UNIVERSITÀ DELLA CALABRIA

PHYSICS DEPARTMENT

Master degree in *Physics*

Study of the production of a W boson in association with two b-jets using machine learning with the ATLAS experiment at the LHC

Supervisor:

Prof. Evelin Meoni

Candidate:

Atena Harareh
Mat. 234144

ACADEMIC YEAR 2023/2024

Dedicated to

My dear mom and dad, In the intricate dance of particles and the exploration of the subatomic realm, your love and encouragement have been my guiding forces, transcending the physical distance that separates us. This master's thesis in nuclear physics is a tribute to your unwavering support and the intellectual curiosity you instilled in me. Even as particles collide in accelerators, your belief in my potential has propelled me forward.
Thank you for being the cosmic constants in my scientific journey.

Summary

This thesis is focused on the study of the production of a W boson in association with jets generated from b-quarks (b-jets) in proton-proton collisions data using the ATLAS detector at the Large Hadron Collider. The final state is constituted by a W boson that decays leptonically into a muon and a neutrino in association with two b-jets. Therefore muons, jets, b-jets, and missing transverse energy are the primary physics objects used in this analysis. The measurement of differential cross-section in this final state has never been performed at the Large Hadron Collider so far. The main challenge is constituted by the huge background containing events originating from top-quark pair production.

This thesis presents a strategy optimized for the rejection of the top background using a powerful machine learning approach based on the development of a Neural Network. This study marks the first step for the first differential cross-section measurement of this process in the near future. The study is based on simulated Monte Carlo samples of the signal and the backgrounds officially produced by the ATLAS Collaboration. The simulated samples reproduce the experimental conditions of the full Run-2 dataset, collected by the experiment from 2015 until 2018 at the centre-of-mass energy of $\sqrt{s} = 13$ TeV. This dataset represents the largest and best understood dataset available so far at the ATLAS experiment and corresponds to an integrated luminosity of 140 fb^{-1}. The simulated samples are processed using the full detector simulation and analysed with the same reconstruction tools used for real collision data, therefore they serve as a faithful emulation of real collision data.

Contents

Introduction

The field of particle physics, which focuses on understanding the fundamental components of matter and their interactions, has made remarkable progress thanks to the construction and operation of the Large Hadron Collider (LHC) at CERN. The LHC, the largest and most powerful particle accelerator ever built, collides protons at unprecedented energies, creating a unique setting to explore and validate the predictions of models. The Standard Model theory (SM) is the theory that currently best describes the phenomenology of particle physics, however, it cannot be the ultimate theory of particle physics, indeed, it does not explain phenomena such as Dark Matter and Matter-Antimatter asymmetry among others. Other theories overcome the dark aspects of the SM, these predict new phenomenologies that have to appear at very high energy, leaving the SM as a special case valid at low energies. The ATLAS experiment at the LHC is focused on the validation of the SM at the new energy scale, on the study of the Higgs boson physics, that in the SM is closely related to the origin of mass of fundamental particles, and on the searches for new phenomena beyond SM.

This thesis focuses on studying the production of a W boson with b-quarks (W+ 2b-quarks) [1] through proton-proton collisions data at the centre-of-mass energy of $\sqrt{s}$= 13 TeV collected by the ATLAS experiment at the LHC. This process is predicted by the SM and it is mediated by the strong interaction occurring between quarks and gluons that constitute the proton. Therefore this investigation contributes to the improvement of the understanding of per-

[1] Unless mentioned, it is implicitly assumed that b-quark (b) refers to both b-quark (b) and $\overline{b}$-antiquark ($\overline{b}$).

turbative QCD and serves to improve also the understanding of the proton fundamental structure. Furthermore, the measurement of this process provides a benchmark to probe and improve the modelling of Monte Carlo (MC) generators used to estimate this process in Higgs boson measurements and searches for new physics where it constitutes a frequent background. In particular the discovery of a neutral boson with a mass of about 125 GeV/c^2 consistent with the Higgs boson predicted by the SM at the ATLAS and CMS experiments at the LHC in 2012 motivates a precise examination of its properties and its interactions with SM particles. The Higgs boson decay into two b-quarks (H(bb)) is the most probable decay for a SM Higgs boson of 125 GeV/c^2 of mass, due to the large multijet background at the LHC it is cannot be studied in inclusive production and the leading production becomes the one in association with a W or Z boson (W/ZH)). The W+ 2 b-quarks process is one of the main backgrounds for the WH(bb) analysis, therefore my study provides crucial information to improve the precision of the WH(bb) measurement and consequently to improve the understanding of the Higgs-boson coupling with b-quark.

The final state of my analysis is constituted by a W boson that decay leptonically into a muon and a neutrino produced in association with two b-quarks. The b-quarks are experimentally reconstructed in the detector as b-jets, that are cones of hadrons with features related to the properties of b-quarks. Therefore the processes is often referred to as W+2b-jets.

The process under study is very similar to the production of a Z boson in association with 2 b-quarks. While differential measurements of Z boson production with 2 b-jets have been published by the ATLAS and CMS Collaborations, there is a lack of similar measurements for the W+ 2 b-jets production. The main challenge lies in reducing the significant background from events originating from top quarks pair production ($t\bar{t}$).

I developed an innovative analysis based on a multivariate technique that surpasses traditional methods relying on cascade cuts on single observables. Exploiting the Neural Network's ability to identify non-linear correlations among

input variables, a classifier is constructed to distinguish the kinematics of signal events from background events. To be more precise, two different strategies on jet counting for W boson event candidates have been developed, one more inclusive (allowing additional jets in the final state with respect to the 2 b-jets required by the signature of the process) and one more exclusive (vetoing additional jets in the final state), this idea is inspired by the fact that $t\bar{t}$ events are expected to present an higher number of jets in the final state than signal events. For each of the two regions, a diverse Neural Network has been developed. This technique constitutes the first step for the first measurement of differential cross-sections of the W+2 b-jets process.

This study is based on simulated state-of-the-art MC samples of the signal and the backgrounds officially produced by the ATLAS Collaboration. The simulated samples reproduce the experimental conditions of ATLAS dataset available at the proton-proton centre-of-mass energy of $\sqrt{s}$= 13 TeV. It is the full Run-2 dataset collected by the experiment from 2015 until 2018. This dataset represents the largest and best understood dataset available so far at the ATLAS experiment and corresponds to an integrated luminosity of 140 fb^{-1}. The simulated samples are processed using the full detector simulation and analyses with the same reconstruction tools used for real collision data; therefore, they constitute a faithful emulation of real collision data.

One of the key aspects of this study is the usage of different MC generators for the simulation of the signal in order to estimate the impact of the different technical choices made at the generation level on the proposed machine learning strategy.

This thesis is structured as follows: Chapter 1 provides an overview of the SM particle physics, including a detailed discussion of the signal process. It also includes an overview of the key aspects of proton-proton collisions and their simulation, and a section dedicated to the status of art on the measurement of a W/Z+b-jets. Chapter 2 describes the LHC and the ATLAS detector, highlighting their key components and functionalities and the main aspects of the reconstruction of physics objects at ATLAS, focusing on muon, jets,

b-jets and missing transverse energy, which are the objects used in this thesis. Chapter 3 focuses on the description of simulated signal and background processes, on the criteria to select W-boson and b-jets candidates, and on the ones that define the inclusive and exclusive regions of the analysis. In this chapter the primary signal observables that will be exploited in future differential cross-section measurements are also discussed. Chapter 4 presents the development of Neural Networks for top background rejection, detailing the architecture and optimization strategies for exclusive and inclusive selections, and assessing the impact of Neural Network cuts on primary signal observables. In the last chapter, the main conclusions of the work are drawn. Finally, the appendix offers a detailed dissertation on machine learning concepts relevant to the analysis, such as supervised and unsupervised learning, artificial Neural Networks, feature scaling techniques, and model evaluation metrics.

Chapter 1

Theoretical and phenomenological overview

Over the past five decades, the Standard Model of Particle Physics (SM) has emerged as a highly successful theoretical framework, akin to the periodic table in chemistry and Newton's laws in classical mechanic. It encloses all our understanding of elementary particles and their interactions, predicting and correlating a wide range of phenomena [1]. The model successfully combines the strong, weak, and electromagnetic forces through a quantum gauge field theory. It is a quantum gauge field theory, that is the gauge symmetry group $SU(3) \times SU(2) \times U(1)$. $SU(3)$ is the symmetry group of the strong interactions, while $SU(2) \times U(1)$ is the symmetry group of the electroweak interactions. The model has been experimentally verified across various energy scales and processes governed by strong and electroweak interactions. However despite the model's success in explaining many phenomena, it fails to include gravity, suggesting the need for possible modifications and the exploration of more fundamental theories.

Section 1.1 described the key aspects of the SM theory, covering the classification of fundamental particles in fermions and bosons and presenting the formalism of electromagnetic, weak, and strong interactions between particles. Section 1.2 central aspects of the phenomenology of proton-proton collisions, including the factorization theorem, the jet formation, and the role of MC gen-

erators in simulating these collisions. Section 1.3 covers the theoretical aspects of W+b-jets production important for the simulation of this process. Finally Section 1.4 reviews the relevant measurements of W/Z+b-jets published so far at the LHC.

1.1 The SM of Particle Physics

In this section the key aspects of the SM theory relevant for this work of thesis are discussed. The classification of fundamental particles in fermions and bosons in presented in section 1.1.1, while the formalism of the interactions is discussed in section 1.1.2.

1.1.1 Fundamental particles

The SM of Particle Physics describes the fundamental particles that make up the universe and the forces through which they interact. These particles can be broadly categorized into two main types: fermions and bosons (Figure 1.1).

Fermions

Fermions are massive particles with a half integer spin ($\frac{1}{2}, \frac{3}{2}, ...$), fundamental fermions (all particles on Figure 1.1) have spin $\frac{1}{2}$. For each elementary SM fermion, there exists a corresponding antiparticle with opposing charges and otherwise identical properties. Fermions are subdivided into quarks and leptons. There are three families of leptons (electron, muon and tau and their corresponding neutrinos), each consisting of charged lepton (the charge is typically identified with e and a corresponding chargeless and massless neutrino. The charged leptons are subject to both the electromagnetic and the weak interaction while the neutrinos can only interact weakly. In SM interactions, the number of leptons per family are always conserved (particles count as 1, antiparticles as -1).

Quarks are electrically charged particles (with charge $\frac{2}{3}$ e or -$\frac{1}{3}$ e) that interact

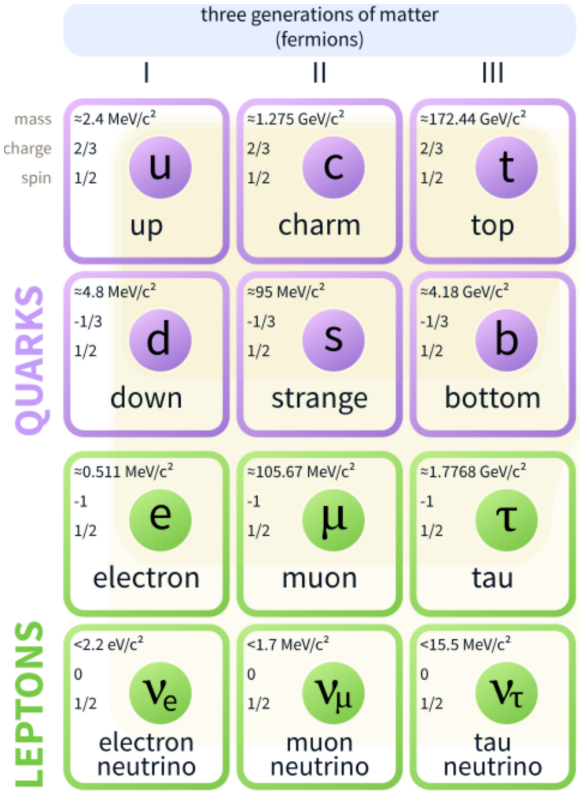

Figure 1.1: The three generation of fermions families in MS with their mass, charge and spin.

via the strong, the weak and the electromagnetic interaction. As with leptons, three families of quarks are predicted by the SM theory. Moving from the first to the last family either for leptons and quarks the mass of the particle increases, expect for neutrinos that are massless.

Bosons

Bosons are particles with integer spin. When two elementary particles interact with each other via one of the fundamental interactions, they do so via the exchange of a gauge boson. In SM theory there are four types of bosons: eight gluons, that are mediators of the strong interactions via Quantum Chromodynamics (QCD) theory, the photon and the W and the Z bosons, that are mediators of the electroweak interactions, and the Higgs boson that is introduced in the theory via the Higgs mechanism to explain how elementary particles get mass. While a comprehensive description of the mechanism is

beyond the scope of this thesis and can be found in [2]. The Higgs boson has been successfully discovered by both ATLAS and CMS experiments at the LHC in 2012 [3, 4]. Gluons, photons and W/Z bosons have spin one, the Higgs is a scalar with spin zero.

Gauge boson	**Spin**	**Mass (GeV/c^2)**
Photon	1	0
W-boson	1	80.4
Z-boson	1	91.2
Gluons	1	0
Higgs	0	125

Table 1.1: The Fundamental Bosons predicted by SM including their Spin and Masses.

1.1.2 Interactions

The SM relies on gauge symmetries to describe the interaction between particles mathematically. The vector bosons are gauge fields that need to be introduced to maintain local gauge invariance, which is in turn needed to keep the Lagrangian symmetric and with respect to the transformations of of a given group (i.e. $U(1)$ for the electromagnetic interaction, or $SU(3)$ for the strong interaction). This symmetry guarantees the conservation of a specific physics quantity in the interaction (i.e. the charge in the electromagnetic interaction, the color in the $SU(3)$ interaction. The SM Lagrangian is invariant under $U(1)$ $\times$ $SU(2)$ $\times$ $SU(3)$ transformations, which correspond to the electromagnetic, weak and strong interactions.

Electromagnetism

The concept of gauge invariance is already familiar in electromagnetism without the addition of quantum mechanics: the vector potentials $\phi = A^0$ and $\mathbf{A}$, associated to the electric and magnetic field respectively, are invariant

under the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \chi \tag{1.1}$$

The concept of gauge invariance translated into quantum mechanics becomes a request of invariance under the local phase transformation $U(1)$. This means that the Lagrangian must be invariant for a $U(1)$ transformation of the ψ field:

$$\psi(x) \rightarrow \psi'(x) = \hat{U}(x)\psi(x) = e^{iq\chi(x)}\psi(x) \tag{1.2}$$

However, the Dirac equation

$$L = (i\gamma^\mu \partial_\mu - m)\psi \tag{1.3}$$

is not invariant under such transformations, because it does not involve interactions. In order to become invariant, a covariant derivative including a gauge field, that is the field of the massless photon, must be introduced:

$$\mathcal{L} = \left[i\hbar c\overline{\psi}\gamma^\mu \partial_\mu \psi - mc^2\overline{\psi}\psi\right] - \left[\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu}\right] - \left(q\overline{\psi}\gamma^\mu\psi\right)A_\mu \tag{1.4}$$

Weak interaction and electroweak theory

Weak interactions were first studied by Enrico Fermi to justify the β-decay $n \rightarrow e^- + p + \overline{\nu}_e$. He proposed an effective point-like theory. Although valid at low energies, this approximation does not consider important features of this interaction, like the massive mediators. A quantum field theory based on a $V-A$ (vector-axial) structure and described by the $SU(2)_L$ symmetry group can be introduced, where the label L indicates the coupling with left-handed fermions only. Fermion fields are represented by the left-handed doublets L and the right-handed singlet R, which, for the first family of fermions, becomes:

$$L = \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L \qquad R = e_R \tag{1.5}$$

$$L = \begin{pmatrix} u \\ d \end{pmatrix}_L \qquad R = u_R, d_R \tag{1.6}$$

The mediators of the weak interaction are three massive bosons: W^+ and W^- that carry electric charge ($\pm$ 1) and Z which is neutral. These particle were initially predicted without mass, in contrast with experimental evidence of $MZ/W >> 0$. The huge mass of these bosons explains the short range of the weak force.

In the late 1960s, Weinberg, Salam and Glashow unified the electromagnetic and weak interactions in the electroweak theory, described by the $SU(2)_L \times U(1)_Y$ symmetry group. Following the same procedure adopted for the QED and QCD theories, the Lagrangian of the left-handed L and right-handed R fermion fields is required to be invariant under global and local transformation of the gauge group. In order to ensure the local invariance, the ∂_μ derivative is replaced by the covariant form. With this approach gauge bosons of the $SU(2)_L$ and $U(1)_Y$ groups are introduced. In the theory the initial Lagrangian of electroweak interactions considers massless fermions and massless gauge fields, in contradiction with experimental observations. The mass terms are generated introducing a $SU(2)_L$ doublet of complex scalar field φ, the Higgs field. Thought a mathematical processes, known as Higgs mechanism, one generates the spontaneous symmetry breaking, that has as main result the appearance of the massive W and Z bosons, the massless photon and the massive Higgs boson.

Strong interaction

The strong interaction is the most intense of the four fundamental forces. It is described by Quantum Chromodynamics (QCD) that is the quantum field theory of the strong interactions, symmetric with respect to gauge rotation of $SU(3)_C$ group. In order to determine the QCD Lagrangian, it is important to consider that each quark field can occur in three colours. Again in the theory a covariant derivative is introduced to guarantee the invariance under $SU(3)_C$, this procedure consequently introduces the the gauge field of the strong interaction, the eight gluons. In QCD there isn't the breaking of symmetry that applies to electroweak interactions, therefore it is a perfect symmetry and

as a consequence, gluons are massless bosons in the SM. Another feature of the theory is that it predicts self-interaction between the mediators, which is why Feynman diagrams that involve 3 or 4 gluons are allowed, as seen in Figure 1.2. Gluons only couple to particles that have non-zero color charges, that is quarks. The QCD coupling constant α_s has a running behaviour as a function of the energy. The divergence of the strong coupling at low energy is an indication of confinement, the mechanism that explains why the free quarks and the gluons have never been observed. Due to confinement, QCD bounds quarks together in zero-color bound states called hadrons: two quarks form a meson, three quarks form a baryon. Protons and neutron for example are baryons. On the other hand, at very high energies when quarks are very close, the intensity of QCD is greatly reduced, and quarks and gluons are free. This concept is called asymptotic freedom. It allows to develop the perturbative QCD (pQCD) approach that is valid at the high energies tested at colliders, therefore experimentalists compare the results of their measurements at colliders with the pQCD predictions.

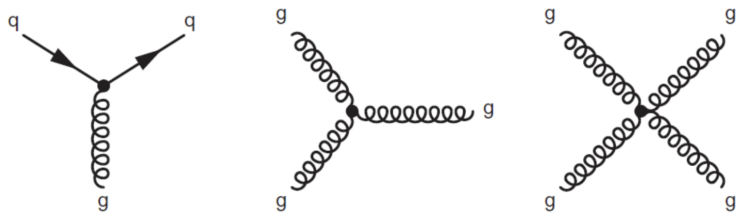

Figure 1.2: The three QCD vertices involving quarks and gluons predicted by $SU(3)$ invariance.

1.2 Phenomenology of proton-proton collisions

Proton-proton collisions are extremely complex phenomena due to the composite nature of the protons and the very complicated dynamics of the interactions between the partons that constitute the proton. The momentum distribution of the partons inside the proton depends on the energy and it is parameterised in the parton distribution functions (PDFs), which cannot be

calculated directly from calculations but have to be derived from global fit to experimental data. When two hadrons collide, several partons may interact. Most of the inelastic processes are soft and give rise to low-momentum final state particles, but occasionally a hard inelastic parton-parton scattering occurs through a large momentum transfer. Interesting physics processes, such as Higgs or W/Z boson production, proceed via hard scatterings, while the soft component always accompany the event. The result of the interactions of other partons in the same proton-proton collision, additional to the parton-parton interaction that generates the hard scattering, is called *underlying event*. Moreover either the incoming or outgoing partons may radiate and all outgoing partons fragment to produce the particles that constitutes the final state observed at colliders. Therefore in the study of proton-proton collisions at high energy both perturbative and non-perturbative QCD processes are involved.

1.2.1 The factorisation theorem

The QCD factorization theorem [5] allows to separate the cross-section of proton-proton collisions into two independent terms: the short-distance parton cross-section, that is process-dependent and it is calculable in pQCD, and the universal long-distance term involving PDFs. The factorization theorem states that the cross-section of the collision process between two hadrons with four momentum p_1 and p_2 can be written as follows:

$$\sigma(p_1,p_2) = \sum_{i,j=q,\bar{q},g} \int dx_1 \int dx_2 f_i(x_1,\mu_F^2) f_j(x_2,\mu_F^2) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}) \tag{1.7}$$

where x_1 and x_2 are the proton momentum fractions carried by the two partons respectively; Q^2 is the transferred momentum; $f_i(x_1,\mu_F^2)$ and $f_j(x_2,\mu_F^2)$ are the PDFs; $f_i(x_1,\mu_F^2)$ represents the probability for the parton i to carry a fraction x_1 of the proton momentum at the factorisation scale μ_F. $f_j(x_2,\mu_F^2)$ represents the probability for the parton j to carry a fraction x_2 of the proton momentum at the same scale; α_s is the strength of the strong force; $\hat{\sigma}_{ij}$ is the interaction

cross-section between the two partons involved in the interaction.
The cross-section $\hat{\sigma}_{ij}$ can be written in a perturbative expansion as:

$$\hat{\sigma}_{ij} = \hat{\sigma}_{ij}^{LO} + \alpha_s \hat{\sigma}_{ij}^{NLO} + \alpha_s^2 \hat{\sigma}_{ij}^{NNLO} + \alpha_s^3 \hat{\sigma}_{ij}^{N^3LO} + O(\alpha_s^4) \tag{1.8}$$

where the leading-order, LO, is the lowest order in the pQCD calculation, followed by the next-to-leading order, NLO, the next-to-next-to-leading-order, $NNLO$, and so on up to a certain perturbative order.
Even though PDFs are extremely non-perturbative and therefore incalculable, they are approximately universal and, once determined in a given process from data obtained in one experiment, can be applied to other hadronic processes. The evolution of the PDFs at any scale μ_F^2 is theoretically predicted by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations ([6]-[8])

1.2.2 Jet formation

Quarks and gluons are abundantly produced in proton-proton collisions at high energy. As mentioned in the previous section due to the confinement, quarks and gluons have never been observed as individual stable particles.
Soon after two quarks are produced in collisions from gluon splitting ($g \to qq$), the distance between them increases and as the distance increases, the energy required to separate them grows linearly, approximately 1 GeV per femtometer. This makes energetically more favorable for quarks rearranging into a colorless state as they move apart. When two quarks drift apart with enough kinetic energy, the potential energy in the gluon string between them eventually exceeds the mass of a quark-antiquark pair. At this point, a new pair is produced from the vacuum, causing the gluon string to split and form two colorless mesons. This process continues until the remaining kinetic energy is insufficient to produce more quark-antiquark pairs. This phenomenon is known as "hadronization" or jet formation.
In high-energy collisions, quarks are produced and undergo hadronization, forming cone of hadrons called jets. The reconstruction of these objects is possible through the use of jet algorithms. This phenomenon, observed in ex-

periments such as those conducted at the LHC, provides valuable insights into the dynamics of particle formation and the properties of quarks and gluons at extreme energies. Within the framework of the SM, the study of hadronic jets offers a deeper understanding of the underlying mechanisms governing particle interactions.

The properties of these jets vary depending on the initiating quark, with "light" jets originating from up or down quarks and "heavy flavor" jets originating from bottom or charm quarks. Discriminating between these jet types, as observed in experiments like those conducted at the ATLAS detector, provides crucial insights into the underlying physics processes.

1.2.3 Monte Carlo Generators

As anticipated the simulation of processes in proton-proton collisions is extremely complex. While fixed order calculations in pQCD are available only up to a limited order in α_s and just focus on the hard-scattering aspect providing results at parton level, MC generators aim to the complete simulation of the processes and their predictions can be directly compared to experimental data. Typically the hard scattering of the two partons is described by a matrix element (ME) that describes the cross-section at a fixed order in pQCD. In the state-of-art MC generators a variety of processes are available with NLO MEs with multiple partons in the final state. Shortly after being produced, hard partons repeatedly radiate low-energy and collinear gluons, that might split in quark and anti-quark pair. This process in MC is called Parton Shower (PS). PS methods in MCs simulate the contributions of higher order that is missing in the ME. The hadronisation in MCs is described by phenomenological models (i.e. cluster model and string model). Also underlying event is included in MCs generators. ME supplemented by PS together with hadronisation models and underlying event description form the basis of widely used general-purpose MC simulation programs such as Pythia, Herwig, Sherpa, and Madgraph which give realistic descriptions of events with all final-state particles. These are the MCs extensively used at the LHC.

1.3 Theoretical aspects of W+b-jets process

The theoretical understanding of W+b-jets production at high-energy within the pQDC framework has advanced significantly over past three decades. Initially, this research focused on its role of W+b-quarks process a background for top-quark final states ($t \to Wb$), and later for Higgs searches ($WH \to Wb\bar{b}$). While currently with the large integrated luminosity of the LHC these processes can be measured accurately at the LHC and this trigger the need of detailed theoretical calculations.

The leading order (LO) production of $W + b\bar{b}$ events is initiated through the process $q\bar{q}' \to Wb\bar{b}$, as depicted in Figure 1.3 [9].

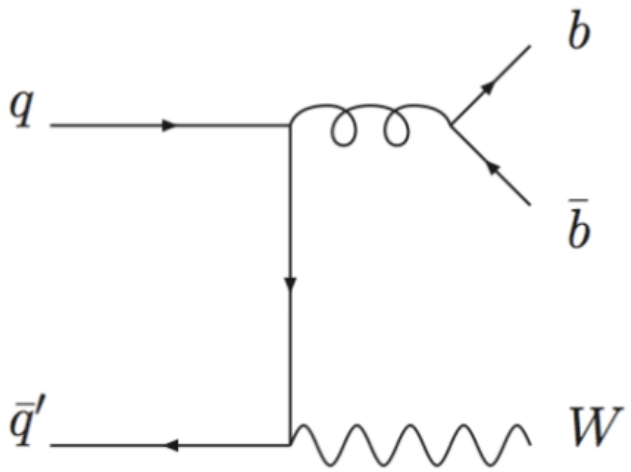

Figure 1.3: Diagram for W+2b-quarks at LO

In this calculation the b-quark is considered massive ($m_b = 4.7$ GeV). This *LO* calculation, as typical of *LO* approach, exhibits significant energy scale dependence.

The scale dependence is reduced when diagrams with additional parton emissions are included as shown in Figure 1.4. These additional diagrams can be seen as approximating an NLO calculation, without including the 1-loop effects.

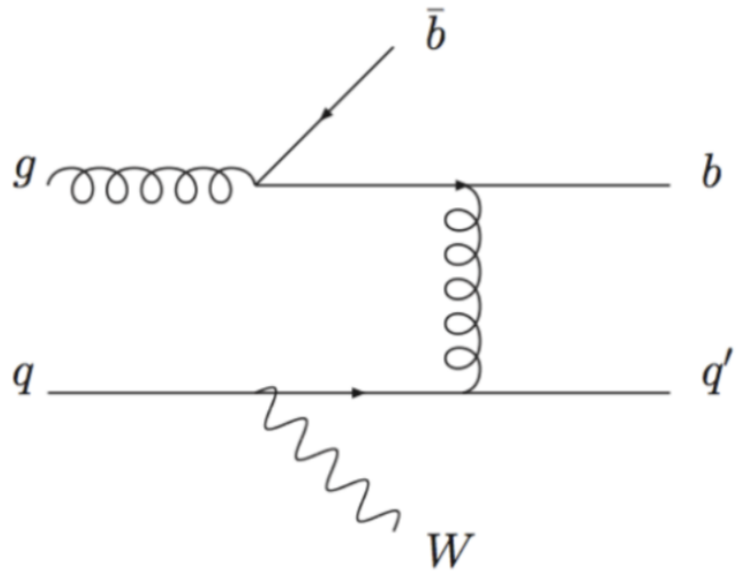

Figure 1.4: Diagram for W $+2b$ with one additional parton emission.

Currently, next-to-leading order (NLO) perturbative QCD calculations for the W+b-jets process have been integrated into MC simulations.
State-of-the-art MC predictions for W+b-jets are typically based on NLO MEx combined with parton showers (PS). MC predictions are available for different flavour and mass schemes. In particular predictions might be derived in either a 4-flavour number scheme (4FNS) or a 5-flavour number scheme (5FNS). In the 4FNS, only u, d, c, s are considered as initial-state quarks, b-quarks are considered massive and since they are more massive than the proton, they do not contribute to the PDFs of the proton and, in the partonic ME they only appear via gluon splitting $g \rightarrow bb$ at high Q^2. In the 5FNS, b-quark is treated as massless. This approach is valid whenever the energy scale is higher than the b-quark mass. Therefore b-quarks contribute to the PDFs of the proton. The ambiguity among the 5FNS and 4 FNS schemes is an intrinsic property of the calculation and is expected to reduce with the inclusion of higher order perturbative corrections.
For analyses with the full Run-2 dataset (proton-proton data at $\sqrt{s}$= 13 TeV with 140 fb^{-1}) ATLAS is employing fully simulated MC samples produced

with generators based on NLO MEs combined with PS using the 5FNS. Indeed the 5FNS shows better agreement with data in Z+b-jets measurements (as discussed in the next section 1.5). In particular two generators are used each with a specific set-up: Sherpa 2.2.11 and MadGraph5_aMC@NLO (v2.6.5). Sherpa 2.2.11 computes matrix elements at NLO accuracy in the strong coupling constant for up to 2 additional partons (meaning that the ME generates processes with a W boson and up to two jets) and LO-accurate MEs for up to 5 partons (meaning up to W+5 jets). MadGraph5_aMC@NLO (v2.6.5) is combined with Pythia (v8.245) parton shower using a merging procedure called FxFx. The MadGraph5_aMC@NLO (v2.6.5) program generates MEs for W-boson production, with up to 3 additional partons in the final state at NLO accuracy in the strong coupling constant. Further details are provided in Chapter 3.

Recently the first calculation for the production in proton-proton collisions of a W boson in association with a massive bottom (b) quark–antiquark pair at NNLO in pQCD has been published ([10],[11]). It will be extremely interesting in the near future compare the results of this high level fixed order calculation with W+2b-jets measurements at the LHC.

1.4 W+b-jets measurements: state of the art

ATLAS and CMS Collaborations at the LHC performed various measurements of W/Z+b-jets processes so far using at the beginning the Run-1 dataset collected between 2010 and 2012 at the proton-proton centre-of-mass energy of $\sqrt{s}$ = 7 TeV (about 5 fb^{-1} of data for each experiment) and of 8 TeV (about 20 fb^{-1} of data per experiment) afterward the huge Run-2 dataset collected between 2015 and 2018 at $\sqrt{s}$ =13 TeV (about 140 fb^{-1}). Below are briefly discussed the latest measurements.

The CMS Collaboration conducted two studies focusing on the measurement of the production cross-section of W+b-jets with Run-1 data. The first study, performed using a proton-proton dataset of 5.0 fb^{-1} at $\sqrt{s}$= 7 TeV, focuses

on events where the W boson decays into a muon and a neutrino ($W \to \mu\nu$) and it is produced in association with 2b-jets [12]. The second study is done using a proton-proton dataset of 19.8 fb^{-1} at $\sqrt{s}$= 8 TeV and it is extended also to the electron channel in addition to the muon one ($W \to \ell\nu$ with $\ell = \mu$ or e). In both cases only the inclusive W+2b-jets cross-section has been measured in a fiducial phase space close to the detector level one to minimise the uncertainties.

ATLAS Collaboration published a study [13] using 5.0 fb^{-1} of proton-proton collisions data at $\sqrt{s}$= 7 TeV where the differential cross-sections of a W-boson production in association with b-jets has been performed, selecting a first fiducial phase-space with exactly 1b-jet and not additional jets and a second fiducial phase-space with 2 jets of which at least one being a b-jet. The study focuses on events where the W boson decays into a lepton (μ or e) and a neutrino. W+ 2b-jets events are expected to represent only the 10% of the total number of events in the 2-jet fiducial region.

Neither ATLAS nor CMS has published measurements yet of the differential cross-section of the W+2b-jets process, due to very large top background. Moreover they aren't available neither inclusive measurements at $\sqrt{s}$=13 TeV, this because while the W+2b-jets cross-section scales linearly with the collision energy of LHC, the scaling of $t\bar{t}$ pair production is quadratic, therefore at 13 TeV the signal cross-section is about 2 times the one at 7 TeV, while the $t\bar{t}$ background is about a factor 4, making the measurement more challenging.

My thesis work is meant as the first step for the first ATLAS differential cross-section measurement of W+ 2b-jets with the largest and best known dataset available so far, full Run-2 dataset.

The twin process of the W+2b-jets production is the Z boson production in association with 2 b-jets (Z+2b-jet), with the Z boson decaying leptonically ($Z \to \ell\ell$). The advantage of this second process is that it is affected by large backgrounds and therefore it has been already measured differentially by ATLAS and CMS. In particular ATLAS published a result with 35.6 fb^{-1} proton-proton collision data at $\sqrt{s}$ = 13 TeV [14] and recently submitted the

follow-up with 140 fb^{-1} at the same collision energy [15], the corresponding CMS measurement has been already published [16]. In these publications both experiments present differential cross-section distributions for Z+2b-jets events as functions of various kinematic observables that are useful for precision tests of pQCD predictions, to improve understanding on PDFs and to improve the modelling of this process in MC generators used in Higgs studies and New Physics studies where Z+2b-jets process constitutes a background.

It is important to mention that also W ($\rightarrow \ell\nu$)+2b-jets and Z($\rightarrow \ell\ell$)+2b-jets process constitutes one of the main backgrounds in the study of the Higgs boson decay into two b-jets ($H \rightarrow b\bar{b}$) in associated production with a W/Z boson (W/ZH). ATLAS Collaboration published a measurement of $H \rightarrow b\bar{b}$ decay in W/ZH production using 140 fb^{-1} proton-proton collision data at $\sqrt{s}$ = 13 TeV in 2022 [9]. The study establishes the production of the Higgs boson in association with a W or Z boson with observed significances of 4.0 and 5.3 standard deviations respectively, with a signal strength compatible with the SM predictions. The uncertainties on the modelling of W/Z+2b-jets result among the main limiting factors of this measurement. Improving the precision of Higgs boson measurements, especially its couplings with b-quarks, will depend significantly on advancing our experimental understanding of W/Z boson production in association with b-jets. Therefore this thesis work, in addition of its importance for validating QCD predictions at the LHC scales, has also a central importance in the studies of the Higgs boson at the LHC.

Chapter 2

LHC and ATLAS

The LHC is currently the largest and most powerful particle accelerator ever built. Four main detectors are situated along the LHC collider's ring, including the ATLAS detector.
In this chapter, the Large Hadron Collider is introduced in Section 2.1. The ATLAS experiment and its sub-detectors are presented in Section 2.2. The reconstruction and identification tools of physics objects developed within the ATLAS experiment are briefly discussed in Section 2.3, focusing on the objects relevant for this thesis.

2.1 The LHC

LHC is a circular ring about 27 kilometers long, located approximately 100 m underground. It's located at CERN in Geneva, across the France-Switzerland border. As a hadron–hadron collider, it is capable of accelerating and colliding protons or heavy ions. In particular it has been designed to collide protons at a maximum centre-of-mass energy of $\sqrt{s}$= 14 TeV.
Protons are produced by ionizing hydrogen gas and accelerated by four different accelerators to an energy of 450 GeV and finally injected into the LHC ring. The LHC further accelerates the two proton beams in opposite directions before colliding them at four interaction points distributed along its ring. At each interaction point, a detector is housed that records the collisions.

ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) are the two multi-purpouse experiments of LHC, while two specific experiments LHCb (Large Hadron Collider beauty) and ALICE (A Large Ion Collider Experiment) are focused respectively on studies of the b-quark and of the heavy ion collisions. Figure 2.1 shows the schematic illustration of the LHC accelerator chain and its experiments. The LHC offers an unprecedented physics potential, from precise measurements of SM processes and their parameters as well as the Higgs boson properties to the search for new physics phenomena.

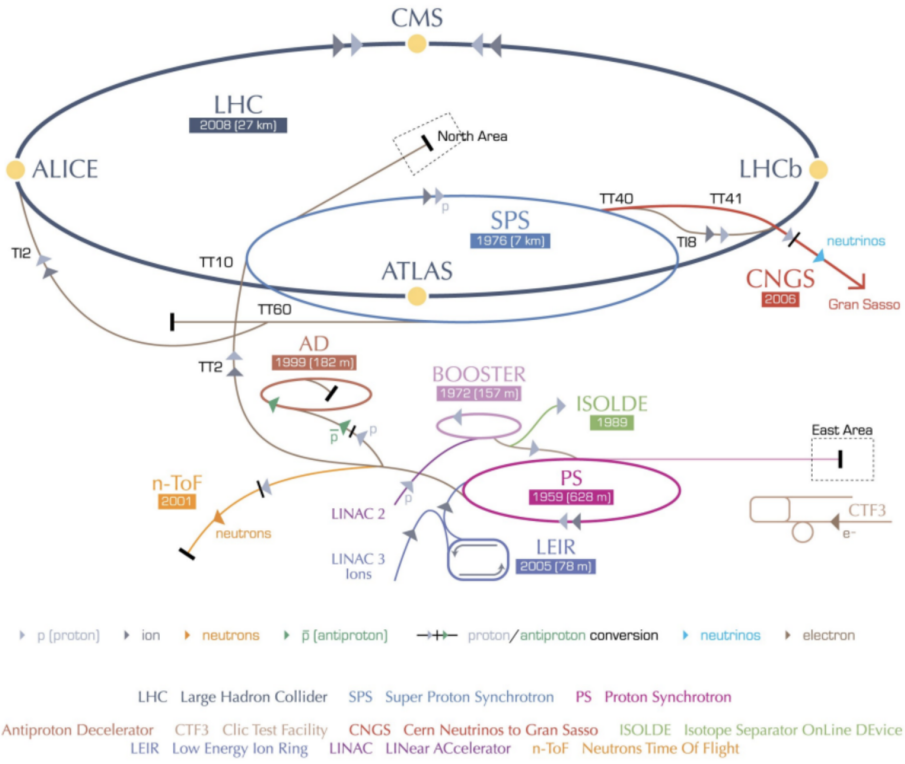

Figure 2.1: Layout of CERN Accelerating Complex. ATLAS and the other three LHC detectors are also shown.

At the LHC energies , the protons travel at speeds close to the speed of light. Protons are accelerated packed into bunches. The bunches contain approximately 10^{11} protons each and fill the LHC accelerator, separated in time by only 25 ns, corresponding to a distance of approximately 7.5 m. The instantaneous luminosity of the machine L is the number of collisions per unit of time per unit of area and it is defined as: $L = R/\sigma$, where R is the event

rate of a given physics process and σ is its cross-section.

The integrated luminosity, L_{int}, is obtained by integrating the instantaneous luminosity over a certain time interval t:

$$L_{int} = \int L dt \tag{2.1}$$

Therefore the number of events of a given physics process with cross-section produced with a given integrated luminosity L_{int} is

$$N = \sigma L_{int} = \sigma \int L dt \tag{2.2}$$

The instantaneous luminosity depends on several variables, including the number of protons in each bunch and the intensity of the beam. The design instantaneous luminosity of LHC is L= 10^{34} cm^{-2} s^{-1}. Table 2.1 shows the most important parameters of the LHC. LHC already completed two data-taking campaigns: Run-1 (from 2010 up to 2012) where the proton collision energy was $\sqrt{s}$= 7 and 8 TeV, Run-2 (from 2015 to 2018) where the proton collision energy was $\sqrt{s}$ = 13 TeV. Since 2022, the Run-3 campaign is on-going at $\sqrt{s}$ = 13.6 TeV. Since 2018, LHC reached the double of its design value, accumulating large datasets.

This thesis focus on the Run-2 dataset, that currently is the largest available. During Run-2, the ATLAS detector collected 147 fb^{-1} from the 156 fb^{-1} of proton-proton collisions delivered by the LHC. The amount of data deemed to be of good quality for physics analyses was about 140 fb^{-1}.

As a consequence of the high luminosity and large proton-proton cross-section, multiple proton-proton collisions simultaneously happen in a single proton bunch crossing. During the LHC Run-2, the average number of interactions per bunch crossing, $\langle\mu\rangle$ (pile-up), was 33.7.

Every collected event at the LHC is accompanied by pile-up and by the underlying event introduced in the previous Chapter. The underlying event can be thought of as all what is seen in a single proton-proton collision which is not coming from the primary hard scattering process of the two colliding partons. It represents the hadronisation of the collided proton remnants, as well as all spectator interactions outside the hard scattering event of interest.

Parameter	**Value**
Circumference	26659 m
Dipole operating temperature	1.9 K
Number of dipole magnets	1232
Number of quadrupole magnets	858
Number of RF cavities	eight per beam
Bunch spacing	25 ns (or multiples thereof)
Peak dipole magnetic field at 7 TeV	8.33 T
Design Luminosity	$10^{34} cm^{-2} s^{-1}$
Energy per beam, protons	7, 8 TeV (Run-1), 13 TeV (Run-2), 13.6 TeV (Run-3)
Bunch crossing frequency	40 MHz
Number of bunches per proton beam (max.)	2808
Number of protons per bunch	1.15 $\times 10^{11}$ (25ns) or 1.6 $\times 10^{11}$ (50ns)
Number of collisions per second	600 millions
Total crossing angle (at the interaction point)	300 μrad

Table 2.1: Main LHC Parameters

2.2 The ATLAS detector

ATLAS, is one of two multi-multipurpose experiments of LHC. It is an international collaboration involving 170 institutes spread all over the world. It is specifically designed to make precise measurements of processes predicted by the SM, including the physics of the Higgs boson, and to explore new phenomena not predicted by the SM, the so-called beyhond SM searches. This complex instrument has a cylindrical shape and is quite massive, weighting over 7000 tons. It is 46 meters long and has a height of 25 meters [17].

