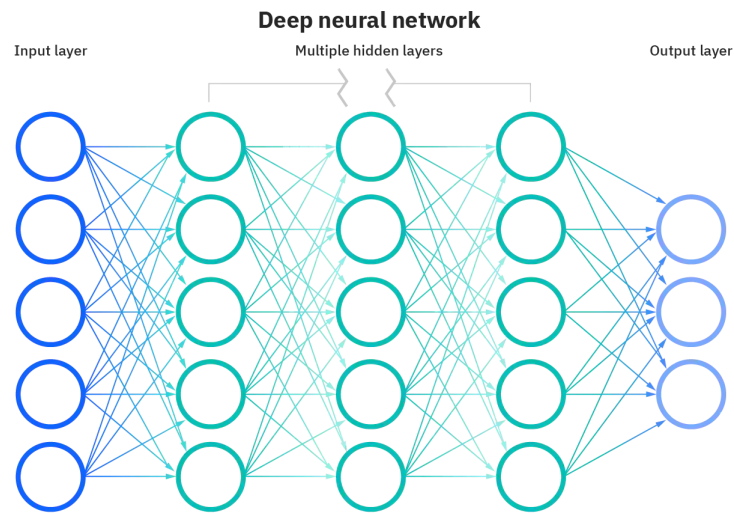
23

Figure 10: Deep neural network representation. [31]

The output then goes to the next node, passing through an activation function whose objective, in this work, is to allow the NN to learn non-linear relations. The activation functions that are important in this work:

- Rectifier linear unit (ReLU) activation function

$$f(x) = \max(0, x), \tag{3.4}$$

- Leaky rectifier linear unit (Leaky ReLU) activation function

$$f(x) = \begin{cases} x & \text{if} x > 0 \\ 0.01x & \text{otherwise.} \end{cases} \tag{3.5}$$

ReLU activates the nodes that have a non-zero output and has better gradient propagation which means that does not fall so easily into vanishing gradient problems. The Leaky ReLU allows the output to be a small positive gradient when the output of the node is less than zero.

As the model is being trained, the evaluation of its accuracy is done recurring to a loss (or cost) function. This function measures the quality of the predictions - as the model adjusts its weights and bias - with a differentiable function. An important loss function for the problem addressed in this work is the mean squared error:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{3.6}$$

Then the weights are iteratively corrected using stochastic gradient descent. This is done over averages of mini-batches:

$$w^{t+i} = w^t - \eta \nabla L. \tag{3.7}$$

The inputs of a neural network need preprocessing - the usual way is to standardise them, setting the mean to 0 and the standard deviation to 1.

$$x \rightarrow \frac{x - \bar{x}}{\sigma_x}. \tag{3.8}$$

where $\bar{x}$ is the mean of the input values and σ_x is their standard deviation.

This needs to be done in order to prevent grandient explosion since stochastic gradient descent updates the weights layer by layer and if there is large numbers, it will lead to large activations and therefore large updates.

3.2.1.1   Avoid overfitting

When training, it is very common for a neural network to learn about statistical fluctuations and noise, leading to overfitting to the training data. One way to avoid these kind of flaws is to use an early stopping method associated with a cross validation. In order to fulfill this, the dataset needs to be divided into three sets - a training set, used for training, a validation set, used to cross-validate, and a test set, used to produce the final results.
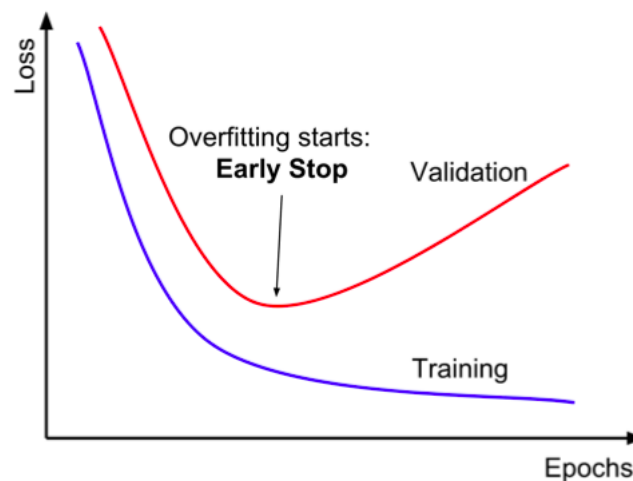

Figure 11: Loss function curves of the train and validation set. [32]

While training on the training set, the loss function will keep minimizing but, if one calculates the loss function on the validation set parallel to the calculation of the loss function on the train set, there will be a point where the validation loss will not improve (see figure 11). It means that the best model is the one that corresponds to the minimum validation set loss. A way of doing this is using an early stopping method that will stop the training when the loss function of the validation set worsens - but it needs to

have a patience associated to make sure that the minimum of the validation set loss is a global minimum and not a local minimum.

Normalization techniques can affect the stabilization of the NN and allow the NN to be more profound. Batch normalization normalizes each feature of the data organized in mini-batches with a certain size, while training, setting the mean of the output to a value close to 0 and the standard deviation to a value close to 1. The layer normalization normalizes each input of the data of the batch over all features.

3.2.1.2   Hyperparameters and Metrics

As in the decision tree, neural networks also have some adjustable hyperparameters. One can define:

- The number of hidden layers.

- The number of units.

- Dropout, used to avoid overfitting, randomly selects neurons to be ignored during training at a given dropout rate.

- The activation function.

- The learning rate, that sets the step of the backpropagation.

- The normalization - batch normalization or layer normalization, for instance.

A way to know if a set of hyperparameters is good is to evaluate the performance of the model using metrics. Some machine learning metrics that are useful are accuracy, precision, recall, receiver operating characteristic (ROC), mean squared error (mse), amongst others.

The ROC curve is a plot of the true positive values (TP) against the false positive values (FP). An ideal classifier would have the TP equal to 1 and the FP equal to 0 while a random classifier would have both at 0.5. A representation of a ROC curve is on figure 12.

Another measure that can be useful when plotting a ROC curve is the area under the ROC curve or AUC. This measure is ideally close to 1 and if it is a random classifier it will be close to 0.5.

The mean squared error can be defined as it was in equation (3.6), and it measures the average of the squares of the errors, or in other words, the average of the squared difference between the estimated value and the true value.

3.2.2   Autoencoders

Autoencoders are a field of study of artificial intelligence and they are applied in several problems, such as feature detection, anomaly detection and others. It is formed by two NN connected by a latent space with variable dimension that affects the performance of the model. The first neural network of the autoencoder is called the encoder and its function is to compress the data until it gets to the latent space or code layer.
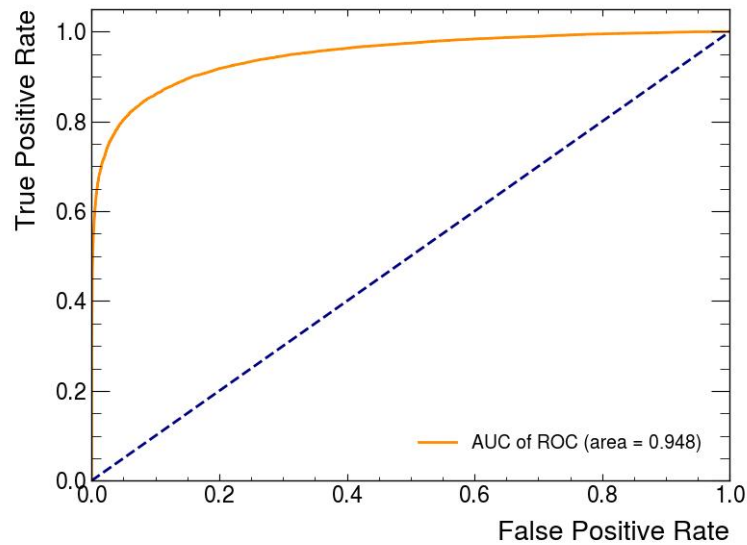
Figure 12: Example of a ROC curve and the area under the ROC curve

