Identificación de procesos tH usando técnicas ML
(Identification of tH processes using ML techniques)

Trabajo de Fin de Máster
para acceder al

MÁSTER EN FÍSICA DE PARTÍCULAS Y DEL COSMOS

Autor: Pablo Herreros Fuentevilla

Director\es: Javier Andrés Brochero Cifuentes

Septiembre - 2022

Resumen

En este proyecto se ha realizado un análisis aplicando técnicas de machine learning, concretamente BDTs, para la discriminación de la señal contra diferentes fondos. El proceso principal que constituye la señal es tHW mientras que como fondo se utilizan los procesos $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$. Los resultados también han sido evaluados para tHq. Para ello, se han utilizado simulaciones de MonteCarlo. Se han presentado las figuras mas importantes relacionadas con los resultados del training, así como los resultados obtenidos. Estos resultados son aceptables basados en las eficiencias de la selección, especialmente para las regiones con más de un leptón.

Palabras clave: quark top, bosón de Higgs, genjets, machine learning, BDT

Abstract

In this project an analysis has been carried out by implementing machine learning techniques, specifically BDTs, for the discrimination of the signal against different backgrounds. The process that constitutes the signal is tHW while the processes $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$ are backgrounds. Results are evaluated also in tHq. For this purpose, data obtained from MonteCarlo simulations have been used. The most important figures related to the training results have been presented, as well as the results obtained, which are acceptable according to the calculated selection efficiencies, especially for regions with more than one lepton.

Keywords: top quark, Higgs boson, genjets, machine learning, BDT

Contents

1 Introduction **5**

2 Standard Model **6**

3 The LHC experiment **8**
3.1 Introduction . . . 8
3.2 CMS . . . 9
3.2.1 Coordinate system . . . 9
3.2.2 Tracker . . . 10
3.2.3 Calorimeters . . . 10
3.2.4 Muons detectors . . . 10

4 Higgs physics **12**
4.1 Higgs mechanism . . . 12
4.2 Higgs production . . . 13
4.2.1 Associated production with top quarks . . . 14
4.3 Higgs decays . . . 14

5 Top quark physics **16**
5.1 Top quark production . . . 16
5.1.1 $t\bar{t}$ production . . . 16
5.1.2 Single top production . . . 17

6 Final states for tHW events **18**
6.1 Background processes . . . 19

7 Analysis **23**
7.1 Object selection . . . 23
7.2 Event selection . . . 23
7.3 Control Plots . . . 24
7.4 Quark-Genjet matching efficiencies . . . 26
7.5 BDT analysis . . . 29
7.5.1 Regions and variables . . . 29
7.5.2 tHW training . . . 29
7.5.2.1 tHW training versus $t\bar{t}$ in 1 lepton channel . . . 30
7.5.2.2 tHW training versus $t\bar{t}W$ in 1 lepton channel . . . 32
7.5.2.3 tHW training versus $t\bar{t}W$ in 2 leptons channel . . . 33
7.5.2.4 tHW training versus $t\bar{t}W$ in 3 leptons channel . . . 35
7.5.2.5 tHW training versus $t\bar{t}Z$ in 1 leptons channel . . . 37
7.5.2.6 tHW training versus $t\bar{t}Z$ in 2 leptons channel . . . 38
7.5.2.7 tHW training versus $t\bar{t}Z$ in 3 leptons channel . . . 40
7.5.2.8 tHW training versus $t\bar{t}Z$ in 4 leptons channel . . . 41
7.5.3 Training tHW vs $t\bar{t}V$ . . . 43
7.5.3.1 1 lepton region . . . 43
7.5.3.2 2 leptons region . . . 44

7.5.3.3 3/4 leptons region 46
7.5.4 BDT response 47
7.5.5 Results 49

8 Conclusions **51**

9 Bibliography **53**

1 Introduction

The Compact Muon Solenoid (CMS) is one of the detectors of the Large Hadron Collider which collected a large amount of experimental data in 2015-2018 in what is called Run 2. Data were produced by proton-proton collisions at the center of mass energy of 13 TeV. Recently has started the Run 3 which has increased the energy to 13.6 TeV and the integrated luminosity to 280 pb^{-1} [1]. In this analysis, it is used MonteCarlo simulations instead of experimental data.

The project is focused on the single top associated with a Higgs boson and it is used machine learning techniques. The study of both particles is very important in High Energy Physics (HEP). The top quark was discovered in 1905 and has some interesting properties which can consolidate our knowledge about the Standard Model and also can guide us to the discovery of physics beyond the standard model or new physics. Some of these properties are its large mass or its short lifetime (it decays before hadronizing). Then, Higgs boson was discovered in 2012 and still has a lot of properties that are unknown which also could lead to the discovery of new physics. The single top associated with a Higgs boson process can be produced in association with a W boson (tHW) or with an additional light jet (tHq). In this case, both are used but mainly tHW. This process is compare with some of its possibles background which are $t\bar{t}V$ ($t\bar{t}W$ and $t\bar{t}Z$) and $t\bar{t}$.

In this project is implemented a signal discrimination analysis using machine learning techniques which is the Boosted Decision Trees in order to separate the signal from the main backgrounds. The signal of the analysis is the tHW and it is trained versus $t\bar{t}V$ process due to their final states are very similar, therefore it is hard to separate them. Then, the training is applied to $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ which are the main backgrounds. Results are also evaluated in tHq. From the BDT responses of each process and region, the results are obtained and presented as well as, the important figures related with the training.

The document is divided into three main parts: the theoretical framework, the LHC experiment and the core of the analysis. In the theoretical framework, it is explained the properties of the top quark and the Higgs boson, as well as, the tHW process and its possible backgrounds. In the LHC experiment, it is briefly explained how the particles are detected and the different parts of the detector. Finally, the analysis and its results.

2 Standard Model

The Standard model of particle physics is a well-tested theory that describes the elementary particles that make up the matter of the universe as well as their interactions following the four fundamental forces. These elementary particles are classified in two different groups based on their spin. When the spin is an integer, the particles are called bosons. If its spin is 0, it is a scalar boson (Higgs) meanwhile the bosons of spin 1 are called gauge bosons and they are the carriers of three of the four fundamental forces. The elementary particles which spin is a half integer are called fermions.

Among fermions are distinguished two sets of particles, quarks and leptons. Both groups consist of six particles divided in three generations where generation I corresponds to stable particles with low masses meanwhile the most unstable and heavy particles are found in generation III. As it can be seen in figure 1, quarks (q) have six different "flavors": up (u), down (d), strange (s), charm (c), top (t) and bottom (b). Depending on their electric charge, quarks are u-type or d-type. As the u, c, t quarks have an electric charge of $\frac{2}{3}$e, they are u-type meanwhile the d, s, b quarks are d-type because their electric charge is $-\frac{1}{3}$e. Also leptons (ℓ) have six flavors: electron (e), muon (μ), tau (τ), electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ). Electrons, muons and tau have charge -1e while neutrinos have no electric charge and their mass is very small (non-zero). Each fermion has its corresponding antiparticle, which has the same mass but opposite quantum numbers.

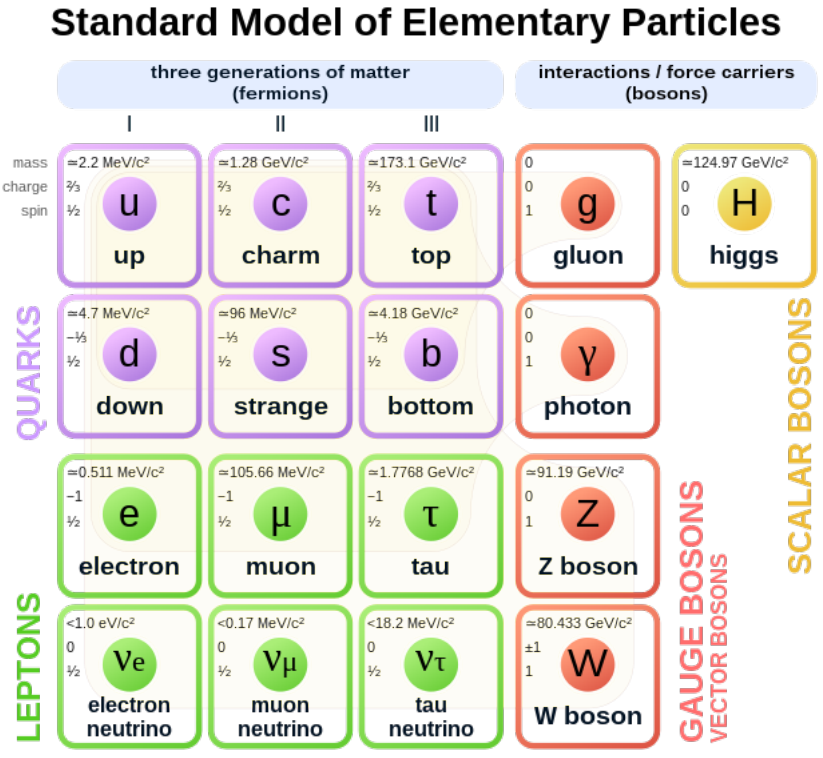


Figure 1: Standard Model of particle physics.[4]

The universe is governed by four fundamental forces: gravitational force, electromagnetic force, weak force and strong force. Each force has different strengths and work over different range. Gravity is not include in the standard model which is an open problem currently in physics. Nevertheless, gravitational force is the weakest and working at this scale, its effect is negligible. The other three interactions are described by a Quantum Field Theory as the exchange of gauge bosons which are the force-carriers particles. Z and W bosons are the carriers of the weak interaction meanwhile the electromagnetic force is carried by photons and the strong force is carried by gluons.

The electromagnetic force is described by Quantum Electrodynamics, acting between charged particles and it is mediated by photons. These bosons have neither electric charges nor mass. Since photons are massless, the working range of this force is infinite.

The weak force is mediated by the exchange of the W and Z bosons and it causes the beta nuclear decay. His working range is finite, around $\sim 10^{-3}$fm [5], because these bosons, unlike photons, have mass. The W boson has electric charge meanwhile Z boson have not electric charge and in its interactions, it does not change the flavor of the particles.

The strong force is responsible for keeping the atomic nucleus stable. It has a limited range (1 fm) [5] and the mediating particles are the gluons, which have no electric charge but they do have color charge. These interactions are described by Quantum Chromodynamics(QCD).

Although it is not part of the Standard Model, gravitational force is one of the four fundamental forces in nature which has an infinite range and it is the weakest force. It is thought than could exist a force-carrier particle for gravity, an hypothetical boson called graviton.

Force	Boson	Mass$[GeV]$	Coupling	Electric charge	Color charge
Strong	gluon (g)	0	1	0	Y
Weak	W$\pm$	80.4	10^{-13}	$\pm$ 1	N
Weak	Z	91.2	10^{-13}	0	N
Electromagnetic	photon (γ)	0	10^{-3}	0	N

Table 1: Properties of the fundamental forces in nature (except gravity) [3].

Quarks can not be found isolated because color confinement establish that they can only be observed in colorless states. Therefore, quarks group forming baryons (three quarks) or mesons (quark-antiquark pair) through strong force. In experiments of high energy physics, like colliders, if quarks and gluons are attempted to separate, this gives rise to the formation of another series of particles, without color charge, such as hadrons or pions. That process is called hadronization and it leads to the production of jets, which are defined as a collimated spray of particles (hadrons) created from hadronization and therefore with the same origin [3] [6].

3 The LHC experiment

3.1 Introduction

The Large Hadron Collider (LHC) is a particle accelerator located in Geneva. It consists on a ring of 27 km of perimeter with superconducting magnets. In this collider are produced proton-proton(pp) collisions that leads to the processes to be studied. To get these collisions, they are send bunches of protons ($\approx 10^{11}$ protons). This way, it is possible to ensure that always will occur a collision between two protons with the correct angle. That is the reason why it is needed to send more than two protons. The center-of-mass energy used in this accelerator has varied over the years, in this case, the data simulations analyzed correspond to 13 TeV.

Depending on the purpose of the study to be carried out, it will be necessary to use different technologies of detection or detectors. Thus, within the LHC there are 4 different detectors. The best known detectors are ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) which were designed for general purposes but they do not work neither the same way nor using the same technologies. This allows to check the results obtained by each detector. In order to study deeper the strong interaction especially at extreme energy densities (quark-gluon plasma) it is used the detector ALICE (A Large Ion Collider Experiment). Then, for studies related to differences between matter and matter (focused on b quark) it was designed the LHCb (The Large Hadron Collider beauty). Also can be found other detectors dedicated to the study of exotic matter and forward physics as LHCf, TOTEM, MOEDAL and FASER. [7].

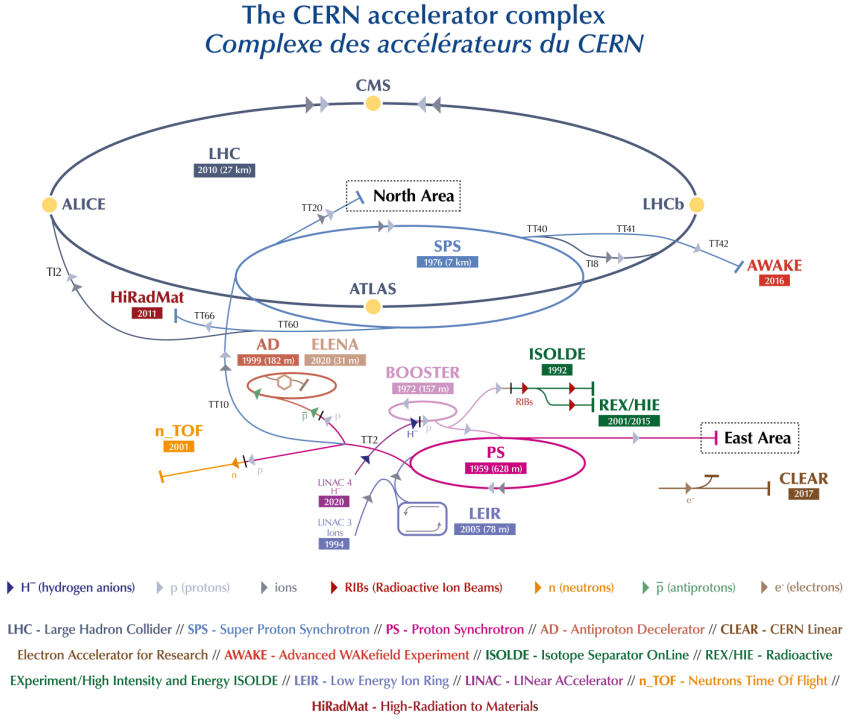


Figure 2: Scheme LHC.[8]

3.2 CMS

This detector is, as well as ATLAS, a general purpose detector. It consists on a large solenoid magnet which it is composed of a cylindrical coil of superconducting fibres which generates a magnetic field of 4 Teslas [9]. At the proton collisions are produced many kinds of particles which have not the same properties. Thus, there are several subdetectors, within the detector, using different detection mechanism and technologies to detect the particles produced in the collisions based on how they interact with matter.

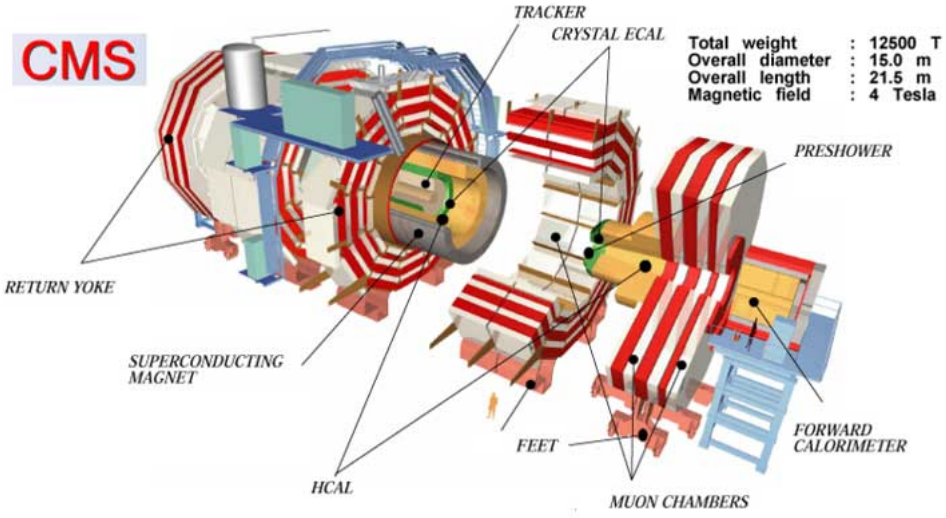


Figure 3: Scheme of CMS.[10]

Figure 3 shows CMS dimensions and technical specifications as well as a schematic representation of the components of CMS including the different subdetectors. These subdetectors are tracker, calorimeters (ECAL and HCAL) and the muon chambers.

3.2.1 Coordinate system

In order to be able to measure the properties of the particles, it is necessary to define a coordinate system. The x-axis points radially inward toward the center of the LHC, the z-axis points along the beam direction (magnetic field direction) and the y-axis points vertically upward. The origin is centered at the theoretical collision spot of the beams.

Using polar coordinates, the polar angle, θ, is measured from the z-axis, while the azimuthal angle, ϕ is measured from the x-axis in the XY plane, and they are defined as $\phi = arctan(y/x)$ and $\theta = arctan(\sqrt{x^2 + y^2}/z)$, respectively. Due to the cylindrical geometry of the detector, the polar angle is replaced by pseudorapidity, η, which is defined as

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

This way, instead of measuring the lineal momentum it is used the transverse momentum, p_T, which is calculated from the transverse energy measured in the calorimeters [11]. The imbalance of the transverse energy is also calculated.

CMS adopted this coordinate system because the high energy particles' multiplicity is roughly constant in η which is very useful in high energy physics. Also, along the z-axis rapidity's intervals are Lorentz-invariant.

3.2.2 Tracker

The nearest subdetector to the collision point is the tracker, therefore it is the first one that particles pass through. In the tracker are reconstructed the trajectories of the particles that goes through it, from the hits (electric signals) that particles leaves in the tracker. Knowing that there are a magnetic field in the detector, charged particles are bend by it. Measuring how much they are bend, it is calculated the transverse momentum (greater curvature lesser momentum and vice versa). Therefore, neutral particles are not detected by tracker.

This subdetector is based on silicon technology and there is divided in two detectors: silicon pixel detector which is located at the core of the detector hence it deals with the highest intensity of particles and silicon strip detector that surround it. The pixels and microstrips produce small electric signals, when particles goes through the detector, that are amplified and detected [12].

3.2.3 Calorimeters

When particles arrive to calorimeters, some of them are absorbed, which means that the energy is deposited by particles and it is measured. There are two differents calorimeters: electromagnetic calorimeters (ECAL) and hadrons calorimeters (HCAL).

ECAL is the inner layer and it is located after the tracker. It measures the energy of light electromagnetic particles as photons and electrons or jets due to electromagnetic showers. The ECAL is made of lead tungsten ($PbWO_4$) crystals which work as absorbers and scintillators and provide precise measurements.[13]

HCAL is the outer layer and it measures the energy of hadrons and its decays products since hadrons are not stopped at the ECAL. It is made of copper layers interleaved with scintillator material.

3.2.4 Muons detectors

The outer subdetector is the muon detector. Its goal is to detect muons which are the particles that goes through all the previous subdetector (do not lose all their energy at calorimeters) besides neutrinos that can not be detected directly. Within muon detector, there are three types of detectors: drift tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). They are based in gas technologies (different gas composition).

