UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Triple Higgs production at LHC using CMS data, decaying into the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state

Mémoire présenté en vue de l'obtention du grade académique de Master [120] en sciences physiques

Auteur : Thomas Mc Cleery
Promoteur : Agni Bethani
Lecteurs : Gwenhaël Wilberts Dewasseige et Andrea Giammanco
École de Physique
Année académique 2024-2025

Acknowledgements

I would first like to thank Andrea Giammanco and Gwenhaël Wilberts Dewasseige for accepting to read this work and to listen to my defense, but also for giving me precious advice.

Secondly, many thanks to Francisco for answering my questions and for providing me with the necessary resources without which I wouldn't have been able to even start.

Finally, my deepest thanks goes to Agni for the great suggestions, the more often than not enlightening conversations and for her continuous support throughout the entire year.

Contents

Chapter 1

Introduction

In 1964, the Brout-Englert-Higgs (BEH) mechanism was proposed and permanently changed the landscape of particle physics. The BEH mechanism is a process that would've happened less than a picosecond after the Big Bang, generating the masses of the now massive elementary particles. This is possible due to the breaking of the electroweak gauge symmetry, implying a radical change in the shape of the potential of the Higgs field, which gives the W and Z bosons their respective masses [1]. The excitations of this field are more commonly known as the newest massive elementary particle, the Higgs boson. Despite the Higgs boson being predicted in 1964, it is only in 2012, at the LHC, that it got officially discovered.

While the Standard Model (SM) has been exceedingly effective at predicting the existence of the Higgs field and its consequences for the Universe, the shape of its potential has not yet been fully established. As of now, the shape is only roughly estimated, as all of the terms of the Higgs potential described in eq. 1.1 [2] are not yet known.

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_3 v H^3 + \frac{1}{4}\lambda_4 H^4 \tag{1.1}$$

The Higgs mass m_H has been measured at 125.38 ± 0.14 GeV by CMS [3], and the vacuum expectation value v at 246.22 GeV [4], but λ_3 and λ_4 have not been measured yet. These parameters are the trilinear (λ_3) and quadrilinear (λ_4) Higgs self-coupling constants, and must be experimentally measured to establish the exact shape of the Higgs potential. Determining its exact shape is of utmost importance due to the large implications it has on the universe, but is also a probe for new beyond the Standard Model (BSM) physics. Indeed, in the SM, λ_3 and λ_4 are equal to each other, which implies that the universe sits at its true minimum as seen in Fig. 1.1. If these couplings are not what is predicted by the SM, this automatically leads to BSM physics and, for some values of λ_3 and λ_4, it could indicate that the universe might actually sit in a false minimum which would render the universe unstable. This is one of the numerous reasons why λ_3 and λ_4 should be measured as precisely as possible.

These couplings quantify the strength of the Higgs self-interaction between 3 and 4 Higgs respectively, and such interactions can be found in double and triple Higgs production events, with examples of such events shown in Fig. 1.2. Therefore, the plan is to detect as many double and triple Higgs events as possible at the LHC to set better and better limits on λ_3 and λ_4. In our case, we will focus solely on triple

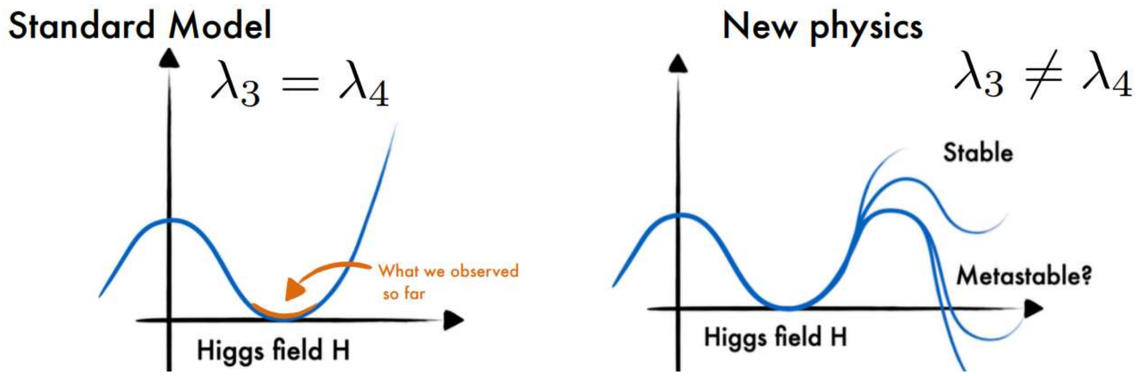
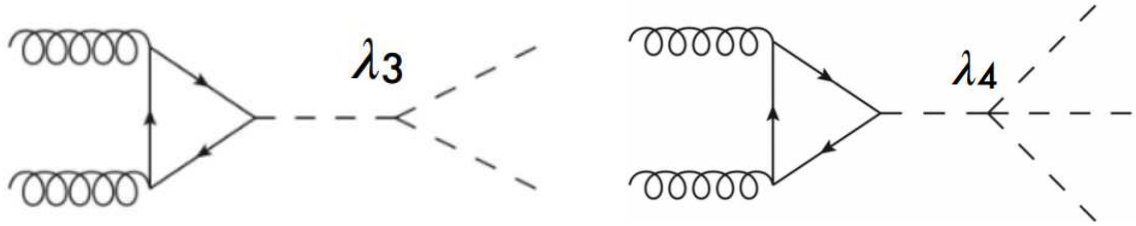


Figure 1.1: Shape of the Higgs potential in the SM (left) and in the case of new physics (right). The universe would be considered unstable if it is able to quantum tunnel to the lower, 'true' minimum. [2]



(a) **(b)**

Figure 1.2: Feynman diagrams for gluon-gluon fusion (a) double and (b) triple Higgs production, with λ_3 and λ_4 the trilinear and quadrilinear Higgs self-couplings respectively. [2]

Higgs production, and more specifically on the HHH $\rightarrow b\bar{b}b\bar{b}\tau^+\tau^-$ process using CMS data at a center-of-mass energy of 13.6 TeV.

Unfortunately, there is a substantial hurdle standing in our way : the extremely low cross section of triple Higgs events. This means that most of the events that are flagged as possible HHH decays in CMS are in fact other processes that decayed to the same final state that we're interested in. The objective of this thesis will then be to see how well we're able to separate our signal (HHH) from the most problematic backgrounds by using machine learning.

Chapter 2 will provide the information that is most relevant to this work about the particles Standard Model, and will also present the necessary information about double and triple Higgs. Chapter 3 introduces the core principles of particle acceleration at the LHC and detection at CMS. It also presents information about b-quark and τ lepton reconstruction due to these particles making up our studied final state. Chapter 4 enters in more detail about the analysis strategy and data samples used in this thesis. Alongside basic event variables, custom event variables are then displayed and discussed, with the process of making those being described in detail. With the objective of separating the HHH signal from other backgrounds, these variables are then used in two different classifiers described in Chapter 5. The resulting outputs of the classifiers are then discussed at length. The thesis is then concluded with a discussion about the possible next steps for this study in Chapter

6. Figures are made by me unless there is a citation at the end of the caption.

Chapter 2

Standard Model

2.1 Particles in the SM

The standard model of particle physics contains 17 elementary particles divided into 4 main categories : leptons, quarks, vector bosons and scalar bosons.

2.1.1 Lepton sector

The lepton sector of the standard model contains 6 elementary particles : 3 leptons and 3 neutrinos. The leptons are 3 negatively charged fermions interacting with the electroweak force, namely the electron (e^-), the muon (μ^-) and the tau (τ^-). The neutrinos are named after the leptons (i.e. the electron (ν_e) neutrino, the muon (ν_μ) neutrino and the tau (ν_τ) neutrino respectively) and also interact through the electroweak force, but they have an extremely low cross section and are considered massless in the SM. It has however been shown that neutrinos do have a mass after all, albeit a very small one. Each lepton forms a doublet with its own neutrino, with these doublets forming the aforementioned lepton sector of the standard model. The τ lepton is the heavier of the 3, having a mass of 1.776 GeV. This particle is very short-lived and can either decay into hadrons, or decay leptonically into either a ν_τ, a ν_e and an electron, either into a ν_τ, a ν_μ and a muon [5]. Muons are much lighter with a mass of 105.7 MeV and can thus only decay to a ν_μ, a ν_e and an electron. It also has a lifetime of 2.2 µs and interacts very lightly when going through matter, meaning that muons can usually travel several meters of matter before decaying [6].

2.1.2 Quark sector

The quark sector of the standard model contains 6 quarks, namely up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Quarks are fermions that interact with the strong force and are one of the two main components of hadrons, the other being the gluon. In the context of this thesis, only the top and bottom quark will be considered, as the four other quarks do not appear in the desired final state due to them coupling extremely weakly to the Higgs boson 2.1, and too light to decay into the b-quarks (4.18 GeV) figuring in our final state unlike the very heavy top quark (172.76 GeV).

Decay channel	Branching ratio	Rel. uncertainty
$H \rightarrow \gamma\gamma$	2.28×10^{-3}	$^{+5.0\%}_{-4.9\%}$
$H \rightarrow ZZ$	2.64×10^{-2}	$^{+4.3\%}_{-4.1\%}$
$H \rightarrow W^+W^-$	2.15×10^{-1}	$^{+4.3\%}_{-4.2\%}$
$H \rightarrow \tau^+\tau^-$	6.32×10^{-2}	$^{+5.7\%}_{-5.7\%}$
$H \rightarrow b\bar{b}$	5.77×10^{-1}	$^{+3.2\%}_{-3.3\%}$
$H \rightarrow Z\gamma$	1.54×10^{-3}	$^{+9.0\%}_{-8.9\%}$
$H \rightarrow \mu^+\mu^-$	2.19×10^{-4}	$^{+6.0\%}_{-5.9\%}$

Figure 2.1: Branching ratios of the Higgs boson. [7]

2.1.3 Vector bosons

Vector, or gauge, bosons are the force carriers of the SM. The strong force is mediated through gluons (g), the electromagnetic force through photons (γ) and the weak force through the W^+, W^- and Z^0 bosons. While the gluons and photons are massless, the $W^\pm$ and Z^0 have a mass of 80.4 and 92.2 GeV respectively. These massive vector boson can constitute possible background for this study since they can decay to the much lighter b-quark.

2.1.4 Higgs boson

The last remaining sector is the scalar boson sector, only containing a single particle, it being the Higgs boson. It is a no charge, spin-0 scalar particle with a mass of around 125 GeV. The Higgs boson is a very unstable particle meaning that only its decay products are detectable, with its most likely decay modes being $H \rightarrow b\bar{b}$ at a rate of 57.7% and $H \rightarrow W^+W^-$ at 21.5% (see Fig.2.1).

2.2 Double Higgs production

Before moving on to triple Higgs production, it is interesting to understand beforehand what are the underlying principles of Higgs pair production. Higgs pair production is a process in which not one, but 2 Higgs bosons are produced in the same event. At the LHC, the 2 main ways of producing a pair of Higgs boson are with gluon-gluon fusion (ggF) or vector boson fusion (VBF). The main ggF production mechanisms involves a top-quark loop coupling with either one or two Higgs bosons, as shown in Fig.2.2. Vector boson fusion Higgs pair production is less common and has a distinctively different signature than its ggF counterpart (this will be shown in Chapter 4). Vector boson fusion happens when two quarks each radiate a vector boson which fuse into a Higgs boson (Fig. 2.3). The quarks that radiated the vector bosons then continue to travel in approximately the same direction before decaying into jets [8], which contributes to the VBF processes having different signatures than ggF processes despite producing two Higgs bosons. The cross section for both processes at 13.6 TeV are $\sigma_{ggFHH} = 34.13\ fb$ and $\sigma_{VBFHH} = 1.87\ fb$.

Having a good knowledge of these double Higgs processes will be of great importance

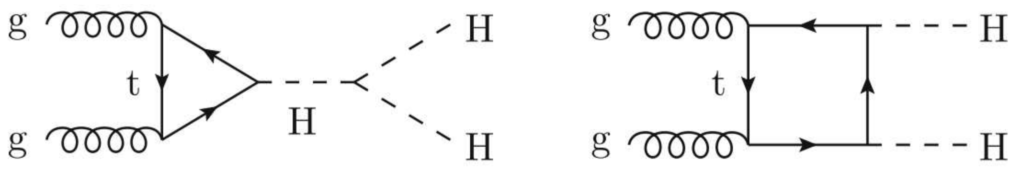


Figure 2.2: Feynman diagrams of the most common gluon-gluon fusion double Higgs production processes. The diagram on the left involves the λ_3 coupling at the vertex between the three Higgs bosons, but the diagram on the right doesn't. This diagram will have to be treated as irreducible background in the search for λ_3. [1]

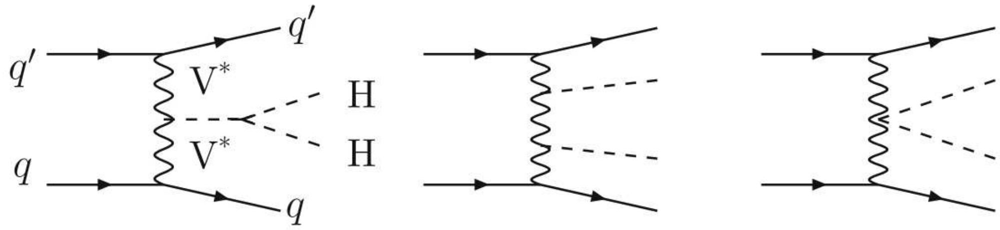


Figure 2.3: Feynman diagrams of the most common vector boson fusion double Higgs production processes. [1]

since they will be the hardest background to separate from our HHH signal due to their Feynman diagrams similarities with the HHH Feynman diagrams.

2.3 Triple Higgs production

Triple Higgs production is similar to double Higgs, but is even rarer with a predicted cross section several orders of magnitude lower than the HH cross section. For this work, we will only use HHH events from the gluon-gluon fusion channel with a cross section of $\sigma_{HHH} = 0.09\ fb$ at a center-of-mass energy of 13.6 TeV. Fig. 2.4 shows the possible Feynman diagrams for ggF HHH in the SM, assuming $h_i = h_j = H$. An important property of these ggF channels is that their amplitudes interfere with each other, and a contribution from BSM processes might alter the interferences which could have major consequences on the cross sections of HHH, making its study very sensitive to BSM physics [1] [9]. The same also applies to HH and the diagrams in Fig. 2.2.