Then, the decompression of the data is done on the second neural network that is called the decoder. In figure 13 there is a representation of an autoencoder.
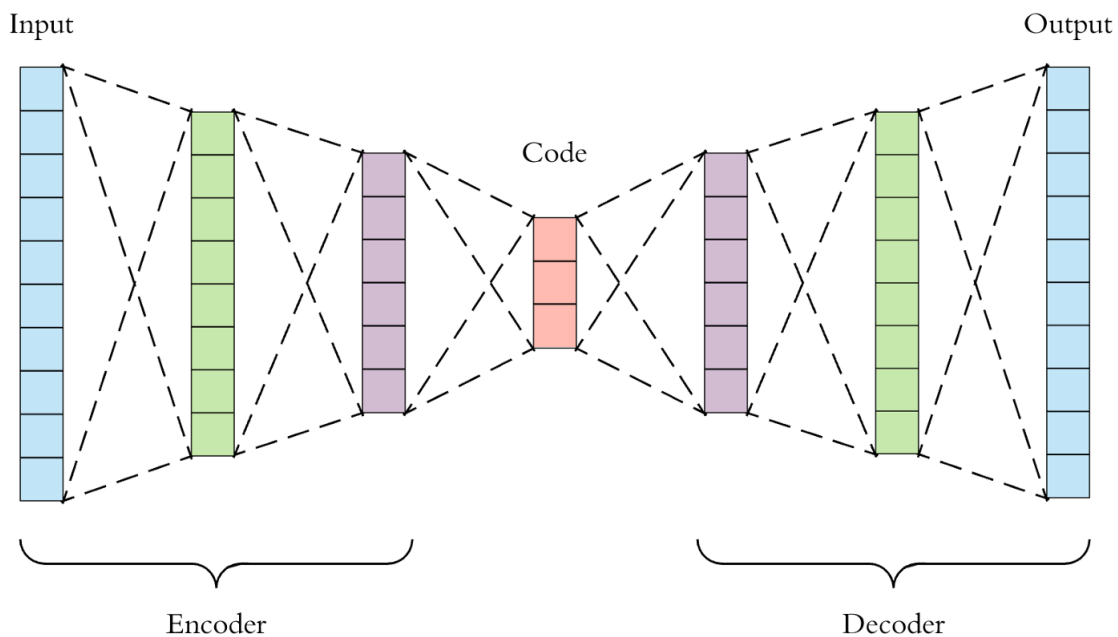

Figure 13: Representation of an autoencoder. [33]

The aim of the autoencoder is to get the output as close as it can be to the input - create a model that can reproduce the input features. Varying the dimension of the latent space will affect the performance of the autoencoder - and the hyperparameters of the encoder and decoder will affect it as well. So it is very important to optimize these hyperparameters - this was done with optuna, which is an hyperparameter optimization software framework (see section 3.2.3).

It is important to notice that, as said before, autoencoders are in literature an unsupervised machine learning method. However, in the study that was performed in this work, the inputs that were fed to the autoencoder were background data - so it was known prior to the training that it corresponded to background. This makes our approach to be that of semi-supervised learning rather than unsupervised.

3.2.3 Optuna optimization

Optuna [34] is a framework designed particularly for machine learning problems. Hyperparameters is the name given for the fixed parameters of the model that are assigned before training. Optuna aims to optimize these hyperparameters.

It defines trial as a single execution of the objective function and study as an optimization that has as a basis an objective function. Overall, the optimization process consists on a study that with many trials tries to find the optimal set of hyperparameters that suits the model. The objective function is the function that the study is going to optimize. In this function, the metrics that are going to be optimized need to be defined - such as the number of layers, number of units, the activation function, normalisation, and others. When the study is created, the number of trials can be defined as well as the aim of the study - one can set the study to maximize the R^2 score, for instance.

It works using *Define-by-run API*, that allows to run deep neural networks dynamically - this means that the user is able to select the search space actively, whilst doing the optimization.

Optuna also has effective ways of wiping out bad trials, using early stopping or pruning methods. The early stopping method was already discussed before. The pruning method consists of dismissing a trial if the predefined condition is not achieved. For example, the median pruner looks at the intermediate value of a trial - if the best intermediate value on a certain step is not better than the median of the intermediate values of the previous trials, then the trial is dismissed. Pruning optimizes the study and makes the optimization process faster.

Results

In this work, using data from simulated pp collisions at 13 TeV, it was done a comparison between a supervised machine learning method - BDT - and an anomaly detection method - AE. [35]

4.1 Dataset

The dataset [35] used contains simulated data of pp collisions at 13 TeV with 2 leptons and 1 b-jet in the final state and selected benchmark BSM signals. The data considered has $H_T > 500$ GeV. It contains SM samples generated at leading order that contain the sub-samples Z plus jets, $t\bar{t}$, WW, WZ and ZZ. To guarantee a good amount of statistics, the processes were generated in kinematic regions. The event generation filters at parton level used for the sampling:

- $t\bar{t}$ with $p_T < 100$ GeV, $100 < p_T < 250$ GeV, $p_T > 250$ GeV,

- S_T - scalar sum of the p_T - of outgoing particles for Z plus jets is $S_T < 250$ GeV, $250 < S_T < 500$ GeV, $S_T > 500$ GeV,

- W or Z p_T for dibosons: $p_T < 250$ GeV, $250 < p_T < 500$ GeV, $p_T > 500$ GeV.

The BSM samples of this dataset contain vector-like T-quarks with masses of 1.0, 1.2 and 1.4 TeV. They were pair produced, result of a SM gluon or a BSM 3 TeV heavy gluon. Samples of a FCNC vertex that results on tZ production were also considered.

The samples were simulated using MadGraph5 2.6.5 and the detector effects were simulated using DELPHES 3 with the default CMS card. The SM background is supplied with Pythia 8.2 and Herwig 7 hadronizations, to compare the background simulation. BSM signals were supplied with Pythia.

Signal	AUC	Feature
HQ$_{1000 \text{ TeV}}$ w/ HG	0.981	H_T
HQ$_{1200 \text{ TeV}}$ w/ HG	0.987	H_T
HQ$_{1400 \text{ TeV}}$ w/ HG	0.990	H_T
HQ$_{1000 \text{ TeV}}$ wo/ HG	0.974	H_T
HQ$_{1200 \text{ TeV}}$ wo/ HG	0.986	H_T
HQ$_{1400 \text{ TeV}}$ wo/ HG	0.991	H_T
FCNC	0.797	Leading lepton p_T

Table 8: Best features AUC and the respective feature for each signal

The dataset produced has a total of 47 features, amongst them are the features represented in figure 14 and figure 15 - for a better look, the SM distributions are in figure 14 and the BSM distributions alongside with the SM are in figure 15. Additional plots are in annex I.

