



UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di scienze e tecnologie
Laurea Triennale in Fisica

Separating the production of single top quarks in association with a Z boson from background events with Machine Learning techniques at the **ATLAS experiment**

Relatore:
Lidia Dell'Asta

Tesi di Laurea Triennale di:
Niccolò Laurora
Matricola: **932540**

Anno Accademico 2021/2022

alla mia famiglia

Abstract

The study of the top quark is essential for the ATLAS (**A** Toroidal **L**H_C Apparatu**S**) experiment, because it could help shedding a light over the electroweak symmetry breaking mechanism.

The top quark was discovered at Tevatron in 1995.

It is the heaviest particle of the Standard Model (SM) with a coupling to the Higgs boson close to one. The overwhelming amount of data collected by the ATLAS Experiment at the LHC at CERN during Run-II has allowed the recording and measuring of rare processes, foreseen by the SM, that have never been observed before.

Recently, the ATLAS Experiment has registered a sporadic process of primary importance, which involves the top quark: the production of a single top quark in association with a Z boson (tZq event). The cross section of this process, predicted by the SM to be 102 fb, has been measured with an uncertainty of 15%. Since then the aim has been to find new ways to measure the process, decreasing the uncertainty. The purpose of this thesis is to investigate Machine Learning techniques which would improve the discrimination of tZq events from background sources in proton-proton collisions at $\sqrt{s} = 13$ TeV in the ATLAS experiment.

In the first and second chapter I will provide an overview on the Standard Model and the top quark physics. In the third chapter I will give a brief description of LHC and the ATLAS experiment. In the fourth chapter I will give a short introduction on Machine Learning. In the fifth and last chapter I will explain the work I have done in the past few months.

Contents

1	The Standard Model	1
1.1	Quantum Chromodynamics	2
1.1.1	Leptons	3
1.1.2	Quarks	3
1.1.3	Gluons	4
1.1.4	QCD interaction	5
1.2	Electroweak sector	5
1.2.1	Electroweak unification	6
1.2.2	Leptons	8
1.2.3	Quarks	8
1.2.4	Electroweak bosons	9
1.3	The Higgs Mechanism	9
1.3.1	Spontaneous symmetry breaking	10
1.3.2	Yukawa term	11
1.4	The Lagrangian of the Standard Model	11
2	Top quark	13
2.1	Top quark properties and decay	13
2.2	Top quark production	14
2.2.1	Top quark production in association with a Z boson	15
3	The Large Hadron Collider and the ATLAS Experiment	19
3.1	The Large Hadron Collider	19
3.1.1	Accelerator structure and fundamental properties	20
3.1.2	ATLAS coordinate system	22
3.2	The ATLAS detector	23
3.2.1	Inner Detector	24
3.2.2	Calorimetry system	25
3.2.3	Muon Spectrometer	26
3.3	Object Reconstruction	27
3.3.1	Primary vertex	27
3.3.2	Electrons	27

CONTENTS

3.3.3	Muons	27
3.3.4	Jets	28
3.3.5	b-jet	28
3.3.6	Missing transverse momentum	29
4	Machine Learning	31
4.1	Supervised Learning	32
4.2	Artificial Neural Networks	34
4.2.1	Binary and Multi-Class Classification	36
4.2.2	Binary Classifier	36
4.2.3	Multi-class Classifier	38
5	Separation of tZq from background with DNNs	41
5.1	Data Samples	41
5.1.1	Signal	41
5.1.2	Background	42
5.2	Event Selections and Signal Regions	43
5.3	Definition of the DNN classes	45
5.4	Variables for the DNN trainings	46
5.5	DNN trainings	56
5.5.1	Loss functions	56
5.5.2	Trainings for the 2j1b signal region	59
5.5.3	Trainings for the 3j1b signal region	71
5.5.4	Trainings for the nj1b signal region	83
5.5.5	Trainings summary	96
5.6	Conclusions	99
	Bibliography	103

Chapter 1

The Standard Model

The Standard Model (SM) is the theory used to describe the interactions between fundamental particles. As far as we know, the interactions between fundamental particles are just four: *gravitational*, *electromagnetic*, *weak* and *strong*. These four forces can be ordered by their strength as shown in table 1.1.

Force	Strength	Theory	Mediator
Strong	10	Chromodynamics (QCD)	Gluon
Electromagnetic	10^{-2}	Electrodynamics (QED)	Photon
Weak	10^{-13}	Flavourdynamics (QFD)	W and Z
Gravity	10^{-42}	Geometrodynamics	Graviton

Table 1.1: Interactions' properties [1].

The SM is a Gauge theory based on the $U(1) \times SU(2) \times SU(3)$ symmetry group. This theory consists of two main sectors: quantum chromodynamics (QCD), which describes the strong force and is based on the $SU(3)$ symmetry, and the Electroweak sector (EW), which describes the weak and electromagnetic forces and is based on the $U(1) \times SU(2)$ symmetry.

The SM does not offer a proper description of *Gravity*, it only suggests the existence of the *Graviton*, the mediator of the gravitational interaction. However, this particle has not been discovered thus far, therefore a completely satisfactory quantum theory of gravity has yet to be formulated.

In the following paragraphs I will briefly discuss the two main sectors of the SM using Griffiths' *Introduction To Elementary Particles* [1] as reference.

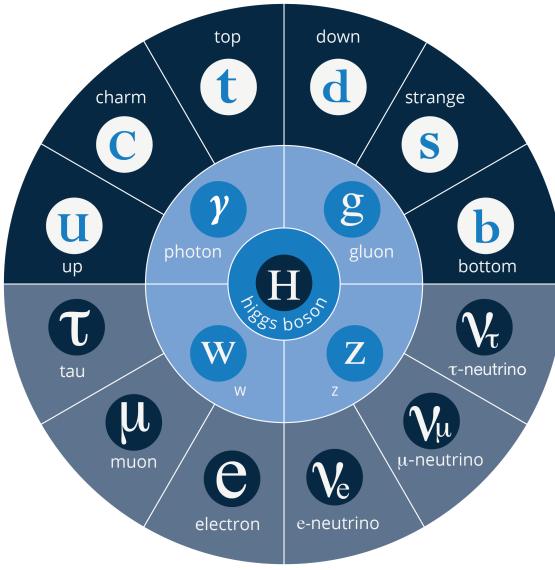


Figure 1.1: The particles of the Standard Model.

1.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory of the strong interaction [2, 3].

QCD is based on the $SU(3)$ symmetry, where the number three refers to the possible charges of *color*: *red, green, blue*.

Unlike $U(1)$ and $SU(2)$, $SU(3)$ is a non-abelian group.

Color is a quantum number introduced to explain the existence of hadrons made up of the same three quarks, such as Ω^- and Δ^{++} [4–6].

Since quarks are fermions, they obey the Fermi-Dirac statistic, which doesn't allow two or more fermions to occupy the same quantum state (Pauli's principle).

The introduction of this new quantum number solves the apparent violation of Pauli's principle in hadrons made up of the same three quarks. Therefore, if we consider the Ω^- (Δ^{++}) hadron, its constituents are three strange (up) quarks in three different color charges.

The experimental evidence of the existence of color is given by the collision of electrons and positrons. This collision can produce four different final states (table 1.2).

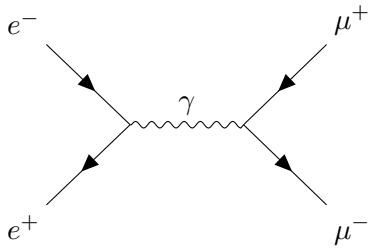
Label	Final State
Bhabha Scattering	$e^- + e^+$
Annihilation	$\gamma + \gamma$
Muons	$\mu^- + \mu^+$
Hadrons	$q + \bar{q}$

Table 1.2: Final states of $e^- + e^+$ collisions.

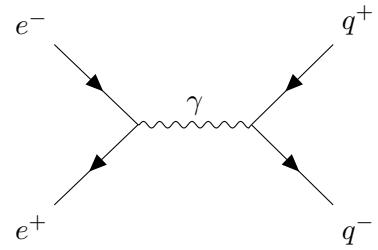
The ratio between the cross-sections of the muons and hadrons final states is compatible with the existence of color.

$$R = \frac{\sigma(e^- + e^+ \rightarrow \text{Hadrons})}{\sigma(e^- + e^+ \rightarrow \mu^- + \mu^+)} = N_C \sum_i Q_i^2 \quad (1.1.1)$$

where Q_i is the charge of i-th quark and N_C is the number of colors (3).



Muonic final state.



Hadronic final state.

1.1.1 Leptons

Leptons are elementary fermions particles. They occur in two main types: charged leptons and neutrinos, which are electrically neutral. Three generations of leptons have been discovered thus far, in order of increasing mass the three families are: *electronic*, *muonic*, *tauonic*.

Lepton	Mass [MeV]	Charge [e]	Mean Lifetime	Generation
e	0.51099895000	-1	6.6×10^{28} yr	1
μ	105.6583755	-1	2.1969811×10^{-6} s	2
τ	1776.86	-1	290.3×10^{-15} s	3
ν_e	$< 1.1 \times 10^{-6}$	0	Unknown	1
ν_μ	$< 1.1 \times 10^{-6}$	0	Unknown	2
ν_τ	$< 1.1 \times 10^{-6}$	0	Unknown	3

Table 1.3: Leptons' properties [7].

Although neutrinos are massive particles, they are considered massless within the SM.

1.1.2 Quarks

Quarks are elementary fermions particles. They occur in six different *flavours* and they are organized in three *generations*. Quarks can be divided in *light-flavour* (up, down, strange) and *heavy-flavour* (charm, bottom, top). Moreover quarks can be classified in two main categories: *Up-Type* and *Down-type*.

Quark	Mass	Charge [e/3]	Generation	Class
Up (<i>u</i>)	$2.16^{+0.49}_{-0.26}$ MeV	2	1	Up-Type
Down (<i>d</i>)	$4.67^{+0.48}_{-0.17}$ MeV	-1	1	Down-type
Charm (<i>c</i>)	1.27 ± 0.02 GeV	2	2	Up-Type
Strange (<i>s</i>)	$93.4^{+8.6}_{-3.4}$ MeV	-1	2	Down-type
Top (<i>t</i>)	172.69 ± 0.30 GeV	2	3	Up-Type
Bottom (<i>b</i>)	$4.18^{+0.03}_{-0.02}$ GeV	-1	3	Down-type

Table 1.4: Quarks' properties [7].

It is noteworthy to highlight that top's mass has been directly measured, unlike other quarks' masses. Quarks' masses are a non trivial measurement considering they do not exist as free particles, but solely as bound states. This property is shared by quarks and gluons and it is known as *confinement*. Quarks are characterized by *color*, an internal quantum number that has three different states: *red*, *green*, *blue*. The quark field can be represented as a colour triplet in QCD:

$$\psi(x) = \begin{pmatrix} \psi_R(x) \\ \psi_G(x) \\ \psi_B(x) \end{pmatrix} \quad (1.1.2)$$

1.1.3 Gluons

Gluon is a massless boson which mediates the strong interaction. Each gluon carries both color and anticolor charge. This gives nine possible combinations of color and anticolor in gluons, organized in an octet and a singlet.

$$\begin{aligned} |1\rangle &= (r\bar{b} + b\bar{r})/\sqrt{2} & |5\rangle &= -i(r\bar{b} - b\bar{r})/\sqrt{2} \\ |2\rangle &= (r\bar{g} + g\bar{r})/\sqrt{2} & |6\rangle &= -i(r\bar{g} - g\bar{r})/\sqrt{2} \\ |3\rangle &= (g\bar{b} + b\bar{g})/\sqrt{2} & |7\rangle &= -i(b\bar{g} - g\bar{b})/\sqrt{2} \\ |4\rangle &= (r\bar{r} - b\bar{b})/\sqrt{2} & |8\rangle &= (r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{6} \end{aligned}$$

Gluons' color octet

The color-octet is made of colored gluons, whereas the color-singlet is a colorless gluon, that has not been observed in nature thus far.

$$|9\rangle = (r\bar{r} + g\bar{g} + b\bar{b})/\sqrt{3}$$

Gluons' color singlet

1.1.4 QCD interaction

The mass of a quark does not depend on its color charge.

Therefore the free Lagrangian for a particular flavour is:

$$\begin{aligned}\mathcal{L} = & [i\bar{\psi}_R \gamma^\mu \partial_\mu \psi_R - m\bar{\psi}_R \psi_R] + [i\bar{\psi}_G \gamma^\mu \partial_\mu \psi_G - m\bar{\psi}_G \psi_G] \\ & + [i\bar{\psi}_B \gamma^\mu \partial_\mu \psi_B - m\bar{\psi}_B \psi_B]\end{aligned}\quad (1.1.3)$$

The quark field (1.1.2) is extremely convenient since it allows to recast the 1.1.3 Lagrangian in the form of the original Dirac Lagrangian for a free fermion.

$$\mathcal{L} = [i\bar{\psi} \gamma^\mu \partial_\mu \psi - m\bar{\psi} \psi] \quad (1.1.4)$$

Both 1.1.3 and 1.1.4 Lagrangians express the free Lagrangian for a particular quark flavour, but the second is more compact. As every gauge theory, the strong interaction is introduced imposing the invariance under a local gauge transformation (1.1.5) of the free Lagrangian (1.1.4). The local gauge invariance, which introduces the strong interaction in the free Lagrangian is the following:

$$\psi \rightarrow S\psi = e^{i\vec{\lambda}\vec{a}}\psi \quad (1.1.5)$$

where \vec{a} is a vector with eight inputs and $\lambda, \lambda_2, \dots, \lambda_8$ are the Gell-Mann matrices.

Therefore, the Lagrangian which describes the strong interaction mediated by eight colored gluons is the following:

$$\mathcal{L} = [i\bar{\psi} \gamma^\mu \partial_\mu \psi - m\bar{\psi} \psi] - q(\bar{\psi} \gamma^\mu \vec{\lambda} \psi) \vec{A}_\mu - \frac{1}{16} F^{\mu\nu} F_{\mu\nu} \quad (1.1.6)$$

where ψ is the quark field, m is the quark mass-term, q is the strong coupling constant and $F^{\mu\nu}$ is the gluon field strength-tensor (1.1.7).

$$F^{\mu\nu} := \partial^\mu A^\nu - \partial^\nu A^\mu - 2q(A^\nu \times A^\mu) \quad (1.1.7)$$

The Lagrangian of the strong interaction (1.1.6) has three main terms. The first term describes the free movement of a quark. The second term indicates the eight color currents which act as *sources* for the color field (\vec{A}_μ): it expresses the QCD interaction between a quarks and a gluons.

Finally the last term expresses the free gluon field.

1.2 Electroweak sector

The Electroweak sector emerges from the unification of two interactions: the weak force and the electromagnetic force. The Electroweak sector is based on the $U(1) \times SU(2)$ symmetry, where $U(1)$ refers to the electromagnetic interaction and $SU(2)$ refers to

the weak interaction. The weak force is carried by three massive bosons, W^+ , W^- and Z , whereas the electromagnetic force is carried by the photon. The electroweak sector contains the Higgs boson, the carriers of the weak force, the photon and the families of leptons and quarks. Before discussing the components of the electroweak sector, in the next paragraph I will briefly describe the Electroweak unification [8–11].

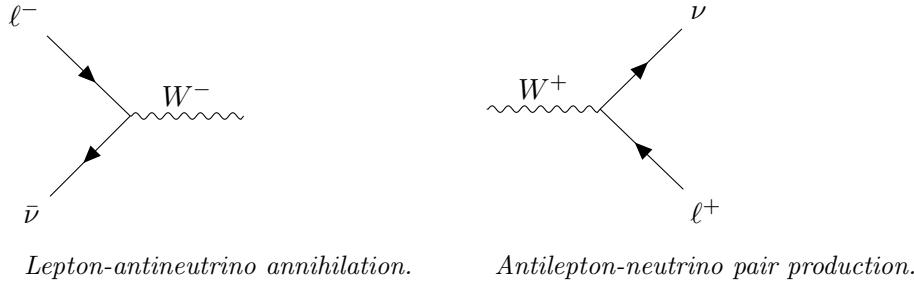


Figure 1.3: Examples of W -boson mediated charge-current interactions.

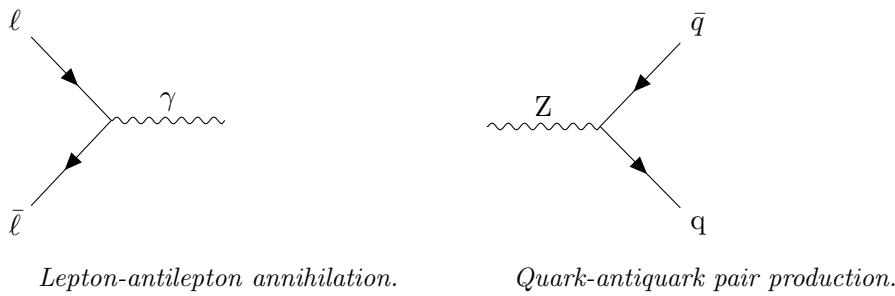


Figure 1.4: Examples of Z -boson and γ -boson mediated neutral-current interactions.

1.2.1 Electroweak unification

The *electroweak* interaction is the unified description of the electromagnetic and weak forces [8–11].

Sheldon Glashow, Abdus Salam, and Steven Weinberg were awarded the 1979 Nobel Prize for their work on the electroweak unification theory.

Their theory asserts that the weak force and the electromagnetic force are two different manifestation of the same interaction: the *electroweak* interaction.

This unification does not take place at ordinary low-level energies, as the two forces merge above the threshold energy of 246 GeV.

The formulation of this unification had to overcome three main apparent obstacles:

1. The *strength* issue: enormous disparity in strength between the two interactions;
2. The *mass* issue: the photon is massless, whereas the W and Z aren't massless;
3. The *current* issue: the electromagnetic current is vectorial (V structure), whereas the weak currents are vectorial-axial (V-A structure);

The first obstacle to unification, the enormous disparity in strength, could be accounted for if the weak mediator was an extremely massive particle.

The second problem can be solved thanks to the *Higgs mechanism*, which explains the existence of two massive bosons, Z and W.

The last issue, the structural difference between the two interactions, would seem to preclude any possibility of unification, but it can be unraveled with the concept of *chirality* and the use of bispinors.

Chirality is the symmetry, which defines the *handedness* of a particle. The solution of the Dirac equation is a spinor made of four components, two of which are associated to a particle and the other two to its antiparticle. The bispinor is a vector made of four components, where the *u* component refers to the particle and *v* component refers to the antiparticle.

$$\psi = \begin{pmatrix} u \\ v \end{pmatrix} \quad (1.2.1)$$

The bispinor notation is extremely powerful, because it allows us to recast the weak and electromagnetic interaction in a form that facilitates their unification. The V-A structure of the weak current can be manipulated to obtain a pure vectorial structure, however this manipulation highlights the fact that weak currents couples only left handed particles, whereas the electromagnetic current couples both left and right handed particles. This manipulation gives a vectorial weak interaction, that can be more easily blended with the electromagnetic interaction. However before combining the two forces it is necessary to analyze the weak currents. The charged weak currents can be compactly expressed by introducing the *chiral weak isospin doublet*:

$$\chi_L^f = \begin{pmatrix} \nu_f \\ l_f \end{pmatrix} \quad (1.2.2)$$

where f indicates the generation (family) and L the left handedness. The compact form of the positive and negative weak currents is shown in 1.2.3.

$$J^\pm = \bar{\chi}_L \gamma_\mu \tau^\pm \chi_L \quad (1.2.3)$$

where τ^\pm are linear combinations of the first two Pauli spin matrices. The final step before mixing the weak and electromagnetic interactions is the completion of the *weak isospin* symmetry with a neutral current. The weak neutral current cannot be introduced using the third Pauli's matrix, because it would give a current that couples both right and left handed particles. The weak neutral current is obtained from the Gell-Mann-Nishijima formula (1.2.4), which connects the *weak isospin* and the *weak hypercharge* to the electric charge:

$$Q = \frac{Y}{2} + I_3 \quad (1.2.4)$$

where Q is the electric charge, Y is the weak hypercharge and I_3 is the third component of the weak isospin. The Gell-Mann-Nishijima formula is the paradigm used to introduce the weak neutral current (J_μ^Y) and it sets off the connection between the electromagnetic and weak interactions:

$$J_\mu^Y := 2J_\mu^{em} - 2J_{mu}^3 \quad (1.2.5)$$

where J_μ^{em} is the electromagnetic current and J_{mu}^3 is the neutral current obtained with the third Pauli matrix. The neutral current represents the *weak hypercharge current*. Finally, after defining the three weak currents, it is possible to merge the weak and electromagnetic interactions. The unification theory of Glashow, Salam and Weinberg asserts that the three isospin currents (J^\pm and J^3) couples to a isotriplet of bosons with strength g_W and the hypercharge current couples with a isosinglet boson, with strength g .

$$\mathcal{L}_{EW} = g_W \vec{J}_\mu \vec{W}_\mu + \frac{g}{2} J_\mu^Y B^\mu \quad (1.2.6)$$

The isotriplet contains the two real W^\pm bosons. The linear combination of the third boson of the isotriplet and the isosinglet boson provides the photon and the Z boson. The electroweak Lagrangian is the following:

$$\mathcal{L}_{EW} = \frac{g_W}{\sqrt{2}} (J_\mu^- W^\mu + J_\mu^+ W^\mu) + \frac{g_W}{\cos\theta_W} (J_\mu^3 - \sin^2\theta_W J_\mu^{em}) Z^\mu + (g_W \sin\theta_W) J_\mu^{em} A_\mu \quad (1.2.7)$$

where θ_W is the Weinberg's angle. The first and the third term are respectively purely weak and purely electromagnetic, whereas the second is the mixing term. This lagrangian does not describe the mass of the involved particles, as that is added by the Higgs Mechanism.

1.2.2 Leptons

The electroweak sector organizes leptons in left-handed isospin doublets and right-handed isospin singlets:

$$\chi_L^f = \begin{pmatrix} \nu_f \\ \ell_f \end{pmatrix} \quad \chi_R^f = \ell_f \quad (1.2.8)$$

where f is the generation number and L/R indicates the handedness.

1.2.3 Quarks

Similarly to leptons, quarks are organized in right-handed isospin singlets and left-handed isospin doublets:

$$\zeta_L^f = \begin{pmatrix} U_f \\ D'_f \end{pmatrix} \quad \zeta_R^f = U_f \quad \zeta_R^f = D_f \quad (1.2.9)$$

where f is the generation number, U is an *Up-type* quark and D' is a *Down-type* quark that has been rotated in flavour space.

$$D_{f'} = \sum_{f'} V_{ff'} D_{f'} \quad (1.2.10)$$

The rotation is expressed by the Cabibbo-Kobayashi-Masukawa (CKM) matrix [7]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \simeq \begin{pmatrix} 0.97373 & 0.2243 & 0.00382 \\ 0.221 & 0.975 & 0.0408 \\ 0.0086 & 0.0415 & 1.014 \end{pmatrix} \quad (1.2.11)$$

where V_{AB} expresses the coupling strength between quarks A and B. The CKM matrix is a unitary matrix which contains information on the strength of the flavour-changing weak interaction. Since the highest values are on the matrix' diagonal, weak interactions tend to change flavour within the same (strong) generation.

1.2.4 Electroweak bosons

The electroweak bosons are three: Z, W and γ . The photon (γ) is massless, electrically neutral and mediates solely neutral-current interactions. The Z boson is massive and, as the photon, mediates only neutral-current interactions. Finally, the W boson is the heaviest, electrically charged and mediates the charge-current interactions.

1.3 The Higgs Mechanism

The Standard Model described thus far does not contemplate the description of leptons, quarks or bosons' masses without violating gauge local invariance, which must be preserved since it allows to introduce the two fundamental interactions (strong and electroweak) within the Dirac's lagrangian.

The *Higgs Mechanism*, proposed by Peter Higgs along with R. Brout and F. Englert [12–14], provides a way for the particles of the SM to acquire mass without violating the gauge local invariance via a process known as *spontaneous symmetry breaking*.

The *Higgs Mechanism* is based on a new particle, the *Higgs boson*, which was theorized in 1964, but was discovered after a 50 years search in 2012 at LHC at CERN [15, 16]. The *Higgs boson*, is a massive (125.25 ± 0.17 GeV [7]) scalar boson.

1.3.1 Spontaneous symmetry breaking

The Higgs field is a doublet of complex-valued scalar fields:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.3.1)$$

where:

$$\phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \quad \phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}} \quad (1.3.2)$$

Then, it is possible to introduce the Higgs Lagrangian, which allows the Z and W bosons to acquire mass within the SM:

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \quad (1.3.3)$$

where:

$$D_\mu = \partial_\mu + i \frac{q}{\hbar c} A_\mu \quad (1.3.4)$$

The Higgs lagrangian (1.3.3) has a potential, the *Higgs potential*, which is characterized by two positive constants (μ^2 , λ). The *Higgs potential* (1.3.5) is $SU(2) \times U(1)$ symmetric.

$$V(\Phi^\dagger \Phi) = -\mu^2 (\Phi^\dagger \Phi) + \lambda (\Phi^\dagger \Phi)^2 \quad (1.3.5)$$

The Higgs field $\Phi = 0$ is not the ground state of the Higgs potential. The ground state of the Higgs potential forms a ring of degenerate minima where the Higgs field has a negative constant value.

$$\Phi^\dagger \Phi = \frac{1}{2} (\phi^+ + \phi^0) = \frac{\mu^2}{2\lambda} \quad (1.3.6)$$

The Higgs lagrangian (1.3.3) presents an even symmetry ($\Phi \rightarrow -\Phi$), which is *broken* by choosing one of the infinite minima of the Higgs potential.

The expression *spontaneous symmetry breaking* refers to the act of picking one minimum from the set of degenerate minima of V to act as the *vacuum state*.

The *Goldstone theorem* affirms that the breaking of a continuous symmetry is always accompanied by the appearance of one or more Goldstone bosons (massless scalar boson). However it is possible to astutely choose a minimum in order to eliminate the Goldstone fields, which are said to be “eaten” by the W and Z bosons, allowing them to acquire mass through the Higgs field.

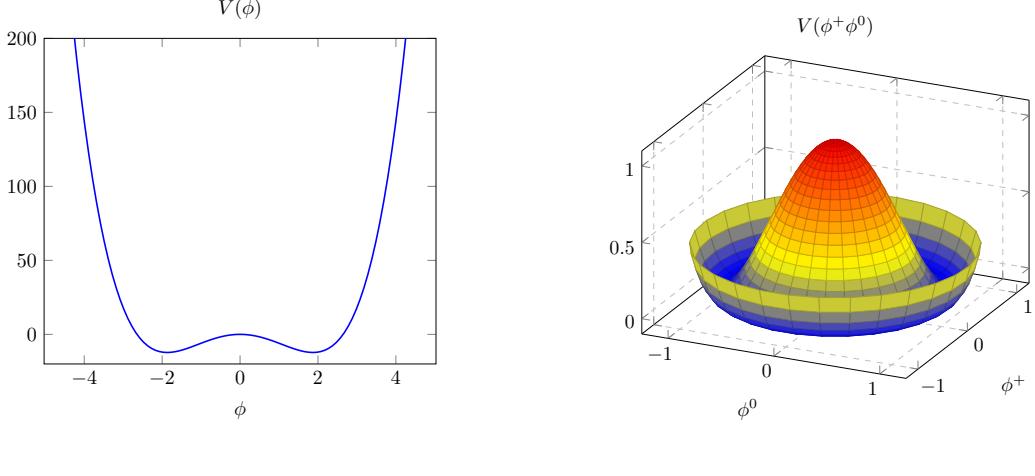


Figure 1.5: Comparison between the 1D and 2D Higgs potentials.

1.3.2 Yukawa term

The masses of quarks and charged leptons (neutrinos are massless within the SM) are introduced in the SM through the *Yukawa coupling*:

$$\mathcal{L}_{Yukawa} = -\alpha_\ell [(\bar{\chi}_L \phi) \chi_R + \bar{\chi}_R (\phi^+ \chi_L)] - \alpha_q [(\bar{\zeta}_L \phi) \zeta_R + \bar{\zeta}_R (\phi^+ \zeta_L)] \quad (1.3.7)$$

where: α_ℓ and α_q provide the mass terms.

1.4 The Lagrangian of the Standard Model

The Lagrangian of the Standard Model has three main terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{Interactions} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \quad (1.4.1)$$

The first term describes the strong (S) and electroweak (E) interactions without taking into account the mass term of fermions and bosons.

This term can be obtained by considering the non-interacting SM Lagrangian:

$$\mathcal{L}_{non-int} = \mathcal{L}_{leptons} + \mathcal{L}_{quarks} = \bar{\chi}_L^f i\gamma^\mu \partial_\mu \chi_L^f + \bar{\chi}_R^f i\gamma^\mu \partial_\mu \chi_R^f + \bar{\zeta}_L^f i\gamma^\mu \partial_\mu \zeta_L^f + \bar{\zeta}_R^f i\gamma^\mu \partial_\mu \zeta_R^f \quad (1.4.2)$$

and then requiring the local gauge invariance of the type $U(1) \times SU(2) \times SU(3)$. The second term of equation 1.4.1, the Higgs Lagrangian, is responsible for the mass of the W and Z bosons. Finally, the last term of equation 1.4.1 is the Yukawa Lagrangian, which takes care of the mass of the fermions, both quarks and leptons.

Chapter 2

Top quark

The top quark was theorized to complete the weak isospin doublet containing the *bottom* (or *beauty*) quark and it was discovered at Tevatron in 1995 [[Abe 1995, 17](#)]. The study of the top quark is essential in particle physics. Since its mass is close to the scale of electroweak symmetry breaking it could help shedding a light over this mechanism. Moreover many SM extensions have final states, where the top quark appears as a background source, hence it is important to deeply understand top production processes.

2.1 Top quark properties and decay

The *top* quark is the heaviest particle of the SM ($m = 172.69 \pm 0.30$ GeV [[7](#)]) with a coupling to the Higgs boson close to 1. It is an up-type quark and it belongs to the heavy flavour family alongside the charm and the bottom quark.

Since it has a short lifetime ($\sim 10^{-25}$ s) it decays before hadronization ($\sim 10^{-24}$ s). This unique feature among quarks allowed physicists to measure its mass directly, unlike every other quark. This peculiar property doesn't allow the top quark to form bound states. Therefore an hadron containing the top quark doesn't exist.

The CKM matrix shows that the top quark interacts almost exclusively with the bottom quarks.

$$|V_{td}| = 0.0086 \pm 0.0002 \quad |V_{ts}| = 0.0415 \pm 0.0009 \quad |V_{tb}| = 1.014 \pm 0.0029$$

The top quark mainly decays in the W boson and bottom quark.

$$t \rightarrow Wb \tag{2.1.1}$$

The decay of the W boson defines two different final states for the top quark decay: *leptonic* and *hadronic* (figure 2.1).

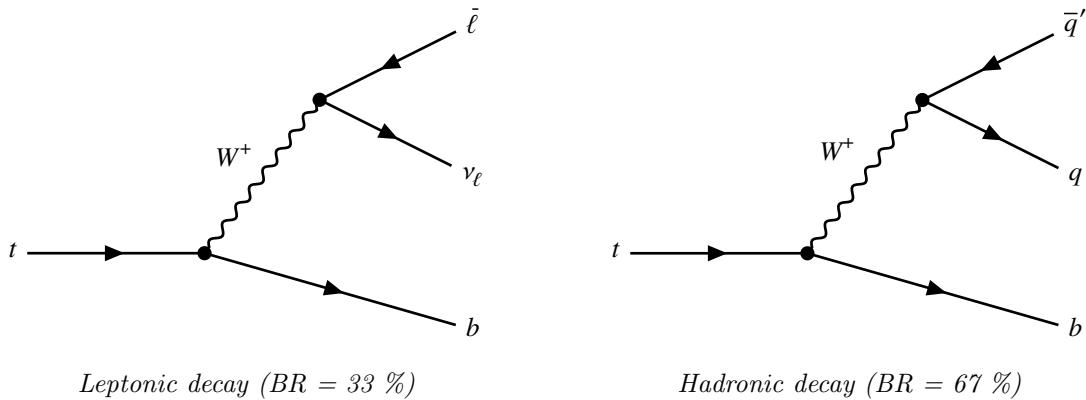


Figure 2.1: The Feynman diagrams of the two possible decays of the top quark

The *hadronic* final state has the largest branching ratio (BR = 67 %) and it is characterized by a quark and an anti-quark. This final state produces three jets, which makes non trivial to distinguish this decay from a large multijet background. The *leptonic* final state is characterized by a charged anti-lepton and its neutrino.

2.2 Top quark production

In hadron collisions, the dominant process for production of top quarks is $p\bar{p} \rightarrow t\bar{t}$, which is a *strong* process called *pair production* (figure 2.2). This process is caused mainly by *quark-antiquark annihilation* or *gluon-gluon fusion*. The relative importance of those two processes depends on the centre of mass energy. The top quark production was dominated by $q\bar{q}$ annihilation at Tevatron, because there the collisions were $p\bar{p}$, whereas at LHC the dominant process is gluon-gluon fusion. Since $t\bar{t}$ is a strong process, parity is conserved, hence top quarks are produced without polarization.