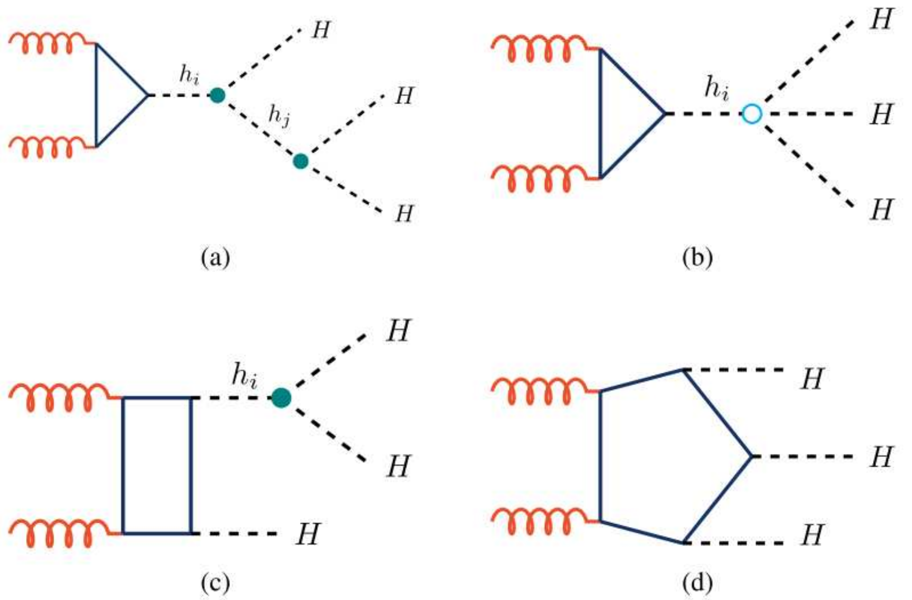


Figure 2.4: Feynman diagrams for gluon-gluon fusion triple Higgs production processes. By setting $h_i = h_j = H$, we get the SM Feynman diagrams and the solid and open circles are the λ_3 and λ_4 coupling constants. If they are not the same, then we enter the realm of BSM physics because the existence of h_i and h_j implies the existence of an extended scalar sector. [9]

Chapter 3

Detecting and reconstructing events in CMS at the LHC

3.1 The Large Hadron Collider

CERN's (Conseil Européen pour la Recherche Nucléaire) Large Hadron Collider, or LHC, is the largest particle accelerator in the world. It is a 27-kilometre ring around which four subatomic particle detectors were built : ATLAS, ALICE, LHCb and CMS (Fig. 3.1) [10]. The entire apparatus has been placed underground near Geneva, directly below the border between France and Switzerland. As the name suggests, the LHC accelerates hadrons, in our case protons, sending them in opposite directions in an attempt to make them collide inside one of the four detectors.

While many parts are involved in a proton's acceleration process, a quick rundown of it is enough to grasp the basics of particle acceleration at CERN. Protons are first accelerated in the Proton Synchrotron Booster (PSB) up to an energy of 1.4 GeV, then in the Proton Synchrotron (PS) to 25 GeV, they are transferred subsequently to the Super Proton Synchrotron (SPS) that accelerates them up to 450 GeV before finally being released in the LHC to be accelerated up to 14 TeV [1]. Those protons then collide into each other in one of the four specific points. Data from these collisions is then collected by the detectors with the objective to reconstruct from the detected final states, the physical processes that occurred in the collisions. All of the data used in this thesis will be data produced by proton-proton collisions and collected by the CMS experiment.

3.2 Compact Muon Solenoid

The Compact Muon Solenoid, or CMS, is a 14 000 ton detector, designed to detect particles produced in the collisions. Outgoing particles first go through a silicon tracker where charged particles, namely electrons, muons and charged hadrons, leave a trace of their passage in the form of a curved track. This curvature of the track is due to a powerful magnetic field of 3.8 T, produced by a superconducting solenoid. Other particles, like photons and neutral hadrons, will go through this first layer undetected due to their absence of charge [6].

The second layer of the detector is the Electromagnetic Calorimeter (ECAL), which measures the energy of the outgoing photons and electrons. Hadrons will pass

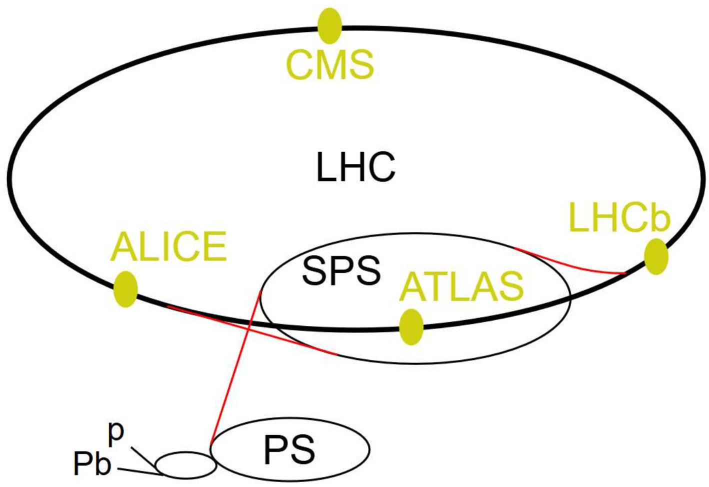


Figure 3.1: Simplified version of the LHC's particle acceleration complex. Protons (p) and sometimes lead ions (Pb) have to first go through different accelerators before reaching the LHC and colliding inside one of the four detectors. [11]

through the ECAL to instead have their energies measured in the third layer of CMS : the Hadron Calorimeter (HCAL). The HCAL will additionally measure the position and arrival time of the hadrons. The HCAL is then itself surrounded by the fourth layer, consisting of the aforementioned superconducting solenoid that creates the 3.8 T magnetic field.

The only remaining particles that go undetected throughout those layers are the muons and the neutrinos. The unique set of properties of the muon allows it to go through the ECAL unscathed, hence the need for a fifth and ultimate layer able to measure the momentum and track the trajectories of muons. This layer is made up of 4 muon stations, themselves each containing hundreds of muon chambers that are able to detect muons passing through them. By checking which chambers are activated, the trajectory of a muon can therefore be reconstructed, as we can see in Figure 3.2.

Despite this dense instrumentation layering, there is still no way to directly detect neutrinos inside CMS as it would require a totally different type of detection process due to their very low cross section. However, it is possible to indirectly quantify the presence of neutrinos in an event by calculating the Missing Transverse Energy (MET) with the help of the measurements provided by the different layers of CMS. This quantity is measured by noticing an imbalance in the kinematics of the collision (i.e. too much transverse momentum on one side and not enough on the other), indicating the production very lightly interacting particles that leave CMS undetected such as neutrinos.

All of this instrumentation allows for extreme precision in measurements of high-energy collider physics by covering a wide range of final states; ideal for the discovery and subsequent study of particles such as the Higgs boson.

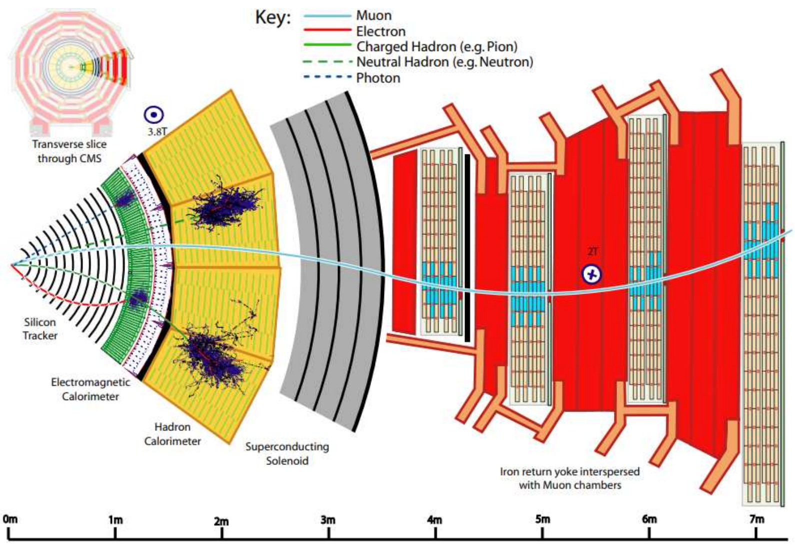


Figure 3.2: Transverse slice of the layers making up the CMS detector. [12]

3.3 Kinematics of collisions inside CMS

The next step after detection is event reconstruction, and understanding the collision's kinematics is therefore paramount. Hadrons entering CMS follow the longitudinal z-axis and collide with the hadrons going in the other direction at the Impact Point (IP), which is the spot on the longitudinal axis around which the CMS reference frame is built around (Fig. 3.3). From there, the outgoing particles produced in the collision travel in any direction and get detected by CMS once they reach the tracker [13].

The main variables taken into account during the reconstruction of a particle are its polar and azimuthal angles $0 < \theta < \pi$ and $0 < \phi < 2\pi$, its pseudorapidity $-\infty < \eta < \infty$, and its transverse momentum $\vec{p_T}$. The angle θ is the angle between the z-axis and the direction of the particle's momentum, which is analogous to its trajectory. The azimuthal angle ϕ is the angle between the transverse momentum $\vec{p_T}$ and the x-axis in the transverse plane. The pseudorapidity η is a less straightforward variable, but remains one of the most important ones in event reconstruction nonetheless. We can see on Fig. 3.3 that η is the angle between $\vec{p}$ and $\vec{p_T}$, although it is not very intuitive geometrically. The mathematical definition of η gives however a much better visualization of the variable :

$$\eta \equiv -\ln\,[\tan(\frac{\theta}{2})] \tag{3.1}$$

Indeed, by looking at Eq.(3.1), we see that for $\theta = \frac{\pi}{2}$, we get $\eta = 0$, indicating that the particle traveled in the transverse plane. For $\theta = 0$ or π, η tends towards infinity. In other words, the pseudorapidity for a particle traveling in the $\hat{z}$ direction will be $\pm\infty$, and it will go undetected since the CMS cannot cover the area of arrival of the hadrons as they cannot go through matter (these are the 2 CMS blindspots). CMS can detect nearly all of the possible trajectories, up to an angle of $|\eta| < 2.5$. In all, the pseudorapidity can be considered as a more convenient way to express θ in accelerator physics.

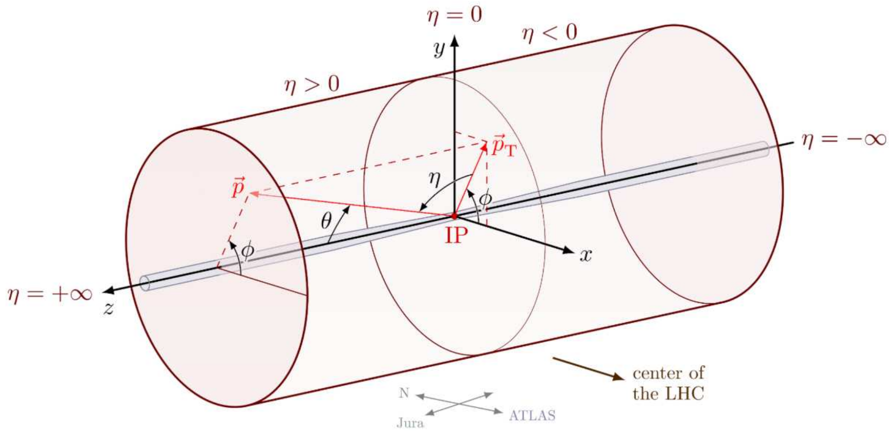


Figure 3.3: Kinematics of a collision inside of CMS. The $\hat{z}$ axis is in fact slightly curved due to it being a small section of the LHC's ring, but is approximated to a straight line in this case. [13]

The transverse momentum $\vec{p_T}$ is the transverse component of the total momentum $\vec{p}$. Transverse momentum is preferred in collider physics because of the nature of the accelerators themselves. Due to the hadrons being sent along the $\hat{z}$ axis, they have a non-zero longitudinal momentum component. Thus, transverse momentum will always be related to the physics of the collision since the hadrons moving along $\hat{z}$ do not have a transverse component to their momentum.

Lastly, we introduce here two variables that are a bit more complex, but will be useful for the remainder of this work : the angular distance ΔR and the invariant mass. Angular distance between two tracks is defined as $\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$, and can be seen as the distance separating two tracks, or 'how much in the same direction they are traveling'. The invariant mass m_0 of a particle is simply its mass when at rest, as defined in the well-known relation $E^2 = p^2c^2 + m_0^2c^4$.

3.4 *b*-tagging

Determining if a jet was produced by the decay of a *b*-quark, more commonly referred to as *b*-tagging, is an essential reconstruction method in high-energy colliders, especially in the realm of Higgs physics. *B*-jets have two main characteristics that are taken advantage of when in the process of *b*-tagging. Firstly, the *b*-quark has a long enough lifetime that it can travel a short distance before decaying [14]. CMS is precise enough to detect if a jet comes from the initial collision (primary vertex), or if it comes instead from a slightly more off-centered secondary vertex (Fig.3.4). If the latter is true, there are strong reasons to believe that this secondary vertex corresponds to a particle with a short (but not too short) lifetime such as the *b*-quark. Secondly, with a mass of 4.18 GeV and only being able to decay into much lighter particles, these particles will thus have high momentum on average. This translates to larger and more energetic jets than most others.

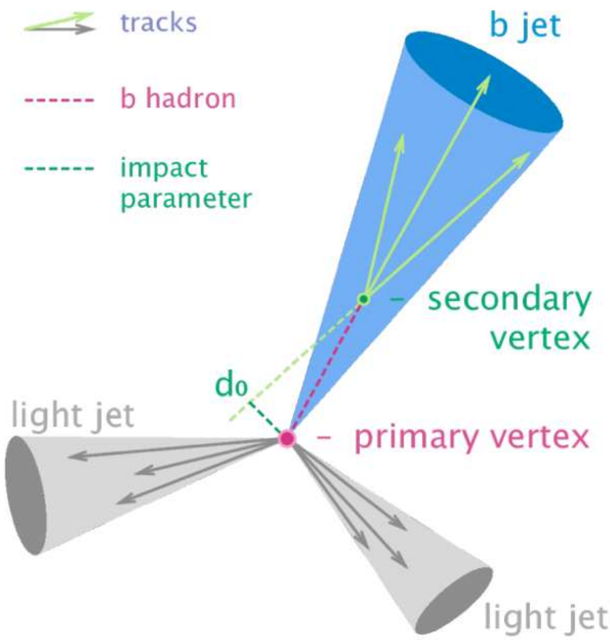


Figure 3.4: Representation of a b-jet. The b-quark created in the primary vertex travels a short distance before decaying at the secondary vertex, into a characteristically energetic jet. [14]

3.5 Hadronic τ tagging

While approximately one third of τ leptons decay leptonically to either a muon or an electron (Fig.3.5) which are then respectively detected by the muon chambers and the ECAL, the remainder of them decay hadronically (we will refer to them as τ_h). This decay mode proves to be a more challenging decay to reconstruct than the leptonic one due to the large amounts of QCD jets that are usually produced in collisions, as they create substantial background [5].

To recontruct τ_h, we must take advantage of the physical nature of the decay. The main feature of τ_h decay is its tendency to produce 'thin' jets and local deposits in the ECAL and HCAL. Additionally, they decay into exactly one, three or five charged particles, and their lifetime is long enough for CMS to detect a displacement (similarly to the b-quark) [15]. The τ_h identification of the τ data samples in this thesis is performed by a deep neural network called DEEPTAU. It is a recently developed algorithm that encompasses electron, muon and jet rejection, whereas previously this would be done by separate algorithms. DEEPTAU is also able to use lower-level information with greater success than its predecessors, since jet decays provide complex enough patterns for a good algorithm to detect the differences between them [5].

Decay mode	Resonance	$\mathcal{B}$ (%)	
Leptonic decays		35.2	
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$			17.8
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$			17.4
Hadronic decays		64.8	
$\tau^- \to h^- \nu_\tau$			11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$		25.9
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$		9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$		9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$			4.8
Other			3.3

Figure 3.5: List of all possible decay modes for the τ lepton. [15]

Chapter 4

Data Analysis

4.1 Analysis strategy

In the objective of being able to separate the HHH signal from other backgrounds, points of comparison between the different types of processes leading to the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state are needed to do so. For this thesis, these points of comparison will be the variables recorded for each event. While the variables in the samples are a good start, we will need better ones to differentiate the processes from one another. Then, with all the variables, we train two different classifiers to separate the signal from the background for the three most problematic backgrounds : top-quark pair ($t\bar{t}$), vector boson fusion double Higgs (VBF HH) and gluon-gluon fusion double Higgs (ggF HH).