Drift tubes are in the barrel and they covers a pseudorapidity's range of $|\eta| < 1.3$. Cathode Strip Chambers are located at endcaps and they work for pseudorapitities $0.9 < |\eta|) < 2.4$. Resistive Plate Chambers are in both barrel and endcaps and they work as a trigger with a maximum aceptance of $|\eta| \approx 1.6$.[14]

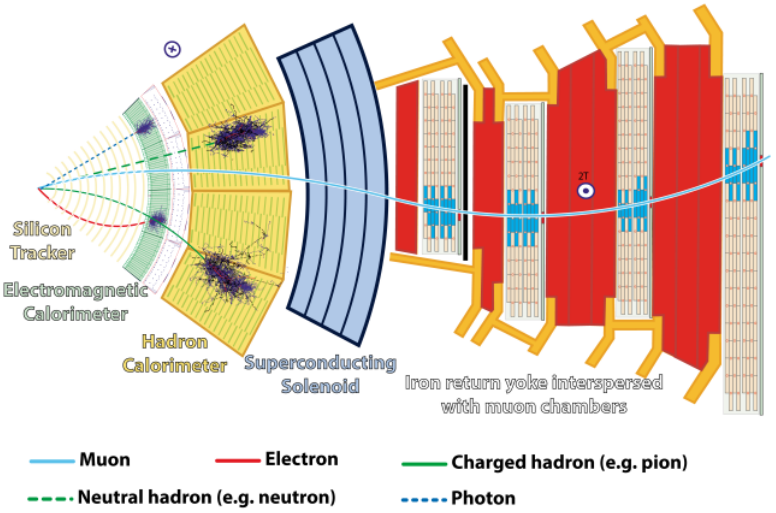


Figure 4: Layout of the CMS detector.[15]

Figure 4 shows how different types of particles behave as they pass through each subdetector. In this illustration, it is displayed the characteristic trace of some particles depending on their properties. The light electromagnetic particles, as electrons, do not pass through the electromagnetic calorimeter meanwhile hadrons stop at the hadron calorimeter. In addition, it is clear that neutral particles do not bend due to the magnetic field. The only detected particles that arrive to the muon chambers are muons. Neutrinos are not detected directly because they do not interact with the detector but its presence can be inferred from the imbalance of the transverse energy. There are more ways to explain this imbalance of energy as bad measurements of the momentum or resolution defects.

4 Higgs physics

Besides gauge bosons, in the Standard Model is found the Higgs boson which is a scalar boson (spin 0) and an essential part of the SM. It was theorized in 1960s and discovered in 2012 at the LHC (CMS and ATLAS). Experimentally has been determined that its mass is 125.35 ± 0.15 GeV [16]. This boson is associated to the Higgs field which is responsible of the mass of bosons and fermions (except neutrinos).

4.1 Higgs mechanism

In Quantum Field Theory (QFT) each particle is associated with a field being these particles a local excitation of the field. In this case, Higgs boson is a local excitation of the Higgs field. The addition of the Higgs field to the Standard Model provides an explanation for the mass of elementary particles. It especially explains how the Z and W bosons have mass because gauge invariance leads to the conclusion that bosons must be massless. The explanation of how these bosons have mass, which was known because the weak interaction has a short range unlike the electromagnetism, is called the Higgs mechanism. It is based on local gauge symmetry breaking and this can be achieved in several ways but what was theorized was the spontaneous symmetry breaking. Therefore, it is conserved the gauge invariance and renormalizability [17].

Electromagnetism and weak interaction was unified in the standard model for energy above 100 GeV and it is called electroweak. It belongs to $SU_L(2)$ X $U(1)_Y$ gauge group. The Higgs field is a weak isospin doublet with four components:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1}$$

Where ϕ^+ is the charged field and ϕ^0 the neutral. Its components are ϕ_1, ϕ_2, ϕ_3 and ϕ_4 [17]. The rotational symmetry of the Higgs field is spontaneously broken by a fluctuation around the minimum vacuum expectation value, ν [18]. This way, the vacuum expectation value, ν, is non-zero and no longer unique, unlike the other quantum fields. In figure 5, is shown the Higgs potential which has the shape of a Mexican hat. Therefore, the interaction between the particles with the non-zero Higgs field gives them mass.

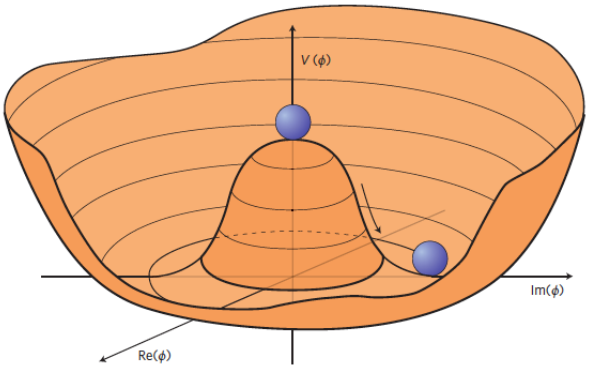


Figure 5: The Higgs potential [19].

The mass of fermions arises from the interaction between Dirac fields and Higgs field in which the Yukawa coupling term is added to the Lagrangian density. Neutrino mass is not explained by this mechanim because they are not right-handed.

4.2 Higgs production

At high energy hadron colliders, like LHC, there are four main Standard Model production mechanisms to produce the Higgs boson: through the fusion of two gluons (ggF), through the fusion of weak vector bosons (VBF), in association with a W or Z boson (VH), or one or more top quarks ($t\bar{t}H$+tH) [21].

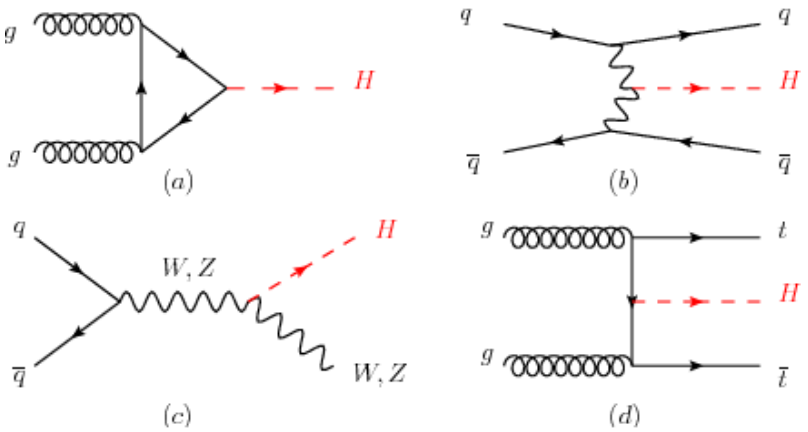


Figure 6: Feynman diagrams of Higgs production mechanism. Process a) is the gluons fusion, b) is the fusion of weak vector bosons, c) is the production of Higgs boson in association with vector bosons and d) in association with top quarks ($t\bar{t}H$).[22]

In figure 6 are shown the feynman diagrams corresponding to the main production mechanisms of the Higgs boson. In order to determine which are the dominant production process of the Higgs boson, it is been presented their cross section in the following table.

Production process	Cross section [pb]
ggF	19.2 $\pm$2.0
VBF	1.57 $\pm$0.04
WH	0.698 $\pm$0.018
ZH	0.412 $\pm$0.013
$b\bar{b}H$	0.202 $\pm$0.028
$t\bar{t}H$	0.128 $\pm$0.014
tH	0.018 $\pm$0.001
Total	22.3 $\pm$2.0

Table 2: Cross section of the Higgs bosons production processes at LHC (8 TeV) [23]

According to this data, the dominant production process is the gluons fusion by far, follow by vector bosons fusion and in association with vector bosons (WH and ZH). The less dominant process is the Higgs boson production in association with top quarks.

4.2.1 Associated production with top quarks

As it is been previously seen, one of the possible production mechanisms of Higgs bosons is the associated production with top quarks. The study of these processes are key to study the Yukawa coupling between top quarks and Higgs boson. As the top quarks is the heaviest particle in SM, the coupling between both particle is strong. The $t\bar{t}$ process allows a direct mesurement of the Yukawa coupling between both particles but as it is shown in table 2, this process has a very small cross section.

On the other side, the associated production of the Higgs boson with top quarks can be with a single top quark. In this case, there are two possible scenarios, one of these consists on the top quark and Higgs boson coupling with a W boson and in the other coupling with a quark. The feynman diagram of these processes are shown in the following figure.

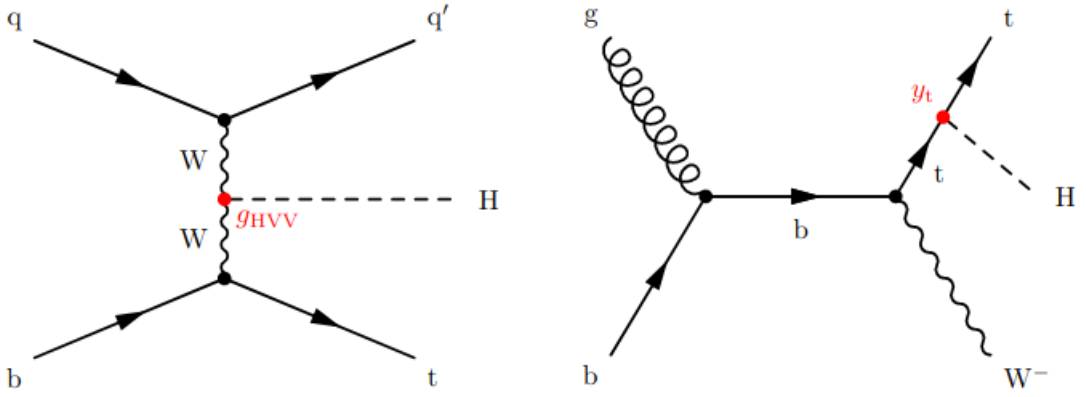


Figure 7: Feynman diagrams of tHq process, on the left and tHW process on the right [24].

4.3 Higgs decays

Higgs bosons can decay into multiple particle including both fermions and boson. It can be classified in three different categories: the decay into two fermion, the decay into two gauge bosons and the decay into four fermions.

1. *decay into two fermion*: Among these processes, it is found the decay into two b quarks, which is the most frequent with a branching ratio around 60%. Also it is usual the decay into two tau.

2. *decay into two gauge bosons*: This category has some considerable decays, as the decay into two gluons or the decay into two Z, but the one with the larger branching ratio is the decay into two W (around 15%).

3. *decay into four fermions*: These processes are not very studied because they have much smaller branching ratios.

To get a better vision of the different decays of the Higgs bosons and their probability, it is been represented the branching ratios of the main possible decays depending on the mass of the Higgs boson.

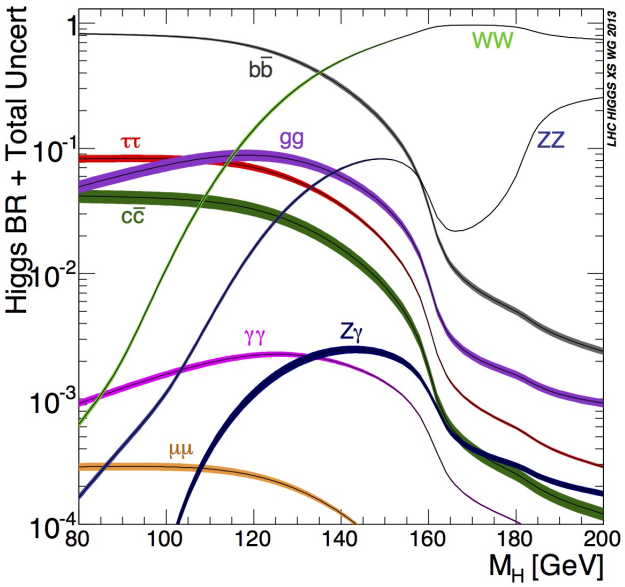


Figure 8: Branching ratios of the decays of the Higgs boson depending on its invariant mass.[21]

Taking that the mass of the Higgs boson is around 125 GeV, from the previous figure it is possible to get an idea of the contribution of each decay. To quantify this, it is been collected in the following table their branching ratio.

Decay channel	Branching ratio [%]
$H \to bb$	57.1 $\pm$1.9
$H \to WW^*$	22.0 $\pm$0.9
$H \to gg$	8.53 $\pm$0.85
$H \to \tau\tau$	6.26 $\pm$0.35
$H \to c\bar{c}$	2.88 $\pm$0.35
$H \to ZZ^*$	2.73 $\pm$0.11
$H \to \gamma\gamma$	0.228 $\pm$0.011
$H \to Z\gamma$	0.157 $\pm$0.014
$H \to \mu\mu$	0.022 $\pm$0.001

Table 3: Cross section of the Higgs boson decays. [23]

In this table is seen how the most frequent decay is the decay into two b quarks followed by the decay into two W boson. Other decays with a considerable branching ratio are the decays into two gluons and into two tau.

5 Top quark physics

The top quark was discovered in 1995 through proton-antiproton collisions at Tevatron. It is a particle whose study is of great interest. It is a u-type quark, so its electric charge is $\frac{2}{3}$. Also, it belongs to the third generation, therefore it is a heavy and unstable particle. Top quarks is the heaviest particle in the standard model and its mass has been experimentally determined obtaining $m_t = 173.34 \pm 0.27(\text{stat.}) \pm 0.71(\text{syst.})$ GeV [25]. Having a mass much greater than that of the W^{+-} boson, the top quark is the only one that decays through weak interactions of first order (W and a down-type quark). The most probable decay (99.8 %) is the b because the factors of the Cabibbo–Kobayashi–Maskawa (CKM) matrix are larger for b quark than for the other down-type quarks (s and d).

One of the reasons for the importance of its study is its mass, since it is the heaviest particle of the standard model. As the mass is so big compare to the rest of particles, this can leads to discover new physics. For example, because of its enormous mass, the coupling with Higgs bosons is large. Another important property of top quarks is that it decays before hadronization occurs. This is due to its short lifetime, since it is shorter than the time require to hadronization, Therefore, jets originated from top quarks can not be found.

5.1 Top quark production

There are three main mechanisms, in SM, to produce top quarks at hadrons colliders. These are the production of $t\bar{t}$ pairs through strong force which it is the dominant process, single top production through electroweak interaction and the associated production of $t\bar{t}$ and a vector boson (W and Z)

5.1.1 $t\bar{t}$ production

In order to produce top-antitop pairs, there are two principal mechanisms: gluon fusion or quark-antiquark collisions. At LHC, the dominant process to produce quarks ($t\bar{t}$) is gluon fusion which accounts for 83%. Therefore, the other way to produce $t\bar{t}$ pairs is through the quark-antiquark annihilation and it represents 17%.

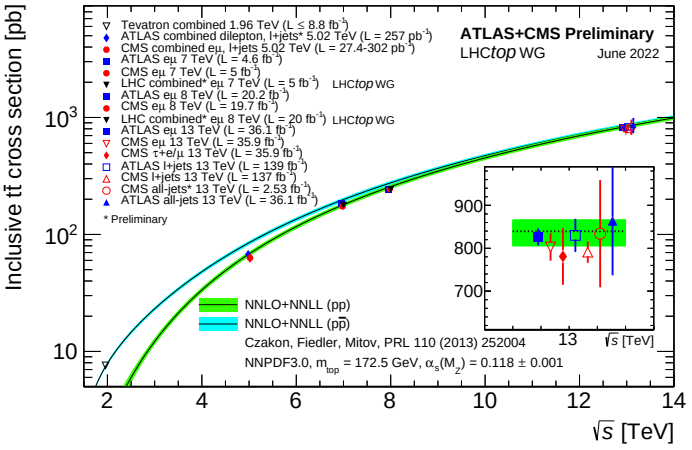


Figure 9: Top quark production's ($t\bar{t}$) cross section.[26]

In figure 54 is shown the cross section of the $t\bar{t}$ production depending on the center of mass energy. The values of center of mass energy used to determine experimental cross section are given by the energy at which the LHC has worked over the years. The continuous lines correspond to the theoretical predictions. Also, there are two different colors that indicates whether proton proton or antiproton proton collisions have been used. Experimental measurements for energies lower than 7 TeV was taken on Tevatron and at the bottom right can be seen how the experimental measurements are in a good agreement with the theoretical predictions.

Energies$[TeV]$	Cross section$[pb]$	Scale$[pb]$	PDF$[pb]$
7	177.31	+4.56 -5.99	+9.02 -9.02
8	252.89	+6.39 -8.64	+11.67 -11.67
13	831.76	+19.77 -39.20	+35.06 -35.06
14	984.50	+23.21 -34.69	+41.31 -41.31

Table 4: $t\bar{t}$ cross sections (for a mass of 172.5 GeV). [27]

On the previous table, are shown cross sections corresponding to the different values of center of mass energy at which the LHC has worked over the years along its respective uncertainties. For the calculations, it is considered that the top quark has a mass of 172.5 GeV.

5.1.2 Single top production

Apart from $t\bar{t}$ production, also top quarks can be obtained from the single top production. This process has three different channels (t-channel, s-channel and associated production with W) [28] and it is mediated by electroweak interaction.

1. *t-channel*: Also known as gluon fusion, is the dominant process to produce single tops. It consists on the interaction of W virtual boson interaction with a b quark which is inside the proton. The b quark is produce from the division of a gluon (in a $b\bar{b}$ pair).
2. *s-channel*: This channel is the less frequent and it consist on the two quarks fusion which produce a W boson. It is a Drell-Yan type of process.
3. *Associated production (tW)*: Unlike the other channels, top quark is produced alongside a W boson.

Energies $[TeV]$	cross section $[pb]$	Scale $[pb]$	PDF $[pb]$	Total unc. $[pb]$
7	63.89	+1.92 -1.25	+2.19 -2.19	+2.91 -2.52
8	84.69	+2.56 -1.68	+2.76 -2.76	+3.76 -3.23
13	216.99	+6.62 -4.64	+6.16 -6.16	+9.04 -7.71
14	248.09	+7.58 -5.40	+6.98 -6.98	+10.30 -8.82

Table 5: Single top production's cross section for the t-channel. [29]

In table 5 are shown the cross section for the t-channel single top production depending on the energies at which the LHC has worked over the years along its uncertainties.

6 Final states for tHW events

The process that has been studied in this project is tHW, which consists on a top quark coupling with a Higgs boson and a W boson. Its final states are determined by the decay of its particles. On one side, top quark decay (99.8%) on a b quark and a W boson. This results on a state with two W boson, a b quark and the Higgs boson. Therefore, depending on the decay products of W bosons and the decay of the Higgs bosons, it can be obtained several final states. Higgs bosons has a probability of around 60 % of decaying into two b quarks. Also, it has been taking into account the decay into two W bosons (around 20%). This means that there are two possible intermediate states: three b quarks and two W bosons (when the Higgs bosons decay into two b quarks) or one b quark and four W bosons. So knowing these intermediate states, all possible final states for tHW events are determined by the combinations of W bosons decays depending on whether it is a leptonic or an hadronic decay.

Higgs decay modes	Products	Signal Region
H$\rightarrow b\bar{b}$	$\ell\nu + b\bar{b} + b + \ell\nu$	$2\ell\ 3q$
	$\ell\nu + b + b\bar{b} + qq$	$1\ell\ 5q$
	$qq + b + b\bar{b} + qq$	$0\ell\ 7q$
H$\rightarrow WW$	$\ell\nu+\ell\nu+\ell\nu+\ell\nu+b$	$4\ell\ 1q$
	$\ell\nu+\ell\nu+\ell\nu+b+qq$	$3\ell\ 3q$
	$\ell\nu+\ell\nu+b+qq+qq$	$2\ell\ 5q$
	$\ell\nu+b+qq+qq+qq$	$1\ell\ 7q$
	$qq+b+qq+qq+qq$	$0\ell\ 9q$

Table 6: Possibles final states (products) and signal regions of tHW process depending on the Higgs decay mode.

In the following figure, it is shown a feynman diagram corresponding to one of the tHW decays.

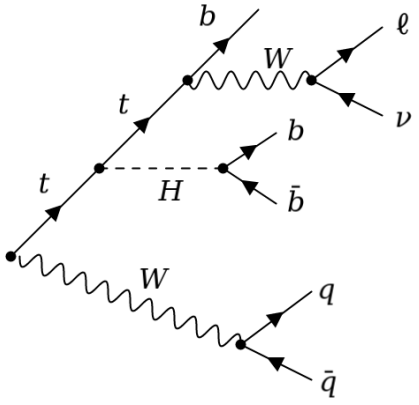


Figure 10: Feynman diagram of tHW decay.

6.1 Background processes

Within the Standard model, there are several processes whose final states are identical or very similar to the one which is studied (tHW). This cause that signal events can be misidentified with those processes, which constitute the background processes. Background processes are classified in two groups: irreducible and reducible. Irreducible background are those processes whose final states are identical to those of the signal or very similar. However, there are other background processes which have different final states (similar) but due to detection system limitation or some other factors, they can be misidentified. These processes are called reducible background. One of the factors that may caused this backgrounds are the non-prompt leptons which are leptons that do not arise from the vertex of the interaction. Also, misidentifying jets or not detecting them due to the geometry of the detector play a role in these reducible backgrounds. The main background processes for tHW are $t\bar{t}V$ ($t\bar{t}W$ and $t\bar{t}Z$), $t\bar{t}$, $t\bar{t}H$ and dibosons. In the case of $t\bar{t}V$, they are irreducible.