4.2   Features Analysis

For each feature, it was computed the ROC and AUC of the ROC to evaluate the discrimination between the signal and background. Computing the AUC of the ROC for each feature and considering the different signals, in table 8 the values of the best AUC of the ROC and the feature to which the value belongs to for each signal is displayed.

The features that have the best AUC of the ROC are the H_T and the leading lepton p_T. With the values on the table 8, one can see that the FCNC has the worst AUC so, it is expected that this signal will be the most challenging one to discriminate. Furthermore, the higher the mass of the heavy quark, the better the AUC - there can be a relation between the mass and the AUC, although we can not reach a conclusion about this with this study.

4.3   Boosted Decision Tree

Before using any machine learning method, the dataset was divided into three sets.

Using the training set, it was created a decision tree to fit the values and it was boosted using XGBoost - a gradient boosting library.

While training, a cross-validation using the validation set was done.

The output of the BDT for the different signals of the test set are in figure 16. The ideal result would be the signal close to 1 and the background close to 0 since 1 corresponds to signal and 0 to background. Since it is hard to tell the difference between the BDT output for the different signals with the plots, one must look at the values of the ROC displayed in table 9 to evaluate the models.

The ROC curves related to the BDT model predictions are in table 9. Analyzing this table, one can see that the values are very close to each other with the exception of the signal FCNC - which has the
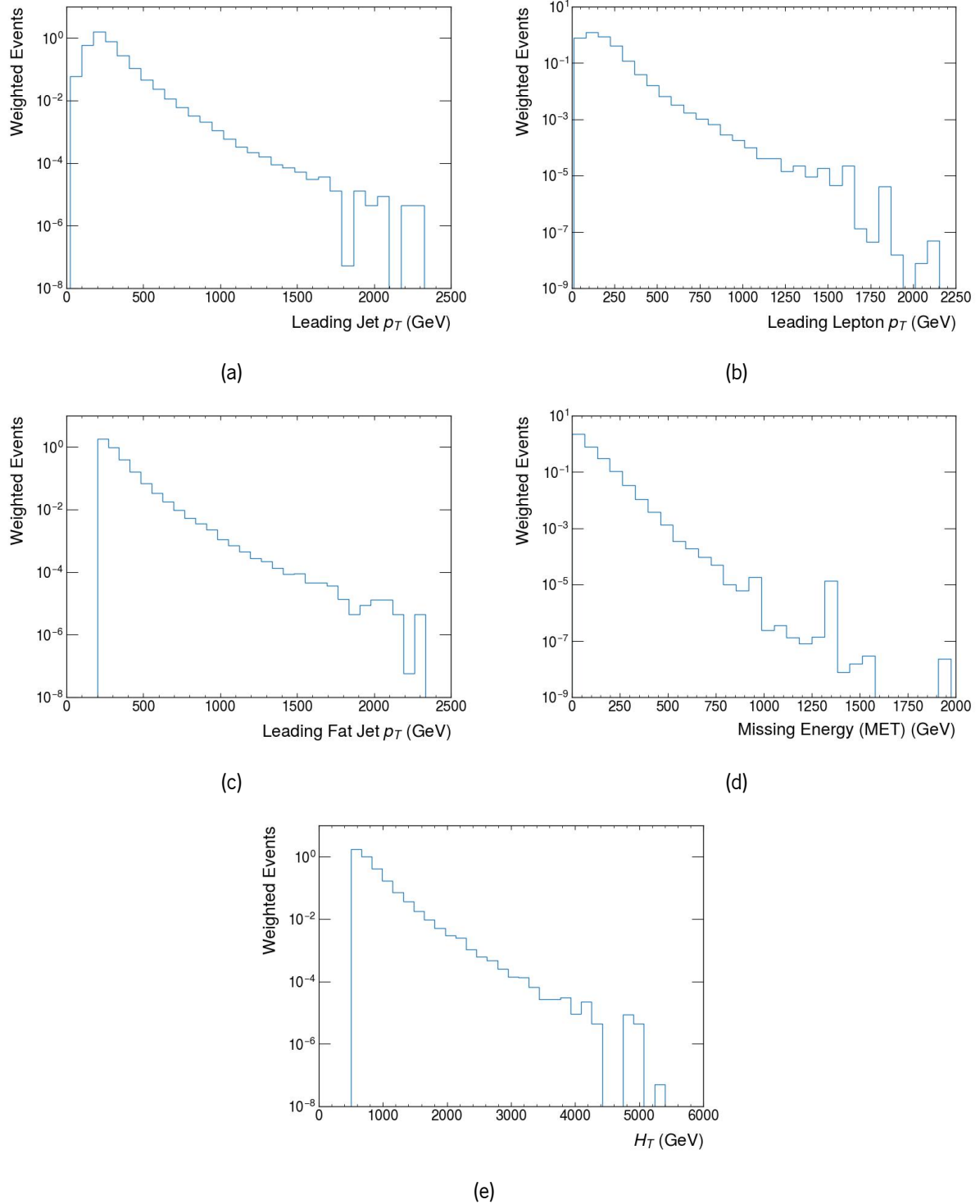
(a)

(b)

(c)

(d)

(e)

Figure 14: Feature plots of the SM data contained in the dataset of simulated pp collisions at $\sqrt{(s)} = 13$ TeV.

(a)

(b)

(c)

(d)

(e)