The reasoning behind the study of the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state is to strike a balance between the amount of signal and background. The most common decay of the Higgs is to a pair of b-quarks, but these b-quarks also tend to suffer from a large number of background events. Conversely, the H $\rightarrow \tau^+\tau^-$ decay produces a much cleaner signature but is also less common. To visualize the likelihood of the possible triple Higgs final states, we can take a look at their branching ratios in Figure 4.1, assuming that the first Higgs decays to a pair of b-quarks. The H $\rightarrow \tau^+\tau^-$ and H $\rightarrow \gamma\gamma$ decay modes provide the cleanest signatures, but they are not produced in abundance. So, if we are to study a final state containing those low-background decay modes, then the main strategy is to only have one of the Higgs decaying to these modes to ensure a high enough signal.

Due to these reasons, we choose to search for the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state, with the additional condition of having one of the τ leptons decaying hadronically and the other one decaying to an electron.

4.2 Data samples

This study uses data from 3 different data sample categories : signal, background and Data (This word will be capitalized to avoid confusion with the general term 'data'). Their cross sections correspond to their SM cross sections for a center-of-mass energy of 13.6 TeV. The signal sample contains Monte Carlo simulated HHH events produced by gluon-gluon fusion and decaying to $4b2\tau$, the Data samples contain events from LHC runs of the latter part of 2022 that were flagged as decaying to

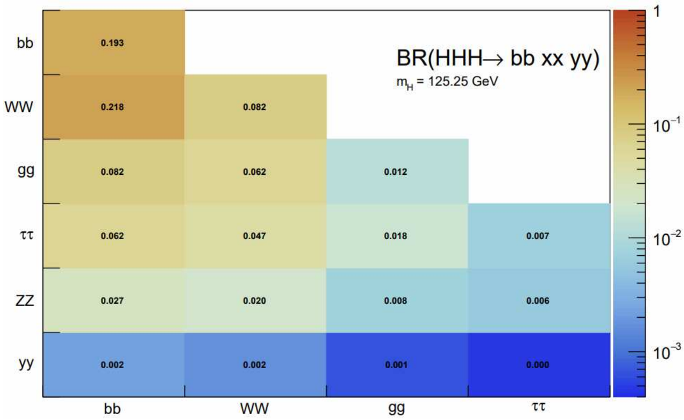


Figure 4.1: Branching ratios for the HHH $\rightarrow b\bar{b}b\bar{b}\tau^+\tau^-$ final states. (Here, 'yy' $=$ $\gamma\gamma$). [16]

the bbbb$\tau^+\tau^-$ final state. And for background, there are 10 MC-simulated samples decaying to the desired final state, namely ggF HH, VBFH HH, single Higgs, $t\bar{t}$, single top, Drell-Yan, single $W^{\pm}$, W^+W^-, $W^{\pm}Z$ and double Z. It is important to note that the Higgs bosons in both HH samples decay specifically into the $b\bar{b}\tau^+\tau^-$ final state, as this will be important when looking for differences in kinematics to separate HHH from HH.

For this thesis, all of the data comes in the form of ROOT trees. ROOT is a data analysis framework, mostly used in the domain of high energy physics, that is capable of dealing with very high amounts of data [17]. A ROOT tree is a data storage structure that to store data very conveniently. For each process, a ROOT tree is created with a number of 'leaves' that each contain the data of all events for a single variable. Thanks to this, it is easy to access the variables individually and iteratively. Each ROOT tree contains the following variables : p_t and its components p_x, p_y, p_z, the η, θ and ϕ angles, and the energy.

4.3 Pre-selection cuts

Pre-selection cuts were made on the events in the data samples to simplify the analysis process. The most important cuts are as follows :

- τ lepton $p_T > 30$ GeV and $\eta < 2.3$
- $\Delta R > 0.4$ between both τ
- Electron $p_T > 20$ GeV and $\eta < 2.1$
- Maximum 1 electron and 0 muon
- τ_e has to pass at least one of two High-Level Triggers (HLT)
- AK4 jets $p_T > 20$ GeV and $\eta < 2.5$

4.4 Pairing of the b-jets

Selecting the events corresponding to the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state leaves us with 2 τ leptons, 4 b-tagged jets and other backgrounds such as $t\bar{t}$. We can easily reconstruct the first Higgs boson with the 2 τ leptons, but this cannot be done in such a straightforward way for the 4 b-tagged jets. Indeed, if those b-jets were produced by a pair of Higgs bosons, it is a non-trivial task to pair a b-tagged jet to the other b-jet produced by the decay of the same Higgs boson.

By convention, the four b-jets will be named j_1, j_2, j_3 and j_4 with j_1 being the most energetic jet, j2 being the second most energetic jet and so on. We can see that 3 possible pairing situations are possible : j_1-j_2/j_3-j_4, j_1-j_3/j_2-j_4 and j_1-j_4/j_2-j_3. With the goal of correctly pairing the b-jets, we will try 3 different pairing algorithms and select the best performing of those to then use for the entirety of this study.

4.4.1 Algorithms

The first method we use is to choose, out of the 3 combinations, the pairings for which the sum of the difference between the Higgs boson mass (125 GeV) and the invariant mass of the di-b-jets is minimal :

$$\Delta_{alg} = |M_1 - 125| + |M_2 - 125| \tag{4.1}$$

where M_1 and M_2 are the invariant masses of the di-b-jets. Hence, the main idea behind this method that we will now refer to as 'simple', is to have the invariant mass of both di-b-jets as close as possible to the mass of the Higgs boson.

The second algorithm uses a χ^2 method [18], which is very similar to the 'simple' method, as shown by the following equation :

$$\Delta_{alg} = (M_1 - 125)^2 + (M_2 - 125)^2 \tag{4.2}$$

We do the same here as with the previous method by selecting the smallest of the three possible Δ_{alg}. The difference between the first two algorithms can be seen mostly in Figure 4.2 where the invariant mass distribution of the highest energy di-b-jet for the 'simple' algorithm seems to be slightly denser around its maximum than 'χ^2' algorithm.

The third and last method we use is a 'closest to diagonal' method, which will refer to as the 'DHH' method. First, we define a point for each pairing combination : A = [M(j_1,j_2);(j_3,j_4)], B = [M(j_1,j_3);(j_2,j_4)] and C = [M(j_1,j_4);(j_2,j_3)] with M(j_m,j_n) the invariant mass of the (j_m,j_n) di-b-jet. Then, we select the closest one to the diagonal in the mass plane, as shown if Fig. 4.3, where the distance is measured by eq. 4.3, with M(H_1) the most massive component of A, B and C. The value of $\frac{X_0}{Y_0}$ was empirically defined at 1.04 for the HH $\rightarrow b\bar{b}b\bar{b}$ decay [19], so we also choose this value since we are looking at this same decay in terms of b-jets. The pairing that is closest to the diagonal, i.e. that has the lowest D_{HH}, is the pairing that the algorithm will select.

$$D_{HH} = \frac{|M(H_1) - \frac{X_0}{Y_0} M(H_2)|}{\sqrt{1 + (\frac{X_0}{Y_0})^2}} \tag{4.3}$$

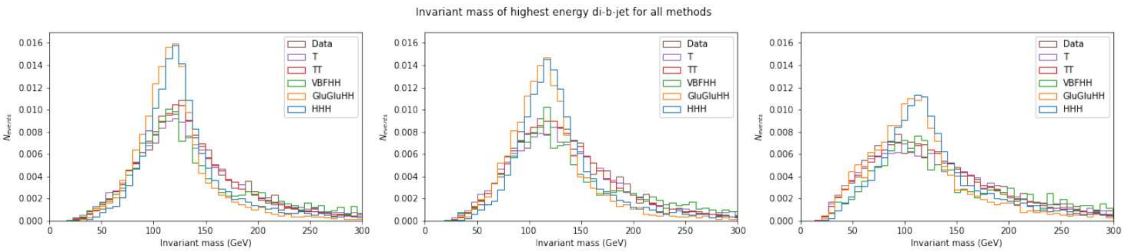


Figure 4.2: Plots of the invariant mass of the highest energy di-b-jet for all 3 algorithms. From left to right : 'simple' method, 'χ^2' method and 'DHH' method. Labels from top to bottom : Data, single top, $t\bar{t}$, vector boson fusion HH, gluon-gluon fusion HH, and triple Higgs.

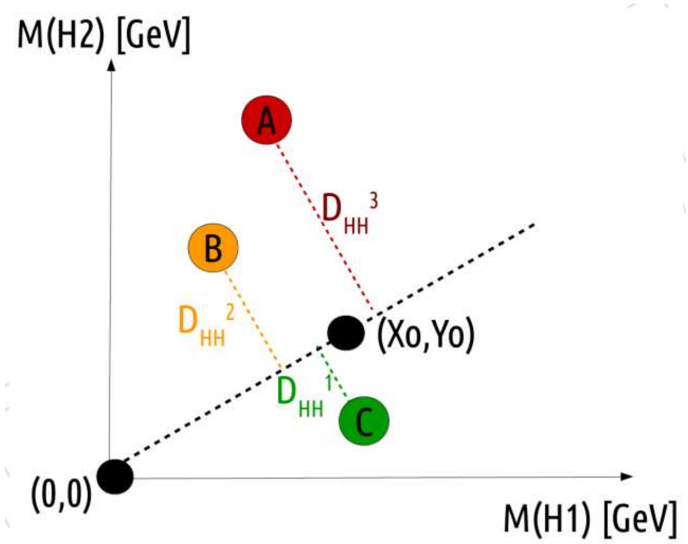


Figure 4.3: Mass plane with M(H_1) the highest energy di-b-jet. In this example the pairing C will be selected by the D_{HH} method. [19]

4.4.2 Algorithm selection

To select the most effective method among the three algorithms, we compare b-jet kinematic variables from the HHH MC-simulated data to those same variables from the 'truth information'. This information corresponds to the exact values simulated by the Monte-Carlo simulation, meaning that we know which b-jet belongs to which Higgs boson. Although they both give the information about the b-jets that were simulated, truth information differs from the MC-simulated data by providing the correct values. The MC-simulated data instead provides the values that we get after applying the CMS reconstruction algorithms, to more accurately match reality. Thus, by comparing the truth information to the MC-simulated values we are working with, we can estimate the correct pairing combination with great precision.

To check if a pairing of 4 'MC' b-jets matches the pairing given by the truth information, we compute the ΔR between each 'MC' b-jet and its 'truth' b-jet counterpart. The larger ΔR is, the less likely that specific b-jet pairing turns out to be correct.

Thus, in the aim of selecting the closest match between the 4 'truth' b-jets and the 4 'MC' b-jets, we need to compare 24 different combinations and select the one with the smallest total ΔR for each event.

Example : Combination 2 is the more likely choice than Combination 1, as we see the 'MC' b-jets (j_1, j_2, j_3 and j_4) being closer on average to the 'truth' b-jets (T_1, T_2, T_3 and T_4).

$$\textit{Combination 1 :}\ \Delta R(T_1, j_2) + \Delta R(T_2, j_4) + \Delta R(T_3, j_3) + \Delta R(T_4, j_1) = 5.34$$
$$\textit{Combination 2 :}\ \Delta R(T_1, j_3) + \Delta R(T_2, j_2) + \Delta R(T_3, j_1) + \Delta R(T_4, j_4) = 2.17$$

We can now compute the efficiency ratio ϵ_{alg} of each algorithm, with N the number of events and N_{corr} the number of events that match the correct pairing combination.

$$\epsilon_{alg} = \frac{N_{corr}}{N} \tag{4.4}$$

The efficiency ratio for the 'simple', 'χ^2' and 'DHH' algorithms are 52.77 %, 52.12 % and 45.99 % respectively. Following these results, the 'simple' algorithm is used for b-jet pairing in this study.

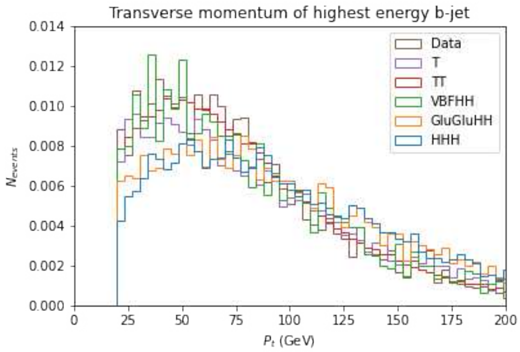
4.5 Variables

4.5.1 Kinematic variables

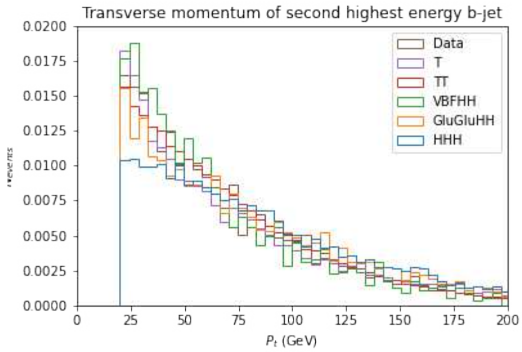
We first take a look at how some of the most relevant kinematic variables mentioned in Chapter 4.2 are distributed. The selected variables are : $\vec{p_T}$, η and θ for the two τ leptons, the four b-jets and the missing transverse energy. Other variables such as ϕ or p_x, p_y and p_z, are not interesting enough to look at due to their very similar shape, independent of the process they represent. The labels for the various distributions in the figures from Fig. 4.4 to Fig. 4.16 are : Data (Data), T (single top), TT ($t\bar{t}$), VBFHH (vector boson fusion double Higgs), GluGluHH (gluon-gluon fusion double Higgs) and HHH (triple Higgs).

The variables surrounding the four b-jets are expected to be relatively impactful in the search for triple Higgs compared to other backgrounds. Indeed, the situation is such that all four b-jets result from Higgs decay in the case of a HHH event. This is however not always true for the background events. For example, double Higgs events might have its first Higgs decaying to a pair of b-quarks and the other one decaying to a pair of τ leptons. We know this to be the case for the double Higgs processes (see Chapter 4.2). This means that the two remaining b-jets were produced in some other process, implying that their variable profile will be different than the b-jets produced in the decay of the Higgs boson. This can be seen across all three aforementioned variables, starting with the transverse momentum. By looking at Fig. 4.4, we observe that the $\vec{p_T}$ is more evenly spread and is, on average, higher for HHH than for the other processes (except for Fig. 4.4d). The Higgs boson being the second heaviest elementary particle (125 GeV) that we know of, it is then expected to produce higher energy (and therefore $\vec{p_T}$) b-jets than other particles on average.