The innermost detector of ATLAS is the Inner Detector, it is responsible for

tracking charged particles right after collisions. It is surrounded by a magnetic field that, bending the track of charged particles, allows to identify the charges and the momenta of the particles. Around this core, there is a calorimeter. The calorimeter is constituted by two detectors: the electromagnetic calorimeter for the measurement of electrons and photons (the electromagnetic showers) and the hadronic calorimeter for the measurement of charged and neutral hadrons (the hadronic showers). The outermost part houses the muon spectrometer, enveloped in a toroidal magnetic field. All these components work together on a significant mission.

The coordinate system of ATLAS, seen in Figure 2.2, is a right-handed coordinate system with the x-axis pointing towards the centre of the LHC ring, the z-axis along the tunnel/beam (counter clockwise) seen from above and the y-axis points upward.

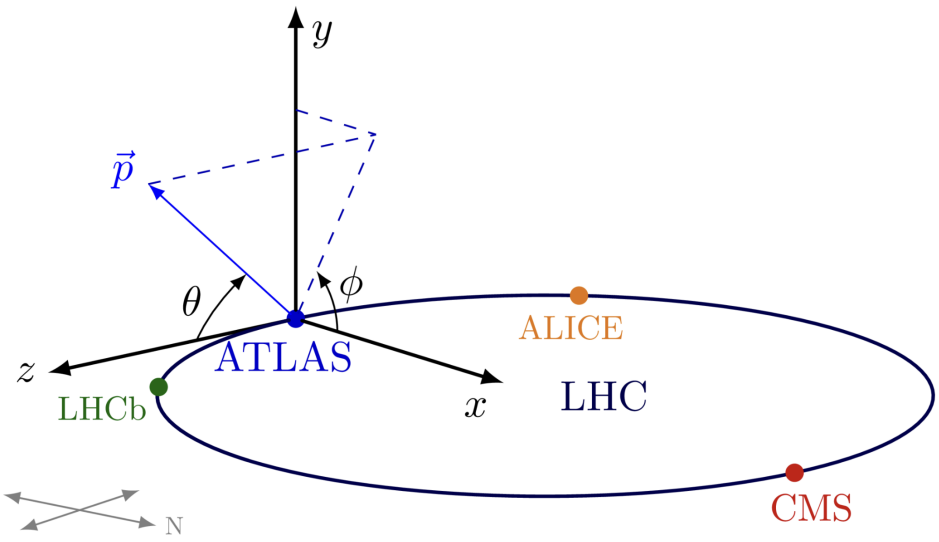

Figure 2.2: Coordinates system of the ATLAS experiment.

The origin is defined as the geometric center of the detector. Polar coordinates (R, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity of a particle is defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.3}$$

where θ is the polar angle measured from the positive z-axis.

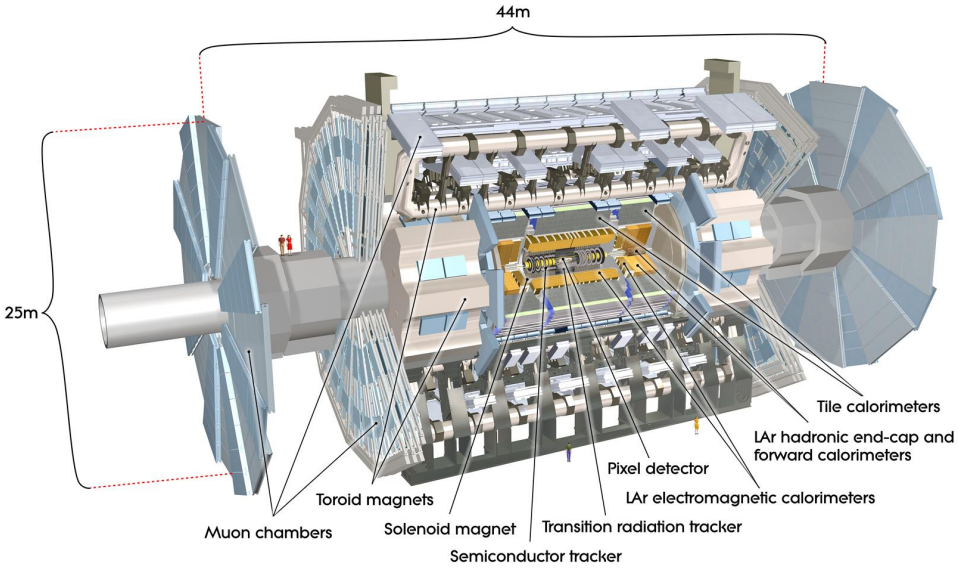

Figure 2.3: A scheme of the ATLAS detector.

It is quite common to calculate the distance between particles and jets in the (η, ϕ) space:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.4}$$

Two physics quantities which are often used in ATLAS analysis and therefore extensively mentioned in this thesis are: the momentum of particles in the transverse plane, referred to as transverse momentum and indicated with p_T and the rapidity of the particles indicated as y. For particles with null masses, "y" coincides with η. That is why for very light particles (i.e. electron and muons) it is used η that has clear kinematical interpretation (being related to the direction of the particle with respect to the z-axis).
Figure 2.3 shows a visual representation of ATLAS and its subdetectors. In the following sections the ATLAS magnetic field and variuos sub-detectors are presented in detail.

2.2.1 Magnet Systems

ATLAS has a large magnet system that bends the paths of charged particles, allowing the measurement of their momentum. The ATLAS magnet

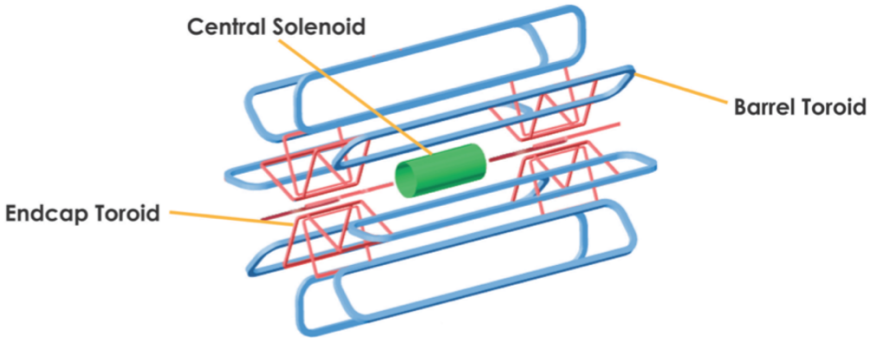

Figure 2.4: Schematic illustration of the the magnetic system of the ATLAS experiment.

system comprises four extensive superconducting magnets generating a magnetic field that encompasses a volume of about 12,000 m^3. This system includes a central solenoid, a barrel toroid, and two endcap toroids (Figure 2.4). The central solenoid encapsulates the Inner Detector (ID) and generates a 2 T axial magnetic field, which alters the paths of charged particles in the transverse plane, enabling the measurement of transverse momentum. This solenoid, weighing 5.4 tonnes and measuring 5.8 meters in length, has inner and outer diameters of 2.46 m and 2.56 m, respectively.

The barrel and endcap toroid magnet systems produce magnetic fields of a maximum value of 0.5 T and 1 T, respectively. Their role is to deviate the paths of charged particles entering the muon spectrometer. The bending plane of the magnetic field is r-z. The barrel toroid, weighing 840 tonnes and measuring 25.3 meters in length, has inner and outer radii of 9.4 m and 20.1 m, respectively. The endcap toroids, totaling 240 tonnes, are 5 meters long, with inner and outer radii of 1.63 m and 10.7 m, respectively.

2.2.2 Inner Detector (ID)

The Inner Detector is the ATLAS sub-detector closer to the beam pipe and therefore to the interaction point [18]. It consists of three different sub-detectors all immersed in a magnetic field parallel to the beam axis. The Inner Detector measures the direction, momentum, and charge of electrically-

charged particles produced in each proton-proton collision in the range $|\eta| < 2.5$ by combining information from these three sub-detectors. The ID consists of a cylindrical central region (full coverage for $|\eta| < 1.5$) arranged around the beam pipe, and two endcaps. Disks in the endcap region are placed perpendicular to the beam axis, covering $1.5 < |\eta| < 2.5$.

The main components of the Inner Detector are: Pixel Detector, Semiconductor Tracker(SCT), and Transition Radiation Tracker(TRT) (Figure 2.5).

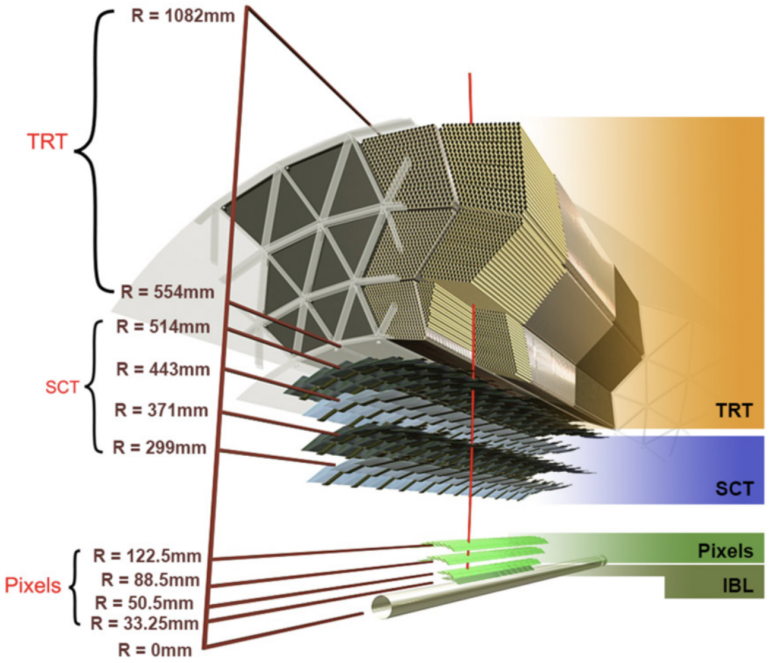

Figure 2.5: Layout of ATLAS inner detector.

Pixel Detector

The Pixel Detector is the first detection point in the ATLAS experiment. It has four layers built of minuscule silicon pixels, each one has the size of a sand grain achieving an excellent resolution of 14 μm $(r - \phi)$ $\times$ 115 μm (z). From Run-2 data-taking the ID includes a new innermost layer, the insertable B-layer (IBL), situated at a distance of only 3.3 cm from the LHC beam line, while the remaining three layers of the pixel system are located at mean radii of 50.5, 88.5, and 122.5 mm respectively. The coverage in the endcap region is enhanced by three disks on either side of the interaction point. The innermost

layers' pixels (IBL) measure 50 (r-phi) $\times$ 250 μm^2 (z), while the outermost layers' pixels measure 50 (r-phi) $\times$ 400 μm^2 (z). It hosts 1744 pixel-sensor modules, and each module contains 46 080 pixels. The Pixel Detector has about 80 million electronic channels. It uses 15 kW of power and covers an area of about 1.9 m^2 in silicon.

Semiconductor Tracker

The Semiconductor Tracker surrounds the Pixel Detector and is crucial for detecting and reconstructing the tracks of charged particles generated in collisions. It consists of over 4,000 modules containing 6 million "micro-strips" of silicon sensors. Specifically, it includes 4,088 two-sided modules and over 6 million implanted readout strips (equivalent to 6 million channels). Each track crosses four (eight) space points located on four (eight) cylindrical double sensor layers for a total of eight (eighteen) strips of the SCT in the barrel (endcaps). In this way, there is access to the two-dimensional information taken from the grid that is created. The silicon strip detector attains a resolution of 17 μm in the transverse plane and 580 μm along the direction of the z-axis.

Transition Radiation Tracker

The third and final sub-detector of the Inner Detector is the Transition Radiation Tracker (TRT). The TRT consists of 300,000 thin-walled drift tubes, known as "straws". The tubes are located both in the barrel region and the end-cap regions. In the former the tubes are tangent to each other and parallel to the direction of the beam, while in the latter the tubes are arranged radially on the caps and vertically to the beam axis. Each straw, with a diameter of just 4 mm, contains a 30 μm gold-plated tungsten wire at its center and is filled with a gas mixture. There are 50,000 straws in the barrel, each 144 cm long, and 250,000 straws in the endcaps, each 39 cm long. The angular coverage of TRT is restricted to $|\eta| < 2.0$ with a resolution of 130 μm in the transverse $(R - \varphi)$ plane in the barrel and a resolution of 130 μm along the z-axis in the

endcap. The TRT is designed to measure the transition radiation, which is produced when a charged particle passes through the straws. When charged particles pass through the straws, they ionize the gas, creating a detectable electric signal that helps reconstruct their tracks. Additionally, the transition radiation provides information about the particle type, identifying whether it is an electron or pion. The TRT features 350,000 read-out channels and occupies a volume of 12 m^3.

2.2.3 Calorimeter System

It measures the energy a particle loses as it passes through the detector. They are designed to absorb most of the particles coming from a collision, forcing them to deposit all of their energy and stop within the detector. ATLAS calorimeters are calorimeters constituted of layers of an "absorbing" high-density material that stops incoming particles, interleaved with layers of an "active" medium that measures their energy. Electromagnetic calorimeters measure the energy of electrons and photons as they interact with matter. Hadronic calorimeters sample the energy of hadrons as they interact with atomic nuclei. Calorimeters can stop most known particles except muons and neutrinos, which interact weakly with matter. The main components of the ATLAS calorimetry system are: the Liquid Argon (LAr) Calorimeter, that is the electromagnetic calorimeter and the Tile Calorimeter, that is the central hadronic calorimeter.

Electromagnetic Calorimeter

The electromagnetic calorimeters are designed to identify and measure the direction and the energy of electrons and photons. It utilizes liquid argon (LAr) as an active material and lead as an absorber, providing precise energy measurements across a wide rapidity range. The EM Calorimeter consists of a barrel section (covering $|\eta| < 1.475$) with an accordion-shaped design and an endcap section (for $1.375 < |\eta| < 3.2$) with a more compact layout.

Hadronic Calorimeters

Beyond the electromagnetic calorimeters lies the hadronic calorimeter system, consisting of Tile calorimeters, two endcap calorimeters, and a forward calorimeter.

The Tile calorimeter, functioning as a sampling calorimeter, employs steel as an absorber and scintillators for detection. Its barrel covers the pseudorapidity range $|\eta| < 1.0$, with an extended barrel slightly overlapping and extending the coverage from $0.8 \leq |\eta| \leq 1.7$. The overall thickness of the tile calorimeters is $9.7\lambda_I$ (interaction length) at $|\eta| = 0$

The endcap hadronic calorimeters cover the range from $1.7 \leq |\eta| \leq 3.2$ and share cryostats with the electromagnetic calorimeters. Each endcap comprises two cylindrical wheels and utilizes copper as the absorber material, with liquid argon serving as the active medium.

The forward calorimeters cover the range from $3.2 \leq |\eta| \leq 4.9$ and share the same cryostat as the endcap calorimeters. The forward calorimeter integrates both electromagnetic and hadronic modules. The electromagnetic modules use copper as the primary absorber material, while the hadronic modules utilize tungsten. In all cases, the active material for the forward calorimeter modules is liquid argon.

2.2.4 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost part of ATLAS and is designed to identify and measure the momentum of muons. It is composed of two kind of precision tracking chambers and two trigger chambers. The measurement of the momentum of muons with tracking chambers is done in the region of $|\eta| < 2.7$, while trigger chambers are located in the region of $|\eta| < 2.4$. The MS has been designed with the requirement of precisely determining the muon kinematics up to the TeV scale achieving a resolution of 10% at 1 TeV. The momentum measurements at the GeV scale is more precise with an average resolution of 3% at approximately 200 GeV. Figure 2.6 shows a

schematic illustration of the MS detector in three dimension and Figure 2.7 shows a schematic layout of the MS in the r-z plane.

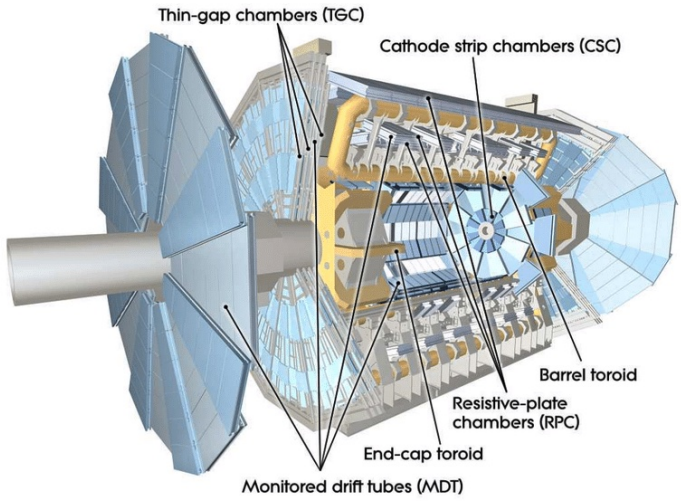

Figure 2.6: Cut-away view of the ATLAS Muon spectrometer.

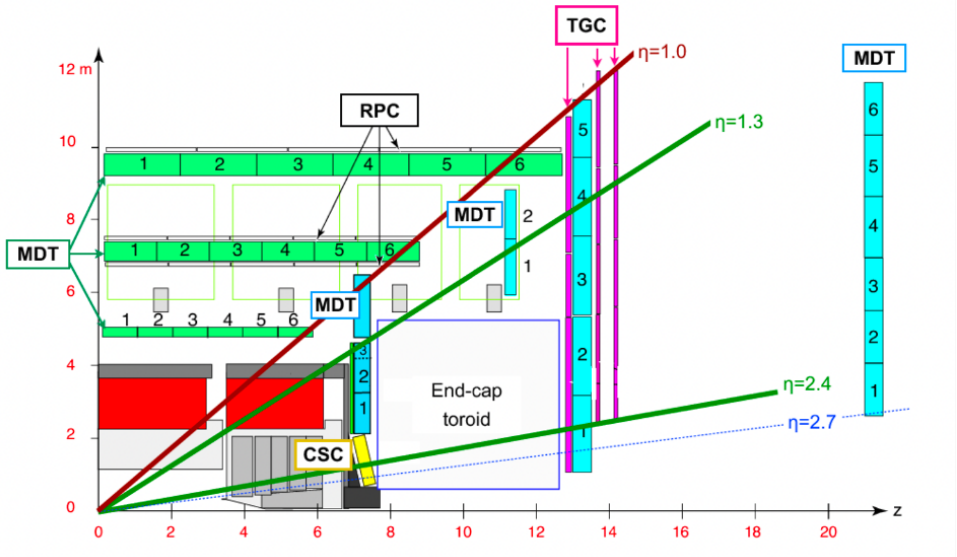

Figure 2.7: Schematic illustration of the ATLAS Muon spectrometer in the r-z plane.

Muon tracking system

The muon tracking system is comprised of a Monitored Drift Tube (MDT) chamber and Cathode Strip Chambers (CSC). The barrel is structured with three concentric cylindrical shells positioned at radii of 5 m, 7.5 m, and 10 m from the geometric center. Endcap muon chambers are situated at ±7.4 m, ±10.8 m, ±14 m, and ±21.5 m along the z-axis relative to the geometric center.

Muon tracking chambers within the barrel are strategically placed between the coils of the barrel toroid magnets, while the endcap chambers are positioned in front of and behind the endcap toroid magnets. The magnetic field generated induces deflection of muon trajectories in the r-z plane, enabling precise charge identification and improving momentum resolution for muons. A small gap in coverage occurs at $\eta = 0$, is intentionally left for accommodating services.

The MDT chambers extend the pseudorapidity coverage up to $|\eta| < 2.7$, employing drift tube chambers in its composition. As muons traverse through, they ionize an Ar/CO_2 gaseous mixture, and the resulting electrons are gathered at an anode tungsten-rhenium wire maintained at a potential of 3080 V. The single-hit spatial resolution, averaged over all drift tube chambers is about 80 μm.

The first layer of the MDT chamber has restricted pseudorapidity coverage ($|\eta| < 2.0$). To complement this region with diminished MDT coverage, CSCs are employed. The CSCs consist of two disks and eight chambers, with each chamber containing four planes. Each plane contributes an independent measurement of η and ϕ. The spatial resolution of the CSCs is 60 μm per CSC plane.

Trigger Chambers

The muon trigger system is employed by the trigger system, conducting an initial muon reconstruction pass. It is comprised of Resistive Plate Chambers (RPC) for the barrel region ($|\eta| < 1.05$) and Thin Gap Chambers (TGC) covering an extended range of $|\eta| < 2.4$.

The RPC are plate detectors constructed with resistive plates interspersed with 2 mm gaps filled with gas, the electric field between the plates is upheld at 4.9 kV/mm. They have an excellent timing resolution of 1.5 ns and are therefore excellent systems to use in triggering. This setup enables avalanches to form along ionizing tracks towards the anode plates, facilitating the detection process. Conversely, the TGC utilize an anode wire maintained at 2900V, situated between resistive cathode plates within a gaseous environment. The TGCs provide a timing resolution

2.2.5 The Trigger and Data Acquisition System

ATLAS employs a sophisticated Trigger and Data Acquisition System (TDAQ) to select events and reduce the event rate from 40 MHz to $\approx$ 1 kHz, which is the maximum capacity for data storage ([19]-[21]). The TDAQ system includes:

• **Level 1 (L1):** This is hardware-based and utilizes information from the calorimeter and fast muon trigger detectors with reduced granularity. The Level-1 trigger performs the initial selection of events, decreasing the event rate from the LHC bunch crossing rate of 40 MHz to about 100 kHz, with a latency (detector read-out time window) below 2.5 μs. L1 identifies Regions of Interest (RoIs) within each event, which are areas of the detector where significant objects like for example high-energy muons might be located. These RoIs are then sent to the High Level Trigger (HLT).

• **Level 2 (L2):** The HLT conducts a full event reconstruction using advanced algorithms tailored for online selection. It is responsible for the final physics selection before further offline analyses. The HLT reduces the rate from the Level-1 output to approximately 1 kHz, with a processing time of around 200 milliseconds and an event size of about 1 megabyte.

2.3 Objects reconstruction in ATLAS

In this section, I provide a brief description of the ATLAS tools for the reconstruction and identification of the physics objects used within this work of thesis. Muons, jets, b-jets, and missing transverse energy are the primary objects used in this analysis.

2.3.1 Muons

Muons reconstruction [22] is performed independently in the Inner Detector and Muon System, then the information from individual sub-detectors is combined, also including the information of the energy loss in the calorimeter, to form the muon tracks used in the physics analyses. The combination just described provides the most precise method to reconstruct and identify muons at ATLAS, but depending on the momentum and the direction of muon it is not always applicable, therefore four muon types are defined:

• Combined muon (CB): track reconstruction is performed independently in the Inner Detector and Muon System, and a combined track is formed with a global fit that uses the hits from both the ID and the MS sub-systems; The method also account for the energy loss in the calorimeter;

• Segment-Tagged muon (ST): a track in the Inner Detector is classified as a muon if, once extrapolated to the Muon System, it is associated with at least one local track segment in the MDT or CSCs;

• Calorimeter-Tagged muon (CT): a track in the Inner Detector is identified as a muon if it can be matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle;

• Extrapolated or Standalone muons (ME): the muon trajectory is reconstructed based only on the Muon System track and a loose requirement on compatibility with origin from the Interaction Point. The parameters of the muon track are defined at the interaction point, taking into account the estimated energy loss of the muon in the calorimeters. In general, the muon is required to traverse at least two layers of Muon System chambers to provide

a track measurement, but three layers are required in the forward region. ME muons are mainly used to extend the acceptance for muon reconstruction into the region $2.5 < |\eta| < 2.7$, which is not covered by the Inner Detector.

There are four kinds of muon identification criteria designed to address the specific needs of physics analyses:

• *Loose* muons are designed to maximise the reconstruction efficiency while provide good-quality muon tracks. All muon types are used;

• $Medium$ muons are designed to minimise the systematic uncertainties associated with muon reconstruction and calibration. Only CB and ME muons are considered. This is the default selection for muons in ATLAS;

• $Tight$ muons are selected to maximise the purity of muons at the cost of some efficiency. Only CB muons with at least two hits in Muon System and satisfying the medium criteria are considered.

• $High-p_T$ muons are selected to maximise the momentum resolution for tracks with transverse momentum above 100 GeV. Only CB muons passing the Medium selection and having at least three hits in three Muon System stations are selected.

In this analysis, muons are required to pass "Medium" identification requirements.

Prompt muons are typically more isolated from other physics objects than non-prompt muons. Isolation variables, based on the amount of energy deposited in the cone around the lepton, are therefore an important feature for the identification of prompt leptons, because they increase the purity of the selection of those produced in the hard interaction against those coming from quark decays. The radius of the isolation cone is defined as $\Delta R = \sqrt{(\Delta\phi^2 + \Delta\eta^2)}$. Isolation can be calorimeter based, by looking at the sum of the transverse energies of clusters inside a cone or track-based, in which the variable is the sum of transverse momenta in a cone. The "FCTight" isolation point is used in this analysis, it is built from tracking and calorimeter information, with a p_T -dependent variable cone. A tight working point is uses to have a large suppression of the multijets background.

2.3.2 Jets

Hadronic collisions in ATLAS detector produce a large number of particles, including quarks and gluons. These particles hadronise, producing a collimated shower of particles with a net momentum equal to the initiating quark or gluon one. These showers of particles are called Jets.

Jets are reconstructed from topo-clusters, three dimensional clusters that are built based on the jet energy deposition in the calorimeters. The reconstruction algorithm is the anti-kt one [23] implemented in the FastJet software package [24]. The algorithm defines the distance d_{ij} between two objects i and j, with the objects being either topo-clusters or previously grouped topo-clusters called *pseudojets*, and the distance d_i between object i and the beam as follows:

$$d_{ij} = min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{2.5}$$

$$d_i = \frac{1}{p_{t,i}^2} \tag{2.6}$$

where: ΔR_{ij} stands for their angular separation defined as

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{2.7}$$

and R is the radius parameter that defines the size of the final reconstructed jet. If d_{ij} is smaller than d_i, objects i and j are combined into a pseudojet; otherwise, object i is called a jet and removed from subsequent iterations of the clustering algorithm. The distances d_{ij} and d_i are redefined and objects are combined or removed until no objects are left. The anti-kt algorithm is preferred by the other sequential recombination algorithms for a number of reasons. Apart from being infrared and collinear (IRC) safe, meaning that the formation of the clustered jet will not change by soft emissions or additional collinear splitting of its constituents, it also creates circular cone-shaped jets. The jet radius parameter R determines the size of a jet and it can affect the jet reconstruction as an input to the clustering algorithm. In the analyses of this thesis small-R jets with $R = 0.4$ were used as the standard option for the jet reconstruction.

The jets reconstructed from electromagnetic topo-clusters are referred to as EMTopo jets. EMtopo jets were the standard collections used in ATLAS analyses until about 2020. The jets which are used in this analysis are reconstructed using the particle flow approach [25], in which information from the tracker is combined with the electromagnetic topo-clusters mentioned above. In EMTopo jets, the energy deposited by charged particles in the calorimeter is used to determine the jet four-momentum. PFlow jets instead use the momenta of the tracks which are associated with the topo-clusters of charge particles, in addition to the topo-clusters of neutral particles. This new approach offers better jet energy resolution by utilising tracking information of the charged particles from the ID in the jet reconstruction. In this way the jet energy resolution is improved especially for low-p_T jets , since for those jets the momentum resolution of the tracker is significantly better than the energy resolution of the calorimeter. Also the accuracy of the jet reconstruction is improved, indeed the association of tracks to vertices can determine whether particles originate from pile-up vertices and therefore should be rejected.
A calibration procedure is needed to account for detector effects which affect the jet energy measurements, including the non-compensation of the calorimeter for hadronic energy deposits, dead material, and leakage of particles beyond the calorimeters. The ATLAS procedure combines several complementary steps in order to calibrate the energies of reconstructed jets, on average, to the energies of their constituent stable particles; this is referred to as the Jet Energy Scale (JES) calibration ([26],[27]). The steps are listed below:

• *Origin correction*: each jet four-vector is recalculated to point to the primary vertex (PV) of the event, instead of the center of the detector. This correction improves the angular jet resolution without a major effect on jet p_T.

• *Pileup correction*: Jet reconstruction in dense environments can be affected by pileup interactions that can potentially modify the area and the energy of the reconstructed jets. In order to correct for this effect the p_T of the redundant contributions is subtracted from the reconstructed jet p_T. The corrections depend on the number of reconstructed primary vertices, the jet

pseudorapidity and the bunch spacing.

• *Absolute calibration*: The absolute JES and eta calibration is derived in order to correct for the difference of the reconstructed jet energy to the jet energy at particle level. This calibration is derived from MC simulated samples with di-jet events.

• *Global sequential calibration*: The global sequential calibration (GSC) is applied on top of the previous calibrations in order to correct JES for residual dependencies on various other jet variables. For example, strong dependence on JES is observed based on the origin of the jet, since quark- and gluon-initiated jets can have differences in the jet shape and particle composition.

• $In-situ$ *calibration*: A final calibration is only applied to data in-situ in order to correct for the differences between the MC simulated samples and the real data. In order to derive these corrections dijets, Z+jets, photon+jets and also multijets events.

In order to distinguish the jets originating from the main primary vertex versus the ones originating from pileup interactions a multi-variate analysis (MVA) discriminant is used called Jet Vertex Tagger(JVT) [28]. The output of the JVT algorithm corresponds to the probability of a jet originating from the main primary vertex and a selection based on this quantity is used in physics analyses.

2.3.3 b-jets

b-jets are jets originated from the decay of a b-hadron and are widely used in ATLAS, both in searches for new physics and in measurements of SM processes, including Higgs boson analyses. The properties of b-hadrons such as a relatively long lifetime, of the order of 1.5 ps, are used to discriminate a b-jet from charm c and light-flavour ((u, d, s)-quark) jets. This leads to reconstruct a secondary vertex, typically a few millimetres from the primary vertex in which the hard-scatter interaction take place. Tracks from the b-hadron decay typically have a large transverse impact parameter, d_0, defined as the distance of closest approach to the primary vertex in the $r\phi$ projection.

Another characteristic propriety of the tracks of b-hadron decay is the large longitudinal impact parameter, z_0, defined as the distance of closest approach along the z-axis.

A schematic diagram of an interaction producing a b-jet plus two light-flavour jets is shown in figure 2.8 and illustrates some of the features that can be used to identify b-jets. These features are used together with the informa-

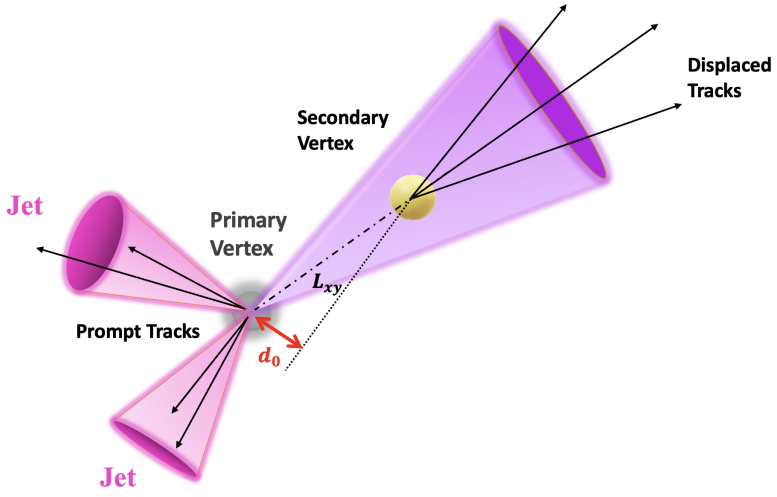

Figure 2.8: A schematic depiction of an interaction producing two light-flavour jet and one b-jet, shown in the transverse plane. The lifetime of b-hadrons corresponds to a transverse decay length, L_{xy} typically a few mm, and produce tracks originating from a secondary vertex. The transverse impact parameter, d_0, and the longitudinal impact parameter are large for tracks originating from the decay of b-hadrons. Jets initiated by light-flavour quarks or gluons do not exhibit these features.

tion of the tracks associated to the jet as input to low-level algorithms. In ATLAS the outputs of the low-level algorithms are included as inputs of high-level algorithms in order to achieve the highest possible b-tagging efficiency, while keeping low mistag rates. High-level algorithms used multivariate analysis techniques to classify the probability of a jet originating from a b-, c- or light-quark. In this analysis the DL1r high-level algorithm is used. The DL1 algorithm [29] employs a DNN approach providing a multi-dimensional output that corresponds to the three probabilities of a jet to originate from a b-, c- or

light-quark, respectively. DL1r is a more recent version of this algorithm. It includes an additional input from the RNNIP algorithm, that is an advanced algorithm for the estimation of the impact parameter based on a recurrent neural network. Using the DL1r output scores, a set of different operating working points can be defined by a set of different cuts on the score, each cut, and consequently each operating working point, correspond to a particular b-jet tagging efficiency. The operating working points cuts and corresponding efficiencies are determined using a simulated $t\bar{t}$ sample. The operating working points are defined for b-jet tagging efficiencies of 60%, 70%, 77%, and 85%, respectively, with the 60% being the tightest and the 85% being the loosest. The operating working points can be used as bin edges and the DL1r b-tagging discriminant output can therefore be binned in four intervals corresponding to certain ranges of b-tagging efficiencies, namely 85-77% (bin 1), 77–70% (bin 2), 70–60% (bin 3), and 60% (bin 4). The operating working point used in this analysis is the 70% one.

As mentioned above the b-tagging efficiencies are established using MC samples, therefore a correction needs to the applied to the simulated data in order to address potential differences with real data. Proper calibrations are derived for the mistag efficiencies as well. These corrections are determined together with their corresponding uncertainties. In particular the calibration of b-jet events is derived comparing data and $t\bar{t}$ MC samples requiring two leptons of leptons of opposite charge and b-jets in the final state.

2.3.4 Missing Transverse Energy

In proton-proton colliders, the exact energy fraction carried by the partons during the collision is unknown. Consequently, the initial energy cannot be determined and the momentum conservation cannot be applied. However, the application of the conservation of momentum is still feasible on the transverse plane, since the total initial momentum in this plane is zero. This allows us to define the missing transverse energy, E_T^{miss} that is used in SM analyses to identify the presence of neutrinos, that cross the entire detector without

interacting with matter and therefore without releasing any signal.
E_T^{miss} is determined as the negative vector sum of the transverse momenta from all identified hard physics objects (such as electrons, muons, jets), along with an extra track-based soft term that includes the contribution of unclustered particles [30]. The E_T^{miss} is calculated as:

$$\vec{E}_T^{\text{miss}} = -\sum \vec{p}_T^{\,e} - \sum \vec{p}_T^{\,\mu} - \sum \vec{p}_T^{\,\tau} - \sum \vec{p}_T^{\,\text{jets}} - \sum \vec{p}_T^{\,\gamma} - \sum \vec{p}_T^{\,\text{soft-terms}}. \quad (2.8)$$

Chapter 3

W+b-jets preselection

The main purpose of this thesis is to study the production of a W boson in association with two b-jets in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. The analysis focuses on the leptonic decay of the W boson into a muon and a neutrino ($W \to \mu\nu$). The critical aspect of this study is developing a method for the suppression of the huge background constituted by events originating from top quark pair production ($t\bar{t}$). This chapter describes the preliminary selection to identify events with a W boson candidate accompanied by 2 b-jets. In this Chapter the different strategies developed on jet counting of events passing the preselection are also presented: one is more inclusive (allowing additional jets in the final state with respect to the 2 b-jets) and one more exclusive (vetoing additional jets in the final state). This is the first step of the analysis, that it followed by a multivariate technique, described in the following chapter, aimed to the rejection of the huge ($t\bar{t}$) background.

Section 3.1 provides a detailed description of the signal and background MC samples utilized in the analysis. Section 3.2 outlines the primary selection criteria applied to these MC samples to identify W+2b-jets event candidates, together with two strategies for the jet counting. The observables of interest for the measurement of the W+2b-jets cross-section are introduced in Section 3.3. Finally, Section 3.4 presents the kinematic distributions of these observables after the event preselection at the detector level.

3.1 Simulated signal and background processes

MC event generators simulate the signal and background processes originating from proton-proton collisions. All MC simulated samples used in this thesis are official samples of the ATLAS Collaboration, generated for the analyses of the full Run-2 dataset.

Two fully simulated signal samples of W boson production in association with jets ($W \to \ell\nu$) are utilized in this analysis:

The nominal sample is produced using the MadGraph5 aMC@NLO (v2.6.5) program that generates MEs for V+0,1,2,3 partons at NLO accuracy and it is based on the 5FS. The showering and subsequent hadronisation has been performed using Pythia8 (v8.245) [31] with the A14 tune [32], using the NNPDF 2.3 LO PDF set [33]. The different jet multiplicities are merged using the FxFx prescription [34] implemented in the MadGraph5 aMC@NLO program. MadGraph5 aMC@NLO performs a 5FNS calculation with massless b- and c-quarks in the ME, and massive quarks in the Pythia8 shower. This sample is referred to as "MGaMC+Py8 FxFx" sample.

An alternative signal sample, based on the 5FS, is produced by Sherpa v2.2.11, using the ATLAS Sherpa v2.2.11 configuration [35]. This configuration performs explicit ME calculations of V+1,2 partons at NLO and from 3 up to 5 partons at LO. The ME are calculated with the Comix [36] and OpenLoops ([37]-[39]) libraries. The PDF set used is the NNPDF 3.0 NNLO set [40]. Higher jet multiplicities can be obtained using the Sherpa PS [41]. This PS model is based on a Catani-Seymour dipole factorization [42]. The merging is performed using the MEPS@NLO prescription ([43]-[46]); the matching is governed by a set of tuned parameters developed by the Sherpa authors. Sherpa uses the cluster fragmentation model for the hadronization of shower particles. In this sample, massless b and c quarks are included at ME level while massive b and c quarks are included in the PS. This sample is referred to as "Sherpa 2.2.11" sample.

The physics processes that constitute the background for this analysis are: $t\bar{t}$,

single top, diboson (WW,ZZ,WZ,WH), Z+jets and multijet events.
The main background is represented by $t\bar{t}$ pair events that are copiously produced at the LHC. In SM model the top quark decays in a W boson and a b-quark ($t \to Wb$) with a branching ratio (BR) close to 100%. Therefore the final states for the $t\bar{t}$ pair-production process can be divided into three classes, depending on the decay of the W boson:

1. $t\bar{t} \to W^+bW^-\bar{b} \to \ell\bar{\nu}\bar{b}q\bar{q}b$ (BR=44%), called semi-leptonic final state, or lepton+jet final state, where in the final state there are: a lepton (muon, electron or tau), a neutrino, four final jets, two of which are b-jets.

2. $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu b\ell^-\nu\bar{b}$ (BR=10%), called leptonic final state, or two lepton final state, where in the final state there are: two leptons and two b-jets and two neutrinos.

3. $t\bar{t} \to W^+bW^-b \to qqbqqb$ (BR=46%), called hadronic final state, where in the final state there are six jets, two of which are b-jets.

Figure 3.1 shows the three possibilities.

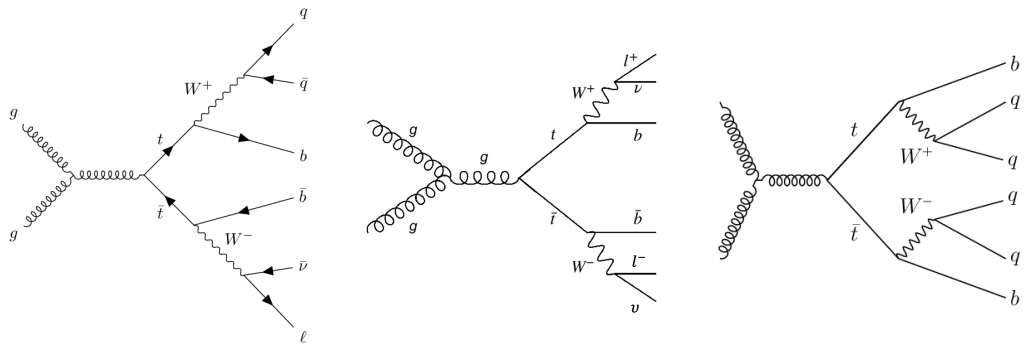

Figure 3.1: Diagram of $t\bar{t}$ in the three final states: semi-leptonic, leptonic, and hadronic final state.

The semi-leptonic final state is the dominant background component in the W+2b-jets analysis, indeed the other two final states are suppressed asking exactly one high p_T lepton in the final state in the preselection as discussed in section 3.2.3.

In this analysis the $t\bar{t}$ background was produced using the PowhegBox v2 ([47]-[49]) generator at NLO with the NNPDF3.0 NNLO PDF set and the h_{damp} parameter set to 1.5 m_{top} [50]. This sample was interfaced to Pythia 8.230, using the A14 tune and the NNPDF 2.3 LO PDF set. This sample is normalized to the NNLO cross-section, including the re-summation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms, calculated with Top++2.0 ([51]-[57]).

Single top quark production also represents a background for this analysis, but thanks to the smaller cross-section it is much smaller than the $t\bar{t}$ one, as shown in section 3.4. Single top quark production, in association with W bosons (tW) as well as in the s- and t- channels, was also simulated using PowhegBox v2 interfaced with Pythia 8.230 at NLO using the NNPDF 2.3 LO PDF set. To remove overlap with the $t\bar{t}$ sample, the diagram-removal [58] procedure was used.

Diboson final states (VV = WW, ZZ, WZ) also constitute background for this analysis in the cases where one W or Z boson decays leptonically and the other decays hadronically, these were simulated using Sherpa v.2.2.11 [59], with NLO MEs for up to one parton emission and LO MEs for up to three parton emissions calculated using the NNPDF 3.0 NNLO PDF set. The ME calculations are matched to the Sherpa 2.2.11 parton shower using the MEPS@NLO prescription.