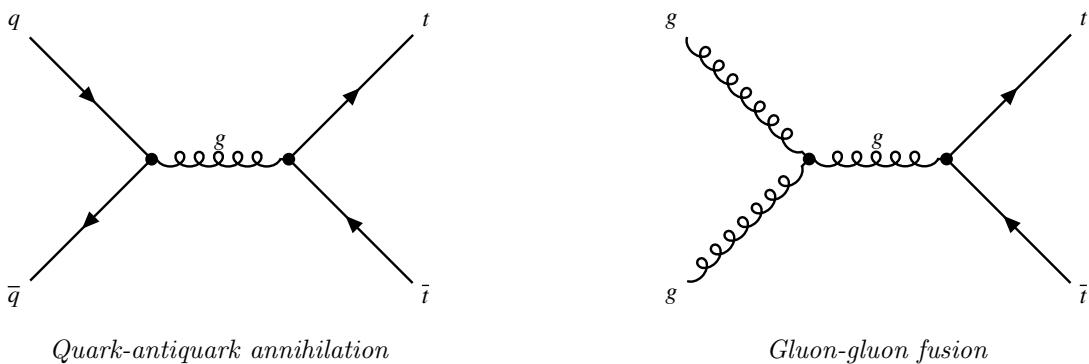


Figure 2.2: The Feynman diagrams of the production of the top quark by $q\bar{q}$ annihilation and gg fusion

The other main production channel for top quarks is the *single top quark production* via weak process (figure 2.3). Since it is a weak process, parity is not conserved, hence

top quarks are produced polarized. This process can occur through three different sub-processes:

1. t-channel: its cross section at $\sqrt{s} = 13$ TeV is $\sigma = 217^{+0.9}_{-0.8}$ pb [18]. It is the most dominant process for single top production (figure 2.3a);
2. tW-channel: its cross section at $\sqrt{s} = 13$ TeV is $\sigma = 71.7 \pm 3.8$ pb [18]. It involves the production of a real W boson by the s-channel exchange of a virtual b-quark (figure 2.3b);
3. s-channel: its cross section at $\sqrt{s} = 13$ TeV is $\sigma = 10.3 \pm 0.4$ pb [18]. It is the least dominant process out of the three (figure 2.3c);

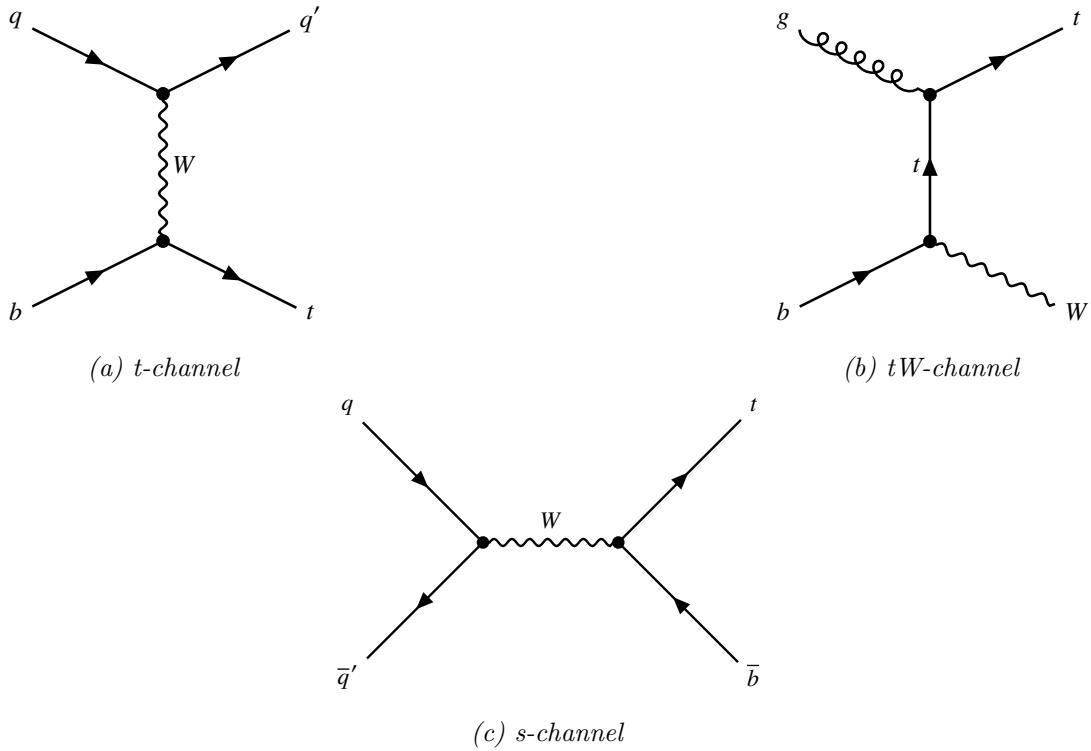


Figure 2.3: Feynman diagrams for the single top quark production

2.2.1 Top quark production in association with a Z boson

In the previous paragraph I described the two main production channels for the top quark. However many other rare top quark production channels exist ($t\bar{t}H$, $t\bar{t}Z$, $t\bar{t}t\bar{t}$, ...) and some of them include the Z boson.

My thesis focuses on the tZq process where a single top quark is produced in association with a Z boson (figure 5.1). It is an electroweak process observed by the ATLAS and CMS experiments at $\sqrt{s} = 13$ TeV with an integrated luminosity of 139 fb^{-1} . Many

different tZq final states are possible, however I will focus on the three-lepton final state, since it is a really clean channel: it has a good signal-background ratio. The three leptons in the final state are produced by the leptonic decay of the Z (2) and the W (1) boson.

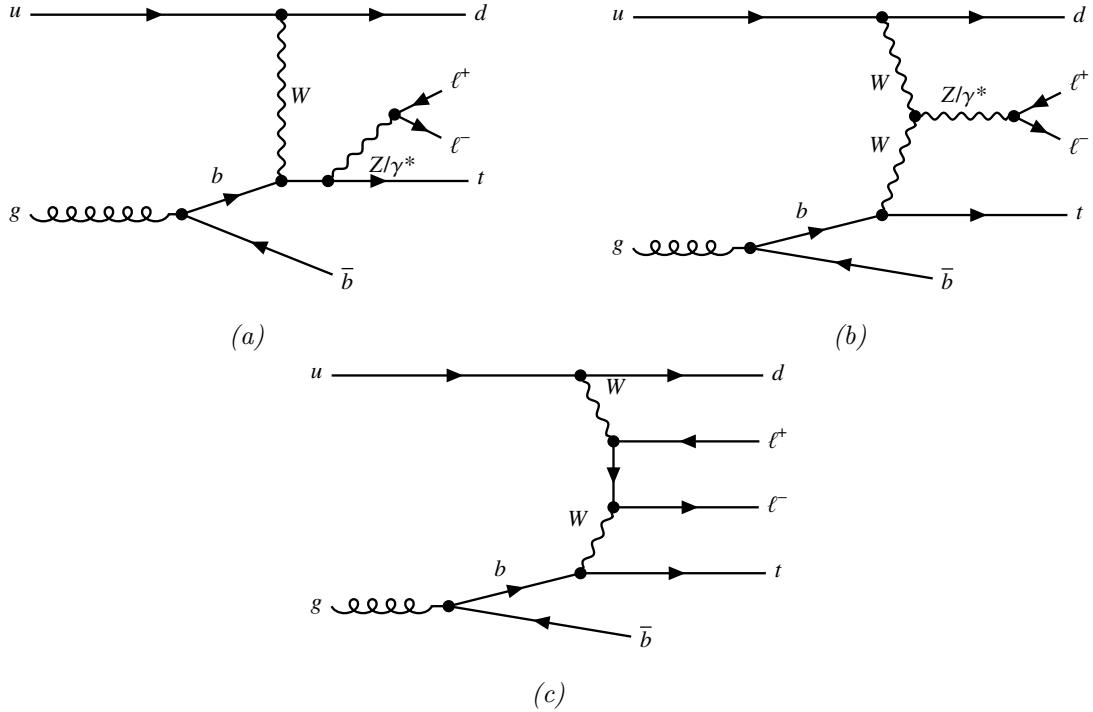
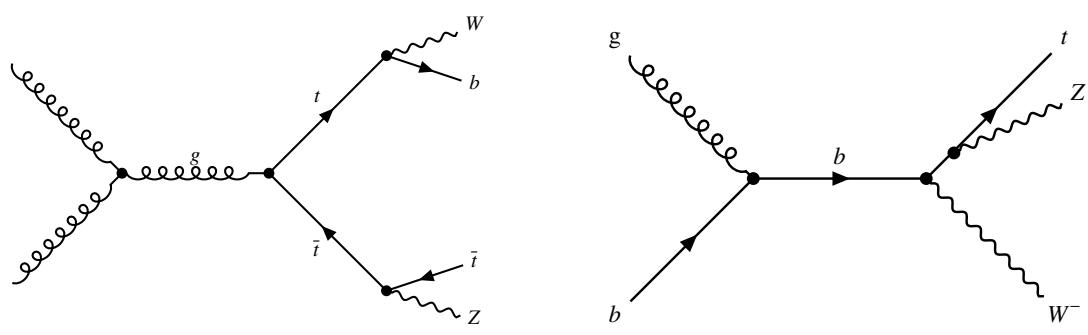


Figure 2.4: Example Feynman diagrams of the lowest-order amplitudes for the tZq process, corresponding to (a, b) resonant $\ell^+\ell^-$ production and (c) non-resonant $\ell^+\ell^-$ production. In the four-flavour scheme, the b-quark originates from gluon splitting.

Moreover, since my thesis focuses on the tZq event where the top quark is produced in association with a Z boson, it is essential to describe every top quark production channel where a Z boson appears (figure 2.5). The most common top and Z boson associated production channels are:

- $t\bar{t}Z$: its cross section at $\sqrt{s} = 13$ TeV and integrated luminosity of 139 fb^{-1} is $(\sigma = 5.273 \pm 0.004) \times 10^{-1} \text{ pb}$ [19].
- tWZ : this process has not been measured yet but its predicted cross section at $\sqrt{s} = 13$ TeV is $\sigma \simeq 15.6 \text{ fb}$ [20].



Feynman diagram for the $t\bar{t}Z$ production.

Feynman diagram for the tWZ production.

Figure 2.5: Feynman diagrams for two single top quark productions in association with a Z boson.

Chapter 3

The Large Hadron Collider and the ATLAS Experiment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [21] is the world's largest and most powerful particle accelerator. The LHC is a circular collider with a circumference of 27 km located near Geneva on the French-Swiss border. It has been active since 2008 as the latest addition to CERN's accelerator complex in Geneva and has continued to run since then, interspersed with shutdown periods for maintenance and upgrades. Thus far it has completed two long periods of operation (Run-I and Run-II), and the third has just started last July (3.1). This third run will be the last operation before an extensive upgrade to reach 3000 fb^{-1} luminosity.

Parameters	Run-I	Run-II	Run-III	HL-LHC
Centre of Mass Energy	7 TeV, 8 TeV	13 TeV	13.6 TeV	14 TeV
Activity period	2009-2012	2015-2018	2022-2025	2029-2040

Table 3.1: Centre of mass energy and period of activity for each operation [22–24].

The CERN's complex accelerates two proton beams in opposite directions through five accelerators: LINAC4 (Linear Accelerator), PSB (Proton Synchrotron Booster), PS (Proton Synchrotron), SPS (Super Proton Synchrotron) and LHC (Large Hadron Collider). These beams then collide in four different locations, where four experiments are located:

1. ATLAS (*A Toroidal LHC Apparatus*) [25];
2. CMS (*Compact Muon Solenoid*) [26];
3. ALICE (*A Large Ion Collider Experiment*) [27];

4. LHCb (*LHC beauty*) [28];

The ATLAS and CMS are multifunctional detectors which aim to improve our understanding of the Standard Model and search for physics beyond the Standard Model (e.g. SUSY).

The ALICE detector studies the properties of the Quark-Gluon plasma (QGP) produced in heavy-ion-collisions, whereas the LHCb experiment focuses on heavy flavour physics.

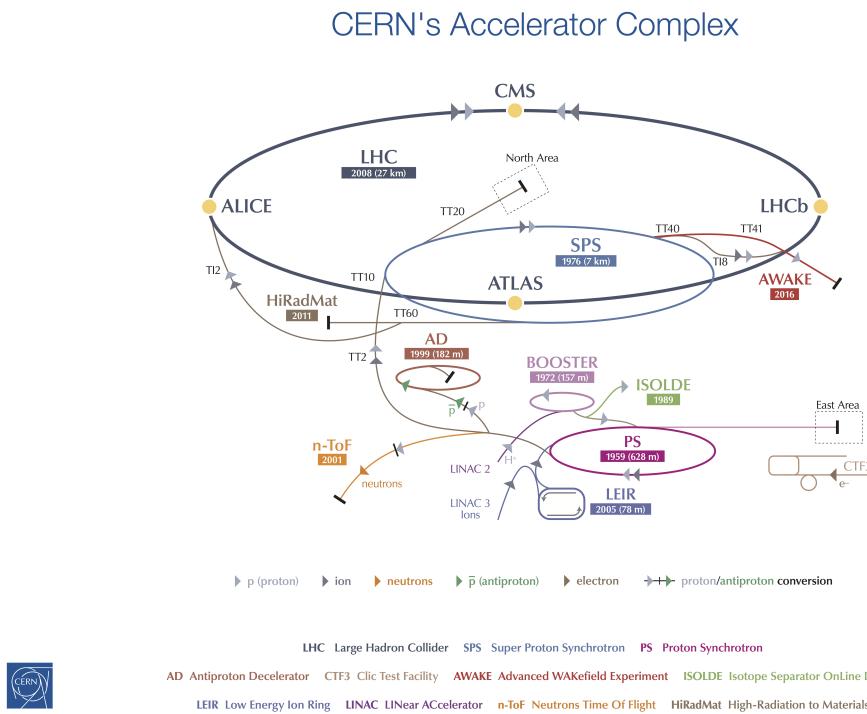


Figure 3.1: Accelerator complex at CERN.

3.1.1 Accelerator structure and fundamental properties

The accelerator complex at CERN is a succession of machines that accelerate protons to increasingly higher energies (figure 3.1). Each machine boosts the energy of a beam of protons before injecting it into the next machine in the sequence.

The proton beams, which are not a continuous stream but are rather made of *bunches* with $\sim 10^{11}$ protons per unit with a spacing between each group of $\sim \text{ns}$, are accelerated through five main stages.

The first step is the acceleration of negative hydrogen ions (a hydrogen atom with an additional electron) to 160 MeV by Linac4 [29], which replaced Linac2 after the long shutdown 2019–21. Subsequently the ions are stripped of their two electrons during injection from Linac4 into the PSB, leaving only protons. The PSB accelerates them

up to 2 GeV [30] before the injection into the PS, which pushes the beam up to 26 GeV [31]. Protons are then sent to the SPS, where they are accelerated up to 450 GeV [32]. Finally the protons are transferred to the pipes of LHC, where they reach their maximum energy of 6.5 TeV in 20 minutes [33].

The LHC has a complicated structure made of 9593 superconducting magnets: 1232 dipoles to deflect the beams, 392 quadrupoles to focalize the beams, sextupoles to control the beams chromaticity, decapoles and dodecapoles. Moreover the accelerator is characterized by eight radio-frequencies ($\nu = 400.8$ MHz) cavities which accelerate the beams.

For a particular process the number of events generated per second (N) at the LHC is:

$$N = L\sigma \quad (3.1.1)$$

where L is the *instantaneous* luminosity (number of collisions per unit of time and transverse section of the beams) and σ is the cross section for the process. For circular colliders which accelerate k_b bunches of N particles at a revolution frequency f, the *instantaneous* luminosity L is:

$$L = \frac{f\gamma}{4\pi\epsilon_n} N^2 k_b \frac{F(\beta^*)}{\beta^*} \quad (3.1.2)$$

where F and β^* adjust the collision area according to the angle of beam crossing, and ϵ_n accounts of the spread of particles in position-momentum space. The *integrated* luminosity is the integral over time of the *instantaneous* luminosity. Figure 3.2 compares the cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018.

Finally it is important to highlight the *pileup* phenomenon. The main quest of ATLAS is to investigate the rarest interactions of proton-proton collisions. To maximize chances for sporadic interactions, the LHC doesn't simply collide single protons, but rather large bunches of protons, such that multiple protons interact when the bunches collide. The particles from the interaction of interest are thus recorded together with particles from multiple additional interactions, so-called *pileup* interactions. Increasing the number of protons per bunch maximizes the number of interesting collisions, however it also increase the *pileup* interference (trade-off). The *pileup* interference also increases with luminosity, because the rate of the events is higher. Table 3.2 sums up the values of the main accelerator parameters used during Run-I and Run-II, and the expected parameters for Run-III and the HL-LHC. Run-II shows a significant peak luminosity improvement with respect to Run-I.

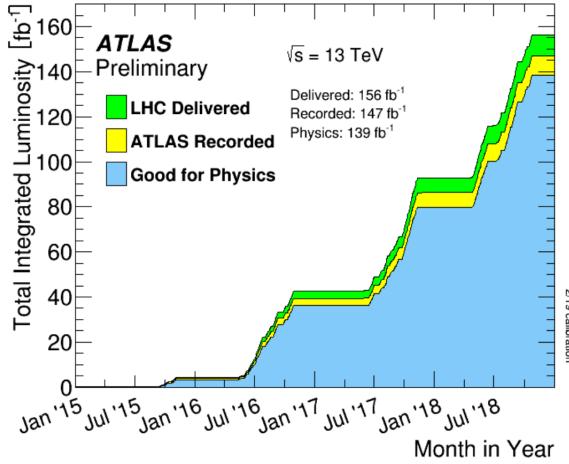


Figure 3.2: Integrated luminosity delivered to ATLAS (green), recorded (yellow), and certified to be good quality data (blue) during 13 TeV collisions in the period 2015–2018.

Parameters	Run-I	Run-II	Run-III	HL-LHC
Bunch spacing [ns]	50	25	25	25
Bunch Intensity [10^{11} ppb]	1.6	1.2	up to 1.8	2.2
Number of bunches	1400	2500	2800	2800
Peak Luminosity [10^{34} cm $^{-2}$ s $^{-1}$]	0.8	2.0	2.0	5.0
Peak pileup	45	60	55	150

Table 3.2: Summary of main accelerator parameters used during Run-I and Run-II, and the expected parameters for Run-III and the HL-LHC [34].

3.1.2 ATLAS coordinate system

The coordinate system adopted by ATLAS has the origin at the centre of the detector (the nominal interaction point). The x-axis is defined perpendicular to the beam pointing to the centre of the LHC, the y-axis is also perpendicular to the beam and points upwards. The resulting x-y plane is regarded as the *transverse plane*. Finally, the z-axis is parallel to the beam direction.

However the cylindrical symmetry of the detector suggests the usage of cylindric coordinates (r, ϕ, θ):

- Radial distance from the nominal interaction point: $r = \sqrt{x^2 + y^2}$;
- Azimuthal angle measured from the z-axis: $\phi = \text{atan}(\frac{y}{z})$;
- Polar angle measured from the z-axis: $\theta = \text{cot}(\frac{x}{y})$;

Moreover some useful quantities are introduced to better describe the events produced by the collision:

- Transverse momentum of a particle: $p_T = p \sin\theta$;
- Transverse energy of a particle: $E_T = E \sin\theta$;
- The *pseudorapidity* $\eta = -\ln(\tan(\frac{\theta}{2}))$ is often used instead of the polar angle (θ), because $\Delta\eta$ has the advantage of being Lorentz invariant;
- Angular distances are often measured in units of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$;

It is worthy to highlight that p_T and E_T are more interesting variables than p and E . The colliding partons have transverse component negligible in the initial state ($p_T^i = 0$ and $E_T^i = 0$), therefore this brings a constrain on the total transverse momentum and energy of the final states.

3.2 The ATLAS detector

The ATLAS detector [25] at the LHC is the largest particle detector ever built for a collider experiment. It has the shape of a cylinder with a diameter of 25 m and a length of 44 m. It has a total weight of around 7000 tonnes and it is placed 100 meters below ground. The detector covers nearly the entire solid angle around the collision point.

It comprises four main layers, moving inside out we find:

- Inner Detector (ID);
- Electromagnetic Calorimeter (ECAL);
- Hadronic Calorimeter (HCAL);
- Muon Spectrometer;

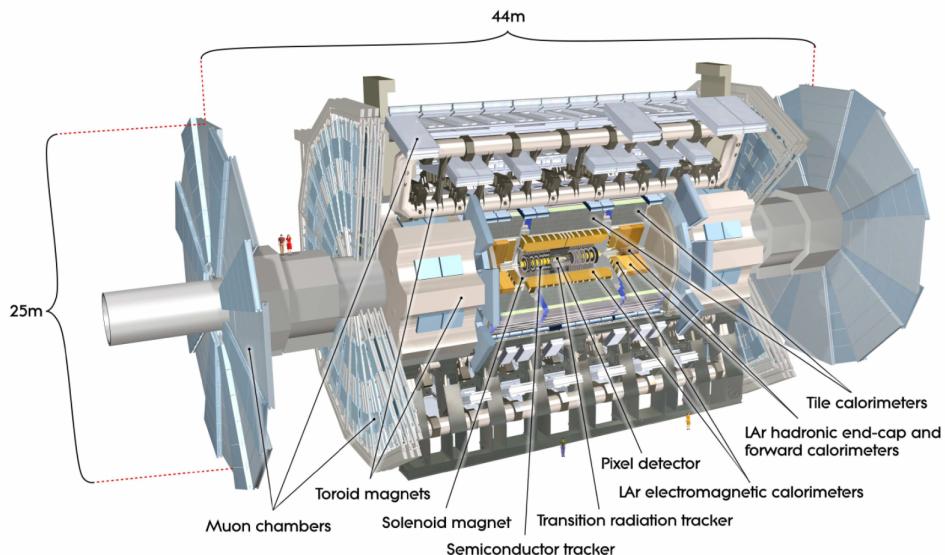


Figure 3.3: Cut away view of the ATLAS detector [25].

3.2.1 Inner Detector

The ATLAS Inner Detector (ID) system is immersed in a 2 T axial magnetic field and provides charged particle tracking in the range $|\eta| < 2.5$. The ID system provides essential information for the reconstruction of physics objects such as electrons, muons, τ -leptons and jets, as well as for identification of jets containing b-hadrons, and for event-level quantities that use charged-particle tracks as input. It comprises three detector types (figure 3.4), moving inside out we find:

- the Silicon Pixel Detector. It is a high-granularity silicon pixel detector which contributes to the accurate measurement of vertices. It provides four measurements per track;
- the SemiConductor Tracker (SCT). It contributes to measure precisely the particle momentum. It provides eight measurements per track;
- the Transition Radiation Tracker (TRT). It contributes to ease the pattern recognition, while also contributing to electron identification;

During the first long shutdown, the Insertable B-Layer (IBL) was constructed, inserted and commissioned to become an additional (innermost) layer of the existing Pixel Detector. The IBL helps the Pixel detector to measure vertices. These subsystems are essential for the reconstruction of charged-particle tracks, the accuracy of which is limited by the finite resolution of the detector elements and by an incomplete knowledge of their positions.

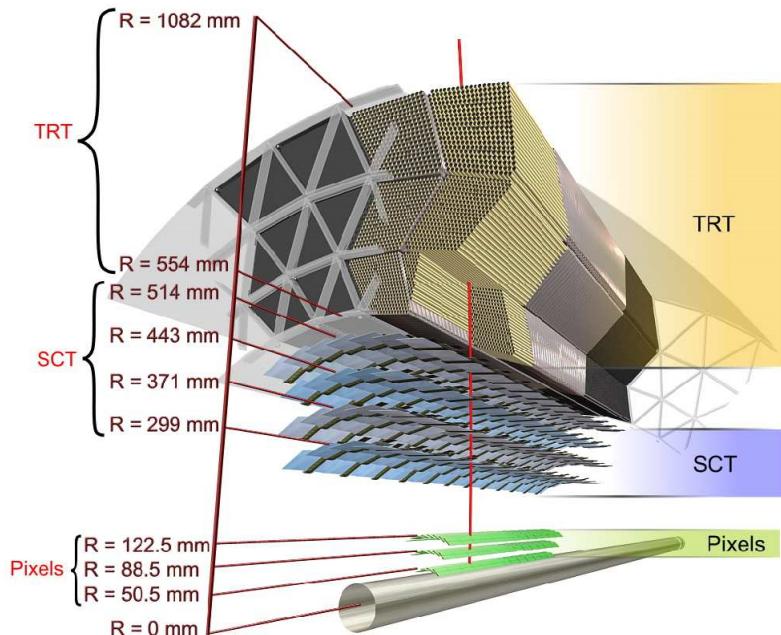


Figure 3.4: Cross section of the Inner Detector.

3.2.2 Calorimetry system

The ATLAS detector has a calorimetry system made of two different types of calorimeters (figure 3.5): the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). They are both designed to absorb most of the particles coming from a collision, forcing them to deposit all of their energy and stop within the detector. Therefore, ATLAS calorimeters are composed by an alternation of layers of *active* and *passive* material. The *passive* (or *absorbing*) layers are made of high density material, that aims to slow down and eventually stop incoming particles, whereas the *active* layers contribute to measure the energy of the particles. The interactions between high energy particles and the detector material lead to particle showers which can be measured up to a pseudorapidity of $|\eta| < 4.9$. However, muons and neutrinos pass the calorimeters without being measured.

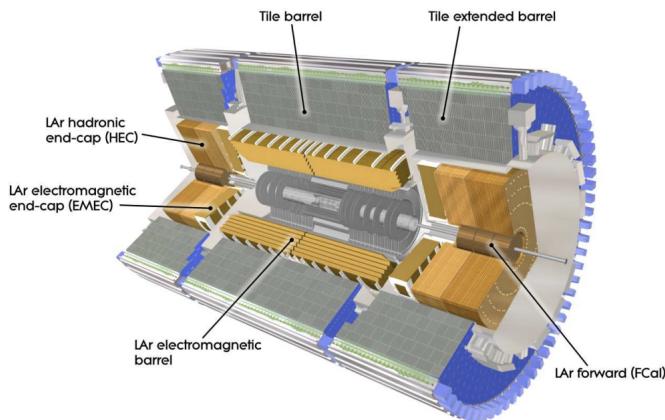


Figure 3.5: Cut away view of the ATLAS calorimeter system.

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) measures the energy of electrons and photons as they interact with matter. The ECAL has a barrel part, which covers the region $|\eta| < 1.475$, and has two end cap components, which cover $1.375 < |\eta| < 3.2$. In the part of the barrel region, a fine granularity is chosen to perform precision measurements of electrons and photons while the rest of the electromagnetic calorimeter features a coarser granularity as this area is mainly used for jet reconstruction and measurements of missing transverse momentum. The barrel calorimeter is divided in two identical half-barrels, separated by a small gap (4 mm) at $z = 0$, whereas each end cap calorimeter is divided into two coaxial wheels: an outer wheel which covers the region $1.375 < |\eta| < 2.5$, and an inner wheel which covers the region $2.5 < |\eta| < 3.2$. The EM calorimeter uses the lead plates as the *absorbing* medium, whereas liquid argon (LAr) at the temperature of -183 °C acts as the *active* medium, which gets ionized by the particle showers.

Hadronic Calorimeter

Hadronic calorimeters (HCAL) sample the energy of hadrons as they interact with atomic nuclei. The HCAL consists in three main components: the tile calorimeter, which is placed outside the ECAL envelope, the LAr hadronic end cap calorimeter (HEC), which is located directly behind the end cap electromagnetic calorimeter, and the LAr forward calorimeter (FCal), which is integrated into the end cap cryostats. The first component, the tile calorimeter, is a sampling calorimeter which uses steel as the *absorbing* material and scintillating tiles as the *active* material. Its barrel covers the region $|\eta| < 1$, and its two extended barrels cover $0.8 < |\eta| < 1.7$. The HEC consists of two independent wheels per end cap and each wheel is divided into two segments in depth, for a total of four layers per end cap. These wheels are composed of interleaved layers of copper, that acts as the *absorbing* medium, and layers of liquid argon, that acts as the *active* medium. The HEC covering region slightly overlaps with that of the tile calorimeter ($|\eta| < 1.7$) and the forward calorimeter (around $|\eta| = 3.1$): $1.5 < \eta < 3.2$. Finally the FCal reduces the radiation background levels for the muon spectrometer and provides an uniform calorimetric coverage. It consists in three main modules in each end-cap: the first (copper) is optimised for electromagnetic measurements, while the other two (tungsten) measure predominantly the energy of hadronic interactions. Each module has a metal matrix, with regularly spaced longitudinal channels filled with concentric rods and tubes parallel to the beam axis. The liquid argon fills the gaps between rods and tubes and acts as the *absorbing* medium.

3.2.3 Muon Spectrometer

Muons do interact electromagnetically, however since they are much heavier than electrons, their energy loss due to bremsstrahlung and ionization is much smaller than for electrons. Therefore they penetrate the electromagnetic calorimeter without losing much energy. Furthermore, muons do not interact via the strong force, hence they also penetrate the hadronic calorimeter without losing much energy. Therefore, the muon's properties (coordinates, charge and momentum) are measured in the muon spectrometer (MS) located around the calorimetry system. The MS comprises separate trigger and high-precision tracking chambers with four different detector subsystems and it is immersed in a magnetic field generated by superconducting toroids, which ranges between 2.0 and 6.0 T across most of the detector. The muon trigger system has two main subsystems: the Resistive-Plate Chambers (RPC) in the barrel, and the Thin-Gap Chambers (TGC) in the endcap regions. The high-precision tracking system has also two main subsystems: the three layered Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC) in the forward region. The muon trigger systems covers $\eta < 2.4$, whereas the high-precision tracking system covers $\eta < 2.7$.

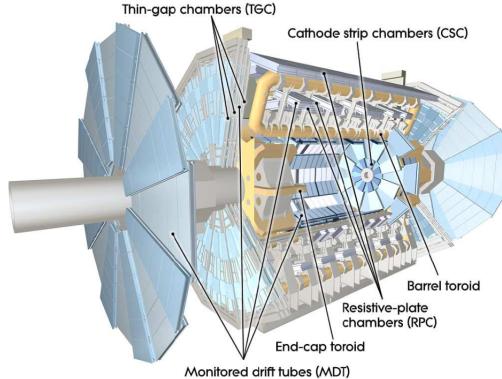


Figure 3.6: Cross section of the muon spectrometer.

3.3 Object Reconstruction

This paragraph is meant to describe the reconstruction of the basic objects used in the analysis.

3.3.1 Primary vertex

The primary vertex [35] of the interaction is selected as the pp vertex candidate with the highest sum of the p_T^2 of all the associated tracks with $p_T > 500$ MeV.

3.3.2 Electrons

Electron candidates are reconstructed from energy deposits in the EM calorimeter associated with ID tracks [36–38].

The electrons have two requirements: $|\eta| < 2.47$ and $E_T > 10$ GeV. However the transition region between the barrel and end-cap EM calorimeters is excluded ($1.37 < |\eta| < 1.52$). The electrons are identified through three different Identification Working Points (IDWPs), which are in order of decreasing background rejection power: **TightLH**, **MediumLH** or **LooseLH**. A Working Point is a set of cuts and selections which gives a certain efficiency on the identification or isolation of an object. For instance, to give an idea of the different efficiency of the three IDWPs, the identification efficiency of a prompt electron with $E_T = 40$ GeV is 93% for Loose, 88% for Medium and 80% for Tight [39]. In this analysis all electron candidates must pass the **TightLH** selection and are required to be isolated, using the **PLVLoose** working point, which has a p_T dependency.

3.3.3 Muons

Muon candidates can be reconstructed both using only the Muon Spectrometer information or combining the MS reconstruction with Inner Detector tracks.

Similar to the electrons, muons are identified through three different working points: **Loose**, **Medium**, and **Tight**. In this analysis the muon candidates must pass the **Medium** selection: $|\eta| < 2.5$ and $p_T > 10$ GeV. The isolation requirement is the same as that of electrons (**PLVLoose** working point).

3.3.4 Jets

Jets reconstruction in the calorimeter starts from grouping together energy deposits in adjacent cells of the calorimeter, creating a topocluster. They are required to have $p_T > 35$ GeV and $|\eta| < 4.5$. All jets with $|\eta| > 2.5$. are required to pass the requirements of the **fJVT Medium** working point. This has an efficiency up to 97% of selecting hard scattered jets and a pileup-jet efficiency of 53.4% for jets with p_T between 40 GeV and 50 GeV [40].

3.3.5 b-jet

The top quark decay almost exclusively in a W boson and a bottom quark. The bottom quark generates a jet of hadrons because of confinement. Hadrons containing bottom quarks have sufficient lifetime ($\tau \sim 10^{-12}$) that they travel some distance before decaying. On the other hand, their lifetimes are not so high as those of light quark hadrons, so they decay inside the detector rather than escape. Therefore a b-jet is characterized by a secondary vertex, which is where the b-hadrons decay.

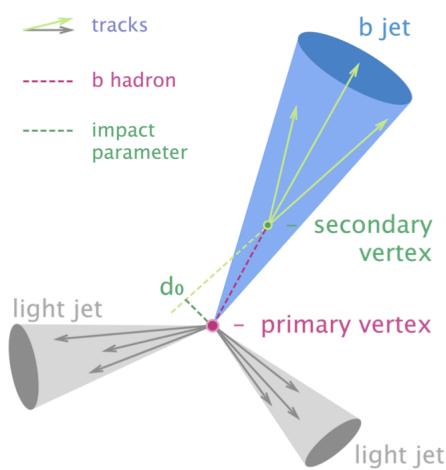


Figure 3.7: Schematic view of a b-hadron decay in a b-jet compared to light jets.

The identification of b-jets is based on several objects reconstructed in the ATLAS detector, mainly the tracks and the vertices. Jet containing a b-hadron are identified using the **MV2c** algorithm. I worked on two selections that were based on two different b-jet efficiency (table 3.3): the first used **eff_70** (*default* selection) and the second used **eff_85** (*loose* selection).