1. ***$t\bar{t}W$***. This is one of the $t\bar{t}V$ processes that can be a background process for to the signal process (tHW). Top quarks decay into bW (99.8%) which results in an intermediate state of 2 b quarks and 3 W bosons. Therefore, depending on whether W bosons decay leptonically or hadronically, the possibles final states are

Products	Signal region
$\ell\nu{+}qq{+}qq{+}b\bar{b}$	$1\ell\ 6q$
$\ell\nu{+}\ell\nu{+}qq{+}b\bar{b}$	$2\ell\ 4q$
$\ell\nu{+}\ell\nu{+}\ell\nu{+}b\bar{b}$	$3\ell\ 2q$
$qq{+}qq{+}qq{+}b\bar{b}$	$0\ell\ 8q$

Table 7: Signal regions and products of all the possible final states of $t\bar{t}W$.

Comparing $t\bar{t}W$ and tHW, it can be seen how the final states are very similar. For example, in the channel with three leptons the main difference is that tHW has a extra quark which may be not detected due to several reasons as it is been previously explained. The same thing happens for the other three channel. The similarities between process (signal are background) is what make it a irreducible background.

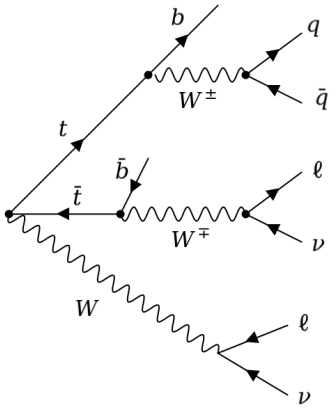


Figure 11: Feynman diagram for a decay channel of $t\bar{t}W$ process.

2. $\boldsymbol{t\bar{t}Z}$. This is the other process that is part of $t\bar{t}V$ and therefore is a irreducible background. In this case, knowing the decay of top quarks, the intermediate state is made of 2 b quarks, 2 W bosons and a Z boson. So, the different decays of the weak vector bosons leads to the following possibles final states

Z decay modes	Products	Signal region
Z$\to q\bar{q}$	$b\bar{b}$+$\ell\nu$+qq+$q\bar{q}$	1ℓ $6q$
	$b\bar{b}$+$\ell\nu$+$\ell\nu$+$q\bar{q}$	2ℓ $4q$
	$b\bar{b}$+qq+qq+$q\bar{q}$	0ℓ $8q$
Z$\to \ell\bar{\ell}$	$b\bar{b}$+$\ell\nu$+qq+$\ell\ell$	3ℓ $4q$
	$b\bar{b}$+$\ell\nu$+$\ell\nu$+$\ell\bar{\ell}$	4ℓ $2q$
	$b\bar{b}$+qq+qq+$\ell\bar{\ell}$	2ℓ $6q$
Z$\to \nu\nu$	$b\bar{b}$+$\ell\nu$+qq+$\nu\nu$	1ℓ $4q$
	$b\bar{b}$+$\ell\nu$+$\ell\nu$+$\nu\nu$	2ℓ $2q$
	$b\bar{b}$+qq+qq+$\nu\nu$	0ℓ $6q$

Table 8: Products and signal regions of the $t\bar{t}Z$ process depending on the Z decay mode.

Although there is one more final state than in the case of $t\bar{t}W$, this process contributes in the same way which consists on very similar final states to tHW but with an extra quark or one less, depending the channel, which can be misidentified or may be not detected.

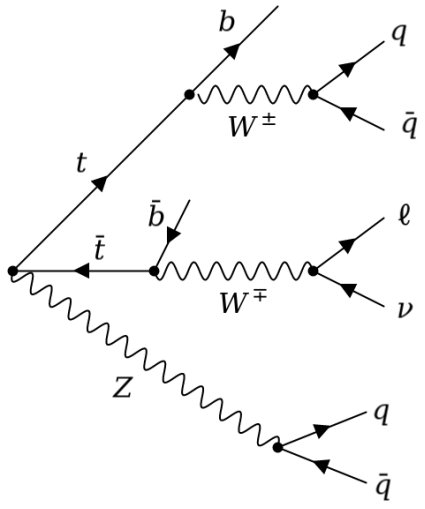


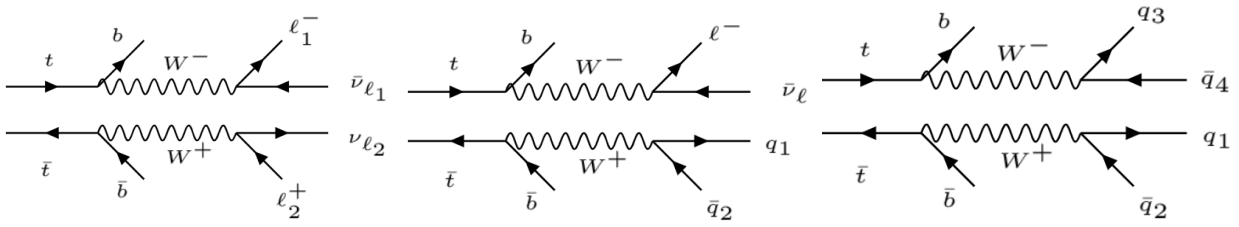
Figure 12: Feynman diagram for a decay channel of $t\bar{t}Z$ process.

3. $\boldsymbol{t\bar{t}}$. Top quarks decay into a W boson and a down-type quark (99.8% probability of being a b quark). This way, the intermediate state of the $t\bar{t}$ process results in 2 b quarks and two W bosons opposite-sign (electric charge). Therefore depending on the decay of the W bosons which can be hadronic or leptonic, four different decay modes can be distinguished:

Decay modes	Products	Signal region
Dileptonic	$b\bar{b}$+$\ell\nu$+$\ell\nu$	2ℓ $2q$
Semileptonic	$b\bar{b}$+$\ell\nu$+qq	1ℓ $4q$
Fully hadronic	$b\bar{b}$+qq+qq	0ℓ $6q$

Table 9: Products and signal region of $t\bar{t}$ for each decay mode.

In this case, for the semileptonic decay the final state is 1 lepton and 4 quarks which is very similar to the channel 1 lepton and 5 quarks of the tHW process. Also the fully-hadronic channel may caused problems if there is a non-prompt lepton. A part from the decay modes of table 9, W bosons can decay into tau which can decay into other particles.



(a) Dileptonic (b) Semi-leptonic (c) Fully-hadronic

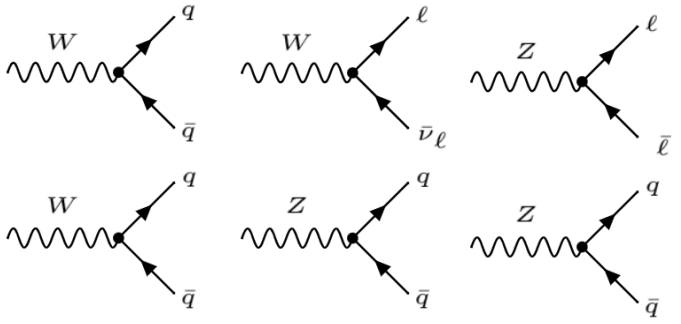
Figure 13: Feynman diagrams of $t\bar{t}$ decay modes.

4. **Dibosons**. Some dibosons processes have final states similar to tHW and therefore, they can be background, specifically WZ and ZZ processes but also WW. Depending on the different decays of the weak vector boson, the possible final states are:

Process	Products	Signal region
WW	$\ell\nu$+$\ell\nu$	2ℓ $0q$
	qq+qq	0ℓ $4q$
	$\ell\nu$+qq	1ℓ $2q$
WZ	$\ell\nu$+$q\bar{q}$	1ℓ $2q$
	$\ell\nu$+$\ell\bar{\ell}$	3ℓ $0q$
	$\ell\nu$+$\nu\nu$	1ℓ $0q$
	qq+$q\bar{q}$	0ℓ $4q$
	qq+$\ell\bar{\ell}$	2ℓ $2q$
	qq+$\nu\nu$	0ℓ $2q$
ZZ	$q\bar{q}$+$q\bar{q}$	0ℓ $4q$
	$q\bar{q}$+$\ell\bar{\ell}$	2ℓ $2q$
	$q\bar{q}$+$\nu\nu$	0ℓ $2q$
	$\ell\bar{\ell}$+$\ell\bar{\ell}$	4ℓ $0q$
	$\nu\nu$+$\nu\nu$	0ℓ $0q$
	$\nu\nu$+$\ell\bar{\ell}$	2ℓ $0q$

Table 10: Products and signal regions produced by each dibosons processes.

Some of the reasons that can produce similar final states to the signal ones could be when ZZ process decay to 4 leptons and an extra quark is detected from another source or in other channels could be due to a combinations of a non-prompt lepton and a misidentified jet.



(a) WW process (b) WZ process (c) ZZ process

Figure 14: Feynman diagrams for dibosons processes.

5. $\boldsymbol{t\bar{t}H}$. Even though it has not been considered in this project, $t\bar{t}H$ can be a background of tHW. The intermediate state of this process consists on two b quarks, two W bosons and a Higgs boson so depending on the decay mode of W and Higgs bosons, the following final states are possible

Higgs decay mode	Products	Signal region
H$\to b\bar{b}$	$\ell\nu$+$\ell\nu$+$b\bar{b}$+$b\bar{b}$	2ℓ $4q$
	$\ell\nu$+qq+$b\bar{b}$+$b\bar{b}$	1ℓ $6q$
	qq+qq+$b\bar{b}$+$b\bar{b}$	0ℓ $8q$
H$\to WW$	$\ell\nu$+$\ell\nu$+$\ell\nu$+$\ell\nu$+$b\bar{b}$	4ℓ $2q$
	$\ell\nu$+$\ell\nu$+$\ell\nu$+qq+$b\bar{b}$	3ℓ $4q$
	$\ell\nu$+$\ell\nu$+qq+qq+$b\bar{b}$	2ℓ $6q$
	$\ell\nu$+qq+qq+qq+$b\bar{b}$	1ℓ $8q$
	qq+qq+qq+qq+$b\bar{b}$	0ℓ $10q$

Table 11: Products and signal region for $t\bar{t}H$ depending on the Higgs decay mode.

Looking at the final states presented in table 11, it is clearly shown that $t\bar{t}H$ is a irreducible background. In all final states with at least 1 lepton, misidentifying a jet would cause the final state of both processes to be the same. The reason why $t\bar{t}H$ is not used in this project is because the study was aiming to separate tHW from $t\bar{t}V$.

7 Analysis

On this project, it has been carried out an analysis of the tHW sample at GEN-level. Generation level (GEN-level) is a description of the simulated process considering all the partons before hadronization. Therefore, it is used MonteCarlo simulation instead of experimental data. The only decays modes of the Higgs bosons taken into account are the two dominant (two b quarks or two W bosons). This way, properties of this sample has been studied in the corresponding possibles decay channels. On the other side, the main part of the study, which includes the signal discrimination using boosted decision trees, are studied in different regions defined based on the number of leptons.

7.1 Object selection

To perform the analysis of this project, it is necessary to use several selection criteria in order to obtained the best possible signal events.

First of all, several cuts have been applied in order to avoid non prompt leptons. In the analysis, leptons electrons and muons only (and its respective antiparticles). The cuts applied require that leptons must have a pT larger than 15 GeV and also that they must have $|\eta| \leq 2.4$.

The analysis is made at GEN-level, so from simulations, the information of quarks is directly available. As the behavior of quarks is different from the jets, it has also been studied more complex objects called genjets which are jets at GEN-level. The cuts applied to these objects are a pT > 15 GeV and a $|\eta| \leq 4.7$. Genjets are classified according to their flavor and their $|\eta|$. Attending to their flavor, as well as quarks, genjets can be divided in heavy flavor and light flavor genjets. Also, they can be divided in genjets and forward genjets. Forward genjets are those that satisfy $2.4 < |\eta| < 4.7$. The last condition applied to genjets is that the Delta R between a genjet and leptons has to be greater than 0.15 in order to avoid misidentify leptons with genjets.

7.2 Event selection

In the study of the tHW events, several regions have been defined taking into account the different possible final states, considering only the two dominant decays of the Higgs boson, which are H $\rightarrow b\bar{b}$ and H $\rightarrow$ WW. Therefore, it is been studied the following regions:

Channel	Products
4ℓ $1q$	$\ell\nu$+$\ell\nu$+$\ell\nu$+$\ell\nu$+b
3ℓ $3q$	$\ell\nu$+$\ell\nu$+$\ell\nu$ b+qq
2ℓ $5q$	$\ell\nu$+$\ell\nu$+b+qq+qq
2ℓ $3q$	$\ell\nu$+$\ell\nu$+b+$b\bar{b}$
1ℓ $7q$	$\ell\nu$+b+qq+qq+qq
1ℓ $5q$	$\ell\nu$+b+$b\bar{b}$+qq

Table 12: Channels defined in order to study (matching efficiencies) tHW process.

On the other side, for the main objective of the project, which is the use of machine learning techniques (BDT) to the signal discrimination (tHW events) from background ($t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$), is defined four new regions only depending on the number of leptons (1 lepton, 2 leptons, 3 leptons and 4 leptons).

7.3 Control Plots

First of all, it has been represented plots of some standard variables in order to check the validity of the MonteCarlo simulation that have been used in the project.

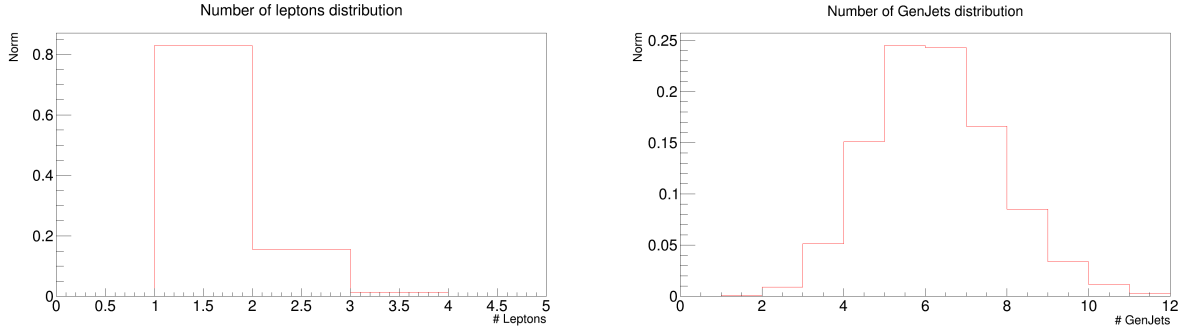


Figure 15: On the left is represented the distribution of number of leptons and on the right is plotted the distribution of number of genjets (all channels). Both distributions are for the tHW sample.

In figure 15, the plot on the left shows a distribution depending on the number of leptons. It is observed that there are not events without leptons which is expected because all the channels that have been studied has at least one lepton. Also, it is appreciated that most of the events has only one lepton (around 80%) which makes sense because the most probable decay of the Higgs boson is into two b quarks (not leptons) and W boson decay preferably hadronically. Also, it is important to take into account that some channels, especially for 4 leptons, have too low statistics which may affect to the analysis of those channels,

Then, in the plot on the right of the figure 15, it is shown a distribution depending on the number of genjets. The channel with a greater number of quarks is the one with 1 lepton and 7 quarks and the channel with a smaller number of genjets is the channel consisting on 4 leptons and 1 quark (see table 12). Knowing this, it makes sense that there are few events with less than four genjets. Also, it can be observed how there are additional genjets because the maximum number of genjets expected, based on the channels studied, are 7. Finally, the most frequent number of genjets match with what it was expected (around 5 or 6 genjets). The peak of the distribution is expected to be around those values because the dominant decays of Higgs boson and W bosons are hadronically.

Once it is been studied the histograms relative to the number of leptons and genjets, it has been checked the distribution of the transverse momentum and pseudorapidity of leptons and genjets.

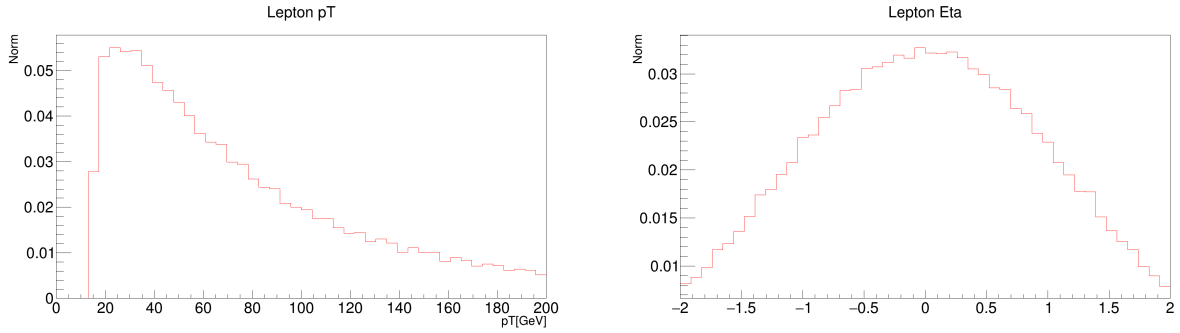


Figure 16: Distribution of the Pt and η of leptons (all channels).

In the case of the distribution of the leptons pt, the minimum value of the transverse momentum matches with the cut established for leptons and for lower values there are no entries, therefore it works correctly. Then, the distribution has the typical shape for the transverse momentum. On the other side, the η distribution of leptons is as expected, the peak of the distribution is found at 0 and the behavior is symmetric. It has been represented from -2.4 to 2.4 which are the cuts that have been applied.

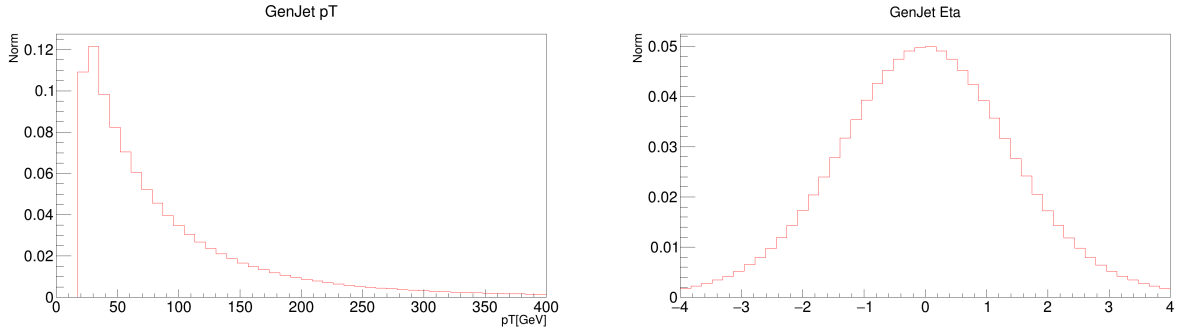


Figure 17: Distribution of the Pt and η of the genjets (all channels).

In figure 17 are shown the distribution of the pt and η of the genjets. As well as in the case of leptons, these genjets distribution are as it was expected. The η distribution is symmetric and the peak is at 0. The cuts on η are so that $-4.7 < \eta < 4.7$ as it can be seen in the histogram. This range of η has been chosen because forward genjets has been studied and they have $|\eta| > 2.4$. On the other side, the pt distribution has the typical behavior of the pt distributions and entries start at the value of the cut applied for genjets.

7.4 Quark-Genjet matching efficiencies

As the analysis carried out is at GEN level, they are studied both quarks and genjets. The information of the quarks is obtained from the GEN info and genjets are studied because they are jets at GEN-level.

Then, in order to know which genjet corresponds to each quark, a matching is made between them with the following requirements: A Delta R < 0.4 between them and both of them have the same flavor. Finally, the matching efficiency is plotted for several variables.