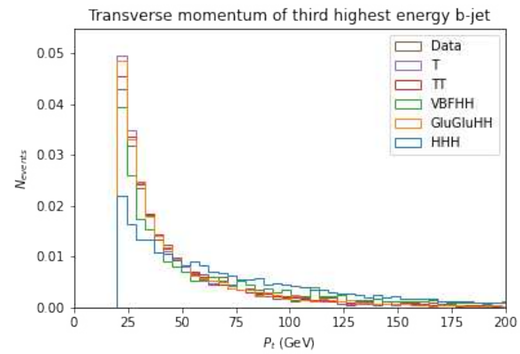
Unlike with the transverse momentum, we can see notable differences in pseudorapidity between HHH and HH (mostly VBF HH in this case). By looking at Fig. 4.5,



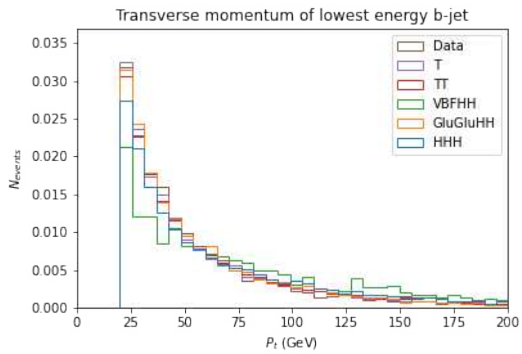
(a)



(b)

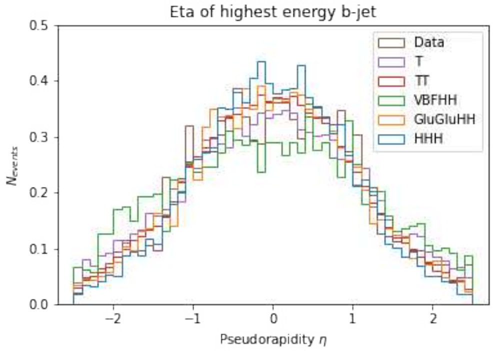


(c)

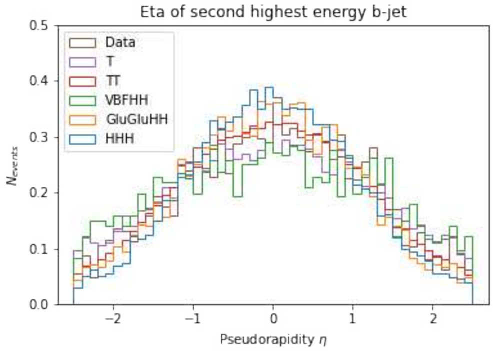


(d)

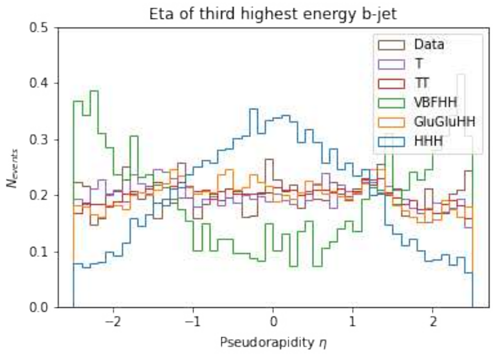
Figure 4.4: Normalized distributions of the transverse momentum of all four b-jets, for HHH signal, relevant backgrounds and data.



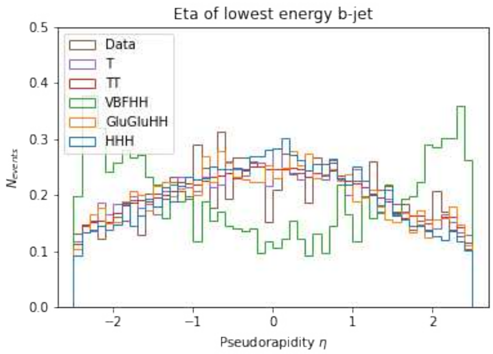
(a)



(b)



(c)



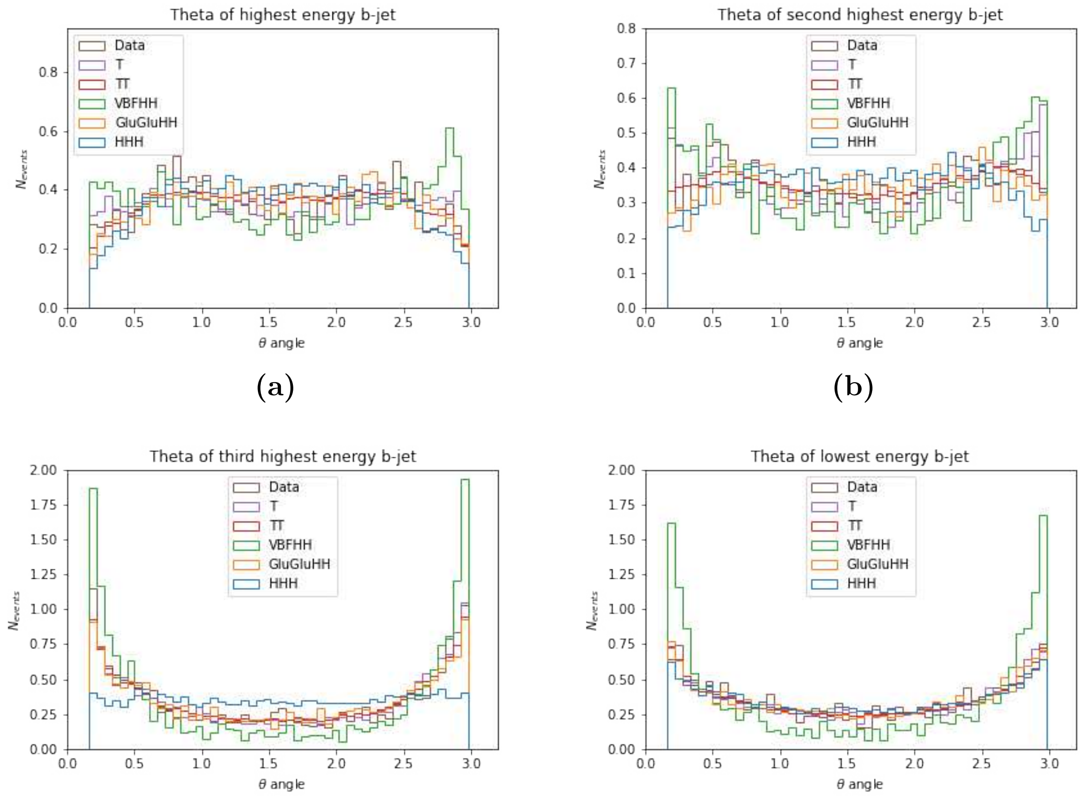
(d)

Figure 4.5: Normalized distributions of the pseudorapidity of all four b-jets, for HHH signal, relevant backgrounds and data.

we see that the η for the b-jets displays very distinguishable pattern for the two lowest energy jets, but similar shapes for the two highest energy jets. In the case of the HH samples, this can once again be explained by the fact that both are MC-generated samples in which the HH specifically decayed to the $b\bar{b}\tau^+\tau^-$ final state, and those b-jets are likely to be the most energetic of the four as mentioned earlier. Turning our attention now to Figures 4.5c and 4.5d, it is clear that the VBF HH exhibits a completely different behaviour, by having its two lower energy b-tagged jets decay much more often in the vicinity of the $\hat{z}$ direction than the other processes. These b-tagged jets are most likely the signature extra jets produced by vector boson fusion. Fig. 4.5c is particularly interesting due to the shape of the HHH being easily discernible from the others, with the b-jet having a tendency to decay close to the transverse plane. As for θ, due to its dependence on η and vice-versa, it is therefore unsurprising to see that the distributions of θ in Fig. 4.6 produce similar results to the distributions of η on Fig. 4.5.

Results for the τ leptons are a bit more mixed. Similarly to the b-jets, the $|\vec{p_T}|$ for the electronic τ_e and the hadronic τ_h all follow the same trend, but are quite different in the low $|\vec{p_T}|$ region, as seen in Fig. 4.7. For η however, the distributions for VBF HH now show no clear separation from HHH (Fig. 4.8). This is because its τ leptons were also produced by a Higgs decay.

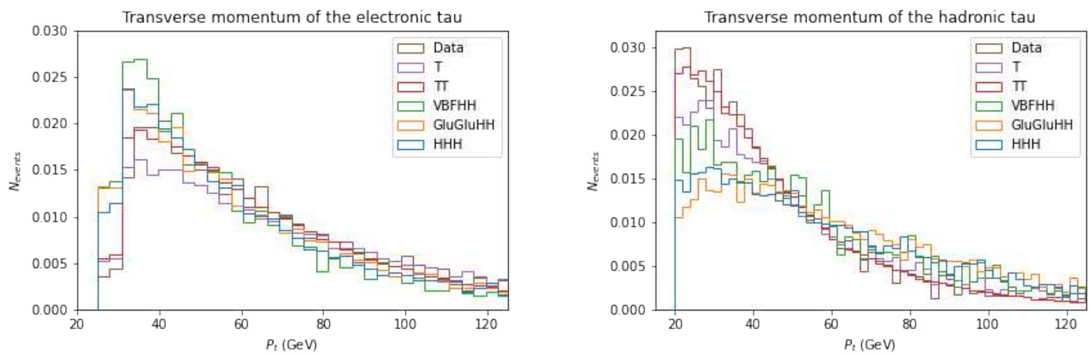
Lastly, missing transverse energy must not be neglected, as it is a key component of the final state, even if not directly mentioned. Like its name indicates, MET is



(a) (b)

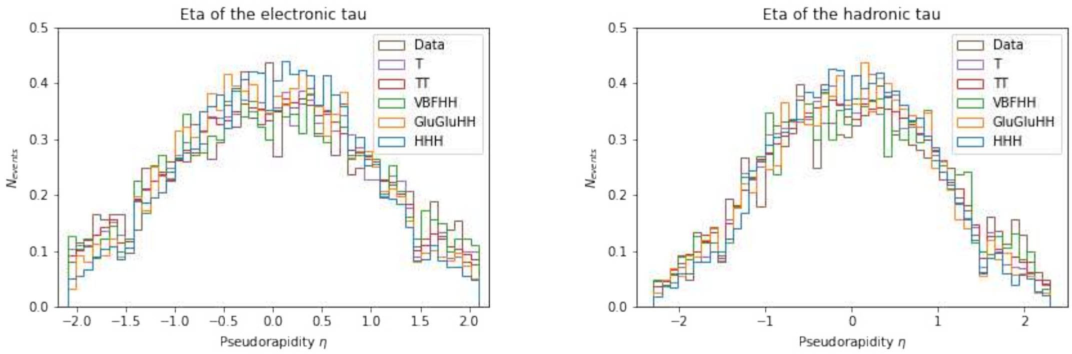
(c) (d)

Figure 4.6: Normalized distributions of the angle θ of all four b-jets, for HHH signal, relevant backgrounds and data.



(a) (b)

Figure 4.7: Normalized distributions of the transverse momentum of the (a) leptonic τ and (b) the hadronic τ, for HHH signal, relevant backgrounds and data.



(a) (b)

Figure 4.8: Normalized distributions of the pseudorapidity of the (a) leptonic τ and (b) the hadronic τ, for HHH signal, relevant backgrounds and data.

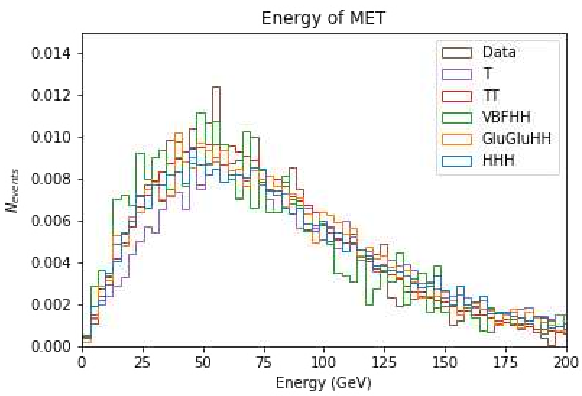


Figure 4.9: Normalized energy distribution of the missing transverse energy. Since neutrinos are almost massless and there is no longitudinal component, MET is nothing more than missing $\vec{p_T}$.

calculated in the transverse plane, implying that η and θ are both zero and that energy would therefore be its best discriminator. Unfortunately, as seen on Fig. 4.9, no distribution seems to exhibit a noteworthy shape.

4.5.2 Event variables

While some of the initial kinematic variables show distinctive shapes and will hence be useful in data classification, most of them do look quite similar across all processes. A sensible strategy is to then create other variables in order to improve the performance of the classifiers, which will be introduced in Chapter 5. Thus, we create and add an additional 10 new variables. Five of them use invariant mass and the other five use the ΔR quantity, both of which are described in Chapter 3.3.

To calculate the invariant mass for the b-jets, τ leptons and MET, we must first recreate their Lorentz vectors. To do this, we use the 'TLorentzVector' class in ROOT, which creates Lorentz vectors if enough information is given for it to be able to do so. Thus, by inputting the p_x, p_y, p_z and E variables, the class creates the Lorentz vectors (or 4-momentum) for any desired track. With the additional

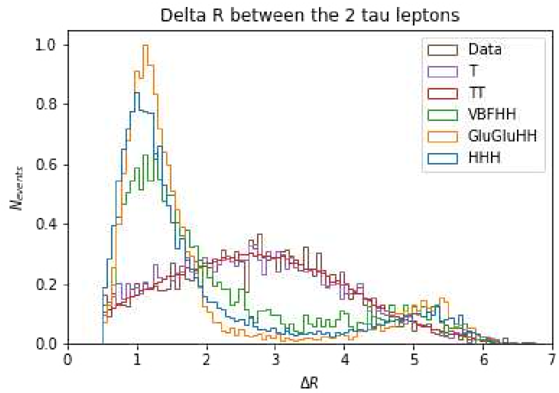


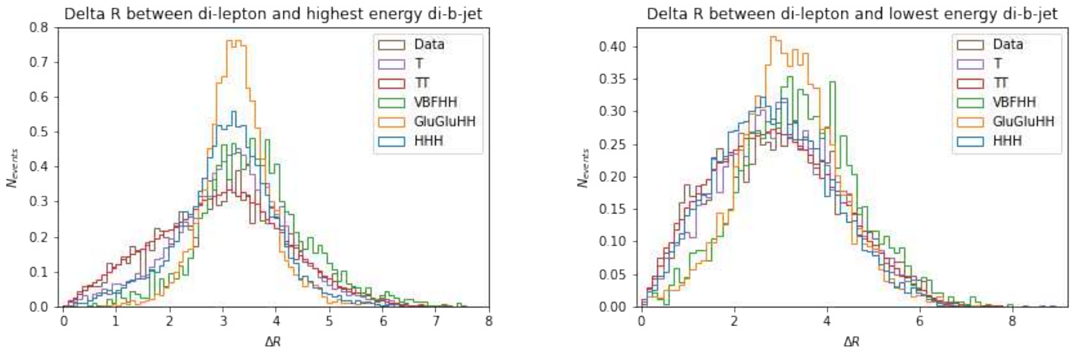
Figure 4.10: Normalized distributions of ΔR between both τ leptons for various processes. This plot presents vastly different shapes for Higgs and non-Higgs related processes.

help of its own invariant mass built-in function, this allows us to directly extract the invariant mass from these vectors. For the case of ΔR, it is a more straightforward calculation because this quantity depends solely on η and ϕ, which are variables that we already have.

Thus, here are 10 new variables :

- ΔR between both τ leptons in Fig. 4.10
- ΔR between the highest energy di-b-jet and the di-lepton (sum of the Lorentz vector of each τ) in Fig. 4.11a
- ΔR between the lowest energy di-b-jet and the di-lepton in Fig. 4.11b
- ΔR between both di-b-jets in Fig. 4.12
- ΔR between the di-lepton and the MET in Fig. 4.13
- Invariant mass of the di-lepton in Fig. 4.14
- Invariant mass of the sum of the di-lepton and the MET (sum of the Lorentz vectors) in Fig. 4.15a
- Invariant mass of the sum of the di-lepton, both di-b-jets and the MET in Fig. 4.15b
- Invariant mass of the highest energy di-b-jet in Fig. 4.16a
- Invariant mass of the lowest energy di-b-jet in Fig. 4.16b

At first glance, we see that these variables look far more interesting than the initial kinematic variables, as they offer clearer discrimination between HHH and the other processes. Figures 4.10, 4.14 and 4.15a all discriminate extremely well HHH from top-quark related processes, but not so much from the double Higgs processes. These three plots being so similar is most likely explained by the fact that τ leptons are



(a) **(b)**

Figure 4.11: Normalized distributions of ΔR between the di-lepton and the (a) highest/(b) lowest di-b-jet for various processes. A clear difference can be seen between the shape of HHH and the shapes of double Higgs processes.