Name	b-jet efficiency (%)	c-jet R	light-jet R	τ -jet R
<code>eff_60</code>	60.61	27	1495	314
<code>eff_70</code>	70.56	9	429	61
<code>eff_77</code>	77.50	4	141	17
<code>eff_85</code>	85.39	2	31	4

Table 3.3: Efficiency for b-jet reconstruction and rejection (R) corresponding to different b-tagging efficiency single-cut operating points for the DL1r algorithm, evaluated on $t\bar{t}$ events [41].

3.3.6 Missing transverse momentum

Neutrinos do not interact with detectors, hence they can be reconstructed only using the difference between the initial state and final state total momentum. The initial transverse momentum is zero, so the final one has to be zero too. Candidate neutrinos are defined by missing transverse energy:

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \quad (3.3.1)$$

where E_x^{miss} and E_y^{miss} are calculated as the magnitude of the negative vector sum of the transverse momentum of all the leptons with $p_T > 25$ GeV, all the jets with $p_T > 20$ GeV and a soft term built from tracks which are associated to the hard-scatter vertex, that are not associated to any of the reconstructed objects.

Chapter 4

Machine Learning

The term *Machine Learning* was coined in 1959 by Arthur Samuel, an IBM employee and pioneer in the field of computer gaming and artificial intelligence. The term *Machine Learning* is often interchanged with other terms, such as *Deep Learning*, *Artificial Intelligence* (AI) and *Artificial Neural Networks* (ANN,) but they have different meanings.

AI is the science and engineering of making intelligent machines (John McCarthy '56). AI has many branches: *Machine Learning*, *Deep Learning* and *Artificial Neural Networks* are three of them.

Actually the hierarchy is a little more complex: *Artificial Neural Networks* is a sub-field of *Deep Learning*, which is a sub-field of *Machine Learning*.

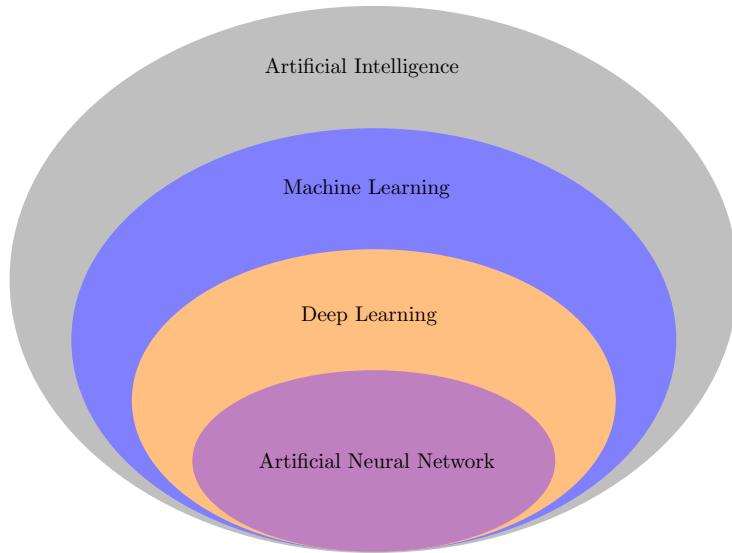


Figure 4.1: Venn diagram representing the relationship between AI, machine learning, deep learning and ANN.

The term *Machine Learning* refers to more than 60 different algorithms and they

can be divided in three main categories:

1. Supervised Learning: the algorithm learns to map an input to an output, based on example input-output pairs. Two main tasks belong to this category: classification and regression;
2. Unsupervised Learning: the algorithm learns patterns from untagged data. Two main tasks belong to this category: clustering, dimensionality reduction;
3. Reinforcement Learning: the algorithm learns from the actions of an agent in an environment. Its main task is real-time decisions;

Since my thesis focuses on a classifier, I will delve into *Supervised Learning* in the next paragraph.

4.1 Supervised Learning

Supervised learning is one of the three main categories of machine learning algorithms. Its task is to infer a function of certain variables in order to predict other variables.

The operative workflow in Machine Learning could be summarized by a *pipeline* of five modules:

1. Data collection and cleaning;
2. Choice of the model, the cost function and the optimizer;
3. Training of the model;
4. Cross-Validation of the model;
5. Optimisation;

Since my thesis focuses on a binary and a multi-class classifier, which belong to the supervised learning category, I will describe the *pipeline* specific to a supervised learning algorithm.

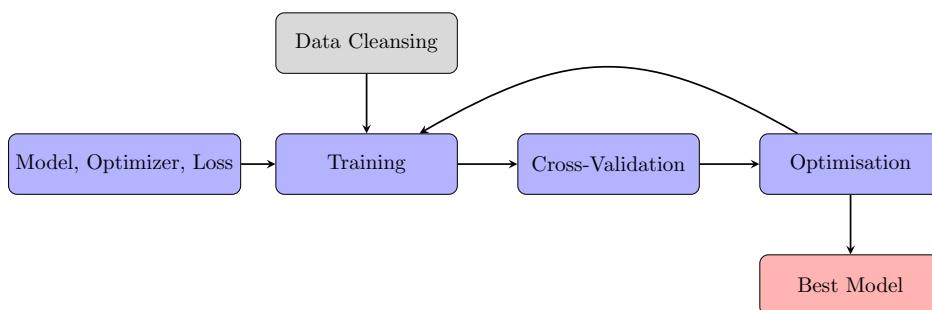


Figure 4.2: Pipeline of a machine learning algorithm.

The first stage is the *collection, evaluation* and *cleaning* of the data. It is essential to study, understand, clean and assemble the training data. The term *data cleaning* or *data cleansing* refers to all kinds of tasks and activities to detect and repair errors in the data.

The second step is the *choice* of the model, the metric and the optimizer. The choice of the model and the metric depend on the specific task. For instance, two possible model choices for a classification task are: Artificial Neural Networks or Boosted Decision Trees.

Many different metrics exist and it is essential to choose the most appropriate for the specific problem.

The final choice is the optimizer.

The most popular optimisation algorithms are *gradient-descent* based (Adam, Nadam, ...). The basic idea underlying these algorithms is to find a minimum (local or global) of the loss function through an iterative process. Therefore, the first step is to select a starting point, an initial vector \vec{w} , and then the algorithm keeps updating the vector of parameters \vec{w} until it finds a minimum of the loss function. Simultaneously each parameter gets updated:

$$w_i := w_i - \eta \frac{\partial}{\partial w_i} J(\vec{w}) \quad (4.1.1)$$

where η is the *learning rate*. The *learning rate* is a tuning parameter which determines the step size at each iteration while moving toward a minimum of the loss function. It represents the *learning speed* of a machine learning model. It is essential to carefully choose the learning rate, since, if η is too small gradient descent could be really slow, if η is too large gradient descent may fail to converge or diverge.

The third step is the *training* of the chosen model.

In this stage the model is provided with observed data (x) and the label of the data (y), also known as *target values*. By providing to the model observed data (x) and the label of the observed data (y), the algorithm learns and adapts to map observed data to the labeled data. For instance, in the training stage of an animal classifier algorithm, x would be images of animals and y the animal breed, hence the algorithm learns to connect the right label to a specific animal.

The fourth step is *Cross-Validation*.

The entire input dataset is usually divided in two groups: *training* set and *testing* set. Therefore the model will have an error associated to the training dataset and a second error associated to the testing dataset. The *bias-variance trade off* is the conflict in trying to simultaneously minimize these two errors that prevent supervised learning algorithms from generalizing beyond their training set. A high bias leads to *underfitting*, which means that the model misses relevant relations between features and

target outputs. A high variance leads to *overfitting*, which means that the model learns the random noise in the training data.

In order to reduce the *bias-variance trade off* it is common to use cross-validation techniques, which divide the dataset in three groups: *training* set, *validation* set and *testing* set.

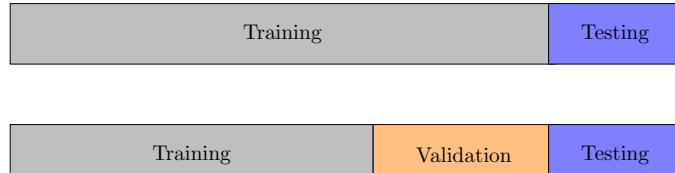


Figure 4.3: Comparison between the division of the data in two sets and three sets.

The cross-validation I used is regarded as *k-fold*.

This technique consists in splitting the training dataset in N folds (groups), where N-1 sets are unified and become the new *training* set and the last one becomes the *validation* set. In rotation every single k-fold will be used as the *validation* set once.

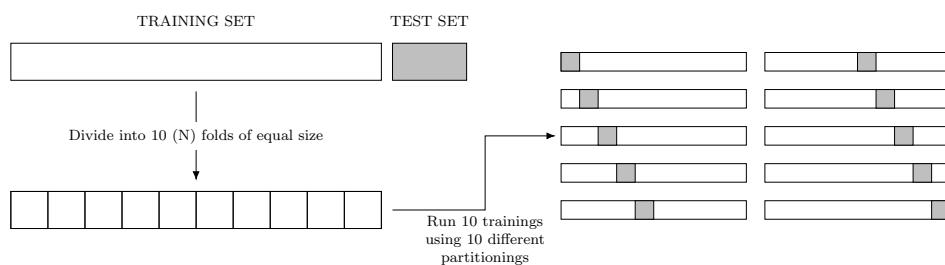


Figure 4.4: Schematic explanation of the *k*-fold cross validation technique with 10 folds.

Finally, the fifth and last step is *Optimisation*.

This step consists in changing the values of the *hyperparameters* of the ANN, in order to find a new ANN which performs better than the previous. This step could be really time consuming.

In conclusion, the third, fourth and fifth steps can be repeated until the *Best Model* is found.

4.2 Artificial Neural Networks

Artificial neural networks (ANNs) are computing systems inspired by the biological neural networks that constitute animal brains.

An ANN can be visualised as *layers of neurons* (or *node*), each of which receives inputs from all neurons in the previous layer. A neural network has an *input* and an *output*

layer and the layers in between are referred to as *hidden*. When the number of hidden layers (the network *depth*) goes beyond one, the network is normally referred to as *deep neural network* (DNN). The number of neurons in a given layer is referred to as its *width*.

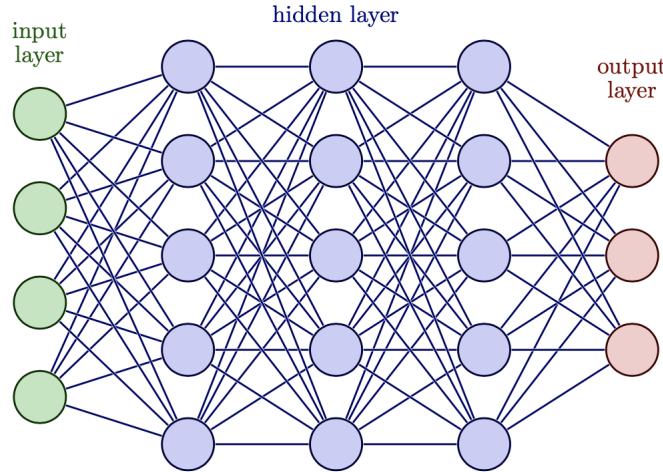


Figure 4.5: Multilayer perceptron with three hidden layers.

The neuron is the fundamental unit of every ANN. Each node combines linearly the input with *weights* and a *bias* and then its output is modulated by an *activation function*, which introduces the non-linearity in the model. Therefore the output of a node would be:

$$a = g(w_0 + w_1x_1 + w_2x_2 + w_3x_3) \quad (4.2.1)$$

where g is the activation function, w_0 is the bias and w_i ($i \neq 0$) are the weights.

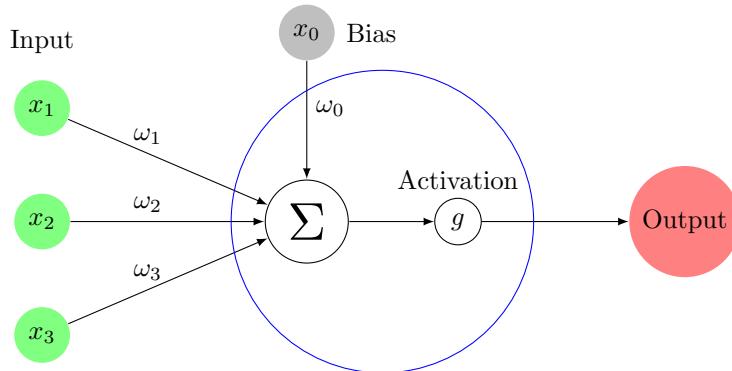


Figure 4.6: Schematic representation of a neuron.

Many different activation functions exist: *tanh*, *sigmoid*, *ReLU*, *ELU*, ...

The training of an ANN is usually performed with gradient descent methods. Since the number of weights could be copious, it is essential to use an algorithm which reduces the large amount of computations. The *backpropagation algorithm* is the most common algorithm to compute the gradient, since it can be used with any gradient-based optimizer.

Many different ANN architectures exist, such as Recurrent neural networks or Generative adversarial network. My thesis focused on a Multilayer Perceptron (MLP), which is a *fully-connected* and *feed-forward* artificial neural network. The term *fully-connected* means that each node of a layer is connected with every node of the previous layer. The term *feed-forward* means that connections between nodes do not form a cycle or a loop, hence the information moves in only one direction (forward) from the input nodes, through the hidden nodes and to the output nodes.

4.2.1 Binary and Multi-Class Classification

Classification problems are supervised learning tasks for predicting variables that consist of a finite number of categories called *classes*. When the number of classes is two the classification problem is called *binary-classification*, when the number of classes is larger than two, the classification problems are called *multi-class classification* problems. Many different models could be used to deal with a classification task, such as K-Nearest Neighbors, Decision Trees or Artificial Neural Networks. The next two paragraphs describe the main features of binary and multi-class classifiers using as model an ANN.

4.2.2 Binary Classifier

A binary classifier has two classes, hence it has just one node in the output layer. Its output is a number between 0 (first class) and 1 (second class), hence it has to be decided a *discrimination threshold*.

The cost function of a binary classifier is the *binary cross-entropy*:

$$J(\vec{w}) = -\frac{1}{n} \sum_{i=1}^n y_i \log(\hat{y}_w(x_i)) + (1 - y_i) \log(1 - \hat{y}_w(x_i)) \quad (4.2.2)$$

where y_i is the target value, $\hat{y}_w(x_i)$ is the model output:

$$\hat{y}_w(x_i) = \frac{1}{1 + e^{-w^T x_i}} \quad (4.2.3)$$

The output of a binary classifier could be visualized in a *confusion matrix*.

		True Class	
		Positive	Negative
Predicted Class	Positive	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
	Negative	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

Figure 4.7: Confusion matrix for a binary classifier.

Each row of the *confusion matrix* represents the instances in a *predicted class* while each column represents the instances in a *true class* (or viceversa). In binary classification tasks the cost function is usually complemented with many other metrics based on the *confusion matrix*:

- Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.2.4)$$

- Precision:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.2.5)$$

- Recall (or True positive rate):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.2.6)$$

- False positive rate:

$$\text{Recall} = \frac{FP}{FP + TN} \quad (4.2.7)$$

- F1 score:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.2.8)$$

However these metrics depend on the discrimination threshold. Therefore it is useful to introduce metrics which show the diagnostic ability of a binary classifier system as its discrimination threshold is varied:

- The *ROC curve* plots the true positive rate against the false positive rate at various threshold settings;
- The *AUC* is the integral of the ROC curve;

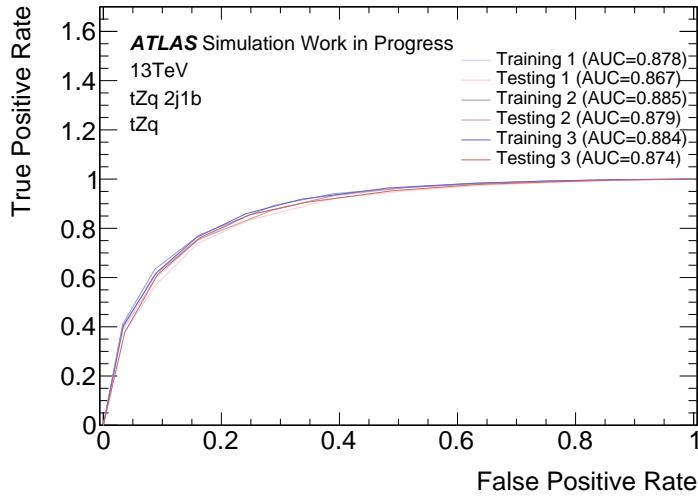


Figure 4.8: ROC curves for a three folded training of a binary classifier.

4.2.3 Multi-class Classifier

There are two main strategies for solving a multi-class classification problem. The first is the use of classification algorithm which can solve the problem directly (K-Nearest Neighbors, Decision Trees, Artificial Neural Networks). The second strategy is the decomposition of the original multi-class classification problem into several binary sub-problems (one-against-one or one-against-rest), and each sub-problem can be solved with a different classification algorithm [42].

In my analysis I used an ANN which solves the problem directly without decomposing it in several binary sub-problems. A multi-class classifier has one node per class in the output layer.

Unlike a binary classifier whose output is modulated by a sigmoid, a multi-class network has as output a *softmax* output. The *softmax* function takes as input the vector of values of the output layer and converts them into probabilities associated to each class. Therefore the output of *softmax* is a probability vector of three components for a multi-class classifier with three classes.

$$\sigma_i(\vec{x}) = \frac{\exp(x_i)}{\sum_{i=1}^N \exp(x_j)} \quad (4.2.9)$$

The sum of the components of the *softmax* output vector is normalized to one by definition.

The cost function of a multi-class classifier is the *categorical cross-entropy*:

$$J(\vec{w}) = - \sum_{i=1}^n y_i \log(\hat{y}_w(x_i)) \quad (4.2.10)$$

where y_i is the i-th element of the vector \vec{y} (the target vector) and represents the actual class, meanwhile $\hat{y}_w(x_i)$ represents the value of the i-th output. The target vectors \vec{y} for a three class classifier are three, one for each class:

$$\vec{y}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{y}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{y}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4.2.11)$$

Chapter 5

Separation of tZq from background with DNNs

The most interesting final states for the tZq process is the trilepton final state, which occurs when both the Z and W bosons decay leptonically. The trilepton channel is the one with the lowest branching ratio, but it is the purest one. In section 5.1 I will briefly discuss the Monte Carlo samples used in the analysis. In section 5.2 I will describe the event selection and the signal regions chosen for my analysis. The multivariate analysis I conducted will be discussed in sections 5.3, 5.4, 5.5. Finally I will sum up the results of my analysis in section 5.6.

5.1 Data Samples

Monte Carlo techniques are used to simulate the events produced by the inelastic collision of protons. The simulated events were grouped in three Monte Carlo campaigns:

- **mc16a**: it simulates the data collected in 2015 and 2016. The total integrated luminosity to which these events have to be scaled to is 36.2fb^{-1} .
- **mc16d**: it simulates the data collected in 2017. The total integrated luminosity to which these events have to be scaled to is 44.3fb^{-1} .
- **mc16e**: it simulates the data collected in 2018. The total integrated luminosity to which these events have to be scaled to is 58.5fb^{-1} .

The events were simulated using as mass of the top quark $m_t = 172.5$ GeV and a branching ratio of the $t \rightarrow Wb$ decay equal to 1. In the following paragraph, I will describe the samples for both the signal and the SM backgrounds.

5.1.1 Signal

The signal sample is tZq .

The signal sample is the single top quark production in t-channel in association with a

Z boson, where the Z boson decays to a pair of charged leptons in the trilepton channel. The tZq final state used for this measurement comprises three charged leptons (electrons or muons), missing transverse momentum, one b-jet from the top-quark decay and an additional jet (un-tagged jet). A second untagged jet is allowed in order to include events with QCD radiation.

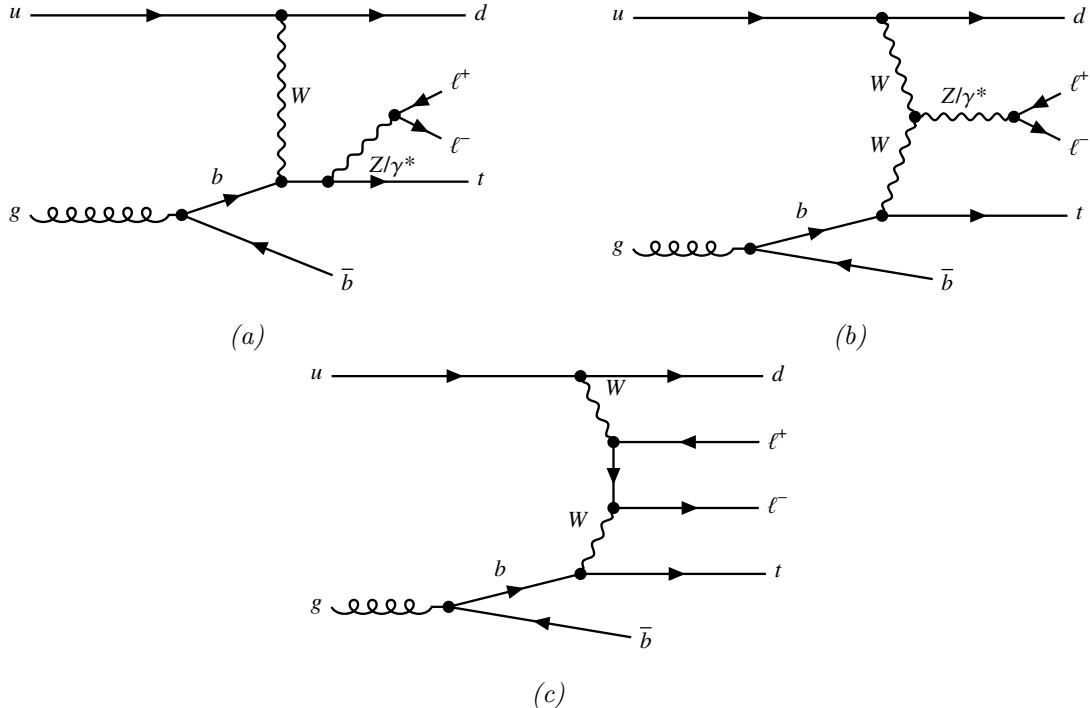


Figure 5.1: Example Feynman diagrams of the lowest-order amplitudes for the tZq process, corresponding to (a, b) resonant $\ell^+\ell^-$ production and (c) non-resonant $\ell^+\ell^-$ production. In the four-flavour scheme, the b-quark originates from gluon splitting.

5.1.2 Background

A variety of background sources are considered. These include SM processes with similar final states as tZq , as well as events in which at least one of the leptons in the final state is ‘fake’ (either a jet misidentified as a lepton or a non-prompt lepton).

tW

Three different single-top channels exist, however the s- and t-channels have no overlap with the tZq final state, so only the tW sample is evaluated. Two samples are used: the first one has a top quark and a W^- boson in the final state, whereas the second has an anti-top quark and a W^+ boson. Both samples have two leptons in the final state, one coming from the leptonic decay of the W boson and the second from the leptonic decay of another W boson, generated by the decay of the top quark.

t \bar{t} H

It is a strong process with three different final states, defined by the decays of the W bosons produced by the decay of the top and the anti-top. We can have a hadronic final state, which is when both t and \bar{t} decay hadronically, a one-lepton final state, that is when only one of the two W bosons decays leptonically and a dilepton final state when both W bosons decay leptonically.

t \bar{t} Z

It is a strong process of a pair production in association with a Z boson. There are five samples because the Z boson has five different possible decays: $e\bar{e}$, $\mu\bar{\mu}$, $\tau\bar{\tau}$, $\nu\bar{\nu}$, $q\bar{q}$.

t \bar{t} W

This process consists in a strong top quark pair production associated with a W boson. There is only one sample, which is inclusive, in that the three W bosons (two from $t\bar{t}$ and the third one) can decay either leptonically or hadronically.

tWZ

This event was described in [2.2.1](#). It is required that the Z boson decays leptonically.

Z + jets

The Z + jets category contains events where a Z boson decays leptonically. This is a background for the tZq event because it has two leptons coming from the leptonic decay of the Z boson and a third fake lepton. Two jets must also be produced in order to fake the signal signature.

Diboson

In the trilepton topology, WZ events are the ones that significantly contribute to the background. Two samples are semi-leptonic decays: $Z \rightarrow (q\bar{q})$ $Z \rightarrow (\ell\bar{\ell})$, $W \rightarrow (q\bar{q})$ $Z \rightarrow (\ell\bar{\ell})$. Three samples are fully-leptonic: $\ell\ell\ell\ell$, $\ell\ell\nu\nu$, $\ell\ell\ell\nu$.

5.2 Event Selections and Signal Regions

My analysis was conducted on two different event selections and three different signal regions (SR).

I focused on two different selections: the *default* selection and the *loose* selection. Table [5.1](#) shows the specifics of both selections.

Selection	Jet p_T	b-tag working point	Leptons p_T
Default	35 GeV	70%	27,20,15 GeV
Loose	20 GeV	85%	27,15,10 GeV

Table 5.1: Specifics of the default and loose selections.

The *default* selection uses as b-tag working point `eff_70`, whereas the *loose* selection `eff_85`. Moreover the p_T thresholds for the jets are 35 GeV (*default*) and 20 GeV (*loose*). Both selections require exactly three leptons (no hadronically-decaying τ). The *default* selection requires as p_T thresholds for the leptons: 27 GeV, 20 GeV, 15 GeV. The *loose* selection requires as p_T thresholds for the leptons: 27 GeV, 15 GeV, 10 GeV. Table 5.2 shows the yield of each event for the *default* and *loose* selections. The *loose* selection has many more background events than the *default* selection.

Event	Default			Loose		
	2j1b	3j1b	nj1b	2j1b	3j1b	nj1b
tZq	118	66	184	93	99	192
Diboson	235	69	304	557	450	1007
$t\bar{t}Z + tWZ$	97	148	245	33	79	112
Z + jets	10	10	20	368	205	573
Others	9	6	15	6	6	12
Total	469	299	768	1057	839	1896

Table 5.2: Event yields for the default and loose selections.

Regarding the signal regions I worked on three different SRs. The first signal region considered is made of 2 jets, one of which is b-tagged and it is regarded as the *2j1b SR*. The second signal region considered is made of 3 jets, one of which is b-tagged and it is regarded as the *3j1b SR*. The third signal region is the sum of the first two signal regions and it is regarded as the *nj1b SR*.

The *2j1b* and *3j1b* SRs are orthogonal. For the *2j1b SR* the b-jet is defined as the b-tagged jet and the *forward* jet as the other jet. For the *3j1b SR* the definition of the b-jet is the same as for *2j1b*. The *forward* and the *radiation* jets are defined respectively as the jet with the highest and the lowest invariant mass with the b-jet. Both SRs required an opposite-sign-same-flavour (OSSF) lepton pair for the reconstruction of the leptons coming from the Z boson decay. The lepton final states combinations are $e e \mu \mu$, $e \mu \mu \mu$, $e e e e$ and $\mu \mu \mu \mu$. In the first two cases it is trivial to determine the OSSF pair, whereas in the last two scenarios the OSSF pair is made of the two leptons with invariant mass closer to the mass of the Z boson. The inferior acceptance threshold on the invariant mass of the OSSF pair is set at 81 GeV, in order to avoid non-resonant OSSF lepton pairs. Table 5.3 shows the specifics of the *2j1b SR* and the *3j1b SR*.

SR 2j1b	SR 3j1b
1 OSSF pair	1 OSSF pair
$ m_{\ell\ell} - m_Z < 10 \text{ GeV}$	$ m_{\ell\ell} - m_Z < 10 \text{ GeV}$
2 jets, $ \eta < 4.5$	3 jets, $ \eta < 4.5$
1 b-jet, $ \eta < 2.5$	1 b-jet, $ \eta < 2.5$

Table 5.3: Overview of the requirements applied when selecting events in the 2j1b and 3j1b signal regions. OSSF is an opposite-sign same-flavour lepton pair.

5.3 Definition of the DNN classes

The separation of the tZq event from the background can be conducted with either a binary or a multi-class classifier.

The goal of a multi-class classifier is to focus on a specific background process by isolating it in a specific class.

Regarding the default selection, table 5.2 shows that *diboson* events are the dominant background source in the 2j1b signal region, whereas $t\bar{t}Z$ and *diboson* events are the dominant background source in the 3j1b SR. Regarding the loose selection, table 5.2 shows that *diboson* and $Z + \text{jets}$ events are the dominant background source in the 2j1b and 3j1b SR. Therefore the binary classifier has just two classes, *signal* and *background*, whereas the multi-class classifier I trained had three classes:

- Signal. The class containing only the tZq events;
- Dominant Background. This class contains the most dominant background;
- Other. The class containing every other background source;

Initially my analysis was based only on the default selection, hence, since the most dominant backgrounds are *diboson* and $t\bar{t}Z$, I trained two different multi-class classifiers. The first one had as classes: *signal*, *diboson*, *other*.

The second one had as classes: *signal*, $t\bar{t}Z + tWZ$, *other*.

Later I started working on the loose selection, which inherited the classes previously defined for the default selection, even though $Z+\text{jets}$ counts many more events than $t\bar{t}Z$.

However the definition of the classes based on the dominance of a specific background event is not the only possible choice. Another reasonable strategy could be to define the classes looking at the promptness of the leptons. Therefore two classes could be defined: the first would contain the events with three prompt leptons and the second with non-prompt leptons. The multi-class classifier would have three classes: *signal*, *non-prompt* class, *prompt* class.

5.4 Variables for the DNN trainings

The first step of the analysis was the choice of the variables to use in the trainings of the DNNs. The goal of this analysis step is to find the minimal set of input variables that achieve a similar separation of signal and background as the full set.

Initially I considered 41 different variables (table 5.4), that contained different physics observables associated to different particles involved in the events.

The variable list could be divided in two main groups: *assigned variables* and *variables ordered by p_T* . The *assigned variables* are variables in which leptons are assigned to the boson that produced them, whereas *variables ordered by p_T* are variables in which leptons and jets are ordered by p_T .

The following table contains the complete list of variables I considered.

Variable	Definition
$m(b, j_f)$	Invariant mass of the b-jet and the j_f
$m(\text{top})$	Reconstructed top quark mass
$m_T(W)$	Transverse mass of the W boson
$m(Z)$	Mass of the reconstructed Z boson
$p_T(Z)$	p_T of the reconstructed Z boson
$p_T(W)$	p_T of the reconstructed W boson
$p_T(\text{top})$	p_T of the top quark
$p_T(\ell_{Z_1})$	p_T of the first lepton from the Z boson decay
$p_T(\ell_{Z_2})$	p_T of the second lepton from the Z boson decay
$p_T(j_f)$	p_T of the j_f jet
$p_T(j_b)$	p_T of the b-jet
$p_T(\ell_W)$	p_T of the lepton from the W boson decay
$ \eta(j_f) $	Absolute value of the η of the j_f jet
$ \eta(j_b) $	Absolute value of the η of the b-jet
$ \eta(\ell_W) $	Absolute value of the η of the lepton from the W-boson decay
$ \eta(Z) $	Absolute value of the η of the Z boson
$ \eta(\text{top}) $	Absolute value of the η of the top
$ \eta(W) $	Absolute value of the η of the W boson
$ \eta(\ell_{Z_1}) $	Absolute value of the η of the first lepton from Z boson decay
$ \eta(\ell_{Z_2}) $	Absolute value of the η of the second lepton from Z boson decay
$\Delta R(j_f, Z)$	ΔR between the j_f jet and the Z boson
$\Delta R(\text{top}, Z)$	ΔR between the top and the reconstructed Z boson
$\Delta R(\ell_W, j_{\text{closest}})$	ΔR between the lepton from W boson decay and closest jet
$\Delta\phi(j_f, Z)$	$\Delta\phi$ between the j_f jet and the Z boson
$\Delta\phi(\text{top}, Z)$	$\Delta\phi$ between the top and the Z boson
$\Delta\phi(W, j_{\text{closest}})$	$\Delta\phi$ between the W boson and the closest jet
$j_f(b - \text{tagging})$	b-tagging score of the j_f
$j_b(b - \text{tagging})$	b-tagging score of the b-jet

<i>mindiffmass</i>	Mass difference between Z boson and leptons of its decay
H_T	Scalar sum of the p_T of the leptons and jets in the event
$q(\ell_W)$	Electric charge of the lepton from the W boson decay
<hr/>	
$p_T(\ell_1)$	p_T of the lepton with the highest p_T
$p_T(\ell_2)$	p_T of the lepton with the second highest p_T
$p_T(\ell_3)$	p_T of the lepton with the third highest p_T
$p_T(j_1)$	p_T of the jet with the highest p_T
$p_T(j_2)$	p_T of the jet with the second highest p_T
$\phi(\ell_1)$	ϕ of the lepton with the highest p_T
$\phi(\ell_2)$	ϕ of the lepton with the second highest p_T
$\phi(\ell_3)$	ϕ of the lepton with the third highest p_T
$\phi(j_1)$	ϕ of the jet with the highest p_T
$\phi(j_2)$	ϕ of the jet with the second highest p_T

Table 5.4: Complete list of variables (41). The variables ordered by p_T are on the bottom of the table.

My first objectives were:

- To rank the variables with respect to their separation power;
- To determine if I should use the *assigned variables* or the *variables ordered by p_T* ;

In order to do so, I estimated the separation strength of each variable between the tZq signal and three classes: *diboson*, $t\bar{t}Z + tWZ$, *other* (definition of the classes in section 5.3). The separation value of each variable depends on the signal region (2j1b and 3j1b) and the classes.