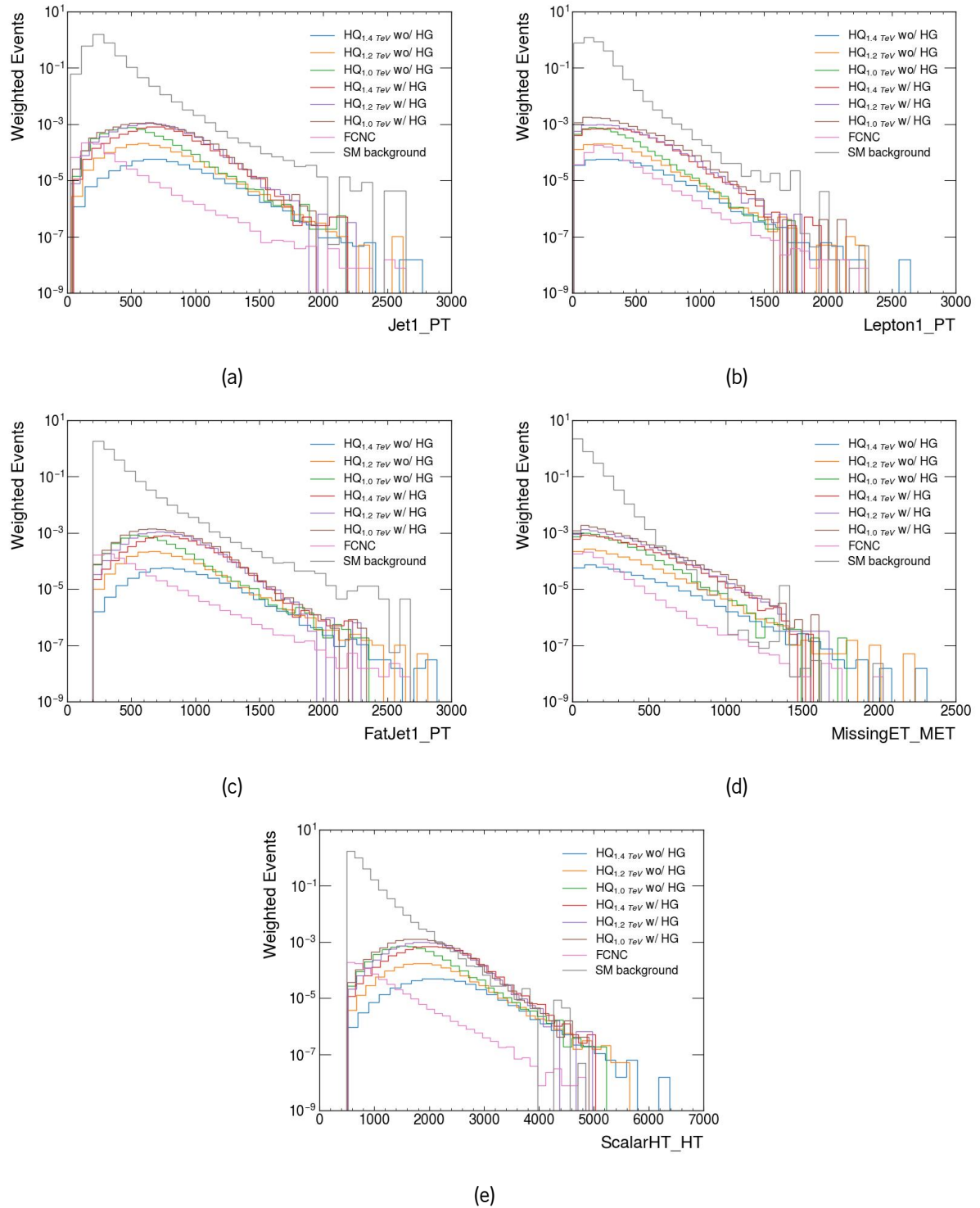Figure 15: Feature plots of the SM and BSM data contained in the dataset of simulated pp collisions at $\sqrt{(s)} = 13$ TeV.
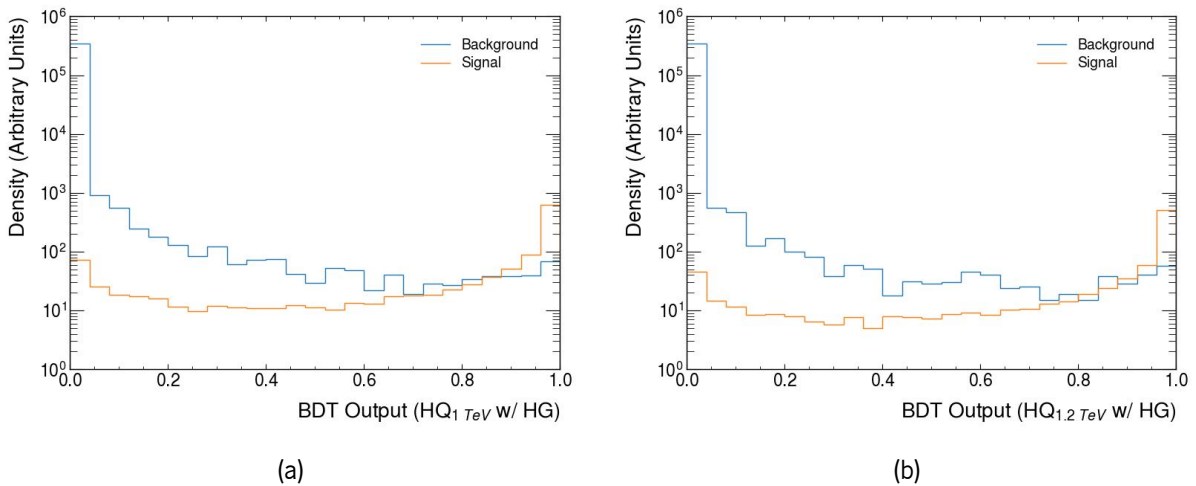
(a)   (b)

Figure 16: Boosted Decision Tree output for the different signals

Signal	ROC AUC
HQ$_{1000\ TeV}$ w/ HG	0.997
HQ$_{1200\ TeV}$ w/ HG	0.998
HQ$_{1400\ TeV}$ w/ HG	0.998
HQ$_{1000\ TeV}$ wo/ HG	0.995
HQ$_{1200\ TeV}$ wo/ HG	0.997
HQ$_{1400\ TeV}$ wo/ HG	0.998
FCNC	0.948

Table 9: The area under the receiver operating characteristic curve for the different signals, using the boosted decision tree model

lower value for the AUC of the ROC. The relation stated before can also be applied here, where the higher the mass of the heavy quark, the better the ROC. So, in sum, it can be more challenging to discriminate signal and background for lower masses of the heavy quark and for the case of the FCNC.

4.4   Autoencoder

After the BDT, it was done a study using an autoencoder, with optimized hyperparameters for each latent space dimension. The goal was to look at the best performance of the model for each latent space dimension, compare them directly and evaluate their outputs. This way, the effect that the latent space has on the signal detection can be studied.

4.4.1   Optuna optimization
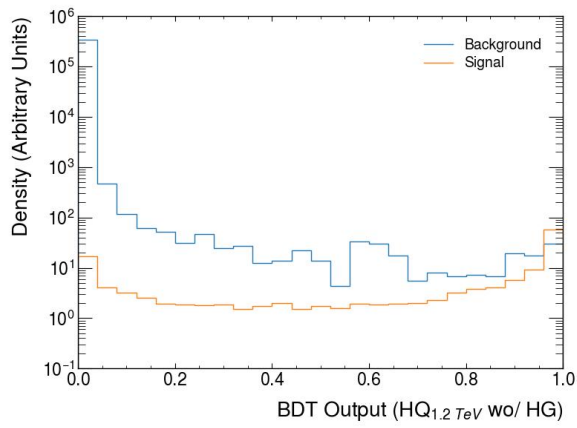
With optuna, the hyperparameters - number of units, number of layers, activation function, normalizations - of the autoencoder were optimized.
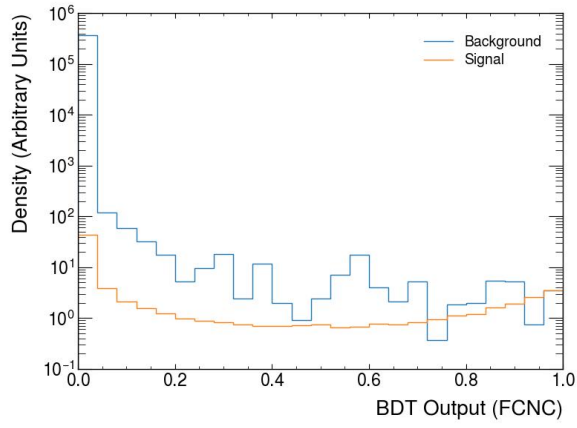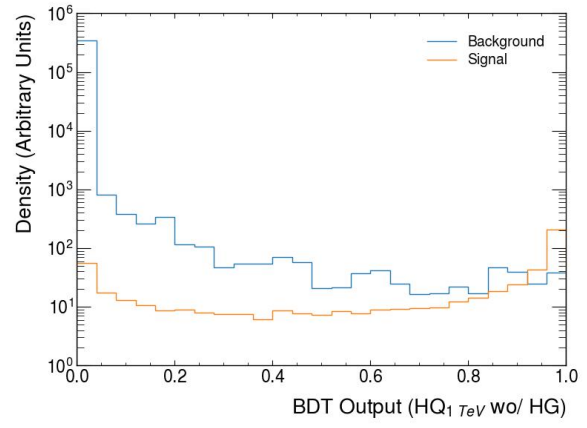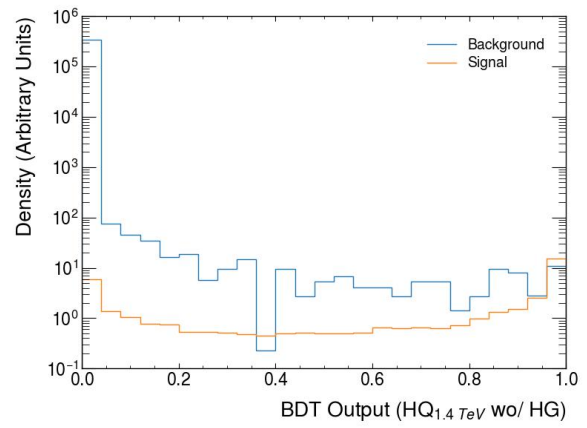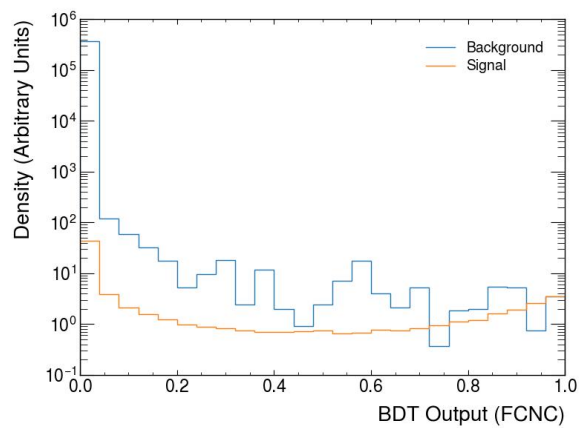
Figure 16: Boosted Decision Tree output for the different signals