POWHEG+Py8 with the NNPDF 3.0 NNLO PDF set was also used to simulate the production of vector bosons W, Z in association with a Higgs boson (VH).

Z+jets events may also contribute to W+2b-jets selection if one of the leptons decays outside the detector acceptance or in the case of $Z \to \tau\tau$ decays, with one of the taus decaying in $\tau \to \mu\nu\nu$. Finally W($\to \tau\nu$)+jets events have to

be considered. W/Z+jets background have been simulated with Sherpa 2.2.11 with the same set-up used for the alternative signal samples. Table 3.1 shows a summary of all MCs samples used in this analysis.
The last background to be considered is the multijet background where a muon can be identified in the final state mainly via decays of b- and c-hadrons that contain jet with muons. The event must also contain some missing transverse energy either from a measurement of the deposited energy or from neutrinos in heavy-flavour decays. MC generators are not reliable enough for the estimation of this background , therefore it has to be estimated directly on data with data-driven techniques, this goes beyond the goals of this thesis, the reader has just to take in mind that the multijet background in W+jets analyses typically amounts to the 10% of the signal, if a tight isolation is applied to the muon as in this thesis, it therefore a small background.
All simulated event samples include the effect of multiple proton-proton interactions in the same and neighbouring bunch crossings (pile-up) by overlaying simulated minimum-bias events on each generated signal or background event. The minimum-bias events were simulated with the single-, double- and non-diffractive proton-proton processes of Pythia 8.186 using the A3 tune [60] and the NNPDF 2.3 LO PDF. For all MadGraph and Powheg samples, the EvtGen v1.6.0 program [61] was used for the bottom and charm hadron decays. The generated samples were processed using the Geant-based ATLAS detector simulation ([62]-[63]) and the same event reconstruction algorithms were used as for the data.

3.2 Primary selection

A crucial aspect of all ATLAS analyses is the efficient management of the vast amounts of data and simulated samples. Initially, data are recorded in a comprehensive format that contains complete event information and a lot of datasets are available encompassing all triggered events and simulated samples. This data must be refined into a format containing only the information

Process	Generator
$W(\rightarrow \mu\nu) + jets$	MGaMC+Py8 FxFx
$W(\rightarrow \mu\nu) + jets$	Sherpa 2.2.11
$t\bar{t}$	Powheg+Py8
single top	Powheg+Py8
$VV(WZ, ZZ, WW)$	Sherpa 2.2.11
$W/HH(b\bar{b})$	Powheg+Py8
Z+jets	Sherpa 2.2.11
$W(\rightarrow \tau\nu) + jets$	Sherpa 2.2.11

Table 3.1: Summary of signal and background MC samples. The generator programs used in the simulation are listed in the second column.

necessary for the specific analysis and a dataset with only the relevant events. This final format must allow for a rapid access and a straightforward usage. ATLAS centrally performs a splitting and a reduction of the amount of recorded data and generated simulated samples, in order to provide the analysers data formats containing only interesting events with useful information for their analysis and to speed up the process of managing data files. The procedure followed in Run-2, which is also used in this thesis, is detailed in section 3.2.1. This section also covers the initial step of my thesis work, where I established and applied the further specific skimming procedure optimal for my analysis. one has to apply requirements to defined the physics objects in order to efficiently identify signal events and to reject background in the W+ $2b$-jets analysis. The definition of these physics objects is described in section 3.2.2, while the detailed selection of the events of interest is covered in section 3.2.3.

3.2.1 ATLAS Derivations and CxAOD format

In Run-2, the reduction of the data (referred to as "derivations") is made by the Derivation Framework, which provides tools to select interesting events. Derivations have been performed starting from the output of the general recon-

struction framework of the ATLAS experiment, ATHENA, in a format called xAOD. The derivations, called DxAOD (Derived-xAOD), have the same format of xAODs, but reduced size. The derivation procedure is based on three operations.

The full set of reconstructed events is first reduced through the skimming procedure, which consists on the removal of not interesting events. It is followed by the thinning, which removes not useful reconstructed objects from an event, but keeping the event, and finally by the slimming, which deletes not necessary information from objects. The format of the derivations have been defined by individual physics groups according to the specific analysis needs. In this thesis, the STDM4 DxAODs are used, produced for a set of SM analyses. They are characterised by the presence of at least one lepton in each event, passing single lepton triggers. In particular the STDM4 DxAODs contain the following events:

- Events with at least one muon with $p_T > 15$ GeV and $|\eta| < 2.6$;
- Events with at least one electron with $p_T > 20$ GeV and $|\eta| < 2.6$;
- Events passing the single electron and muon High Level Trigger selection.

Then, by means of the CxAOD Framework, the analysers apply the latest calibrations to physics objects and a futher event reduction. The reduction I applied at this stage for W+2b-jets analysis is the following:

- Events are required to have at least one primary vertex with at least two associated tracks.
- Events are required to have at least 1 jet (≥ 1 jet) with $p_T > 20$ GeV and $|\eta| < 4.7$.
- Events are required to have at least 1 electron with loose identification with $p_T > 7$ GeV and $|\eta| < 2.47$ or at least 1 muon with with loose identification $p_T > 7$ GeV and $|\eta| < 4.7$.

The events are finally saved in a format similar to DxAODs called CxAODs (Calibrated xAOD) that is the one used for the analysis. Figure 3.2 shows the schematization of the full procedure.

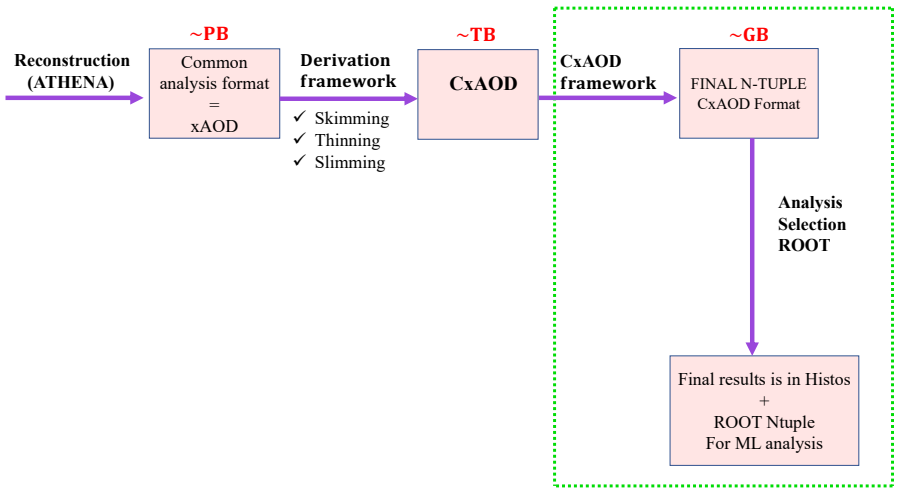

Figure 3.2: Scheme of the Derivation Framework used by the ATLAS Collaboration in Run-2 together with scheme of the CxAOD framework: Data are reconstructed with ATHENA and derived by the Derivation Framework in DxAOD samples with smaller size and processed with the CxAOD framework for final size reduction.

3.2.2 Object definition

Muons are reconstructed and identified in ATLAS as described in section 2.8.1. Medium muons are used in this analysis. They have to come from the primary vertex, therefore they need to meet the criteria: $|d_0|/\sigma(d_0) < 3.0$ and $|z_0 \sin\theta| < 0.5$ mm. Additionally, they must be isolated (meeting the "FCTight FixedRad" working point): the p_T sum within a variable-radius cone in the ID system around the combined track must be less than 0.04 times the muon transverse momentum, the maximum cone size is $\Delta R = 0.3$, decreasing for higher muon p_T [64]. Muons must have $p_T > 27$ GeV and $|\eta|$ <2.5. Table 3.2 outlines the criteria for defining muon candidates.

Muon Selection	
ID	Medium
Isolation	FCTight
vertex track association	$\lvert d0signBL\rvert < 3$, $\lvert z0BL * \sin\theta\rvert < 0.5$ mm
p_T	27 GeV
η	$\lvert\eta\rvert < 2.5$

Table 3.2: Overview of the muon selection criteria

Jets are reconstructed from PFlow objects using the anti-k_t [65] algorithm in the FastJet package [66], as described in section 2.8.2. They must have a $p_T > 20$ GeV and $|y(jet)| < 2.5$. Jets originating from pileup vertices are excluded by using the "Tight" working point of JVT discriminant discussed in Section 2.3.2 [67], which requires JVT to be above 0.5 for jets within the 20-60 GeV p_T range and $|\eta| < 2.4$.

Jet Selection	
Algorithm	anti-k_t, R=0.4, AntiKt4EMPFlowJets
p_T	20 GeV
$\lvert y\rvert$	< 2.5
JVT	Tight WP
***b*-jet Selection**	
WP	DL1r@70%
p_T	20 GeV
$\lvert\eta\rvert$	< 2.5

Table 3.3: Overview of the jet selection criteria.

Jets are classified as b-tagged if they meet the 70% efficiency working point, meaning the algorithm has an 70% efficiency rate for identifying jets with a b-hadron. The b-tagging selection results in a misidentification efficiencies of 12.5% for c-jets and a misidentification efficiencies of 0.3% for light-jets.

Additionally, jets that meet the b-tagging criteria must have $p_T > 20$ GeV and be within the pseudorapidity range of $|\eta|$ <2.5.
E_T^{miss} is determined as detailed in section 2.8.4. The "Tight" Emiss operating point is utilised, which includes forward jets with $2.4 < |\eta| < 4.5$ and $p_T > 30$ GeV.

3.2.3 Primary Event selection

Events stored in the CxAODs have to be further selected in order to collect W+2b-jets candidates. After preliminary studies, I developed the analysis strategy described below, it has been inspired to preliminary studies done in the ATLAS experiment for the analysis of partial Run-2 dataset (36 fb^{-1}), that wasn't finalised due to the challenges related to the $t\bar{t}$ background rejection.
Events containing a W-boson candidate are selected by requiring exactly one muon, defined as detailed in Table 3.2, and E_T^{miss} >25 GeV. Due to the pres-

$W \to \mu\nu$ Selection	
N leptons	exactly 1 good muon (as defined in Table 3.2)
2nd lepton veto	no additional isolated muons or electrons with $p_T >$ 7 GeV and $\|\eta\| < 2.5$
E_T^{miss}	> 25 GeV
m_T^W	> 60 GeV

Table 3.4: Selection criteria for W boson

ence of the neutrino the invariant mass of W candidates cannot be reconstructed, therefore the transverse mass (m_T^W) is used for the selection, m_T^W must be greater than 60 GeV, where m_T^W is defined as follows:

$$m_T^W = \sqrt{2p_T^\mu p_T^\nu [1 - \cos(\phi^\mu - \phi^\nu)]}. \tag{3.1}$$

p_T^ν and ϕ_ν are the measured E_T^{miss} and its angle, $\phi_{E_T^{miss}}$. To reject the possibility of di-leptonic $t\bar{t}$ decays, the selection process strengthens the veto on additional leptons rather than simply requiring exactly one signal muon, also

no extra leptons meeting looser criteria are permitted. Vetoed leptons are isolated muons and electrons with $p_T > 7$ GeV and $|\eta| < 2.5$. The identification criteria for W candidate events are summarized in Table 3.4.

After the definition and selection of the W-boson, I developed two possible strategies for the b-jets selection. The first one (referred to as Geq2BiJ) is inspired to the published ATLAS papers on the Z+2b-jets either with partial Run-2 dataset and with full dataset, the second one (referred to as 2B2J) is an alternative inspired to the preliminary studies done in the ATLAS experiment for the analysis of partial Run-2 dataset. The explored selections and the underlying ideas supporting them are explained below:

Geq2BiJ Strategy: The most inclusive selection applicable to the W+2bjets analysis is a selection asking for events with at least 2 bjets (as defined in Table 3.3) after the W selection described in Table 3.4. This is exactly the b-jets strategy applied by ATLAS for the measurement of Z+2bjets analysis in Run-2 ([68],[15]), where the inclusive b-jets selection has been kept both at detector level and at particle level for the final cross-section measurement. Applying the same strategy on W analysis would be ideal in order to study the modelling on W/Z+2 b-jets production in a very similar phase-space. Table 3.5 shows the number of expected signal and $t\bar{t}$ events after the Geq2BiJ selection normalised to the full Run-2 dataset (140 fb^{-1}). The ratio of signal to $t\bar{t}$ background is 5.3-5.4% depending on the signal MC used, MGaMC+Py8 FxFx sample is the nominal signal sample of the analysis, while Sherpa 2.2.11 is the alternative signal sample. The ratio has a very small decreases if one considers the ratio of signal to sum of all background (S/B), indeed it is 4.8%, indicating as expected that the $t\bar{t}$ is the dominant background ($t\bar{t}$ is the 90% of the total background). A S/B of about 5% makes a precise differential cross-section impossible. This strategy is applicable only applying a further selection at detector level optimised for the rejection of the $t\bar{t}$ background. A neural network classifier has been developed with this scope as detailed in the next Chapter.

2B2J Strategy: To reject the significant background from semi-leptonic

Selection	**Signal MGaMC +Py8 FxFx**	$t\bar{t}$	**S/$t\bar{t}$(%)**	**S/B(%)**	Eff_S(%)
Geq2BiJ	1006830 ±580	1887700 ±500	5.3	4.8	100
2B2J	28580±360	73900±100	38.6	25.8	28.4
Selection	**Signal Sherpa 2.2.11**	$t\bar{t}$	**S/$t\bar{t}$(%)**	**S/B(%)**	Eff_S(%)
Geq2BiJ	101050 ±880	1887700 ±500	5.4	4.8	100
2B2J	25000±500	73900 ±100	33.8	22.2	25.1

Table 3.5: Signal and $t\bar{t}$ event number after the Geq2BiJ and 2B2J selections together with S/$t\bar{t}$ and S/B(all background), and signal efficiency with respect to Geq2BiJ selection. Results are shown for 2 signal MC samples.

$t\bar{t}$ decays ($t\bar{t} \to WbWb \to \ell\nu bjetjetb$) that are expected to present a larger number of jets in the final state with respect to the W+2b-jets signal, a specific b-jets counting and veto strategy is employed. Events with exactly two b-jets (and no additional b-jets) are selected. The definition of b-jets is detailed in Table 3.3. The following jet-veto strategy is then applied:

- **Central jet veto**: For selected events, no additional good jets are allowed with $p_T > 20$ GeV and $|y| < 2.5$ (using the jet collection detailed in Table 3.3)

- **Forward jets veto**: Selected events must have no good jets with $2.5 < |y| < 4.5$ with $p_T > 30$ GeV. The selection of forward jet is detailed in Table 3.6. The higher p_T cut of forward jets with respect to central jets is meant to limit the jets from pileup that are copiously produced in the forward region of the detector.

Forward Jet Selection	
Algorithm	anti-k_t, R=0.4, AntiKt4EMPFlowJets
p_T	30 GeV
$\|y\|$	$2.5 < \|y\| < 4.5$

Table 3.6: Overview of the Forward jet selection criteria.

Table 3.5 shows the performances on this strategy. The ratio of signal to $t\bar{t}$ background range between 35-50% depending on the signal MC used and it decreases to 22-26% if one considers the ratio of signal to sum of all background (S/B), the cost to pay is a large reduction of the signal, it is reduced at a level of 25-28% with respect to the Geq2BiJ selection.
This selection was initially developed for the first analysis attempt using a partial Run-2 dataset (36 fb^{-1}). It was noted at that time that achieving a precise measurement of the differential cross-section required further suppression of the top background, prompting the exploration of a BDT approach. In order to have an effective BDT cut a further and too large reduction of the signal was needed, making the measurement not feasible. In this thesis, a neural network classifier has been designed in place of the BDT approach, indeed currently in ATLAS neural network analyses are showing better performances than the BDT ones. Moreover this thesis is exploiting the full Run-2 dataset, that is about four times larger that the dataset employed in the first attempt of the analysis, this allows to apply tighter cut on the signal still keeping high the number of final signal events allowing a differential cross-section measurement.

3.3 Primary observables for the measurement of the W+2bjets

In this section, I discuss the main observables that it is interesting to measure in the W+2b-jets cross-section measurement, the differential cross-section measurement will be measured as a function of each of them in the

final ATLAS measurement, and their physics motivations:

1. The p_T of the leading b-jet [1] and the second leading b-jet: The measurements of jet kinematics, particularly the p_T of the leading jets, are crucial for testing QCD predictions and improving the modeling of MC event generators.

2. The y of the leading b-jet and the second leading b-jets: These measurements not only aid in understanding QCD predictions but also provide valuable inputs for fitting PDFs.

3. The invariant mass of the leading b-jets pair (m_{bb}), the azimuthal angle separation between the two leading b-jets ($\Delta\phi_{bb}$) and the angular separation between the two leading b-jets (ΔR_{bb}): The distributions of these dijet variables are essential for studying Higgs boson topologies and searching for new physics scenarios (e.g., resonances decaying into b-jets pairs), since W+2b-jets are backgrounds for these topologies. Consequently, they play a critical role in refining MC event generator models.

These variables are the main focus of investigation in this analysis, being the observables that will be unfolded for the differential cross-section measurements in the future.

3.4 Detector level distributions after preselection

In this section I present the kinematic distributions of the physics objects on which I applied cut to defined the pre-selected W+2b-jets events and I also show the kinematic distributions of the primary observables discussed in the previous section as provided by the MC simulations. The shown distributions include in addition to the signal and $t\bar{t}$ backgrounds also smaller backgrounds. Specifically, the distributions are showcased for the two specific selections: the

[1] The leading b-jet is the one with the highest p_T.

Geq2BiJ and the 2B2J.

To provide a comprehensive overview, Table 3.7 displays the expected number of events categorized by signal and all backgrounds for both selection criteria. These tables and distributions are normalized to represent the full Run-2 dataset corresponding to integrated luminosity of 140 fb^{-1}.

Process	**Generator**	**N. of evts in 2B2J**	**N. of evts in Geq2BiJ**
W($\rightarrow \mu\nu$)+$2b-jets$	MGaMC+Py8 FxFx	28580 ± 360	100680 ± 580
W($\rightarrow \mu\nu$) + $1b-jet$	MGaMC+Py8 FxFx	2250 ± 130	6100 ± 200
W($\rightarrow \mu\nu$) + $c-jets$	MGaMC+Py8 FxFx	5340 ± 290	21270 ± 470
W($\rightarrow \mu\nu$)+$light$-$jets$	MGaMC+Py8 FxFx	770 ± 170	2900 ± 300
W($\rightarrow \mu\nu$)+$2b-jets$	Sherpa 2.2.11	25000 ± 500	101050 ± 880
W($\rightarrow \mu\nu$) + $1b-jet$	Sherpa 2.2.11	2900 ± 200	9220 ± 360
W($\rightarrow \mu\nu$) + $c-jets$	Sherpa 2.2.11	6340 ± 300	25430 ± 490
W($\rightarrow \mu\nu$)+$light$-$jets$	Sherpa 2.2.11	1360 ± 260	2860 ± 370
$t\bar{t}$	Powheg+Py8	73900 ± 100	1887700 ± 500
Single top	Powheg+Py8	19210 ± 30	167240 ± 120
Z+$jets$	Powheg+Py8	6890 ± 140	26600 ± 240
Diboson	Powheg+Py8	1760 ± 30	5890 ± 60
W ($\rightarrow \tau\nu$) +$jets$	MGaMC+Py8 FxFx	440 ± 60	1880 ± 110

Table 3.7: Expected number of events for signal and background for each selection criterion. Results are shown for 2 signal MC samples.

Figure 3.3 shows the distribution of E_T^{miss}, p_T of the muon and m_T^W after the preselection, these distributions are done as sanity checks, for example the peak of m_T^W as expected is well visible, also E_T^{miss} and p_T^μ present the expected trends.

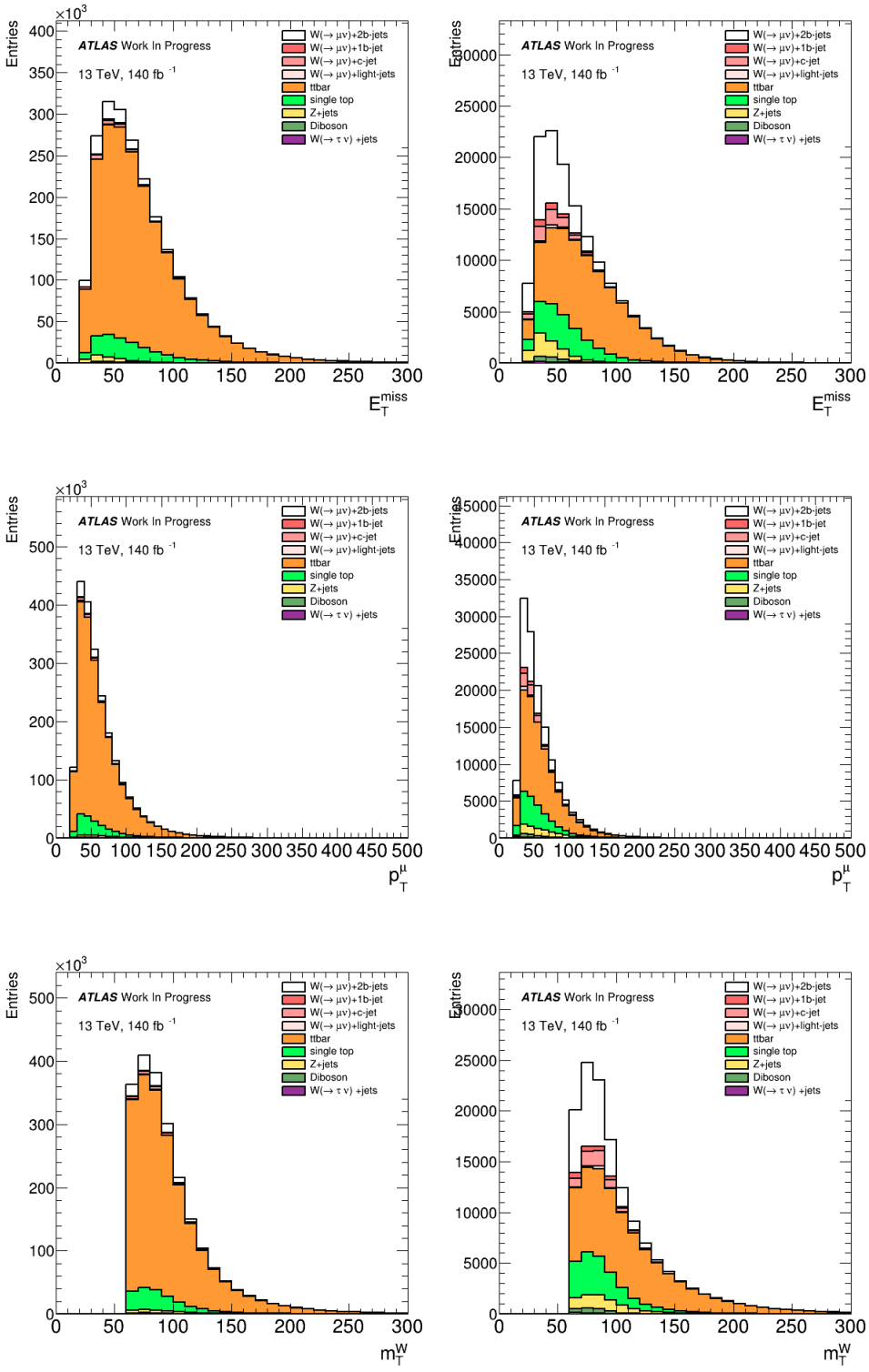

Figure 3.3: Distributions of E_T^{miss} (first row), p_T^{μ} (second row), and m_T^W (third row) in events of Geq2BiJ (left) and 2B2J (right) regions.

Figure 3.4 shows the distributions of the p_T of the leading and subleading b-jets, while Figure 3.5 shows the distributions of rapidities for these jets.

Figure 3.6 presents the distributions of m_{bb}, $\Delta\phi_{bb}$, and ΔR_{bb}. These figures cover both Geq2BiJ and 2B2J. The figures are done using the nominal signal MC sample (MGaMC+Py8 FxFx), similar results are obtained with the alternative Sherpa 2.2.11 signal MC sample. As expected passing from the Geq2BiJ selection to the 2B2J the signal component is enhanced if compared with the background, but the total amount of data is strongly reduced.

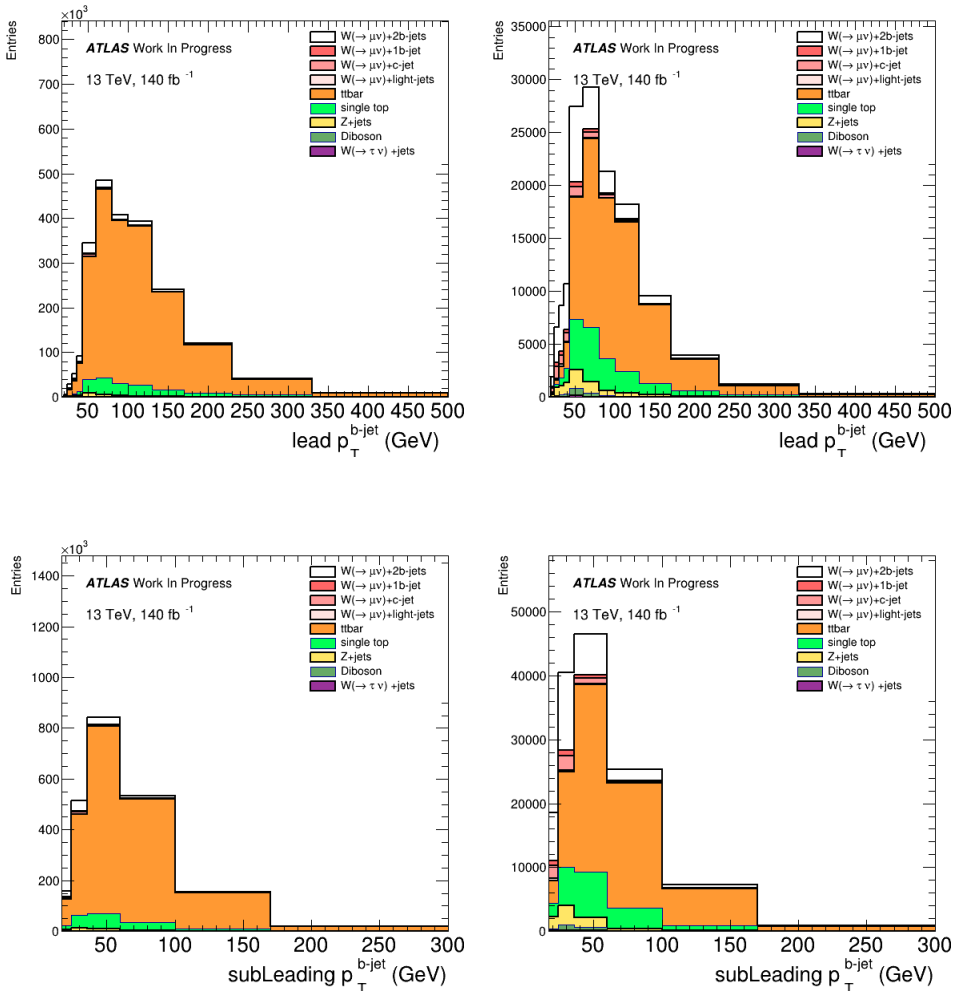

Figure 3.4: Distributions of p_T of leading b-jet (first row), p_T of subleading b-jet (second row) in events of Geq2BiJ (left) and 2B2J (right) regions.

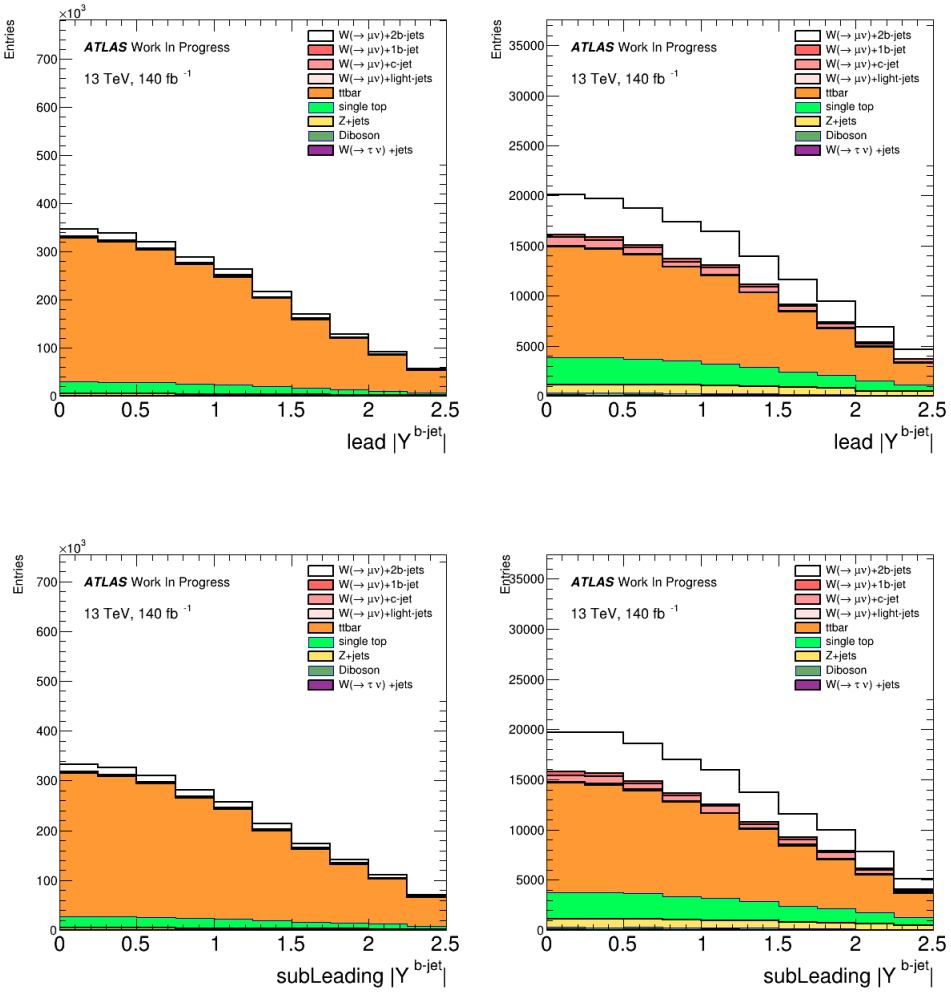

Figure 3.5: Distributions of Y of leading b-jet (first row), Y of subleading b-jet (second row) in events of Geq2BiJ (left) and 2B2J (right) regions.

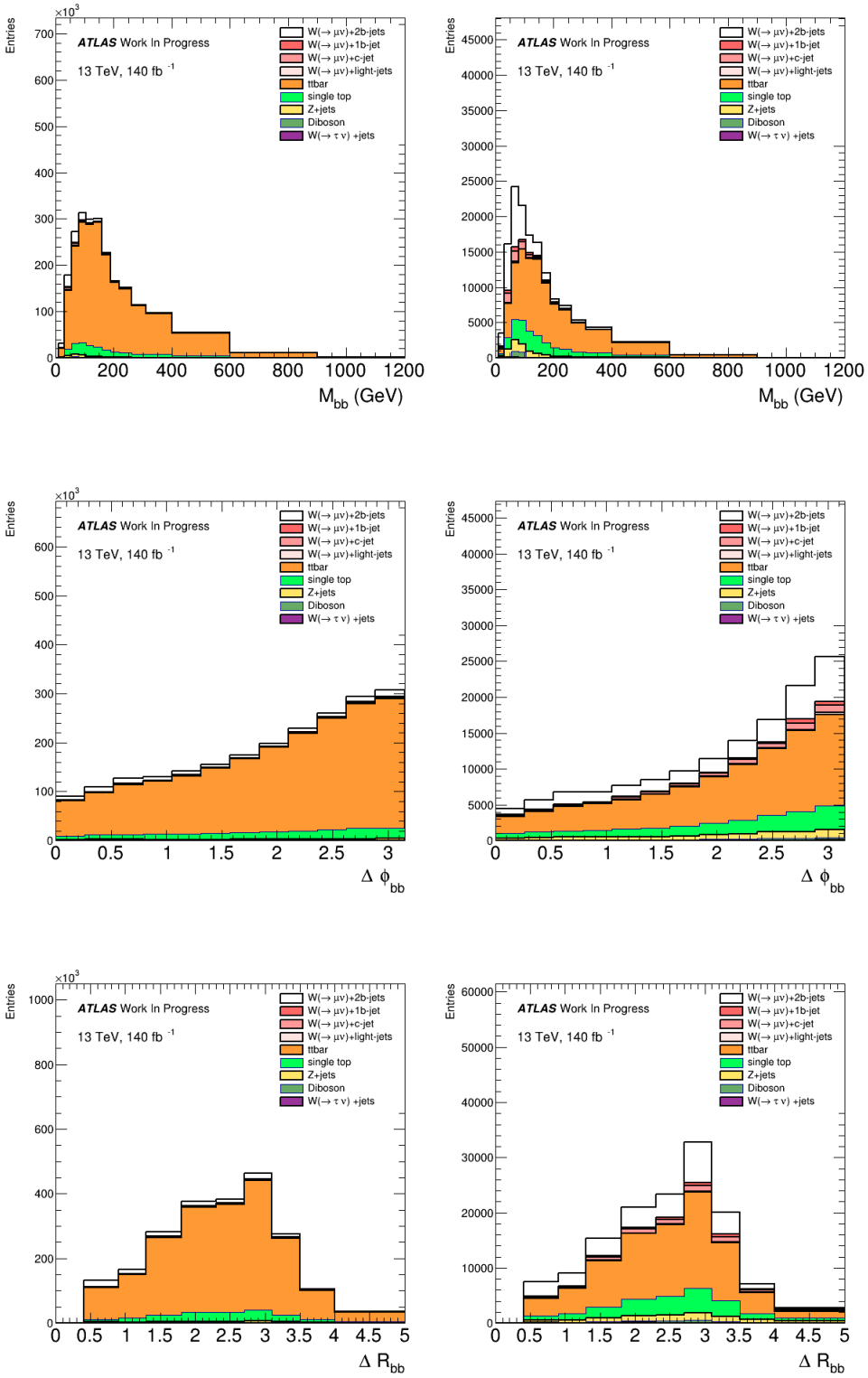

Figure 3.6: Distributions of m_{bb} (first row), $\Delta\phi_{bb}$ (second row), and ΔR_{bb} (third row) in events of Geq2BiJ (left) and 2B2J (right) regions.

Chapter 4

Neural Network for $t\bar{t}$ background rejection

The initial event selection described in the previous chapter allows a first selection of signal candidates and a first rejection of the background. This stage is cut-based and involves applying a series of cuts on single observables. The observables employed in the selection are related to the kind of physics objects present in the events (i.e., one muon, E_T^{miss} coming from the W boson decay plus two b-jets) and their kinematic properties (i.e., the transverse momentum, the rapidity). These cuts are meant to select only events with one W decaying leptonically and two b-jets.

However, the cut-based selection alone is not sufficient to obtain a significant rejection of all backgrounds, particularly events from top quark production, which mimic the signal. Therefore, a more sophisticated method is employed in the second stage. This involves using a Neural Network (NN), a machine learning model which is able to identify complex patterns in data.

The NN is trained on a labeled dataset, where events are tagged as either signal or backgrounds. Through this training process, the NN learns to distinguish between the subtle differences in the features of signal and background events. Once trained, the NN is applied to the selected events that have passed the first-stage selection, assigning a probability to each event of being signal or background. By using this NN-based approach, the second stage of the

analysis significantly improves the rejection of top background events, thereby enhancing the purity of the selected signal sample. This combination of cut-based selection followed by NN-based classification ensures a more efficient identification of W and b-jet candidates while minimizing contamination from top quark backgrounds.

The chapter is organized as follows: Section 4.1 introduces the methodology for NN construction, covering architecture design and data preprocessing. Section 4.2 details the NN developed specifically for the exclusive selection (2B2J), discussing variable selection, optimization strategies, and model performances. Section 4.3 extends this analysis to inclusive selection (Geq2BiJ), mirroring the structure of Section 4.2 but making use of more variables.

4.1 How to Build the Neural Network

I developed the NN to distinguish events where a W boson decays leptonically in a muon and a neutrino and it is accompanied by 2 b-jets ($W(\mu\nu) + 2b - jets$) from background events. The NN can understand different features of the events and use them to separate the signal from the background. In particular, the NN is optimised to reject $t\bar{t}$ background that is the most copious. As already discussed in Section 3.1 the $t\bar{t}$ pair events present a final state constituted by 2 W bosons and 2b-jets and the $t\bar{t}$ final state, that is harder to reject, is the one coming from the semileptonic decay, where in the final state there is one muon and one neutrino (coming from the decay of one of the two W bosons), 2 jets (coming from the decay of the second W) plus 2 b-jets.

A NN is a model that makes decisions in a manner similar to the human brain by using processes that mimic the way biological neurons work together to identify phenomena, weigh options, and arrive at conclusions. A NN is composed of layers of interconnected nodes (neurons). The primary types of layers are: the input layer, hidden layers, and the output layer. The input layer is constituted of a set of neurons corresponding to the features of the input data (each neuron corresponds to a feature). The output layer produces the final

prediction of the NN. The number of neurons in this layer depends on the nature of the problem. For example, in a binary classification task where the NN is optimised to distinguish between two distinct classes (i.e., the separation between signal and background), typically there is just one neuron providing the probability that the data belong to one of the two classes. Between the input and output layers, there are typically hidden layers. These layers perform complex transformations of the input data. In a feed-forward NN, as developed for this thesis, each neuron in a hidden layer receives inputs from all the neurons of the previous layer, applies a weighed sum of the inputs, and performs its output through an activation function. Figure 4.1 shows the typical structure of this kind of NNs.

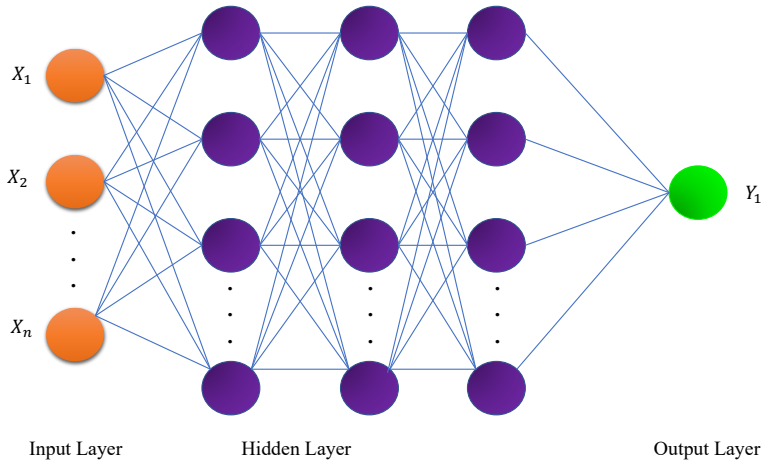

Figure 4.1: Example of a NN with n features, and 3 hidden layers, 1 output layer with 1 node.

The weights of a NN are learned during a training procedure. During the training, a set of labeled examples is provided to the NN and the predictions of the NN are compared to the labels. A function loss is defined to evaluate the discrepancy between the predictions and the labels, an optimizer uses the values of the losses to minimize the function loss (via backpropagation) and to update the weights. During the backpropagation, the network learns from

its mistakes and updates its weights to minimize the loss. A hyperparameter, known as the learning rate, controls the size of weight updates. This cycle is repeated several times, using different groups of examples. Each iteration over the entire training set is called an epoch. A detailed description of NN models is provided in Appendix A.

In this study, the signal and background samples are split into multiple subsets to realize training, validation, and testing samples that are independent without bias. The strategy developed allows at the same time to exploit all available MC samples. Signal and background events are numbered. Based on the parity of the Event Number associated with each event, each sample is divided into two sets of even and odd events. One NN is trained on even events and used to analyse the odd events, and a second NN is trained on odd events and used to analyze the even events. During training, the model learns from a labeled dataset, adjusting its parameters to minimize the difference between predicted and actual outputs. To check the performance of the NNs, the set of simulated events not used for the training is further divided into equal-sized "test" and "validation" sets. The validation sample allows to assess the model's performance on a separate dataset that was not used during training, it is used to optimise the NN hyperparameters and prevent overfitting by checking how well the model generalizes to new data. Finally, the test sample is used to evaluate the expected performance on another independent dataset. It ensures the model's reliability and measures its ability to make accurate predictions on unseen data, providing a final assessment of its effectiveness and generalization capabilities.

Due to the splitting applied to signal and background events, in this analysis there are: 2 training samples (labeled as: train_0 and train_1), 2 validation samples (labeled as: validation_0 and validation_1) and two test samples (labeled as: test_0 and test_1). If the label of the sample is not explicitly mentioned (0 or 1) in the text and in the plots, it means that the sum of the two is considered.

The inputs of the NN model, the so-called features, are provided by the observ-

ables that have discriminating power between signal and background. Before feeding these input variables into the networks, they are preprocessed in order to harmonize the inputs. The preprocessing of the input variables consists of a scaling procedure. The scaling procedure is needed because the NN works best with input variables whose values have the same order of magnitude. Each input variable is scaled separately by subtracting its mean and dividing by its standard deviation, where the mean and the standard deviation are computed for each observable on the set of signal and background events used in the training of each network.

The implementation of the feed-forward NNs developed for this study has been performed using Keras 2.15.0 [69] and TensorFlow 2.11 [70], both of which are open-source software libraries. Keras serves as a high-level Application Programming Interface (API) for building and deploying NNs, written in Python and designed to run on top of TensorFlow. Keras is known for its user-friendly interface, which simplifies the construction and deployment of machine learning models, while TensorFlow provides robust computation capabilities.