I used the TRExFitter framework to estimate the separation strength of a variable y , defined by the integral:

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S - \hat{y}_B)^2}{\hat{y}_S + \hat{y}_B} \quad (5.4.1)$$

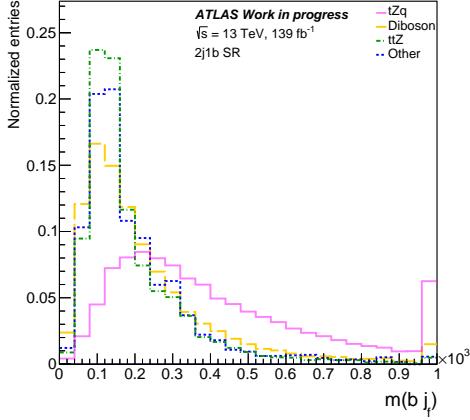
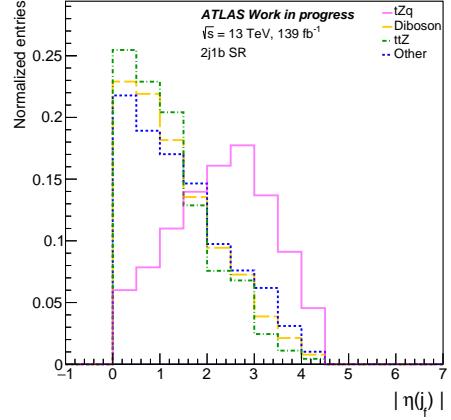
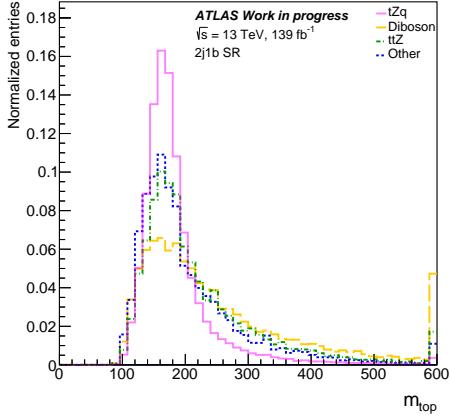
where \hat{y}_S and \hat{y}_B are the signal and background PDFs of y respectively. The separation is zero for identical signal and background shapes, and it is one for shapes with no overlap.

The following tables contains the separation strength for each variable in 2j1b SR (table 5.5) and 3j1b SR (table 5.6).

Variable	Diboson	ttZ+tWZ	Other
$m(bj_f)$	17.6	27.3	23.8
$ \eta(j_f) $	18.4	23.8	14.2
m_{top}	16.1	7.1	5.14
$m_T(W)$	1.47	4.57	8.26
$p_T(\ell_W)$	5.1	3.68	5.31
$p_T(j_f)$	2.23	2.64	6.84
j_b (b-tagging)	10.3	0.0786	0.261
H_T	2.15	2.01	3.47
$p_T(Z)$	1.49	4.79	0.577
$p_T(W)$	2.1	3.03	0.943
j_f (b-tagging)	0.256	3.36	1.53
$ \eta(Z) $	0.261	3.01	1.55
$p_T(\ell_{Z_1})$	1.17	2.95	0.502
$p_T(\ell_{Z_2})$	0.525	2.93	0.495
$p_T(j_b)$	3.16	0.688	0.227
$q(\ell_W)$	0.421	1.39	1.45
$\Delta R(\ell_W, j_{closest})$	2.13	0.695	0.932
$\Delta R(j_f, Z)$	0.551	2.6	0.282
$\Delta\phi(top, Z)$	0.289	2	0.97
$\Delta\phi(Z, j_f)$	1.66	0.563	0.815
$ \eta(top) $	1.69	0.128	0.713
$\Delta R(topZ)$	0.122	1.49	1.03
$p_T(top)$	0.688	0.352	1.59
$ \eta(\ell_W) $	1.63	0.309	0.476
$ \eta(\ell_{Z_1}) $	0.0218	1.09	1.12
m_Z	0.0294	0.0891	2.1
$ \eta(W) $	0.914	0.124	0.912
$mindiffmass$	0.00374	0.0522	1.89
$ \eta(\ell_{Z_2}) $	0.131	0.698	0.693
$\Delta\phi(W, jet)$	0.127	0.344	0.5
$ \eta(j_b) $	0.549	0.0711	0.0952
$p_T(\ell_3)$	1.01	2.9	5.8
$p_T(\ell_2)$	2.98	4.64	1.66
$p_T(\ell_1)$	3.4	4.48	0.848
$p_T(j_2)$	2.3	0.747	3.36
$p_T(j_1)$	2.19	0.349	3.04
$\phi(\ell_1)$	0.0087	0.0771	0.249
$\phi(j_2)$	0.0232	0.0583	0.168
$\phi(\ell_2)$	0.0052	0.068	0.117
$\phi(j_1)$	0.0166	0.0667	0.0912

$\phi(\ell_3)$	0.0129	0.0212	0.0992
----------------	--------	--------	--------

Table 5.5: Separation values of the variables in the 2j1b signal region.

Invariant mass of the b -jet and the j_f .Absolute value of the η of the j_f jet.

Reconstructed top quark mass.

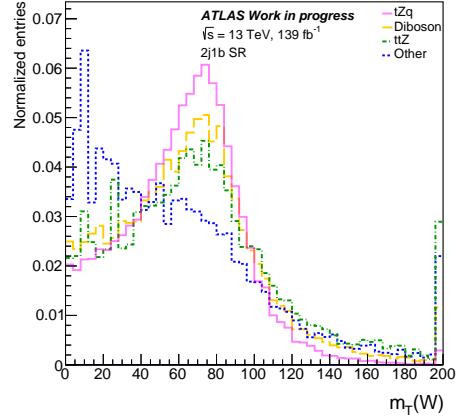
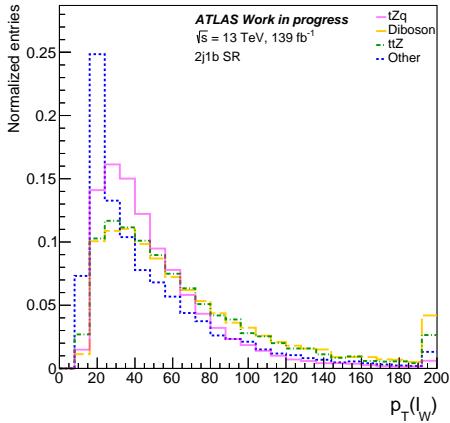
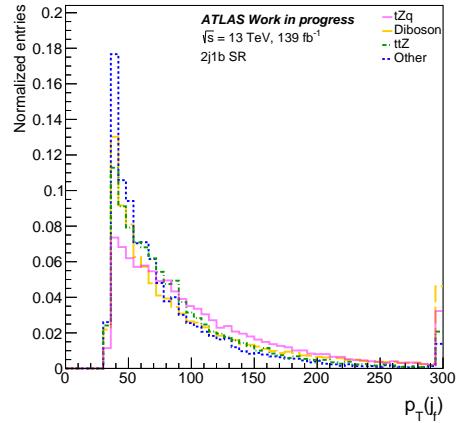
Transverse mass of the W boson. p_T of the ℓ_W . p_T of the j_f .

Figure 5.2: Normalized distribution for the best variables in the 2j1b signal region. The pink line shows the distribution of the variable for the tZq event.

Variable	Diboson	ttZ+tWZ	Other
$m(bj_f)$	12.3	25.4	22.5
$ \eta(j_f) $	14.8	20	16.4
m_{top}	14.6	3.04	2.34
$p_T(\ell_W)$	6.03	2.39	1.62
j_b (b-tagging)	9.16	0.0291	0.0476
$p_T(Z)$	2.02	4.32	2
j_f (b-tagging)	0.396	4.06	3.74
$m_T(W)$	1.7	1.99	2.88
$p_T(j_f)$	1.72	1.89	2.21
$p_T(\ell_{Z_1})$	1.63	2.65	1.23
$p_T(W)$	3.08	1.95	0.627
H_T	3.72	1.06	0.292
$ \eta(Z) $	0.283	2.73	2.02
$q(\ell_W)$	0.682	2.03	1.99
$p_T(\ell_{Z_2})$	0.876	2.62	1.31
$\Delta R(top, Z)$	0.788	1.86	1.61
$\Delta R(j_f, Z)$	0.319	2.29	1.33
$\Delta R(\ell_W, j_{closest})$	2.1	0.37	0.618
$p_T(j_b)$	2.35	0.27	0.103
$\Delta\phi(W, j)$	0.525	0.956	0.948
$ \eta(\ell_W) $	1.43	0.237	0.475
$ \eta(W) $	0.835	0.595	0.428
$ \eta(\ell_{Z_1}) $	0.0335	0.94	0.776
$\Delta\phi(topZ)$	0.186	1.01	0.481
$ \eta(\ell_{Z_2}) $	0.108	0.666	0.475
$p_T(top)$	0.542	0.258	0.273
$\Delta\phi(Zj_f)$	0.441	0.23	0.258
$ \eta(j_b) $	0.318	0.307	0.18
m_Z	0.0691	0.0414	0.663
$ \eta(top) $	0.78	0.0287	0.0829
$mindiffmass$	0.0135	0.0136	0.572
$p_T(\ell_1)$	4.39	3.81	1.52
$p_T(\ell_2)$	3.83	3.95	1.51
$p_T(\ell_3)$	1.4	2.22	1.58
$p_T(j_1)$	2.53	0.604	0.926
$p_T(j_2)$	0.891	0.147	0.31
$\phi(\ell_2)$	0.0674	0.0317	0.0161
$\phi(\ell_3)$	0.0422	0.0103	0.0304
$\phi(\ell_1)$	0.0222	0.0188	0.043
$\phi(j_2)$	0.0112	0.0269	0.0129

$\phi(j_1)$	0.0193	0.0145	0.00581
-------------	--------	--------	---------

Table 5.6: Separation values of the variables in the 3j1b signal region.

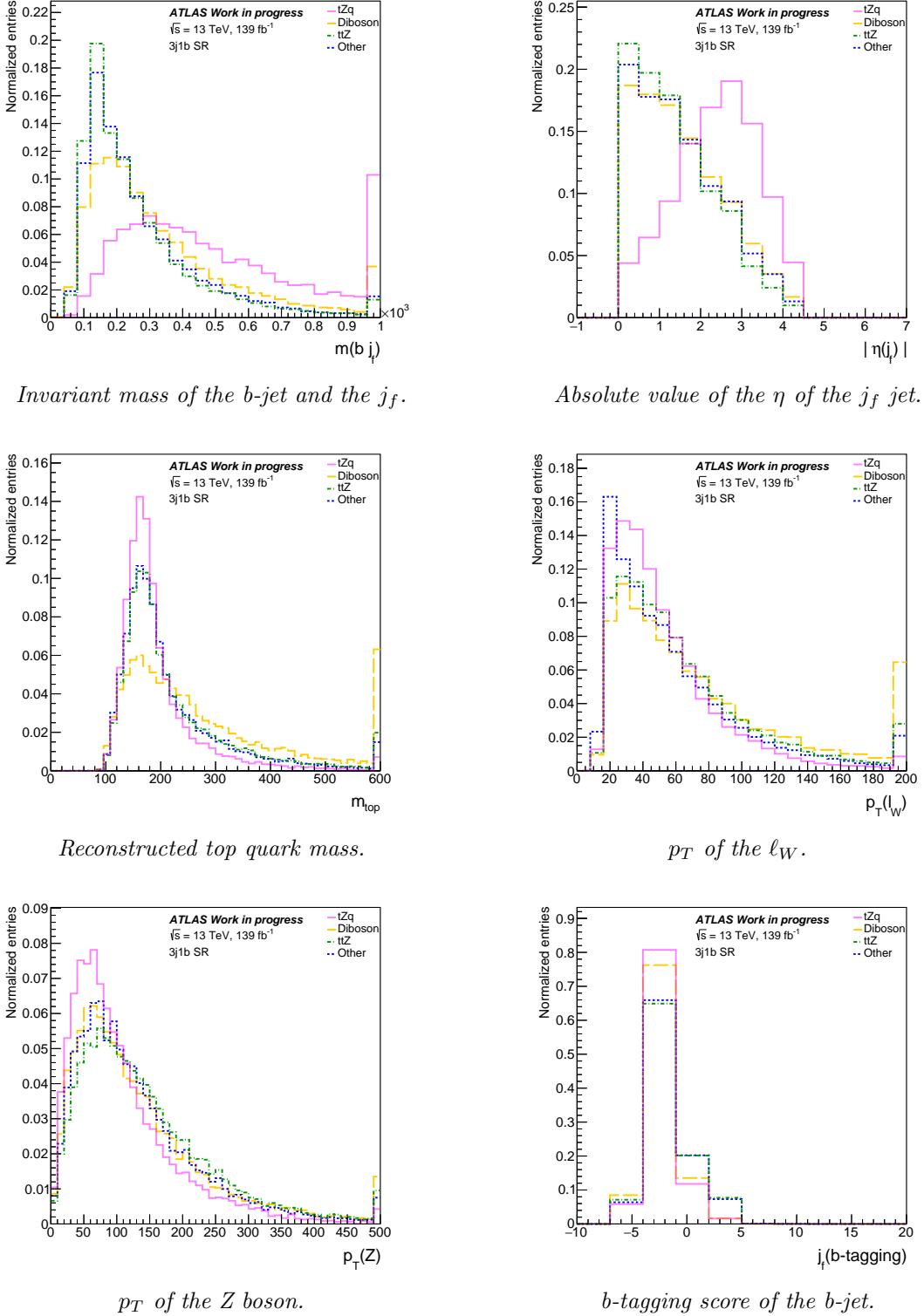


Figure 5.3: Normalized distribution for the best variables in the $3j1b$ signal region. The pink line shows the distribution of the variable for the tZq event.

Both tables (5.5 and 5.6) show that $m(b, j_f)$, $|\eta(j_f)|$ and $m(\text{top})$ have the best separation values. Moreover these tables (5.5 and 5.6) show that the *variables ordered by p_T* do not have a significant separation value with respect to the *assigned variables*. This is due to their definition: *variables ordered by p_T* do not carry with them informations related to the particular process I am analyzing (tZq event). Moreover defining *variables ordered by p_T* adds an unnecessary complication to the analysis: it is way more straightforward to use *assigned variables*. Therefore I decided to exclude the *variables ordered by p_T* and to base the entire analysis on the *assigned variables*, reducing the list of variables to 31.

After that I evaluated the Pearson's correlation between the 31 remaining variables, in order to shrink the list even further:

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (5.4.2)$$

where X, Y are two different variables, cov(X, Y) is the covariance and σ_i is the standard deviation.

The Pearson's correlation of the variables was evaluated on every sample (both signal and background). Even though the correlation does not heavily depend on the signal region, I evaluated the Pearson's correlation for the 2j1b and 3j1b signal regions separately. Figure 5.4 and 5.4 shows only the correlations higher than 0.5.

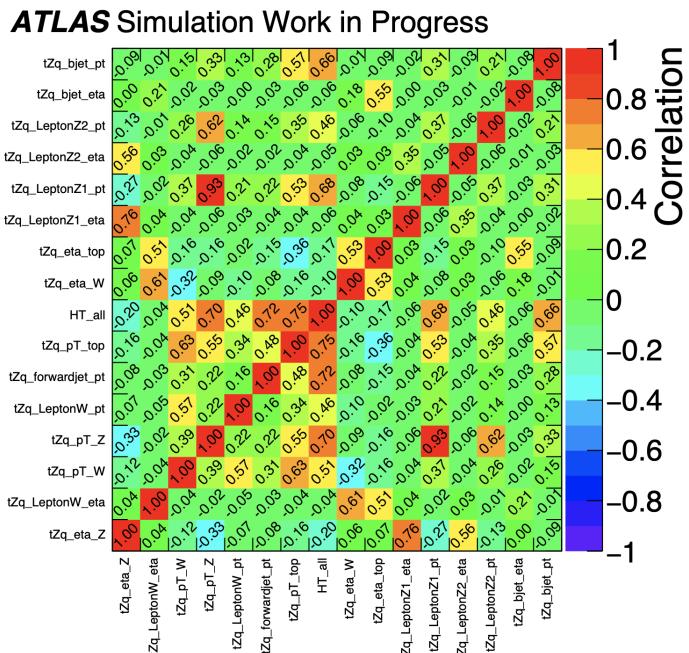


Figure 5.4: Correlation matrix for the 2j1b signal region. This matrix shows only the variables containing at least one correlation's value higher than 0.5.

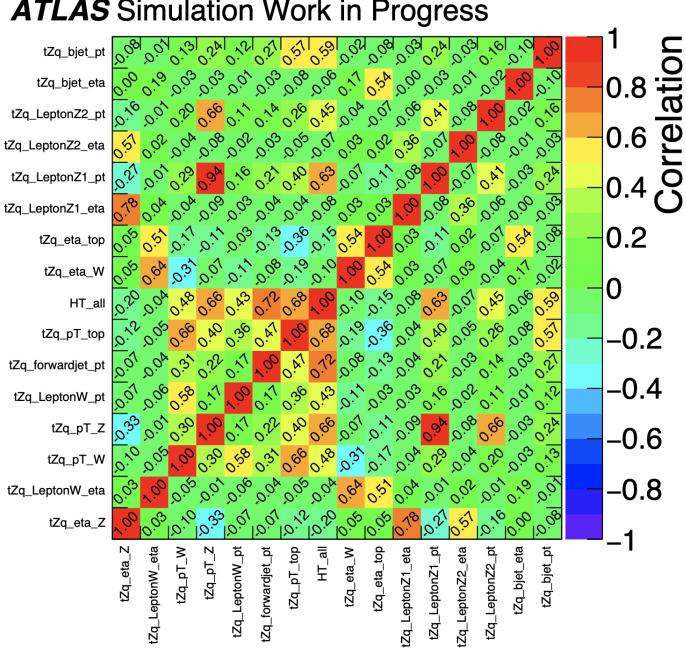


Figure 5.5: Correlation matrix for the 3j1b signal region. This matrix shows only the variables containing at least one correlation's value higher than 0.5.

As expected the variables of the leptons coming from a boson decay (both p_T and $|\eta|$) are highly correlated to the ones of the boson which generated them. Moreover H_T is highly correlated to the p_T of almost every object, but that is not a surprise looking at its definition (table 5.4).

Therefore I could shrink the list down to 23 elements. The final list of variables used in the DNN trainings is shown in table 5.7.

$m(b, j_f)$	$m(\text{top})$	$q(\ell_W)$
$m_T(W)$	$m(Z)$	$j_b(b - \text{tagging})$
$p_T(j_b)$	$ \eta(j_b) $	$j_f(b - \text{tagging})$
$ \eta(j_f) $	$ \eta(Z) $	mindiffmass
$ \eta(\ell_W) $	$p_T(\ell_W)$	$\Delta\phi(W, j_{\text{closest}})$
$p_T(j_f)$	$p_T(\text{top})$	$\Delta R(\text{top}, Z)$
$\Delta R(j_f, Z)$	$\Delta\phi(\text{top}, Z)$	$\Delta R(\ell_W, j_{\text{closest}})$
$p_T(Z)$	$p_T(W)$	

Table 5.7: List of variables (23) used in the DNN trainings.

5.5 DNN trainings

I studied three different NNs strategies for each signal region (3) for both selections (2): a binary classifier that discriminated between signal and background, a multiclass network that distinguished between signal, diboson and all other background processes and another multiclass network that distinguished between signal, $t\bar{t}Z$ and all other background processes. In summary I have trained and optimised a total of 18 neural networks.

These eighteen architectures shared some common features that I will describe in this paragraph, whereas the differences will be discussed contextually in specific paragraphs. These architectures were fully-connected feed-forward NNs consisting of one input layer with one node for each input variable, followed by five hidden layers (20, 30, 30, 30, 20) and one output layer. The output layer consisted of one node for the binary classifier and three nodes representing the three classes for the multi-class network.

The *sigmoid* was used as the output function of the output node for the binary classifier. The *softmax* was used as the output function for the output layer for the multiclass classifier.

The *rectified linear unit* was used as the non-linear activation function of the nodes in all hidden layers:

$$ReLU(x) = \max(0, x) \quad (5.5.1)$$

During the training, the weights were adapted with the *Adam* optimisation algorithm [43, 44]. Each training was characterized by 3 *folds*, 2000 *epochs*, *validation size* of 25 % and *batch size* equal to 1000.

Moreover *early stopping* was implemented with a patience of 100 epochs and $\Delta_{min} = 0.0005$ to avoid *overfitting*.

Furthermore four *dropout* layers have been used as a regularisation method with a *dropout probability* of 30 %.

The *learning rate* differed from one architecture to another, hence it will be later specified for each DNN.

Finally all variables were preprocessed before they were used as input to the DNNs. All variables were transformed with the *MinMax* scaler into the range [0,1], meaning that the minimum and maximum value of each variable is going to be 0 and 1, respectively:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} + x_{min}} \quad (5.5.2)$$

5.5.1 Loss functions

Since I have trained eighteen DNNs and each one was subjected to a three-folded training I would have to show 54 loss functions.

However, in order to avoid being too verbose, in this paragraph I will describe briefly the general behaviour of the loss functions of each training for each signal region showing

few examples.

Firstly I will describe the loss functions of the binary classifiers. I trained six different binary classifiers and each one of them showed a well converging model. The validation and the training loss converged: there was not any sign of overtraining. Figure 5.6 shows two examples: the plot on the left is the second fold of the default binary classifier of the 2j1b signal region, whereas the plot on the right is the first fold of the loose binary classifier of the nj1b signal region.

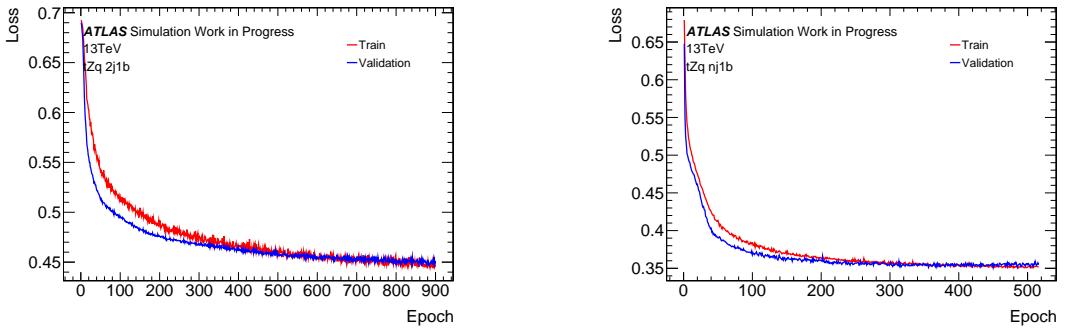


Figure 5.6: The plot on the left is the loss function of the second fold of the binary default classifier of the 2j1b signal region. The plot on the right is the first fold of the loss function of the binary loose classifier of the nj1b signal region.

Regarding the loss functions of the diboson multi-class classifiers, I have to point out that in each fold of the nj1b signal region the validation loss of the default classifier was slightly lower than the training loss. This behaviour could be due to the fact that the validation loss is always calculated after one epoch, whereas the training loss is calculated as the average across the same epoch. Therefore, since the validation loss is calculated once the weights have already been updated it is expected to be slightly lower than the training loss. This could be also accentuated by the high number of dropout layers (4 out of five hidden layers).

Figure 5.7 shows two examples for the diboson multi-class classifiers.

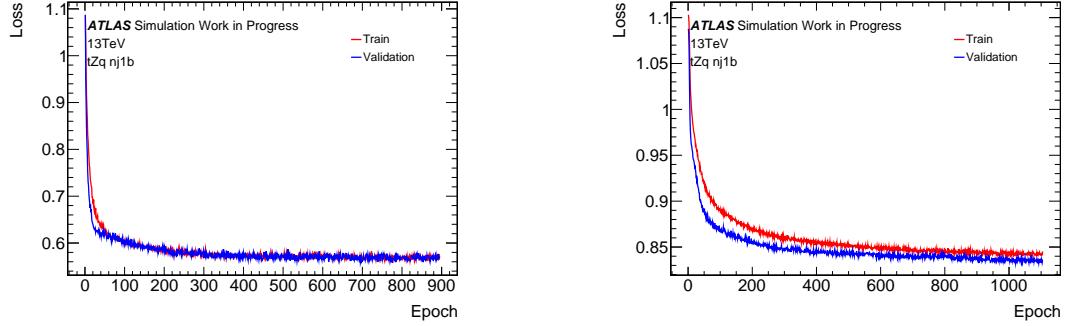


Figure 5.7: The plot on the left is the loss function of the third fold of the diboson loose classifier of the nj1b signal region. The plot on the right is the loss function of the second fold of the diboson default classifier of the nj1b signal region.

Finally the loss functions of the $t\bar{t}Z$ multi-class classifiers showed six well converging models. Every fold did not show signs of overtraining. Figure 5.8 shows two examples for the $t\bar{t}Z$ multi-class classifiers.

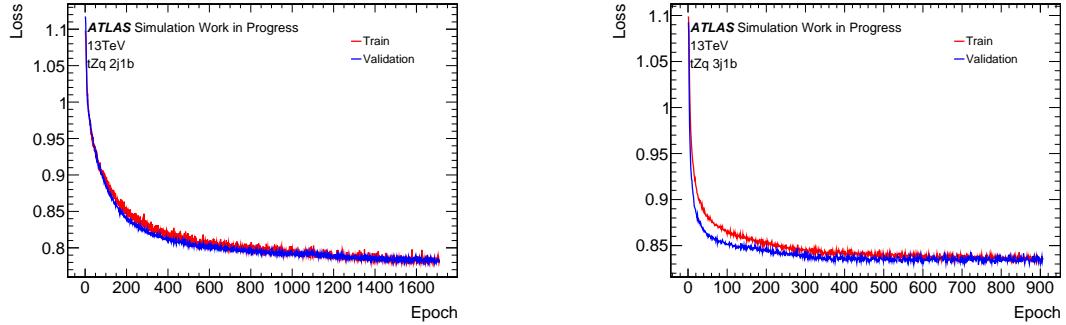


Figure 5.8: The plot on the left is the loss function of the first fold of the $t\bar{t}Z$ default classifier of the 2j1b signal region. The plot on the right is the loss function of the third fold of the $t\bar{t}Z$ loose classifier of the 3j1b signal region.

The next three paragraphs (5.5.2, 5.5.3, 5.5.4) describe the trainings for each signal region comparing the results obtained for the two different selections. Therefore each paragraph will be divided in three sub-paragraphs: the binary training, the multi-class training with diboson as the dominant background class and the multi-class training with $t\bar{t}Z$ as the dominant background class. In each sub-paragraph the results obtained in the default and loose selections will be compared.

5.5.2 Trainings for the 2j1b signal region

This paragraph shows and compares the trainings of the 2j1b signal region for both selections (default and loose). Throughout this paragraph I will refer to the binary and multi-class classifiers of the default and loose selections as the *default* and *loose* classifier. The main features of each architecture have been previously described in section 5.5.

Binary classifier

The learning rate of the default and loose classifier is 0.0001 and 0.0002 respectively. The confusion matrices show that the loose classifier performs better than the default classifier (figure 5.9).

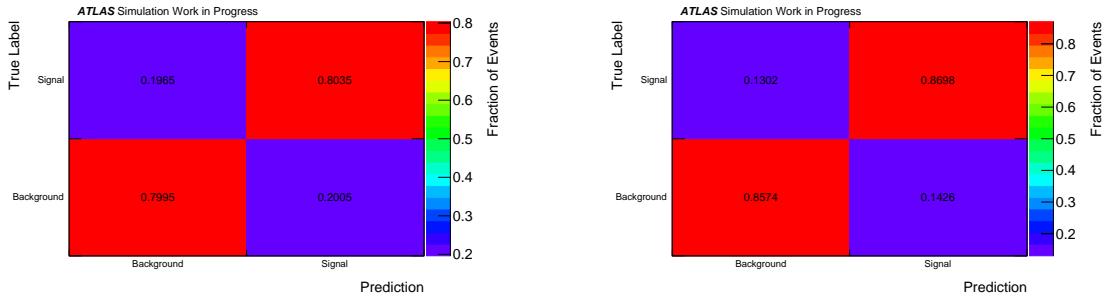


Figure 5.9: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 2j1b SR.

Moreover the separation plots show that the default classifier has a separation of 43.77 %, whereas the loose classifier of 60.11 % (figure 5.10).

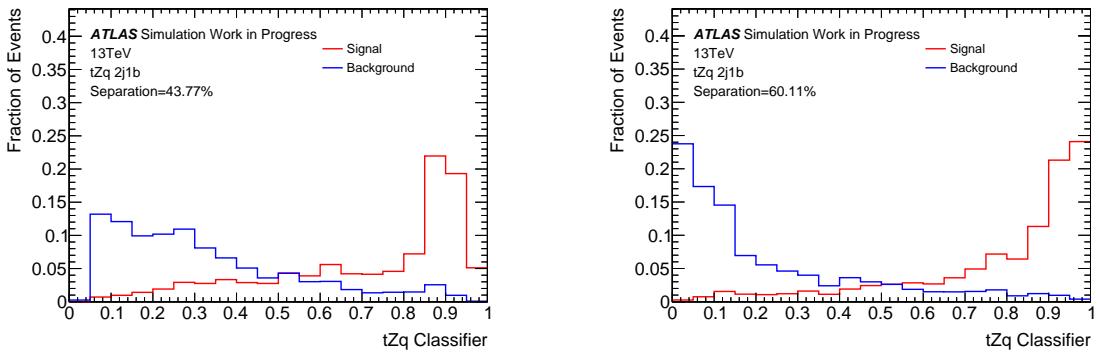


Figure 5.10: Comparison of the normalized distributions of the default (left) and loose (right) classifier in the 2j1b SR.

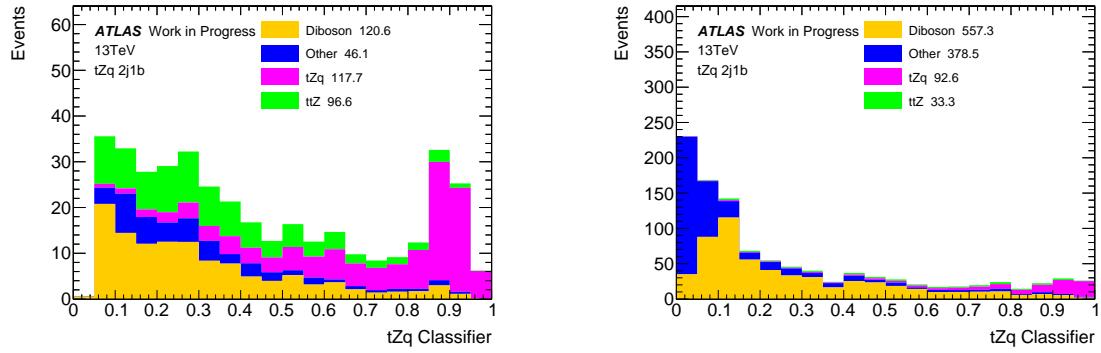


Figure 5.11: Comparison of the stacked plots of the default (left) and loose (right) classifier in the 2j1b SR.

As explained in section 4.2.3, in a binary classifier every single event will receive a score between [0,1]. Therefore looking at figure 5.12 I determined the threshold score with the highest $S/\sqrt{S + B}$ significance. Table 5.8 shows the max $S/\sqrt{S + B}$ value and its threshold score.

	max $S/\sqrt{S + B}$	Score
Default	7.82	0.58
Loose	6.09	0.82

Table 5.8: Maximum values of $S/\sqrt{S + B}$ for the default and loose binary classifier in the 2j1b SR.

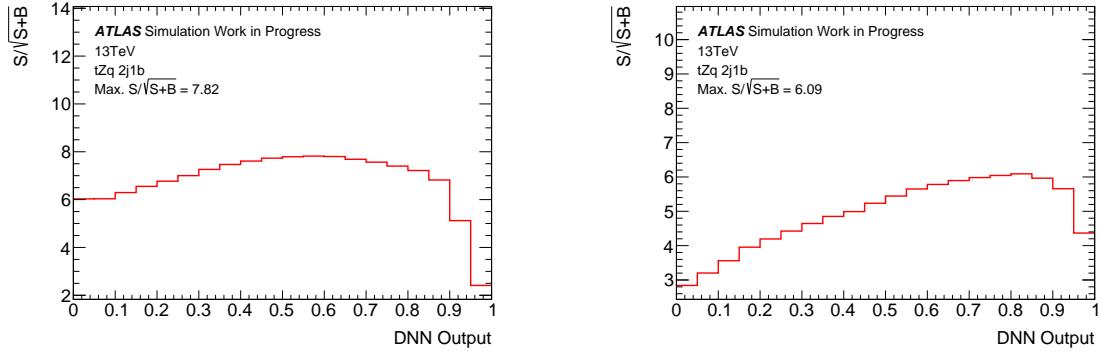


Figure 5.12: Comparison of the $S/\sqrt{S + B}$ plots for the default (left) and loose (right) binary classifier in the 2j1b SR.

Table 5.8 shows that the default classifier has a higher max $S/\sqrt{S + B}$ significance

than the loose classifier, even though its performance (as shown by the confusion matrices) is significantly worse. This might be due to the fact that the loose selection has many more background events than the default selection.

Multi-class classifier: signal, diboson, other

The learning rate of the default and loose classifier is 0.0005 and 0.0004 respectively. The confusion matrices show that the loose classifier performs better than the default classifier across all three classes (figure 5.13).