In the objective function it was defined the values that the hyperparameters could take whilst the study was being performed:

- activation function: ReLU or LeakyReLU,

- normalization: batch normalization, layer normalization or none,

- number of layers: 1 to 5

- number of units: 8 to 256 with steps of 8

Setting the correct amount of trials needed to do the study is very important. When a study finishes, a plot of the optimization history of all trials can be done in order to appraise the performance of the study and if it converged. The optimization history plot for four distinct latent space dimensions are in figure 17. Looking at these plots, one can infer that the higher the latent space dimension, the quicker the study converges.

After finding the optimal hyperparameters (table 10), the models corresponding to them were used to calculate some metrics on the test set.

The statistical measure R^2 of the background is in figure 18. In this figure, the curve for the R^2 is increasing as the latent space dimension increases. Since this quantity measures how well the model replicates the inputs, the increasing values of the R^2 are the ones expected since the least compression that the data suffers should give rise to better results. On the other hand, if the latent space dimension is small, the compression is high and it becomes harder to decompress and reconstruct the inputs.

The AUC of the ROC for this study from dimensions 1 to 27 is in figure 19a. To see in more detail the variations of the AUC of the signals excluding FCNC, another plot was done (figure 19b).

In figure 19b, the signals follow the same pattern and are all peaking at the same latent dimension - 18 - and FCNC is peaking at latent space dimension - 17 - (see table 11). These latent space dimensions are more sensitive in the exclusion of those signals. This is of huge importance, since it shows that the choice of the latent space dimension should not be taken so lightly. It is important to study if there is a latent space that is more sensitive to the detection of the signal. This choice is critical to the overall performance of the model.

In order to find the relation between the R^2 and the AUC of the ROC, it was done a correlation study. This tells how much these two quantities are linearly related to each other. If they are highly correlated, we can use one of them to predict the value of the other and vice-versa. The correlation between R^2 and AUC is in figure 20. FCNC is the signal that has the higher correlation between the R^2 and the AUC. On the other hand, the signal that has a lower correlation is the heavy quark with 1.4 TeV with heavy gluon.

Gathering the information on table table 12. The model that produced the best results was the BDT, the supervised model. The anomaly detection model, AE, despite being worse than the supervised model, it was better than doing nothing - the AUC of the features was lower most of the times with the exception of the signal FCNC.
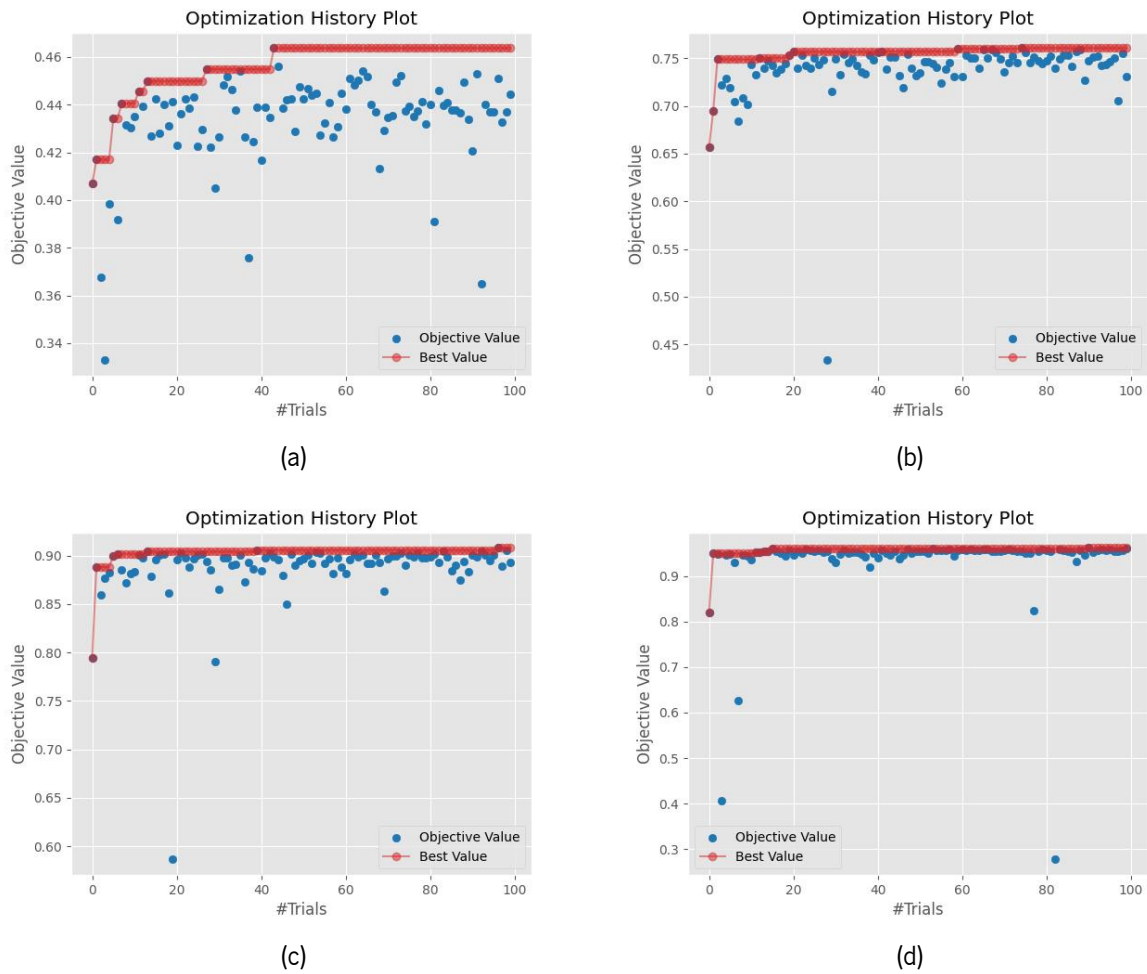
Figure 17: Optimization history plot for latent space dimensions of 3 (figure 17a), 11 (figure 17b), 19 (figure 17c) and 27 (figure 17d).
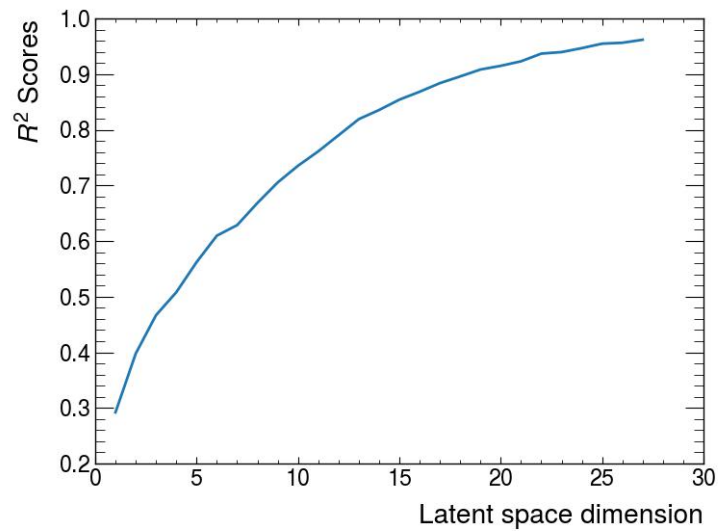