The number of hidden layers, the number of nodes in each layer, the number of epochs, and the learning rate have been optimised based on the expected classification performance on the validation sample. The activation function chosen for hidden layers is a Rectified Linear Unit (RELU), one of the most common for binary classification tasks. It is shown in Figure A.9 of Appendix A. The activation function of the output layer is a Sigmoid, as is usually done for binary classification tasks, its value indeed ranges between 0 and 1 and it can be interpreted as a probability. The sigmoid function is shown in Figure A.7 of Appendix A. The loss function used for binary classification tasks is the binary cross-entropy loss shown in Equation A.7 of Appendix A, and the optimizer is Adam. The technical aspects of the NN development are discussed in detail in Appendix A for further understanding. These technical details apply to all the NNs developed in this thesis and presented in the following sections.

4.2 NN for the exclusive selection (2B2J)

In this section, the development of the NN used for the exclusive selection (2B2J) described in the previous chapter is presented. The events passing the 2B2J selection contain a W candidate decaying into a muon and a neutrino, exactly 2b-jets in the central part of the detector and no additional jets neither in the forward nor in the central region. The primary signal MC sample used to perform this study is the MGaMC+Py8 FxFx sample, that is the nominal one of the analysis. The input observables of the NN model are presented in Section 4.2.1. The work on the optimisation of the NN is shown in Section 4.2.2. The performance of the model is presented in Section 4.2.3. The studies on the second signal MC sample available (Sherpa 2.2.11) is reported in Section 4.2.4.

4.2.1 Selected features for the model

After applying criteria that define the 2B2J selection, a study has been conducted to identify the observables that can effectively distinguish signal events from top background events. The physics objects present in events passing the 2B2J selection are: the muon, the missing transverse energy (that represents the neutrino), and the 2 b-jets. Using these objects a set of kinematic variables has been studied.

The shapes of the signal and background event distributions of the various observables have been compared to identify the ones that can be used for the separation. At this stage both the dominant $t\bar{t}$ background and the smaller single top background are considered. Figure 4.2 illustrates the distributions of muon transverse momentum (p_T^{μ}), muon pseudorapidity ($|\eta_{\mu}|$), missing transverse energy (E_T^{miss}), the angle between the muon and E_T^{miss} in the transverse plane ($\Delta\phi(\mu, E_T^{miss})$), and the transverse mass of the W boson (m_T^W).

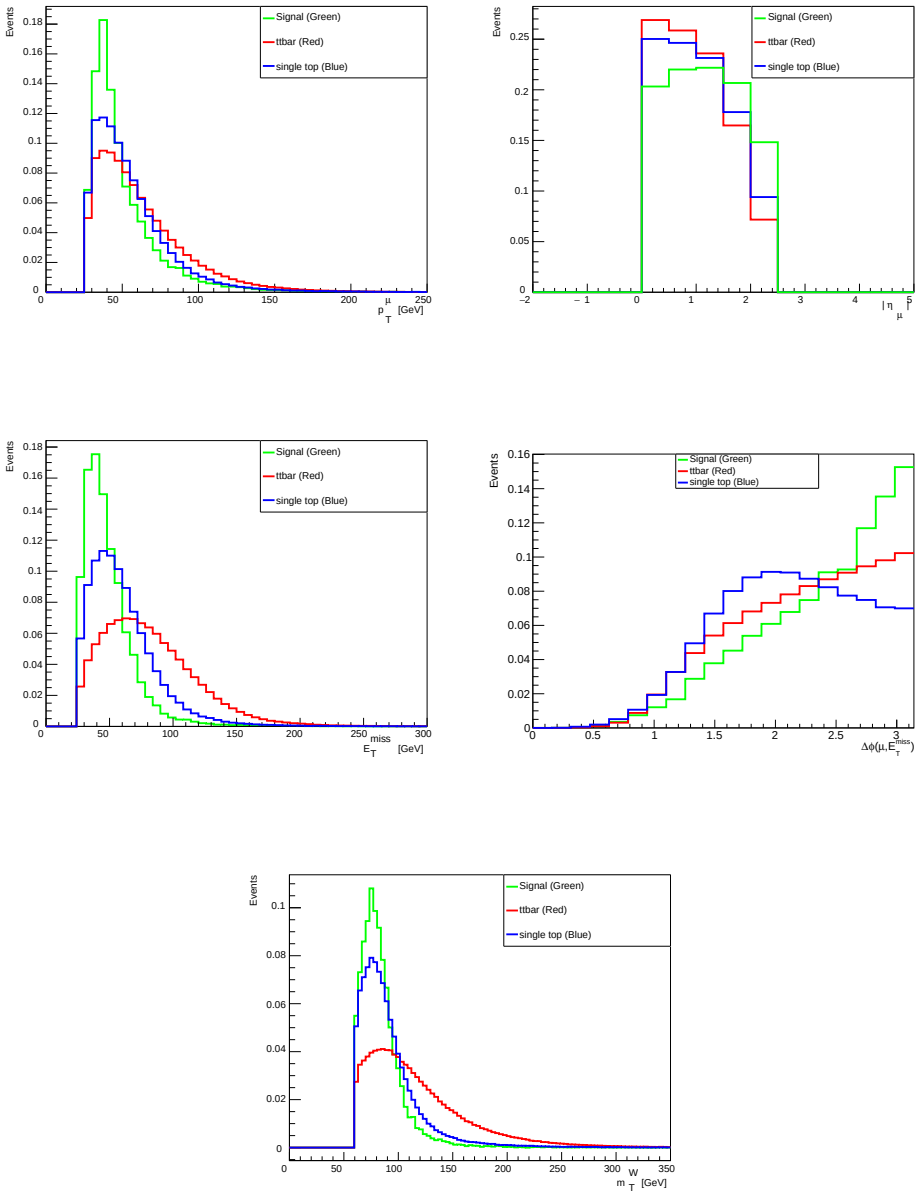

Figure 4.2: Distributions of: p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, and m_T^W in the 2B2J region for: nominal signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

Additionally, Figure 4.3 shows the distributions of the transverse momentum (p_T) and rapidity (Y) of the leading and sub-leading b-jets, and the *di-b*jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$.

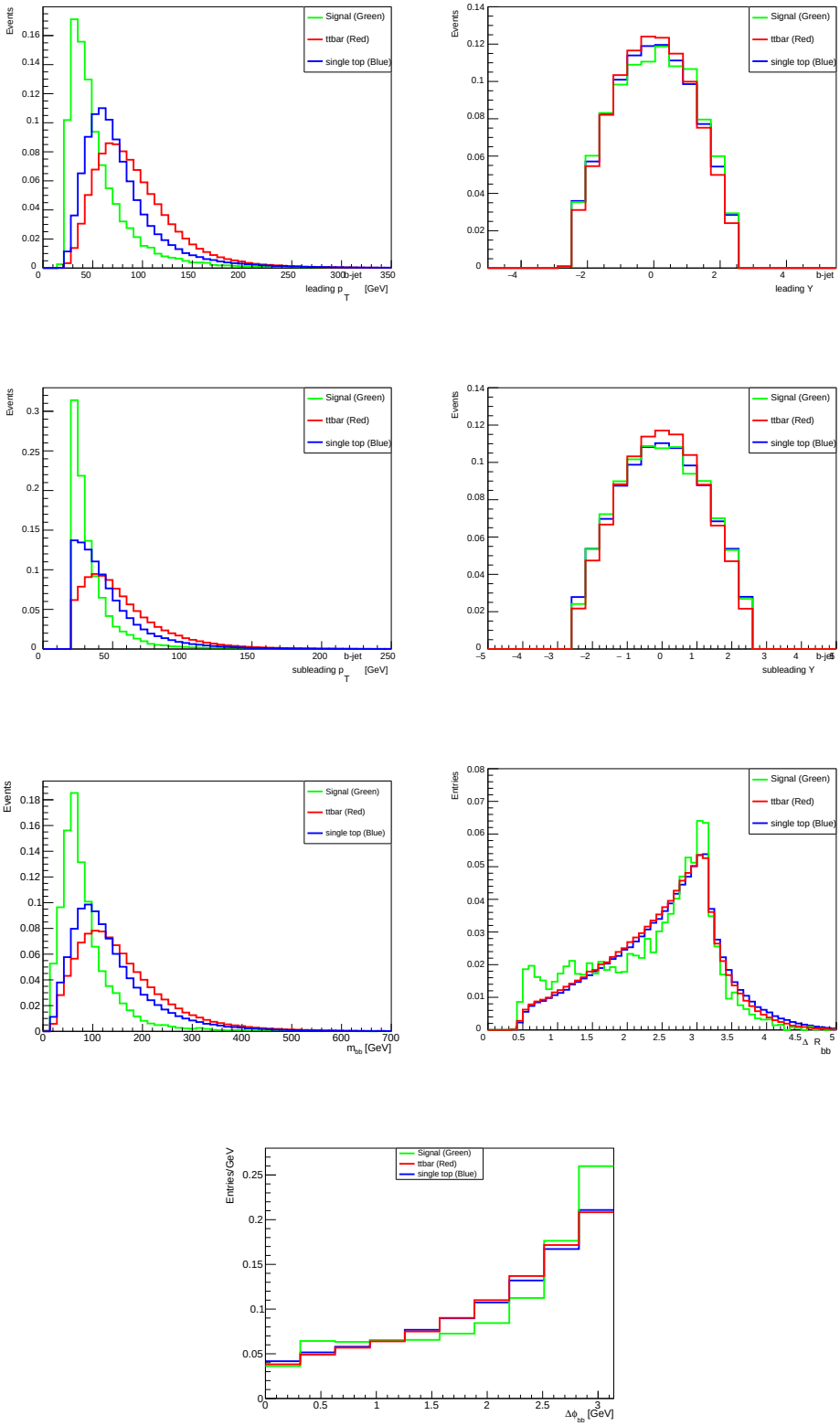

Figure 4.3: Distributions of: p_T and Y of the leading and sub-leading b-jets, as well as the *di*-$b$jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$ in the 2B2J region for: nominal signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

All the shown observables have some discriminating power. The signal (in green) and $t\bar{t}$ background (in red) distributions are the most distinct, with the single top process (in blue) exhibiting intermediate characteristics. Most of the b-jet observables shown in Figure 4.3 demonstrate good separation between signal and background. However, these observables are crucial for the final cross-section measurement. A specific cut on a NN, that uses the b-jet observables as input, might strongly distort the shape of the b-jet observables, which are the ones used for the differential cross-section measurements. Cross-section measurements in these analyses are performed in a fiducial phase space close to the detector level. An unfolding procedure is applied to correct detector-level distributions for efficiencies, resolutions, and small phase-space effects, transitioning from detector-level to particle-level (free from detector effects) distributions, which represent the cross-sections. The NN cut is treated as an inefficiency in the unfolding, which corrects for potential sculpting effects introduced by the NN. However, specific systematic uncertainties must be estimated for this. To minimize these uncertainties, large sculpting effects should be avoided. Therefore, as it is typically done in NN analyses in ATLAS, the observables measured in the cross-sections are not used as NN inputs variables. This is the reason why the b-jets observables are not included as inputs of the NN and the NN is trained on the five observables shown in Figure 4.2. Being the $t\bar{t}$ background the most copious (it is about four times larger than the single top in the 2B2J region) it has been decided to use only this background for training and testing the NN.

Scatter plots provide a visual representation of the relationships between different variables in the datasets, helping to identify patterns, correlations, and outliers in the data. Figure 4.4 shows the relationships between the five variables for the signal sample used as NN inputs: p_T^{μ}, E_T^{miss}, m_T^W, $|\eta_{\mu}|$, and $\Delta\phi(\mu, E_T^{miss})$. While Figure 4.5 shows the same relationship to the $t\bar{t}$ background.

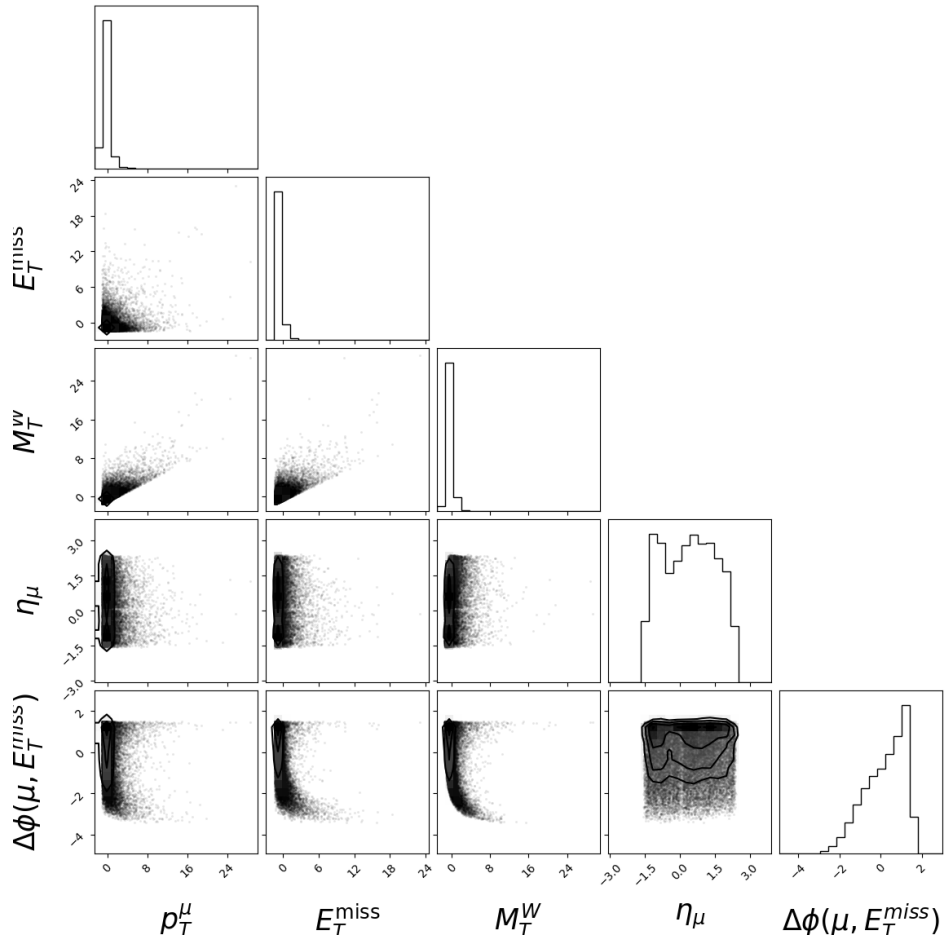

Figure 4.4: Scatter plots and distributions of the 5 features for the nominal signal MC sample.The distributions are shown after the scaling procedure described in Section 4.1

The diagonal plots from top left to bottom right show the distributions of each individual variable after the scaling procedure described in Section 4.1. These histograms provide an understanding of the distribution of each variable. Below the diagonal are scatter plots of two variables at a time. These plots can reveal any potential correlations or patterns between pairs of variables. For example, a scatter plot with points closely following a line suggests a strong linear correlation between the two variables. The scatter plots show different correlations between the observables in signal and $t\bar{t}$ samples.

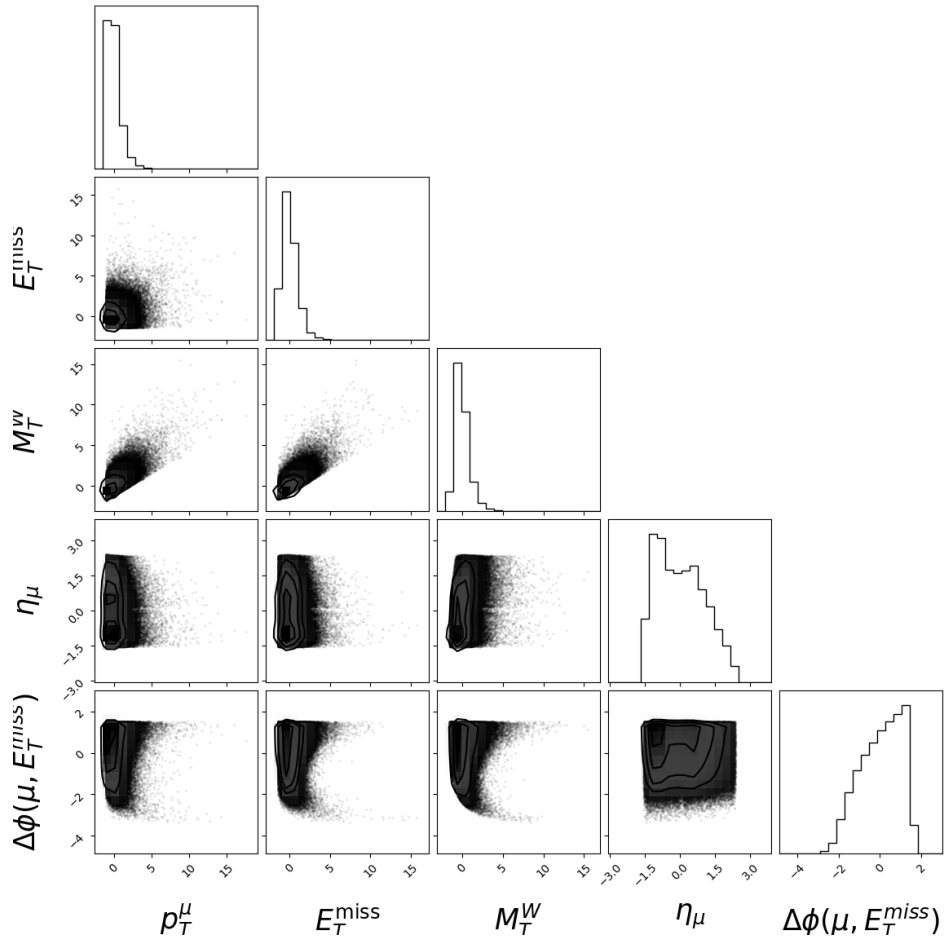

Figure 4.5: Scatter plots and distributions of the 5 features for the $t\bar{t}$ MC sample. The distributions are shown after the scaling procedure described in Section 4.1.

4.2.2 Optimization strategy for the model

An NN model is implemented using the five variables as inputs, discussed in the previous section. The technical details have been explained in Section 4.1. The number of signal events used for the training is 50k while 600k $t\bar{t}$ events are used, as explained in Section 4.2 these events correspond to the total amount of MC events available in the 2B2J region.

I performed an optimisation study to establish the best hyperparameters of the NN. I used a scan search approach. The hyperparameters considered for the scan are: the number of hidden layers, the number of nodes per layer, the

number of epochs and the learning rate. The chosen figure of merit for the optimisation is the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. This curve is generated by computing the rate of true positive signal events as a function of the false positive background events, and it is obtained by changing the NN value used to classify signal against background events. As explained in Appendix A, this is a method commonly used to evaluate the efficiency of a classifier. The best NN is the one providing the highest AUC on the validation sample. Table 4.1 shows the grid of values tested in the scan procedure. A total of 600 combinations have been tried.

Hyperparameters	**Values selected for optimization**
Hidden layers	1, 2, 3, 4, 5
Nodes per layer	5, 10, 15, 20, 25
Epochs	80, 100, 180, 250
Learning rate	$10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$

Table 4.1: Grid of values of the hyperparameters used in the optimisation procedure.

Figure 4.6 shows the AUC values found as a function of the combinations, that are labeled from 1 (i.e. 1 hidden layer, 5 nodes per layer, 80 Epochs, learning rate of 10^{-1}) to 600 (i.e. 5 hidden layers, 25 nodes per layer, 250 Epochs, learning rate of 10^{-6}). The plot shows that the maximum AUC is around 0.83 and more than one combination of the hyperparameters provides this value. Among the best combinations, I identified the most frequently occurring values for each hyperparameter, suggesting the best combination for optimal performance. I verified that the NN built with the most frequently occurring hyperparameter values is one of the NNs with the maximum AUC (0.83). This approach ensured a thorough search of the hyperparameter space, resulting in the identification of the best-performing configuration for the NN. The chosen set of hyperparameters is as follows: 2 hidden layers, 15 nodes

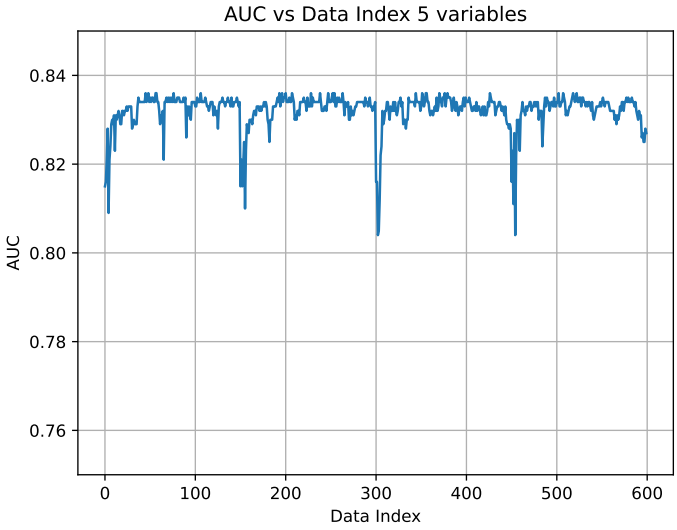

Figure 4.6: Variation in AUC for different hyperparameter configurations defined as described. These configurations are produced using 5 features with the nominal signal MC sample.

per layer, 100 epochs, and a learning rate of 10^{-3}. The fact that a shallow network (despite with a quite large number of nodes) is performing best, it is most probably due to the limited size of the training samples, deep and large NN are not performing better and are more prone to over-training. Then I reran the code using this optimised set of hyperparameters. The model was then trained using the identified optimal hyperparameters, and the resulting performance was evaluated to ensure that it indeed provided the best results.

4.2.3 Performances of the model

Loss plot

The plot of the loss (or error) as a function of the number of epochs is a crucial tool for evaluating the performance and stability of NN models. It illustrates how the loss (error), decreases with the increasing number of iterations (epochs), indicating how well the results of the training samples agree with the ones of the validation samples.

Figure 4.7 shows the loss as a function of the number of epochs. For this

analysis, four lines are displayed, corresponding to the training and validation losses for two different samples(train_0, valid_0, train_1, valid_1). These lines overlap almost perfectly, with only minor fluctuations, demonstrating that the model's performance is consistent and stable across both training and validation datasets. This close alignment and minimal fluctuation indicate robust learning and generalization capabilities, confirming that the model is well-optimized and reliable.

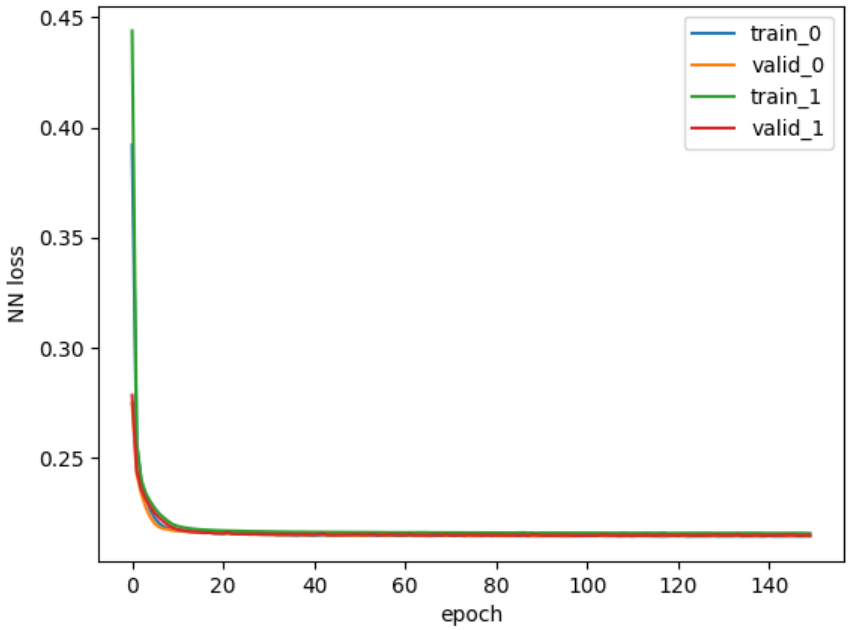

Figure 4.7: NN loss as a function of the training epochs in the training and validation samples. NN is trained using 5 features within the 2B2J region with the nominal signal MC sample.

Accuracy plot

Also accuracy plots as a function of the number of epochs offer valuable insights into the performance of the NN model across different training and validation scenarios. As explained in Appendix A, accuracy is defined as the fraction of correct predictions (true positive + true negative) with respect to the total number of predictions (true positive + true negative + false posi-

tive + false negative). By visually depicting the model's accuracy on both the training and validation sets at an increasing number of epochs it is possible to reach a comprehensive understanding of the model's learning dynamics and generalization capabilities. Specifically, the plots show trends in accuracy improvement, potential convergence issues, and signs of overfitting.

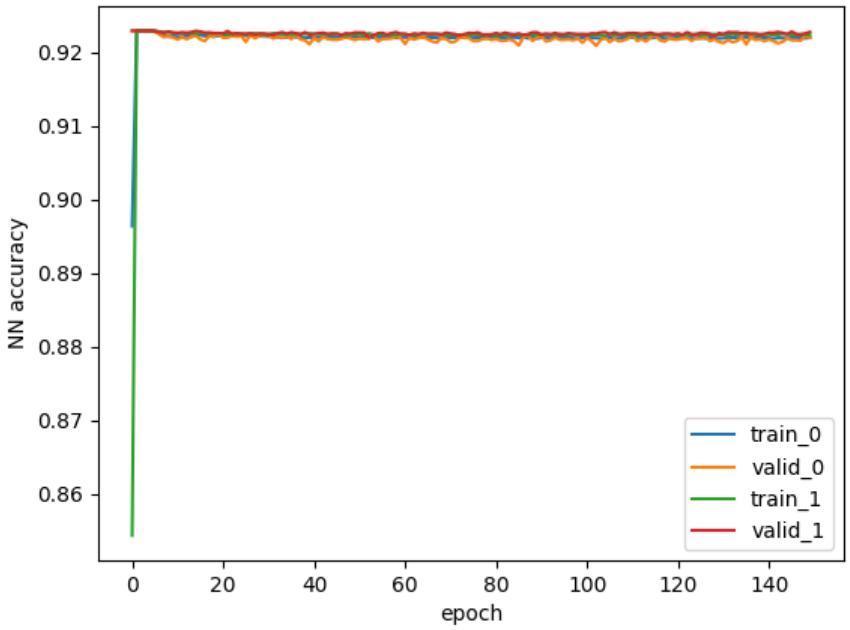

Figure 4.8: NN accuracy as a function of the training epochs in training and valid samples. NN is trained using 5 features within the 2B2J region with the nominal signal MC sample.

Figure 4.8 shows the accuracy as a function of the number of epochs for this analysis. As before, four lines are displayed, corresponding to the accuracy for the two training and the two validation samples built into the analysis. The training and validation accuracy lines are almost identical, with only very small fluctuations. This indicates that the model is performing consistently well on both the training and validation datasets, it is stable and not overfitting.
In general, the higher the accuracy and the lower the loss, the better the model. It is important to check that with the chosen number of epochs, the accuracy and the loss reach stable plateau values and the values of these quantities

evaluated on the validation samples are not diverging from those evaluated on the training samples. This is to avoid overtraining. The value of 100 epochs guarantees these conditions.

NN plot

Figure 4.9 shows the distributions of the NN scores as obtained for the test samples using signal and $t\bar{t}$ background shown in different colours. Signal and background distributions have their own distributions. Background distribution peaks at zero. The fact that the signal distribution doesn't peak at one should be due to the the limited size of the training sample.

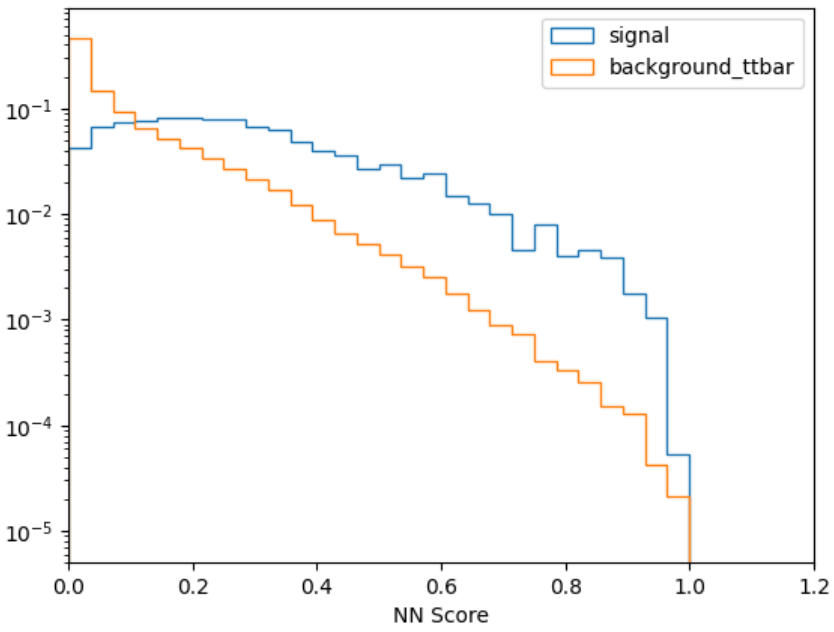

Figure 4.9: NN score distributions for nominal signal MC and $t\bar{t}$ background in the test sample. NN is trained using 5 features within the 2B2J region with the nominal signal MC sample.

This plot provides a clear visual representation of the model's ability to distinguish between the two categories. This pattern of scoring is consistent across all datasets (test, train, validation), further demonstrating the model's

robust performance. This result is a promising indicator of the model's potential utility in applications requiring accurate signal-background discrimination.

AUC plot

As discussed earlier in this chapter, AUC is a tool for evaluating the performance of binary classification models. Figure 4.10 shows the ROC curve of the NN model developed for this analysis using the hyperparameters defined in the optimisation procedure.

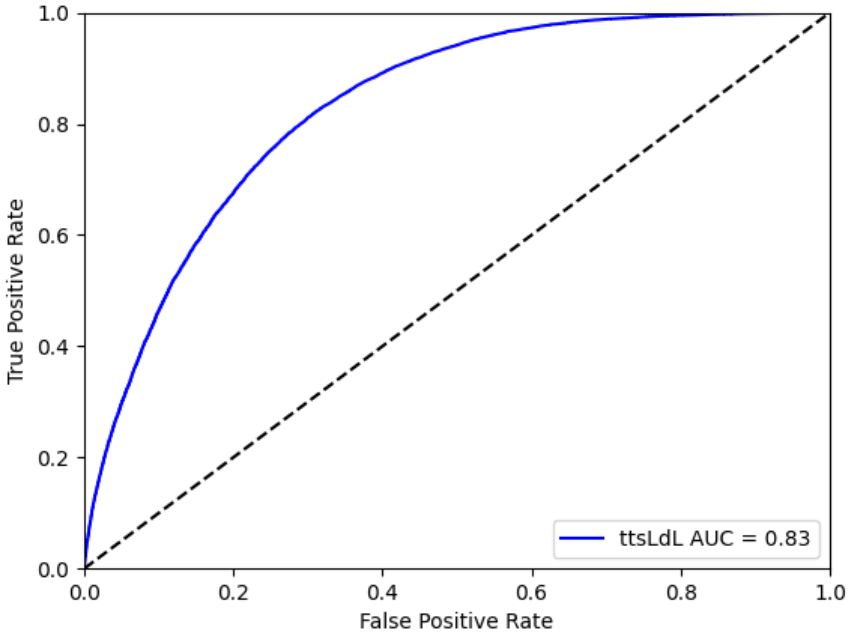

Figure 4.10: NN performance in terms of signal efficiency ("True Positive Rate") and $t\bar{t}$ background mis-identification efficiency ("False Positive Rate") obtained on the test sample. As figure of merit, AUC is also given. The nominal signal MC sample is used. NN is trained using 5 features within the 2B2J region.

It is evaluated using the testing sample. The AUC also on the testing sample is 0.83. The high AUC value suggests that the model has a very good ability to distinguish between the two classes of events.

Features importance

Feature importance is a technique that helps identify the contribution of each feature to the model's prediction performance. The methodology adopted here involves the permutation importance technique, which assesses the impact of shuffling individual features on the model's performance.
The permutation importance method involves the following steps:

• *Baseline Performance*: Initially, the baseline performance of the model is evaluated using a chosen metric, such as the AUC score, using the validation sample.

• *Feature Shuffling*: The values of a given feature are shuffled across all events in the validation sample, one feature is shuffled at a time. This step disrupts the relationship between that specific feature and the target variable while keeping the rest of the features intact.

• *Performance Evaluation*: After shuffling a feature, the performance of the model is revaluated using the validation sample. The decrease in performance compared to the baseline indicates the importance of the shuffled feature.

Features whose removal causes a larger decrease in the model's performance are more critical for the model's predictive power.

Figure 4.11 shows the importance of the features, ranked from the most significant, the m_T^W, to the least one, $|\eta_\mu|$.

4.2.4 Second signal simulated samples

In the final analysis it is important to assess the systematics related to the NN cut that has been developed using a given signal MC sample. The systematics on the NN cut can be addressed using a second signal MC sample with different characteristics with respect to the initial one. This is done in this analysis using the other MC sample available. In this section, all the main studies performed for the first signal sample (as presented in Sections 4.2.1, 4.2.2, and 4.2.3) are repeated and presented for the second signal sample

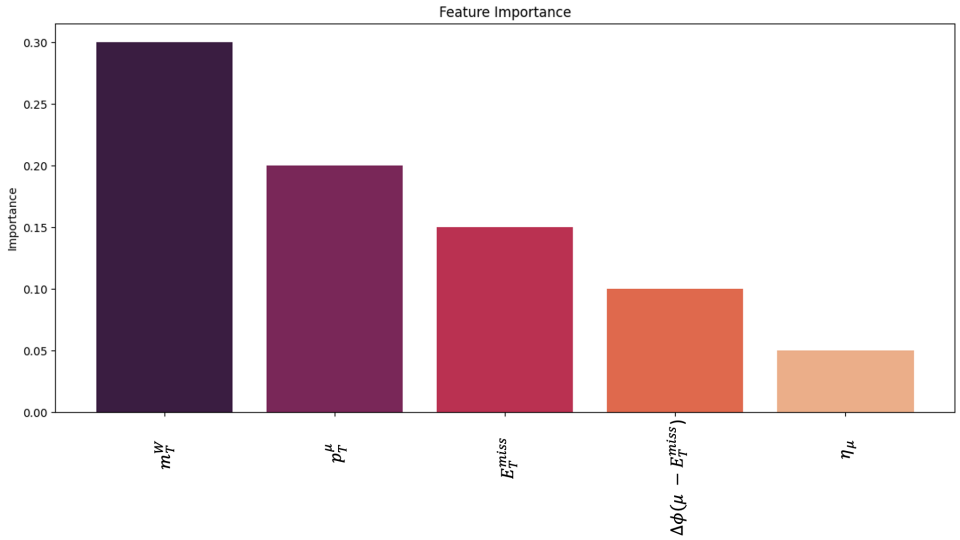

Figure 4.11: Feature importance plot for NN trained with 5 variables in 2B2J region. The nominal signal MC sample is used to train the NN.

(Sherpa 2.2.11).

Figure 4.12 illustrates the distributions of p_T^μ, $|\eta_\mu|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, and m_T^W. Additionally, Figure 4.13 shows the distributions of p_T and Y of the leading and sub-leading b-jets, as well as the *di-b*jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$. Comparing these figures with the ones shown for the nominal signal sample (MGaMC+Py8 FxFx), one can observe that the modeling of the five variables used as inputs for the NN looks very similar, as similar looks the modeling of the p_T and y for the leading and subleading b-jets, while some differences are noticed between the shape of the di-bjets observables (m_{bb}, $\Delta\phi_{bb}$, ΔR_{bb}). The reason of these differences needs to be investigated in the different technical choices of the two MC generators detailed in the previous Chapter. It will be extremely interesting to see how the measured data differential cross-sections of a function of these observables will appear to understand if data agree with one of the two MCs or if they disagree with both, this is a clear example that shows how these measurements provide precious inputs for the MC tuning.

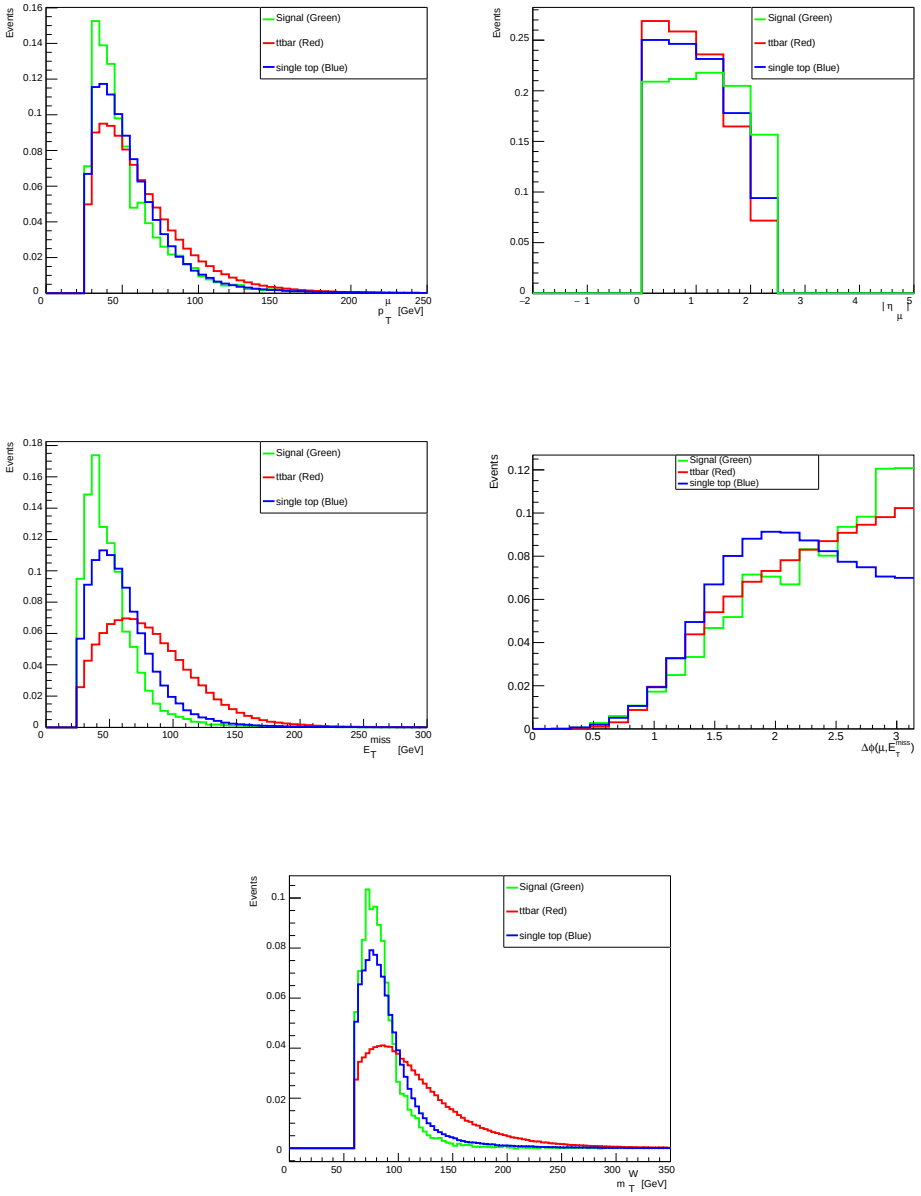

Figure 4.12: Distributions of: p_T^μ, $|\eta_\mu|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, and m_T^W in the 2B2J region for: second signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

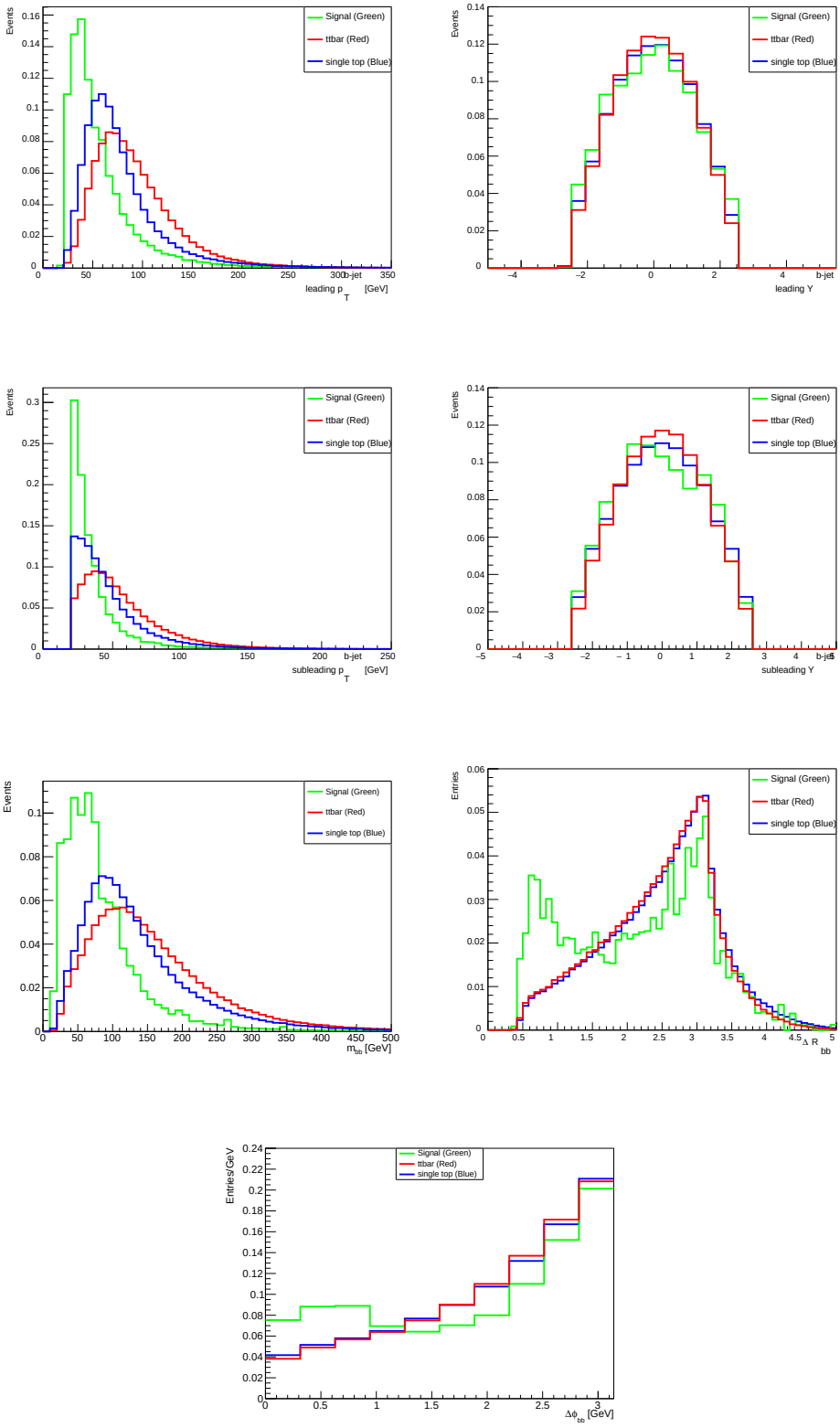

Figure 4.13: Distributions of: p_T and Y of the leading and sub-leading b-jets, as well as the *di*-$b$jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$ in the 2B2J region for: second signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

Figure 4.14 shows the relationships between five different variables : p_T^{μ}, E_T^{miss}, m_T^W, $|\eta_\mu|$, and $\Delta\phi(\mu, E_T^{miss})$ for signal dataset. The correlation path between the various observables looks similar between the two signal MC samples.