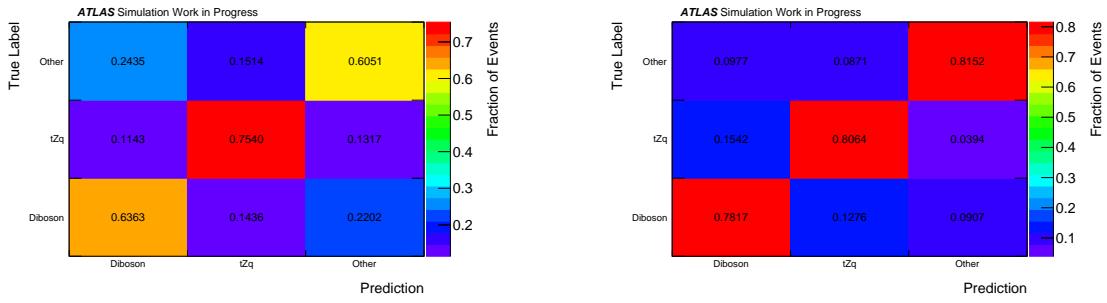


Figure 5.13: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 2j1b signal region.

The separations for each class are summarized in table 5.9 and the distributions of the classifiers are shown in figure 5.14. The loose classifiers have a better separation across all classes and the tZq class has the best for both the default and loose selection.

	tZq	Diboson	Other
Default	44.47 %	28.32 %	27.04 %
Loose	57.08 %	52.94 %	59.91 %

Table 5.9: Separation for the tZq , diboson and other class in the 2j1b signal region.

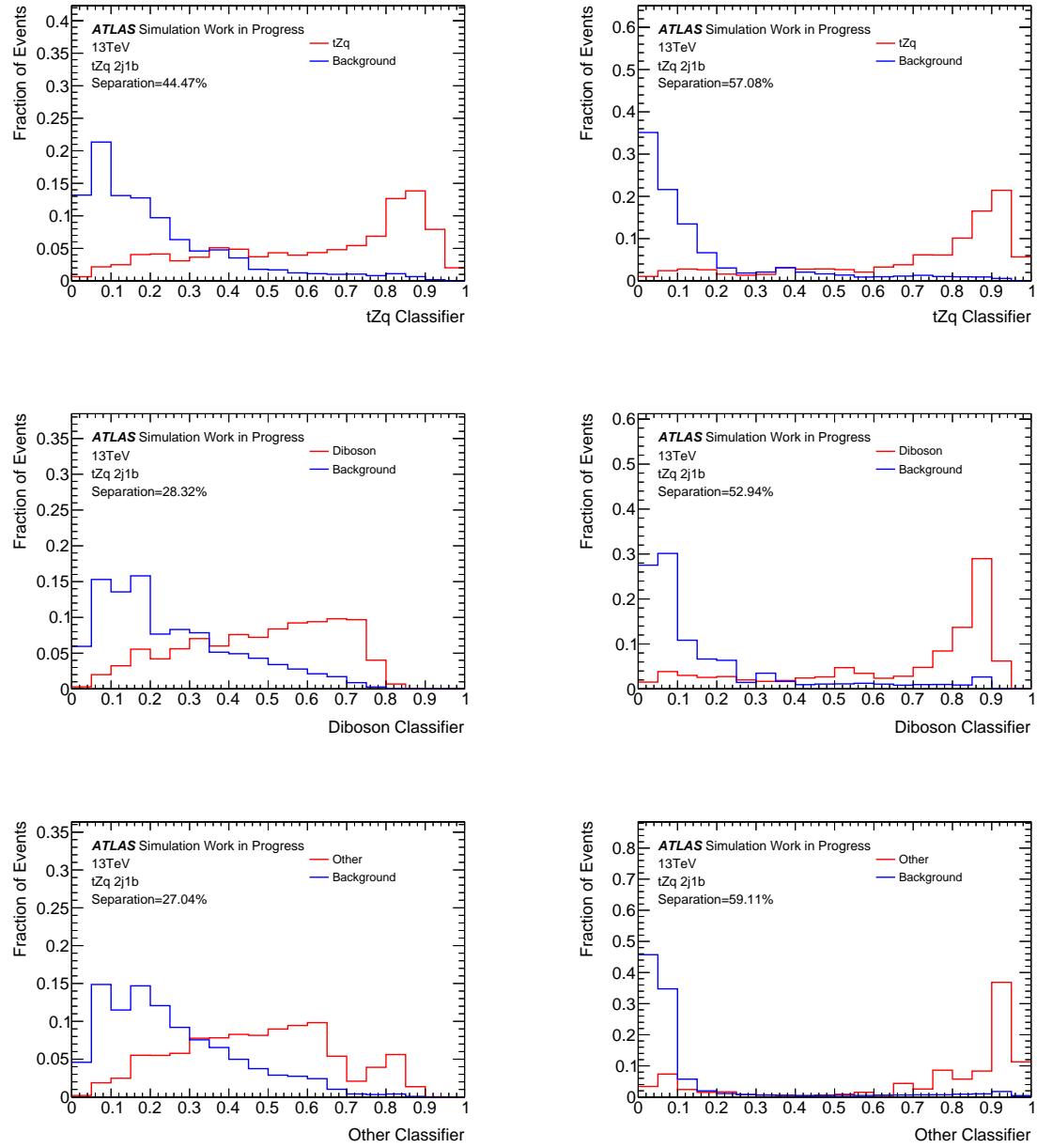


Figure 5.14: Comparison of the normalized distributions of the classifiers (tZq , diboson, other) of the default (left) and loose (right) selection in the $2j1b$ signal region.

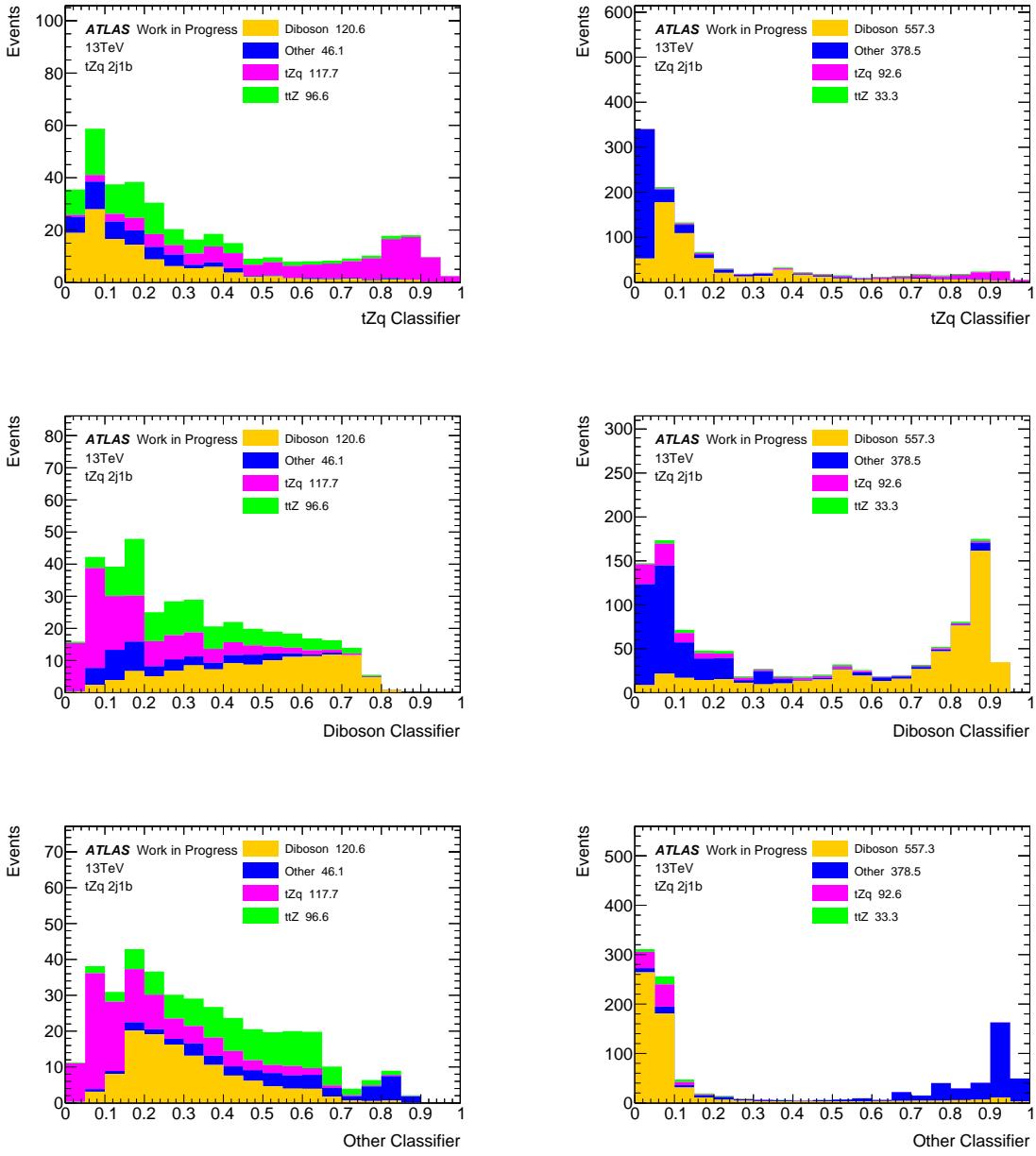


Figure 5.15: Comparison of the stacked plots for the classifiers (tZq , diboson, other) of the default (left) and loose (right) selection in the $2j1b$ signal region.

The *response* plots (figure 5.16) show the scores assigned to a particular class of events by the diboson and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The comparison between the response plots of both classifiers shows that the loose classifier isolates better than the default classifier the response of

each class of events. Therefore the loose classifier allows a purer definition of the signal and control region.

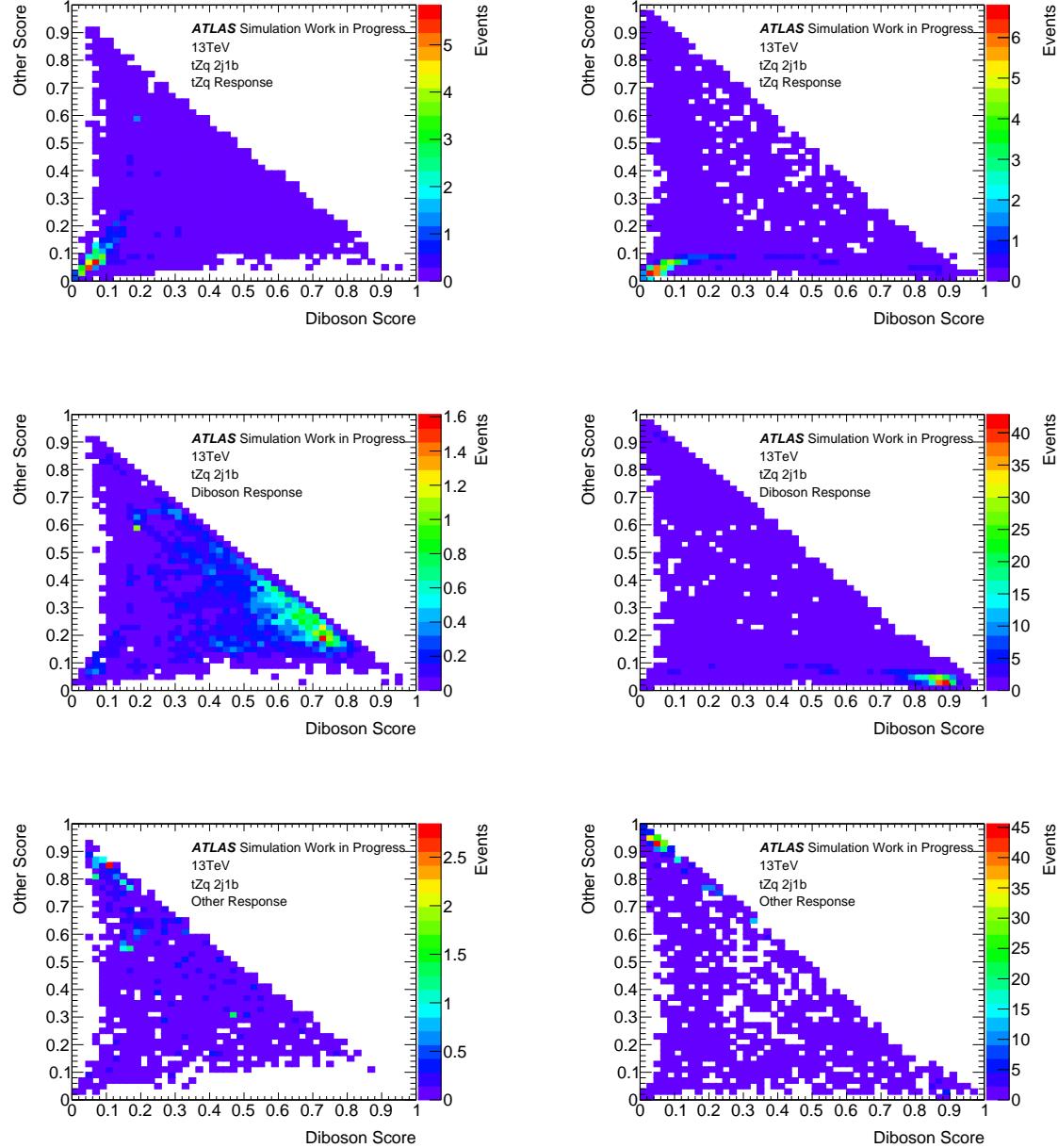


Figure 5.16: Comparison of the 2D plots showing the signal response in the nodes associated to the diboson and other class in the 2j1b signal region.

Therefore looking at figures 5.17 and 5.18 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the

region with the highest signal significance.

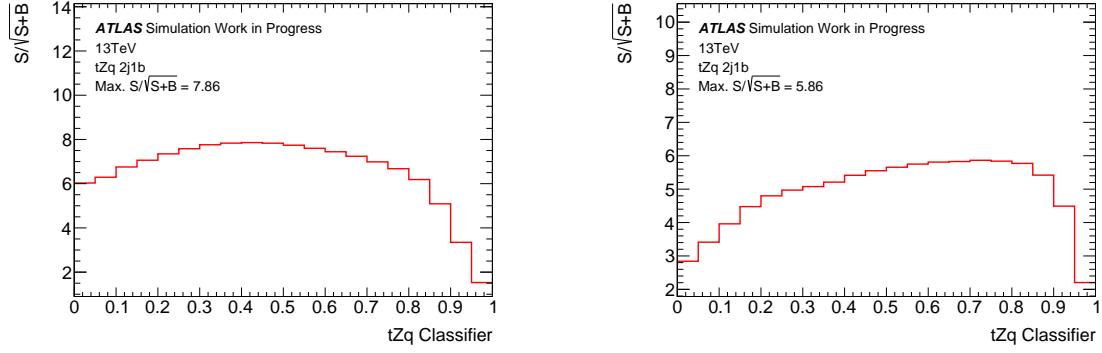


Figure 5.17: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) tZq classifier in the $2j1b$ SR.

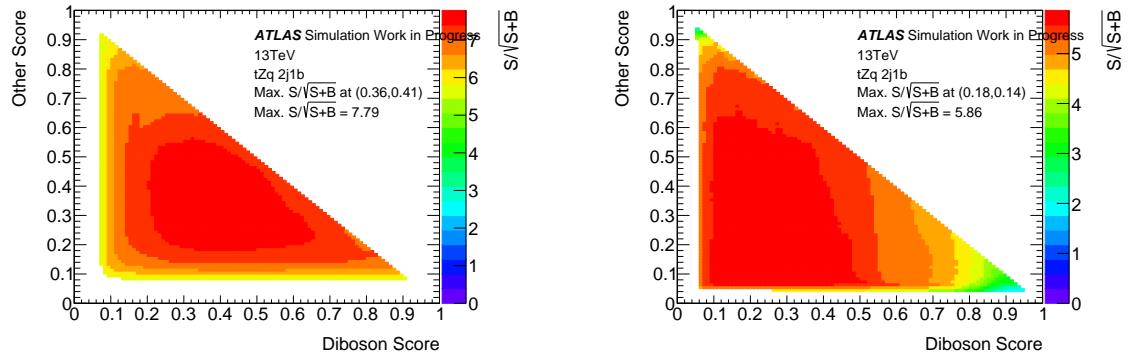


Figure 5.18: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the diboson and other nodes for the default (left) and loose (right) classifier in the $2j1b$ signal region.

Table 5.10 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	Diboson Score	Other Score
Default	7.79	0.36	0.41
Loose	5.86	0.18	0.14
	Max $S/\sqrt{S+B}$	tZq Score	
Default	7.86	0.42	
Loose	5.86	0.72	

Table 5.10: Maximum values of $S/\sqrt{S+B}$ for the default and loose diboson multi-class classifier in the 2j1b SR.

Multi-class classifier: signal, $t\bar{t}Z$, other

The learning rate of the default and loose classifier is 0.0002 and 0.0002 respectively. The confusion matrices show that the loose classifier performs better than the default classifier for the *Other* and $t\bar{t}Z$ class, whereas their performance is almost the same for the tZq class (figure 5.19).

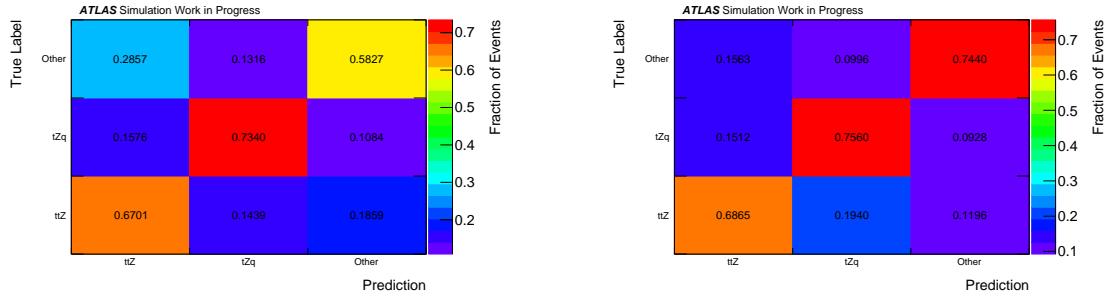


Figure 5.19: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 2j1b SR.

The separation for each class are summarized in table 5.11 and the distributions of the classifiers are shown in figure 5.20. The loose classifiers have a better separation across all classes and the tZq class has the best for both the default and loose selection.

	tZq	$t\bar{t}Z$	Other
Default	44.08 %	26.46 %	30.84 %
Loose	54.20 %	40.48 %	53.99 %

Table 5.11: Separation efficiencies for the tZq , $t\bar{t}Z$ and other class in the 2j1b SR.

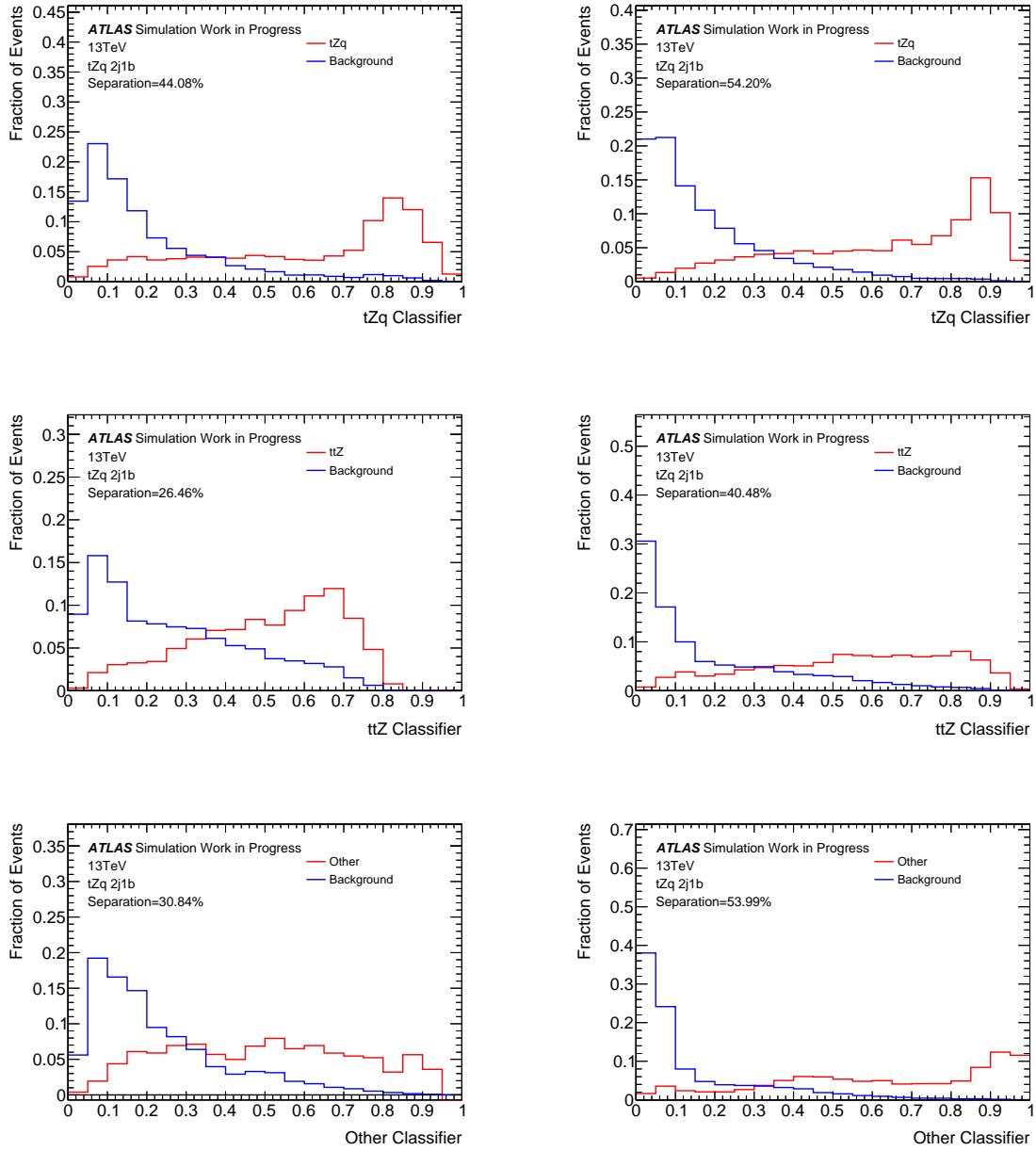


Figure 5.20: Comparison of the normalized distributions of the classifiers (tZq , $t\bar{t}Z$, other) of the default (left) and loose (right) selection in the 2j1b SR.

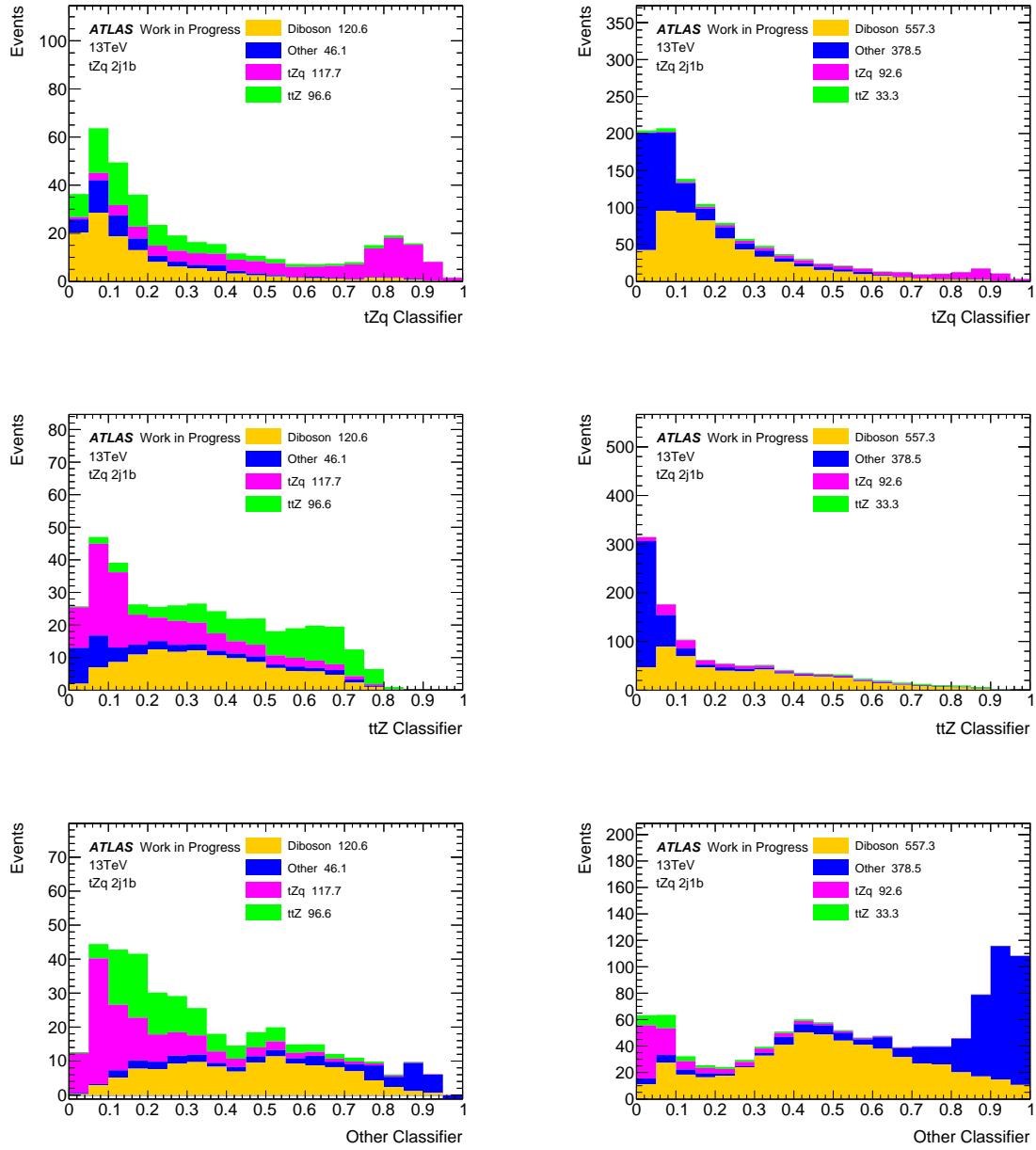


Figure 5.21: Comparison of the stacked plots for the classifiers (tZq , $t\bar{t}Z$, other) of the default (left) and loose (right) selection in the 2j1b SR.

The *response* plots (figure 5.22) show the scores assigned to a particular class of events by the $t\bar{t}Z$ and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The response plots of both classifiers show that the tZq signal is clustered in the bottom left corner, where the scores of the $t\bar{t}Z$ and other node are below 0.2.

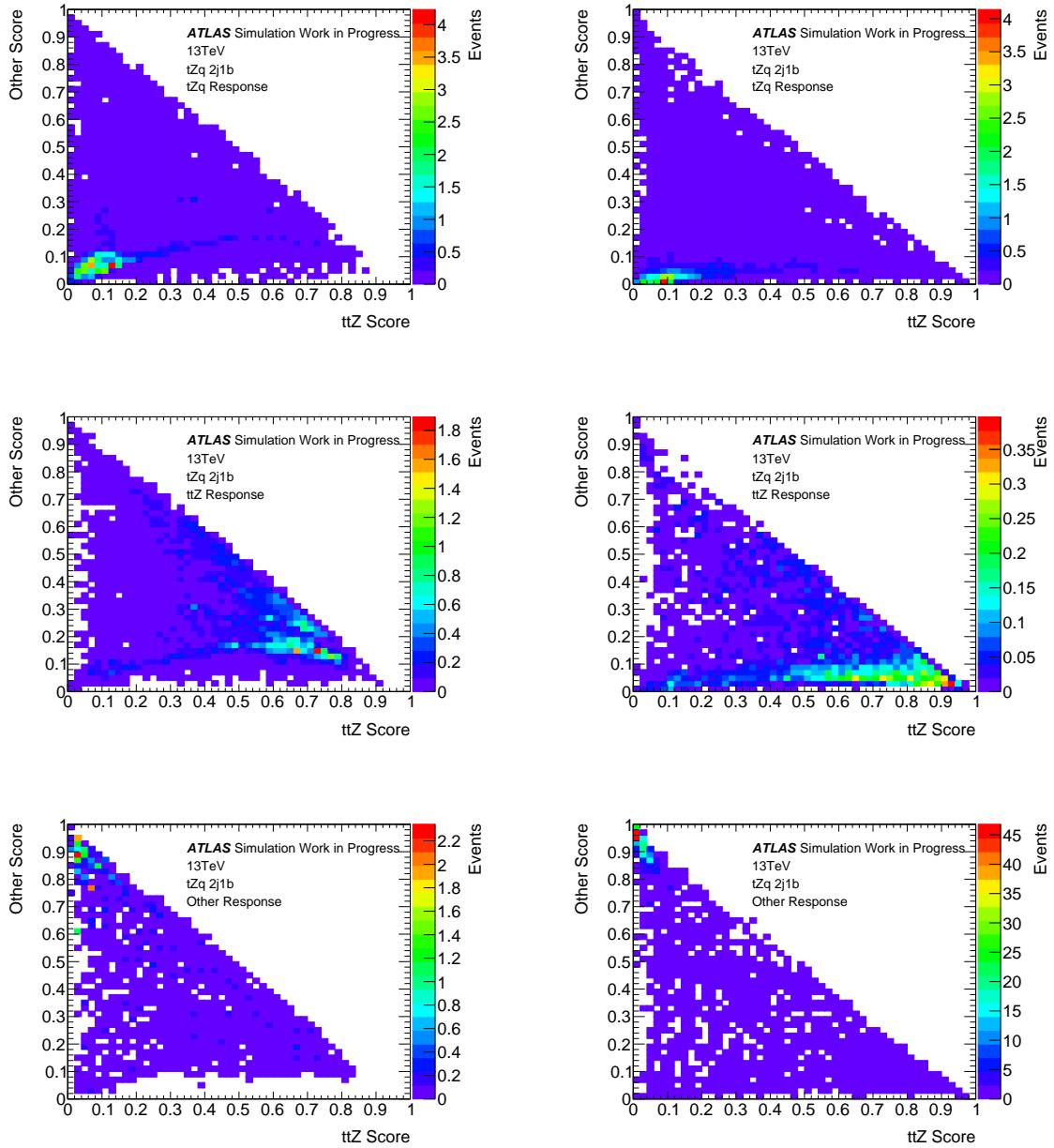


Figure 5.22: 2D plots showing the event (tZq , $t\bar{t}Z$, other) response in the $t\bar{t}Z$ and other nodes in the 2j1b SR.

Therefore looking at figures 5.23 and 5.24 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the region with the highest signal significance.

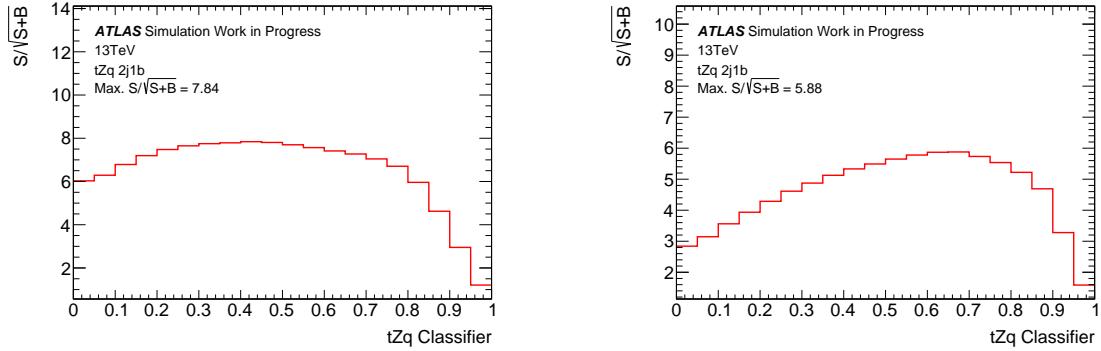


Figure 5.23: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) $t\bar{Z}q$ classifier in the 2j1b SR.

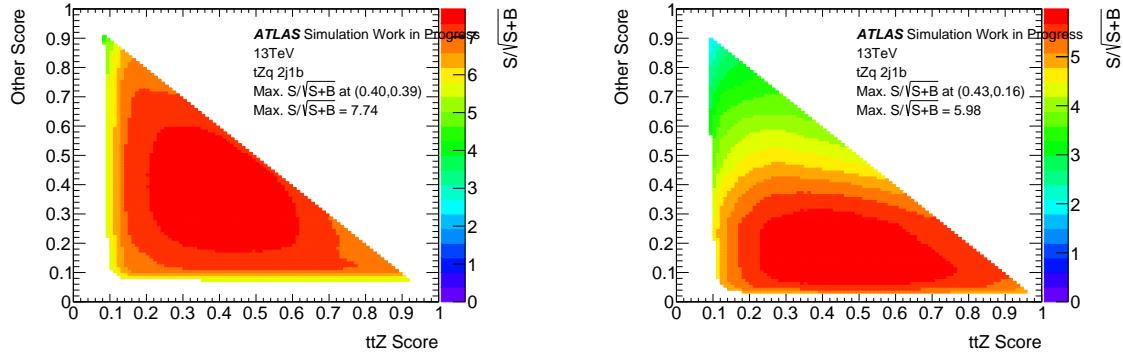


Figure 5.24: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the $t\bar{Z}Z$ and other nodes for the default (left) and loose (right) classifier in the 2j1b SR.