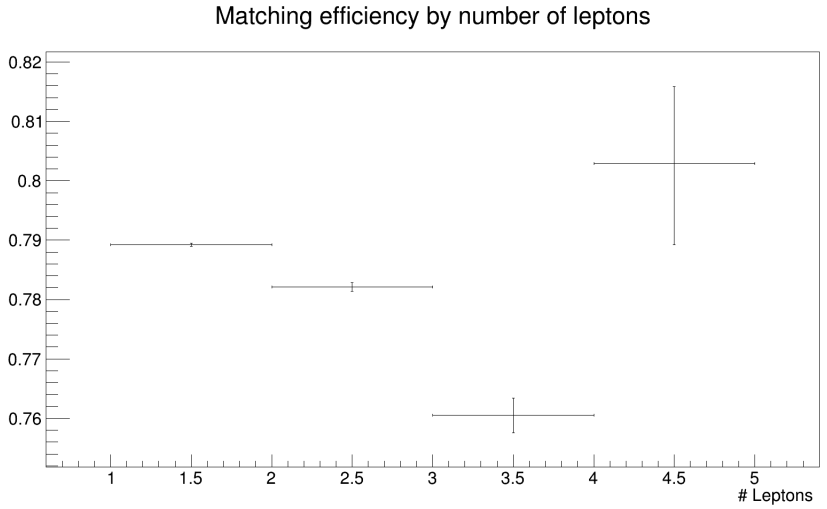


Figure 18: Matching efficiency depending on the number of leptons.

In this figure, it is shown the matching efficiency depending on the number of leptons. The larger matching efficiency seems to correspond to the channel of 4 leptons. First of all, it is necessary to take into account that there are few events of this channel which produce large error bars and it makes difficult its study. Observing the other three cases, the lower the number of leptons, which is also related to a larger number of quarks, the higher the matching efficiency. Nevertheless, the matching efficiency for all four cases is higher than 75%.

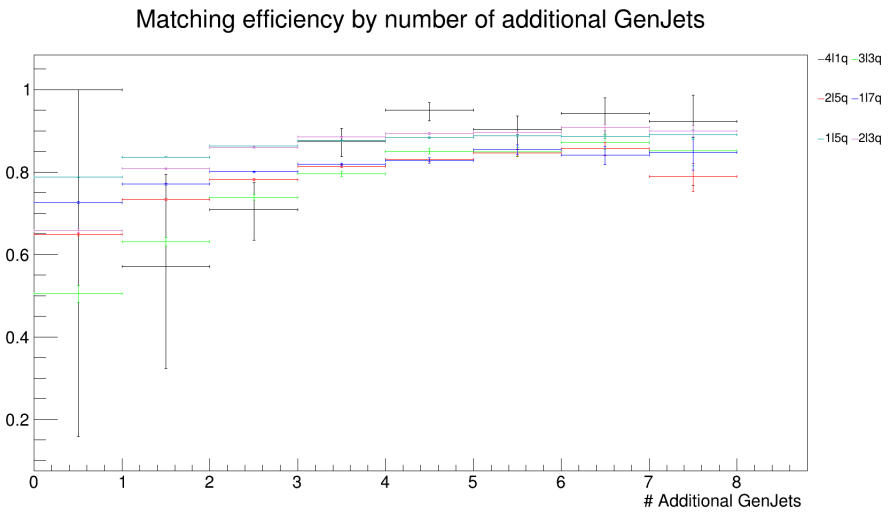


Figure 19: Matching efficiency depending on the number of additional genjets.

In this figure, it is represented the matching efficiency depending on the number of additional genjets for each channel. In the case of the channel of four leptons and one quark, it is observed large error bars again due to its few events. Also, it can be observed that the larger the number of additional genjets, the higher the matching efficiency. The lowest matching efficiency corresponds to the case without additional genjets while with more than 3 additional genjets the matching efficiencies remains more or less stable. This makes sense because it is more likely to have a match between a quark and a genjet, the more genjets there are. However there will be some point where adding genjets will have no effect because all possible quarks have been matched.

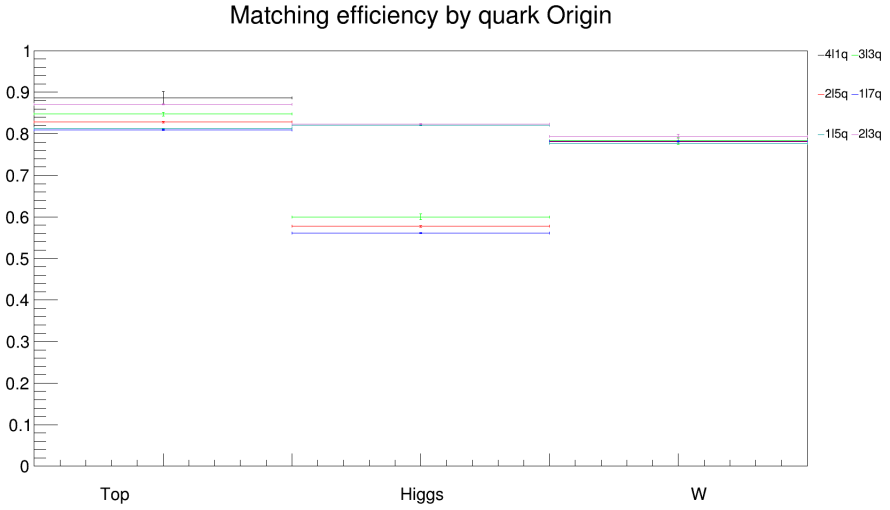


Figure 20: Matching efficiency depending on the origin of the particle.

Then, it has been plotted the matching efficiency depending on the origin of the particle as it is shown in the previous figure. There are three possible origins: the top quark, the Higgs

boson or the initial W boson. As expected, the matching efficiency for the three different cases are similar, larger than the 75%. However, there are unusually low matching efficiency values in some of the channels for the Higgs boson. These channels corresponds to the Higgs decaying into two W bosons. Therefore, it could be possible that genjets from the decay H→ WW were harder to match.

Finally, it has been plotted the matching efficiency of each channel in order to check the matching performance.

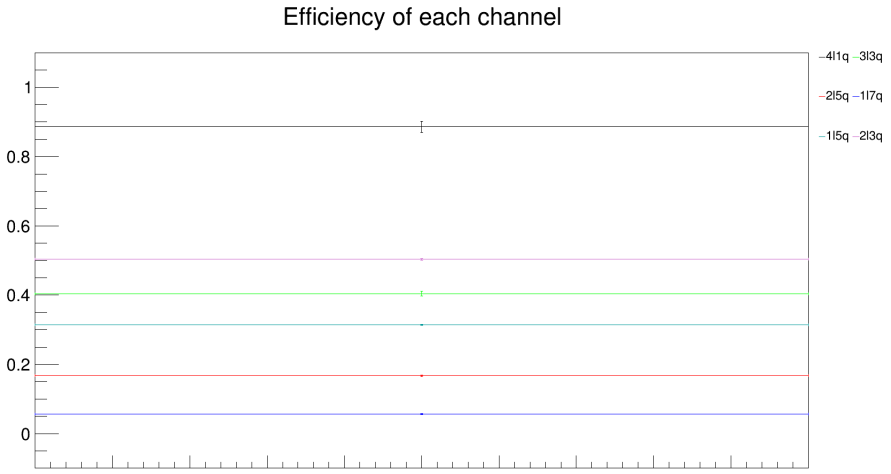


Figure 21: Matching Efficiency for each channel (all quarks-genjets matched).

In this figure, it has been represented the matching efficiency of each channel. This consists on see how often in a process all quarks are matched with a genjet and their origin is what is supposed to be depending on the channel. The results obtained are surprisingly less efficient than expected. The most efficient channel is 4 leptons and 1 b quark which makes sense because only one quark-genjet has to match. On the other side, due to the low matching efficiency of quarks which origin is Higgs bosons (H→ WW), make sense that those channels have a matching efficiency lower than 0.6. The least efficient channel corresponds to the channel with 1 lepton and 7 genjets what was expected because is the channel with a larger number of genjets. In any case, the matching efficiency of that process is extremely low.

7.5 BDT analysis

7.5.1 Regions and variables

In order to distinguished tHW events from their main backgrounds ($t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$), it has been applied machine learning techniques, in particular boosted decision tress (BDT). To perform this signal discrimination, it has been used four different channels because the previous ones were specific to the tHW decays. The channels used in this analysis are 1 lepton, 2 leptons, 3 leptons and 4 leptons.

Once channels are defined, it is necessary to provide to the machine learning technique several discriminant variables in order to perform the signal discrimination.

Variable	Definition
NJets	Number of genjets.
SumLepCharges	Sum of the electric charges of the leptons.
pTsubLep	Pt of the 3rd highest pt lepton (2nd highest pt lepton).
DeltaPhiSS	$\Delta\phi$ between two same-sign lepton (highest pt).
minLepDR	Minimum ΔR between any two leptons.
MaxEtaJet	Maximum $\|\eta\|$ of any light flavor (no b jet) genjet.
NUntaggedJet	Number of light flavor genjets.
detaFwdJetBJet	$\Delta\|\eta\|$ between forward light jet and leading b jet
detaFwdJet2BJet	$\Delta\|\eta\|$ between forward light jet and subleading b jet
detaFwdJetClosestLep	$\Delta\|\eta\|$ between forward light jet and closest lepton

Table 13: Input observables to the signal discrimination classifier.

These inputs are selected so that signal discrimination is as good as possible attending to the characteristics of each process. For example, the number of genjets and the number of light flavor genjets are useful to distinguish between processes with different number of genjets in their final states. Also, the sum of the electric charges of the leptons could help to distinguish processes when the origin of their leptons is different (like opposite-sign W bosons and any two W bosons). This way, the other variables also contribute to the signal discrimination using other properties as $|\eta|$ or pT.

7.5.2 *tHW* training

The first part of the analysis consists on the training of the signal process (tHW) versus each background process ($t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$) and for the different possible regions. That way, it is observed how well the signal discrimination classifier works in each case. To study the results of the training, it has been analyzed, for each training, three figures which are: the inputs variables, the inputs correlation matrix (both signal and background) and the BDT itself. The plot of the inputs variables is used to observe the difference between variables distributions of each process (signal and background) and therefore, it can be noted which input variables has a greater discriminating power. Then, input correlation matrix is useful to check the correlation between the inputs variables so they can be made changes if they are very correlated or at least this correlation is taking into account. The signal input correlation

matrix do not depends on the background process, only depends on the region, so this figure is shown once for each region. The last figure consists on the BDT itself where it can be seen both distributions (signal and background) and it is possible to get an idea of how well the signal discrimination works. If the BDT response is close to 1, the events are signal like and if it is close to -1, they are background like events. Also, it is useful to check the existence (or no existence) of overtraining.

In the case of $t\bar{t}$ events, the sample corresponds to the semileptonic decay and therefore, only it is possible to study the region of 1 lepton. As it was explained before, there is not a final state for $t\bar{t}W$ with four leptons, so this sample is analyzed in the other three regions (1 lepton, 2 leptons and 3 leptons). Finally, $t\bar{t}Z$ can be studied in all the regions defined but not all of the regions have the same amount of events.

7.5.2.1 tHW training versus $t\bar{t}$ in 1 lepton channel

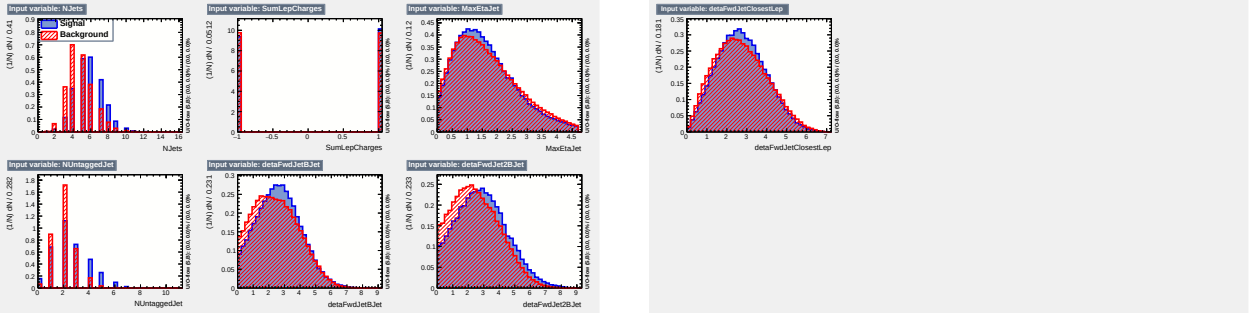


Figure 22: Distribution of the inputs variables for the training of tHW versus $t\bar{t}$ (1 lepton channel).

In the previous figure, it can be appreciated that inputs variables distributions are very similar except in the case of the number of genjets and the number of light flavor genjets. The peak of both input variables is located at lower values for the $t\bar{t}$ process. That was expected because the semileptonic decay of $t\bar{t}$ has fewer quarks (around 4) than tHW events which some decay channels have up to 7 quarks. The number of light flavor quarks is explained by the same reasoning to which should be added that in tHW the number of b quarks is usually higher than in $t\bar{t}$ events.

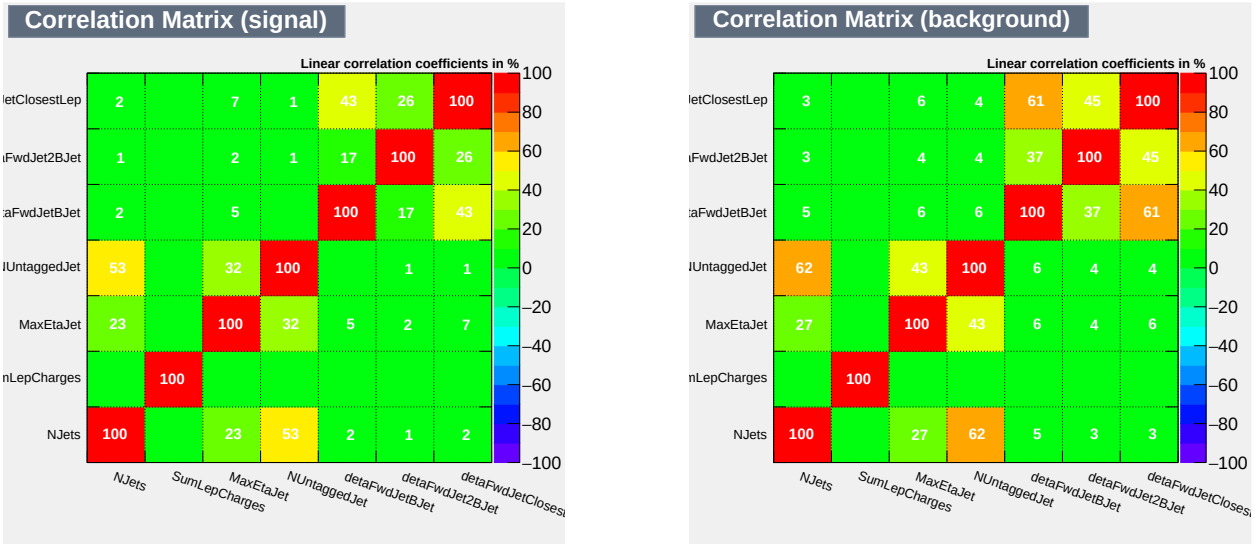


Figure 23: Input correlation matrix corresponding to the training of tHW versus $t\bar{t}$ (1 lepton channel) for both signal and background.

Looking at the left figure, it can be appreciated the correlation between the different input variables for the signal events, The highest correlation is between the number of genjets and the number of light flavor genjets. This high correlation was expected because by definition they are related to each other.

In the case of the background correlation matrix, correlations between the inputs variables are very similar to the ones commented for the signal. Comparing both correlation matrix, the main difference is that for the background the correlation between the number of light flavor genjets and the maximum $|\eta|$ is more relevant than before. Also, the correlation between "detaFwdJetBJet" and "detaFwdJetClosestLep" has increased considerably.

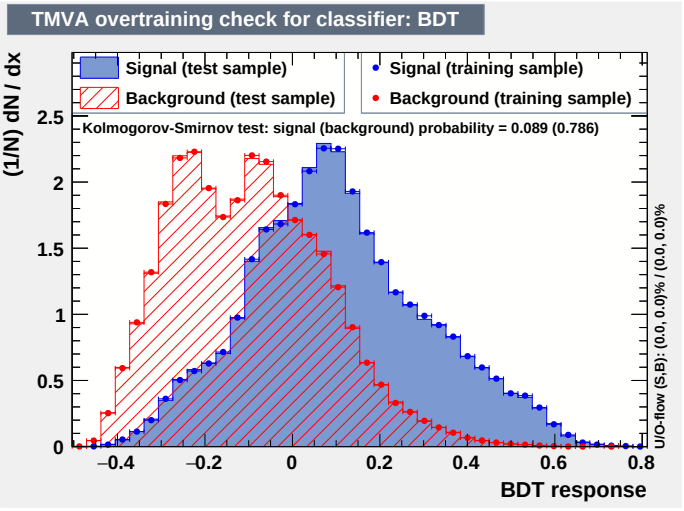


Figure 24: BDT output and overtraining check for the training of tHW versus $t\bar{t}$ (1 lepton channel).

In this figure, it is shown the BDT response for both signal and background and it is also checked for overtraining. To evaluate whether overtraining is occurring, it is used the Kolmogorov-Smirnov (KS) statistical test. As KS is greater than 0.01, it can be confirmed that there is no evidence of overtraining. Looking at the shape of the distribution, they are not completely discriminated but it is possible to establish a cut good enough to minimize the background events, keepin enough signal events.

7.5.2.2 tHW training versus $t\bar{t}W$ in 1 lepton channel

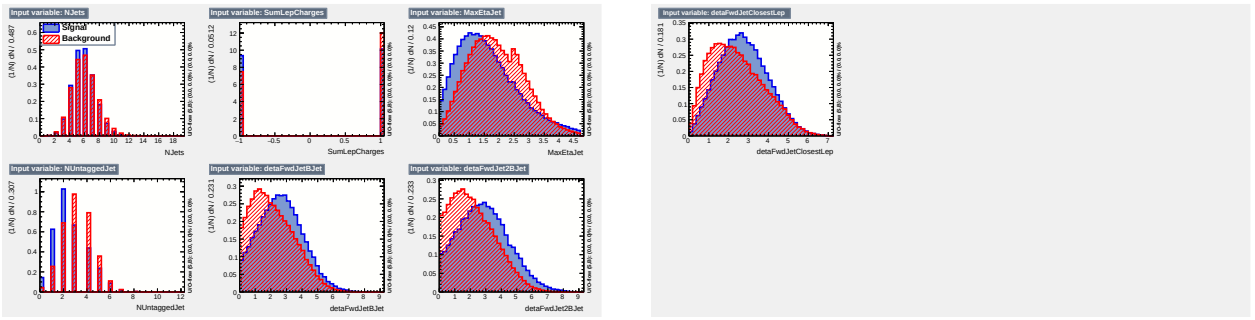


Figure 25: Input variables for the training tHW versus $t\bar{t}W$ (1 lepton channel).

In this case most of the distribution of the input variables are similar but with small differences. An example of this are the variables related with $\Delta\eta$ in which the peak of the background distribution correspond to lower values than the ones for the signal. The biggest difference between both distribution occur in the case of the number of light flavor genjets where the peak for the signal is around 2 genjets meanwhile for the background is around 3 genjets.

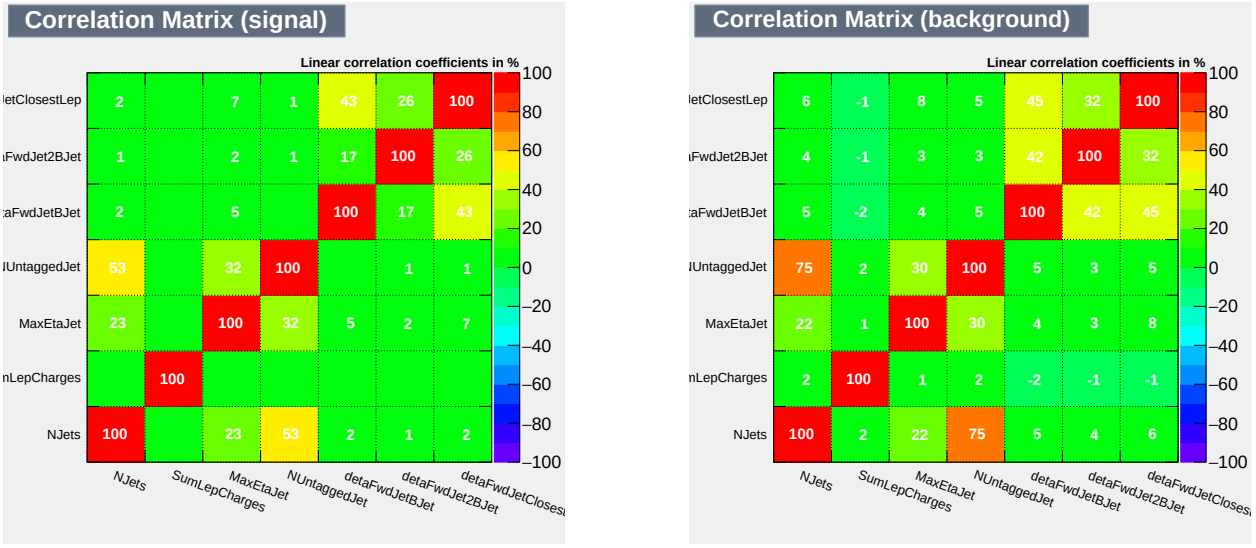


Figure 26: Inputs correlation matrix for the training tHW versus $t\bar{t}W$ (1 lepton channel).