Figure 18: R^2 for the background data

| Latent dimension	Number of units	Number of layers	Activation	Normalisation
1	56	4	ReLU	Batch
2	240	5	LeakyReLU	Batch
3	240	5	LeakyReLU	Batch
4	216	5	ReLU	None
5	248	3	ReLU	None
6	224	5	ReLU	None
7	176	5	LeakyReLU	Layer
8	224	3	ReLU	Batch
9	152	4	LeakyReLU	Layer
10	144	3	ReLU	Layer
11	216	5	LeakyReLU	None
12	224	2	ReLU	Layer
13	96	2	LeakyReLU	Batch
14	216	3	ReLU	Batch
15	208	3	ReLU	Layer
16	160	3	ReLU	None
17	240	5	ReLU	Layer
18	184	5	ReLU	Layer
19	216	5	LeakyReLU	Layer
20	240	5	ReLU	Batch
21	176	4	ReLU	Layer
22	256	3	ReLU	Layer
23	232	4	ReLU	Layer
24	248	3	ReLU	Layer
25	216	3	ReLU	Layer
26	224	5	ReLU	Layer
27	248	5	ReLU	Layer

Table 10: Best hyperparameters found with the Optuna optimization for each latent space dimension

Signal	Best AUC	Dimension
$HQ_{1000\ TeV}$ w/ HG	0.982	18
$HQ_{1200\ TeV}$ w/ HG	0.987	18
$HQ_{1400\ TeV}$ w/ HG	0.990	18
$HQ_{1000\ TeV}$ wo/ HG	0.978	18
$HQ_{1200\ TeV}$ wo/ HG	0.987	18
$HQ_{1400\ TeV}$ wo/ HG	0.992	18
FCNC	0.697	17

Table 11: Best value of the Receiver Operating Characteristic for the different signals and the corresponding latent space dimension
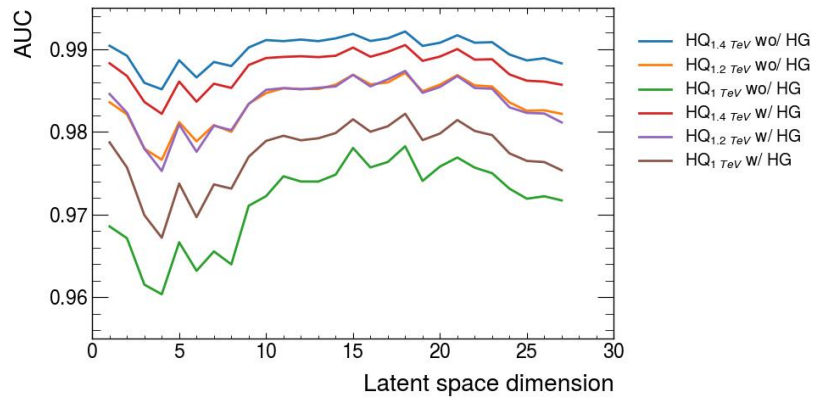
(a) AUC for all signals
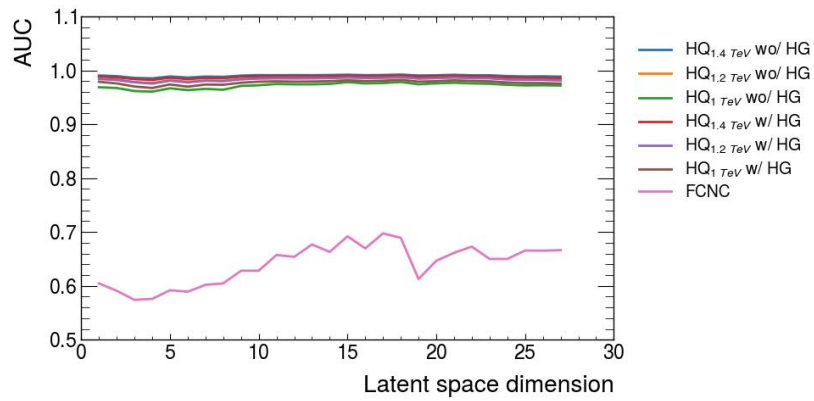
(b) AUC excluding FCNC

Figure 19: Area under the Receiver Operating Characteristic for the different signals from dimensions 1 to 27.
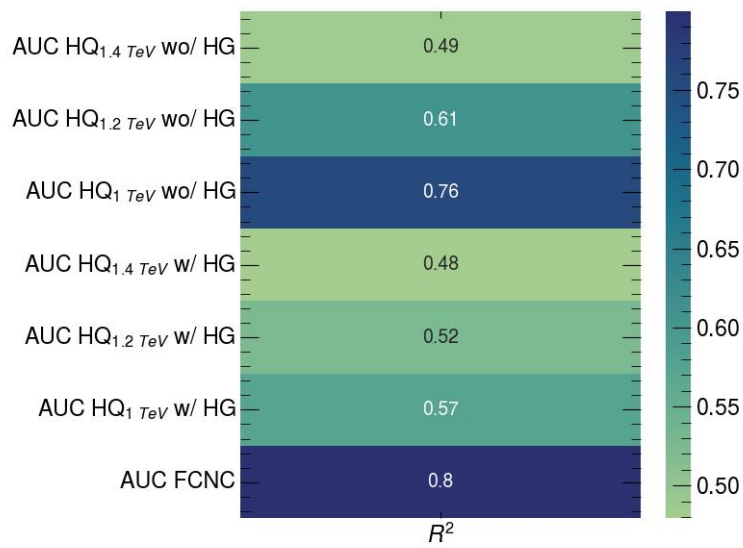

Figure 20: Correlation between R^2 and the area under the receiver operating characteristic

Signal	AUC Feature	AUC BDT	AUC AE
$HQ_{1000\ TeV}$ w/ HG	0.981	0.997	0.982
$HQ_{1200\ TeV}$ w/ HG	0.987	0.998	0.987
$HQ_{1400\ TeV}$ w/ HG	0.990	0.998	0.990
$HQ_{1000\ TeV}$ wo/ HG	0.974	0.995	0.978
$HQ_{1200\ TeV}$ wo/ HG	0.986	0.997	0.987
$HQ_{1400\ TeV}$ wo/ HG	0.991	0.998	0.992
FCNC	0.797	0.948	0.697

Table 12: Area under the Receiver Operating Characteristic for the Features, Boosted Decision Tree model and Autoencoder model

5

Conclusions

In this work, a supervised model, BDT, and an anomaly detection model, AE, were compared when applied to high-energy physics, particularly with BSM physics.