Table 5.12 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	$t\bar{Z}Z$ Score	Other Score
Default	7.74	0.40	0.39
Loose	5.98	0.43	0.16
	Max $S/\sqrt{S+B}$		tZq Score
Default	7.84		0.42
Loose	5.88		0.68

Table 5.12: Maximum values of $S/\sqrt{S+B}$ for the default and loose $t\bar{Z}Z$ multi-class classifier in the 2j1b SR.

5.5.3 Trainings for the 3j1b signal region

This paragraph shows and compares the trainings of the 3j1b signal region for both selections (default and loose). Throughout this paragraph I will refer to the binary and multi-class classifiers of the default and loose selections as the *default* and *loose* classifier. The main features of each architecture have been previously described in section 5.5.

Binary classifier

The learning rate of the default and loose classifier is 0.0005 and 0.0004 respectively. The confusion matrices show that the loose classifier performs better than the default classifier (figure 5.25).

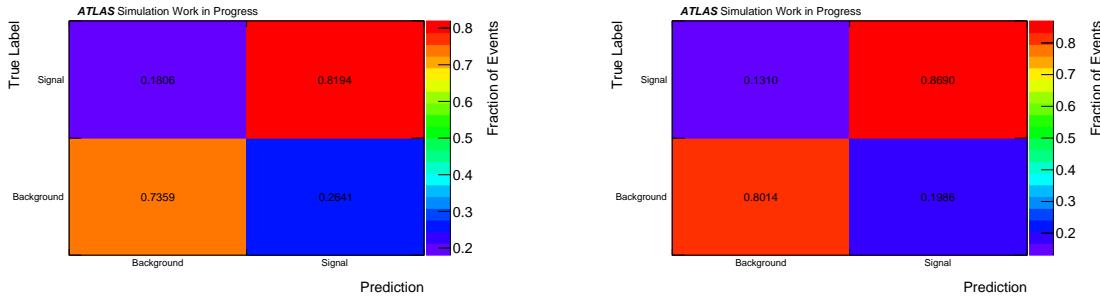


Figure 5.25: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 3j1b SR.

Moreover the separation plots show that the default and loose classifier have a separation of 39.15 % and 53.03 % respectively (figure 5.26). Figure 5.26 shows that the default classifier has more overlap in the central region than the loose classifier, which causes a lower separation.

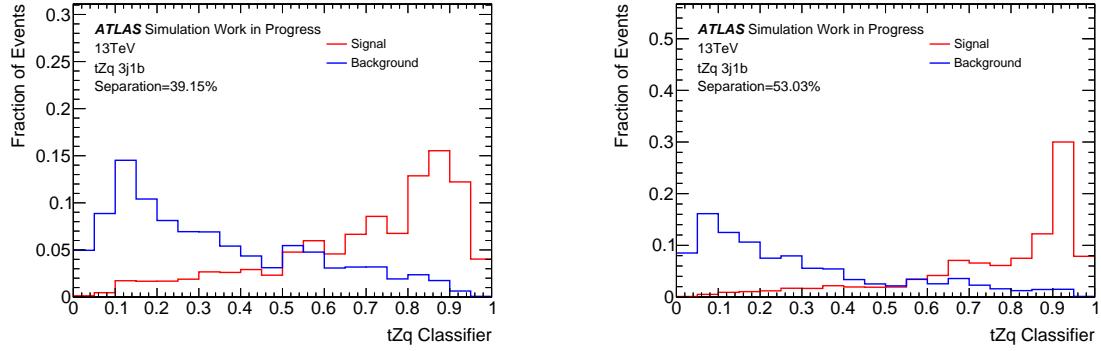


Figure 5.26: Comparison of the normalized distributions of the default (left) and loose (right) classifier in the 3j1b SR.

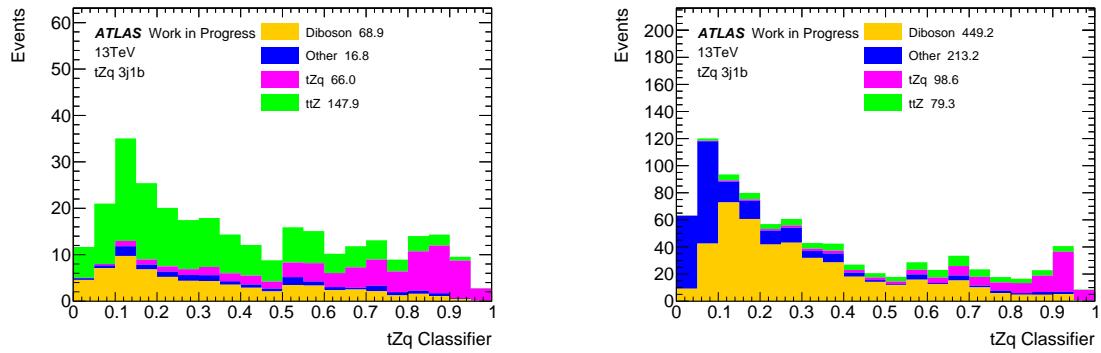


Figure 5.27: Comparison of the stacked plots of the default (left) and loose (right) classifier in the 3j1b SR.

As explained in section 4.2.3, in a binary classifier every single event will receive a score between [0,1]. Therefore looking at figure 5.28 I determined the threshold score with the highest $S/\sqrt{S+B}$ significance. Table 5.13 shows the max $S/\sqrt{S+B}$ value per classifier and its threshold score.

	max $S/\sqrt{S+B}$	Score
Default	5.10	0.62
Loose	6.09	0.78

Table 5.13: Maximum values of $S/\sqrt{S+B}$ for the default and loose binary classifier in the 3j1b SR.

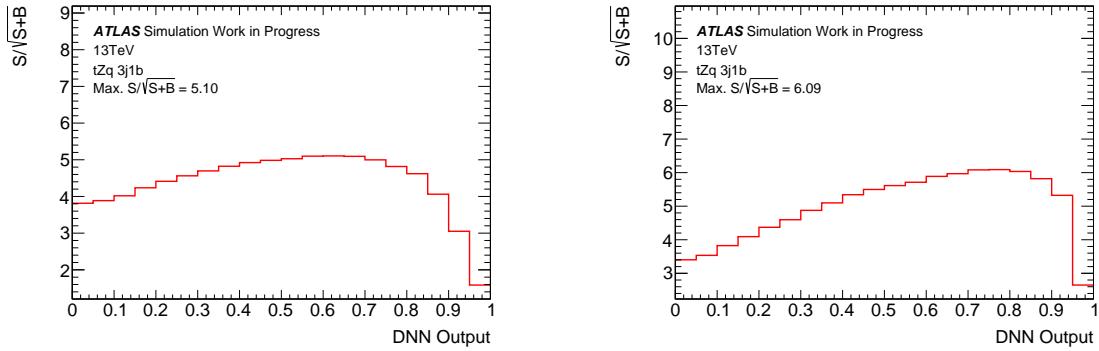


Figure 5.28: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) binary classifier in the 3j1b SR.

Unlike the binary training in the 2j1b SR, table 5.13 shows that the loose classifier has an higher max $S/\sqrt{S+B}$ significance than the default classifier.

Multi-class classifier: signal, diboson, other

The learning rate of the default and loose classifier is 0.0003 and 0.0001 respectively. The confusion matrices show that the loose classifier performs better than the default classifier across all three classes (figure 5.29).

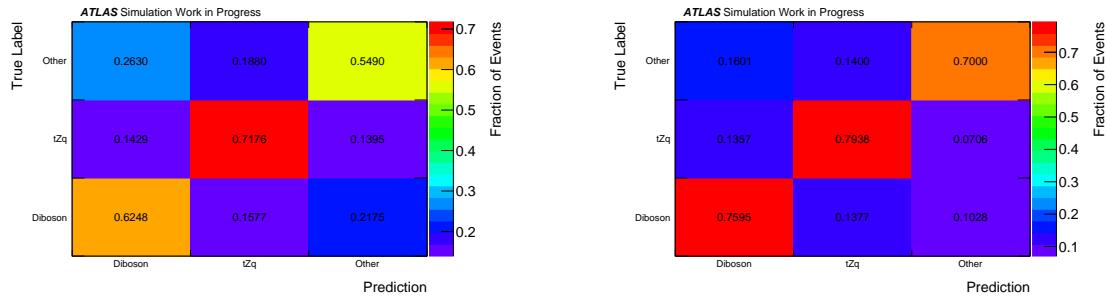


Figure 5.29: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 3j1b signal region.

The separation for each class are summarized in table 5.14 and the separation plots are shown in figure 5.30. The loose classifiers have a better separation across all classes and the tZq class has the best for both the default and loose selection.

	tZq	Diboson	Other
Default	37.11 %	21.28 %	21.60 %
Loose	51.43 %	45.48 %	46.32 %

Table 5.14: Separation for the tZq, diboson and other class in the 3j1b signal region.

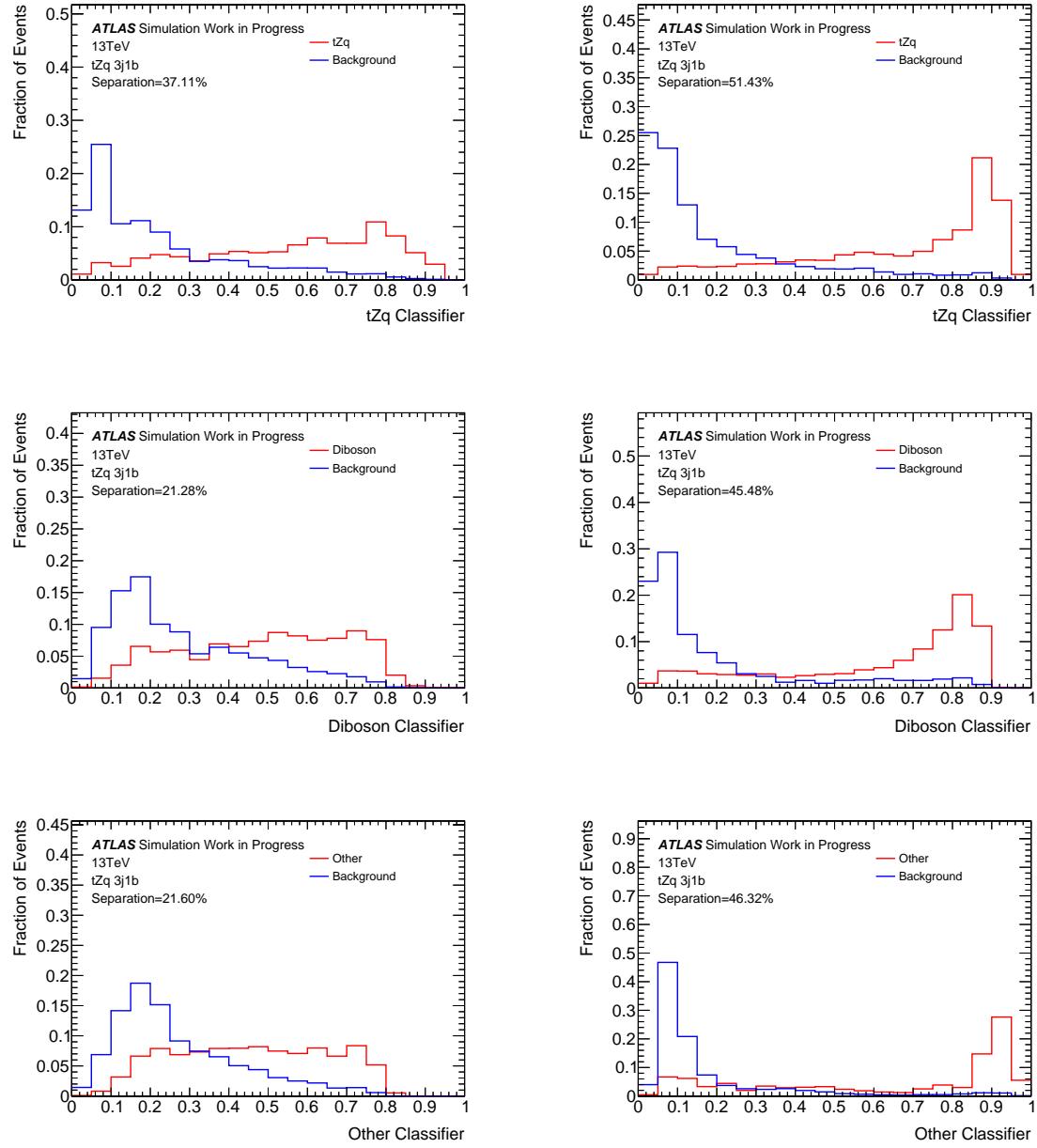


Figure 5.30: Comparison of the normalized distributions of the classifiers (tZq, diboson, other) of the default (left) and loose (right) selection in the 3j1b signal region.

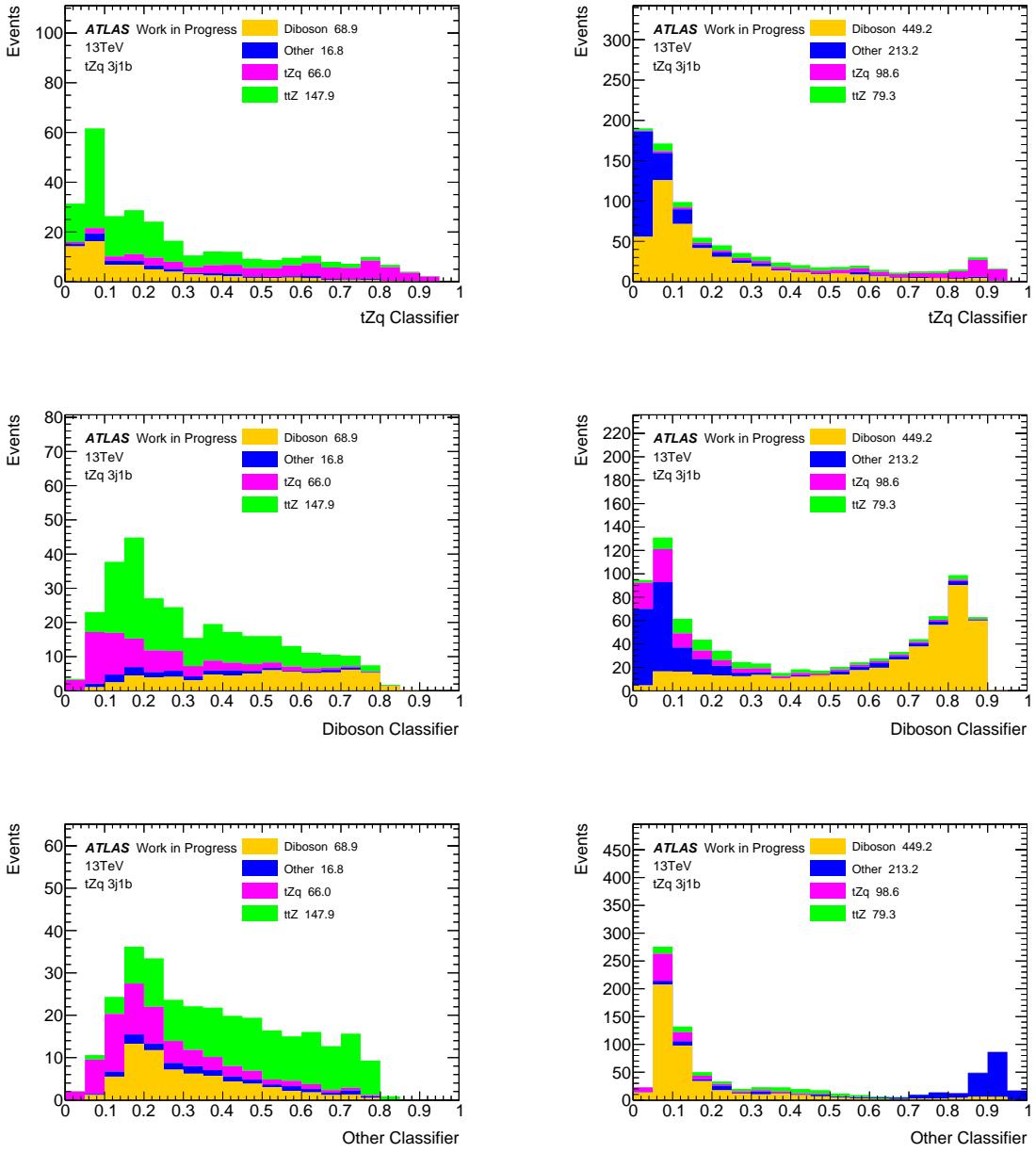


Figure 5.31: Comparison of the stacked plots for the classifiers (tZq , diboson, other) of the default (left) and loose (right) selection in the $3j1b$ signal region.

The *response* plots (figure 5.32) show the scores assigned to a particular class of events by the diboson and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The comparison between the response plots of both classifiers shows that the loose classifier isolates better than the default classifier the response of

each class of events. Therefore the loose classifier allows a more precise definition of the signal and control region.

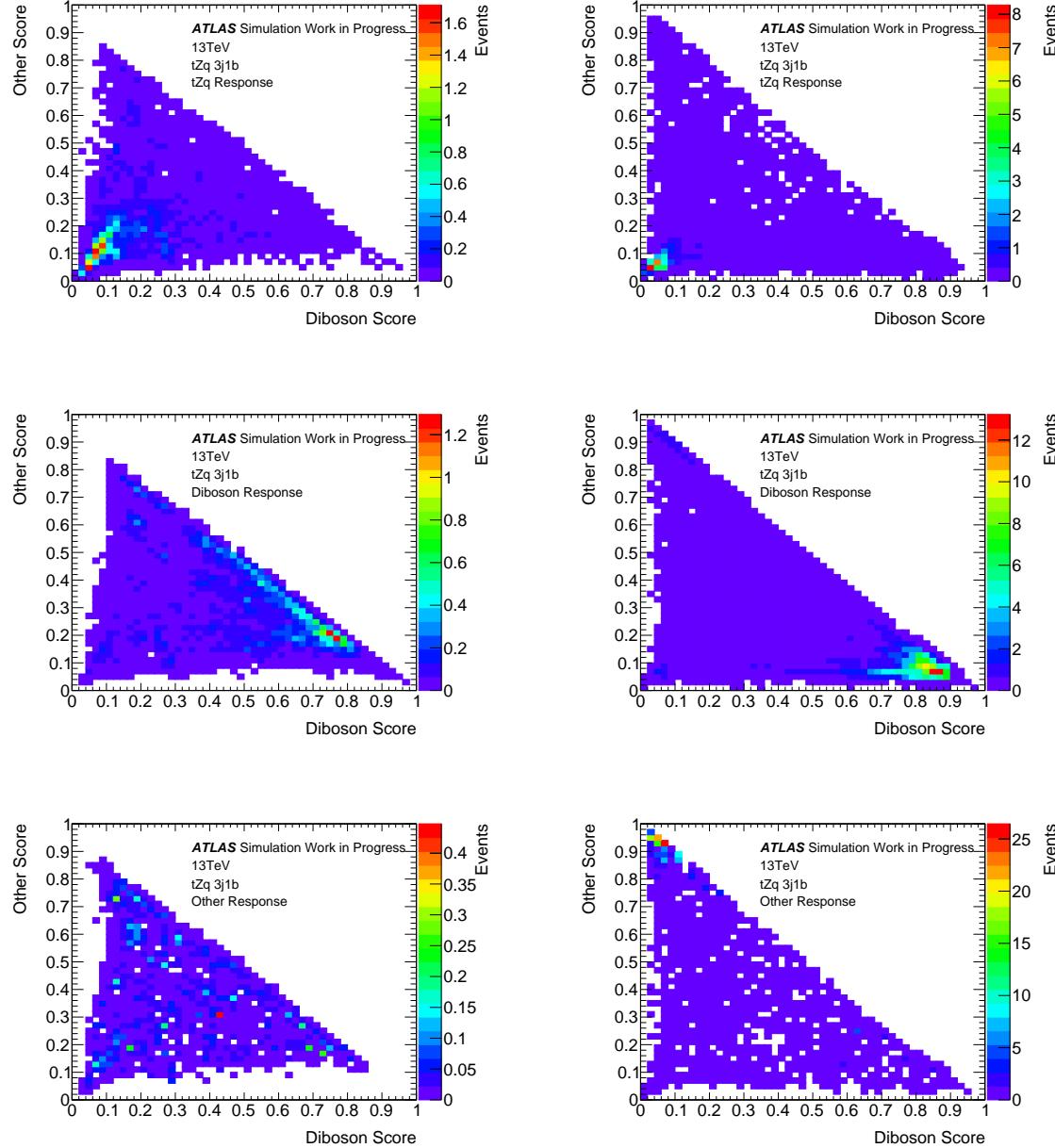


Figure 5.32: Comparison of the 2D plots showing the signal response in the nodes associated to the diboson and other class in the nj1b signal region.

Therefore looking at figures 5.33 and 5.34 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the

region with the highest signal significance.

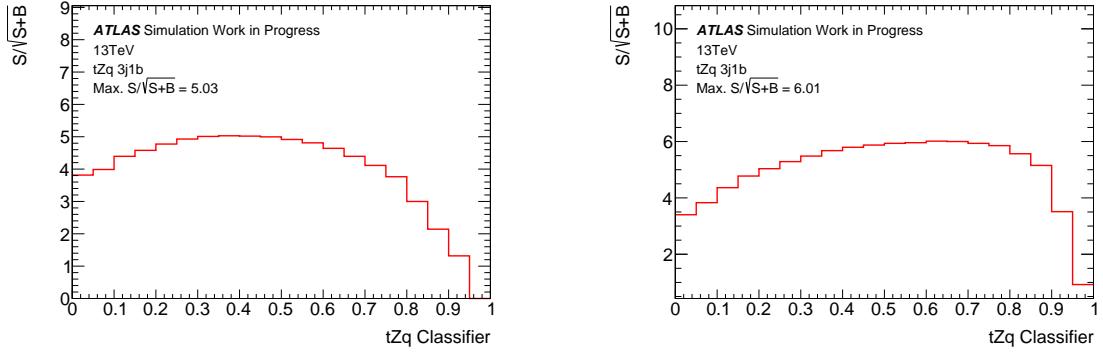


Figure 5.33: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) tZq classifier in the 3j1b SR.

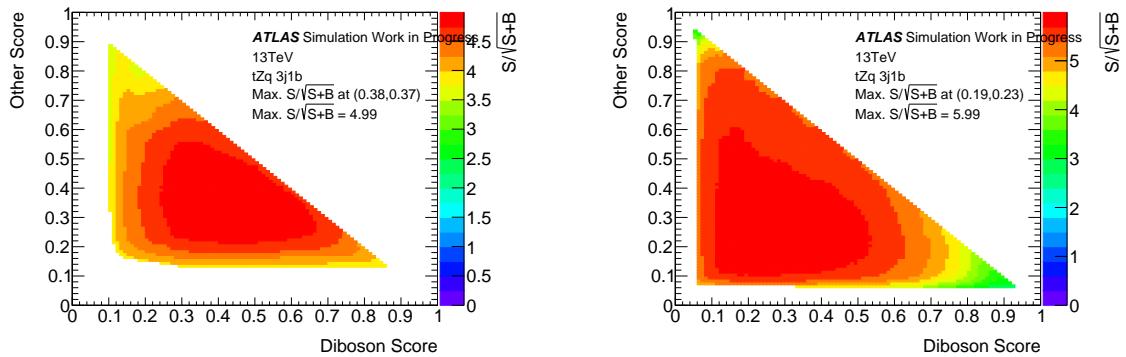


Figure 5.34: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the diboson and other nodes for the default (left) and loose (right) classifier in the 3j1b signal region.

Table 5.20 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	Diboson Score	Other Score
Default	4.99	0.38	0.37
Loose	5.99	0.19	0.23
	Max $S/\sqrt{S+B}$	tZq Score	
Default	5.03	0.38	
Loose	6.01	0.62	

Table 5.15: Maximum values of $S/\sqrt{S+B}$ for the default and loose diboson multi-class classifier in the 2j1b SR.

Multi-class classifier: signal, $t\bar{t}Z$, other

The learning rate of the default and loose classifier is 0.0001 and 0.0002 respectively. The confusion matrices show that the loose classifier performs better than the default classifier for the *Other* and $t\bar{t}Z$ class, whereas their performance is almost the same for the tZq class (figure 5.35).

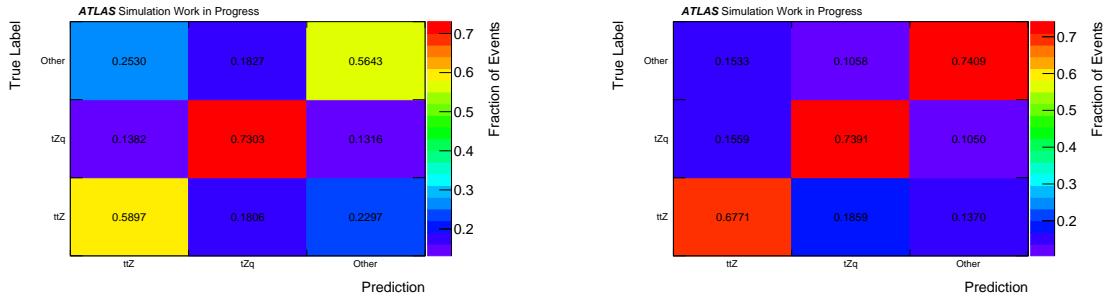


Figure 5.35: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the 3j1b signal region.

The separation for each class are summarized in table 5.16 and the separation plots are shown in figure 5.36. The loose classifier has a better separation across all classes than the default classifier.

	tZq	$t\bar{t}Z$	Other
Default	38.42 %	23.25 %	29.90 %
Loose	49.05 %	37.62 %	48.39 %

Table 5.16: Separation for the tZq , $t\bar{t}Z$ and other class in the 3j1b signal region.

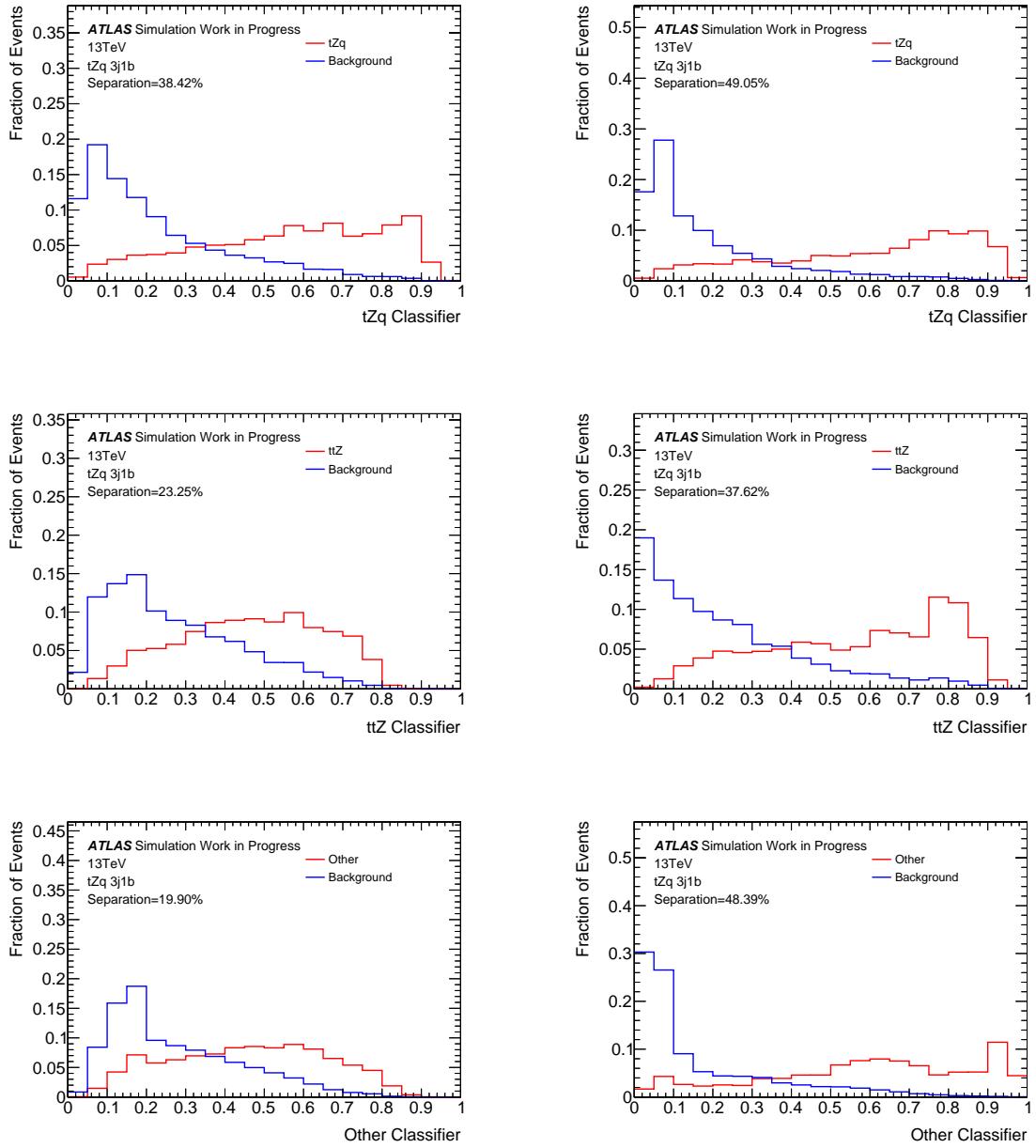


Figure 5.36: Comparison of the normalized distributions of the classifiers (tZq , $t\bar{t}Z$, other) of the default (left) and loose (right) selection in the 3j1b signal region.

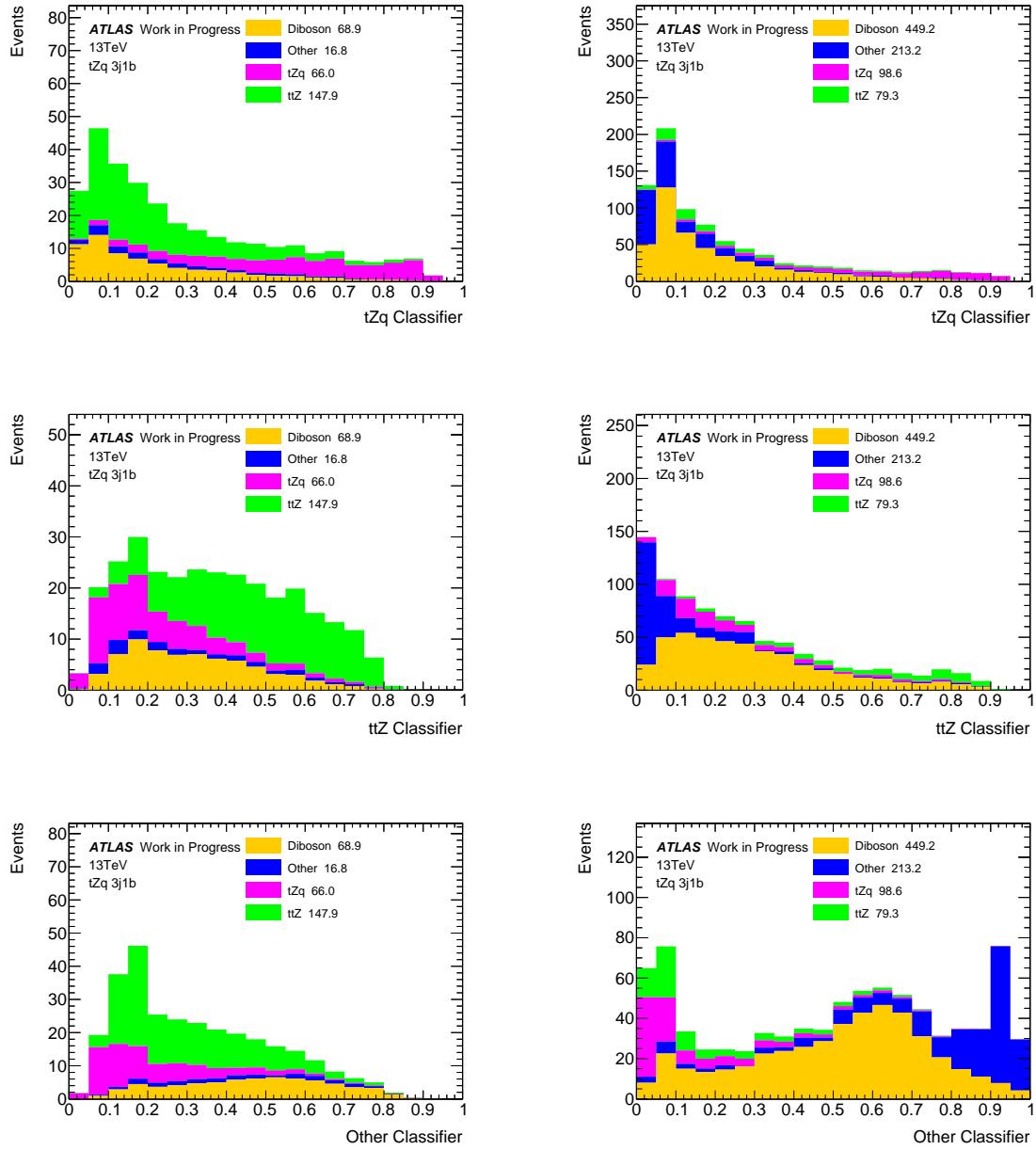


Figure 5.37: Comparison of the stacked plots of the classifiers (tZq , $t\bar{Z}$, other) of the default (left) and loose (right) selection in the $3j1b$ signal region.

The *response* plots (figure 5.38) show the scores assigned to a particular class of events by the $t\bar{Z}$ and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The response plots of both classifiers show that the tZq signal is clustered in the bottom left corner, where the scores of the $t\bar{Z}$ and other node are below 0.2.