The input correlation matrix (signal) is the same as before because the region is the same

(1 lepton). Observing the input correlation matrix of the background, it can be appreciated that is very similar to the signal one except for the correlation between the number of genjets and the number of light flavor genjets which has increased significantly.

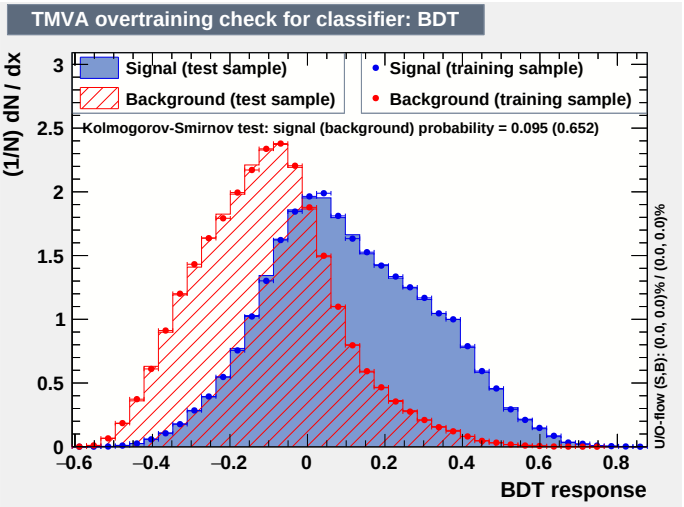


Figure 27: BDT output and overtraining check for the training of tHW versus $t\bar{t}W$ (1 lepton channel).

Observing the figure can be seen distributions of the BDT response of both signal and background. As it is shown in the figure, distributions are not completely separated but at some range of BDT response values, both of them well distinguished. It can be seen how for larger values of the BDT response, there are much more signal events than background events. Checking the overtraining, it is conclude using KS test that there is no evidence of it.

7.5.2.3 tHW training versus $t\bar{t}W$ in 2 leptons channel

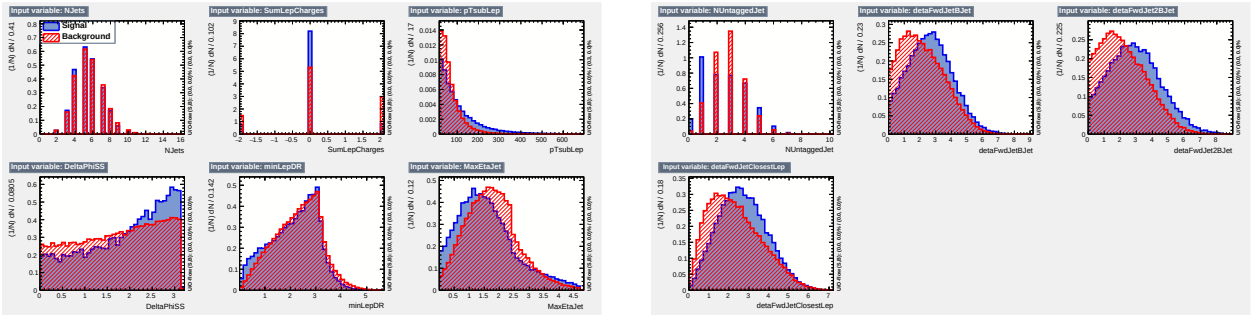


Figure 28: Input variables for the training tHW versus $t\bar{t}W$ (2 leptons channel).

In this case, the region of 2 leptons is studied, therefore there are a few more variables than before. The main differences between signal and background distributions are presented in the sum of the lepton electric charges and in the number of light flavor. Looking at the other

variables, distributions have a few differences between signal and background except for the number of genjets and the minimum ΔR between leptons which are almost identical.

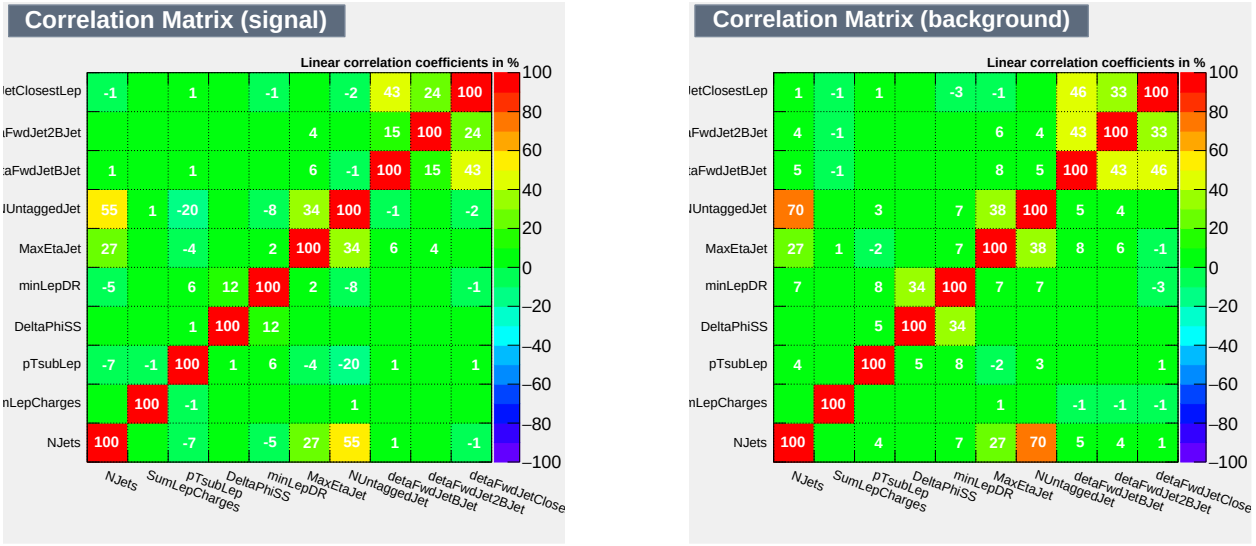


Figure 29: Inputs correlation matrix for the training tHW versus $t\bar{t}W$ (2 lepton channel).

Although variables have been added, inputs correlation matrix are similar to the ones before and none of those new variables have a high correlation with another one. In the signal matrix, as well as before, only the number of genjets and the number of light flavor genjets have a high correlation (above 50%). It is also considerable the correlation between "detaFwdJetBJet" and "detaFwdJetClosestLep". Other significant correlation (although they have not too high correlation) are between the number of light flavor genjets and the maximum $|\eta|$ of genjets, and it can be also appreciated a not too high correlation between the number of light flavor genjets and the pT of the subleading lepton.

In the case of the background, there is a high correlation between the number of genjets and the number of light flavor genjets as expected. As well as in the signal matrix, there are observed a considerable correlation between "detaFwdJetBJet" and "detaFwdJetClosestLep" and between the number of light flavor genjets and the maximum $|\eta|$ of genjets. In this background, it is also necessary taking into account the correlation between the minimum ΔR of leptons and the $\Delta\phi$ of same-sign leptons and it is also significant the correlation between "detaFwdJetBJet" and "detaFwdJet2BJet".

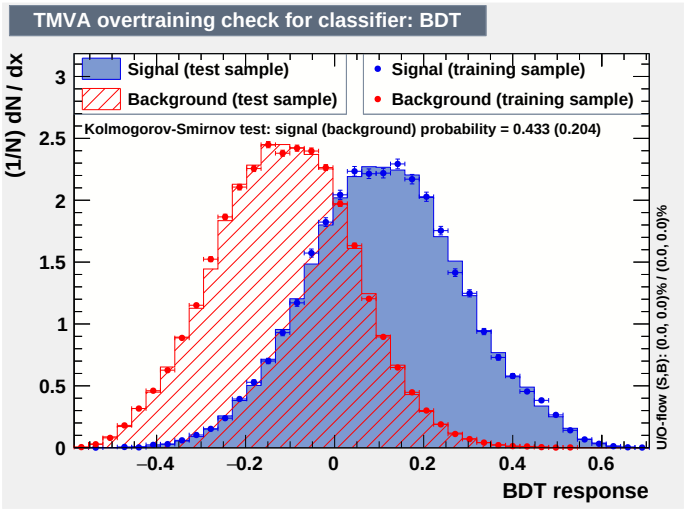


Figure 30: BDT output and overtraining check for the training of tHW versus $t\bar{t}W$ (2 leptons channel).

In this figure, it is shown the BDT reponse of $t\bar{t}W$ in the region of 2 leptons. Although distributions are not perfectly discriminated, both of them can be distinguished clearly and it could be performed a good cut in order to reject all the possible background keeping enough signal events. Attending to the KS test there is no overtraining.

7.5.2.4 tHW training versus $t\bar{t}W$ in 3 leptons channel

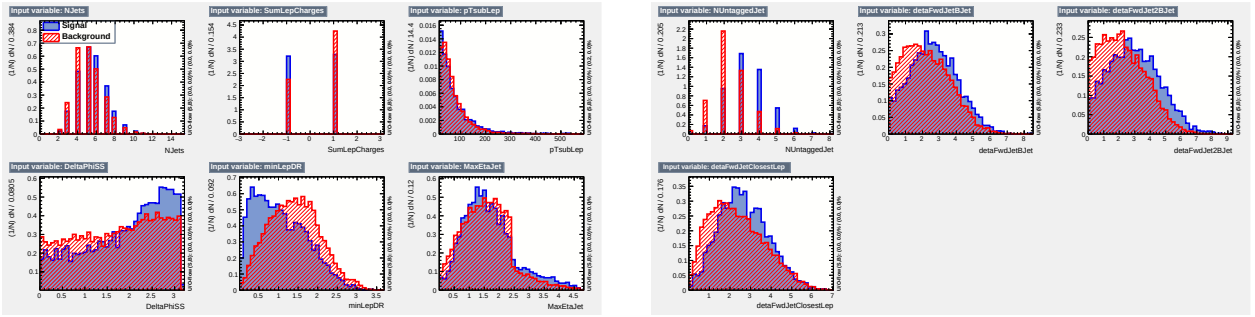


Figure 31: Input variables for the training tHW versus $t\bar{t}W$ (3 leptons channel).

In most of the distributions of the inputs variables shown in this figure, background and signal are very similar. Furthermore, some input variables as the maximum η of the genjets or the pT of the subleading lepton are almost identical. Nevertheless, there are some exception as the number of genjets and the sum of the lepton charges which distribution has significant differences between signal and background. On the other side, the most discriminating variables are the number of light flavor genjets, as well as in other cases, and the minimum ΔR between leptons. The difference in the distribution of the number of light flavor genjets was expected. This is because, in this channel, the three W bosons of the $t\bar{t}W$ process decay leptonically, therefore there are only two b quarks (they could be misidentified).

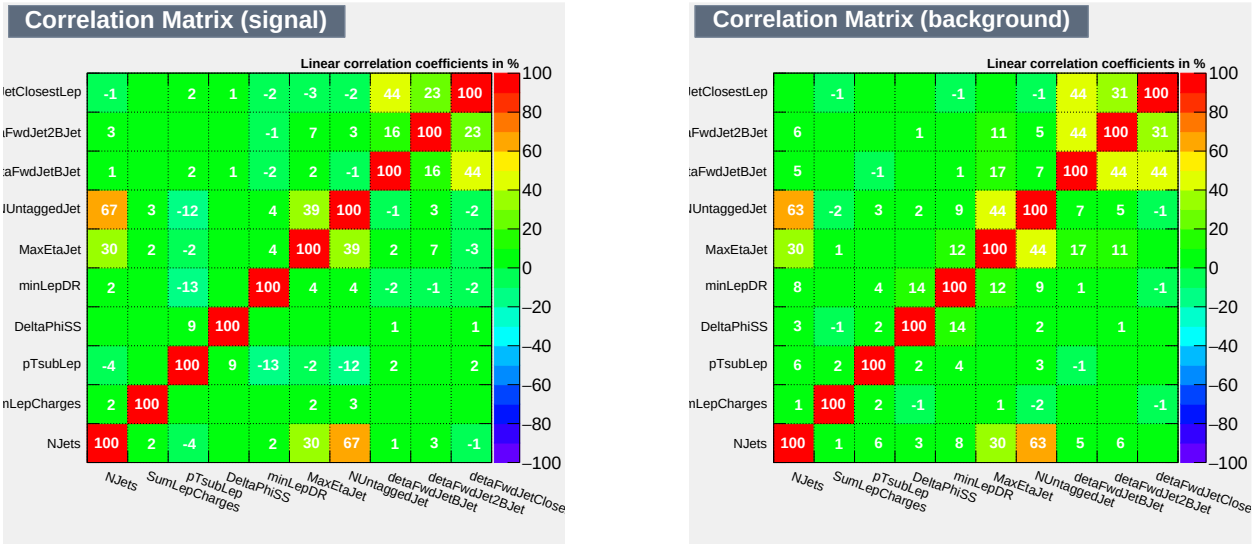


Figure 32: Inputs correlation matrix for the training tHW versus $t\bar{t}W$ (3 lepton channel).

For the region of 3 leptons, the signal input correlation matrix is very similar to the region of 2 leptons. It has the usual high correlation between the number of genjets and the number of light flavor genjets which has increased for 3 leptons. It is also high the correlation between "detaFwdJetBJet" and "detaFwdJetClosestLep" and between the number of light flavor genjets and the maximum $|\eta|$ of genjets. It is also noteworthy than in the region of 3 leptons is increased the correlation between the number of genjets and the maximum $|\eta|$ of genjets which makes sense because is also a bit correlated with "NUntaggedJet".

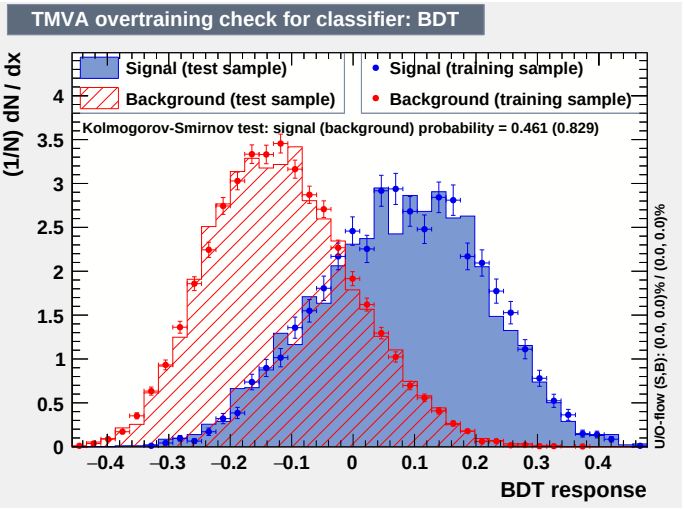


Figure 33: BDT output and overtraining check for the training of tHW versus $t\bar{t}W$ (3 leptons channel).

First of all, it is check for overtraining using the Kolmogorov-Smirnov test and as its larger than 0.01 it can be conclude that there is not overtraining. Looking at the shape of the distributions, both of them are well distinguished although they are not completely separate

which would be the ideal case. In addition, it can be observed that there are not too many events (low statistics) especially for the signal process.

7.5.2.5 tHW training versus $t\bar{t}Z$ in 1 leptons channel

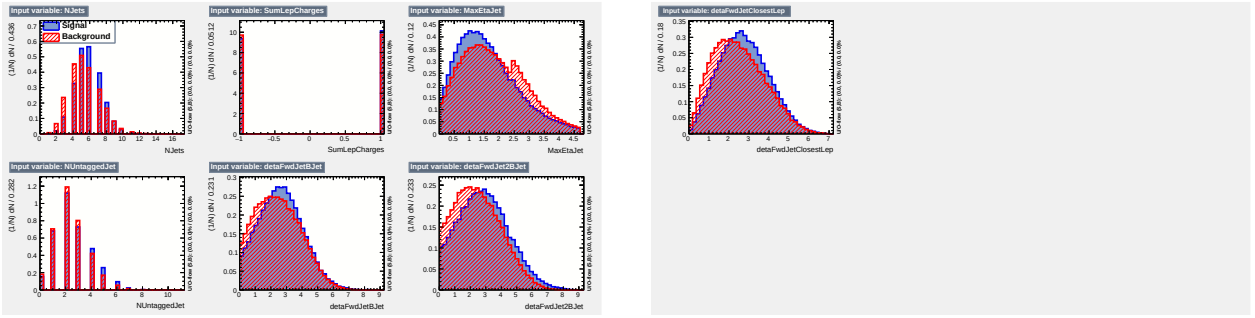


Figure 34: Input variables for the training tHW versus $t\bar{t}Z$ (1 lepton channel).

From these distribution, can not be inferred too much information because most of the inputs variables are almost identical. The distribution of the number of jets could be the most discriminated variable in this case.

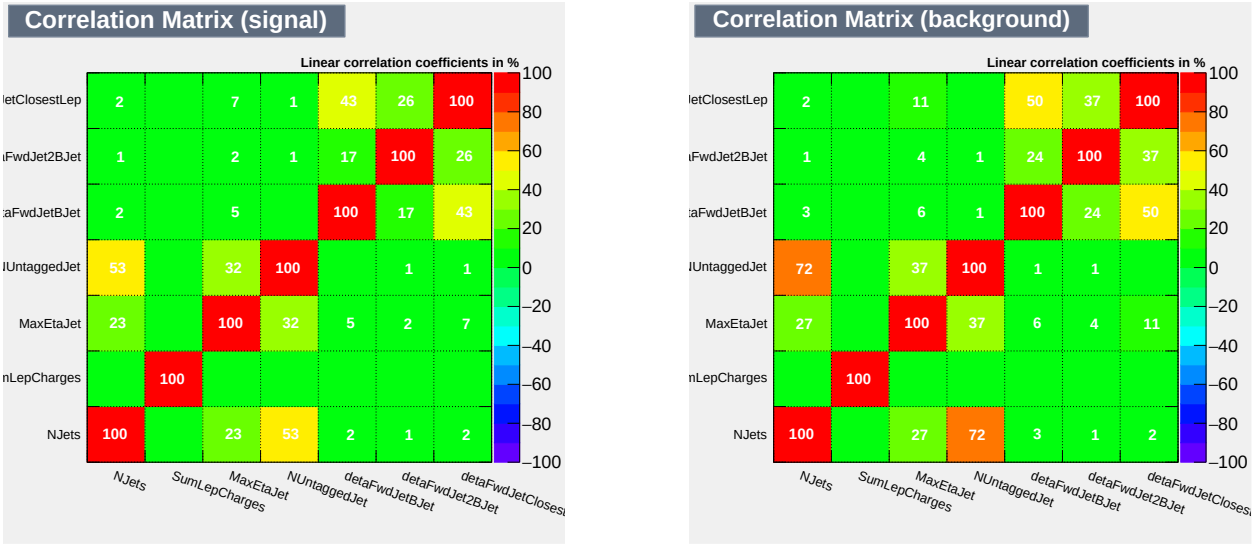


Figure 35: Inputs correlation matrix for the training tHW versus $t\bar{t}Z$ (1 lepton channel).

The signal correlation matrix is the same that in the other 1 lepton regionss explained before. Observing the background correlation matrix, the highest correlated variables are the number of genjets and the number of light flavor genjets which happens in most of the cases because for definition they are highly correlated. Then, as well as in other cases, another variables significantly correlated are for example "detaFwdJetBJet" and "detaFwdJetClosestLep".