With the supervised model, BDT, the results achieved were good and the AUC was really close between the signals with the exception of FCNC - that was a harder signal to discriminate.

This difficulty was also shown with the AE, where FCNC also had the worst result - worse than the AUC of the features alone. With the other signals, the results were close to the results of the BDT, being shortly lower.

One interesting result of this work was the latent space dimension - the expected would be that with higher latent space, better the results. However, this study was done from latent spaces of dimensions from 1 to 27 and the dimension that gave the best results was mostly the 18th. This means that for future studies, the dimension should not be chosen so freely - it needs to be in consideration that it could be dimensions that are more sensitive to certain kinds of signals.

Furthermore, studies of the correlations between the R^2 and the AUC can be done to understand how they vary and therefore, it can be used in order to predict the values of one another and achieve good results. The signal which had the AUC of the ROC more correlated with the R^2 of the background was the FCNC, followed by the heavy quark with 1 TeV without heavy gluon. On the other hand, the one that was less correlated was the heavy quark with 1.4 TeV with heavy gluon followed by the heavy quark with 1.4 TeV without heavy gluon.

For future work, this study can be extended for more latent space dimensions up to the dataset dimension. There was a time and computational limitation in this work so this was not possible. Other machine learning methods such as variational autoencoders or deep SVDD should be performed and evaluated. It is also important to consider systematic uncertainties in future work, as well as testing the models on real data obtained from collider experiments.

Bibliography

[1] M. Tanabashi et al. "Review of particle physics". In: *Physical Review D* 98.3 (2018), p. 030001.

[2] T. W. B. Kibble. "The Standard Model of Particle Physics". In: Dec. 2014. arXiv: `1412.4094` `[physics.hist-ph]`.

[3] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear physics* 22.4 (1961), pp. 579–588.

[4] A. Salam. "Weak and electromagnetic interactions". In: *Selected Papers Of Abdus Salam: (With Commentary)*. World Scientific, 1994, pp. 244–254.

[5] S. Weinberg. "A model of leptons". In: *Physical review letters* 19.21 (1967), p. 1264.

[6] C.-S. Wu et al. "Experimental test of parity conservation in beta decay". In: *Physical review* 105.4 (1957), p. 1413.

[7] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s= 540 GeV". In: *Physics letters B* 122.1 (1983), pp. 103–116.

[8] M. Banner et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5-6 (1983), pp. 476–485.

[9] CERN. *The origins of the Brout-Englert-Higgs mechanism.* `https://home.cern/science/physics/origins-brout-englert-higgs-mechanism`. Online; accessed in 18 March 2022.

[10] F. Englert and R. Brout. "Broken symmetry and the mass of gauge vector mesons". In: *Physical review letters* 13.9 (1964), p. 321.

[11] P. W. Higgs. "Broken symmetries and the masses of gauge bosons". In: *Physical Review Letters* 13.16 (1964), p. 508.

[12] J. Lykken and M. Spiropulu. *The future of the higgs boson.* Dec. 2013. url: `https://physicstoday.scitation.org/doi/10.1063/PT.3.2212`.

[13] J. A. Aguilar-Saavedra et al. "Handbook of vectorlike quarks: Mixing and single production". In: *Phys. Rev. D* 88 (9 Nov. 2013), p. 094010. doi: `10.1103/PhysRevD.88.094010`. url: `https://link.aps.org/doi/10.1103/PhysRevD.88.094010`.

[14] V. M. Abazov et al. "Search for flavor changing neutral currents via quark–gluon couplings in single top quark production using 2.3 fb- 1 of $p\bar{p}$ collisions". In: *Physics Letters B* 693.2 (2010), pp. 81–87.

[15] T. Aaltonen et al. "Search for the Flavor-Changing Neutral-Current Decay $t \rightarrow Zq$ in $p\bar{p}$ Collisions at $\sqrt{s}$= 1.96 TeV". In: *Physical review letters* 101.19 (2008), p. 192002.

[16] ATLAS. "Search for flavour-changing neutral current top-quark decays to qZ in pp collision data collected with the ATLAS detector at $\sqrt{s}$= 8 TeV". In: 76.1 (2016), pp. 1–24.

[17] CMS. "Search for Flavor-Changing Neutral Currents in Top-Quark Decays $t \rightarrow Zq$ in pp Collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev. Lett.* 112 (17 May 2014), p. 171802. doi: 10 . 11 03 / PhysRevLett . 112 . 171802. url: https : / / link . aps . org / doi / 10 . 1103 /PhysRevLett.112.171802.

[18] CMS. "Search for Flavour Changing Neutral Currents in single top events". In: (2013).

[19] CMS. "Search for associated production of a Z boson with a single top quark and for tZ flavour-changing interactions in pp collisions at $\sqrt{s} = 8$ TeV". In: *Journal of High Energy Physics* 2017.7 (July 2017). doi: 10 . 1007 / jhep07(2017)003. url: https : / / doi . org / 10 . 1007%2 Fjhep07%282017%29003.

[20] ATLAS. "Search for single top-quark production via flavour-changing neutral currents at 8 TeV with the ATLAS detector". In: *The European Physical Journal C* 76.2 (2016), pp. 1–30.

[21] V. Khachatryan et al. "Search for anomalous single top quark production in association with a photon in pp collisions at $\sqrt{s} = 8$ TeV". In: *Journal of High Energy Physics* 2016.4 (2016), pp. 1–38.

[22] *CERN accelerating science*. url: https://home.cern/science/accelerators/accelerator-complex.

[23] ATLAS. "The ATLAS experiment at the CERN large hadron collider". In: *Journal of instrumentation* 3.S08003 (2008).

[24] A. La Rosa. "The ATLAS Insertable B-Layer: from construction to operation". In: *JINST* 11.12 (2016). Ed. by C. Gemme and L. Rossi, p. C12036. doi: 10 . 1088 / 1748 − 0221 / 11 / 12 /C12036. arXiv: 1610.01994 [physics.ins-det].

[25] H. Pernegger. "The Pixel Detector of the ATLAS experiment for LHC Run-2". In: *Journal of Instrumentation* 10.06 (2015), p. C06012.

[26] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. doi: 10.1088/1748-0221/3/08/S08004.

[27] T. Sakuma and T. McCauley. "Detector and Event Visualization with SketchUp at the CMS Experiment". In: *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 022032. doi: 10.1088/1742-6596/513/2/022032. arXiv: 1311.4942 [physics.ins-det].