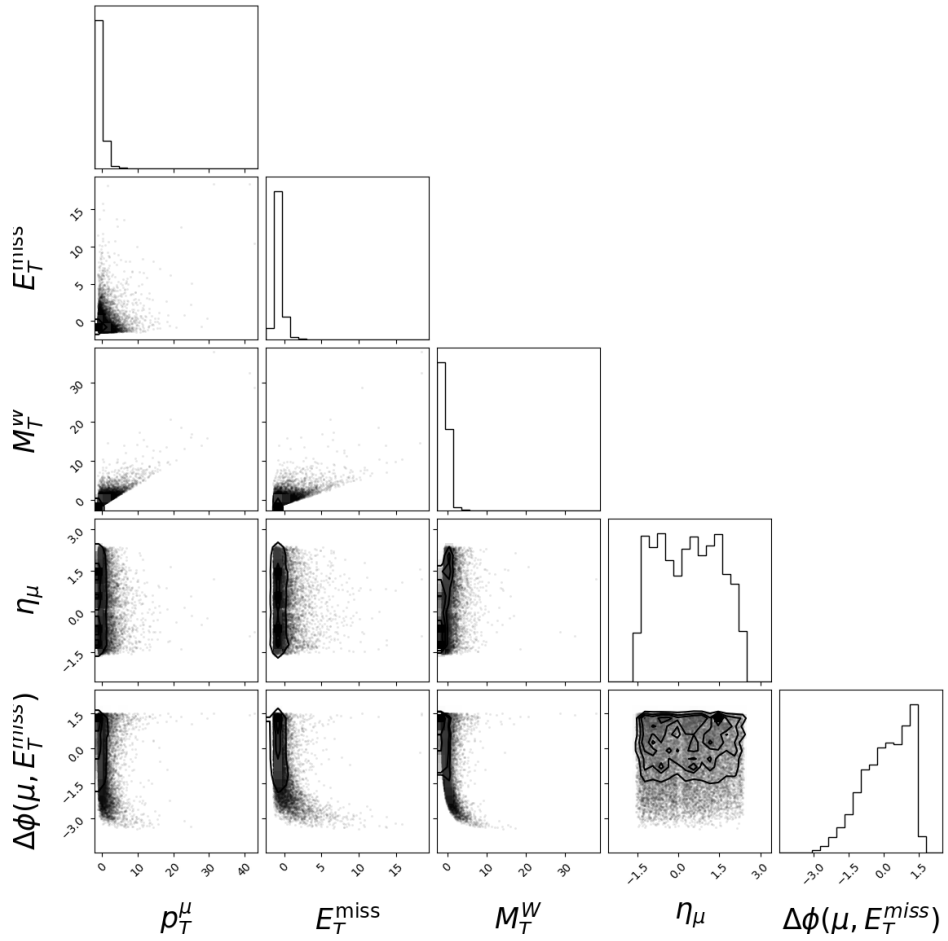

Figure 4.14: Scatter plots and distributions of the 5 features for the second signal MC sample.The distributions are shown after the scaling procedure described in Section 4.1

The same optimisation procedure described in Section 4.2.2 is performed here. The number of signal events used for the training is 30k while 600k $t\bar{t}$ events are used, again these events correspond to the total amount of MC events available in the 2B2J region. The chosen set of hyperparameters is as follows: 2 hidden layers, 15 nodes per layer, 80 epochs, and a learning rate of 10^{-3}.

Figure 4.15 shows the loss as a function of the number of epochs and Figure 4.16 shows the accuracy as a function of the number of epochs for the two training and validation samples, as before the accuracy and the loss reach stable plateau values at the number of chosen epochs and the values of these quantities evaluated on the validation samples are not diverging from those evaluated on the training samples.

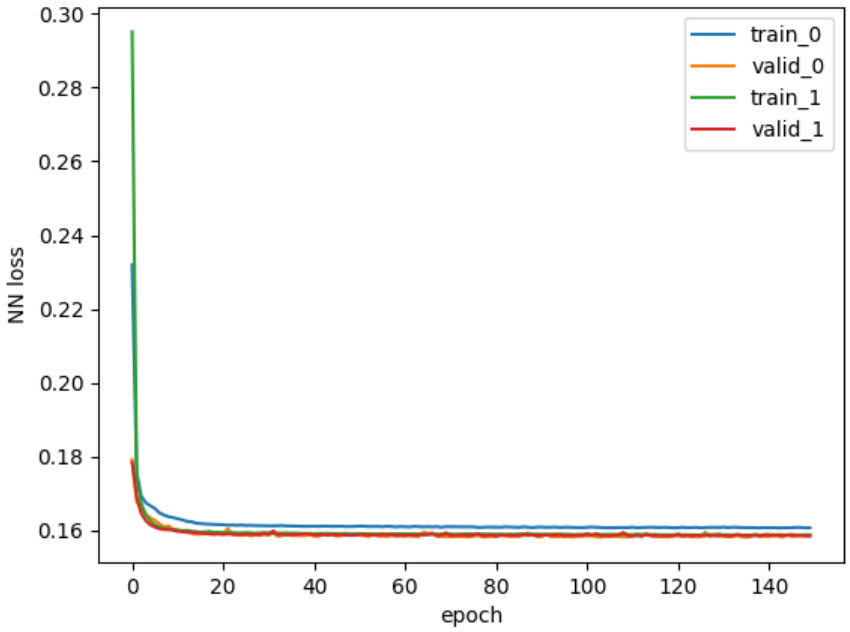

Figure 4.15: NN loss as a function of the training epochs in the training and validation samples. NN is trained using 5 features within the 2B2J region with the second signal MC sample.

Therefore the model shows consistent and reliable performances across different training and validation conditions and it is not overfitting to the training data.

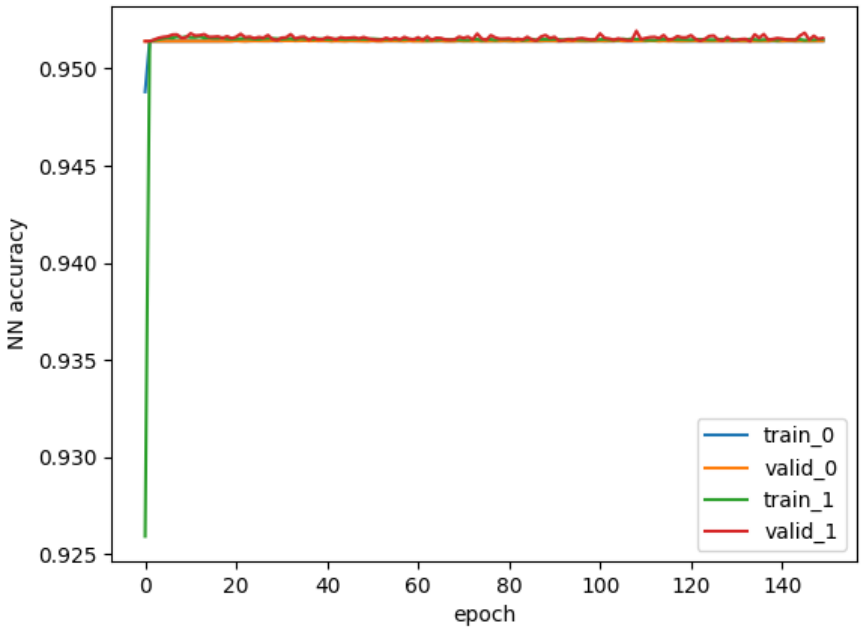

Figure 4.16: NN accuracy as a function of the training epochs in training and valid samples. NN is trained using 5 features within the 2B2J region.

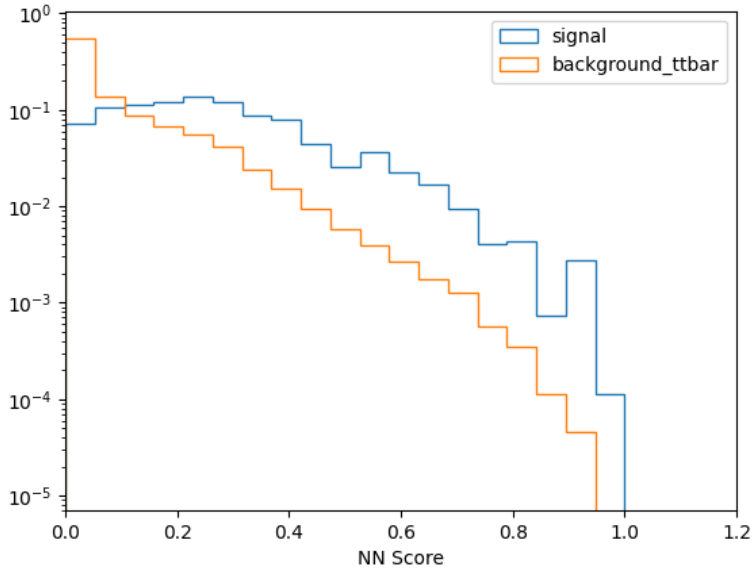

Figure 4.17: NN score distributions for signal MC and $t\bar{t}$ background in the test sample. NN is trained using 5 features within the 2B2J region.

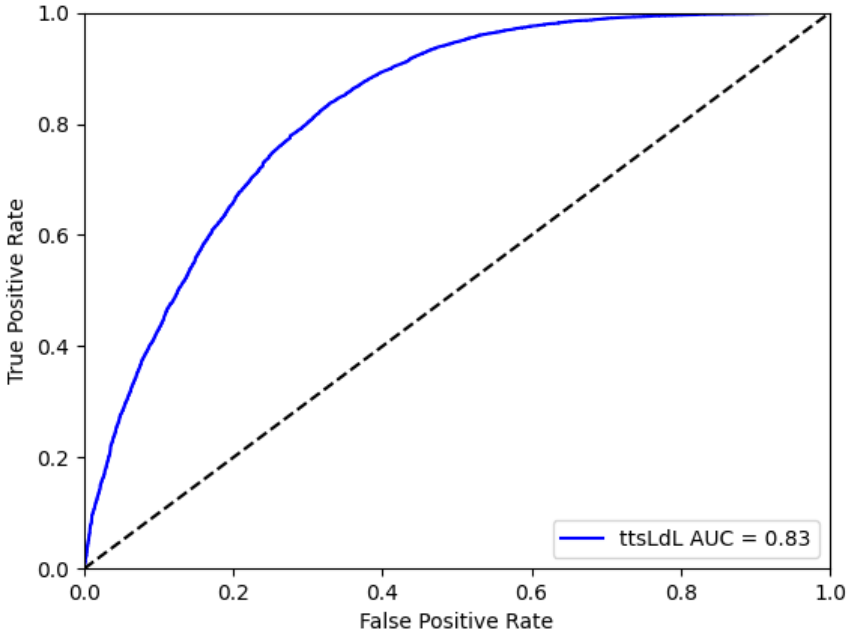

Figure 4.18: NN performance in terms of signal efficiency ("True Positive Rate") and $t\bar{t}$ background mis-identification efficiency ("False Positive Rate") obtained on the test sample. As figure of merit, AUC is also given. The signal MC sample is used. NN is trained using 5 features within the 2B2J region.

Figure 4.17 shows the distributions of the NN scores using the signal and $t\bar{t}$ test samples, while Figure 4.18 shows the ROC curve of the NN model developed with the hyperparameters defined in the optimisation procedure. The AUC on the testing sample is 0.83, matching the value obtained from the first signal Monte Carlo (MC) sample.

Figure 4.19 shows the importance of the features, ranked from the most significant to the least one. The trend of feature importance for the 5 variables is consistent across both signal MC datasets.

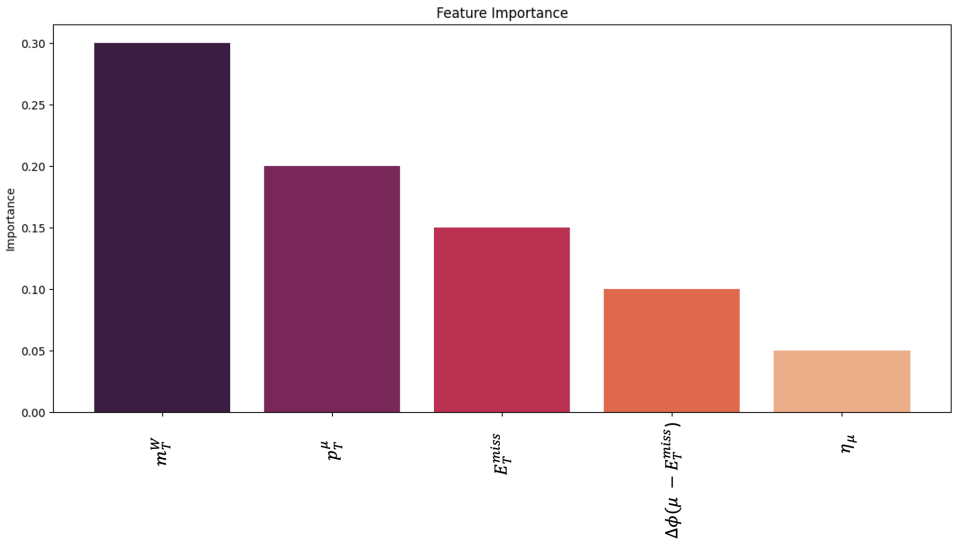

Figure 4.19: Feature importance plot for NN trained with 5 variables in 2B2J region. The signal MC sample is used to train the NN.

The NN developed with the two signal MC datasets looks very similar either in terms of optimal hyperparameters (both uses 2 hidden layers and 15 nodes) and in terms of performances: same AUC, and same ranking of feature importance. This implies that the systematics related to the NN cut on the final cross-section analysis will not be large.

4.2.5 5-features NN working point

In order to identify good working points for the NN cut to be applied on the detector level distributions, Table 4.2 reports some of the results of ROC curve obtained using the nominal MC samples (shown in Figure 4.10).

Signal Efficiency(%)	$t\bar{t}$ **Rejection**(%)
70	78.6
80	71.0
85	65.7
90	58.7
95	48.1
99	26.8

Table 4.2: Signal efficiency versus $t\bar{t}$ rejection for different NN cuts. NN is trained with 5 variables in 2B2J region. The nominal signal MC is used.

The Table reports, for a given value of the NN, the signal efficiency (here computed exactly as the fraction of True Positive) and the corresponding background rejection (here computed as the unity minus the fraction of False Positive). This table helps to identify possible working points for the NN cut to be applied at detector level in the measurements. The 2B2J selection is already rejecting a large fraction of signal events (the expected number of signal events with the full Run-2 dataset is about 28K, as shown in Table 3.5), therefore a tight cut (at very large background rejection, but also quite reduced signal efficiency) would make impossible the measurement of differential cross-sections. Considering that the first Z+2b-jets differential cross-section measurement in Run-2 (with 36 fb^{-1}), that has features very similar to this final state, has been conducted with about 20K events using a binning for the differential cross-section close to the one shown in Section 3.4. A NN cut, selecting signal events between 80% and 70%, with a $S/t\bar{t}$ ratio between 1.1 and 1.3 and an overall S/B between 0.5 and 0.6, as shown in Table 4.3, seems adequate for this analysis. In the next, the working point at 70% signal efficiency is used for the comparisons with the other approaches.

Table 4.3 is obtained with 5-features NN using the nominal signal MC sample, while Table 4.4 reports the results obtained using the second MC. As expected the results of the second MC sample are very close to those of the first sam-

ple, indicating that the systematics related to the signal model in the final cross-section measurement will be small.

S.Eff(%)	$t\bar{t}$ **Rej.**(%)	S **evts**	$t\bar{t}$ **evts**	B **evts**	$S/t\bar{t}$	S/B
70	78.6	20010	15810	33920	1.3	0.6
80	71.0	22860	21431	44820	1.1	0.5

Table 4.3: Signal efficiency versus $t\bar{t}$ rejection for different NN cuts. NN is trained with 5 variables in 2B2J region. This is done with training the NN on the nominal signal MC.

S.Eff(%)	$t\bar{t}$ **Rej.**(%)	**S evts**	$t\bar{t}$ **evts**	S/$t\bar{t}$
70	77.6	17500	16550	1.2
80	70.5	20000	21800	0.9

Table 4.4: Signal efficiency versus $t\bar{t}$ rejection for different NN cuts. NN is trained with 5 variables in 2B2J region. This is done with training the NN on the second signal MC.

4.3 NN for the inclusive selection (Geq2BiJ)

In this section, it is discussed the development of the NN for the inclusive selection, Geq2BiJ, which has already been described in chapter 3. Events passing the selection Geq2BiJ contain a W candidate decaying in a muon and a neutrino and at least 2 b-jets in the central part of the detector and no additional cut on the number of jets in the central or forward region is performed. The primary signal MC sample used to perform this study is MGaMC+Py8 FxFx. The input observables of the NN model and the optimisation of the NN are presented in Section 4.3.1. Two NN models have been built in this region: a NN with 7 features, discussed in Section 4.3.2, and a more efficient NN with 11 features, discussed in Section 4.3.3. The studies based on a second signal MC sample (Sherpa 2.2.11) are reported in Section 4.3.4.

4.3.1 Selected features for the model

After applying criteria that define the Geq2BiJ selection, a study has been conducted to identify the observables that can effectively distinguish signal events from top background events. The physics objects present in events passing the Geq2BiJ selection are: the muon, the missing transverse energy, the 2b-jets, plus in case other jets, some of them even tagged as b-jets. Using these objects a set of kinematic variables has been studied. As done for the work on the 2B2J region, also in this case the shapes of the signal and background event distributions of the various observables have been compared to identify the ones that can used for the separation. The five observables used in the 2B2J case are again investigated in the new phase space, and in addition thanks to the presence of additional jets other observables have been explored. At this stage both the dominant $t\bar{t}$ background and the smaller single top background are considered.

Figure 4.20 illustrates the distributions of p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, and m_T^W. In addition to these five observables, other six new ones have been explored, they are: the number of central jets (i.e. jets with $|y| < 2.5$), N_{jet}^{centr}, the number of forward jets (i.e. jet with $2.5 < |y| < 4.5$), N_{jet}^{fwd}, the scalar sum of the transverse momenta of central jets H_T^{centr} ($H_T^{centr} = \sum_{i=1}^{N_{jets}} |p_{T,i}|$), the sum of the masses of central jets M_{centr} ($M_{centr} = \sum_{i=1}^{N_{jets}} m_{jet}$), the scalar sum of the transverse momenta of all jets H_T^{all}, and the sum of the masses of all jets M_{all}. Also these additional six observables are shown in Figure 4.20.

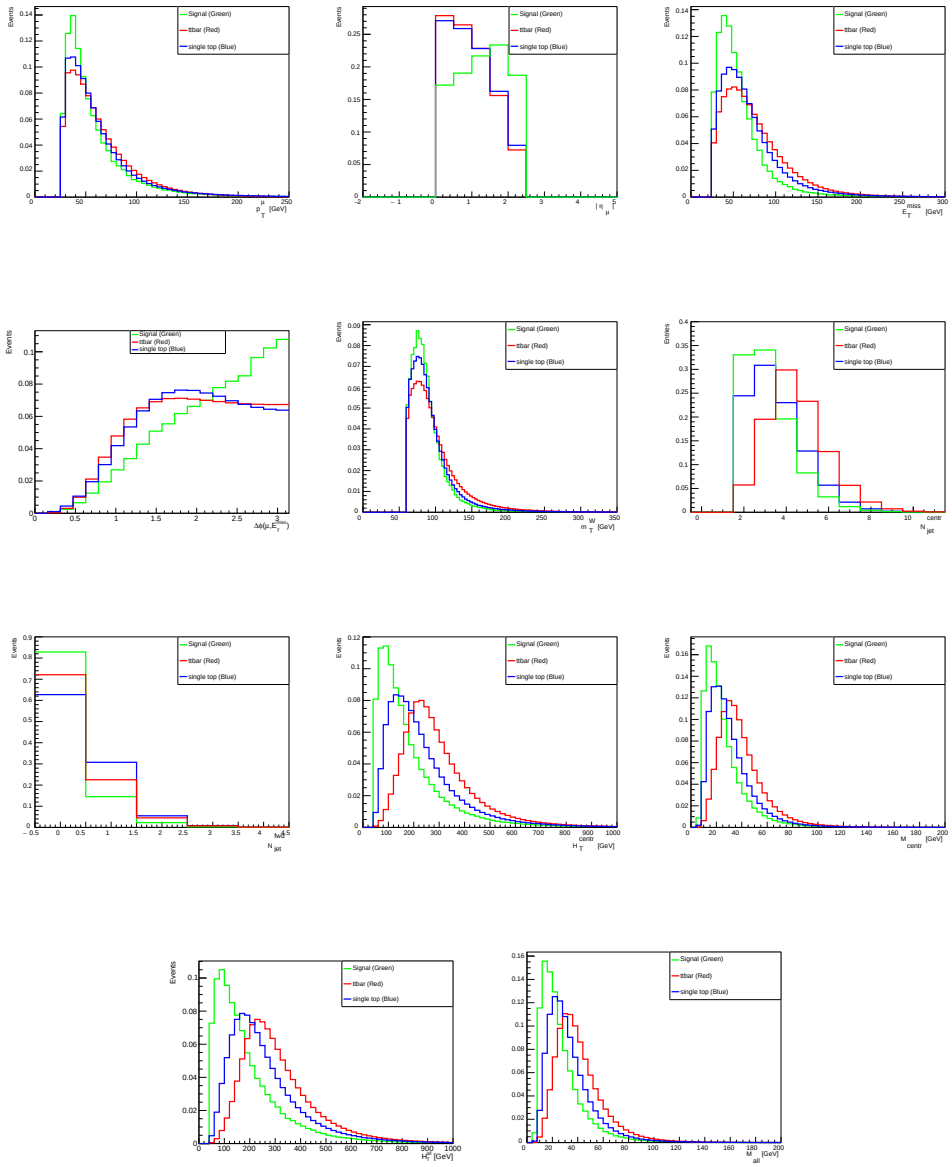

Figure 4.20: Distributions of: p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, m_T^W, N_{jet}^{centr}, N_{jet}^{fwd}, H_T^{centr}, M_{centr}, H_T^{all}, and M_{all} in the Geq2BiJ region for: nominal signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

Additionally, Figure 4.21 shows the distributions of p_T and Y of the leading and sub-leading b-jets, as well as the *di-b*jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$.

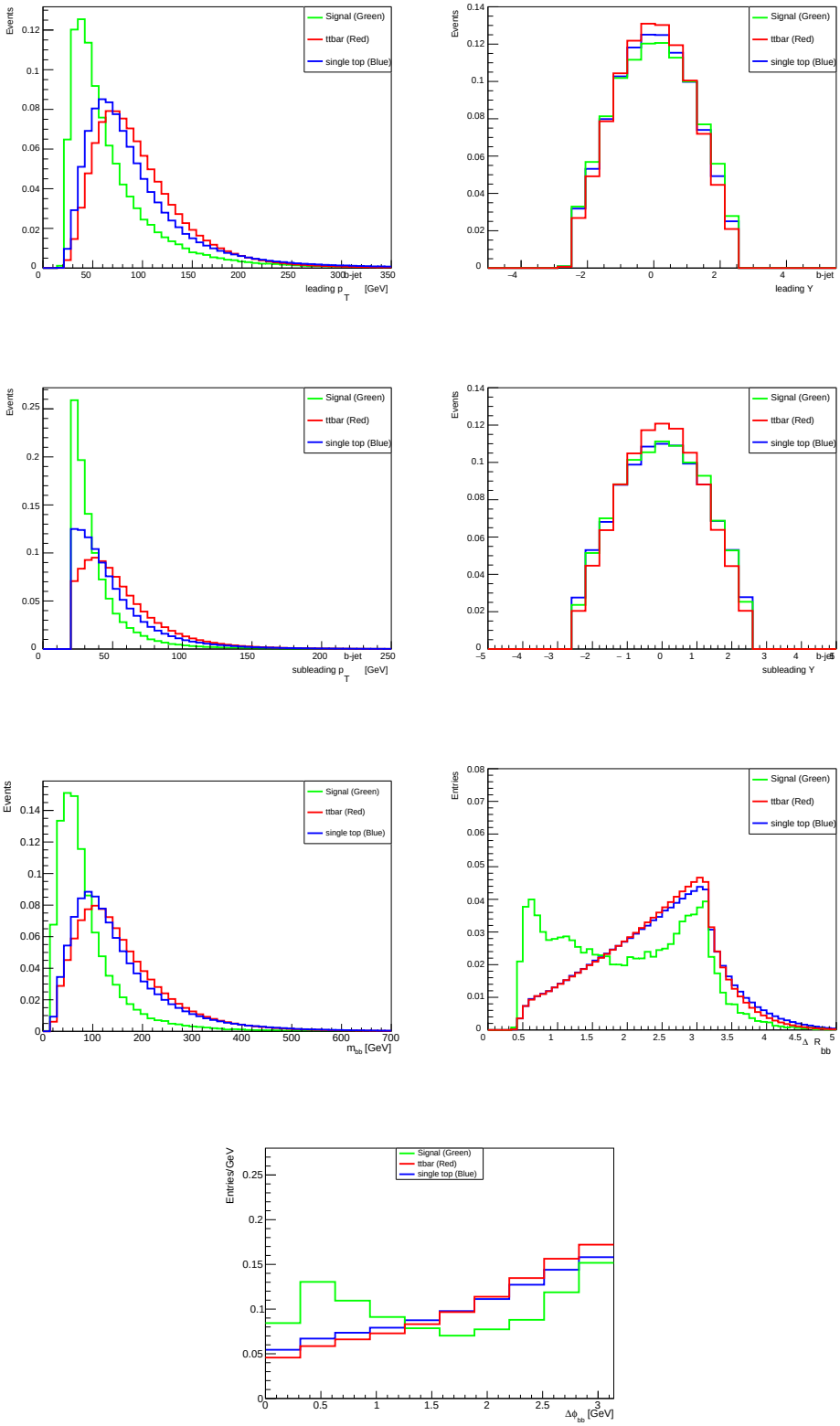

Figure 4.21: Distributions of: p_T and Y of the leading and sub-leading b-jets, as well as the *di*-$b$jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$ in the Geq2BiJ region for: nominal signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

As expected also in this region, all the shown observables have some discriminating power. The signal (in green) and $t\bar{t}$ background (in red) distributions are the most distinct, with the single top process (in blue) exhibiting intermediate characteristics. Again the b-jet variables are not used as NN input variables for the reason explained in the previous section. Comparing the b-jets distribution of this region with the ones shown for the 2B2J region (Figure 4.3) one might notice a difference in the angular 2-b-jets distributions, this is somehow expected since in the 2B2J region in the final state there are only 2 b-jets and no additional jets, while in the Geq2BiJ region the number of jets is larger, therefore the angular correlation between b-jets changes. In general measurements of the angular separation between the two leading b-jets allow for the characterisation of the hard radiation at large angles and the soft radiation for collinear emissions, in the Geq2BiJ region due to the presence of additional jets, the soft radiation at smaller angles is enhanced with respect to the 2B2J region, dominated by the hard radiation at large angles.
The scatter plots together with the distributions of the variables after the scaling are shown in Figure 4.22 for the nominal signal and in Figure 4.23 for the $t\bar{t}$.

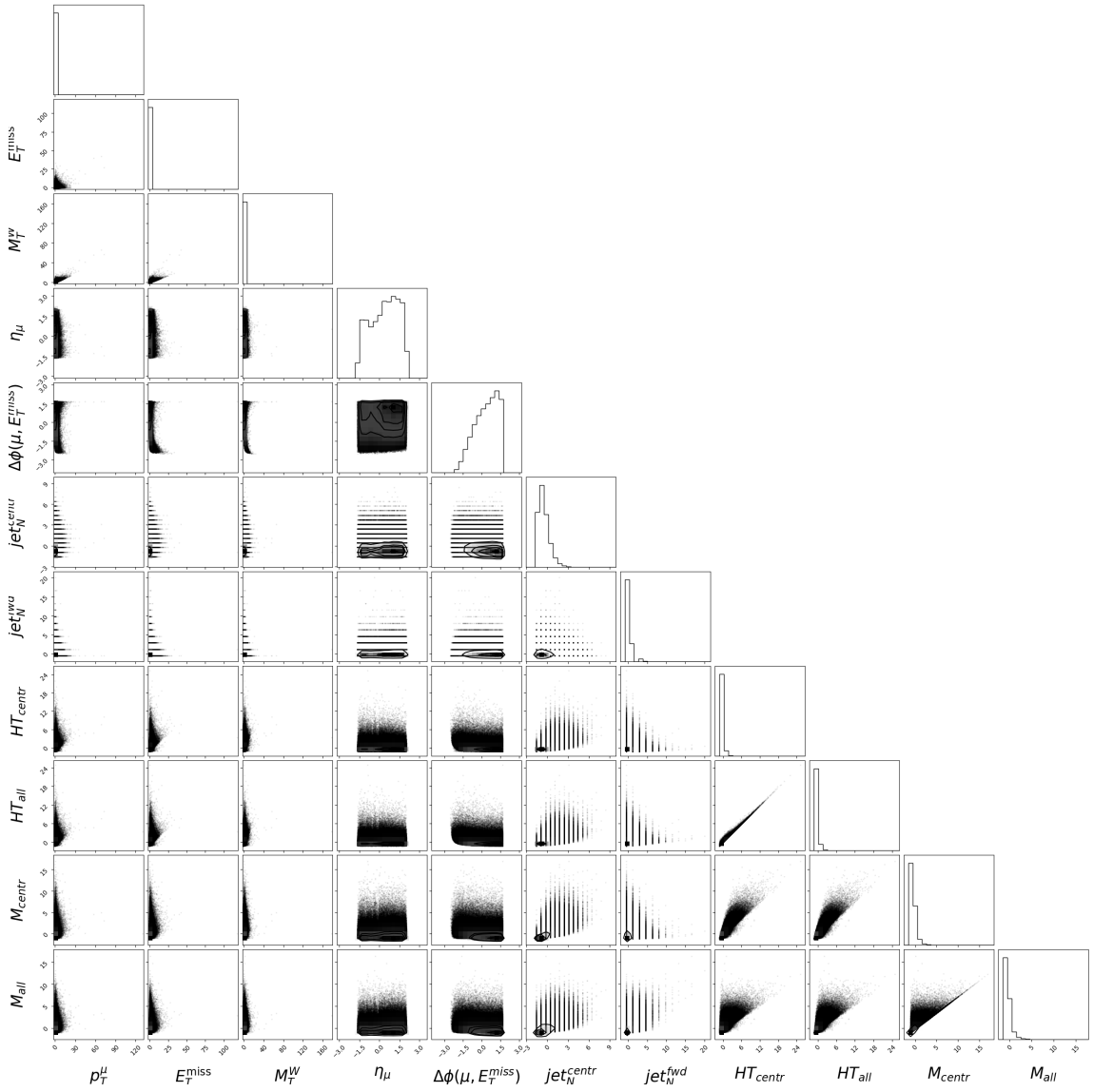

Figure 4.22: Scatter plots and distributions of the 11 features for the nominal signal MC sample.The distributions are shown after the scaling procedure described in Section 4.1

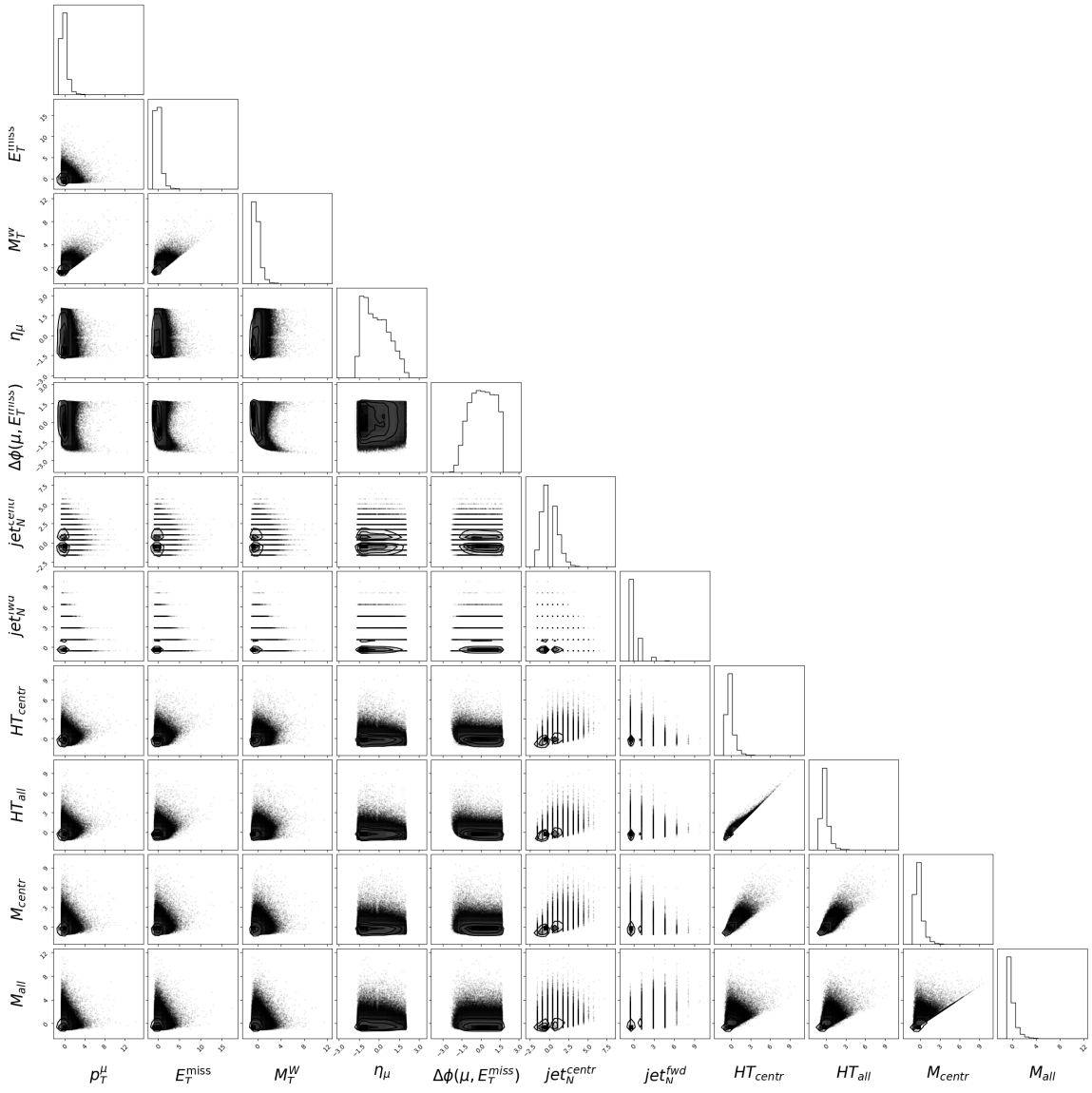

Figure 4.23: Scatter plots and distributions of the 11 features for the $t\bar{t}$ MC sample.The distributions are shown after the scaling procedure described in Section 4.1.

4.3.2 NN with 7 features

A first NN model has been implemented using seven observables: the five ones already used for the NN of the 2B2J region plus the number of central and forward jets, N_{jet}^{centr} and N_{jet}^{fwd}. The number of signal events used for the training is 0.8M while 15M $t\bar{t}$ events are used, as explained in Section 4.2 these events correspond to the total amount of MC events available in the Geq2BiJ

region. The chosen set of hyperparameters after the optimisation study is as follows: 2 hidden layers, 25 nodes per layer, 100 epochs, and a learning rate of 10^{-5}.

Figure 4.24 shows the loss as a function of the number of epochs, while Figure 4.25 shows the accuracy as a function of the number of epochs.

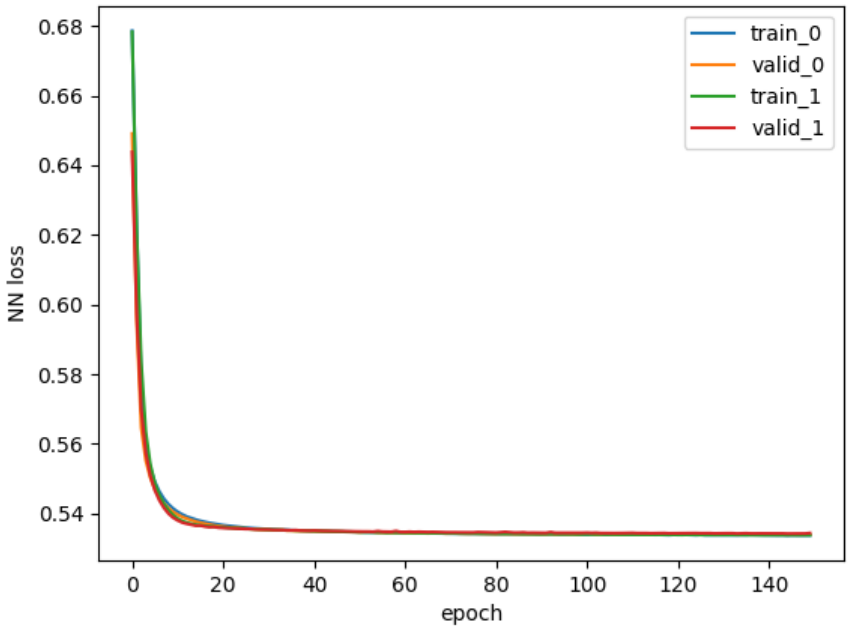

Figure 4.24: NN loss as a function of the training epochs in the training and validation samples. NN is trained using 7 features within the Geq2BiJ region.

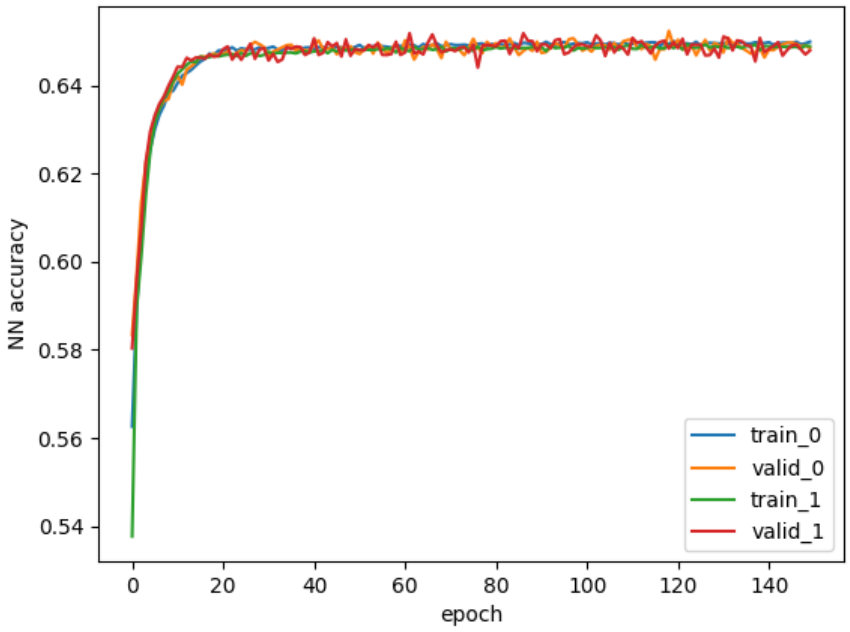

Figure 4.25: NN accuracy as a function of the training epochs in training and valid samples. NN is trained using 7 features within the Geq2BiJ region.

As before the plots are done for the two training and the two validation samples, and again the accuracy and the loss reach stable plateau values at the number of chosen epochs and the values of these quantities evaluated on the validation samples are not diverging from those evaluated on the training samples. Therefore the model is robust and it is not overfitting to the training data.

Figure 4.26 shows the distributions of the NN scores using the signal and $t\bar{t}$ test samples. Again good separation is observed.