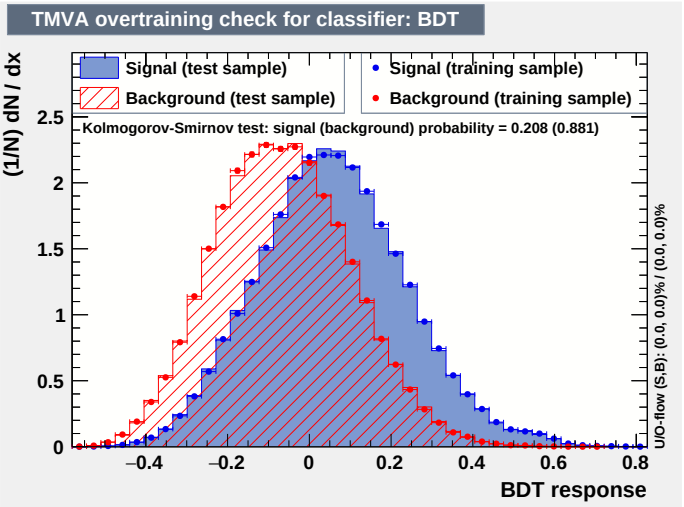


Figure 36: BDT output and overtraining check for the training of tHW versus $t\bar{t}Z$ (1 lepton channel).

Observing the BDT response, it can be appreciated how the signal discrimination seems not to be very effective because both distributions are almost overlapping. Nevertheless, applying a good cut maybe it is possible to obtained a good enough signal/background ratio. There is not overtraining according to the KS test.

7.5.2.6 tHW training versus $t\bar{t}Z$ in 2 leptons channel

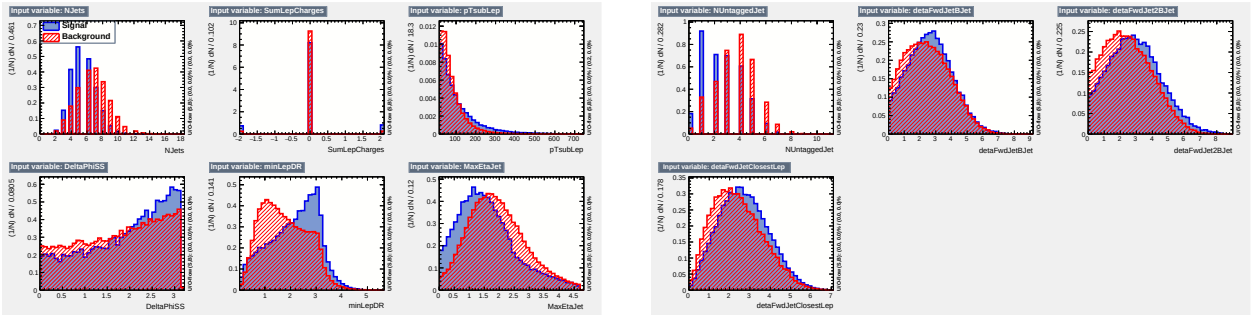


Figure 37: Input variables for the training tHW versus $t\bar{t}Z$ (2 leptons channel).

Looking at the input variables, it can be observed that some of them have a great discrimination power. These variables are the number of genjets, the number of light flavor genjets and the minimum ΔR between two leptons. The difference on the number of genjets are due to the decay channels of $t\bar{t}Z$ in this region. The background process can results on 2 genjets on the final states or 4 which explained the differenced seen in this distribution. This is also related to the number of light flavor genjets.

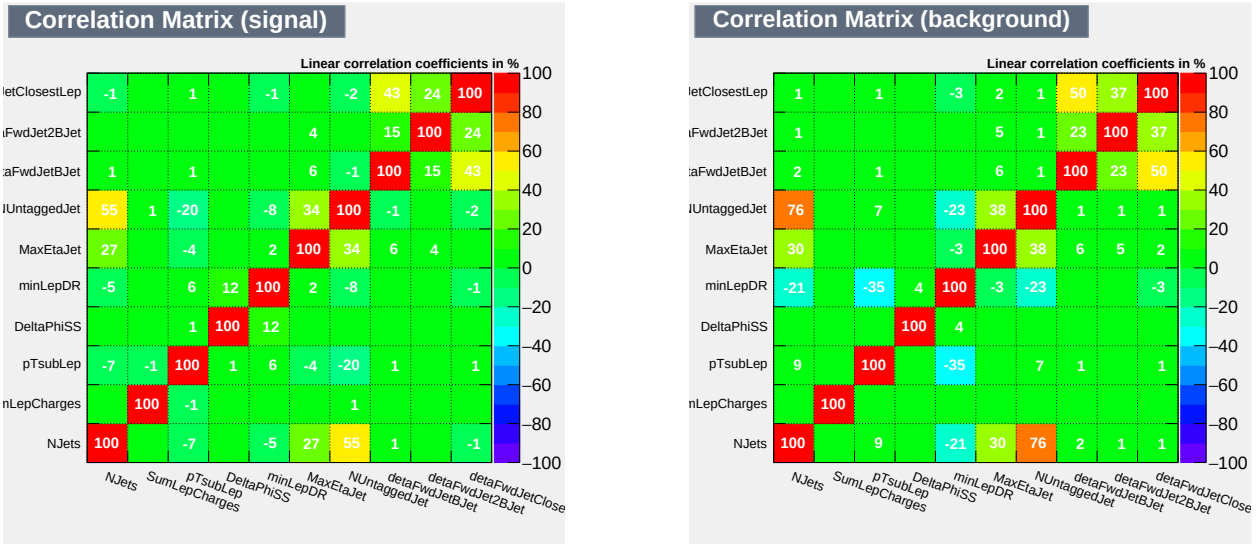


Figure 38: Inputs correlation matrix for the training tHW versus $t\bar{t}Z$ (2 lepton channel).

In this figure, it can be seen the inputs correlation matrix for signal and background. In the case of the signal correlation matrix, it is the same that the one for $t\bar{t}W$ in the region of 2 leptons. On the other side, in the background correlation matrix can be seen the correlation between all the variables. It is noteworthy the evidence of a small correlation between minimum ΔR of leptons and the pT of the subleading lepton.

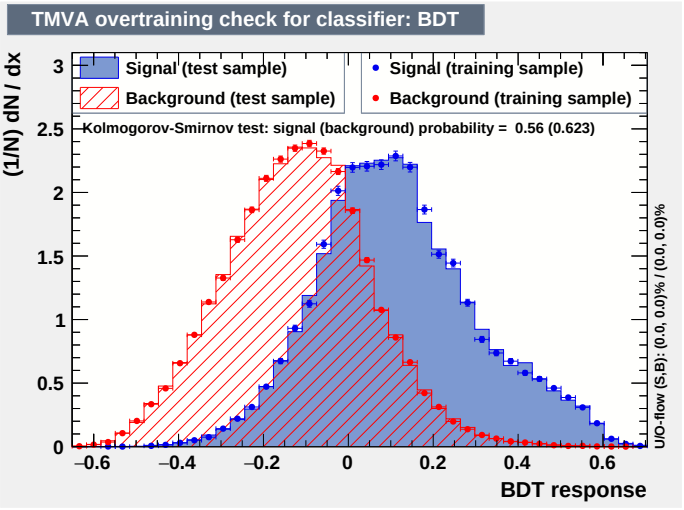


Figure 39: BDT output and overtraining check for the training of tHW versus $t\bar{t}Z$ (2 leptons channel).

According to the KS test, it can be determined that there is not overtraining. Broadly speaking, the signal discrimination seems to be well separated in order to obtain a great signal/background ratio although it could be better.

7.5.2.7 tHW training versus $t\bar{t}Z$ in 3 leptons channel

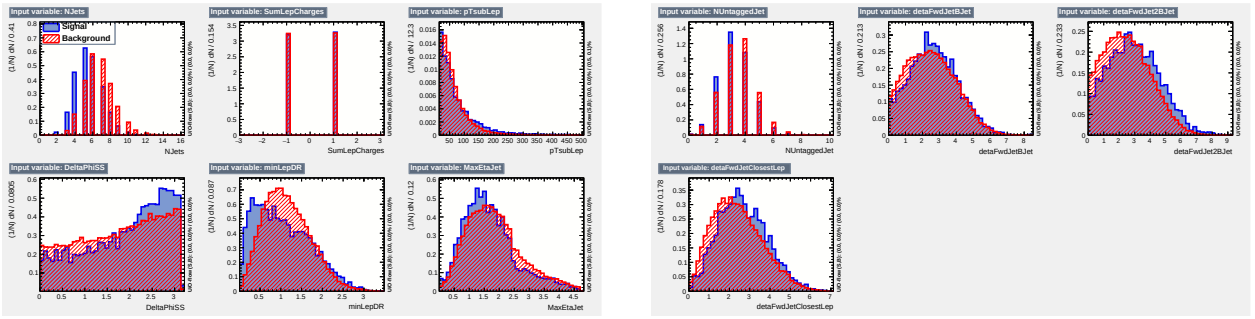


Figure 40: Input variables for the training tHW versus $t\bar{t}Z$ (3 leptons channel).

In this figure, it is shown the different inputs variables used in the analysis. It can be observed as most of them are identical or very similar. The number of genjets and the minimum ΔR of leptons seems to have the greatest discriminating power,

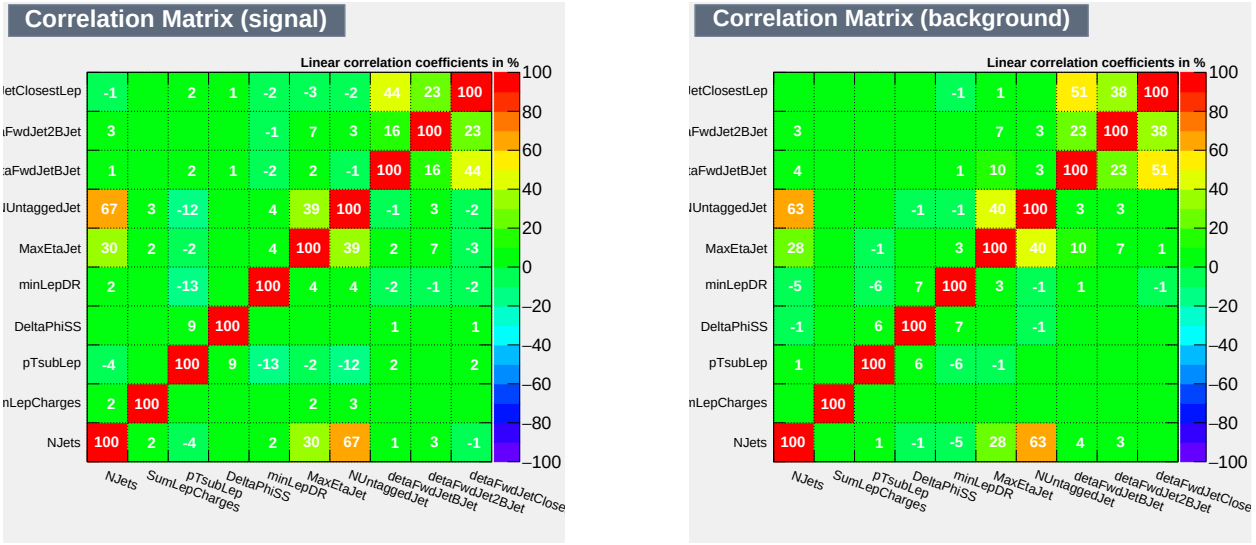


Figure 41: Inputs correlation matrix for the training tHW versus $t\bar{t}Z$ (3 lepton channel).

The signal correlation matrix is the same as the one for the training versus $t\bar{t}W$ in the region of 3 leptons. On the other side, it can be seen, in the background correlation matrix, the correlation between the different input variables. In this case, the highest correlated variables are the number of genjets and the number of light flavor genjets as expected,

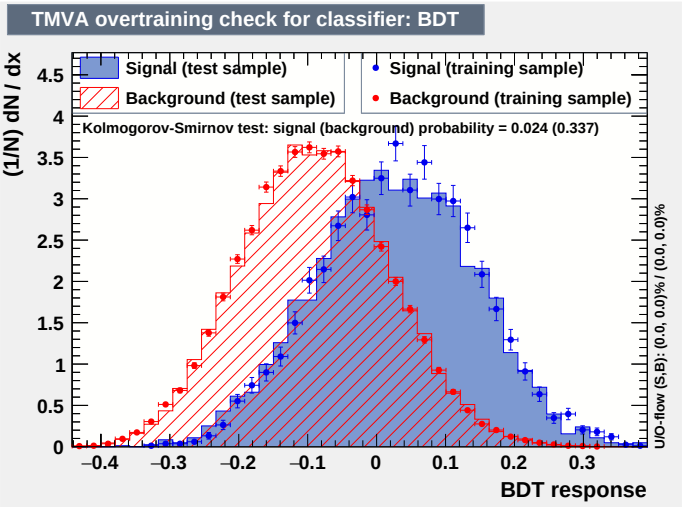


Figure 42: BDT output and overtraining check for the training of tHW versus $t\bar{t}Z$ (3 leptons channel).

Considering that with a KS greater than 0.01 there is no overtraining, it can be confirmed that there is not evidence of it. Looking at the shape of the distribution, as well as in other cases, distributions are good enough separated although there are nor perfectly separated which would be the ideal case.

7.5.2.8 tHW training versus $t\bar{t}Z$ in 4 leptons channel

Finally, it is shown the same figures that before but in the case of $t\bar{t}Z$ in the region of 4 leptons.

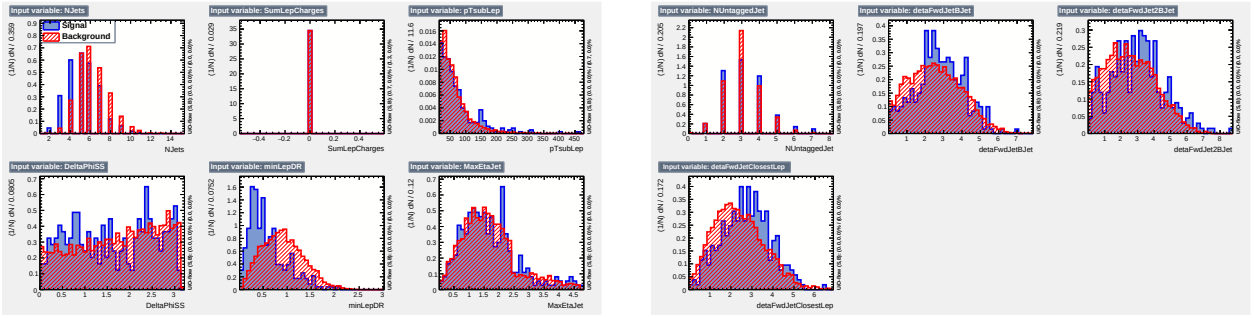


Figure 43: Input variables for the training tHW versus $t\bar{t}Z$ (4 leptons channel).

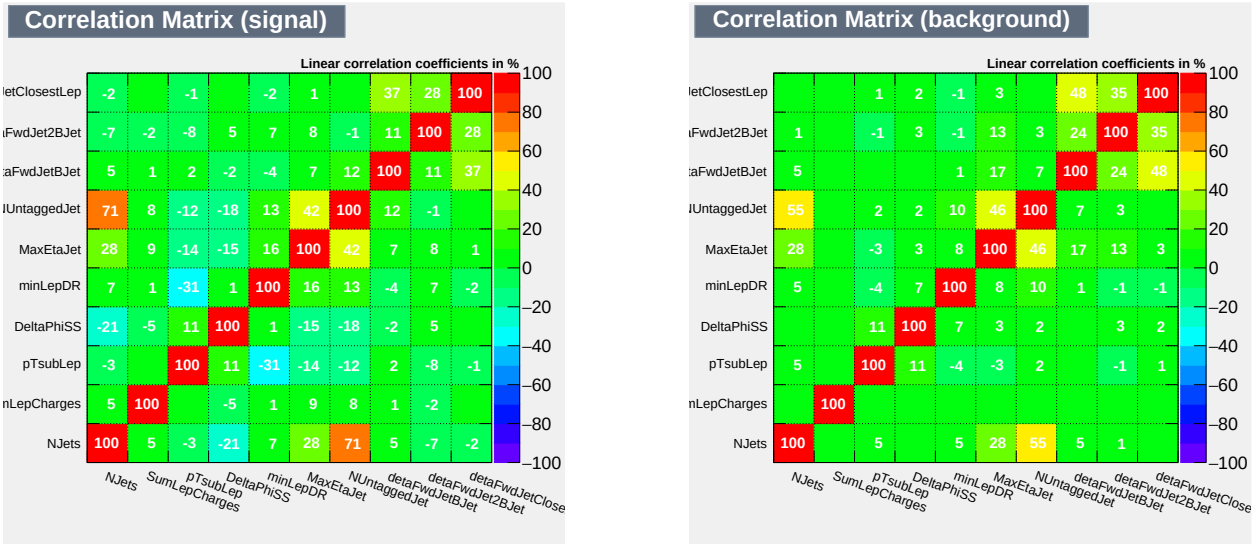


Figure 44: Inputs correlation matrix for the training tHW versus $t\bar{t}Z$ (4 lepton channel).

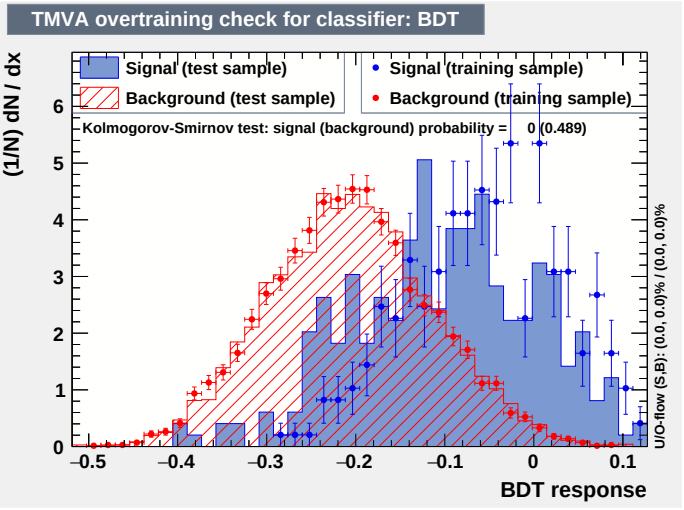


Figure 45: BDT output and overtraining check for the training of tHW versus $t\bar{t}Z$ (4 leptons channel).

On the previous figures can be seen how the few number of events prevents to analyze this region. Looking at the Kolmogorov-Smirnov test, it can be concluded that there is overtraining which is due to the few signal events available for this regions. The distributions of the inputs variable provides a hint of the effects of the few events in the analysis.

7.5.3 Training tHW vs $t\bar{t}V$

Once it has been studied the trainings of tHW versus each background, it has been decided to make the analysis versus $t\bar{t}V$. In addition, it has been noted that the region corresponding to 4 leptons has few signal events. Therefore, it has been chosen to modify the regions and instead of using the regions of 3 and 4 leptons separately, they have been merged into one. So the new regions are: 1 lepton, 2 leptons and 3 or 4 leptons.

7.5.3.1 1 lepton region

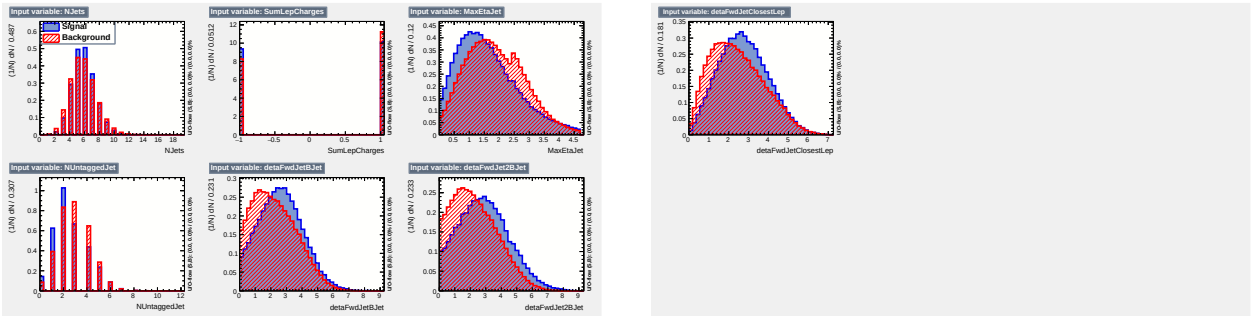


Figure 46: Input variables for the training tHW versus $t\bar{t}V$ (1 lepton channel).