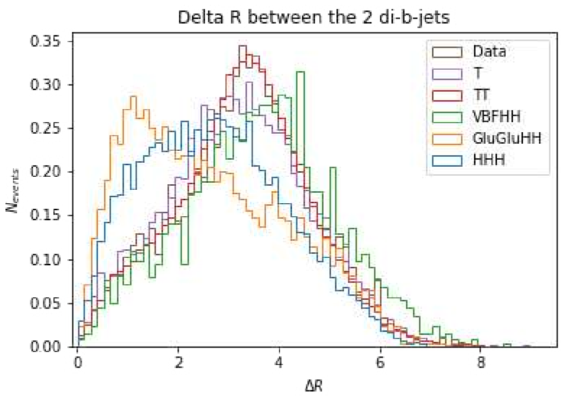


Figure 4.12: Normalized distributions of ΔR between both di-b-jets for various processes. This variable is quite unique due to the HHH, ggF HH and VBF HH all having different shapes.

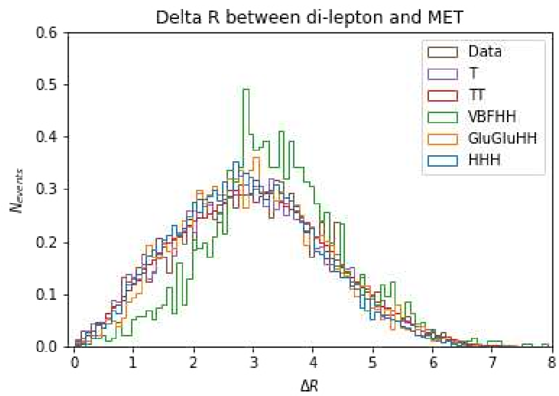


Figure 4.13: Normalized distributions of ΔR between the di-lepton and the MET for various processes.

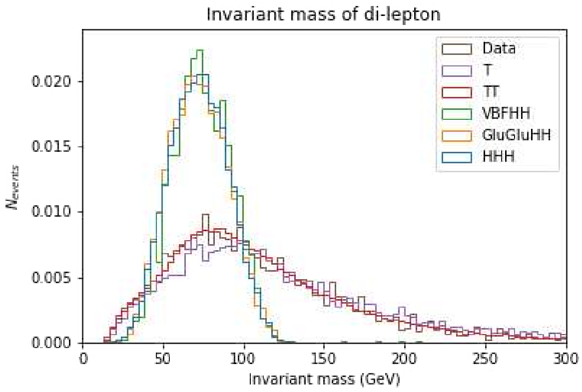
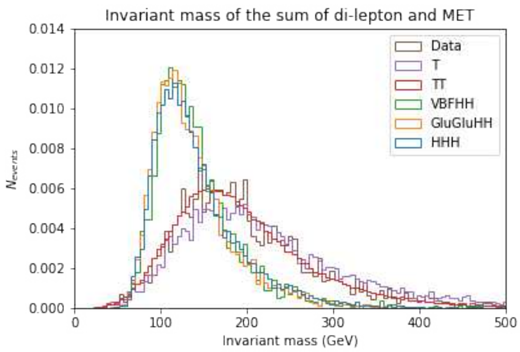
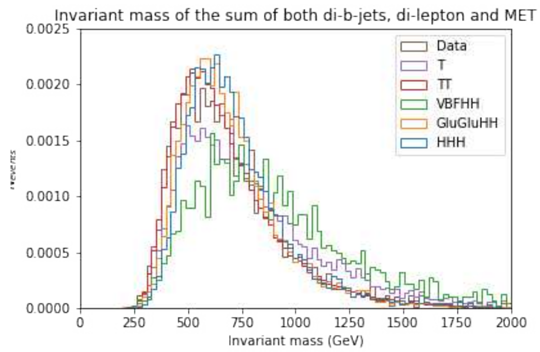


Figure 4.14: Normalized distributions of the invariant mass of the di-lepton for various processes.

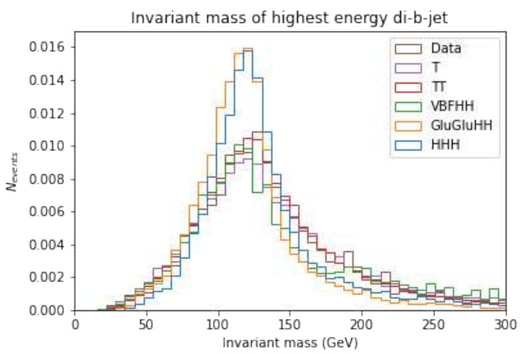


(a)

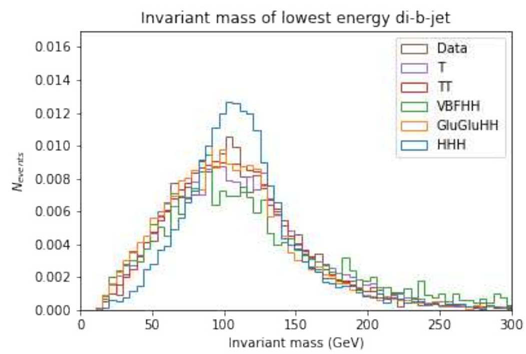


(b)

Figure 4.15: Normalized distributions of the invariant mass of the sum of the (a) di-lepton and the MET / (b) di-lepton, the MET and both di-b-jets for various processes.



(a)



(b)

Figure 4.16: Normalized distributions of the invariant mass of the (a) highest / (b) lowest energy di-b-jet for various processes.

produced by the decay of a Higgs in the case of HHH, ggF HH and VBF HH, as opposed to the top-quark related processes. By adding the invariant mass of the four b-jets to Fig. 4.15a, we obtain a very different plot in Fig. 4.15b. Indeed, the b-quarks are much heavier than the τ leptons and the MET, so the distributions are now dominated by their invariant mass and VBF HH is the only process that stands out.

Figures 4.11 and 4.16 compare the highest and lowest energy di-b-jets for the ΔR between them and the di-lepton, and their invariant masses respectively. Focusing on the HHH in Fig. 4.16, we see that it follows the ggF HH distribution in Fig. 4.16a and clearly stands out in Fig. 4.16b, because the highest energy di-b-jet most likely resulted from the Higgs decay for both processes unlike the lowest energy di-b-jet. One would then expect the same conclusion for Fig. 4.11, and although Fig. 4.11b behaves according to expectation, we observe in Fig. 4.11a that the HHH seems to be more spread out than its HH counterparts. While some figures might produce some slightly unexpected results, it is also wise to keep in mind that the b-jet pairing algorithm matched b-jets together with an accuracy of 52.77%, meaning that we must be careful when looking at plots involving di-b-jets. This applies especially to Fig. 4.12 in which, despite showing clear separation for HHH, both di-b-jets are involved, which could lead to larger errors.

Lastly, Fig. 4.13 will most likely be useful to discriminate HHH from VBF HH. It has to be noted that this plot is a bit curious because one would expect that all Higgs processes would follow roughly the same distribution since the di-lepton results from a Higgs decay. Multiple reasons could explain this discrepancy. First, the quarks produced in vector boson fusion can decay leptonically and therefore produce neutrinos which would change the direction of the MET. Second, as we have seen in Figs. 4.5c and 4.5d, these same quarks have the tendency to travel close to the $\hat{z}$ axis, and that could lead to some of the particles leaving the detector undetected and thus affecting the direction of the MET.

While these figures are very instructive about the behavior of the relevant processes, they are not indicative of the data that will be collected by CMS. To represent the actual distributions, we must first scale the MC-simulated samples to match the real quantities by multiplying them by the weight factor in eq. 4.5,

$$W = \frac{Br.\sigma.L}{N_{gen}} \tag{4.5}$$

with Br the branching ratio of that process, σ its cross section, L the luminosity and N_{gen} the number of generated events.

Then, we create stacked histograms of the now-weighted distributions, with the data plotted against it. These are made up from multiple histograms whose contents were added together into a single histogram. The stacked plots of the 10 variables can be seen in Fig. 4.17, as we can confidently conclude that, of all the events in the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state, an overwhelming majority of the events will come from the $t\bar{t}$ process. This also explains why the data follows $t\bar{t}$ very closely in each of the normalized plots. It has to be noted that some mismodelling of the $t\bar{t}$ background seems to have happened in the stacked plots. This mismodelling originates from the data samples themselves and the exact cause of this is unknown at this time, and thus needs to be further investigated. A way to fix these plots for the time being would be to apply some data-driven corrections.

In summary, the main takeaway is that every one of these 10 extra variables discriminate HHH from at least one other process, so this is a very useful step forward for the main objective of separating HHH from the other backgrounds.

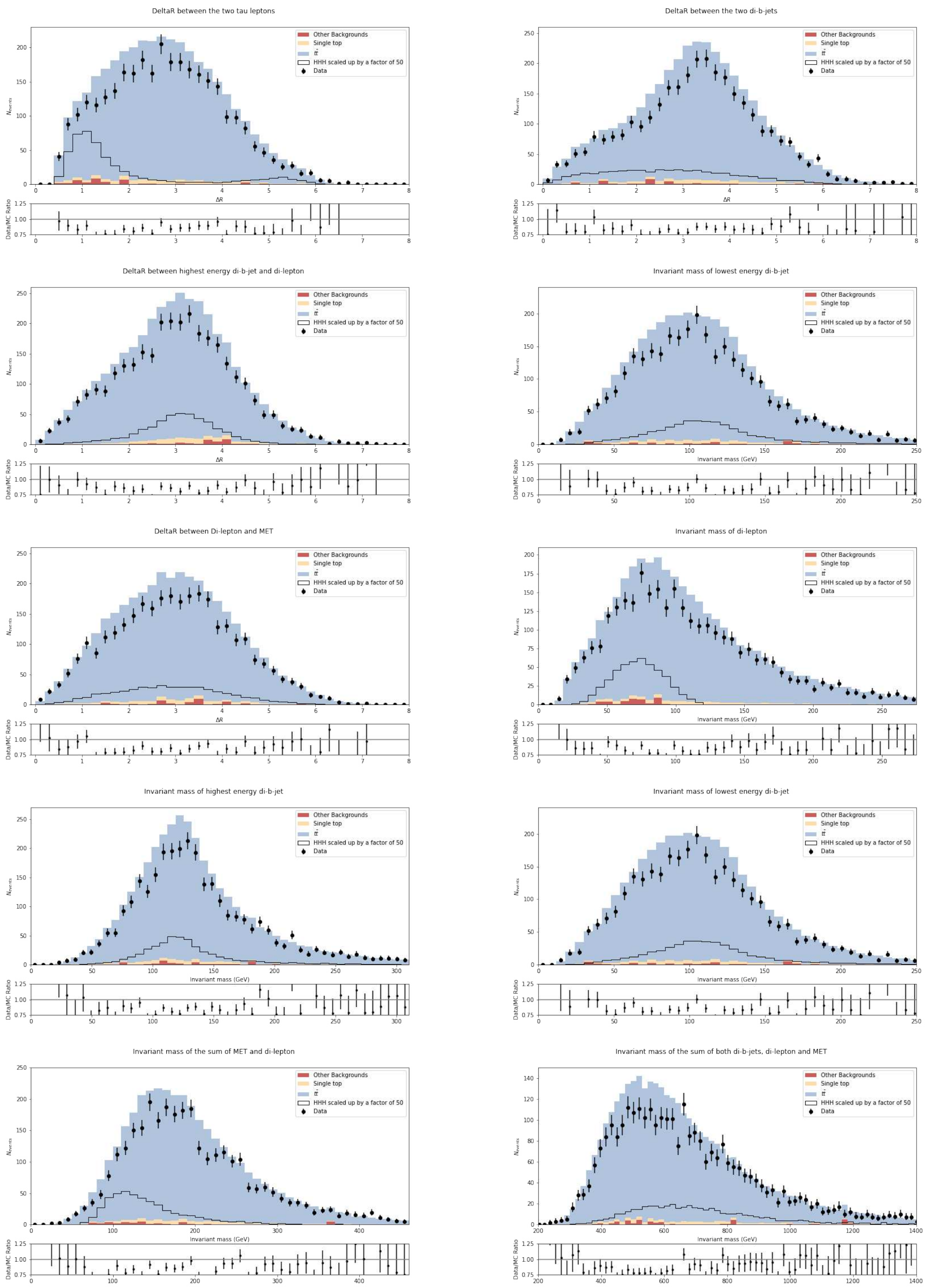


Figure 4.17: Stacked plots of all 10 variables. Each process has been scaled to their standard model cross section (except HHH which has been set by default to 1 pb).

Chapter 5

Machine learning

To discriminate triple Higgs from the other backgrounds, we will use the Toolkit for Multivariate Data Analysis with ROOT, also simply known as TMVA [20]. This toolkit contains many classification methods, including the 2 methods that will be used for this study : a Boosted Decision Tree (BDT) and a Neural Network (NN). We will use those two classifiers to compare triple Higgs to double Higgs produced by gluon-gluon fusion, to double Higgs produced by vector boson fusion and to $t\bar{t}$. This choice is made due to the similarity between HHH and HH, and due to the domination of the $t\bar{t}$ process in the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state.

5.1 Boosted Decision Tree

A decision tree is a machine learning classifier that resembles a genealogical tree, but where each node splits into exactly two other nodes. Each node possesses a criterion, or cut, that an event passing through the tree will either fulfill or not fulfill. Accordingly to their outcome, the node separates the events into the next two nodes, where the same process happens again with a different criterion for each node. The desired result is to end up with some predominantly background end nodes (or 'leaf' nodes) and some predominantly signal end nodes. We can see this overall structure of a decision tree in Fig. 5.1.

On its own, a decision tree is a rather weak classifier, so we can improve its performance by 'boosting' it. This is done by creating many trees iteratively, weighting them according to their performance and adding them together one by one so that the next tree can focus on the mistakes of previous trees [21]. Boosting a decision tree significantly enhances its performance, as well as stabilizing it by reducing the statistical fluctuations that can happen in a single tree.

To build a BDT, we make the data undergo a 'training' procedure. This training is done by selecting data from the HHH signal sample and from the background samples that we wish to compare the signal to. In our case, around 65% to 70% of the signal and background data samples were used to create the dataset that will undergo the training process of the BDT.

Firstly, the building of the BDT starts with a single node that uses an initial criterion to separate the dataset into two subsets, sending each subset to a different node. These nodes will then select the best criterion to separate the signal and the background, meaning that the program will go through hundreds of possible cuts for

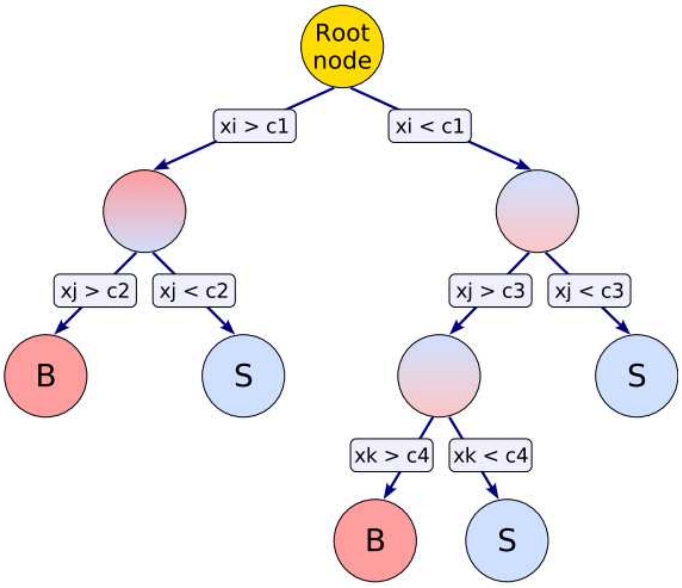


Figure 5.1: Basic illustration of a decision tree. The x_i are the variables on which some cuts c_n are applied. When a data subset is predominantly signal or background, the node does not apply a new cut and becomes a 'leaf' node (S and B nodes). [20]

dozens of variables to select the cut that will produce the best possible separation between signal and background. This process is then once again repeated for the next set of nodes, as their number keeps growing exponentially. The training stops when the tree has attained a satisfactory signal/background separation or when it has reached the desired size, which is set by the user before the training. The number of iterations is to be chosen carefully, as too few would not sufficiently separate the signal from the background and too many could lead to the overtraining of the BDT.