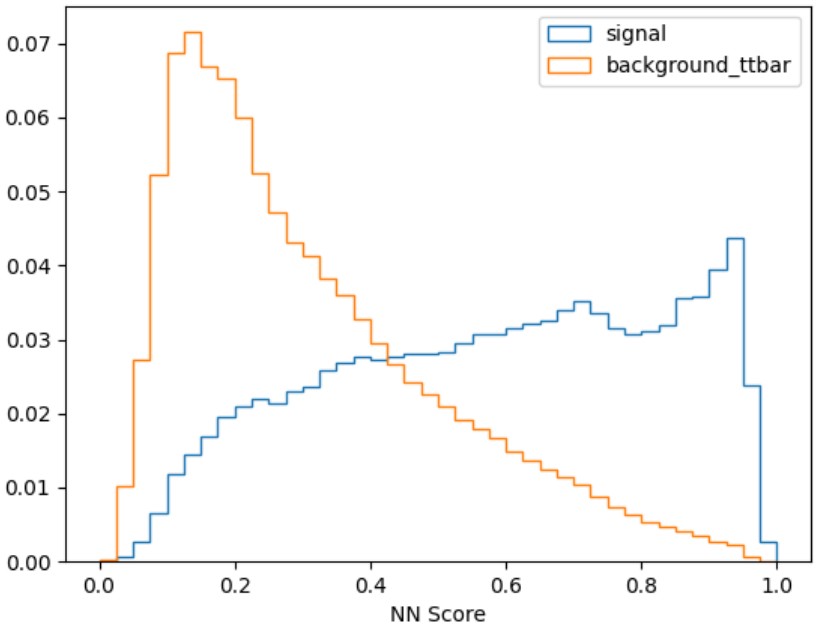

Figure 4.26: NN score distributions for nominal signal MC and $t\bar{t}$ background in the test sample. NN is trained using 7 features within the Geq2BiJ region.

Figure 4.27 shows the ROC curve of the NN model evaluated using the testing sample, the AUC on the testing sample is 0.80.

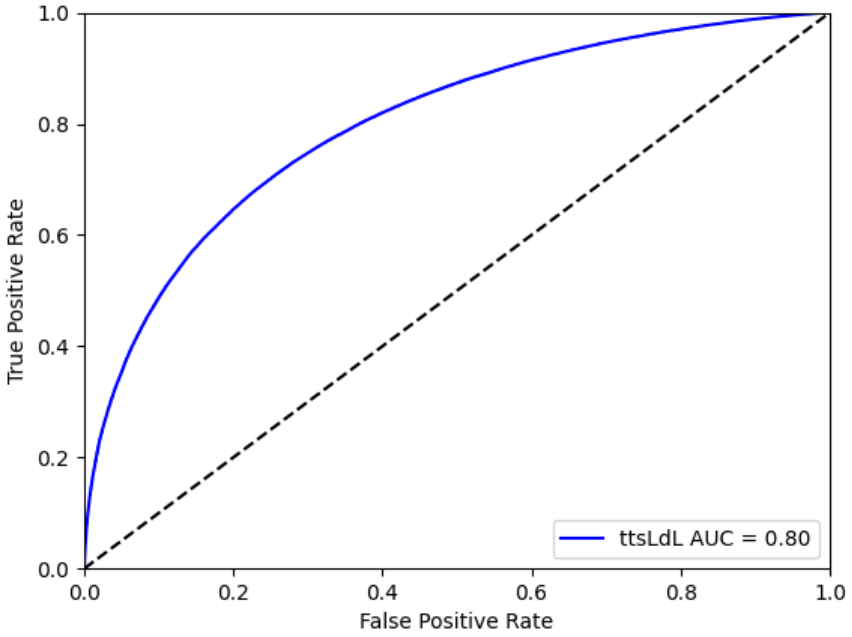

Figure 4.27: NN performance in terms of signal efficiency ("True Positive Rate") and background efficiency ("False Positive Rate") obtained on the test samples against all backgrounds considered in this study. As figure of merit, the Area Under the Curve (AUC) is also given. NN is trained using 7 features within the Geq2BiJ region.

Figure 4.28 shows the importance of the features, the most significant is η_μ, while the least important is N_{jet}^{fwd}.

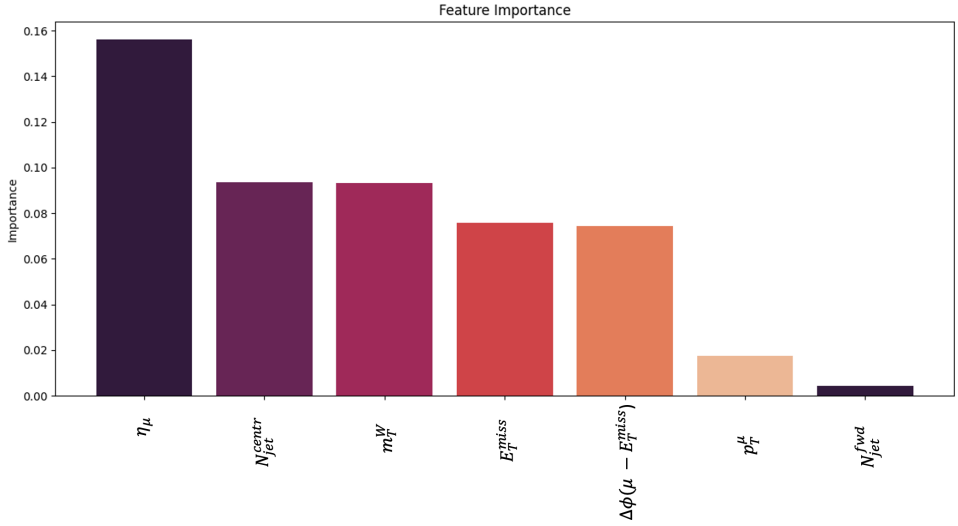

Figure 4.28: Feature importance plot for NN trained with 7 variables in Geq2BiJ region. The nominal signal MC sample is used to train the NN.

Comparison with 5-features NN approach

It is interesting to compare the 7-features NN approach developed for the Geq2BiJ region with the approach of the 5-features NN developed for 2B2J region, this is done by selecting two signal efficiency working point in the 7-features NN approach: the first corresponding to the initial number of signal events in the 2B2J region shown in Table 3.7 (28580 events corresponding an efficiency of 28.4% with respect to the signal events in the Geq2BiJ region) and the second corresponding to the number of events in the 2B2J region after applying an NN cut at 70% signal efficiency (20010 events corresponding to an efficiency of 0.7×28.4% with respect to the signal events in the Geq2BiJ region). Results are shown in Table 4.5. One can notice that although, as expected, the 7-features NN approach improves the $S/t\bar{t}$ ratio with respect to the 2B2J cut (0.5 against the 0.4 shown in Table 3.5), the combination of 2B2J plus the 5 variable NN cut provides a better $S/t\bar{t}$ ratio with respect to the 7 variable NN cut (1.3 shown in Table 4.3 to the current o.7).

S.Eff(%)	$t\bar{t}$ **Rej.(%)**	**N. of signal evnt**	**N. of** $t\bar{t}$ **evnt**	$S/t\bar{t}$
28.4	96.8	28580	60410	0.5
20 (=28.4× 0.7)	98.4	20010	30200	0.7

Table 4.5: Signal efficiency cut versus $t\bar{t}$ rejection for NN trained with 7 variables in Geq2BiJ region. This is done with training the NN on the nominal signal MC.

4.3.3 NN with 11 features

Considering that in the Geq2BiJ region in addition to the decay products of the W boson and the 2 b-jets there are additional jets in the final state, I developed a NN including in addition to the seven observables presented in the previous section also other ones related to jets, they are the last ones presented in Figure 4.20, namely: H_T^{center}, H_T^{all}, M_{centr}, and M_{all}.
The number of signal events used for the training is 0.8M while 15M $t\bar{t}$ events are used, as explained in Section 4.2 these events correspond to the total amount of MC events available in the Geq2BiJ region. The chosen set of hyperparameters after the optimisation study is as follows: 2 hidden layers, 20 nodes per layer, 250 epochs, and a learning rate of 10^{-3}.
Figure 4.29 shows the loss as a function of the number of epochs, while Figure 4.30 shows the accuracy as a function of the number of epochs. As before the plots are done for the two training and the two validation samples, and again the accuracy and the loss reach stable plateau values at the number of chosen epochs and the values of these quantities evaluated on the validation samples are not diverging from those evaluated on the training samples. Therefore the model is robust and it is not overfitting to the training data.

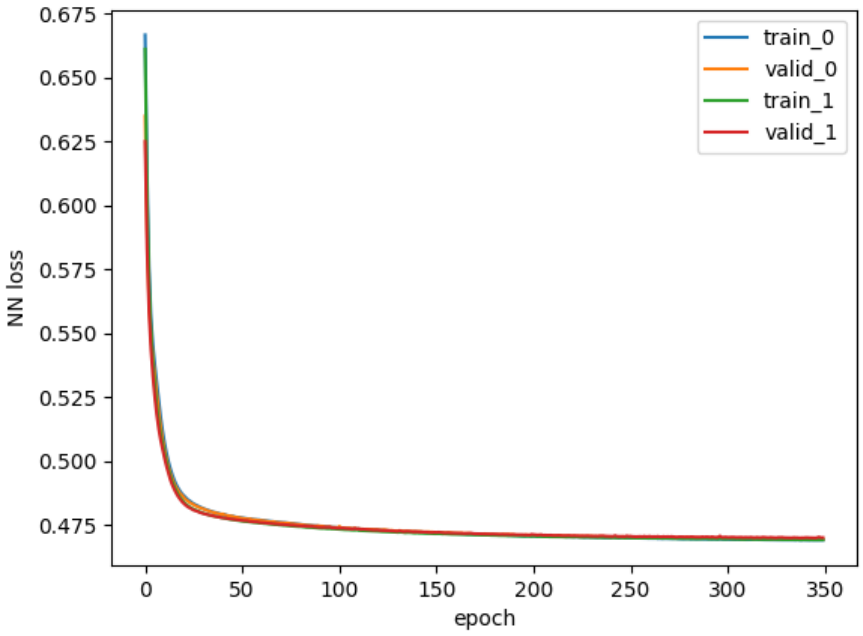

Figure 4.29: NN loss as a function of the training epochs in the training and validation samples. NN is trained using 11 features within the Geq2BiJ region.

Figure 4.31 shows the distributions of the NN scores using the signal and $t\bar{t}$ test samples. A good separation is observed, with the distribution of the background peaking at 0 and the one of the signal peaking at 1.

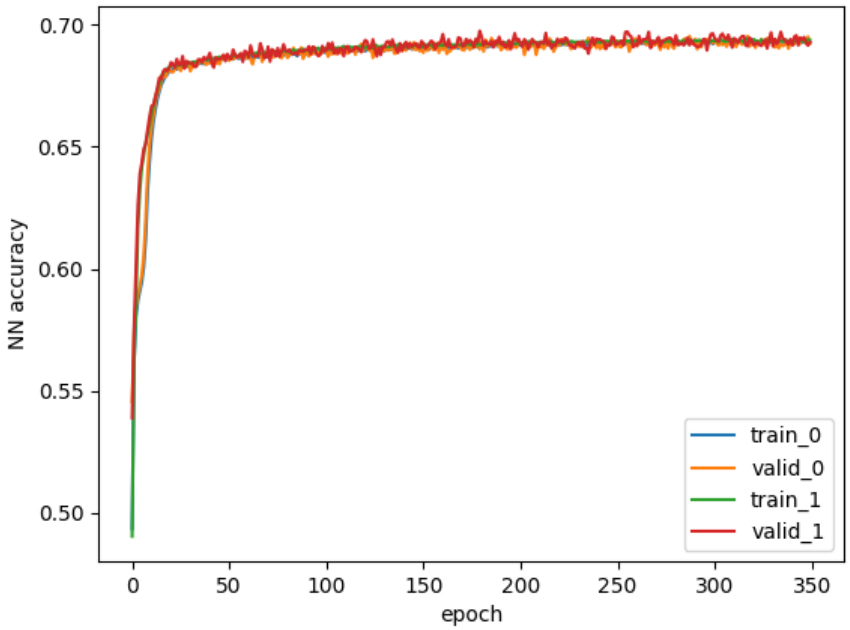

Figure 4.30: NN accuracy as a function of the training epochs in training and valid samples. NN is trained using 11 features within the Geq2BiJ region.

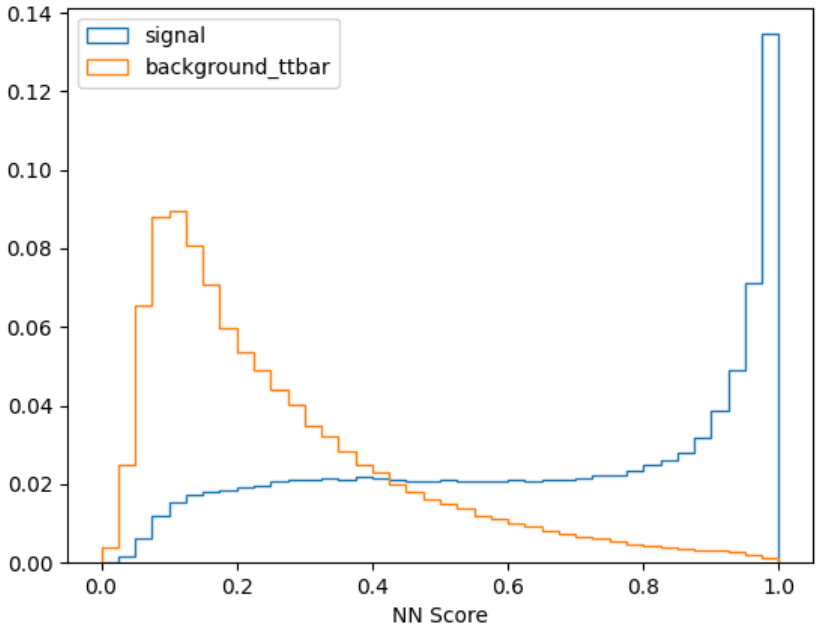

Figure 4.31: NN score distributions for nominal signal MC and $t\bar{t}$ background in the test sample. NN is trained using 11 features within the Geq2BiJ region.

Figure 4.32 shows the ROC curve of the NN model evaluated using the testing sample, the AUC on the testing sample is 0.85.

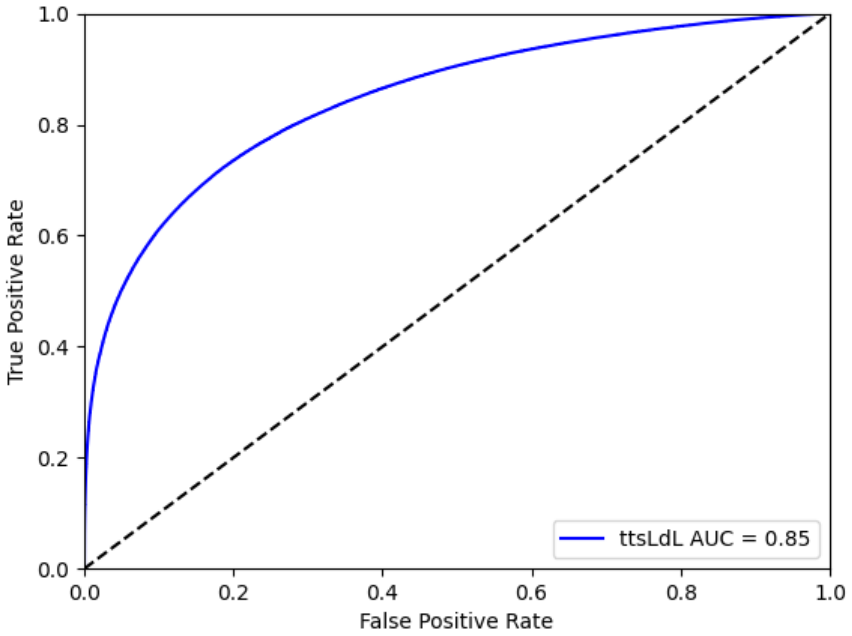

Figure 4.32: NN performance in terms of signal efficiency ("True Positive Rate") and background efficiency ("False Positive Rate") obtained on the test samples against all backgrounds considered in this study. As figure of merit, the Area Under the Curve (AUC) is also given. NN is trained using 11 features within the Geq2BiJ region.

Figure 4.33 shows the importance of the features, the most significant is H_T^{Centr}, while the least important is N_{jet}^{fwd}.

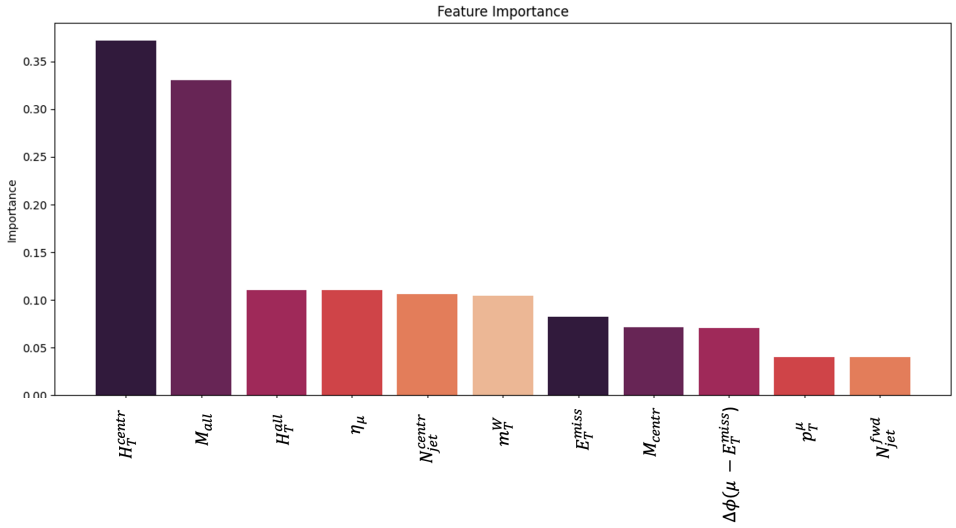

Figure 4.33: Feature importance plot for NN trained with 11 variables in Geq2BiJ region. The nominal signal MC sample is used to train the NN.

Comparison between 5-features NN approach and 11-features NN working point

As done in the case of 7-features NN approach, here the results of the 11-features NN approach in the Geq2BiJ region are compared with the 5-features NN ones in the 2B2J region, this test is done using the NNs built with the nominal signal MC, selecting in the 11-features NN approach a signal efficiency working point corresponding to the number of events in the 2B2J region after an NN cut at 70% signal efficiency (20010 events corresponding an efficiency of 0.7×28.4% with respect to the signal events in the Geq2BiJ). Results are shown in Table 4.6. One can notice that the 11-features NN approach largely improves the $S/t\bar{t}$ ratio with respect to the 2B2J cut (3.6 against 1.3 shown in Table 4.3), and an improvement is observed as expected also in the S/B ratio (0.95 against 0.6).

One of the advantage of employing the Geq2BiJ region is the possibility to enhance the number of signal events, while still keeping a favorable S/B ratio. Table 4.7 provides the signal efficiency and the related background rejection for different cuts of the NN score. It seems from the results that one might work at a signal efficiency between 25% and 40% keeping the number of signal

events between 25k and 40k with an extremely favorable $S/t\bar{t}$ ranging between 2.2 and 0.9, and still favorable S/B ranging between 0.8 and 0.5.

S.Eff(%)	$t\bar{t}$ **Rej.**(%)	S **evts**	$t\bar{t}$ **evts**	B evts	S/$t\bar{t}$	S/B
20	99.7	20010	5660	20960	3.6	0.95

Table 4.6: Signal efficiency cut, $t\bar{t}$ rejection, number of signal and number of $t\bar{t}$ related to this cut for NN trained with 11 variables in Geq2BiJ region. Results are shown for nominal signal MC samples.

S.Eff(%)	S **evts**	$t\bar{t}$ **Rej.**(%)	$t\bar{t}$ **evts**	$S/t\bar{t}$	B **evts**	S/B
25	25170	99.4	11330	2.2	31460	0.8
30	30200	99.1	16990	1.8	43920	0.7
35	35240	98.5	28310	1.2	60900	0.6
40	40270	97.7	43420	0.9	82830	0.5
45	45310	96.6	64180	0.7	110280	0.4
50	50340	95.1	92500	0.5	154640	0.3

Table 4.7: Signal efficiency versus $t\bar{t}$ rejection for different NN cuts. NN is trained with 11 variables in Geq2BiJ region. This is done with training the NN on the nominal signal MC.

4.3.4 Second signal simulated samples

As mentioned in Section 4.2, for the final cross-section measurement it is important to assess the systematics related to the NN cut that has been developed using a given signal MC sample. The systematics on the NN cut can be addressed using a second signal MC sample with different characteristics with respect to the initial one. This is done in this section using the other MC sample available (Sherpa 2.2.11). In this section, all the main studies performed for the nominal signal sample are repeated and presented using the second signal sample.

Figure 4.34 illustrates the distributions of p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, m_T^W, H_T^{center}, H_T^{all}, M_{centr}, M_{all}, N_{jet}^{fwd}, and N_{jet}^{centr}.

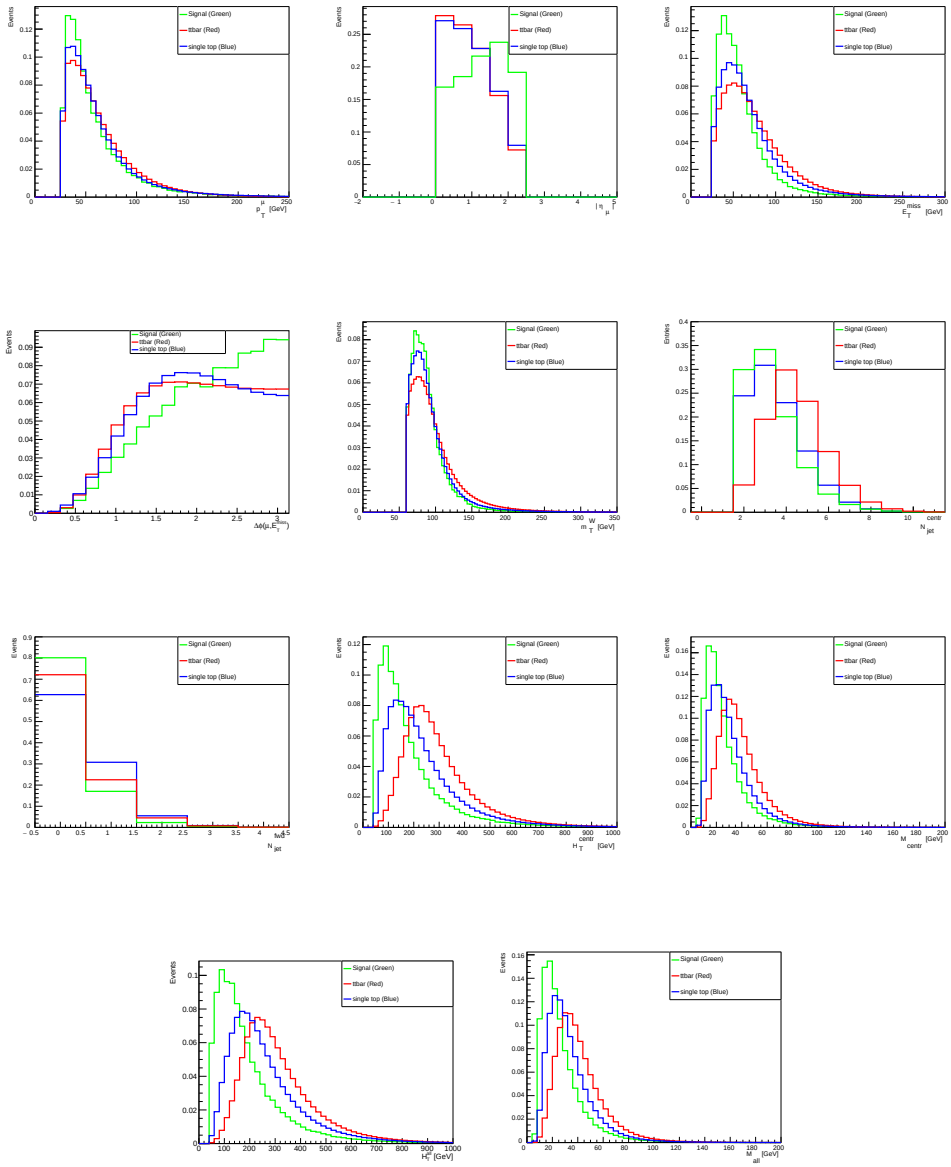

Figure 4.34: Distributions of: p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, m_T^W, N_{jet}^{centr}, N_{jet}^{fwd}, H_T^{centr}, M_{centr}, H_T^{all}, and M_{all} in the Geq2BiJ region for: second signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue)

Additionally, Figure 4.35 shows the distributions of p_T and Y of the leading and sub-leading b-jets, as well as the *di-b*jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$. Again some difference can be noticed between the shape of the di-bjets observables between the two signal MCs, related to their different characteristics.

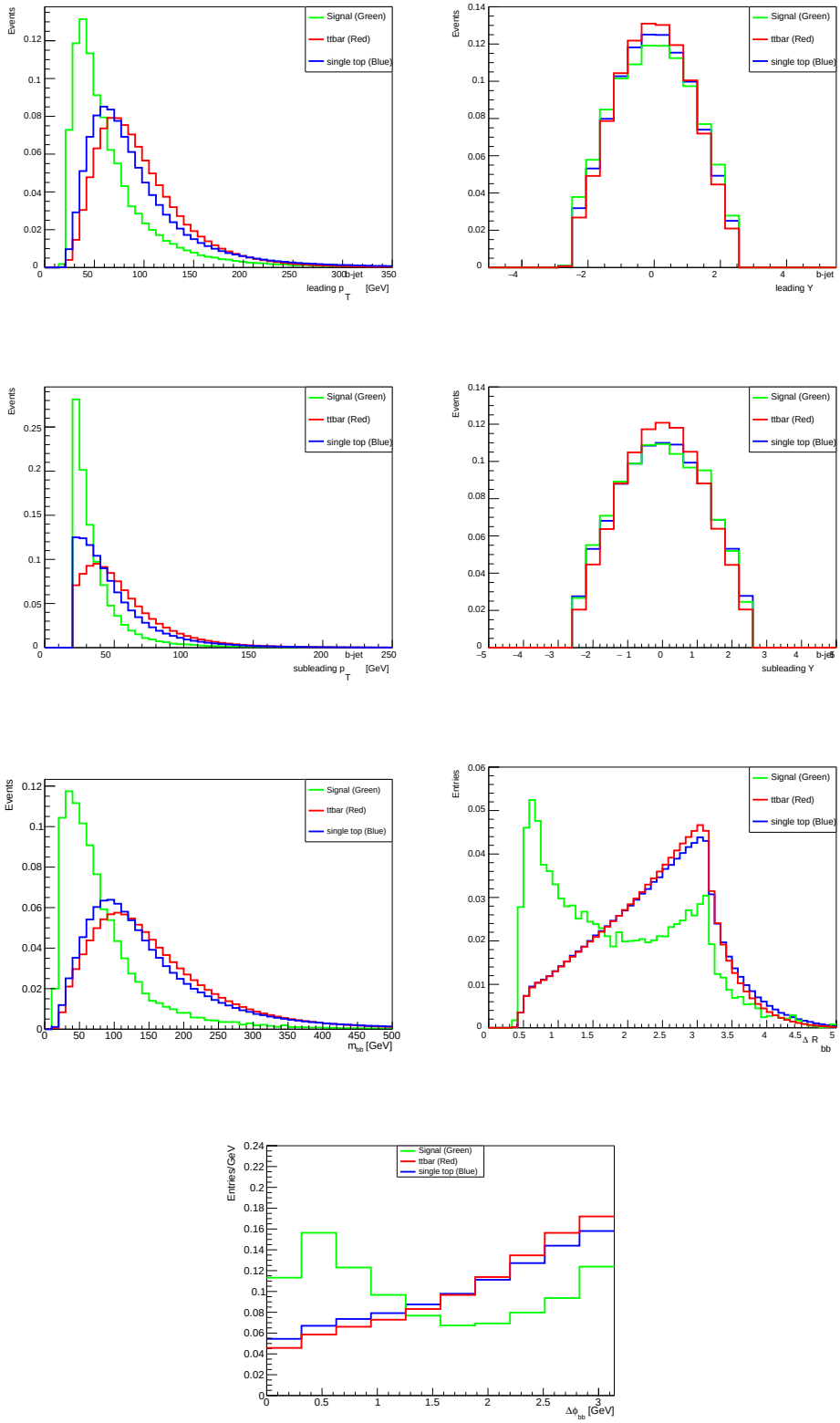

Figure 4.35: Distributions of: p_T and Y of the leading and sub-leading b-jets, as well as the *di*-$b$jet variables: m_{bb}, ΔR_{bb}, and $\Delta\phi_{bb}$ in the Geq2BiJ region for: second signal MC sample(Green), $t\bar{t}$ (Red), and single top (Blue).

Figure 4.36 shows the relationships between 11 different observables: p_T^{μ}, E_T^{miss}, m_T^W, $|\eta_\mu|$, $\Delta\phi(\mu, E_T^{miss})$, H_T^{center}, H_T^{all}, M_{centr}, M_{all}, N_{jet}^{fwd}, and N_{jet}^{centr} for signal dataset.

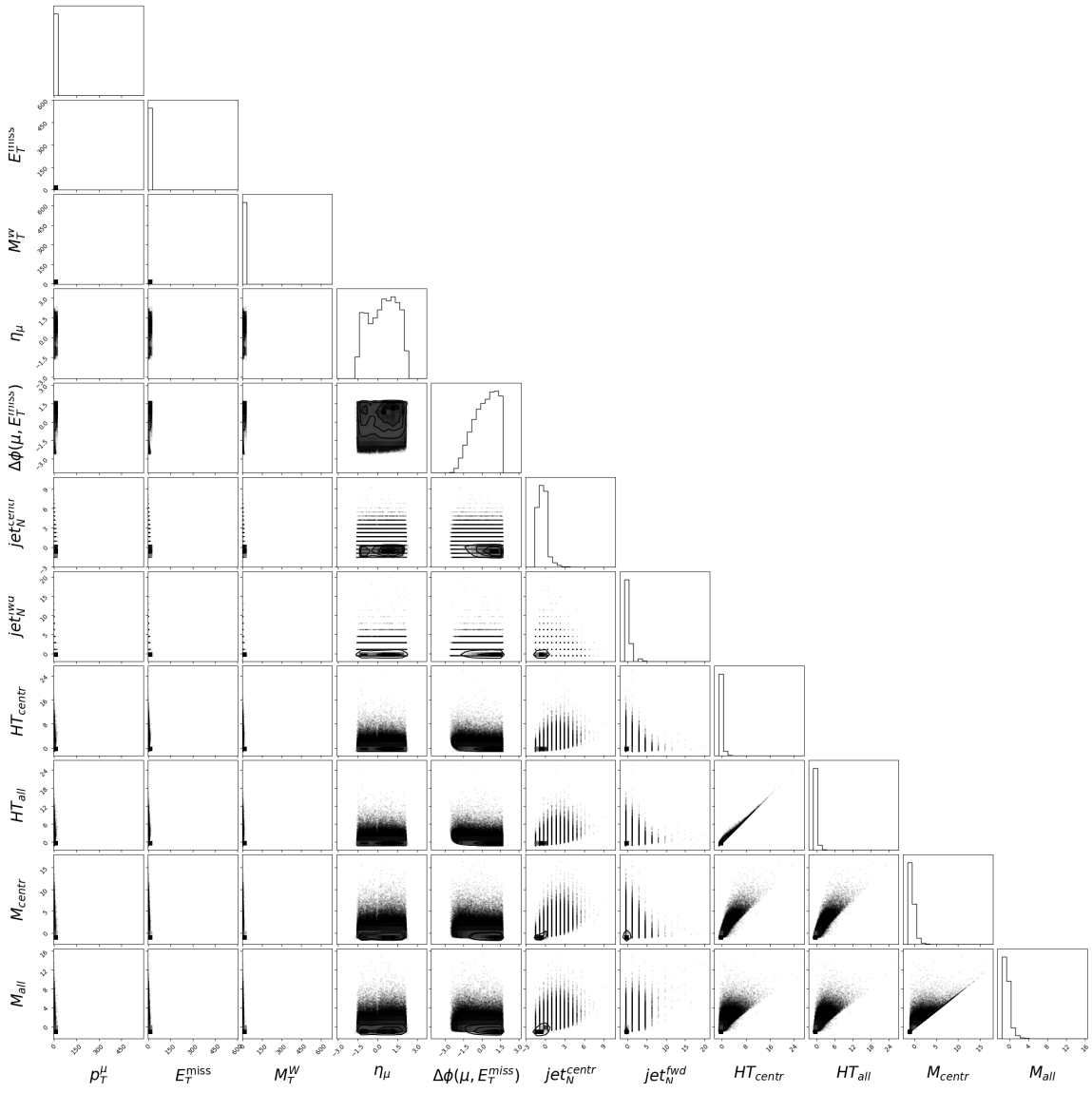

Figure 4.36: Scatter plots and distributions of the 11 features for the second signal MC sample.The distributions are shown after the scaling procedure described in section 4.1.

The optimation of the hyperparameters has been developed for the NN with 11 features and the chosen set after the optimisation study is as follows: 2 hidden layers, 20 nodes per layer, 250 epochs, and a learning rate of 10^{-3}. The

number of signal events used for the training is 0.5M while 15M $t\bar{t}$ events are used, as explained in Section 4.2 these events correspond to the total amount of MC events available in the Geq2BiJ region.

Figure 4.37 shows the loss as a function of the number of epochs, while Figure 4.38 shows the accuracy as a function of the number of epochs.

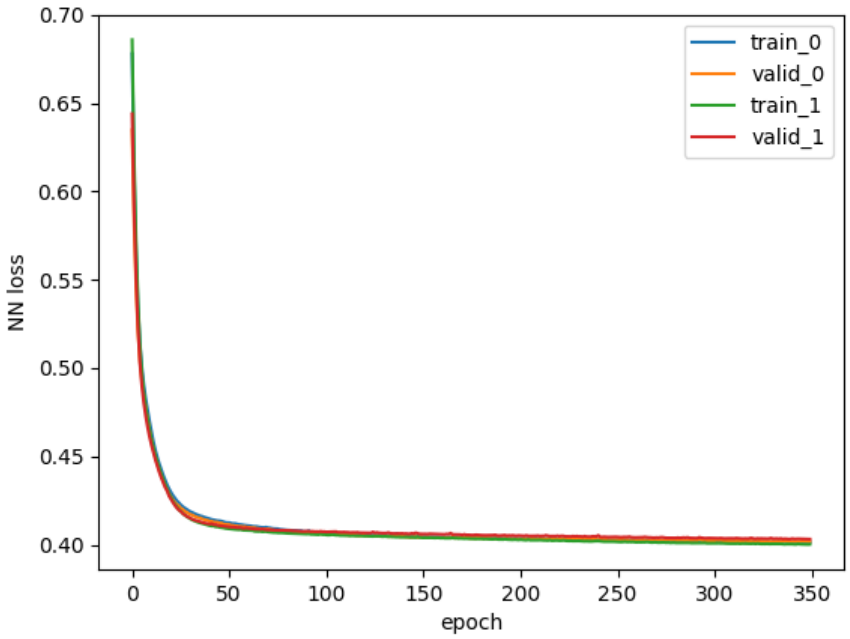

Figure 4.37: NN loss as a function of the training epochs in the training and validation samples. NN is trained using 11 features within the Geq2BiJ region.

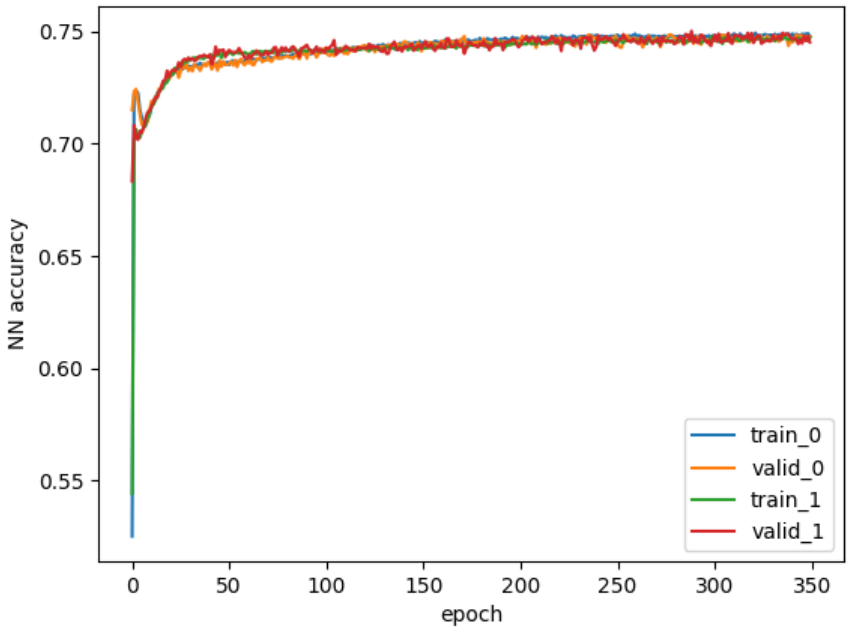

Figure 4.38: NN accuracy as a function of the training epochs in training and valid samples. NN is trained using 11 features within the Geq2BiJ region.

As before the plots are done for the two training and the two validation samples, and again accuracy and the loss reach stable plateau values at the number of chosen epochs and the values of these quantities evaluated on the validation samples are not diverging from those evaluated on the training samples. Therefore the model is robust and it is not overfitting to the training data. Figure 4.39 shows the distributions of the NN scores using the signal and $t\bar{t}$ test samples. Again good separation is observed. Figure 4.40 shows the ROC curve of the NN model evaluated using the testing sample, the AUC on the testing sample is 0.85.

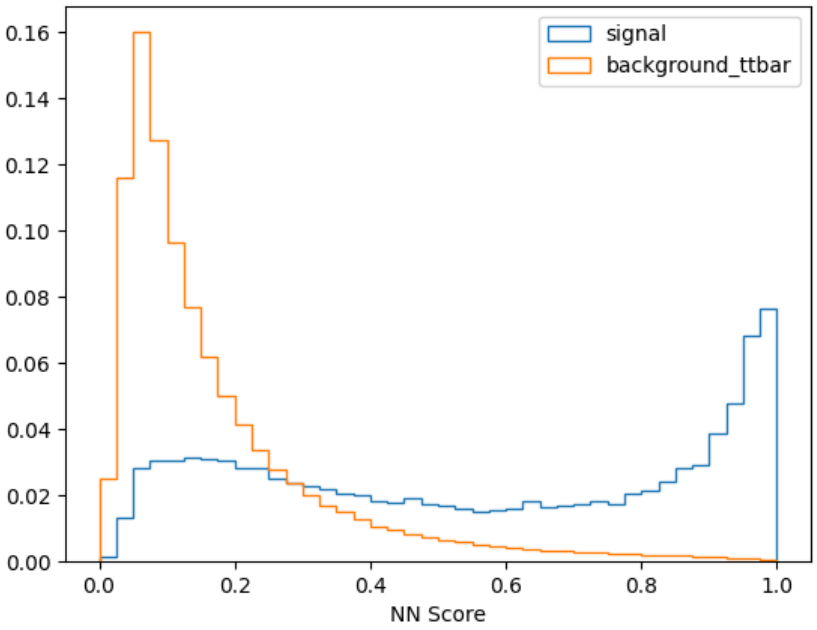

Figure 4.39: NN score distributions for signal MC and $t\bar{t}$ background in the test sample. NN is trained using 11 features within the Geq2BiJ region.

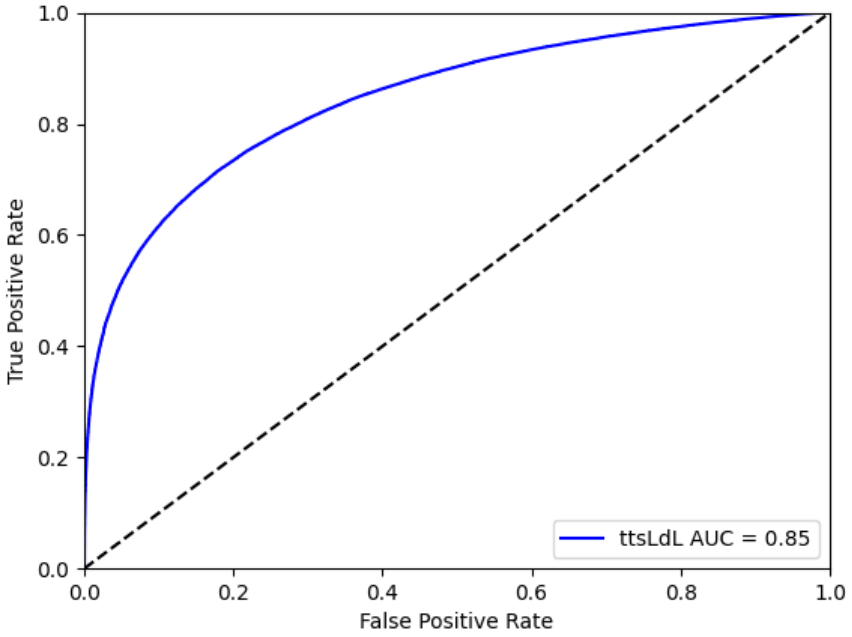

Figure 4.40: NN performance in terms of signal efficiency ("True Positive Rate") and background efficiency ("False Positive Rate") obtained on the test samples against all backgrounds considered in this study. As figure of merit, the Area Under the Curve (AUC) is also given. NN is trained using 11 features within the Geq2BiJ region.

Figure 4.41 shows the importance of the features, the most significant is H_T^{centr}, while the least important is N_{jet}^{fwd}. Finally Table 4.8 reports, for a given value of the NN, the signal efficiency and the corresponding background rejection. Comparing this table with results shown in Table **??** obtained with the NN developed using the nominal MC samples, one can see a very good agreement between the results of the two generators, suggesting again a small impact of the uncertainty on the signal MC modeling on the final measurement.