In this figure are presented the inputs variables used in the analysis. The distributions are very similar (between signal and background) except for the number of light flavor genjets which seems to be the distribution with a greater discriminating power,

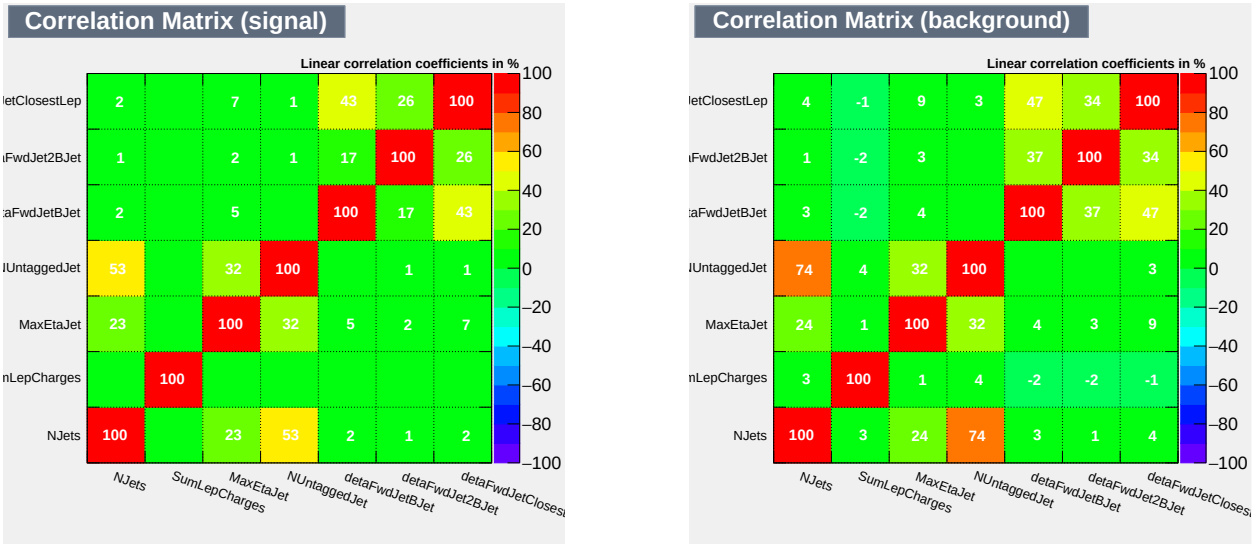


Figure 47: Inputs correlation matrix for the training tHW versus $t\bar{t}V$ (1 lepton channel).

Then, they are shown the input correlation matrix for both signal and background. The signal correlation matrix is the same as the previous ones which have been performed in the same region (1 lepton). Observing the background matrix, the highest correlated variables

(around 74%) are the number of genjets with the number of light flavor genjets which was expected. There are other variables with a correlation around 40% which is significant but it is not a very high correlation. The other variables are not significantly correlated with each other.

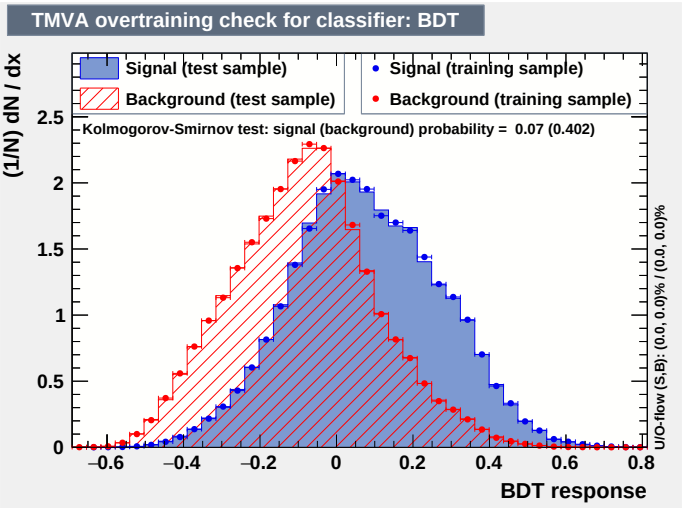


Figure 48: BDT output and overtraining check for the training of tHW versus $t\bar{t}V$ (1 lepton channel).

In this figure, it is plotted the BDT response of the training and it is also checked for overtraining. As the Kolmogorov-Smirnov test is greater than 0.01, there is no overtraining. Observing the shape of the distribution, both of them are a bit overlap but broadly speaking, it seems to be more signal events that background from 0 to 0.8.

7.5.3.2 2 leptons region

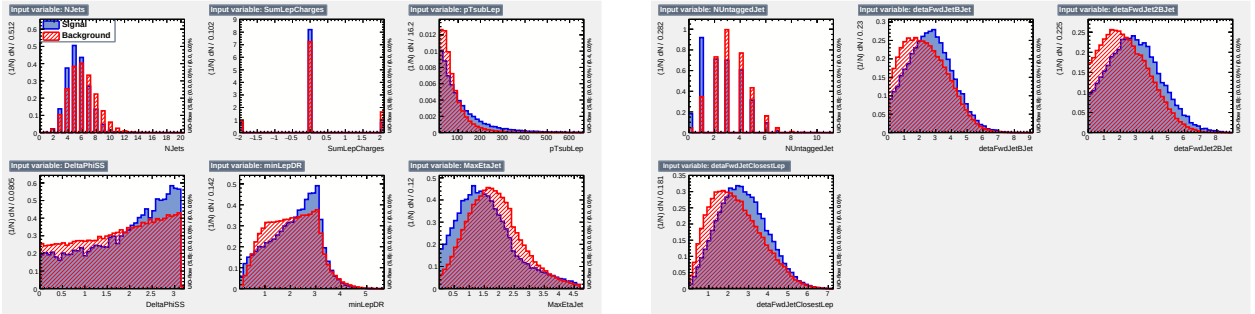


Figure 49: Input variables for the training tHW versus $t\bar{t}V$ (2 leptons channel).

As in the previous case, distributions of inputs variables are very similar for signal and background. At first sight, the input variables with a greater discriminating power seem to be the number of genjets and the number of light flavor genjets. It is also noteworthy, that some distributions vary slightly between signal and background as the minimum ΔR between leptons.

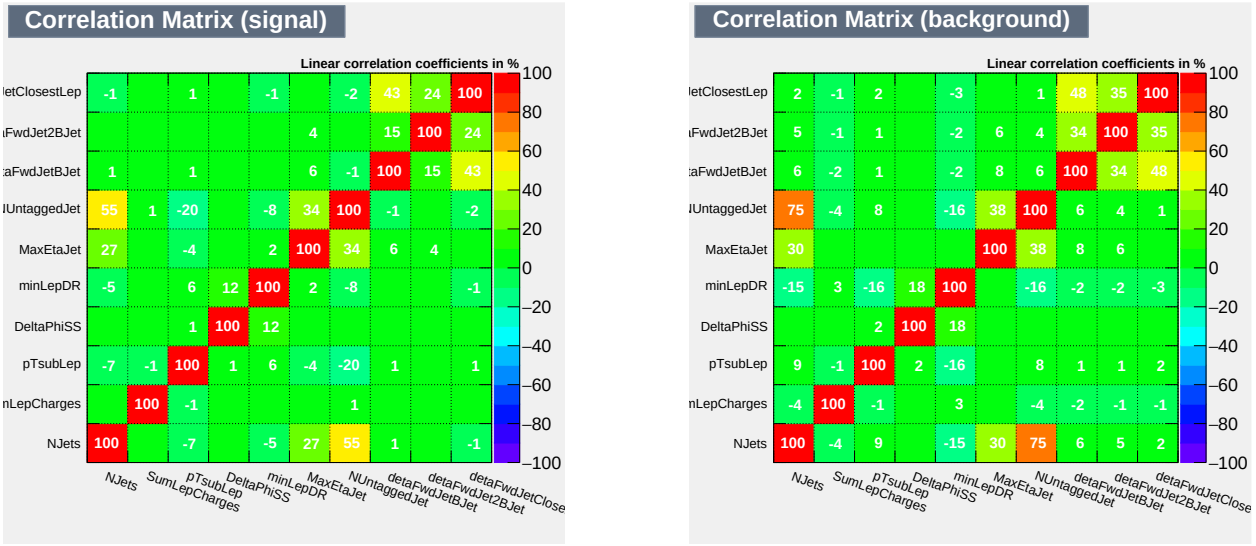


Figure 50: Inputs correlation matrix for the training tHW versus $t\bar{t}V$ (2 leptons channel).

Then, it is shown the inputs correlation matrix for both signal and background. In this figure, can be seen how input variables are correlated between them. The signal matrix is the same for all the trainings that have been performed in the 2 leptons regions. On the other side, it is also shown the matrix of the background process which their corresponding linear correlation coefficients for each pair of input variables.

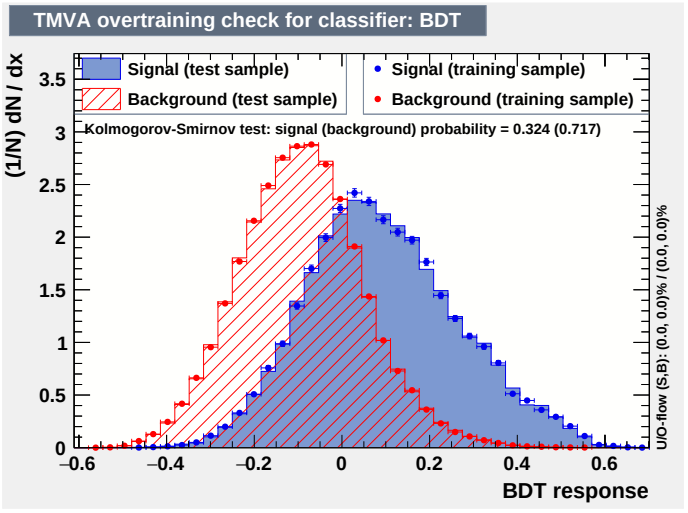


Figure 51: BDT output and overtraining check for the training of tHW versus $t\bar{t}V$ (2 leptons channel).

In this case, there is no overtraining either what it can be concluded because of the KS test. Observing the shape of the distributions, it seems that they are discriminated enough to obtain a good signal/background ratio although they slightly overlap.

7.5.3.3 3/4 leptons region

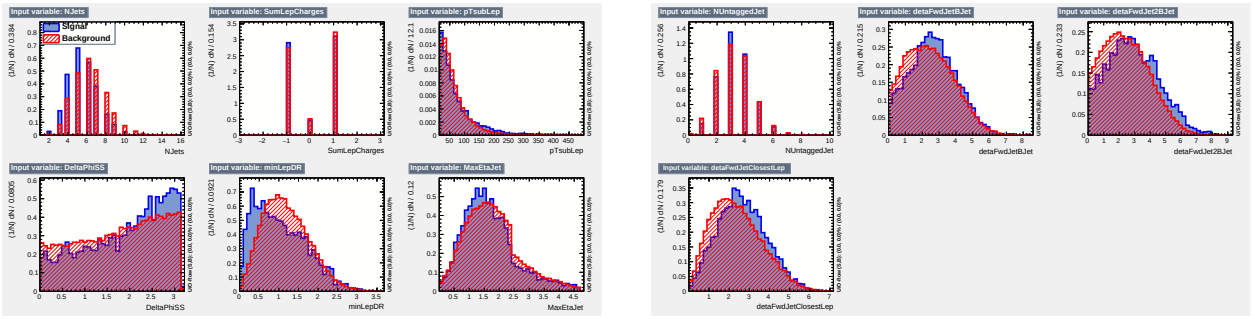


Figure 52: Input variables for the training tHW versus $t\bar{t}V$ (3 or 4 leptons channel).

In this figure, it is plotted the distributions of the inputs variables for signal and backgrounds. As well as in the previous cases, it is hard to determine, at first sight, which are the variables with a greater discrimination power. In this case, it seems to be the number of genjets.

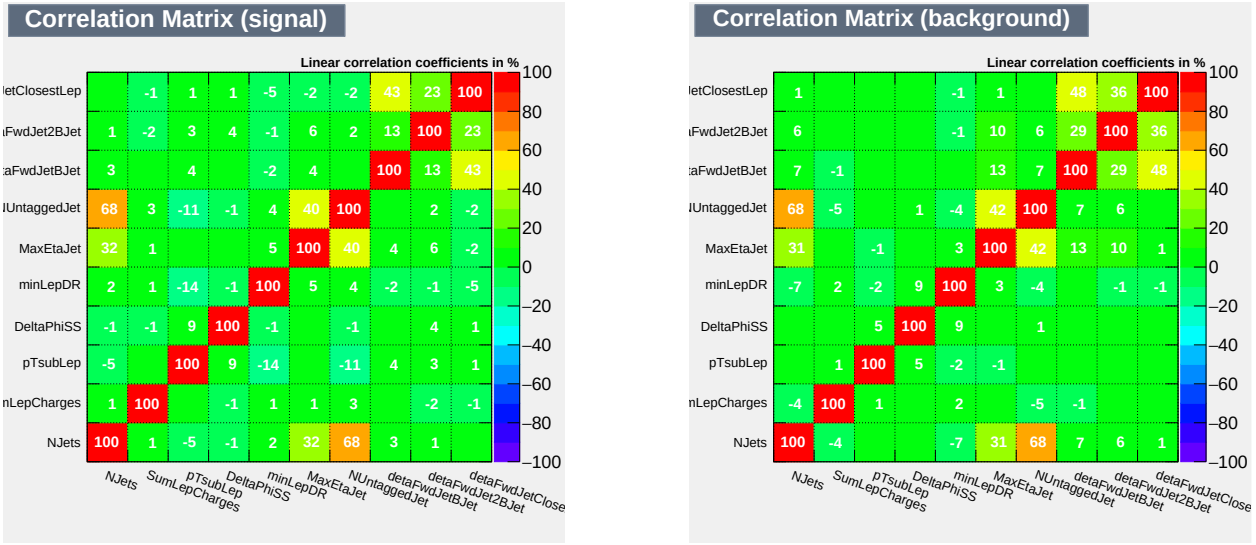


Figure 53: Inputs correlation matrix for the training tHW versus $t\bar{t}V$ (3 or 4 lepton channel).

Then, it is shown the input correlation matrix for both signal and background process. In this plots, it can be appreciated the correlation between the input variables. The highest correlation of any two variables is 68% for both cases between the number of genjets and the number of light flavor genjets.

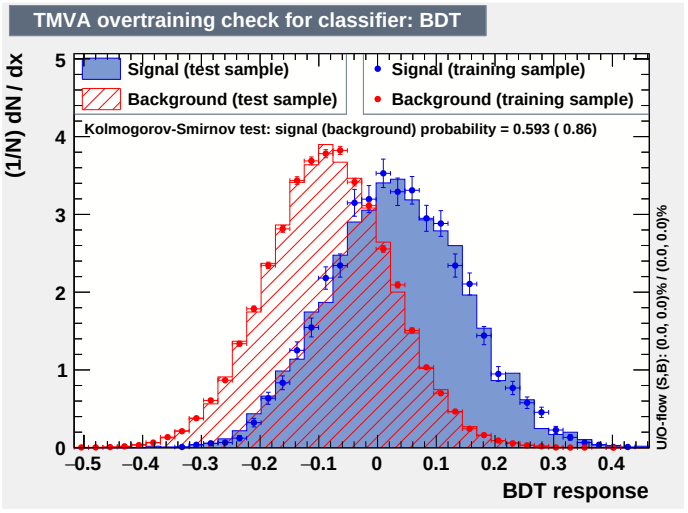


Figure 54: BDT output and overtraining check for the training of tHW versus $t\bar{t}V$ (3 or 4 leptons channel).

First of all, it is checked for overtraining and according to the KS test there is no overtraining. Then, looking at the shape of both distributions, it is found a similar case to the previous one where both of the distribution are a bit overlap but it seems to be possible to obtained a good enough signal/background ratio.

7.5.4 BDT response

Once it has been made the tHW training versus $t\bar{t}V$ for the regions 1 lepton, 2 leptons and 3 or 4 leptons, it has been used the weights of these trainings to obtain the BDT response of the other background processes. The main objective is to achieve a great signal discrimination performance for all the possible backgrounds studied on this project from the $t\bar{t}V$ trainings.

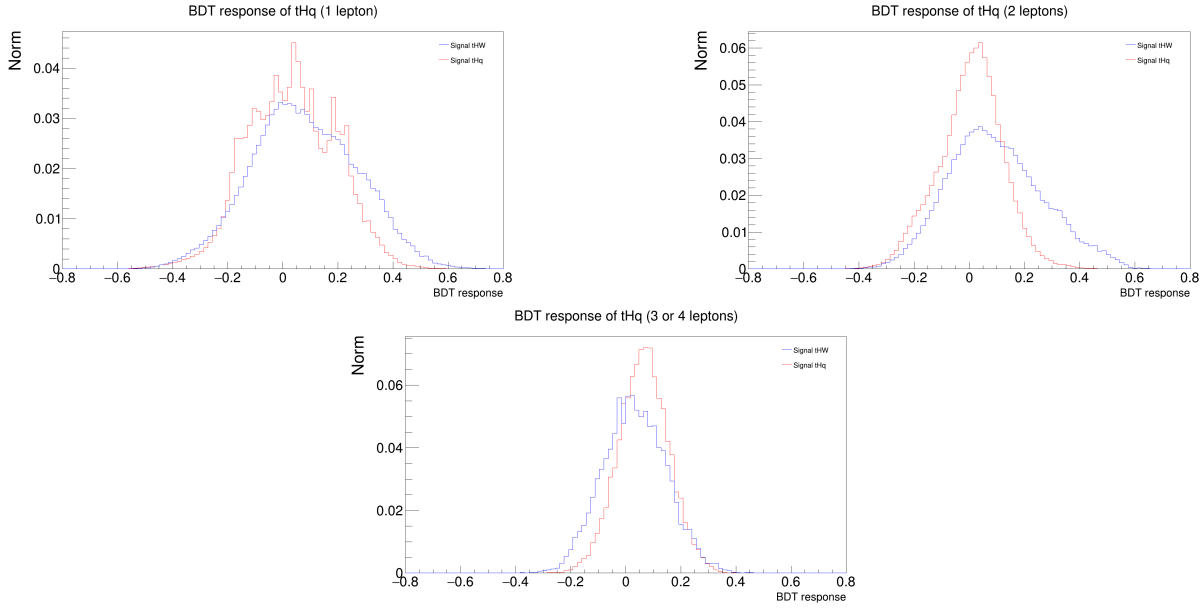


Figure 55: BDT response of tHq process.

First of all, it is obtained the BDT response of the tHq process for the three regions. As tHq is not a background process but is part of the signal (single top and Higgs), both distributions are expected to overlap. This would indicate that this process it is not separated from the signal as if it were a background process. Observing the figure, it can be seen that this is satisfied especially for the region of 1 lepton and the 3 or 4 leptons regions. In the case of the region of 2 leptons, both distributions are slightly different.

Then it has been determined the BDT response for the background processes which are: $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$. They have been grouped by region.