Overtraining can be described as a phenomenon that happens in machine learning where the method (in our case the BDT) has learned the dataset 'too well'. This means that when the BDT is used, for example, on real data from CMS runs, the decision tree might not recognize the signal due to rejecting it because of the small differences that might occur between the signal dataset it trained on and the signal found in the data. Overtraining is therefore caused by too many cuts made on a too small sample size, which implies that BDTs are particularly susceptible due to their large number of nodes.

After the training is done, we can test the BDT on the remaining 30 to 35% of the data that was left untouched in the initial data samples. This is done by applying the BDT to this yet untouched dataset. The obtained results will be interpreted in the form of TMVA response plots and ROC curves, all of which will be detailed in chapter 5.3.

5.2 Multilayer Perceptron

The Multilayer Perceptron (MLP) is an artificial neural network that is set up the same way as a BDT, i.e. we split the signal and background to form a 'training'

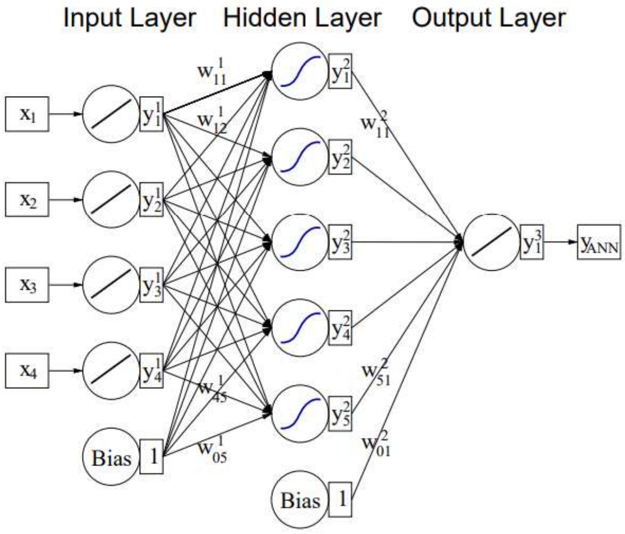


Figure 5.2: Illustration of an MLP neural network with a single hidden layer. [20]

and a 'testing' dataset, but functions fundamentally differently than it.

An artificial neural network (ANN) is a machine learning structure that takes inspiration from how the human brain operates. Instead of neurons, we use a set of interconnected nodes that answer more or less strongly to different types of input information. The MLP is an ANN that has its nodes organized in an input layer, an output layer and hidden layers to reduce complexity [20]. The input layer is the layer that contains the input variables, the output layer returns, in our case, the score of each event passing through it, and the 'hidden' layers are the nodes that will do the classification (Fig. 5.2). Neurons send and receive signals to and from other neurons, with the strength of those signals depending on weights that get adjusted throughout the learning process, similarly to the boosting process of a decision tree. The MLP will try to minimize the errors during training by using these weights.

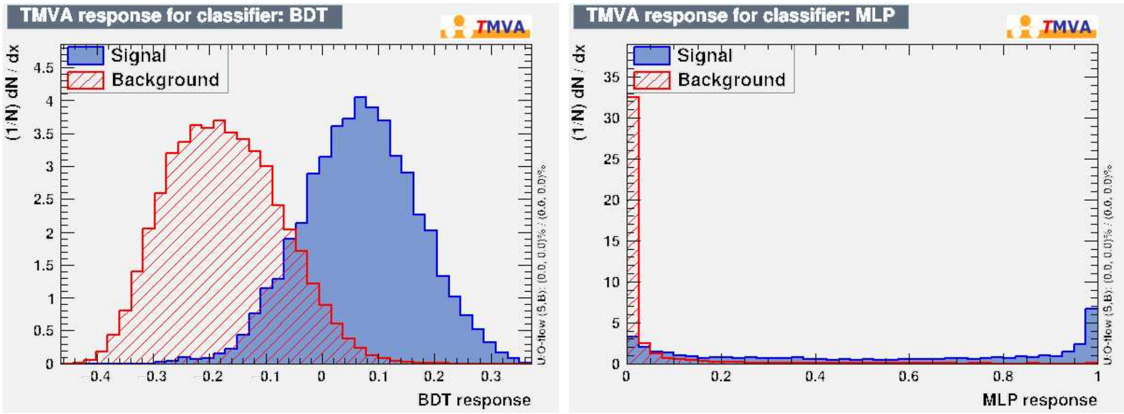
5.3 Results

In this section, results will be presented for the BDT and MLP performances for 3 different signal/background combinations, i.e. HHH vs $t\bar{t}$, HHH vs ggF HH and HHH vs VBF HH. The HH processes were chosen as backgrounds to discriminate due to their similarity with HHH, while $t\bar{t}$ was selected due to its prevalence in the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state.

We then conclude this chapter with a discussion on the impact of the 10 new variables mentioned in Chapter 4.

5.3.1 HHH vs $t\bar{t}$

Let's start first by looking at the results of the BDT and MLP for the HHH vs $t\bar{t}$ scenario. We can estimate the performance of the classifiers by looking at the TMVA response for both classifiers. Each event in the 'testing' sample will be given a score expressing the level of confidence of the classifier on whether the event is signal or background. By convention, events plotted towards the right side (higher



(a) **(b)**

Figure 5.3: The (a) BDT and (b) MLP response plots for the HHH vs $t\bar{t}$ scenario. It is important to note that these distributions are normalized.

score) of the classifier response plots are more likely to be signal, while events on the left side (lower score) are more likely to be background.

Even though testing was done on a dataset made up of both signal and background events, the classifiers are of course oblivious to this information while testing. There is however no problem in revealing this information after the testing to check how well the signal was separated from the background. So, we can see on Figure 5.3 that the signal got rather well separated from the background for both classifiers.

A more intuitive way to visualize this separation, would be to extract a ROC curve from these plots. A Receiver Operating Characteristic curve, or ROC curve, is used to indicate how well separated the signal is from the background for multiple classifiers by drawing a single curve. The ROC curve is obtained by plotting the signal efficiency against the background rejection, which we can directly get from the classifier response plots. Indeed, let's imagine that we keep only the events in the bins on the right hand side of 0.2 in Fig. 5.3a. Then our background rejection would be near perfect as almost no background event would be taken in the selected sample, but our signal efficiency would be very low considering most of the signal events are not selected. On the contrary, selecting all of the events in the bins on the right hand side of -0.2 does improve signal efficiency up to nearly 100%, but it comes at the cost of also selecting over half of the background events. By repeating this operation for every bin, we do find the ROC curve on Figure 5.4. The perfect curve, indicating perfect separation, would make a right angle in the top right corner at $(x, y) = (1, 1)$.

We now look at possible overtraining of the classifiers. Overtraining can be detected by applying the classifiers to the 'training' dataset and by comparing its response with the response of the 'testing' dataset. If the distributions are different, it means that the classifier has learned the data too well, i.e. there is some overtraining, and it overperformed when testing the same events that trained it.

Looking at Figure 5.5, we can see slight overtraining in both classifiers for the signal. This overtraining is however small enough to be neglected.

Lastly, we apply the classifiers on the data, as well as every background and create stacked histograms for both methods. These plots are shown at Fig 5.6. The figures

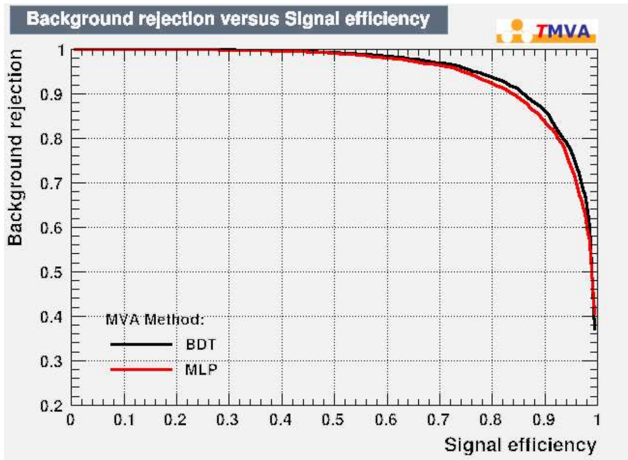
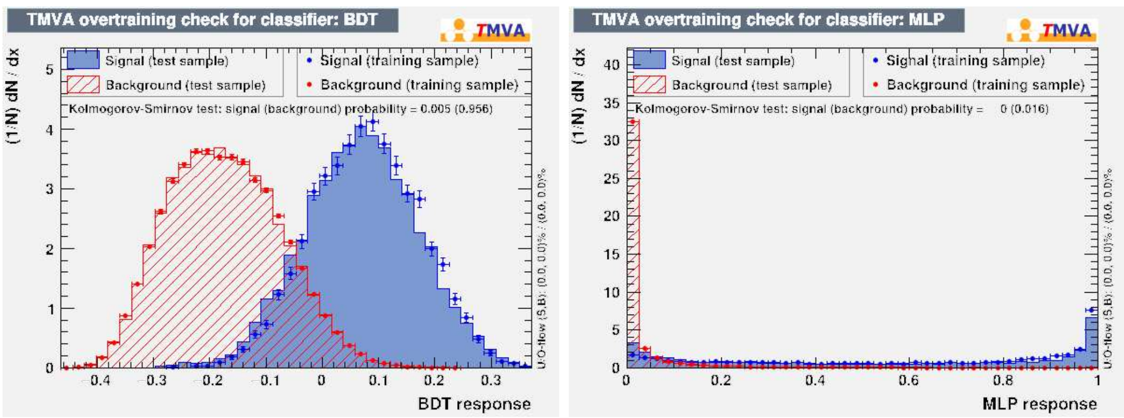
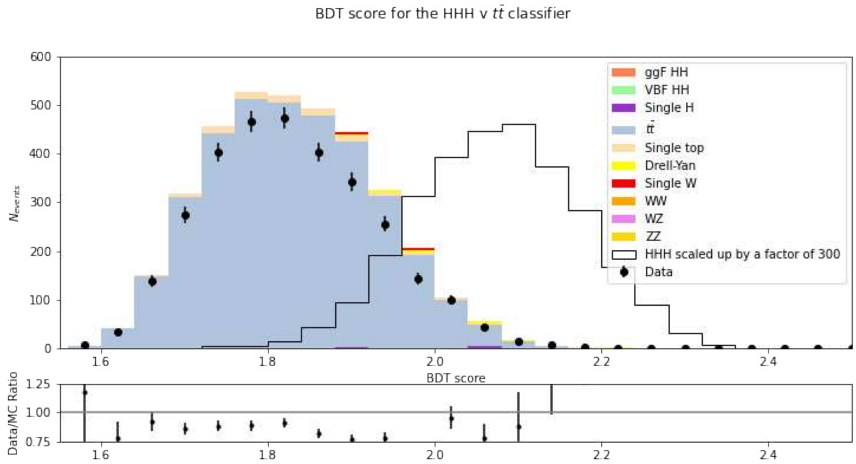


Figure 5.4: ROC Curve of the BDT and the MLP for the HHH vs $t\bar{t}$ scenario.

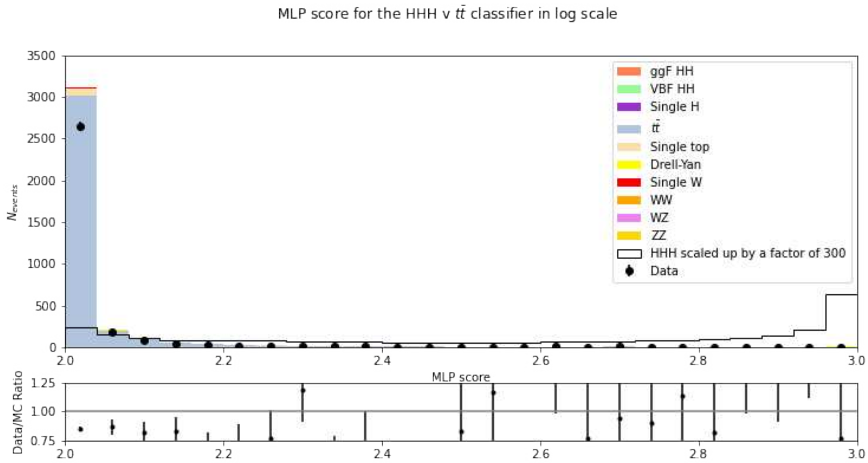


(a) **(b)**

Figure 5.5: Overtraining plots for the (a) BDT and (b) MLP in the HHH vs $t\bar{t}$ scenario. The training sample distributions are plotted against the test sample distributions which are the same as in Fig. 5.3. Slight overtraining can be observed for the signal in both classifiers.



(a)



(b)

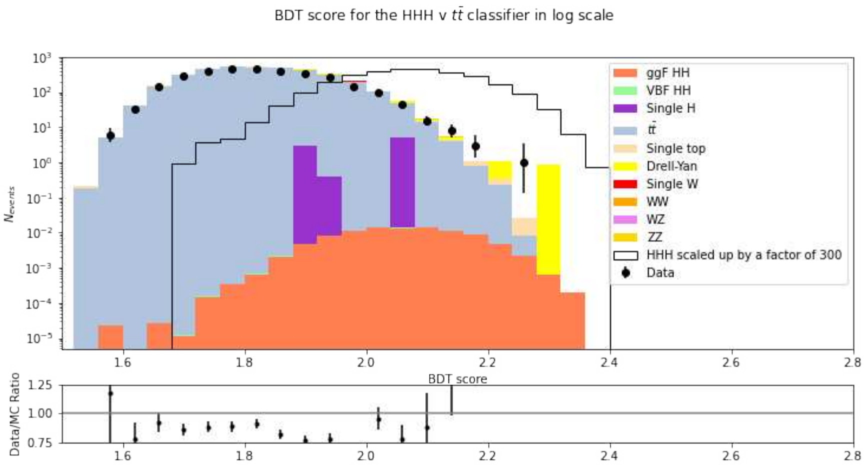
Figure 5.6: Stacked plots of the (a) BDT and (b) MLP response for all processes for the HHH v $t\bar{t}$ scenario.

correspond almost exactly to what the training looked like and it is very easy to see the overall good separation of HHH and $t\bar{t}$ for these classifiers. We however cannot see well how it performed for the other backgrounds due to them being an insignificant portion of the total events. Thus, from now on, these plots will be in log scale. The results of this change of perspective can be seen in Fig. 5.7 where we notice that by using the HHH v $t\bar{t}$ classifiers, other background are not well separated at all from HHH.