[28]   C. Biino. "The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections". In: *Journal of Physics: Conference Series*. Vol. 587. 1. IOP Publishing. 2015, p. 012001.

[29]   *Machine learning decision tree classification algorithm - javatpoint*. url: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.

[30]   V. K. Vemuri. *The Hundred-Page Machine Learning Book: by Andriy Burkov, Quebec City, Canada, 2019, 160 pp., 49.99(Hardcover); 29.00 (paperback); 25.43(KindleEdition),(Alternatively,canpurchaseatleanpub.comataminimumpriceof 20.00), ISBN 978-1999579517*. 2020.

[31]   IBM. *What are neural networks?* url: https://www.ibm.com/cloud/learn/neural-networks.

[32]   G. User. *Imperial College Machine Learning - Neural Networks*. May 2022. url: https://www.doc.ic.ac.uk/~nuric/teaching/imperial-college-machine-learning-neural-networks.html.

[33]   *Applied deep learning - part 3: Autoencoders*. 2022. url: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798.

[34]   T. Akiba et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. 2019. doi: 10.48550/ARXIV.1907.10902. url: https://arxiv.org/abs/1907.10902.

[35]   Crispim Romão, Miguel and Castro, Nuno F. and Pedro, Rute. *Simulated pp collisions at 13 TeV with 2 leptons + 1 b jet final state and selected benchmark Beyond the Standard Model signals [Data set]*. doi: https://doi.org/10.5281/zenodo.5126747.

Annex 1

Additional feature plots of the SM data contained in the dataset, such as the electron multiplicity, the muon multiplicity, the leading jet mass and the leading fat jet mass are in figure 21.
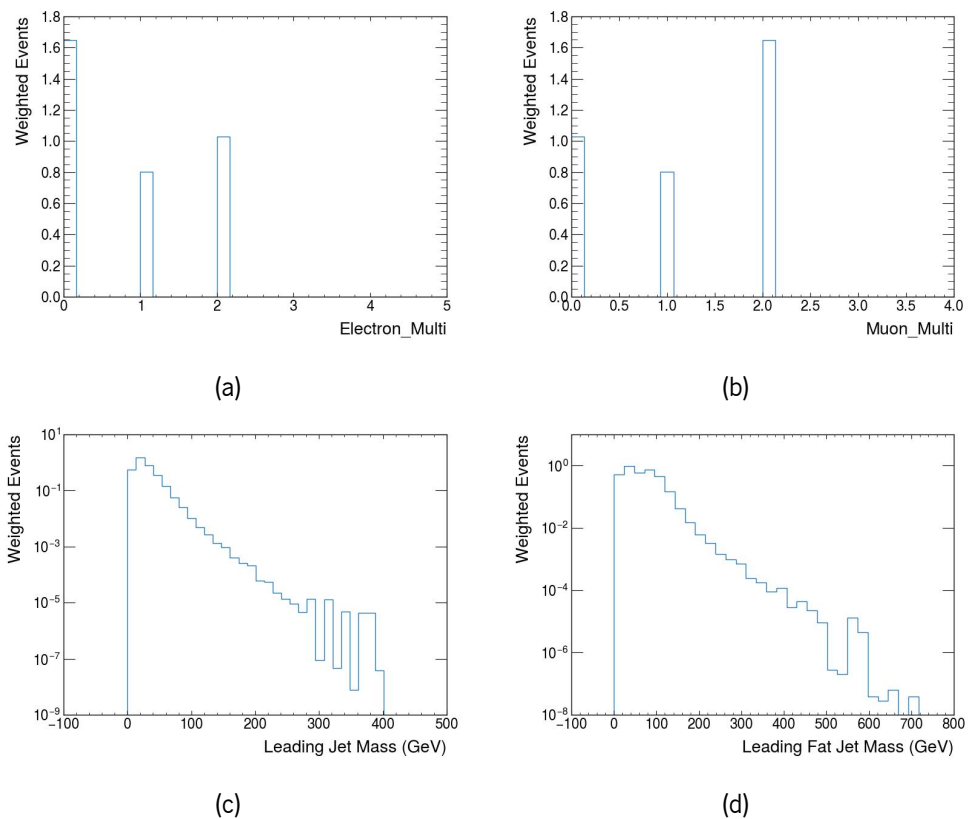

(a)

(b)

(c)

(d)

Figure 21: Feature plots of the SM data contained in the dataset of simulated pp collisions at $\sqrt{(s)} = 13$ TeV.

Considering the signals, the plots for the electron multiplicity, the muon multiplicity, the leading jet

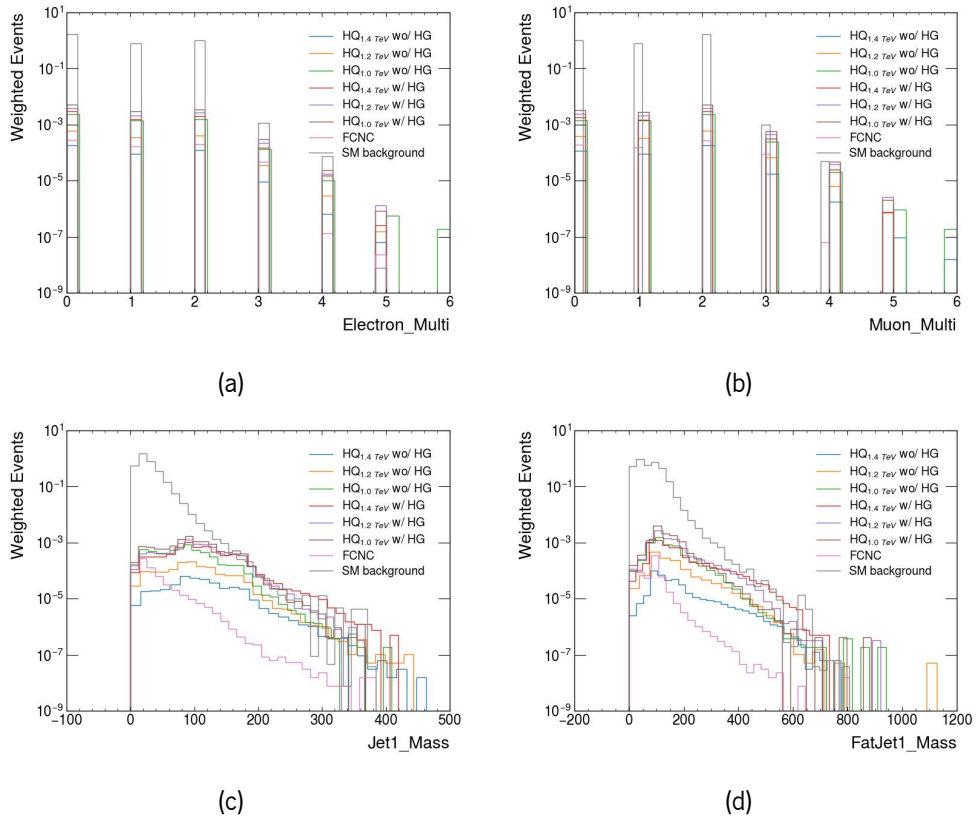mass and the leading fat jet mass are in figure 22.

(a)

(b)

(c)

(d)

Figure 22: Feature plots of the SM and BSM data contained in the dataset of simulated pp collisions at $\sqrt{(s)} = 13$ TeV.