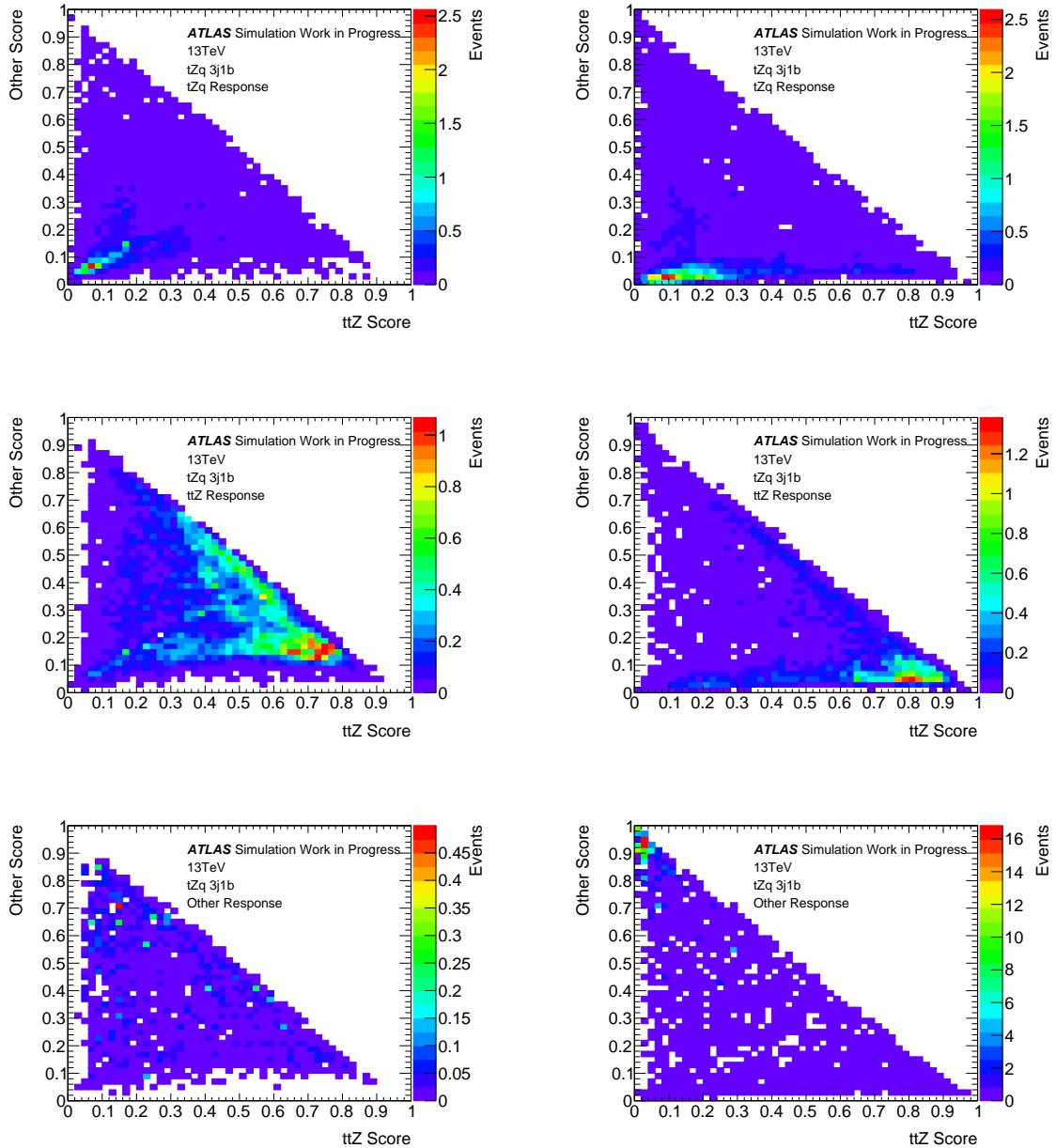


Figure 5.38: Comparison of the 2D plots showing the signal response in the nodes associated to the $t\bar{t}Z$ and other class in the 3j1b signal region.

Therefore looking at figures 5.40 and 5.39 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the region with the highest signal significance.

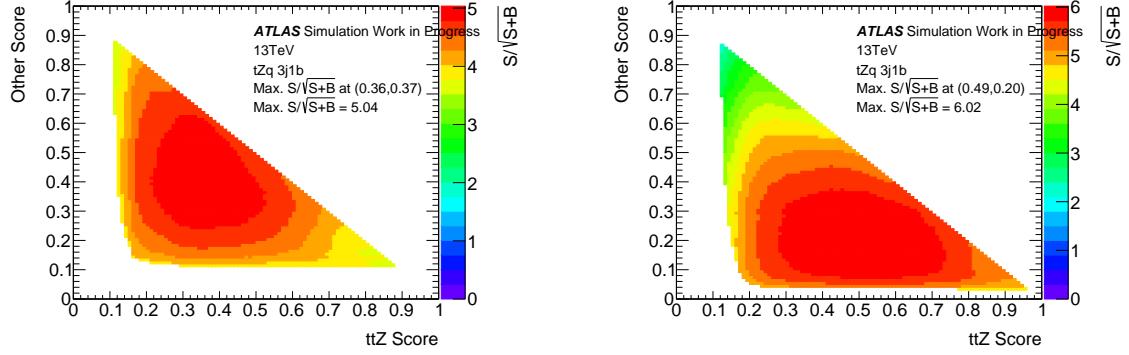


Figure 5.39: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the $t\bar{t}Z$ and other nodes for the default (left) and loose (right) classifier in the $3j1b$ signal region.

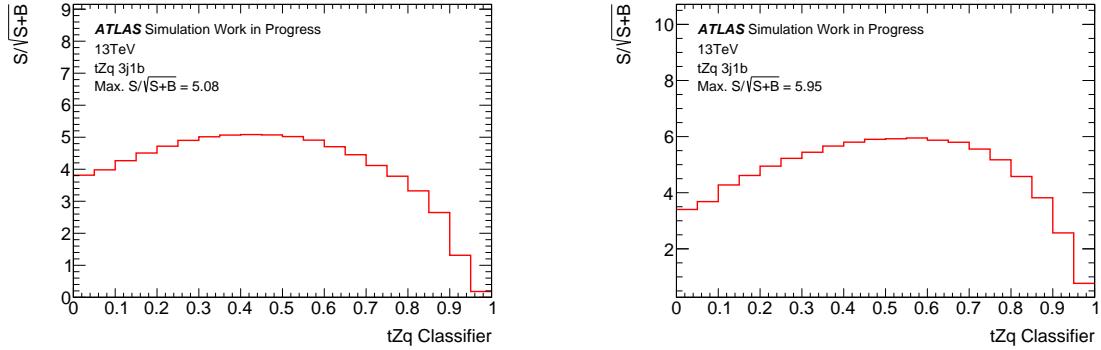


Figure 5.40: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) tZq classifier in the $3j1b$ SR.

Table 5.17 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	$t\bar{t}Z$ Score	Other Score
Default	5.04	0.36	0.37
Loose	6.02	0.49	0.20
	Max $S/\sqrt{S+B}$		tZq Score
Default	5.08		0.44
Loose	5.95		0.58

Table 5.17: Maximum values of $S/\sqrt{S+B}$ for the default and loose $t\bar{t}Z$ multi-class classifier in the $3j1b$ SR.

5.5.4 Trainings for the nj1b signal region

This paragraph shows and compares the trainings of the nj1b (2j1b and 3j1b) signal region for both selections (default and loose). Throughout this paragraph I will refer to the binary and multi-class classifiers of the default and loose selections as the *default* and *loose* classifier. The main features of each architecture have been previously described in section 5.5.

Binary classifier

The learning rate of the default and loose classifier is 0.0005 and 0.0002 respectively. Similarly to the previous two binary trainings (2j1b SR and 3j1b SR) the confusion matrices show that the loose classifier performs better than the default classifier (figure 5.41).

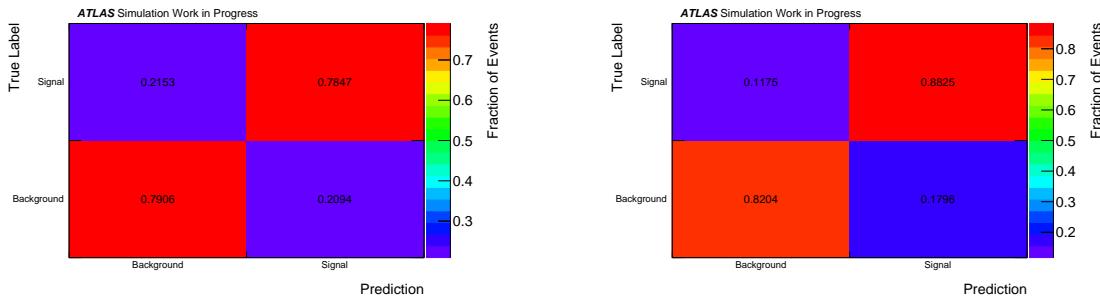


Figure 5.41: Comparison of the confusion matrices of the default (left) and loose (right) binary classifier in the nj1b SR.

Moreover the separation plots confirm that the loose classifier outperforms the default classifier: the default classifier has a separation of 40.60 %, whereas the loose classifier of 57.17 % (figure 5.42). Figure 5.42 shows that the background and signal distributions have a large overlap in the default classifier, which causes a lower separation with respect to the loose classifier.

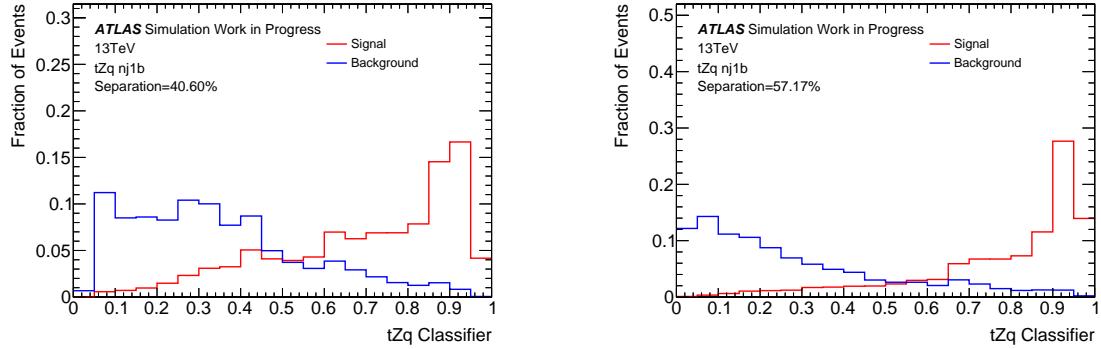


Figure 5.42: Comparison of the normalized distributions of the default (left) and loose (right) binary classifier in the nj1b SR.

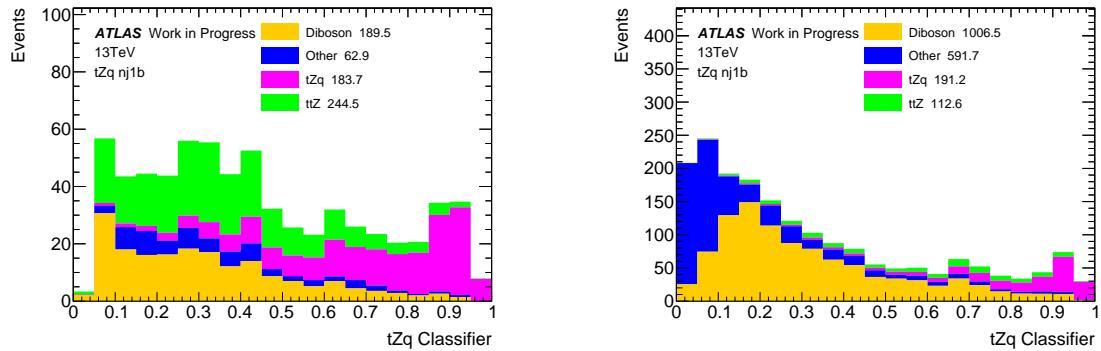


Figure 5.43: Comparison of the stacked plots of the default (left) and loose (right) binary classifier in the nj1b SR.

As explained in section 4.2.3, in a binary classifier every single event will receive a score between [0,1]. Therefore looking at figure 5.44 I determined the threshold score with the highest $S/\sqrt{S+B}$ significance. Table 5.18 shows the max $S/\sqrt{S+B}$ value and its threshold score.

	max $S/\sqrt{S+B}$	Score
Default	9.18	0.58
Loose	8.67	0.76

Table 5.18: Maximum values of $S/\sqrt{S+B}$ for the default and loose binary classifier in the nj1b SR.

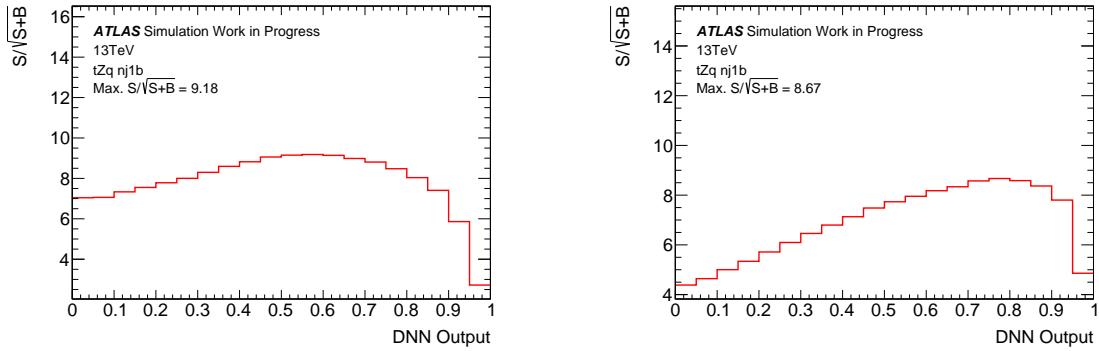


Figure 5.44: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) binary classifier in the nj1b SR.

Table 5.18 shows that the default classifier has a higher max $S/\sqrt{S+B}$ significance than the loose classifier, even though its performance (as shown by the confusion matrices) is significantly worse. This might be due to the fact that the loose selection has many more background events than the default selection, whereas both classifiers share almost the same number of signal events. Therefore, even if the loose classifier shows a significantly better performance than the default classifier, the overwhelming number of background events reduces the $S/\sqrt{S+B}$ significance. Meanwhile the default classifier has a slightly higher $S/\sqrt{S+B}$ value than the loose classifier (even if its performance is worst), because its selection has fewer background events.

Multi-class classifier: signal, diboson, other

The learning rate of the default and loose classifier is 0.0001 and 0.0004 respectively. The confusion matrices show that the loose classifier performs better than the default classifier across all three classes (figure 5.45).

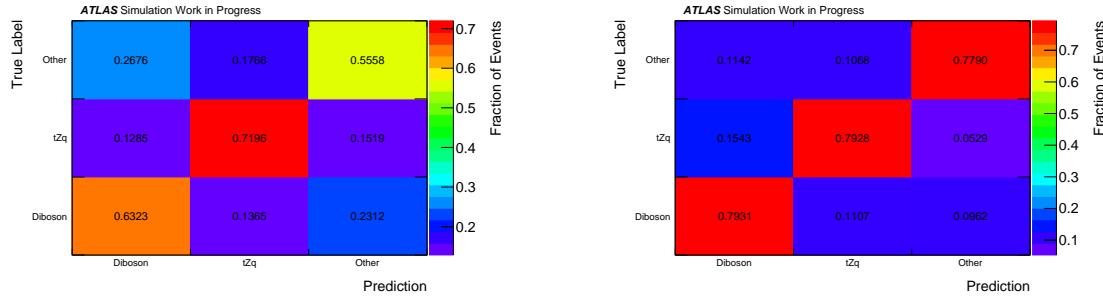


Figure 5.45: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the nj1b signal region.

The separation for each class are summarized in table 5.19 and the separation plots are shown in figure 5.46. The loose classifiers have a better separation across all classes than the default classifier.

	tZq	Diboson	Other
Default	38.72 %	24.26 %	20.46%
Loose	56.22 %	52.39 %	55.87 %

Table 5.19: Separation for the tZq, diboson and other class in the nj1b signal region.

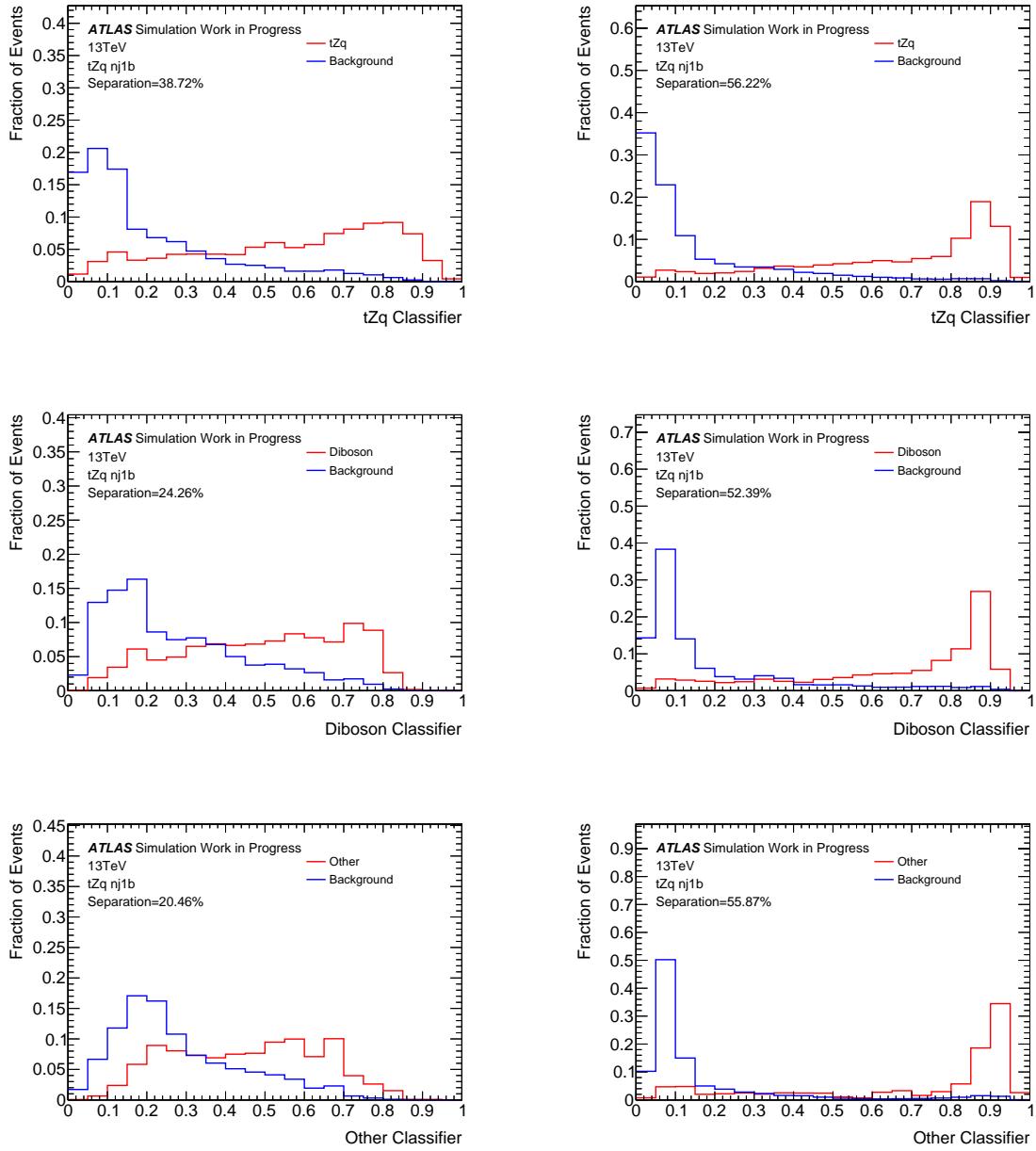


Figure 5.46: Comparison of the normalized distributions of the classifiers (tZq , diboson, other) of the default (left) and loose (right) selection in the $nj1b$ signal region.

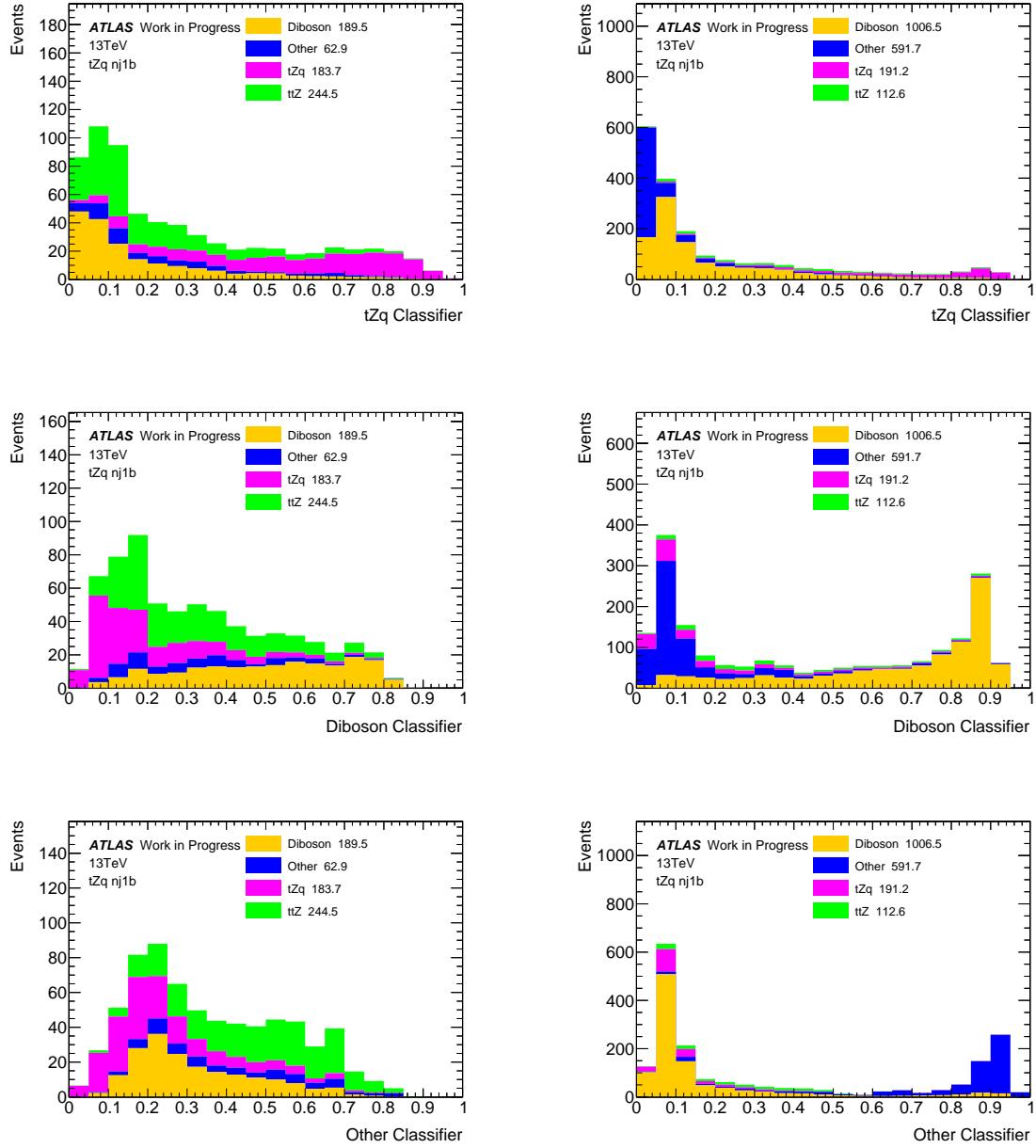


Figure 5.47: Comparison of the stacked plots for the classifiers (tZq , diboson, other) of the default (left) and loose (right) selection in the $nj1b$ signal region.

The *response* plots (figure 5.48) show the scores assigned to a particular class of events by the diboson and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The comparison between the response plots of both classifiers shows that the loose classifier isolates better than the default classifier the response of

each class of events. Therefore the loose classifier allows a more precise definition of the signal and control region.

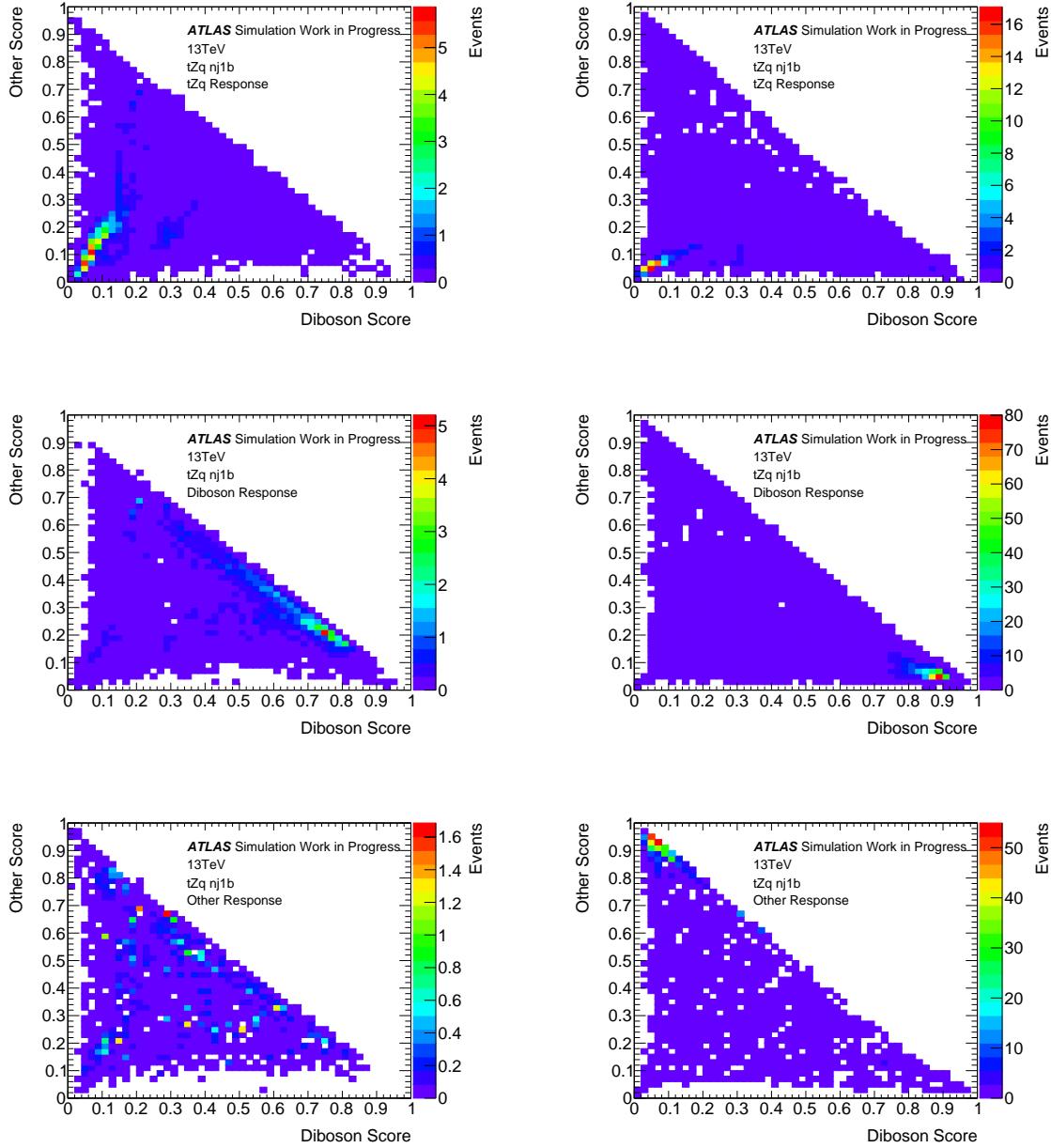


Figure 5.48: Comparison of the 2D plots showing the signal response in the nodes associated to the diboson and other class in the nj1b signal region.

Therefore looking at figures 5.49 and 5.50 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the

region with the highest signal significance.

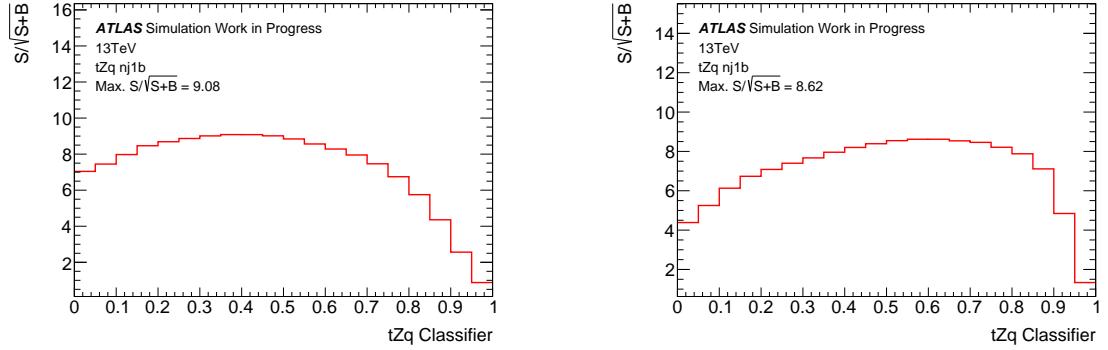


Figure 5.49: Comparison of the $S/\sqrt{S+B}$ plots for the default (left) and loose (right) tZq classifier in the $nj1b$ SR.

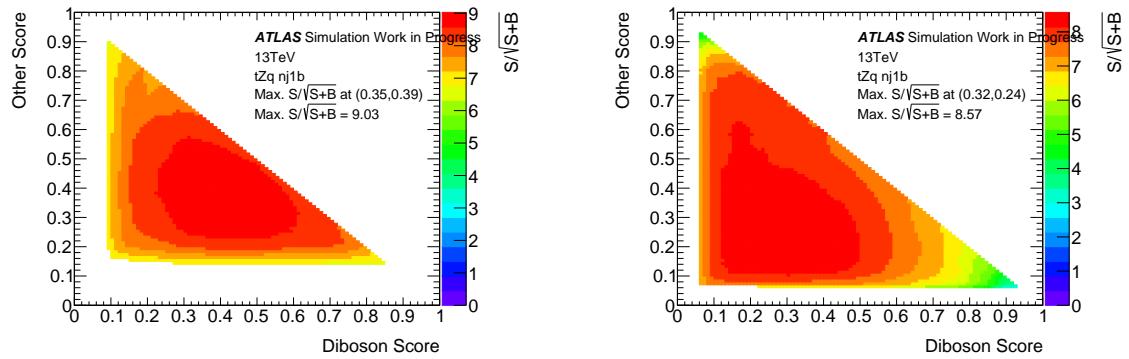


Figure 5.50: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the diboson and other nodes for the default (left) and loose (right) classifier in the $nj1b$ signal region.

Table 5.20 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	Diboson Score	Other Score
Default	9.03	0.35	0.39
Loose	8.57	0.32	0.24
	Max $S/\sqrt{S+B}$		
Default	9.08		0.4
Loose	8.62		0.6

Table 5.20: Maximum values of $S/\sqrt{S+B}$ for the default and loose diboson multi-class classifier in the $2j1b$ SR.

Multi-class classifier: signal, $t\bar{t}Z$, other

The learning rate of the default and loose classifier was 0.0002 and 0.0001 respectively. The confusion matrices show that the loose classifier performs better than the default classifier across all three classes (figure 5.51). Both classifiers have similar performances in distinguishing tZq events, whereas the loose classifier outperforms the default classifier in the other two classes of events.

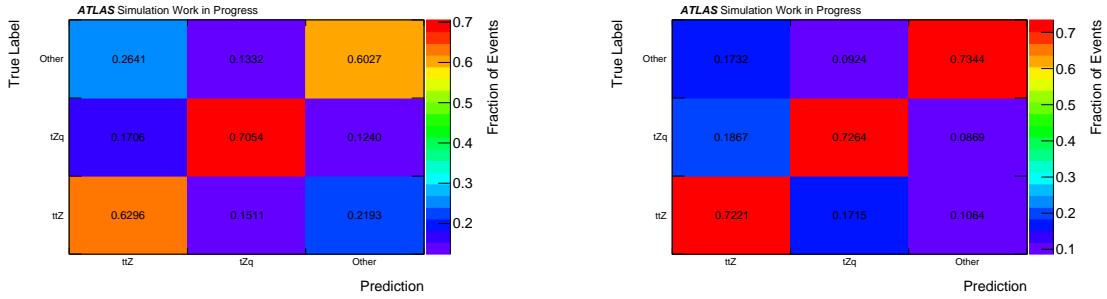


Figure 5.51: Comparison of the confusion matrices of the default (left) and loose (right) classifier in the $nj1b$ SR.

The separation for each class are summarized in table 5.21 and the separation plots are shown in figure 5.52. The loose classifier has a better separation than the default classifier across all classes.

	tZq	$t\bar{t}Z$	Other
Default	40.48 %	24.00 %	26.66 %
Loose	52.84 %	41.11 %	52.83 %

Table 5.21: Separation for the tZq , $t\bar{t}Z$ and other class in the $nj1b$ SR.