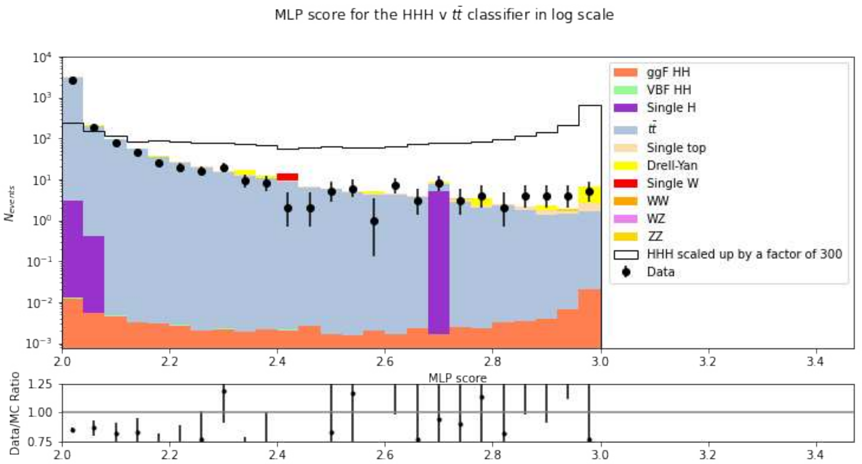
The main takeaway for the HHH vs $t\bar{t}$ scenario is that the classifiers performed well in separating HHH from $t\bar{t}$ because the processes are quite different, as we have already established.

5.3.2 HHH vs VBF HH

Comparing triple Higgs to double Higgs increases the difficulty of separating signal from background due to the similarity between those two processes. By looking

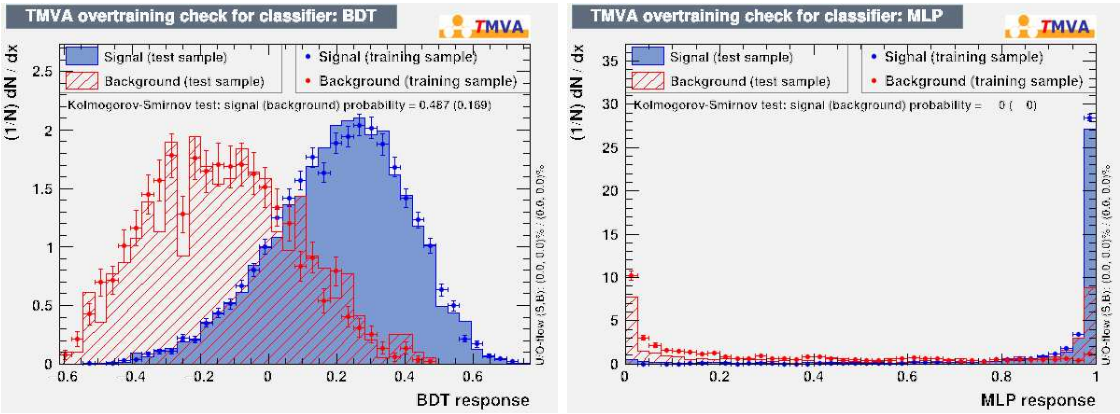


(a)



(b)

Figure 5.7: Stacked plots of the (a) BDT and (b) MLP response for all processes in log scale for the HHH v $t\bar{t}$ scenario.



(a) **(b)**

Figure 5.8: The (a) BDT and (b) MLP response and overtraining plots for the HHH vs VBF HH scenario.

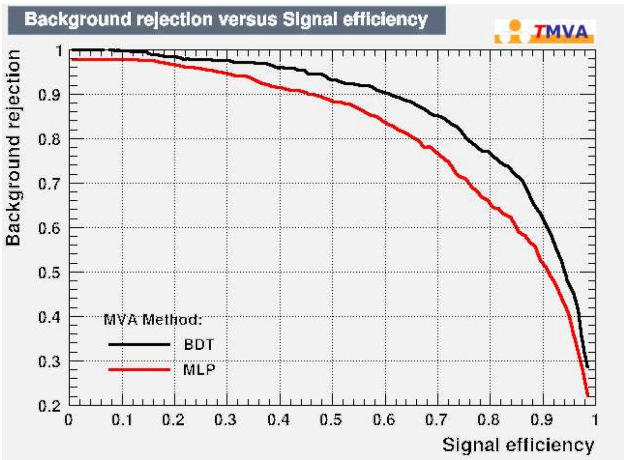
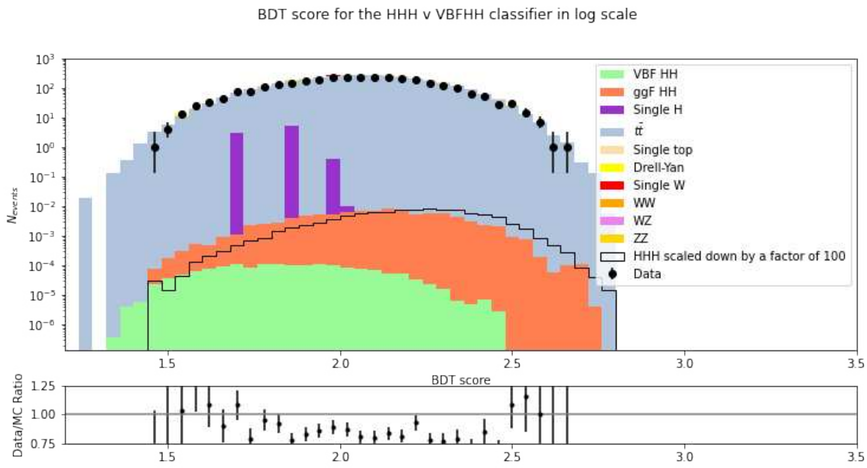


Figure 5.9: ROC Curve of the BDT and the MLP for the HHH vs VBF HH scenario.

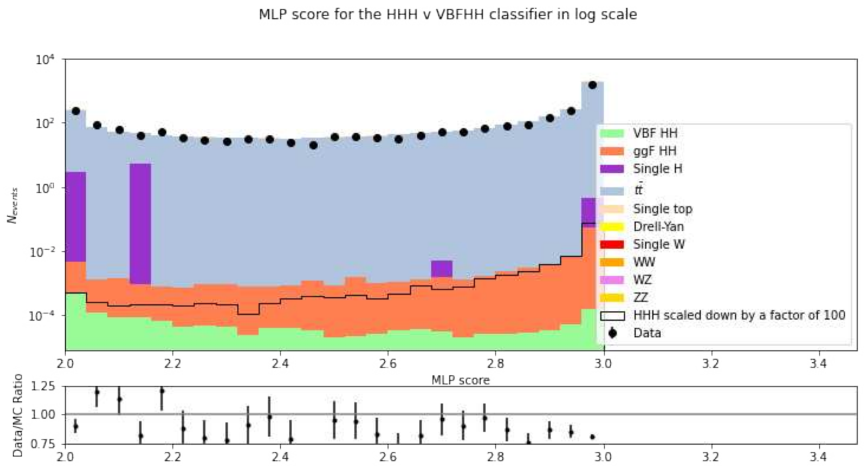
at the classifier response plots on Figure 5.8, it is clear that the separation is of lower quality, as there is a sizeable overlap of the signal and background. Slight overtraining seems to have taken place again, but this time for the background. The shape of the ROC curve (Fig. 5.9) confirms these observations, with the MLP looking worse than the BDT, most likely due to non-optimized parameters in the training process. Figures 5.10a and 5.10b are also in agreement with the previous figures, and shows poor performance in classifying the other processes.

5.3.3 HHH vs ggF HH

As mentioned previously, the triple Higgs events used for this study were produced by gluon-gluon fusion. This means that differentiating the HHH process from the ggF HH process would be one of, if not the main hurdle to overcome if we are to study triple Higgs production going forward. Hence, as predicted, Figures 5.11, 5.12 and 5.13 do show results that indicate an even less-than-ideal separation of signal and background.

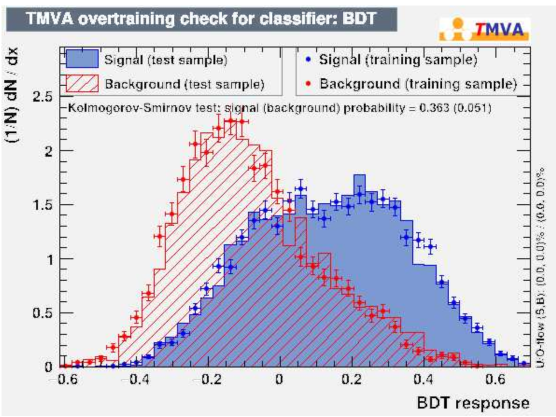


(a)

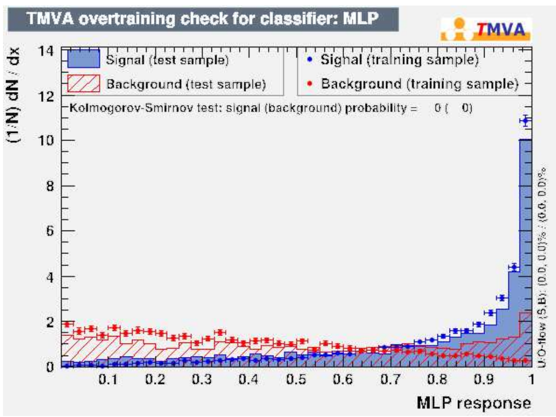


(b)

Figure 5.10: Stacked plots of the (a) BDT and (b) MLP response for all processes in log scale for the HHH v VBF HH scenario.



(a)



(b)

Figure 5.11: The (a) BDT and (b) MLP response and overtraining plots for the HHH vs ggF HH scenario.

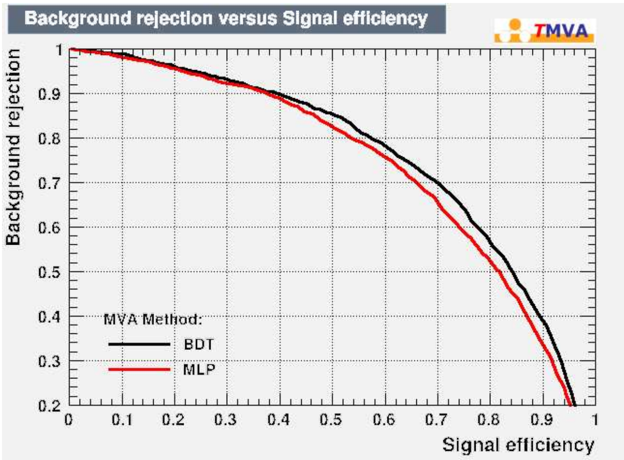
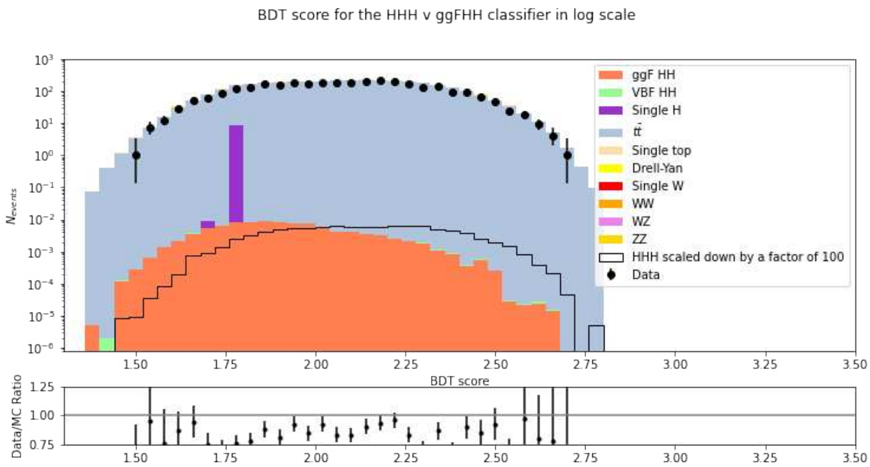
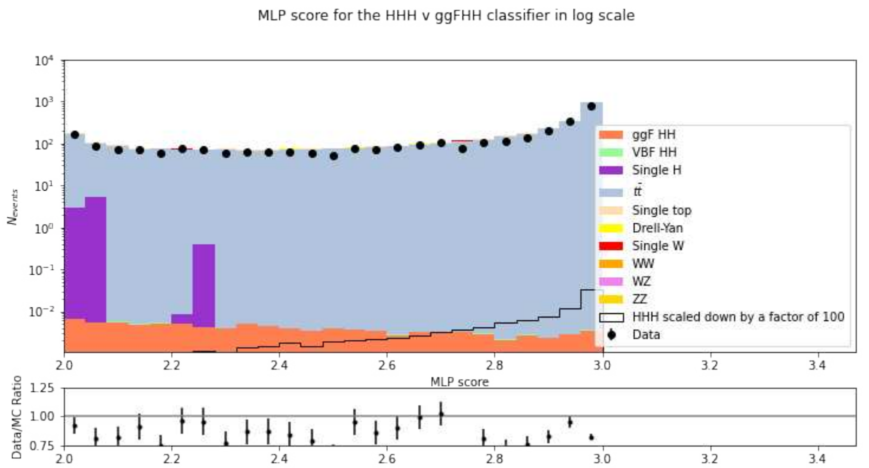


Figure 5.12: ROC Curve of the BDT and the MLP for the HHH vs ggF HH scenario.



(a)



(b)

Figure 5.13: Stacked plots of the (a) BDT and (b) MLP response for all processes in log scale for the HHH v ggF HH scenario.

	90%	80%	60%	40%	20%
HHH v $t\bar{t}$	2.188×10^{-6}	4.18×10^{-6}	9.095×10^{-6}	5.812×10^{-5}	1.701×10^{-4}
HHH v ggF	1.301×10^{-2}	1.615×10^{-2}	2.542×10^{-2}	3.786×10^{-2}	5.877×10^{-2}
HHH v VBF	1.245	2.031	3.659	8.721	12.737

Table 5.1: The $\frac{Signal}{Background}$ ratios if we make a score cut in a way that the selected sample would contain roughly the highest-scoring 90, 80, 60, 40 and 20% of the signal for all three scenarios.

A more mathematical way of interpreting these stacked plots would be to look at how pure our signal would be if we only accepted the highest scoring portion of it. Normalizing the processes to their SM cross section at a center-of-mass energy of 13.6 TeV ($\sigma_{HHH} = 0.09\ fb$, $\sigma_{ggF} = 34.13\ fb$, $\sigma_{VBF} = 1.87\ fb$ and $\sigma_{t\bar{t}} \propto mb$), Table 5.1 shows the $\frac{Signal}{Background}$ ratios for 5 different signal cuts for each scenario (i.e. HHH vs $t\bar{t}$, ggF HH and VBF HH). In other words, this table acts a like a weighted ROC curve, as it shows how much of a specific background we would have if we made a cut on the score such as the selected sample would contain roughly the highest-scoring 20, 40, 60, 80 or 90% of our signal. While the HHH vs VBF HH $\frac{S}{B}$ ratios indicates that there is more signal than background for each cut, the same thing cannot be said for the two remaining backgrounds. For HHH vs ggF HH, the low values are mostly explained by the poor separation by the classifiers. For HHH vs $t\bar{t}$ however, it is due to the sheer quantity of $t\bar{t}$ in that decay into the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state. The HHH vs VBF HH ratios looking that much better than the ratios for HHH vs ggF HH could be explained by the fact that its cross section σ_{VBF} is 20 times lower than σ_{ggF}, but is also more efficiently separated from the signal.

Ways to improve those results could be to introduce additional variables, optimize the training parameters or work with a higher number of events. Indeed, more events could allow us, for example, to increase the number of layers in the BDT without running the risk of severe overtraining.

5.3.4 Impact of the event variables

Having created and used 10 new variables for the training of the classifiers, it is natural to wonder about the actual size of their impact on them, if they had any at all. To measure their influence, we will look at same plots as for the HHH vs $t\bar{t}$ scenario, but this time with the classifiers having been trained on basic kinematic variables only.