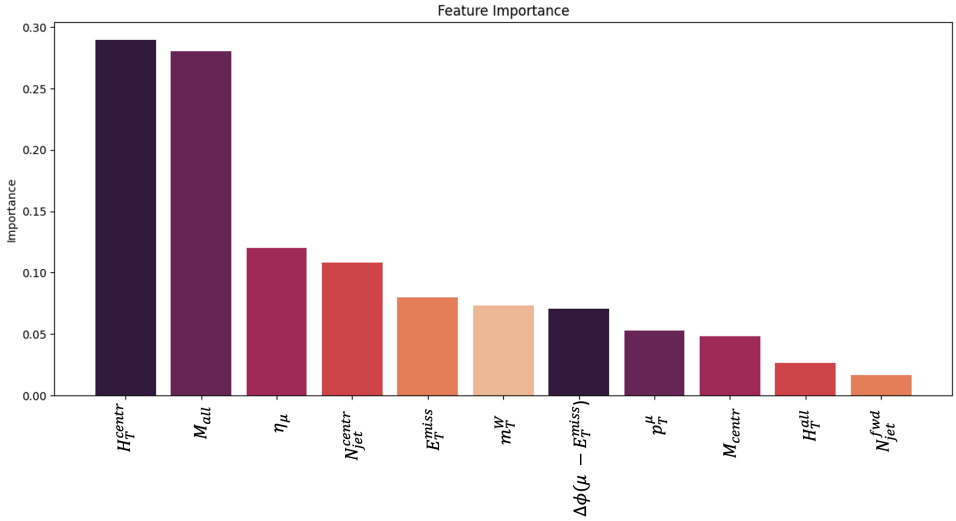

Figure 4.41: Feature importance plot for NN trained with 11 variables in Geq2BiJ region. The signal MC sample is used to train the NN.

Signal Efficiency(%)	$t\bar{t}$ **Rejection**
25	99.5
30	99.2
35	98.7
40	98.0

Table 4.8: Signal efficiency versus $t\bar{t}$ rejection for different NN cuts. NN is trained with 11 variables in Geq2BiJ region. Results are shown for second signal MC samples.

4.4 Detector level distributions after NN cut

In this section, I present the kinematic distributions of the b-jets observables as a function of which the differential cross-sections will be measured. The distributions are shown at detector level after applying the NN cut and they are shown both for the Geq2BiJ selection and for the 2B2J one. The value of the NN cut used is the one corresponding at a 70% signal efficiency for the 2B2J region and the one corresponding to a 30% signal efficiency for

the Geq2BiJ selection (each signal efficiency is computed with respect to the initial number of signal events in each of the two regions respectively). These cuts are just exemplary, they are in the optimal range that has been suggested in the previous sections on the basis of the $S/t\bar{t}$ and S/B ratios. The final NN cut indeed can be fixed only after emulating the full cross-section measurement with proper estimation of all backgrounds [1]and with the proper estimations of all systematic uncertainties related to the measurement. The full analysis has to be repeated for a few NN-working points and finally the one providing the lower systematic uncertainty is chosen.[2] The exemplary cuts chosen provide the same S/B ratio of 0.7 in both regions. The presentation of the results performed here is very similar to the one shown in Section 3.4. Therefore the distributions include in addition to the signal and $t\bar{t}$ background also the other backgrounds. Table 4.9 displays the expected number of events categorised by signal and all backgrounds for both selection criteria. These tables and distributions are normalised to represent the full Run-2 dataset corresponding to integrated luminosity of 140 fb^{-1}. Figure 4.42 shows the distributions of the p_T of the leading and subleading b-jets, while Figure 4.43 shows the distribu-

[1]In particular the W+light-jets and W+c-jets backgrounds have to be estimated starting from the MC results shown here but performing an adjustment using data looking at a flavour sensitive observable, in order to fix potential mis-modeling present in the MC; while detailed studies on the modeling of the top background in control region fully, dominated by this background, are performed.

[2]The formula of the cross-section as a function of an observable X in a given bin (i) is provided by: $\frac{\Delta\sigma}{\Delta X_i} = \frac{Data(i)-B(i)}{L\Delta X_i} \times U(i)$ where $Data(i)$ are the data in the bin (i), $B(i)$ is the total background in that bin, L is the integrated luminosity, ΔX_i is the size of the bin (i), $U(i)$ is the unfolding factor that corrects for inefficiency and small acceptance factors, as for example in this case the efficiency loss due to the NN cut. Therefore if the level of background is small even if the uncertainty on it is large, this doesn't impact much the uncertainty of the cross-section. The other important element is constituted by the uncertainty on U that includes all experimental uncertainty (i.e., uncertainty on b-tagging, jet energy scale, jet resolution, muon, missing transverse energy) and uncertainty on the signal modeling. If the final number of signal events predicted by MC is small, all these uncertainties have a statistical component that results in an overall increase of the uncertainty. This is why it is convenient to work with high signal efficiency and small background levels.

tions of the rapidities for the same jets. Figure 4.44 presents the distributions of m_{bb}, $\Delta\phi_{bb}$, and ΔR_{bb}. The figures are done using the nominal signal MC sample (MGaMC+Py8 FxFx).

As already anticipated in the previous section, after NN cut, the $t\bar{t}$ background is strongly reduced and the other backgrounds play a role. For the chosen working point, the $t\bar{t}$ is about 50% of the total background in the 2B2J region, while it is 40% of the total background in the Geq2BiJ region, and the single top is the second main background.

Process	**Generator**	**N. of evts in 2B2J**	**N. of evts in Geq2BiJ**
W($\rightarrow \mu\nu$) + $2b - jets$	MGaMC+Py8 FxFx	20010 ± 440	30200 ± 550
W($\rightarrow \mu\nu$) + $1b - jet$	MGaMC+Py8 FxFx	1420 ± 100	1900 ± 120
W($\rightarrow \mu\nu$) + $c - jets$	MGaMC+Py8 FxFx	3380 ± 230	5820 ± 290
W($\rightarrow \mu\nu$) + *light-jets*	MGaMC+Py8 FxFx	470 ± 130	1450 ± 190
$t\bar{t}$	Powheg+Py8	15810 ± 70	16990 ± 70
Single top	Powheg+Py8	7360 ± 20	9650 ± 20
Z +*jets*	Powheg+Py8	4140 ± 110	6290 ± 140
Diboson	Powheg+Py8	1100 ± 20	1390 ± 20
W ($\rightarrow \tau\nu$)+*jets*	MGaMC+Py8 FxFx	240 ± 40	450 ± 60
Total B		33920	43920
S/B		0.7	0.7

Table 4.9: Expected number of events for signal and background for each selection criterion. Results are shown for nominal signal MC samples after the NN cut.

Comparing the distributions of Geq2BiJ with 2B2J one might notice that the shape of the observable is different. In the case of the distributions of p_T of the leading and sub-leading b-jet (Figure 4.42) and in the case of the m_{bb} distributions for middle and high values this is due to the fact that the NN of the Geq2BiJ strongly reject $t\bar{t}$ ($S/t\bar{t}$ 1.8 in the Geq2BiJ region against 1.3

in the 2B2BJ) that is dominant in this region. One can also observe that the analysis at high m_{bb} looks still challenging even after the NN cut, while the region below 300 GeV (more relevant for Higgs analyses) looks feasible. In the case of the angular distributions the difference (also visible when comparing Figure 4.3 with Figure 4.13) is due to the fact that the two selections explore different topologies of events. The 2B2J explores a region containing 2 high p_T b-jets and no additional radiation, while the Geq2BiJ explores a region where in addition to the 2 b-jets, there are also additional radiation in the events, therefore the angular distributions change.

Overall after the NN cut, the signal contribution is well visible on top of the background in each bin, except, as already mentioned, in the bins at high m_{bb}.

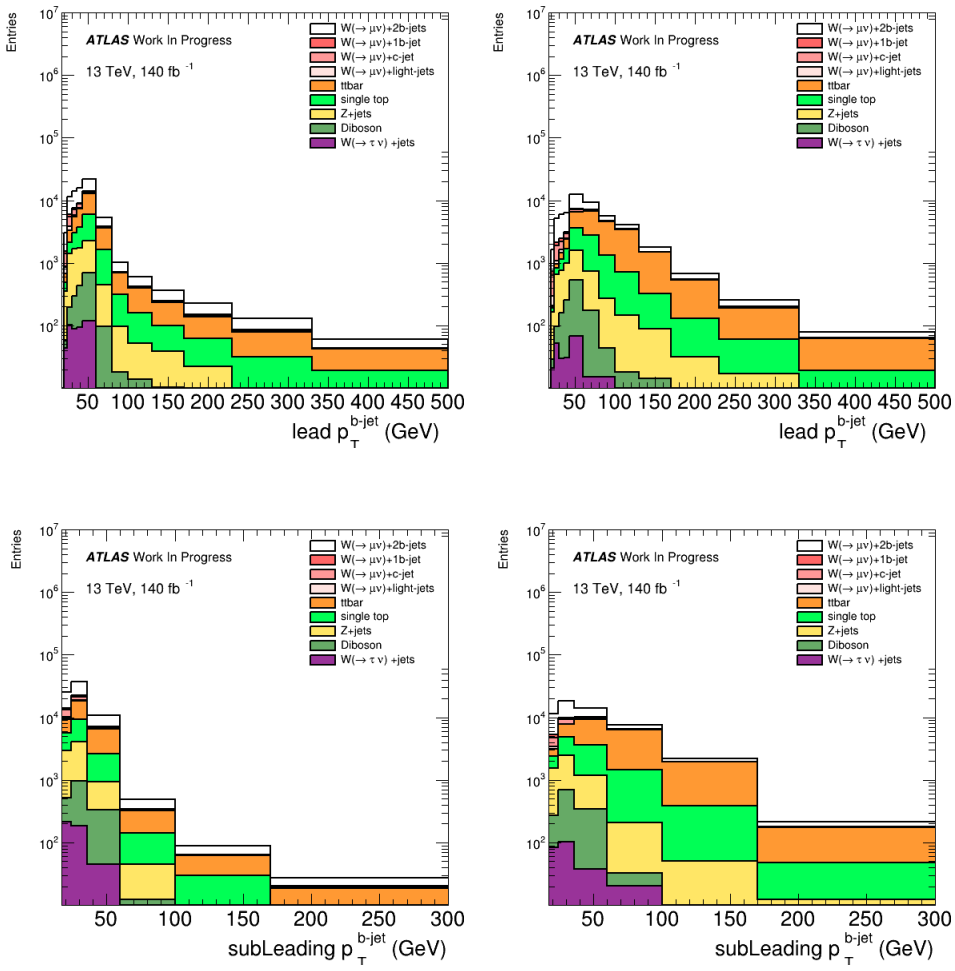

Figure 4.42: Distributions of p_T of leading b-jet (first row), p_T of subleading b-jet (second row) in events of Geq2BiJ (left) and 2B2J (right) regions.

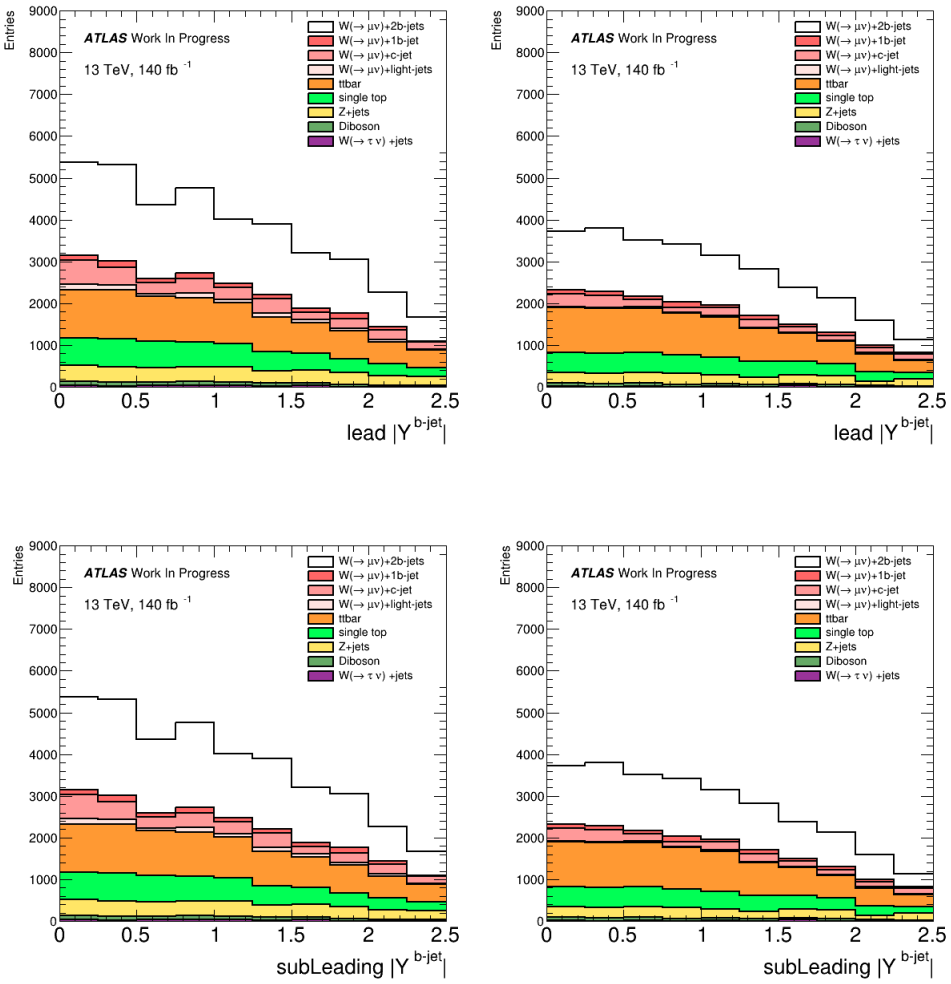

Figure 4.43: Distributions of Y of leading b-jet (first row), Y of subleading b-jet (second row) in events of Geq2BiJ (left) and 2B2J (right) regions.

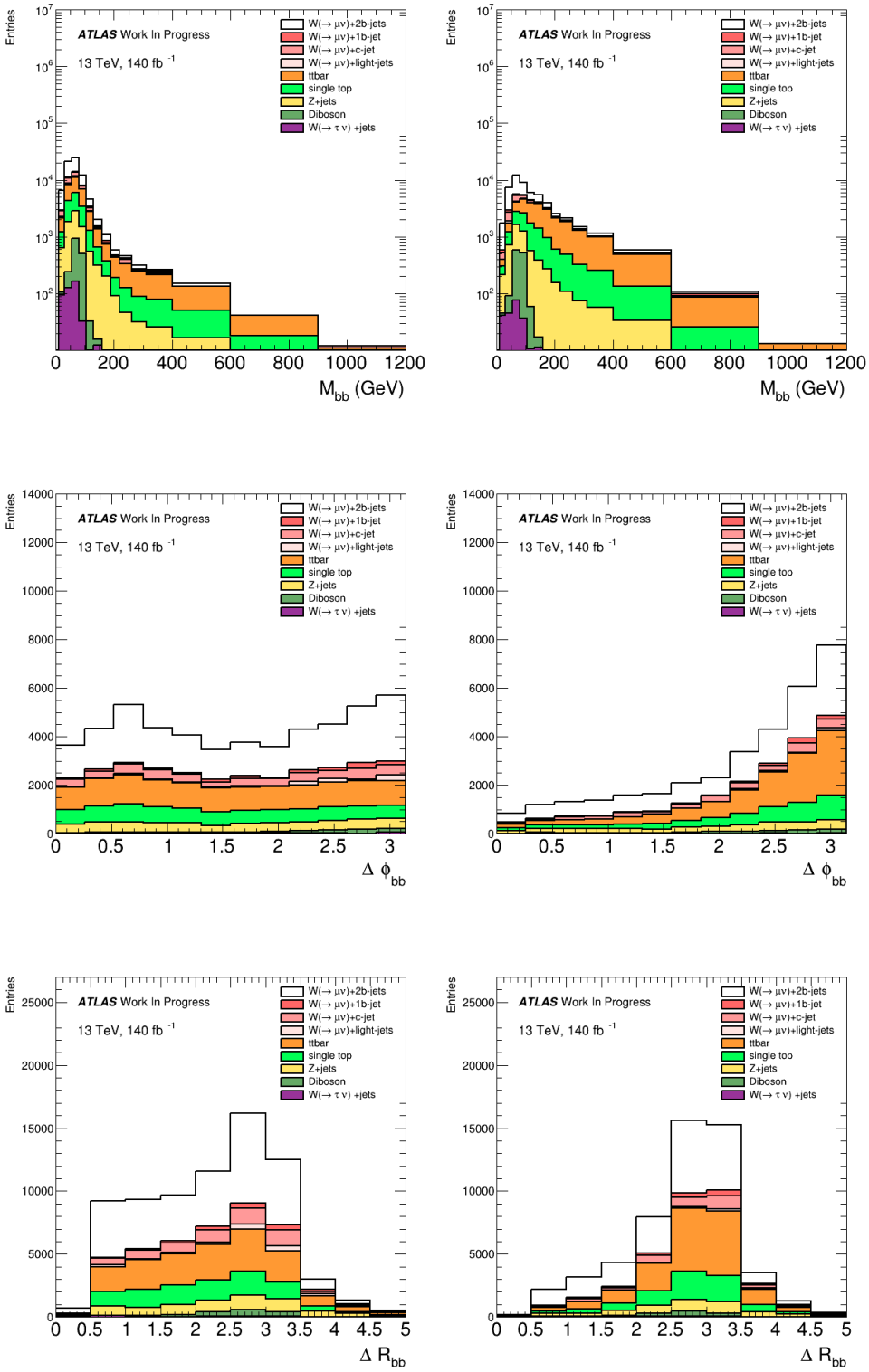

Figure 4.44: Distributions of m_{bb} (first row), $\Delta\phi_{bb}$ (second row), and ΔR_{bb} (third row) in events of Geq2BiJ (left) and 2B2J (right) regions.

Chapter 5

Conclusion

This study focuses on investigating the production of a W boson in association with b-quarks using proton-proton collisions data collected by the ATLAS experiment at the LHC. The specific process under examination involves the W boson decaying leptonically into a muon and a neutrino produced with two b-quarks, which are detected as b-jets in the experimental setup. The study is meant for the analysis of the data collected at a center-of-mass energy of $\sqrt{s}$ = 13 TeV from 2015 to 2018 (the full Run-2 dataset) corresponding to an integrated luminosity of 140 fb^{-1}. The signal process in mediated by strong interaction, therefore this study allows to enhance the understanding of pQCD, provides inputs for the further understanding of the proton's fundamental structure, and it provides critical insights into modeling of the W+2b-quark process, which is a significant background for Higgs boson studies, particularly in measuring the WH(bb) channel.

The measurement of the differential cross-sections in this final state has never been performed at the LHC so far, and the inclusive cross-section in a fiducial phase-space has been performed only with Run-1 data at $\sqrt{s}$=7 and 8 TeV, therefore at lower $\sqrt{s}$ and with a smaller dataset with respect to Run-2. The main challenge of the analysis is constituted by the huge background containing events originating from top-quark pair production, which are abundantly generated at the LHC. The top quark decays into a W boson and a $b-$quark, leading to various final states based on the decay modes of the W bosons.

The semi-leptonic decay mode, where one W boson decays into a lepton and a neutrino and the other into a pair of quarks, is particularly relevant. This mode, that has a branching ratio of 44%, results in a final state comprising a lepton, a neutrino, and four jets, including two b-jets. This characteristic final state is frequently encountered in LHC experiments, making it a significant component of the background in the analysis. The other decay modes, namely the leptonic and hadronic final states, also contribute to the background but to a lesser extent.

This study is based on simulated MC samples of the signal and the backgrounds officially produced by the ATLAS Collaboration. The simulated samples reproduce the experimental conditions of the full Run-2 dataset. They are processed using the full detector simulation and analyses with the same reconstruction tools used for real collision data, therefore they represent a fair simulation of real collision data. Two different MC samples have been used for the simulation of the signal. The nominal sample is obtained using the Madgraph generator, it generates the matrix element of the process up to 3 jets at NLO accuracy and for the higher multiplicities it uses the parton shower provided by Pythia (MGaMC+Py8 FxFx). The alternative sample is generated with Sherpa 2.2.11, in this case the the matrix element of the process is up to 2 jets at NLO accuracy and from 3 up to 5 jets at LO accuracy, for the higher multiplicities the Sherpa parton shower is used. Both MC samples are done within the 5 Flavour Number Scheme, therefore allowing the b-quark to be present in the initial state.

Events containing a W-boson candidate were selected by requiring exactly one muon with $p_T > 27$ GeV and $|\eta| < 2.5$, no additional isolated muons or electrons with $p_T > 7$ GeV and $|\eta| < 2.5$, and missing transverse energy $E_T^{miss} > 25$ GeV. The transverse mass (m_T^W) of the W boson had to be greater than 60 GeV. After the selection of the W-boson candidates, two different strategies have been developed to select events containing 2 b-jets, which result in the definition of two different regions: the Geq2BiJ and the 2B2J regions.

The Geq2BiJ strategy involves selecting events with at least two b-jets allowing

other jets to be present in the event. This inclusive approach is similar to the one used in the ATLAS analysis for Z+2b-jets. The results show that after the Geq2BiJ selection, the ratio of signal to $t\bar{t}$ background is approximately 5%, with $t\bar{t}$ events being the dominant background, comprising about 90% of the total background. This strategy provides a good starting point but requires further selection to improve the signal-to-background ratio (S/B), which can be achieved using a neural network classifier.

The 2B2J strategy aims to reduce the significant background from semi-leptonic $t\bar{t}$ decays, which typically produce more jets than the W+2b-jets signal. This strategy involves selecting events with exactly two b-jets and applying a jet veto. The central jet veto excludes events with additional jets ($p_T > 20$ GeV, $|y| < 2.5$), and the forward jet veto excludes events with jets in the forward region ($p_T > 30$ GeV, $2.5 < |y| < 4.5$). The 2B2J strategy results in a higher ratio of signal to $t\bar{t}$ background, at level of about 40%, but at the cost of reducing the signal to 25% of the selected events with the Geq2BiJ strategy. This approach was initially developed for a partial Run-2 dataset and highlighted the need for further background suppression using machine learning techniques to achieve a precise differential cross-section measurement.

By employing neural network classifiers, this analysis identifies a strategy of the event selection at detector level that is the first step of the first differential cross-section measurements of W+2b-jets at the LHC.

A NN model has been developed for each of the two regions, the inclusive and exclusive one. In both cases a comprehensive study was conducted to identify the most discriminative kinematic variables using events passing the respective selections.

To optimize the NN models in both regions, a comprehensive study was conducted to identify the most discriminative kinematic variables using events passing respectively the 2B2J and the Geq2BiJ selections. In the case of the 2B2J region, the chosen input variables include: p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, and m_T^W. These variables were selected based on their ability to distinguish signal events from background events, primarily the $t\bar{t}$ background, which is

the most copious. The developed NN model demonstrates good separation between signal and background, with an AUC of 0.83 on the test and validation samples. This study indicates that an NN cut selecting signal events at an efficiency between 80% and 70% would select about 20-23k signal events that should be enough for a first differential cross-section measurement. This working-point range would guarantee a $t\bar{t}$ rejection at level of 70-80%, resulting in a $S/t\bar{t}$ ratio around 1.1-1.3, while the overall S/B would be at the level of 0.7. Indeed after the NN cut, other backgrounds (initially completely masked by the $t\bar{t}$) play a role. After the NN, the $t\bar{t}$ is still the major background, but it is a level of about the 50% of the total background, while the single top represents the second main background, accounting for about the 25% of the total.

The development and optimization of a NN model for the inclusive selection, Geq2BiJ, has been done in a similar way. Also in this case a thorough investigation has been done to identify kinematic variables that effectively separate signal from background events. A total of 11 variables has been selected for the development of the NN model, comprising the 5 observables previously used for 2B2J and new observables. The complete set is constituted by: p_T^{μ}, $|\eta_{\mu}|$, E_T^{miss}, $\Delta\phi(\mu, E_T^{miss})$, m_T^W, N_{jet}^{centr}, N_{jet}^{fwd}, H_T^{centr}, M_{centr}, H_T^{all}, and M_{all}. A first model has been developed using only 7 of these observables, namely the original 5 features previously used for 2B2J model plus the number of central jet ($p_T >$ 20 GeV and $|y| < 2.5$) and the number of forward jet ($p_T > 30$ GeV and $2.5 < |y| < 4.5$). This initial model did not improve the performance with respect to the approach developed for the 2B2J region. A second high-performing NN model has been instead obtained using all the 11 variables. In this case indeed the NN model is able to obtain a very good separation between signal and background, with an AUC of 0.85 on the test and validation samples. The study indicates that an NN cut selecting signal events at an efficiency between between 20% and 40%, corresponding of an amount of signal events ranging between 20k and 40k signal events, a $t\bar{t}$ rejection higher than 97% is obtained, with a very good $S/t\bar{t}$ ratio, ranging between 3.6 and 1.3, while S/B ratio

ranges between 0.95 and 0.50. As already observed in the 2B2J case, the large suppression of $t\bar{t}$ background after the NN cut, makes the role of other backgrounds more relevant. In the working-point range just discussed the level of $t\bar{t}$ varies between the 30% and the 50% of the total background. Again as for the 2B2J the single top is the second main background, and it is at the level of about 60% of $t\bar{t}$ at the signal efficiency working point of 30%.

In the final cross-section analysis a systematic uncertainty due to the usage of the NN model in the event selection chain has to be addressed. The primary source of uncertainty comes from the modeling of the signal MC sample used to train the NN, this can be estimated employing a different generator with different characteristics with respect to the nominal one. Therefore in this thesis work the development of the 5-features NN in the 2B2J region and the 11-features NN in the Geq2BiJ has been repeated using the alternative signal sample, Sherpa 2.2.11. The two NNs developed with the second signal sample looks very similar to the nominal one in terms of performances: same AUC, similar $t\bar{t}$ rejection at the same signal efficiency, and same ranking of feature importance. This implies that the systematics related to the NN cut on the final cross-section analysis will be small. The second source of uncertainty comes from the modeling of the $t\bar{t}$ sample, this can be studied selecting in data a region fully dominated by $t\bar{t}$ (with contributions from other processes almost null) and comparing the NN distribution in data with the one provided by the $t\bar{t}$ generator. [1]

This thesis work also proposes the first cross-section differential distributions that would nice to measure as a function of few important observables. The proposed observables are: the transverse momentum and the rapidity of the leading and sub-leading jet (p_T^{b-jet} and Y^{b-jet}), the di-b-jet invariant mass (m_{bb}), the angular separation of the 2 b-jets in the transverse plane ($\Delta\phi_{bb}$) and the ΔR_{bb}. A study of the shapes of these distributions has been done leading

[1] This step involving a different event selection and usage of real data, that has to be preliminary slimmed, thinned and skimmed which goes beyond the scope of this master's thesis.

to a few interesting points:

• The shape of these variables is different between the signal and the $t\bar{t}$ samples, but they haven't been used as NN features in order to avoid huge bias in their shapes after the NN cut. The observables as a function of which the differential cross-section are measured are typically not used as features of the NN.

• While the modeling of the observables used to train the NN is very similar between the 2 tested signal samples, some differences are observed employing the b-jets observables, particularly the di-b-jets angular variables, this highlights the importance of a future cross-section measurement, where data will tell what is the best modeling approach.

• Finally comparing the shapes of the same b-jets observables in the two explored regions, the 2B2J and the Geq2BiJ, differences appears, mainly in the di-b-jets angular observables. This is due to the fact that the two regions explore different topologies: in the exclusive region there are just 2 high p_T b-jets, while in the inclusive region the presence of additional jets changes the correlation between angular observables.

Between the two strategies proposed for the suppression of the $t\bar{t}$ background (the exclusive 2B2J selection and the inclusive Geq2BiJ one), the inclusive one is preferable, despite both present good final results. In the inclusive region a larger number of signal events can be selected allowing a larger range of possibilities for the final NN cut to be applied [2]. Moreover, since the fiducial phase-space of the final cross-section measurement will be done in the Geq2BiJ region and since the shape of b-jets observables in the the 2B2J and the Geq2BiJ show some differences, the usage also at detector level of the same phase-space will avoid shape corrections based on MC in the unfolding procedure.

To conclude this thesis is the first documented study of the event selection

[2]For the final analysis one has to repeat the full cross-section measurement with detailed estimation of the all backgrounds and all systematics effects with different NN working points and identify the one that guarantees the smaller total uncertainty.

of the W+2 b-jets final state using machine learning technique for the bachground suppression. It represents the basis of the first measurement of the cross-section production of a W boson in association with 2 b-jets using the full Run-2 dataset at 13 TeV.

Appendix A

Machine learning: Neural Networks

In this appendix after a general introduction on machine learning methods in section A.1, a detailed dissertation on neural networks is presented, in particular; in section A.2 generalities on the neural networks are presented, including discussions on the main hyper-parameters; in section A.3 details on the techniques used for the manipulation of the inputs are provided and finally in section A.4 the main methods used to address the performances of the method are presented.

A.1 Machine learning

Machine learning is a field of artificial intelligence (AI) that focuses on the development of algorithms and models that learn from and make predictions or decisions based on data. In essence, it involves the creation of systems that can automatically learn and improve from experience.
The primary goal of machine learning is to develop algorithms that can identify patterns, extract meaningful insights, and make predictions or decisions without explicit human intervention. This is achieved by training models on large datasets, allowing the algorithm to learn and generalize patterns, which can then be applied to new, unseen data.

Machine learning in physics involves the application of computational techniques to analyze and model physical phenomena. It helps in performing tasks like data classification, regression, and feature recognition by extracting patterns, relationships, and predictions from huge datasets using algorithms. In the field of physics, machine learning finds applications in diverse areas, including particle physics, astrophysics, materials science, and quantum mechanics. It enhances data analysis, accelerates simulations, and contributes to the discovery of novel physical insights. The combination of machine learning and physics not only helps to make sense of complicated data but also enables exploration of areas that were previously out of reach, while also refining how experiments are designed and conducted.

Machine learning is classified into three basic types based on how algorithms learn to improve predictions: supervised learning, unsupervised learning, and reinforcement learning (Figure A.1). The choice of algorithm depends on the nature of the data, and many algorithms can be adapted to multiple types, depending on the problem and dataset.

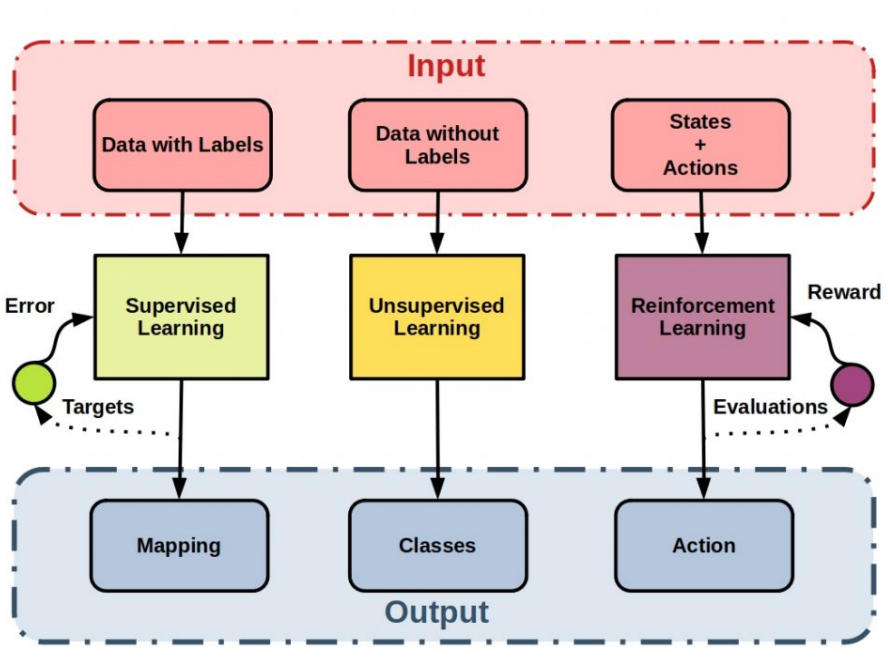

Figure A.1: The three main categories of machine learning.

A.1.1 Supervised learning

Supervised learning is the first type of machine learning, in which labelled data are used to train the algorithms. Therefore in supervised learning, algorithms are trained using marked data, where the input and the output are known. A set of inputs, called features, for each labeled dataset are provide to the learning algorithm , the algorithm has to find the relationships between the features and the label (meant as output) and it learns it by comparing its actual output with the correct output, it leans from its errors and modifies the model accordingly. In this way after the training it is able to make predictions of the response values for new datasets. The raw data are divided into two parts. The first part is for training the algorithm, and the other part is used for test the trained algorithm.
Supervised learning uses the data patterns to predict the values of additional data for the labels. This method is commonly used in applications where historical data predict likely upcoming events. It can anticipate when transactions are likely to be fraudulent or which insurance customer is expected to file a claim. Supervised learning in machine learning comprises two main types: Regression and classification [71].

Regression involves using labeled data to make predictions in a continuous form. It is a predictive modeling technique investigating the relationship between dependent and independent variables (Figure A.2). Examples include Simple Linear Regression, Multiple Linear Regression, Polynomial Regression and etc.

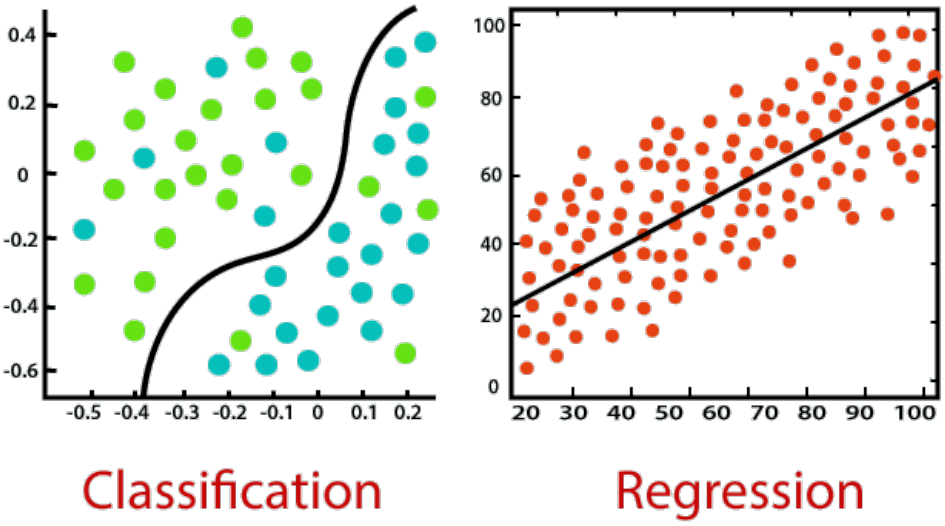

Figure A.2: Comparing Classification and Regression in data analysis. The left graph illustrates classification with a non-linear boundary separating data points. The right graph represents regression with an orange data trend line

Classification uses labeled data to make predictions in a non-continuous form. The algorithm learns from input data and classifies new observations (Figure A.2). Examples include Logistic Regression, K-Nearest Neighbours, Support Vector Machines, and etc.
In High Energy Physics (HEP), machine learning applications predominantly rely on supervised learning techniques.

A.1.2 Unsupervised Learning

Unsupervised Learning is the second type of machine learning, in which unlabeled data are used to train the algorithm, which means it used against data that has no historical labels. What is being showing must figure out by the algorithm. The purpose is to explore the data and find some structure within. In unsupervised learning, the data is unlabeled and provided directly into the algorithm without preprocessing or without knowing the output, which makes it unsuitable for dividing into train and test sets. The algorithm figures out the data and according to the data segments, it makes clusters of data with new labels. As mentioned above, the unsupervised learning mainly uses

clustering algorithms (such as K-Means Clustering or Hierarchical Clustering) for the grouping of similar entities.

A.1.3 Reinforcement Learning

In contrast to supervised and unsupervised learning, reinforcement learning (RL) focuses on maximizing rewards rather than finding patterns or predicting outcomes. RL assumes data is interdependent, and learns through interaction with the environment. It balances exploring new actions and exploiting known ones, often mimicking real-world biological learning via positive reinforcement. In reinforcement learning, an autonomous agent learns to perform a task by trial and error in the absence of any guidance from a human user like Robots.

A.2 Artificial Neural networks

Artificial Neural Networks (ANN) are an information processing paradigm that was originally inspired by the way the biological nervous system such as brain process information. It is composed of large number of highly interconnected processing elements (neurons or nodes) organised in layers working in unison to solve a specific problem [72]. A node is just a place where computation happens, loosely patterned on a neuron in the human brain, which fires when it encounters sufficient stimuli. A node combines input from the data with a set of coefficients, or weights, that either amplify or dampen that input, thereby assigning significance to inputs with regard to the task the algorithm is trying to learn, (e.g., which input is most helpful is classifying data without error?) These input-weight products are summed and then the sum is passed through a node's so-called activation function, to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome, that allows the classification. Figure A.3 represents the general schematisation of a NN.

There are various types of neural networks. I will explain feedforward neu-

ral networks, which I used in my thesis. A feedforward neural network (or fully connected neural network) is one of the earliest neural network models invented in the field of artificial intelligence [73]. They can learn and model complex, non-linear relationships in the data, which is essential for accurate classification. Feedforward consists of three main layers:

Input Layer:

The input layer is where data is introduced into the neural network. It consists of nodes (or neurons), each node of the fist layer representing a feature (input variable of a specific data array from the dataset). For instance, in image recognition, each node might correspond to a pixel's intensity value of a given image. The input layer's purpose is to pass this information forward through the network for processing.

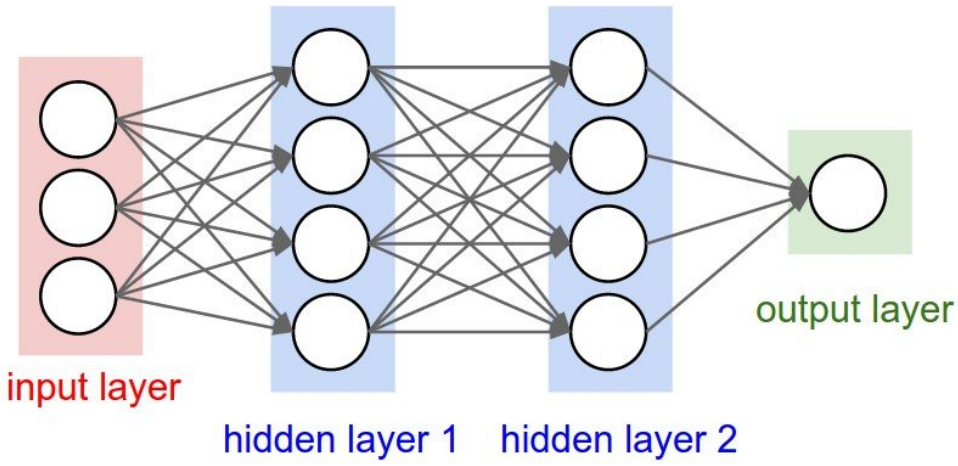

Figure A.3: 3-layer neural network with three inputs, two hidden layers consisting of four neurons each and one output layer

Hidden Layers:

Hidden layers are the core computational units of a neural network. They exist between the input and output layers and are composed of multiple neurons that are indeed arranged in a sequence of layers. Each neuron in a hidden layer receives inputs from all neurons in the previous layer and performs a

transformation that is the output of the neuron applying weights (learned parameters) to the inputs and passing them through an activation function. This process allows the network This process allows the network to learn non-linear relationships between the input and output data.
The architecture and training of hidden layers significantly influence the network's ability to learn and generalize from data. Below, we delve into key aspects of hidden layers:

Output Layer:

The output layer is where the neural network produces its final predictions or outputs based on the processed information from the hidden layers. The number of nodes in the output layer depends on the nature of the problem being solved. For instance, in a binary classification task (e.g., signal and background), the output layer might have just one node representing the probability of one class (e.g., signal) and another node representing the probability of the other class or two nodes one representing the probability of each class (e.g. one node for the signal and one node for the background). In a regression task (e.g., predicting house prices), the output layer might have a single node outputting the predicted value.
The output layer typically applies an activation function tailored to the specific problem, such as softmax for multi-class classification or sigmoid for binary classification. These functions ensure that the outputs are meaningful and fall within a desired range (e.g., probabilities between 0 and 1). This is explained in detail in the subsection A.2.3 dedicated to the activation functions. In the following, the most important elements of NNs are explained.

A.2.1 Weights

They are numerical values associated with the connections between neurons (or nodes) across different layers of the network. Each connection from one neuron to another has an associated weight that signifies the strength and direction (positive or negative) of the influence one neuron has on another

(Figure A.4). When an input signal passes through the network, it gets multiplied by these weights, which cumulatively determine the final output of the network. There is also and additional weight called "Bias" which serve as offsets or thresholds, allowing neurons to activate even when the weighted sum of inputs is insufficient.

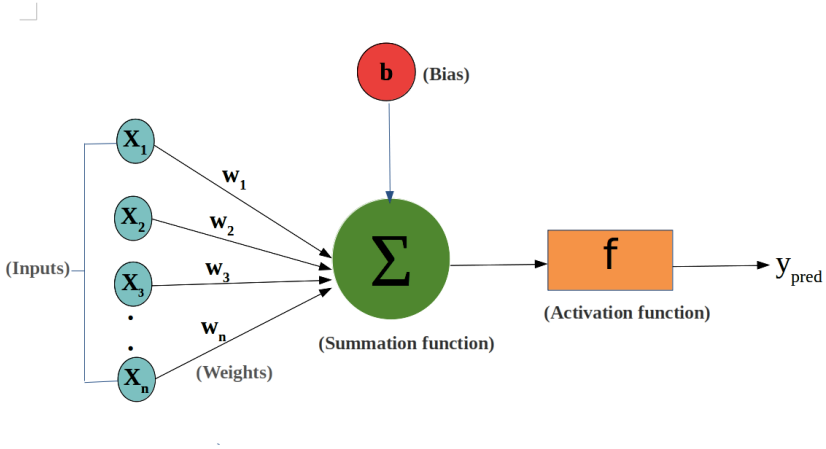

Figure A.4: Representation of a simple neural network with one node.

In the context of artificial neural networks(ANNs), weights are a fundamental component that play a crucial role in the network's ability to learn and make predictions. During the training phase, the ANN learns by iteratively adjusting these weights to predict the correct output for a given set of inputs. The set of weights in the network encapsulates what the network has learned from the training data, effectively encoding the knowledge necessary for making predictions or decisions.