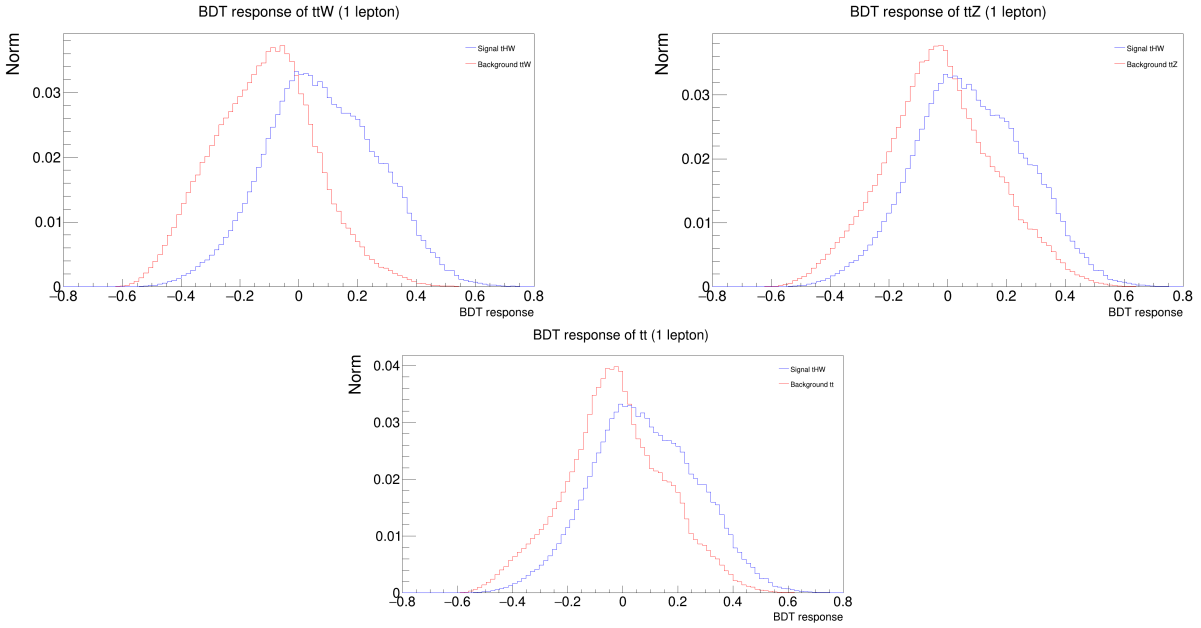


Figure 56: BDT response of $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ processes in the region of 1 lepton.

This plots represent the BDT response for $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$ processes corresponding to the 1 lepton region. In the case of the $t\bar{t}W$ process, it seems to be the best result obtained in this region. Although both distributions are slightly overlapped, there is a range of values for which the signal clearly dominates over the background. In the case of the other two processes, distributions are more overlapped and the signal/background ratio do not seems as good as in $t\bar{t}W$.

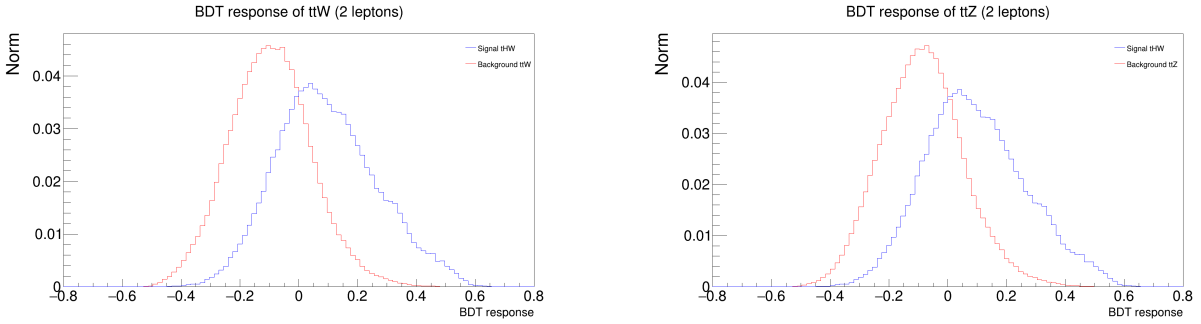


Figure 57: BDT response of $t\bar{t}W$ and $t\bar{t}Z$ processes in the region of 2 leptons.

In this figure, they are shown the BDT response corresponding to the region of 2 lepton. Both of the BDT response are similar but looking closely, some slightly differences can be observed. The signal discrimination for both processes is not bad at all because, although a part of the signal distribution overlaps with the background, there is range of values where the background is almost rejected and there is still a significant amount of signal events.

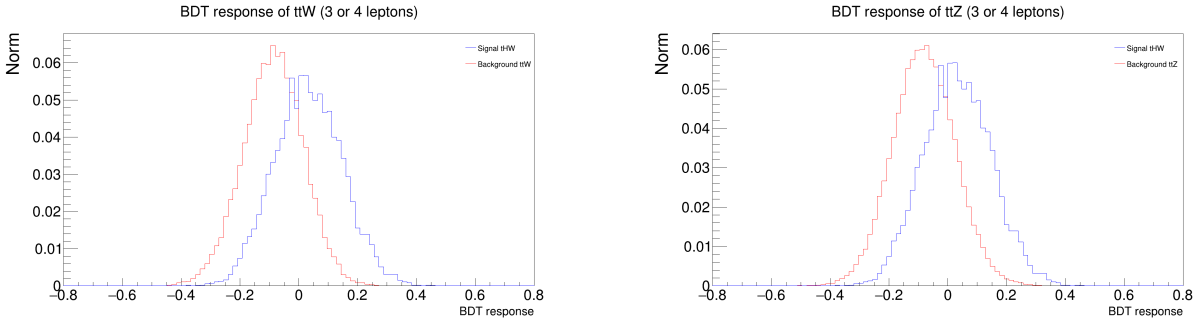


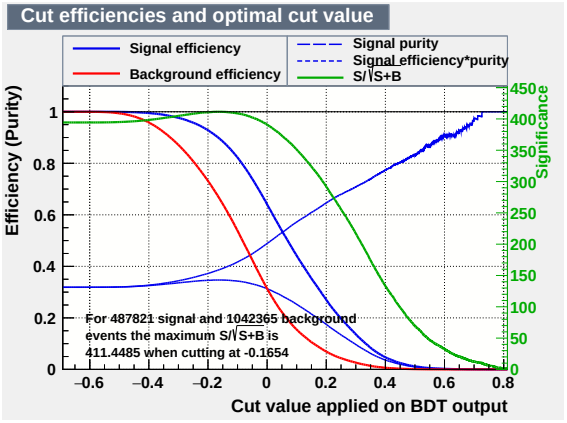
Figure 58: BDT response of $t\bar{t}W$ and $t\bar{t}Z$ processes in the region of 3 or 4 lepton

Then, it has been plotted the BDT response for the region of 3 or 4 leptons ($t\bar{t}W$ and $t\bar{t}Z$ processes). This case is similar the previous one, because there is a portion of the distributions that overlap but both peaks are separated and the shape of the distributions are very similar so it is likely to have a good signal/background ratio at some range which leads to a high efficiency signal discrimination.

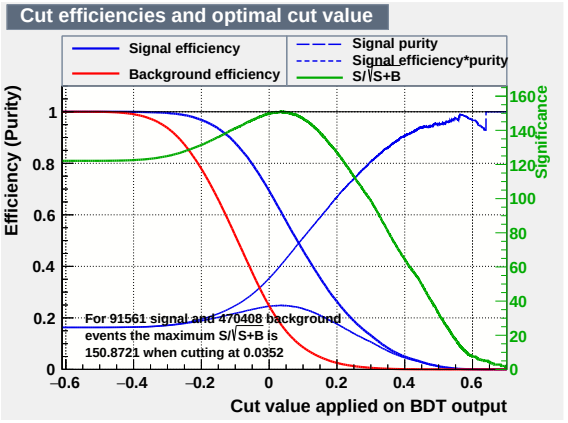
7.5.5 Results

Once it has been obtained the BDT response for all the processes and regions, it has been evaluated quantitatively how efficient the signal discrimination is for the different processes and regions. To do this, it has been applied a cut to the BDT responses in order to obtain the best possible signal discrimination. The way these cuts are selected depends on several factors as the number of events or the goal of the study so in each case will be different.

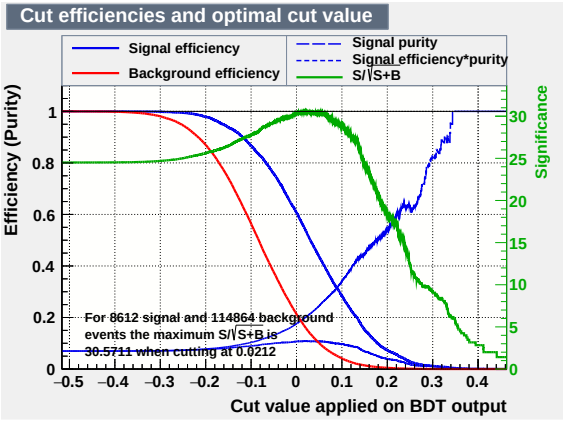
Now, it has been represented a figure where is shown the cut efficiencies in order to obtain an optimal cut value.



(a) 1 lepton region



(b) 2 lepton region



(c) 3 or 4 lepton region

Figure 59: Cut efficiencies and optimal cut value for the tHW training versus $t\bar{t}V$.

In this figure are plotted the signal and background efficiency, as well as, the significance and the signal efficiency purity. Significance is given by

$$\frac{S}{\sqrt{S+B}}$$

Where S are the signal events and B the background events. This variable maximizes the signal events over background. Therefore, significance is used as reference to select the most efficient cut. However, this cut may not be the best for the analysis being carried out. For example, if the process that is studied has many events, perhaps it is preferable to use another cut in which the signal purity is higher.

In the case of the region of 2 leptons and in the 3 or 4 leptons regions, it has been chosen to use the cut given by the significance due to there are not to many events, especially in the signal process. Both cuts have been rounded to 0 and therefore the same cut is applied to both regions. However, in the region of 1 lepton, as there are a large number of events, it is possible to applied a different cut in order to improve the signal purity, because there

are still a considerable number of events left to analyzed. The optimal cut recommended is -0.1654 therefore it has been applied a cut in -0.1.

Using these cuts, it has been calculated the ratio between the number of events of each process after applying the cut and the number of events before. Those values are the selection efficiencies and they have been collected in the table 14.

Regions	tHW [%]	tHq [%]	$t\bar{t}W$ [%]	$t\bar{t}Z$ [%]	$t\bar{t}V$ [%]	$t\bar{t}$ [%]
1 lepton	83.98	80.48	50.34	67.72	56.46	68.99
2 leptons	69.18	54.32	23.55	25.84	24.65	-
3 or 4 leptons	60.72	76.00	19.37	21.86	21.31	-

Table 14: Selection efficiencies obtained for each process after applying the different cuts.

As the goal of the analysis is to separate the signal for the background, it is needed to obtain the highest possible selection efficiencies for signal processes and the lowest possible selection efficiencies for background. Looking at table 14, it can be seen how signal processes (tHW and tHq) has high selection efficiencies meanwhile the selection efficiencies for background are lower. In most of the processes and regions, the difference between the selection efficiencies of signal and background are significant (around 40%). However, in some cases, especially the region of 1 lepton for $t\bar{t}$ and $t\bar{t}Z$, the difference between the selection efficiencies of signal and background is only around 15%.

8 Conclusions

On this project, I studied the production of single top quarks in association with a Higgs boson. This process can occur either in association with a W boson (tHW) or with an additional light jet (tHq). The analysis is focused on the separation of tHW signal events from the most relevant backgrounds ($t\bar{t}V$ and $t\bar{t}$). These results are evaluated also in the other single top associated with a Higgs process (tHq).

In the analysis, it is used MonteCarlo simulations of proton-proton collisions (2017 CMS) corresponding to a center of mass energy of 13 TeV. In addition to analyzing quarks, genjets are also studied which are jets at GEN-level.

On the first part of the project, it is studied the quark-genjet matching efficiency for each channel. These first set of channels were defined attending to the possibles final states of the tHW process (see table 12). It was only considered two of the Higgs decay modes, Higgs boson decaying into two b quarks or into two W bosons. This matching efficiency has been studied depending on the number of leptons, the number of additional genjets or depending on their origin. Then, it has been calculated how often all the quarks are matched to a genjet and their origin matched with each channel. The quark-genjet matching efficiency obtained depending on the number of leptons is adequate since for all 4 different cases, the efficiency is above 76%. Something similar happens for the matching efficiency depending on the number of additional genjets. For more than 2 additional genjets the matching efficiency is still high but for fewer genjets the matching efficiency drops around a 20%. In the case of

the matching efficiency depending on the quark origin, when they come from a top quark or W boson, the matching efficiency is high (around 80%) meanwhile for quarks which origin is a Higgs boson the results are diverse. For channels "2l3q" and "1l5q", related to the H→ $b\bar{b}$ decay mode, the matching efficiency is also high but for the other channels, the matching efficiency decreases significantly. This indicates that genjets produced from the H→ WW decay mode are harder to match. Then, observing the matching efficiency of each channel, unexpectedly low values are obtained. One of the possible reasons is the large number of genjets in some cases and it is also affected by the low matching efficiency of quarks which origin is the Higgs boson (channels related to H→ WW decay mode).

The main objective of the project was the application of machine learning techniques, in this case Boosted Decision Trees (BDT), in order to create a signal discriminator of the single top quark associated to a Higgs boson (tH). In the training, the signal process that has been used is tHW and the background processes are $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$. First of all, it was trained the tHW sample versus each background to check how well the signal discrimination works for each background and region. However, the analysis itself was performed training the tHW process against $t\bar{t}V$ because this background is harder to separate than $t\bar{t}$. Using this training, the BDT response of each process and regions are obtained. I also evaluate the BDT response of this BDT in the tHq process.

The results of the signal discrimination, apart from the several figures presented, are collected in table 14. Looking at the ratio between the number of events of each process after applying the cut and the number of events before, it can be extracted several conclusions. First of all, it can be observed that both signal processes (tHW and tHq) have similar selection efficiencies, therefore with this method the tHq process can be separated from main background ($t\bar{t}V$ and $t\bar{t}$) using the same BDT as tHW. Then, the least efficient signal discrimination corresponds to the 1 lepton region. It can be appreciated that although there is a high selection efficiency (around 80%) of signal events, there is also a high selection efficiency of background events (50-70%). The signal discrimination for the other two regions are high efficient because rejected a considerable amount of background (around 20% left) and keeps a relative high selection efficiency for signal (above 50%). In addition, according to the values obtained, it seems that the signal discrimination for the region of 3 or 4 leptons works better when the signal is tHq and the same happens for tHW and the region of 2 leptons.

It is unexpected that the least efficient signal discrimination corresponds to having $t\bar{t}$ events as background. Although, the training was made versus $t\bar{t}V$ process, it was expected to work well for other background processes, especially for backgrounds that are easier to reject. This low performing signal discrimination related with that process and also with the region of 1 lepton (all processes) may be caused because there are fewer inputs variables for that specific channel, therefore the discriminating power could be lower.

The results obtained show that the discriminating power is reasonable, especially for the regions with more than 1 lepton. However, it may be possible to obtained results with a higher efficiency. In the case of the region of 1 lepton, adding new input variables, as it could be the missing transverse energy, could increase the discriminating power. This variable may be useful because the number of neutrinos could be different between signal and background, especially for $t\bar{t}Z$ due to one of the Z boson decay modes is Z→ $\nu\nu$.

9 Bibliography

[1] *LHC run 3: Physics at record energy starts Tomorrow*. CERN. (n.d.). Retrieved September 21, 2022, from https://home.cern/news/news/physics/lhc-run-3-physics-record-energy-starts-tomorrow

[2] *The Standard Model*. (2022, September 20). CERN. Retrieved September 21, 2022, from https://home.cern/science/physics/standard-model

[3] Thomson, M. (2013). *Modern particle physics*. Cambridge University Press.

[4] Gagnon, P. (2014, March 14). *The Standard Model: a beautiful but flawed theory*. Quantum Diaries. https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/.

[5] Rod Nave, C. (n. d.). *Fundamental Forces*. Hyperphysics. Retrieved August 18, 2022, from http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/funfor.html

[6] *Jets at CMS and the determination of their energy scale*. CMS Experiment. Retrieved September 14, 2022, from https://cms.cern/news/jets-cms-and-determination-their-energy-scale

[7] *Experiments*. (2021, September 8). CERN. https://home.cern/science/experiments

[8] Mobs, E. (2019, July 29). *The CERN accelerator complex - 2019*. CERN Document Server. https://cds.cern.ch/record/2684277.

[9] *Detector — CMS Experiment*. (n.d.). Retrieved September 21, 2022, from https://cms.cern/index.php/detector

[10] *CMS starts underground* . (2022, April 25). CERN Courier. Retrieved September 21, 2022, from https://cerncourier.com/a/cms-starts-underground/

[11] Cid Vidal, X., Cid Manzano, R. (n. d.). *Taking a closer look at LHC - Momentum*. Taking a closer look at LHC. Retrieved August 24, 2022, from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum.

[12] *Tracking — CMS Experiment*. (n.d.). Retrieved September 21, 2022, from https://cms.cern/detector/identifying-tracks

[13] *Energy of Electrons and Photons (ECAL) — CMS Experiment*. (n.d.). Retrieved September 21, 2022, from https://cms.cern/detector/measuring-energy/energy-electrons-and-photons-ecal

[14] Sobrón, M. (n.d.). *Geometría del detector CMS reconstruida con el sistema de alineamiento Link*. [PhD dissertation]. University of Cantabria.

[15] Kuusela, M., Panaretos, V. M. (2014). *Empirical Bayes unfolding of elementary particle spectra at the Large Hadron Collider*. arXiv preprint arXiv:1401.8274.

[16] *CMS precisely measures the mass of the Higgs boson — CMS Experiment*. (n.d.). Retrieved September 21, 2022, from https://cms.cern/news/cms-precisely-measures-mass-higgs-boson

[17] Logan, H. E. (2014). *TASI 2013 lectures on Higgs physics within and beyond the Standard Model*. arXiv preprint arXiv:1406.1786.

[18] *Lecture 17 - The Higgs Boson*. (n.d.). [Slide show]. University of Edinburgh website. https://www2.ph.ed.ac.uk/~playfer/PPlect17.pdf

[19] Ellis, J. (2015, April 27). *A Historical Profile of the Higgs Boson*. CERN Document Server. Retrieved September 21, 2022, from https://cds.cern.ch/record/2012465/plots

[20] Bettoni, D. (n.d.). *Masses and the Higgs Mechanism* [Slide show]. Istituto Nazionale di fisica Nucleare. https://www.fe.infn.it/~bettoni/particelle/Strong/HiggsMechanism.pdf

[21] *Brief introduction to the physics of the Higgs boson* - ATLAS Open Data 13 TeV Documentation. (n.d.). Retrieved September 21, 2022, from http://opendata.atlas.cern/release/2020/documentation/physics/the-higgs-boson.html

[22] Grojean, C. (2017). *Higgs Physics*. arXiv preprint arXiv:1708.00794.

[23] Aad, G., Abbott, B., Abdallah, J., Abdinov, O., Aben, R., Abolins, M., AbouZeid, O. S., Abramowicz, H., Abreu, H., Abreu, R., Abulaiti, Y., Acharya, B. S., Adamczyk, L., Adams, D. L., Adelman, J., Adomeit, S., Adye, T., Affolder, A. A., Agatonovic-Jovin, T., . . . Zwalinski, L. (2016, January). *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at* $\sqrt{s} = 7$ *and 8 TeV in the ATLAS experiment*. The European Physical Journal C, 76(1). https://doi.org/10.1140/epjc/s10052-015-3769-y https://cds.cern.ch/record/2034254/files/arXiv:1507.04548.pdf

[24] Sirunyan, A. M., Tumasyan, A., Adam, W., Ambrogi, F., Asilar, E., Bergauer, T., ... Iaydjiev, P. (2019). *Search for associated production of a Higgs boson and a single top quark in proton-proton collisions at s= 13 TeV*. Physical Review D, 99(9), 092005.

[25] ATLAS , CDF , CMS , DØ Collaboration, "First combination of Tevatron and LHC measurements of the top-quark mass", arXiv:1403.4427.

[26] Aldaya, M. (2022, August 29). *LHCTopWG Summary Plots*. Twiki Cern. ttps://twiki.cern.ch/twiki/pub/LHCPhysics/TopPairCrossectionSqrtsHistory/tt_curve_toplhcwg_jun22.pdf

[27] Czakon, M., Mitov, A. (2013). *NNLO+NNLL top-quark-pair cross sections*. Twiki page. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top_quark_pair_cross_sections_at.

[28] Najafabadi, M. M. (2006, May 12). *Single Top production at LHC*. ArXiv.Org. https://arxiv.org/abs/hep-ex/0605034

[29] Escobar, C. (2017, September 19). *NLO single-top channel cross sections*. Twiki page. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec?sortcol=0;table=1;up=0#sorted_table.