By looking at the ROC curve on Fig. 5.14 and comparing it to Fig. 5.4, we can see a small but non-negligible improvement between both curves. The same conclusion can be made by comparing Figures 5.15a and 5.15b to Figures 5.5a and 5.5b respectively. These observations are further confirmed by the variable ranking list, which orders the variables from most to least relevant during the training process. For BDTs, the relevance is defined by counting the number of times the variable has been used to split the nodes in two, weighted by how well it performed at separating signal from background. For MLPs, the relevance is related to the sum of the weights-squared of the connections between the input (variable) layer and the first hidden layer [20]. By looking at Tables 5.2 and 5.3, we see that, out of the 10

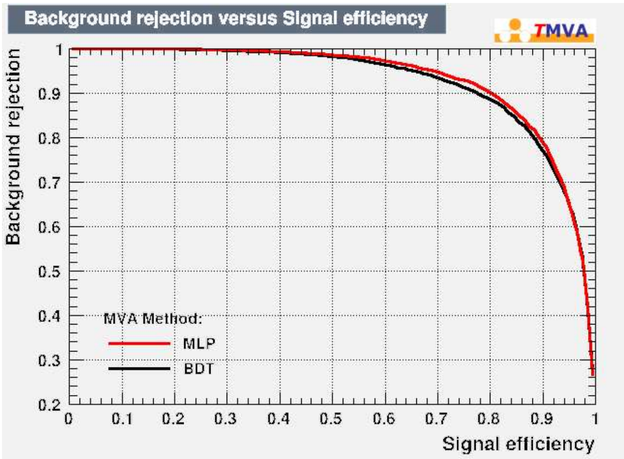


Figure 5.14: ROC Curve for the HHH v $t\bar{t}$ scenario with only the basic kinematic variables introduced in Chapter 4.5.1.

event variables, 50% of them are in the top 10 most relevant variables for the BDT training and 40% are in the top 10 for the MLP training. It is also noteworthy to mention that the MLP seems to favor the 'energy' variables, i.e. the invariant mass, the transverse momentum and the energy, whereas the angular variables seem to not be relevant in the training.

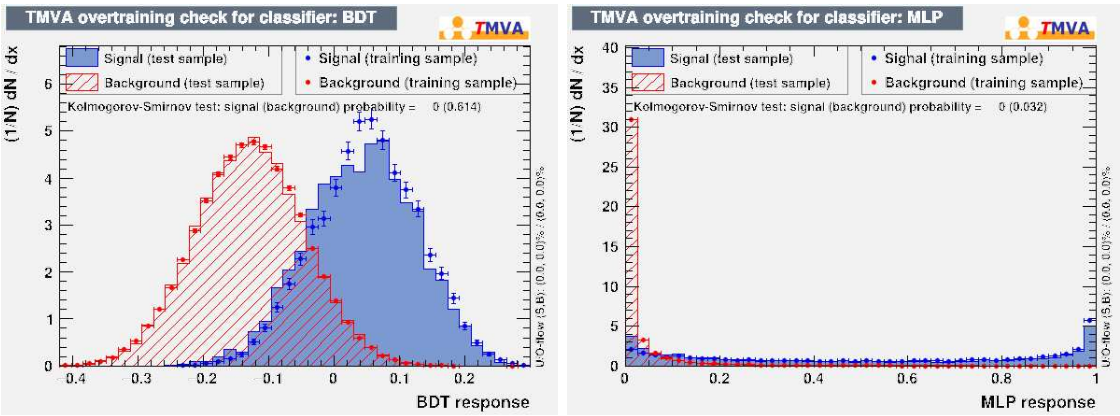


Figure 5.15: The (a) BDT and (b) MLP response and overtraining plots for the HHH v $t\bar{t}$ scenario with only the basic kinematic variables introduced in Chapter 4.5.1

Table 5.2: Top 30 most relevant variables for BDT in the HHH v $t\bar{t}$ scenario

Rank	Variable	Relevance
1	Invariant mass of di-lepton	4.068e-02
2	ΔR between both leptons	3.122e-02
3	Invariant mass of sum of di-lepton and MET	3.122e-02
4	Energy of MET	2.332e-02
		Continued on next page

Table 5.2 – continued from previous page

Rank	Variable	Relevance
5	ΔR btwn di-lepton and highest E di-b-jet	2.244e-02
6	ϕ of third highest E b-jet	2.178e-02
7	Transverse momentum of τ_h	2.143e-02
8	θ of highest E b-jet	2.088e-02
9	ΔR between both di-b-jets	2.055e-02
10	θ of τ_e	2.053e-02
11	Invariant mass of highest E di-b-jet	2.036e-02
12	ϕ of τ_e	2.027e-02
13	θ of τ_h	1.978e-02
14	Transverse momentum of highest E b-jet	1.951e-02
15	ϕ of lowest E b-jet	1.951e-02
16	ϕ of τ_h	1.948e-02
17	ϕ of highest E b-jet	1.892e-02
18	ΔR between di-lepton and MET	1.883e-02
19	Invariant mass of lowest E di-b-jet	1.882e-02
20	θ of τ_h	1.879e-02
21	ϕ of second highest E b-jet	1.874e-02
22	ΔR btwn di-lepton and lowest E di-b-jet	1.874e-02
23	Transverse momentum of third highest E b-jet	1.872e-02
24	p_y of MET	1.857e-02
25	η of highest E b-jet	1.853e-02
26	η of third highest E b-jet	1.850e-02
27	η of τ_h	1.844e-02
28	θ of third highest E b-jet	1.843e-02
29	p_x of τ_e	1.843e-02
30	p_x of MET	1.832e-02

Table 5.3: Top 30 most relevant variables for MLP in the HHH v $t\bar{t}$ scenario

Rank	Variable	Relevance
1	Invariant mass of di-lepton	6.535e+02
2	Invariant mass of sum of di-lepton and MET	2.605e+02
3	Transverse momentum of third highest E b-jet	1.996e+02
4	Transverse momentum of τ_h	1.556e+02
5	Energy of τ_h	1.138e+02
6	Energy of third highest E b-jet	1.102e+02
7	Inv. mass of di-lepton, MET and all b-jets	1.077e+02
8	Energy of MET	1.075e+02
9	Energy of lowest E b-jet	9.824e+01
10	Invariant mass of highest E di-b-jet	9.661e+01
11	Invariant mass of highest E di-b-jet	9.295e+01
12	Transverse momentum of highest E b-jet	9.292e+01
13	Energy of highest E b-jet	9.269e+01
14	Transverse momentum of second highest E b-jet	9.045e+01
15	Transverse momentum of lowest E b-jet	9.024e+01

Continued on next page

Table 5.3 – continued from previous page

Rank	Variable	Relevance
16	Energy of second highest E b-jet	7.939e+01
17	Transverse momentum of τ_e	6.644e+01
18	Invariant mass of lowest E di-b-jet	5.938e+01
19	ΔR between both leptons	4.665e+01
20	θ of third highest E b-jet	4.303e+01
21	θ of τ_e	3.497e+01
22	ϕ of highest E b-jet	3.079e+01
23	ϕ of lowest E b-jet b-jet	2.868e+01
24	θ of τ_h	2.851e+01
25	θ of second highest E b-jet	2.672e+01
26	η of third highest E b-jet	2.607e+01
27	ϕ of τ_e	2.461e+01
28	θ of lowest E b-jet	2.359e+01
29	ϕ of τ_h	2.295e+01
30	ΔR between both di-b-jets	2.288e+01

Chapter 6

Conclusion and future prospects

In this thesis, we studied and tested an analysis strategy for the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state, with the objective of training and using classifiers to discriminate HHH from other backgrounds. We first presented the basic variables collected by CMS and simulated by the Monte Carlo algorithm. We then identified and computed additional useful variables, to further improve the discrimination quality of the classifiers. These classifiers were then trained and tested against the $t\bar{t}$, ggF HH and VBF HH backgrounds. They are subsequently used on these backgrounds, which allows us to compute their $\frac{S}{B}$ ratios for various cuts.

What about the future of this study ? First, there are many improvements that can be done on most of the things that were presented in this thesis, such as using data from the events where the τ decayed into a muon, creating a better b-jet pairing algorithm, finding and fixing the aforementioned scaling problem of $t\bar{t}$ in the stacked plots, creating more variables to enable better separation with the classifiers, or optimizing the parameters of the classifiers such as the number of nodes.

The next step for this study is to optimize and apply selection cuts on all three discriminators. As we have seen in Chapter 5, the classifiers are designed to perform well in discriminating the signal from the specific background that they were trained to discriminate, but we noticed that they are not effective at separating HHH from the other backgrounds. The idea would then be to separate HHH from one background at a time. This could be done by applying classifiers that were trained to discriminate HHH from a specific background on the data, and then making a first cut to separate as most background as possible. Then, the remaining data is trained again with the appropriate classifier and a new cut is applied to once again separate the biggest background from the signal. These steps can be repeated until the remaining data has large enough $\frac{S}{B}$ ratio. The next step after this would then be to perform statistical interpretations on this remaining data, by setting limits on the HHH cross section and constraints on the λ_3 and λ_4 parameters.

For a clearer view of the context surrounding this thesis, Fig. 6.1 shows a flowchart of the entire HHH $\rightarrow b\bar{b}b\bar{b}\tau^+\tau^-$ analysis process.

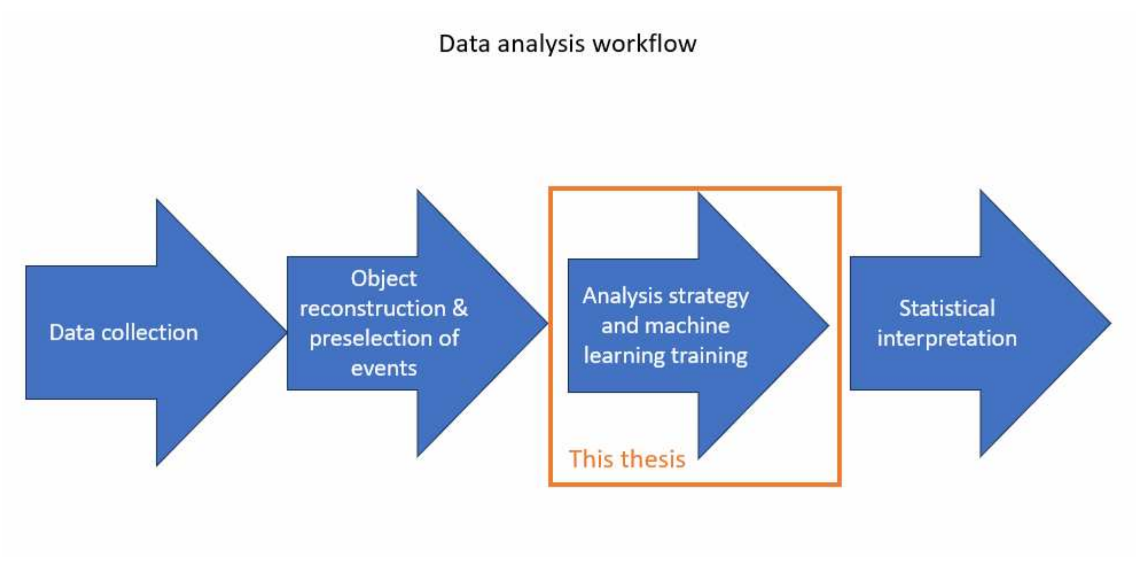


Figure 6.1: Flowchart putting this thesis in the context of the entire analysis process of the $b\bar{b}b\bar{b}\tau^+\tau^-$ final state, from data collection by CMS to the statistical interpretation and paper publishing.

Chapter 7

Contributions

To write this thesis, I had to learn new sets of skills and to go through numerous steps. Listed here are these learnings and accomplishments :

- Learning how to use the fundamentals of ROOT (manipulation of ROOT Trees, processing of the data, ...)
- Creation of the Lorentz vectors using the TLorentzVector ROOT class
- Pairing of the b-jets using 3 different algorithms
- Check of the accuracy of the b-jet pairings with truth information
- Creation of all of the variable distribution plots (normalized and stacked)
- Creation of my own ROOT Tree to allow the classifiers to work
- Training and testing of the BDT and MLP classifiers for HHH against 3 different processes to create the response plots
- Computation of the S/B ratios

Francisco Casalinho must also be credited for creating a python environment that allowed me to use ROOT with python, for making the pre-selection cuts on the data and sending me this data through the ROOT Trees that he made.

Bibliography

[1] L. Cadamuro, *Experimental Apparatus*. Cham: Springer International Publishing, 2018.

[2] M. Stamenkovic, "Hh overview at cms: lessons learned for hhh." `https://indico.cern.ch/event/1232581/contributions/5442567/attachments/2684233/4656849/20230714_HH_overview_HHH_WS.pdf`, 2023.

[3] T. et al., "A portrait of the higgs boson by the cms experiment ten years after the discovery," *Nature*, vol. 607, p. 60–68, July 2022.

[4] P. D. G. et al., "Review of particle physics," *Progress of Theoretical and Experimental Physics*, vol. 2020, p. 083C01, 08 2020.

[5] C. Collaboration, "Identification of hadronic tau lepton decays using a deep neural network," *Journal of Instrumentation*, vol. 17, p. P07023, July 2022.

[6] CERN, "Detector | CMS Experiment — cms.cern." `https://cms.cern/detector`. [Accessed 06-06-2025].

[7] A. Collaboration, "Atlas open data : Higgs boson." `https://opendata.atlas.cern/docs/visualization/the-higgs-boson`.

[8] A. Experiment, "Atlas experiment glossary : Vector boson fusion (vbf)." `https://atlas.cern/glossary/vector-boson-fusion`.

[9] G. A. et al., "Search for triple higgs boson production in the $6b$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector," *Phys. Rev. D*, vol. 111, p. 032006, Feb 2025.

[10] CERN, "The Large Hadron Collider — home.cern." `https://home.cern/science/accelerators/large-hadron-collider`. [Accessed 06-06-2025].

[11] Wikipedia, "Large Hadron Collider — Wikipedia, the free encyclopedia." `http://en.wikipedia.org/w/index.php?title=Large\%20Hadron\%20Collider&oldid=1292355439`, 2025. [Online; accessed 07-June-2025].

[12] S. et al., "Particle-flow reconstruction and global event description with the cms detector," *Journal of Instrumentation*, vol. 12, p. P10003–P10003, Oct. 2017.

[13] L. A. et al., "Triggering dark showers with conditional dual auto-encoders," *Machine Learning: Science and Technology*, vol. 5, p. 035064, sep 2024.

[14] C. LI, “B-tagging in atlas experiment.” https://indico.in2p3.fr/event/13330/contributions/13775/attachments/11586/14280/Changqiao_TPJRJC.pdf.

[15] G. Bagliesi, “Tau tagging at atlas and cms.” https://arxiv.org/abs/0707.0928, 2007.

[16] H. A. et al., “Hhh whitepaper,” *The European Physical Journal C*, vol. 84, Nov. 2024.

[17] R. Brun and R. Fons Rademakers, “An object oriented data analysis framework.” https://root.cern.

[18] V. B. et al., “Search for non-resonant hhh in gluon fusion production mode in 6 b-jets,” 2024.

[19] L. C. et al., “Search for nonresonant higgs boson pair production in the four b quarks decay channel,” 2019.

[20] K. A. et al., “Tmva 4 toolkit for multivariate data analysis with root users guide.” https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf, 2020.

[21] K. Woodruff, “Introduction to boosted decision trees.” https://indico.fnal.gov/event/15356/contributions/31377/attachments/19671/24560/DecisionTrees.pdf, 2017.

UNIVERSITE CATHOLIQUE DE LOUVAIN

Faculté des sciences

Place des sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/sc**