Initialization of Weights: Before training begins, weights in an ANN must be initialized to some starting values. Proper weight initialization is vital for the convergence and performance of the network. Common initialization methods include random initialization, where weights are assigned small random values,

Adjusting Weights with Learning Algorithms: The process of learning in an ANN involves adjusting the weights to reduce the error between the predicted output and the actual output. This is typically done using a learning

algorithm like backpropagation combined with an optimization technique such as gradient descent. Backpropagation is a training algorithm for neural networks that minimizes the error by adjusting the weights of the network which is explained in Section A.2.6. During backpropagation, the error is calculated at the output and propagated back through the network to update the weights in a way that minimally reduces the overall error.

Learning Rate and Weight Updates: The learning rate is a hyperparameter that determines the size of the steps taken during the weight update process. A higher learning rate may lead to faster convergence but risks overshooting the minimum error, while a lower learning rate ensures more stable convergence but may be slow and get stuck in local minima. The learning rate, therefore, plays a critical role in determining how significantly the weights are adjusted during each iteration of the training process.

Overfitting and Regularization of Weights: Overfitting is a common issue where a neural network learns the training data too well, including the noise and outliers, which results in poor performance on new, unseen data. To prevent overfitting, where the ANN performs well on the training data but poorly on unseen data, regularization techniques can be applied to the weights. Regularization methods like L1 and L2 add a penalty term to the loss function based on the magnitude of the weights. This encourages the network to maintain smaller weights, leading to simpler models that generalize better to new data.

A.2.2 Activation functions

An activation function is a function that is added into an artificial neural network in order to help the network learn complex patterns in the data. When comparing with a neuron-based model that is in our brains, the activation function is at the end deciding what is to be fired to the next neuron. That is exactly what an activation function does in an ANN as well. It takes in the output signal from the previous cell and converts it into some form that can be taken as input to the next cell. It maps the resulting values in between 0 to

1, or -1 to 1, or other ranges (depending upon the functions). The activation functions can be basically divided into two types: Linear Activation Function and Non-linear Activation Functions:

Linear or Identity Activation Function: As you can see in Figure A.5 the function is linear. Therefore, the output of the functions will not be confined between any range (it is between ∞ to $+\infty$) and the equation is simply f(x)= x.

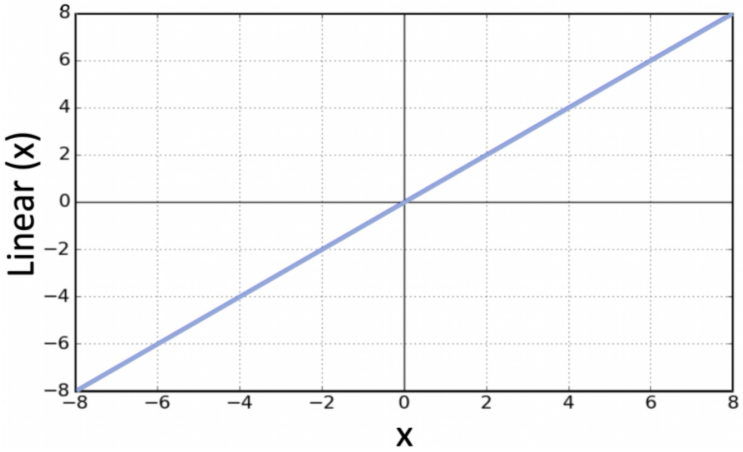

Figure A.5: Linear activation function

It doesn't help with the complexity or various parameters of usual data that is fed to the neural networks.

Non-linear Activation Function: The Nonlinear Activation Functions are the most used activation functions. Non-linearity helps to makes the graph look something like what you can see in Figure A.6.

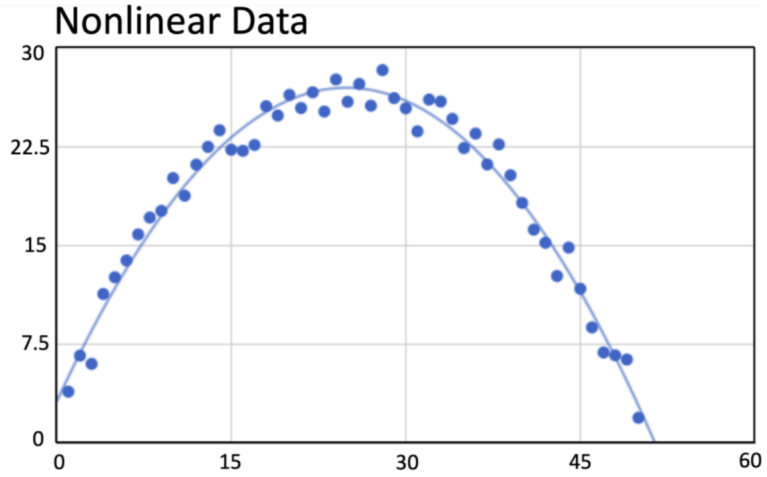

Figure A.6: Non-linear activation function

It makes it easy for the model to generalize or adapt with variety of data and to differentiate between the output. In the following the most common activation functions used in ANN are described.

Sigmoid or Logistic Activation Function: The Sigmoid Function curve looks like a S-shape in the Figure A.7. The main reason why sigmoid function is used is because it exists between 0 to 1. Therefore, it is especially used for models where prediction of the probability as an output is favored. Since probability of anything exists only between the range of 0 and 1, sigmoid is the right choice. The function is differentiable. That means, the slope of the sigmoid curve at any two points is easy to find. The function is monotonic, but function's derivative is not.

The **softmax function** is a more generalized logistic activation function which is used for multiclass classification, and it is computed using the following relationship:

$$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{A.1}$$

Where (z_i) represents the values from the neurons of the output layer.

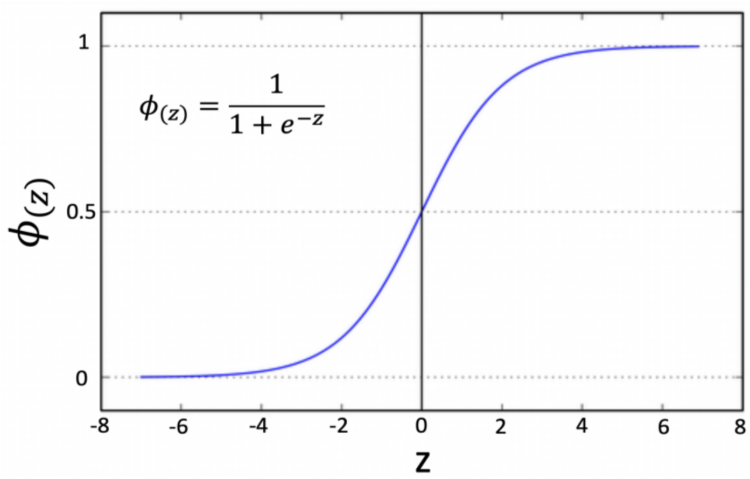

Figure A.7: Sigmoid activation function

The exponential acts as the non-linear function. Later, these values are divided by the sum of exponential values in order to normalize and then convert them into probabilities.

Tanh or hyperbolic tangent Activation Function: Tanh is similar to sigmoid but better. The range of the tanh function is from -1 to 1. Tanh is also sigmoidal (S-shaped) and you can see it in Figure A.8.

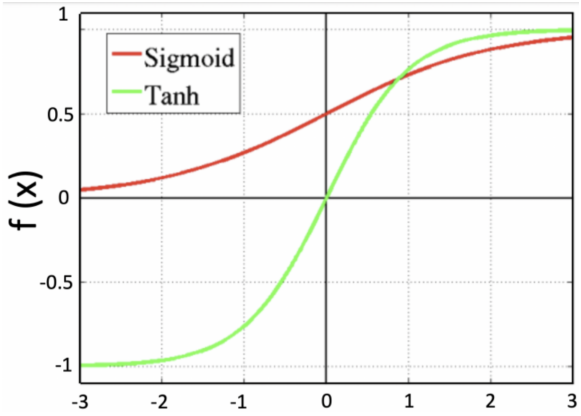

Figure A.8: Tanh vs. logistic sigmoid

The advantage is that the negative inputs will be mapped strongly negative, and the zero inputs will be mapped near zero in the tanh graph. The function is differentiable. The function is monotonic while its derivative is not monotonic. The tanh function is mainly used for classification between two

classes. Both tanh and logistic sigmoid activation functions are used in feed-forward NN.

ReLU (Rectified Linear Unit) Activation Function: The ReLU is one of the most used activation function. As you can see in the Figure A.9, the ReLU is half rectified (from bottom). f(z) is zero when z is less than zero and f(z) is equal to z when z is above or equal to zero. The range is [0, $+\infty$).

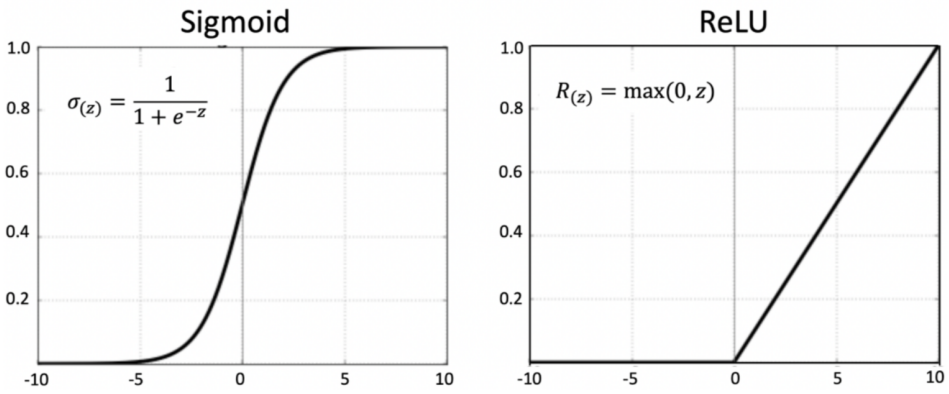

Figure A.9: ReLU vs. logistic sigmoid

The function and its derivative both are monotonic. But the issue is that all the negative values become zero immediately which decreases the ability of the model to fit or train from the data properly. That means any negative input given to the ReLU activation function turns the value into zero immediately in the graph, which in turns affects the resulting graph by not mapping the negative values appropriately.

A.2.3 Gradient Descendent and Optimizers

Optimizers are algorithms used to find the optimal set of parameters for a model during the training process. These algorithms adjust the weights and biases in the model iteratively until they converge on a minimum loss value. Some of the most used ML optimizers are listed below:

Stochastic Gradient Descent (SGD): It is an iterative optimization algorithm commonly used in machine learning and deep learning. It is a variant of gradient descent that performs updates to the model parameters (weights)

based on the gradient of the loss function computed on a randomly selected subset of the training data, rather than on the full dataset.

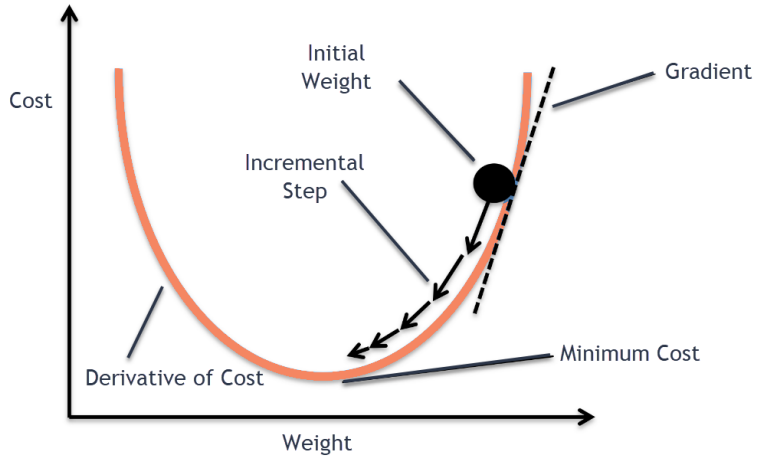

Figure A.10: Stochastic Gradient Descent

The basic idea of SGD is to sample a small random subset of the training data, called a mini-batch, and compute the gradient of the loss function with respect to the model parameters using only that subset. This gradient is then used to update the parameters. The process is repeated with a new random mini-batch until the algorithm converges or reaches a predefined stopping criterion (Figure A.10).

Visualization of SGD. The plot illustrates the process of updating weights to minimize the cost function. Starting from the initial weight, incremental steps are taken in the direction of the negative gradient, representing the derivative of the cost. The trajectory shows the path towards the minimum cost, highlighting how SGD iteratively reduces the cost by adjusting weights.

Stochastic Gradient descent with gradient clipping: Stochastic Gradient Descent with gradient clipping (SGD with GC) is a variant of the standard SGD algorithm that includes an additional step to prevent the gradients from becoming too large during training, which can cause instability and slow convergence.

Gradient clipping involves scaling down the gradients if their norm exceeds a predefined threshold. This helps to prevent the "exploding gradient" problem,

which can occur when the gradients become too large and cause the weights to update too much in a single step.

In SGD with GC, the algorithm computes the gradients on a randomly selected mini-batch of training examples, as in standard SGD. However, before applying the gradients to update the model parameters, the gradients are clipped if their norm exceeds a specified threshold. This threshold is typically set to a small value, such as 1.0 or 5.0.

In training a neural network, an optimization algorithm, or optimizer, is typically used, which consists of a variant of the SGD algorithm. I explain Adam, a widely used optimizer alongside Adagrad, Adadelta, and RMSProp. These optimizers enhance stochastic gradient descent (SGD) by incorporating momentum, which accelerates convergence by directing the descent and reducing oscillations. They update parameters using both current and past gradient components, optimizing the learning process.

Adam: Adam (Adaptive Moment Estimation) is an optimization algorithm used in machine learning and deep learning to optimize the training of neural networks. It maintains a moving average of the gradient's first and second moments, which are the mean and variance of the gradients, respectively. The moving average of the first moment, helps the optimizer to continue moving in the same direction even when the gradients become smaller. The moving average of the second moment, helps the optimizer to scale the learning rate for each parameter based on the variance of the gradients.

Adam also includes a bias correction step to adjust the moving averages since they are biased towards zero at the beginning of the optimization process. This helps to improve the optimization algorithm's performance in the early stages of training.

A.2.4 Backpropagation

To efficiently calculate the gradient of the loss function with respect to the model parameters, in the case of densely connected neural networks, the back propagation algorithm is used. Writing an analytical expression for the

gradient is relatively simple, but evaluating it numerically can require very long calculation times, since the parameters of a neural network are numerous. The backpropagation algorithm makes this procedure simpler and less expensive. It calculates the gradient of the loss function with respect to the parameters using the chain rule, iterating the procedure backwards one level at a time starting from the output level, to avoid redundant calculations of intermediate terms in the chain rule. It is important to specify that this algorithm refers only to the gradient calculation method, while another algorithm, such as stochastic gradient descent, is used to perform learning using the calculated gradient.

A.2.5 Loss Function

Loss functions are one of the most important aspects of neural networks, as they (along with the optimization functions) are directly responsible for fitting the model to the given training data.

A loss function is a function that compares the target and predicted output values; measures how well the neural network models the training data. When training, the aim is to minimize this loss between the predicted and target outputs. The hyperparameters are adjusted to minimize the average loss. The weights, W^T, and biases, b, that minimize the value of J (average loss) can be found.

$$J(W^T, b) = \frac{1}{n} \sum_{i=1}^{n} L\left(\hat{y}^{(i)}, y^{(i)}\right) \tag{A.2}$$

where n is the number of training samples, $y^{(i)}$ is the true label of the ith sample, and $\hat{y}^{(i)}$is its predicted label ($i = 1, 2, ..., n$). We can think of this akin to residuals, in statistics, which measure the distance of the actual y values from the regression line (predicted values), the goal being to minimize the net distance. In supervised learning, there are two main types of loss functions. These correlate to the two major tasks of neural networks: regression and classification loss functions:

Regression Loss Functions: Used in regression neural networks; given an input value, the model predicts a corresponding continuous-valued output value

(rather than pre-selected labels). Examples included Mean Squared Error, Mean Absolute Error, etc.

Mean Squared Error (MSE): One of the most popular loss functions, MSE finds the average of the squared differences between the target and the predicted outputs. This function has numerous properties that make it especially suited for calculating loss.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2 \tag{A.3}$$

The difference is squared, which means it does not matter whether the predicted value is above or below the target value; however, values with a large error are penalized. MSE is also a convex function with a clearly defined global minimum. This allows us to more easily utilize gradient descent optimization to set the weight values. However, one disadvantage of this loss function is that it is very sensitive to outliers; if a predicted value is significantly greater than or less than its target value, this will significantly increase the loss.

Mean Absolute Error (MAE): MAE finds the average of the absolute differences between the target and the predicted outputs.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y^{(i)} - \hat{y}^{(i)}| \tag{A.4}$$

This loss function is used as an alternative to MSE in some cases. As mentioned previously, MSE is highly sensitive to outliers, which can dramatically affect the loss because the distance is squared. MAE is used in cases when the training data has a large number of outliers to mitigate this. It also has some disadvantages; as the average distance approaches 0, gradient descent optimization will not work, as the function's derivative at 0 is undefined (which may result in an error).

Huber Loss: Huber Loss is a combination of MAE and MSE, but it depends on an additional parameter call delta(δ) that influences the shape of the loss function. This parameter needs to be fine-tuned by the algorithm:

$$\text{Huber Loss} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2 & \text{if } |y^{(i)} - \hat{y}^{(i)}| \leq \delta \\ \frac{1}{n}\sum_{i=1}^{n}\delta\left(|y^{(i)} - \hat{y}^{(i)}| - \frac{1}{2}\delta\right) & \text{if } |y^{(i)} - \hat{y}^{(i)}| > \delta \end{cases} \tag{A.5}$$

When the values are large (far from the minima) the function has the behavior of the MAE, closed to the minima, the function behaves like the MSE. So, the δ parameter is the sensitivity to outliers. If the absolute difference between the actual and predicted value is less than or equal to a threshold value, σ, then MSE is applied. Otherwise, if the error is sufficiently large, MAE is applied.

The most common loss functions used for classification tasks are discussed in the following:

Binary Cross-Entropy/Log Loss: This is the loss function used in binary classification models,where the model takes in an input and has to classify it into one of two pre-set categories (for example, these categories could be signal and background).

$$\text{CE Loss} = \frac{1}{n}\sum_{i=1}^{N} - \left(y^{(i)}.\log(p_i) + (1-y^{(i)}).\log(1-p_i)\right) \tag{A.6}$$

Classification neural networks work by outputting a vector of probabilities: the probability that the given input fits into each of the pre-set categories; then selecting the category with the highest probability as the final output. In binary classification, there are only two possible actual values of $y^{(i)}$: 0 or 1. Thus, to accurately determine loss between the actual and predicted values, it needs to compare the actual value (0 or 1) with the estimated probability that the input aligns with that category (p_i = probability that the category is 1; 1 p_i = probability that the category is 0).

Categorical Cross-Entropy Loss: In cases where the number of classes is greater than two, categorical cross-entropy is utilized. This follows a very similar process to binary cross-entropy.

$$\text{CE Loss} = -\frac{1}{n}\sum_{i=1}^{N}\sum_{i=1}^{M} y^{(ij)}.\log(p_{ij}) \tag{A.7}$$

where y_{ij} is equal to 1 if the ith sample belongs to the jth class, and to 0 otherwise ($i = 1, 2, ..., N, j = 1, 2, ..., M$). Binary cross-entropy is a special case of categorical cross-entropy, where $M = 2$ (the number of categories is 2).

A.3 Feature Scaling Techniques in NN

In neural networks, the features (inputs) fed into the network can often have different scales and ranges. This variation in feature scales can present challenges during the training process, particularly in optimization algorithms like gradient descent, where it's crucial for convergence to occur efficiently towards the minima of the loss function.

The quality of data used for training models is key. While complex algorithms get a lot of attention, the golden step is data preprocessing. This crucial step turns raw data into usable features for effective machine learning.

Most machine learning algorithms assume that all features contribute equally to predictions. But when features have different ranges and units, this assumption breaks down, impacting their importance. That's where normalization comes in. It's a crucial part of data preprocessing that ensures all features have similar numerical magnitudes. This prevents features with larger values from overshadowing others. By normalizing the feature values, they are transformed to a standardized range, typically between 0 and 1, facilitating smoother and faster convergence during training.

When data is fed directly to a machine learning model without preprocessing, the algorithm may prioritize features with larger scales, potentially overlooking the importance of smaller-scale features. This bias can result in less accurate predictions and suboptimal model performance. Normalizing the features ensures that each feature contributes proportionally to the model's learning process. This allows the model to effectively learn patterns across all features, leading to a more accurate representation of the underlying relationships in the data.

Min-Max scaling and Z-score normalization (standardization) are the two fundamental techniques for normalization.

Min-max scaling: Often simply referred to as "normalization," adjusts features to fit within a specific range, typically between 0 and 1. The formula for

min-max scaling is:

$$X_{\text{normalized}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{A.8}$$

Where X is a feature value to be normalized, X_{min} is the minimum value of the feature in the dataset, and X_{max} is the maximum value.

When X is the minimum value, the numerator becomes zero ($X_{\text{min}} - X_{\text{min}}$), resulting in a normalized value of 0. When X is the maximum value, the numerator equals the denominator ($X_{\text{max}} - X_{\text{min}}$), resulting in a normalized value of 1. When X is neither the minimum nor the maximum, the normalized value falls between 0 and 1.

Min-max scaling is particularly effective when the dataset's approximate upper and lower bounds are known, and it contains few or no outliers. it is also effective when the data distribution is unknown or non-Gaussian, and the data is roughly uniformly distributed.

Z-Score Normalization (Standardization): Z-score normalization, or standardization, is based on the assumption that the data follows a Gaussian (bell curve) distribution. It transforms features so that they have a mean (μ) of 0 and a standard deviation (σ) of 1. The standardization formula is:

$$X_{\text{standardized}} = \frac{X - \mu}{\sigma} \tag{A.9}$$

Unlike the min-max scaling technique, standardization does not restrict feature values to a specific range. Instead, it expresses features in terms of the number of standard deviations from the mean.

A.4 Evaluation of the Model Performances

Model evaluation and metrics are crucial aspects of developing and deploying machine learning models across various applications. They provide essential checkpoints for assessing how well models perform in real-world situations. By examining metrics like accuracy, precision, recall, and F1 score, practitioners can understand both the strengths and weaknesses of their models, allowing for iterative improvements. Additionally, model evaluation helps

ensure that models can make accurate predictions on new data they haven't seen before. However, there are challenges involved, including issues with data quality, concerns about bias and fairness, and the ongoing effort to balance between overfitting and underfitting. Despite these challenges, a solid understanding of model evaluation and metrics empowers practitioners to create reliable machine learning solutions that positively impact diverse fields.

A.4.1 Confusion Matrix

The confusion matrix is a fundamental tool in evaluating the performance of classification models in machine learning. It takes the form of an N × N matrix, where N represents the number of predicted classes. In binary classification problems, where for example there are two possible classes, N equals 2, resulting in a 2 x 2 matrix.

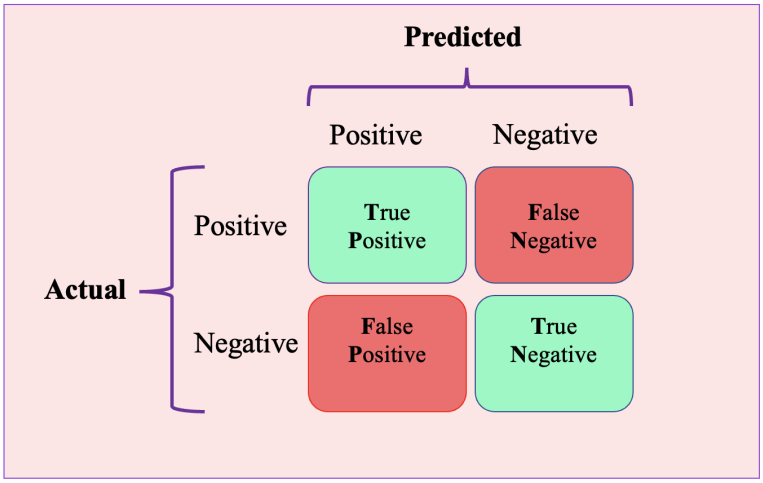

Figure A.11

This matrix provides a comprehensive overview of the model's predictions by displaying four different combinations of predicted and actual values: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These combinations are instrumental in computing various performance metrics such as precision, recall, specificity, accuracy, and area under the receiver operating characteristic (AUC-ROC) curve.

By analyzing the entries of the confusion matrix, practitioners can gain insights into the model's ability to correctly classify instances and identify areas for improvement. Additionally, the confusion matrix serves as the foundation for constructing other evaluation metrics and visualizations, making it an indispensable tool in the model evaluation process.

A.4.2 Accuracy

Accuracy is one metric for evaluating classification models. In other words, accuracy answers the question: how often the model is right? Formally, accuracy has the following definition:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{A.10}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{A.11}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

While it seems like the ideal goal would be to achieve 100% model accuracy when developing a model, getting this result is not something to look forward to. Achieving 100% machine learning model accuracy is typically a sign of some error, such as overfitting; that is, the model learns the characteristics of the training set so specifically that it cannot generalize to unseen data in the validation and evaluation sets. It can also be a sign of a logical bug or data leakage, which is when the feature set contains information about the label that should not be present as unavailable at prediction time. This is because it is practically impossible to have perfect prediction machine accuracy given that complex systems are inherently underdetermined. If you can fully determine the system and achieve 100% accuracy, then you should not be using machine

learning. Instead, you should consider classical modeling or a heuristic.
We can measure the accuracy on a scale of 0 to 1 or as a percentage. The higher the accuracy, the better. We can achieve a perfect accuracy of 1.0 when every prediction the model makes is correct.

A.4.3 The ROC Curve and the AUC

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds as shown in Figure A.12. This curve plots two parameters:

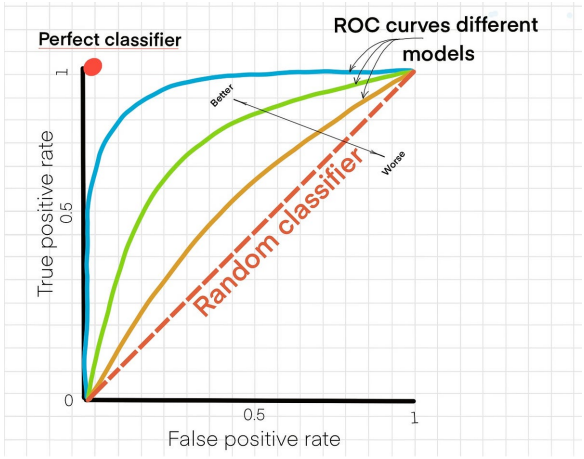

Figure A.12: The ROC space for a "better" and "worse" classifier.

True Positive Rate (TPR) is defined as:

$$TRP = \frac{TP}{TP + FN} \tag{A.12}$$

and the False Positive Rate (FPR) that is defined as:

$$FPR = \frac{FP}{FP + TN} \tag{A.13}$$

For example in a signal-background classification problem, TPR is the fraction of signal events properly classified as signal over the total number of signal events, this is often referred to as signal efficiency in HEP analyses; while the FPR is the fraction of background events misclassified as signal over the total

number of background events, in HEP analyses it is often mentioned and used the background rejection that is defined as: REI= 1-FPR).
The metric connected to the ROC curve is the area under the curve (AUC). AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example. The higher the value of AUC coefficient, the better. AUC = 1 means a perfect classifier, AUC = 0.5 is obtained for purely random classifiers. AUC <0.5 means the classifier performs worse than a random one (Figure A.12).

A.4.4 Precision

Precision is a metric that measures how often a machine learning model correctly predicts the positive class. It can calculated by dividing the number of correct positive predictions (true positives) by the total number of instances the model predicted as positive (both true and false positives):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{A.14}$$

In other words, precision answers the question: how often the positive predictions are correct? we can measure the precision on a scale of 0 to 1 or as a percentage. The higher the precision, the better. To achieve a perfect precision of 1.0, the model should be always right when predicting the target class: it never flags anything in error.

A.4.5 Recall

Recall is a metric that measures how often a machine learning model correctly identifies positive instances (true positives) from all the actual positive samples in the dataset. You can calculate recall by dividing the number of true positives by the number of positive instances. The latter includes true positives (successfully identified cases) and false negative results (missed cases).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FN}} \tag{A.15}$$

In other words, recall answers the question: can an ML model find all instances of the positive class?
We can measure the recall on a scale of 0 to 1 or as a percentage. The higher the recall, the better. To achieve a perfect recall of 1.0, the model should find all instances of the target class in the dataset.

A.4.6 F1 Score

The F1 score or F-measure is described as the harmonic mean of the precision and recall of a classification model. The two metrics contribute equally to the score, ensuring that the F1 metric correctly indicates the reliability of a model.
The F1 score formula is:

$$\mathrm{F}_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} \tag{A.16}$$

It's important to note that calculating the F1 score using arithmetic mean may not appropriately represent the model's overall performance, especially when precision and recall have considerably varied values. That's because the arithmetic mean focuses on the sum of values and their average.
On the other hand, the harmonic mean emphasizes the reciprocal of values. It is computed by dividing the total number of values by the sum of their reciprocals. Hence, it enhances the effect of the smaller value on the overall calculation to achieve a balanced measurement. As a result, the F1 score takes into account both precision-recall while avoiding the overestimation that the arithmetic mean might cause.
The F1 score ranges between 0 and 1, with 0 denoting the lowest possible result and 1 denoting a flawless result, meaning that the model accurately predicted each label.
A high F1 score generally indicates a well-balanced performance, demonstrating that the model can concurrently attain high precision and high recall. A low F1 score often signifies a trade-off between recall and precision, implying

that the model has trouble striking that balance.

Bibliography

[1] M. K. Gaillard, P. D. Grannis and F. J. Sciulli, Rev. Mod. Phys. **71** (1999), S96-S111 doi:10.1103/RevModPhys.71.S96 [arXiv:hep-ph/9812285 [hep-ph]].

[2] M. Thomson, Modern Particle Physics. Cambridge University Press, 2013.

[3] The ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," Physics Letters B, vol. 716, no. 1, pp. 1–29, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[4] The CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the cms experiment at the lhc," Physics Letters B, vol. 716, no. 1, p. 30–61, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[5] J. C. Collins, D. E. Soper, and G. F. Sterman. "Factorization of Hard Processes in QCD". In: Adv. Ser. Direct. High Energy Phys. 5 (1989), pp. 1–91. DOI: 10.1142/ 9789814503266_0001. arXiv: hep-ph/0409313 [hep-ph].

[6] V. N. Gribov and L. N. Lipatov, Tech. Rep., Inst. of Nuclear Physics, Leningrad (1972).

[7] Y. Dokshitzer, Sov. Phys. JETP p. 641 (1977).

[8] G. Altarelli and G. Parisi, Nuclear Physics B 126, 298 (1977).

[9] ATLAS Collaboration, "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13TeV with the ATLAS detector." , Eur. Phys. J. C 81 (2021) 178.

[10] H. B. Hartanto, R. Poncelet, A. Popescu and S. Zoia, Next-to-next-to-leading order QCD corrections to Wbb^{-} production at the LHC, Phys. Rev. D **106** (2022) no.7, 074016 doi:10.1103/PhysRevD.106.074016 [arXiv:2205.01687 [hep-ph]].

[11] L. Buonocore et al., Associated production of a W boson and massive bottom quarks at next-to-next-to-leading order in QCD, Phys. Rev. D 107 (2023) 074032 [arXiv:2212.04954] [INSPIRE].

[12] CMS Collaboration, "Measurement of the Production Cross Sections for a W Boson and and two b jets in pp collisions at $\sqrt{s}$ =7 TeV", PLB 735 (2014) 204.

[13] ATLAS Collaboration, "Measurement of the cross-section for W boson production in association with b jets in pp collisions at $\sqrt{s}$ =7 TeV", JHEP 06, 084 (2013).

[14] ATLAS collaboration, Measurements of the production cross-section for a Z boson in association with b-jets in proton–proton collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector, JHEP 07 (2020) 044.

[15] G. Aad *et al.* [ATLAS], "Measurements of the production cross-section for a Z boson in association with b- or c-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," [arXiv:2403.15093 [hep-ex]].

[16] CMS Collaboration, Measurement of the production cross section for Z+b-jet in proton-proton collisions at $\sqrt{s}$= 13 TeV, Phys. Rev. D 105 (2022) 092014.

[17] The ATLAS Collaboration. The atlas experiment at the cern large hadron collider. Journal of Instrumentation, 3(08):S08003. 437 p, 2008. URL https://cdsweb.cern.ch/record/1129811/.

[18] ATLAS Experiment, Inner Detector, https://atlas.cern/Discover/Detector/Inner-Detector.

[19] ATLAS Collaboration, Performance of the ATLAS Trigger System in 2015, In: Eur. Phys. J. C77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[20] ATLAS Collaboration, Performance of the ATLAS Trigger System in 2010, In: Eur. Phys. J. C72 (2012), p. 1849. DOI: 10.1140/epjc/s10052-011-1849-1. arXiv: 1110.1530 [hep-ex].

[21] The ATLAS TDAQ Collaboration, The ATLAS Data Acquisition and High Level Trigger system, In: Journal of Instrumentation 11.06 (2016), P06008. http://stacks.iop.org/1748-0221/11/i=06/a=P06008 .

[22] ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$ =13 TeV, Eur. Phys. J. C 76 (2016) 292, arXiv:1603.05598 [hep-ex].

[23] M. Cacciari, G. P. Salam and G. Soyez, The antikt jet clustering algorithm, JHEP 04 (2008) 063, arXiv: 0802.1189 [hep-ph].

[24] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, Eur. Phys. J. C, 72(3), Mar. 2012. arXiv: 1111.6097 [hep-ph].

[25] ATLAS Collaboration, Jet reconstruction and performance using particle flow with the ATLAS Detector, Eur. Phys. J. C 77 (2017) 466, arXiv: 1703.10485 [hep-ex].

[26] ATLAS Collaboration, Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ =7 TeV, Eur. Phys. J. C, 73(3): (2304), Mar. 2013. arXiv: 1112.6426 [hep-ex].

[27] ATLAS Collaboration, Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Eur. Phys. J. C, 81(689), 2021. arXiv: 2007.02645 [hep-ex].

[28] ATLAS Collaboration, Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ =8 TeV using the ATLAS detector, Eur. Phys. J. C 76 (2016) 581, arXiv: 1510.03823 [hep-ex].

[29] ATLAS Collaboration, ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset, CERN-EP-2022-226; FTAG-2019-07, 2019, arXiv: 2211.16345v1 [physics.data-an], https://cds.cern.ch/record/2842028.

[30] ATLAS Collaboration, E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC pp collisions, ATLAS-CONF-2018-023, 2018, url: https://cds.cern.ch/record/2625233.

[31] T. Sjöstrand et al., An introduction to PYTHIA 8.2, Comput. Phys. Commun. 191 (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on p. 10).

[32] ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021, 2014, Available: https://cds.cern.ch/record/1966419 (cit. on p. 10).

[33] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio et al., Parton distributions with LHC data, Nucl. Phys. B 867 (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on p. 10).

[34] R. Frederix and S. Frixione, Merging meets matching in MC@NLO, JHEP 12 (2012) 061, arXiv: 1209.6215 [hep-ph] (cit. on p. 10).

[35] ATLAS Collaboration, Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment, JHEP 08 (2021) 089, arXiv: 2112.09588 [hep-ex] (cit. on p. 10).

[36] T. Gleisberg and S. Höche, Comix, a new matrix element generator, JHEP 12 (2008) 039, arXiv: 0808.3674 [hep-ph] (cit. on p. 11).

[37] F. Buccioni et al., OpenLoops 2, Eur. Phys. J. C 79 (2019) 866, arXiv: 1907.13071 [hep-ph] (cit. on p. 11).

[38] Cascioli, Fabio and Maierhofer, Philipp and Pozzorini, Stefano, Scattering Amplitudes with Open Loops, Phys. Rev. Lett. 108 (2012) 111601, arXiv: 1111.5206 [hep-ph] (cit. on p. 11).

[39] A. Denner, S. Dittmaier and L. Hofer, Collier: A fortran-based complex one-loop library in extended regularizations, Comput. Phys. Commun. 212 (2017) 220, arXiv: 1604.06792 [hep-ph] (cit. on p. 11).

[40]] NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II, JHEP 04 (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on pp. 11, 42, 206).

[41] S. Schumann and F. Krauss, A Parton shower algorithm based on Catani-Seymour dipole factorisation, JHEP 03 (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on p. 11).

[42] J.-C. Winter, F. Krauss and G. Soff, A modified cluster-hadronization model, Eur. Phys. J. C 36 (2004) 381, arXiv: hep-ph/0311085 (cit. on p. 7).

[43] S. Höche, F. Krauss, M. Schönherr and F. Siegert, A critical appraisal of NLO+PS matching methods, JHEP 09 (2012) 049, arXiv: 1111.1220 [hep-ph] (cit. on p. 11).

[44] F Siegert et al., QCD matrix elements + parton showers: The NLO case, JHEP 04 (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 11).

[45] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, QCD Matrix Elements + Parton Showers, JHEP 11 (2001) 063, arXiv: hep-ph/0109231 (cit. on p. 11).

[46] S. Höche, F. Krauss, S. Schumann and F. Siegert, QCD matrix elements and truncated showers, JHEP 05 (2009) 053, arXiv: 0903.1219 [hep-ph] (cit. on p. 11).

[47] S. Frixione, P. Nason and G. Ridolfi, A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction, JHEP 09 (2007) 126, arXiv: 0707.3088 [hep-ph] (cit. on p. 11).

[48] S. Frixione, P. Nason and C. Oleari, Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, JHEP 11 (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 11).

[49] S. Alioli, P. Nason, C. Oleari and E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, JHEP 06 (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 11).

[50] ATLAS Collaboration, Studies on top-quark Monte Carlo modelling for Top2016, ATL-PHYS-PUB-2016-020, 2016, https://cds.cern.ch/record/2216168 (cit. on p. 11).

[51] M. Beneke, P. Falgari, S. Klein and C. Schwinn, Hadronic top-quark pair production with NNLL threshold resummation, Nucl. Phys. B 855 (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 11).

[52] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation, Phys. Lett. B 710 (2012) 612, arXiv: 1111.5869 [hep-ph] (cit. on p. 11).

[53] P. Bärnreuther, M. Czakon and A. Mitov, Percent level precision physics at the Tevatron: First genuine NNLO QCD corrections to $q\bar{q} \to t\bar{t} + X$, Phys. Rev. Lett. 109 (2012) 132001, arXiv: 1204.5201 [hep-ph] (cit. on p. 11).

[54] M. Czakon and A. Mitov, NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels, JHEP 12 (2012) 054, arXiv: 1207.0236 [hep-ph] (cit. on p. 11).

[55] M. Czakon and A. Mitov, NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction, JHEP 01 (2013) 080, arXiv: 1210.6832 [hep-ph] (cit. on p. 11).

[56] M. Czakon, P. Fiedler and A. Mitov, Total top-quark-pair-production cross section at hadron colliders through $O(\alpha_S^4)$, Phys. Rev. Lett. 110 (2013) 252004, arXiv: 1303.6254 [hep-ph] (cit. on p. 11).

[57] M. Czakon and A. Mitov, Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders, Comput. Phys. Commun. 185 (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 11).

[58] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber and C. D. White, Single-top hadroproduction in association with a W boson, JHEP 07 (2008) 029, arXiv: 0805.3067 [hep-ph] (cit. on p. 11).

[59] E. Bothmann et al., Event Generation with Sherpa 2.2, SciPost Phys. 7 (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on p. 11).

[60] ATLAS Collaboration, The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model, ATL-PHYS-PUB-2016-017, 2016, https://cds.cern.ch/record/2206965 (cit. on p. 11).

[61] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instrum. Meth. A 462 (2001) 152 (cit. on p. 11).

[62] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 11).

[63] GEANT4 Collaboration, GEANT4: A Simulation toolkit, Nucl. Instrum. Meth. A506 (2003) 250 (cit. on p. 11).

[64] ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV, Eur. Phys. J. C 76 (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 14).

[65] M. Cacciari, G. P. Salam and G. Soyez, The anti-k_t jet clustering algorithm, JHEP 04 (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on p. 14).

[66] M. Cacciari, G. P. Salam and G. Soyez, FastJet User Manual, Eur. Phys. J. C 72 (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 14).

[67] ATLAS Collaboration, Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector, Eur. Phys. J. C 76 (2016) 581, arXiv: 1510.03823 [hep-ex](cit. on p. 14).

[68] G. Aad *et al.* [ATLAS],"Measurements of the production cross-section for a Z boson in association with b-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," JHEP **07** (2020), 044 doi:10.1007/JHEP07(2020)044 [arXiv:2003.11960 [hep-ex]].

[69] Keras, url: https://keras.io/.

[70] Martin Abadi et al., 'TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems', Software available from tensorflow.org, 2015, https://www.tensorflow.org/.

[71] P. Flach, The Art and Science of Algorithms that Make Sense of Data, Cambridge: Cambridge University Press, 2012, pp. 367–382.

[72] Lek, Sovan, and Jean-François Guégan. "Artificial neural networks as a tool in ecological modelling, an introduction." Ecological modelling 120.2-3 (1999): 65-73.

[73] Sazli, Murat. (2006). A brief review of feed-forward neural networks. Communications Faculty Of Science University of Ankara. 50. 11-17. 10.1501/commua1-2_0000000026.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Evelin Meoni. She has been more than just a supervisor; she has been a friend on this journey. Her constant support, guidance, and encouragement have been instrumental in the successful completion of this thesis.

Furthermore, I would like to thank Dr. Daniele Zanzi from Freiburg University and Professor Federico Sforza from Università di Genova. Although I did not have the opportunity to meet them in person, their pioneering work has been influential in guiding my research at the beginning and shaping my understanding of the subject.