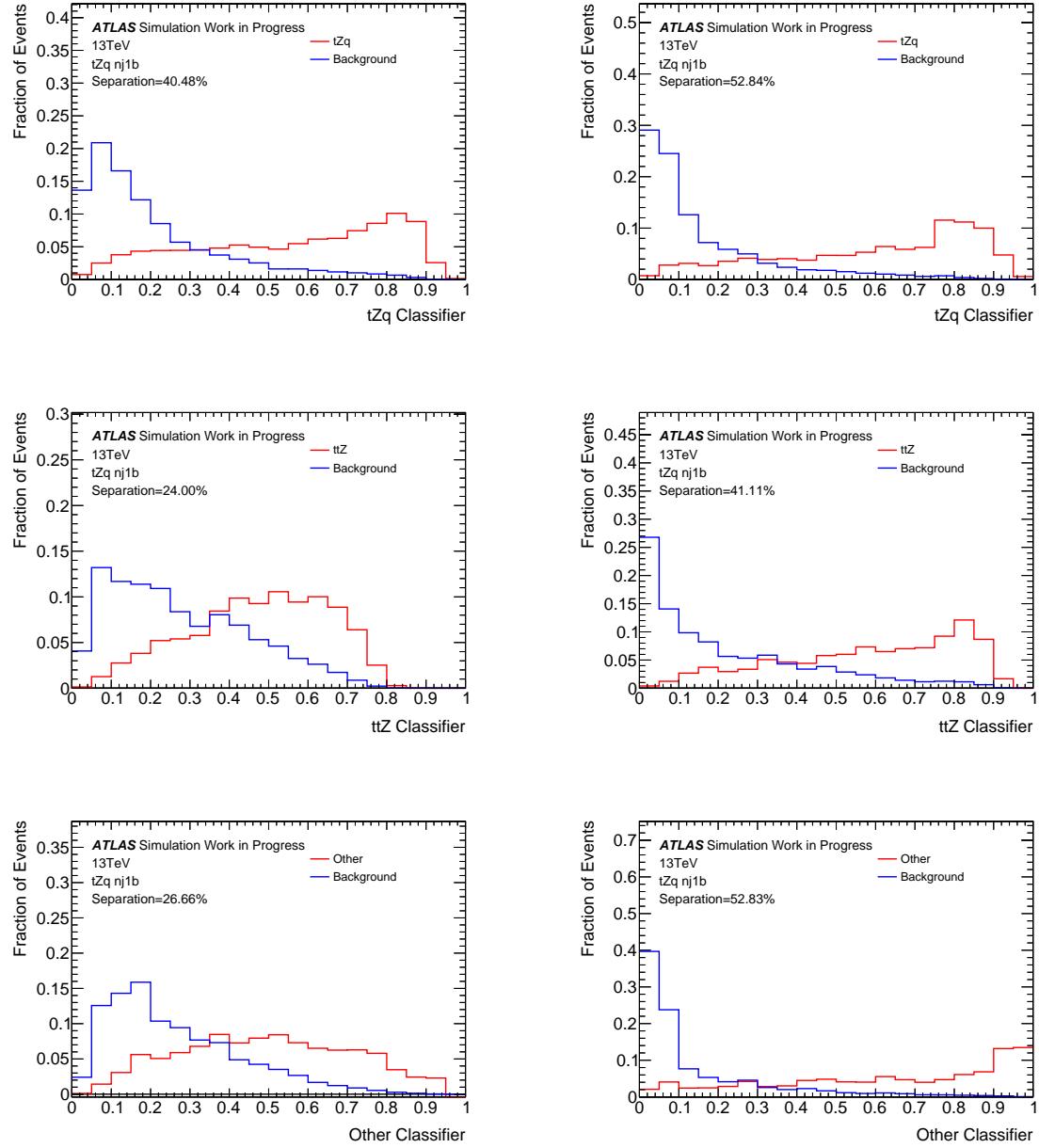


Figure 5.52: Comparison of the normalized distributions of the classifiers (tZq , $t\bar{t}Z$, other) of the default (left) and loose (right) classifier in the $nj1b$ signal region.

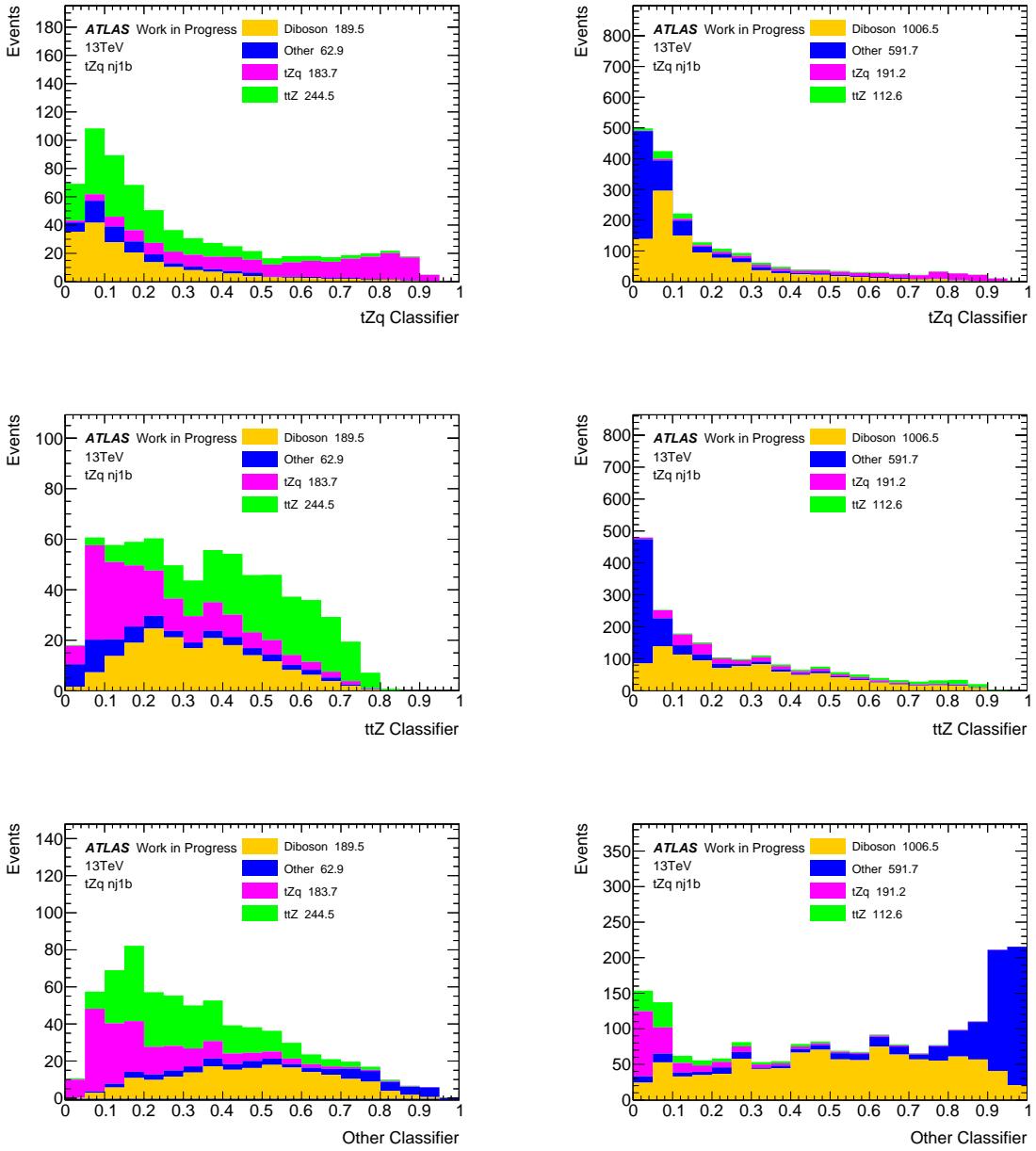


Figure 5.53: Comparison of the stacked plots for the classifiers (tZq , $t\bar{Z}$, other) of the default (left) and loose (right) selection in the $nj1b$ signal region.

The *response* plots (figure 5.54) show the scores assigned to a particular class of events by the $t\bar{Z}$ and other node. These plots are designed to give an idea on how to define the region of scores of a signal event (signal region). Furthermore the remaining area could be divided in sectors to define control regions for the processes of the other two classes. The response plots of both classifiers show that the tZq signal is clustered in the bottom left corner, where the scores of the $t\bar{Z}$ and other node are below 0.2.

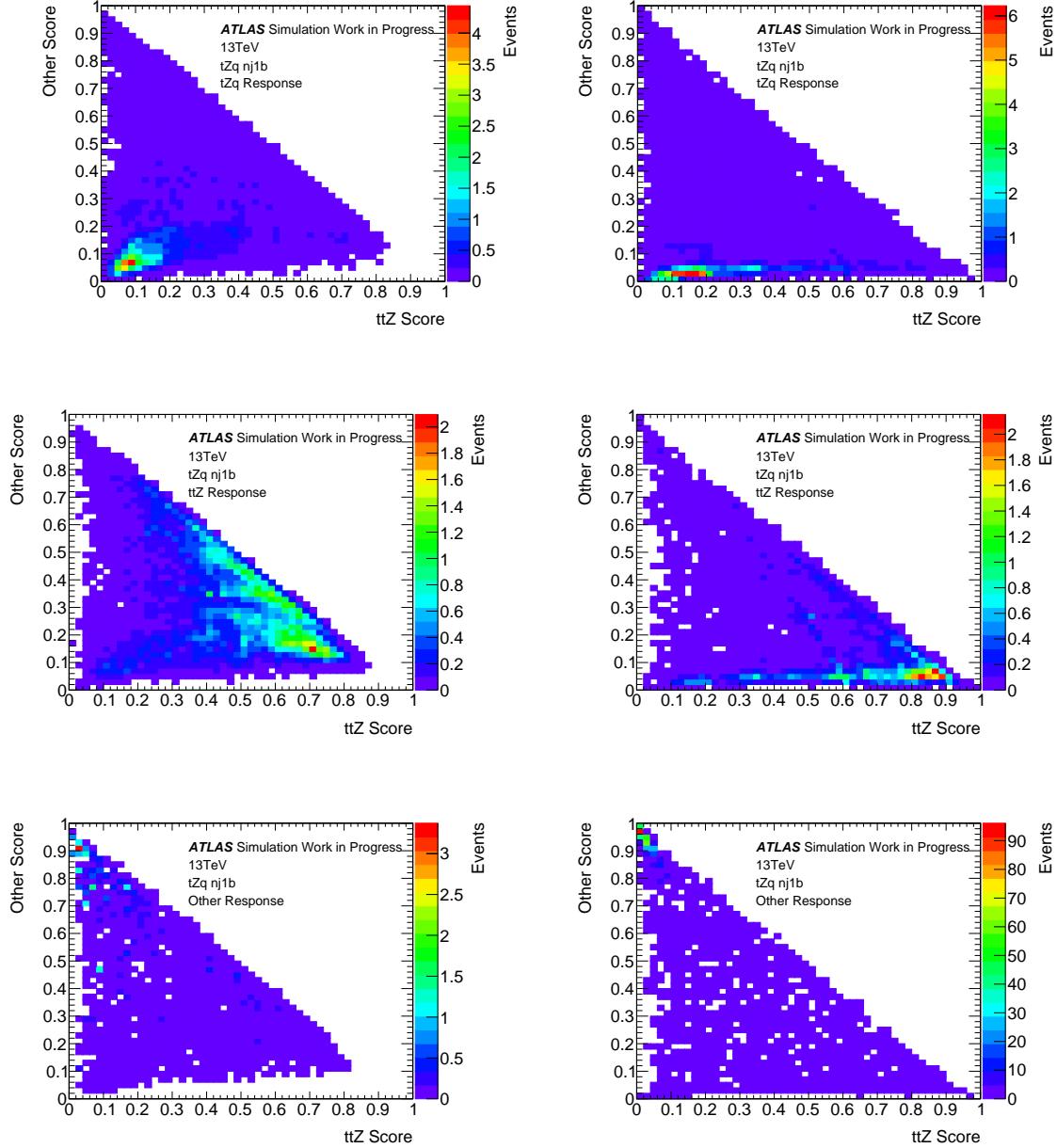


Figure 5.54: 2D plots showing the event (tZq , $t\bar{t}Z$, other) response in the $t\bar{t}Z$ and other nodes in the $nj1b$ signal region.

Therefore looking at figures 5.55 and 5.56 I determined the threshold score with the highest $S/\sqrt{S+B}$ ratio for each node. The ratio $S/\sqrt{S+B}$, where S is the count of signal events per bin and B is the count of background events per bin, indicates the region with the highest signal significance.

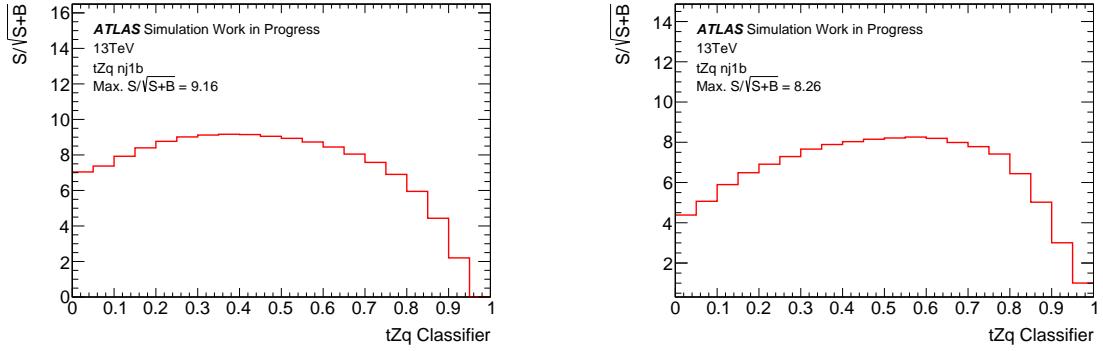


Figure 5.55: Comparison of the plots showing the S over $\sqrt{S+B}$ ratio of the tZq node for the default (left) and loose (right) classifier in the $nj1b$ signal region.

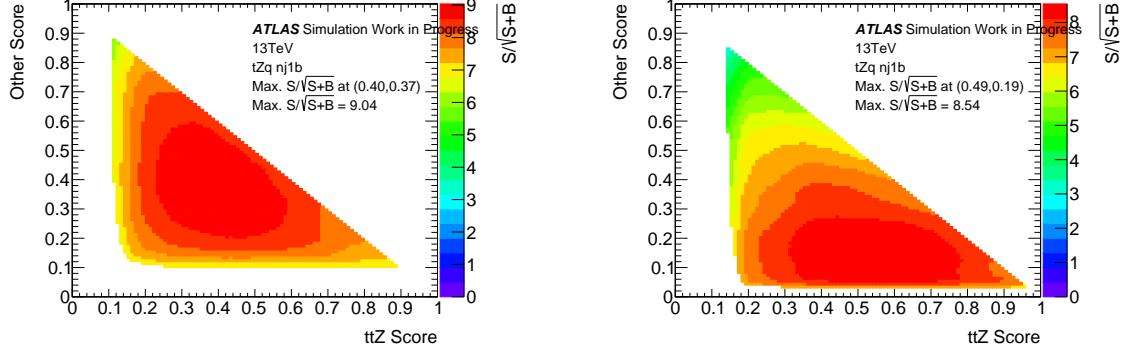


Figure 5.56: Comparison of the 2D plots showing the S over $\sqrt{S+B}$ ratio across all possible scores of the $t\bar{t}Z$ and other nodes for the default (left) and loose (right) classifier in the $nj1b$ signal region.

Table 5.22 shows the max $S/\sqrt{S+B}$ values and their threshold scores.

	Max $S/\sqrt{S+B}$	ttZ Score	Other Score
Default	9.04	0.40	0.37
Loose	8.54	0.49	0.19
	Max $S/\sqrt{S+B}$		tZq Score
Default	9.16		0.38
Loose	8.26		0.58

Table 5.22: Maximum values of $S/\sqrt{S+B}$ for the default and loose $t\bar{t}Z$ multi-class classifier in the $nj1b$ SR.

5.5.5 Trainings summary

Table 5.23 sums up the $S/\sqrt{S+B}$ maximum significance through all trainings. This table shows that the default classifier has an higher $S/\sqrt{S+B}$ maximum significance than the loose classifier in the nj1b and 2j1b SRs, whereas in the 3j1b SR the loose classifier outperforms the default classifier because it counts nearly 40 % more events and it has a better separation. This could be due to the fact that the loose selection has many more background events (964 in the 2j1b SR and 740 in the 3j1b SR) than the default selection (351 in the 2j1b SR and 233 in the 3j1b SR), hence the DNNs' performance improvements are hidden by the overwhelming number of background events.

Moreover table 5.23 shows clearly that the nj1b signal region has the best $S/\sqrt{S+B}$ significance for each classifier (binary, diboson multi-class and $t\bar{t}Z$ multi-class).

Since the nj1b signal region is the sum of the 2j1b and 3j1b signal regions, it has the highest number of events, hence the DNNs have a larger dataset to train and improve. The final observation over table 5.23 is that the 3j1b signal region has the lowest $S/\sqrt{S+B}$ significance.

Signal region	Classes	$S/\sqrt{S+B}$	
		Default	Loose
2j1b	Binary	7.82	6.09
	Multi-class diboson	7.86	5.86
	Multi-class $t\bar{t}Z$	7.84	5.88
3j1b	Binary	5.10	6.09
	Multi-class diboson	5.03	6.01
	Multi-class $t\bar{t}Z$	5.00	5.95
nj1b	Binary	9.18	8.67
	Multi-class diboson	9.08	8.62
	Multi-class $t\bar{t}Z$	9.16	8.26

Table 5.23: Summary of the $S/\sqrt{S+B}$ significance through all trainings.

Tables 5.24, 5.25 and 5.26 show that the loose classifier outperforms the default classifier in terms of separation.

First of all let's have a look at the binary classifiers. The loose binary classifier outperforms the default classifier by a 14 % (at least) separation gap.

Moreover it is clear that both (loose and default) binary classifiers tend to perform better in the 2j1b signal region. This could be due to the fact that the 2j1b signal region has more events for both selections (table 5.2): the DNNs have more data to train and improve.

Signal Region	Classifier	Separation
2j1b	Default	43.77 %
	Loose	60.11 %
3j1b	Default	39.15 %
	Loose	53.03 %
nj1b	Default	40.60 %
	Loose	57.17 %

Table 5.24: Summary of the separation percentages of the binary classifiers for each signal region.

Regarding the multi-class trainings, tables 5.25 and 5.26 show an interesting trend: the default classifier has a respectable separation strength for tZq events, but its performance collapses for the other two classes of events (*diboson/t $\bar{t}Z$* and *other*). On the contrary the loose classifier shows a solid performance throughout all classes of events. The responsibility of this behaviour could be attributed to the larger number of background events in the loose selection.

Signal Region	Classifier	Separation		
		tZq	Diboson	Other
2j1b	Default	44.47 %	28.32 %	27.04 %
	Loose	57.08 %	52.94 %	59.91 %
3j1b	Default	37.11 %	21.28 %	21.60 %
	Loose	51.43 %	45.48 %	46.32 %
nj1b	Default	38.72 %	24.26 %	20.46 %
	Loose	56.22 %	52.39 %	55.87 %

Table 5.25: Summary of the separation percentages of the diboson multi-class classifiers for each signal region.

Finally, I compare the multi-class classifiers. The diboson and $t\bar{t}Z$ multi-class classifiers have similar tZq separation values in the default selection (the differences are roughly 1%), whereas the diboson multi-class classifiers slightly outperform the $t\bar{t}Z$ multi-class classifiers in the loose selection (the differences are between 2% and 4%).

Signal Region	Classifier	Separation		
		tZq	t <bar>t>Z</bar>	Other
2j1b	Default	44.08 %	26.46 %	30.84 %
	Loose	54.20 %	40.48 %	53.99 %
3j1b	Default	38.42 %	23.25 %	29.90 %
	Loose	49.05 %	37.62 %	48.39 %
nj1b	Default	40.48 %	24.00 %	26.66 %
	Loose	52.84 %	41.11 %	52.83 %

Table 5.26: Summary of the separation percentages of the $t\bar{t}Z$ multi-class classifiers for each signal region.

Regarding the separation values of the dominant class ($t\bar{t}Z$ or *diboson*), the diboson and $t\bar{t}Z$ classifiers have similar performances in the default selection. On the contrary the diboson classifier outperforms the $t\bar{t}Z$ classifier in each signal region for the loose selection. This might be due to the fact that the loose selection has many more diboson events than $t\bar{t}Z$ events. Therefore the diboson DNNs have more training samples.

Regarding the separation values of the *Other* class, the $t\bar{t}Z$ multi-class classifiers outperform the diboson multi-class classifiers. However the *Other* class of the $t\bar{t}Z$ multi-class classifiers counts many more events than the *Other* class of the diboson multi-class classifiers. Moreover the *Other* class of the $t\bar{t}Z$ multi-class classifiers is not a miscellaneous class of events: it is roughly uniform because it dominated by the presence of diboson events. Therefore, since the *Other* class of the $t\bar{t}Z$ classifiers is more uniform than the *Other* class of the diboson classifiers, the separation is easier for the DNNs.

Diboson multi-class classifier				
Class	Default		Loose	
	SR 2j1b	SR 3j1b	SR 2j1b	SR 3j1b
tZq	118	66	93	99
Diboson	235	69	557	450
Other	116	164	407	290

Table 5.27: Events distribution in the diboson multi-class classifier.

$t\bar{t}Z$ multi-class classifier				
Class	Default		Loose	
	SR 2j1b	SR 3j1b	SR 2j1b	SR 3j1b
tZq	118	66	93	99
$t\bar{t}Z$	97	148	33	79
Other	254	85	931	661

Table 5.28: Events distribution in the $t\bar{t}Z$ multi-class classifier.

Tables 5.28 and 5.27 show that the events' distribution in the diboson multi-class classifier is more balanced than the $t\bar{t}Z$ classifier's distribution. The classes of the $t\bar{t}Z$ multi-class classifier are extremely unbalanced for the loose selection.

The consequence of an unbalanced dataset is a lower quality DNN.

Let's examine in depth the balance problem for the diboson multi-class classifier. The loose 2j1b SR has the best balance through all three classes, in fact the best separation values are obtained exactly for that specific training. The loose 3j1b SR is slightly unbalanced with respect to the loose 2j1b SR, in fact the trainings are slightly worse. Finally, the loose nj1b SR is balanced, in fact its training has similar performances to the loose 2j1b SR training.

5.6 Conclusions

The previous analysis used a cut based method to separate the signal from the background events. However it was insufficient.

Therefore it was used a multivariate approach to improve the separation between signal and background events. The multivariate analysis was based on Machine Learning techniques, in particular a binary classifier was used.

The goal of my analysis was to explore a more refined multivariate analysis based on a multi-class classifier and to answer to three main questions:

1. The *classifier* problem: which classifier should be used? Binary or Multi-Class?
2. The *class* problem: if the multi-class classifier is chosen, should the analysis be based on a *diboson* or $t\bar{t}Z$ classifier (meaning that the class of the dominant background would be *diboson* or $t\bar{t}Z$ respectively)?
3. The *selection* problem: which selection should be used? Default or Loose?

My dissertation shows that the binary classifier has consistently a better separation and significance ($S/\sqrt{S+B}$) than the multi-class classifier. The binary classifiers showed

a better signal-background separation and signal over background significance than the multi-class classifiers in each SR for both selections, except for the 2j1b SR of the default selection where the multi-class network performed slightly better.

However one motivation to use the multi-class classifier, even though its separation performances are not better than those of the binary classifier, could be the following. The 2D response plots showing the response (scores) of the *Other* and the *Dominant background* nodes to a specific class of events (*tZq*, *dominant background* and *Other*) could be used to define:

- A *Signal region* (SR) of scores. The signal region is defined by some arbitrary cuts and it is enriched in the signal process, meaning it usually has a high $S/\sqrt{S+B}$ significance;
- A *Control region* (CR) of scores. The control region is defined by some arbitrary cuts and it is any region where a background process is dominant. These regions have a very low $S/\sqrt{S+B}$ significance, because the signal contribution should be low here;

The SR and CR would then be used in a profile likelihood fit. The SR should provide an improvement on sensitivity over the measurement of the signal process, whereas the CR should provide constraints on systematic uncertainties arising from the background processes.

Therefore, even if the separation signal-background suggests that the binary classifier is superior to the multi-class network, I would try this attempt before giving a definitive judgement.

Regarding the second question, the goal of my analysis was to determine the best process (between *diboson* and $t\bar{t}Z$) to be used as the dominant class in a three-classes classifier.

However I discovered there is not a physics' criteria with the available statistics which leads to a choice or the other.

My analysis shows that it must be taken into account the *balance* of the classes. My analysis shows that a diboson multi-class classifier has better performances over the $t\bar{t}Z$ multi-class classifier in the loose selection. This was the consequence of a more balanced distribution of the events in the classes of the diboson classifier.

Therefore my thesis suggests to choose the multi-class classifier which has the best events' distribution between the three defined classes.

Regarding the third question, the goal of my analysis was to determine the best selection (between *default* and *loose*) to work with. Each classifier showed better separation values in the loose selection. This was a direct consequence of the fact that the loose classifier is a larger dataset than the default one, hence the DNNs have more data to train and to improve.

However the significance tells a different story.

In the 2j1b SR the $S/\sqrt{S+B}$ ratio is higher for the loose selection, whereas in the 3j1b SR the significance is higher for the default selection. This opposite behaviour is a consequence of the number of signal events. In the 2j1b SR the loose selection has 30 % more signal events than the default selection, whereas in the 3j1b SR the default selection has 40 % more signal events than the loose selection. Since the nj1b SR has approximately the same number of signal events (about 180) in both selections, it is the right signal region to uphold the best SR to work with. The nj1b SR shows that the classifiers have a slightly better significance in the default selection with respect to the loose selection. Therefore the high separation performances of the loose classifiers do not compensate the overwhelming quantity of background events of the loose selection. In conclusion my suggestion is to train the DNNs on a large dataset, such as the loose selection, and then those well trained models should be used on a clean dataset (few background events), such as the default selection.

Bibliography

- [1] D. Griffiths. *Introduction To Elementary Particles*. John Wiley & Sons, 1987.
- [2] H David Politzer. “Asymptotic freedom: An approach to strong interactions”. In: *Physics Reports* 14 (1974), pp. 129–180. ISSN: 0370-1573.
DOI: [https://doi.org/10.1016/0370-1573\(74\)90014-3](https://doi.org/10.1016/0370-1573(74)90014-3). URL:
<https://www.sciencedirect.com/science/article/pii/0370157374900143>.
- [3] W. Marciano and H. Pagels. “Quantum chromodynamics”. In: *Physics Reports* 36 (1978), pp. 137–276. ISSN: 0370-1573.
DOI: [https://doi.org/10.1016/0370-1573\(78\)90208-9](https://doi.org/10.1016/0370-1573(78)90208-9). URL:
<https://www.sciencedirect.com/science/article/pii/0370157378902089>.
- [4] O. W. Greenberg. “Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons”. In: *Phys. Rev. Lett.* 13 (1964), pp. 598–602.
DOI: [10.1103/PhysRevLett.13.598](https://doi.org/10.1103/PhysRevLett.13.598).
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.598>.
- [5] M. Y. Han and Y. Nambu. “Three-Triplet Model with Double SU(3) Symmetry”. In: *Phys. Rev.* 139 (1965), B1006–B1010. DOI: [10.1103/PhysRev.139.B1006](https://doi.org/10.1103/PhysRev.139.B1006).
URL: <https://link.aps.org/doi/10.1103/PhysRev.139.B1006>.
- [6] H. Fritzsch, M. Gell-Mann, and H. Leutwyler.
“Advantages of the color octet gluon picture”. In: *Phys. Lett. B* 47 (1973), pp. 365–368. ISSN: 0370-2693.
DOI: [https://doi.org/10.1016/0370-2693\(73\)90625-4](https://doi.org/10.1016/0370-2693(73)90625-4). URL:
<https://www.sciencedirect.com/science/article/pii/0370269373906254>.
- [7] R. L. Workman and Others. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [8] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.
DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [9] A. Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C* 680519 (1968), pp. 367–377.
DOI: [10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).

- [10] S. L. Glashow. “Partial-symmetries of weak interactions”.
In: *Nuclear Physics* 22 (1961), pp. 579–588. ISSN: 0029-5582.
DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL:
<https://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [11] S. L. Glashow, J. Iliopoulos, and L. Maiani.
“Weak Interactions with Lepton-Hadron Symmetry”.
In: *Phys. Rev. D* 2 (1970), pp. 1285–1292. DOI: [10.1103/PhysRevD.2.1285](https://doi.org/10.1103/PhysRevD.2.1285).
- [12] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”.
In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509.
DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [13] F. Englert and R. Brout.
“Broken Symmetry and the Mass of Gauge Vector Mesons”.
In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323.
DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [14] P. W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”.
In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: [10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).
URL: <https://link.aps.org/doi/10.1103/PhysRev.145.1156>.
- [15] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”.
In: *Phys. Lett. B* 716 (2012), p. 1. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- [16] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 30.
DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235).
- [17] D0 Collaboration. “Observation of the Top Quark”.
In: *Physical Review Letters* 74 (1995), pp. 2632–2637.
DOI: [10.1103/physrevlett.74.2632](https://doi.org/10.1103/physrevlett.74.2632).
URL: <https://doi.org/10.1103%2Fphysrevlett.74.2632>.
- [18] LHC Top Working Group. URL:
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [19] J. Alwall, R. Frederix, S. Frixione et al.
“The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”.
In: *JHEP07(2014)079* (). URL: [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [20] O. B. Bylund. *Modelling Wt and tWZ production at NLO for ATLAS analyses*.
2016. DOI: [10.48550/ARXIV.1612.00440](https://doi.org/10.48550/ARXIV.1612.00440).
URL: <https://arxiv.org/abs/1612.00440>.

- [21] L. R. Evans and P. Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. 164 p.
DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
URL: <http://cds.cern.ch/record/1129806>.
- [22] *High-Luminosity LHC*.
URL: <https://home.cern/science/accelerators/high-luminosity-lhc>.
- [23] *Run-III*. URL: <https://home.cern/events/launch-lhc-run-3>.
- [24] *Run-I and Run-II*. URL: <https://home.cern/news/news/accelerators/lhc-report-final-days-run-2>.
- [25] ATLAS Collaboration.
“The ATLAS Experiment at the CERN Large Hadron Collider”.
In: *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [26] CMS Collaboration. “The CMS Experiment at the CERN LHC”.
In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [27] ALICE Collaboration. “The ALICE experiment at the CERN LHC”.
In: *JINST* 3 (2008), S08002–S08002. DOI: [10.1088/1748-0221/3/08/s08002](https://doi.org/10.1088/1748-0221/3/08/s08002).
URL: <https://doi.org/10.1088/1748-0221/3/08/s08002>.
- [28] LHCb Collaboration. “The LHCb Detector at the LHC”.
In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005.
DOI: [10.1088/1748-0221/3/08/s08005](https://doi.org/10.1088/1748-0221/3/08/s08005).
URL: <https://doi.org/10.1088/1748-0221/3/08/s08005>.
- [29] *Linac4*.
URL: <https://home.cern/science/accelerators/linear-accelerator-4>.
- [30] *Proton Synchrotron Booster*. URL:
<https://home.cern/science/accelerators/proton-synchrotron-booster>.
- [31] *Proton Synchrotron*.
URL: <https://home.cern/science/accelerators/proton-synchrotron>.
- [32] *Super Proton Synchrotron*. URL:
<https://home.cern/science/accelerators/super-proton-synchrotron>.
- [33] *Large Hadron Collider*.
URL: <https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
- [34] J. T. Boyd. *LHC Run-2 and Future Prospects*. 2020.
DOI: [10.48550/ARXIV.2001.04370](https://arxiv.org/abs/2001.04370).
URL: <https://arxiv.org/abs/2001.04370>.
- [35] ATLAS Collaboration.
Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV.
Geneva, 2015. URL: [https://cds.cern.ch/record/2037717](http://cds.cern.ch/record/2037717).

- [36] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data”. In: *JINST* 14 (2019), P03017. DOI: [10.1088/1748-0221/14/03/P03017](https://doi.org/10.1088/1748-0221/14/03/P03017). arXiv: [1812.03848 \[hep-ex\]](https://arxiv.org/abs/1812.03848).
- [37] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*. Tech. rep. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2157687>.
- [38] ATLAS Collaboration. *Electron identification measurements in ATLAS using $\sqrt{s} = 13$ TeV data with 50 ns bunch spacing*. Geneva, 2015. URL: <https://cds.cern.ch/record/2048202>.
- [39] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 639. DOI: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6). arXiv: [1902.04655 \[hep-ex\]](https://arxiv.org/abs/1902.04655).
- [40] ATLAS Collaboration. “Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 580. DOI: [10.1140/epjc/s10052-017-5081-5](https://doi.org/10.1140/epjc/s10052-017-5081-5). arXiv: [1705.02211 \[hep-ex\]](https://arxiv.org/abs/1705.02211).
- [41] ATLAS Collaboration. “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). arXiv: [1907.05120 \[hep-ex\]](https://arxiv.org/abs/1907.05120).
- [42] S. Kang, S. Cho, and P. Kang. “Constructing a multi-class classifier using one-against-one approach with different binary classifiers”. In: *Neurocomputing* 149 (2015), pp. 677–682. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2014.08.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0925231214010200>.
- [43] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2014. DOI: [10.48550/ARXIV.1412.6980](https://arxiv.org/abs/1412.6980). URL: <https://arxiv.org/abs/1412.6980>.
- [44] S. J. Reddi, S. Kale, and S. Kumar. *On the Convergence of Adam and Beyond*. 2019. DOI: [10.48550/ARXIV.1904.09237](https://arxiv.org/abs/1904.09237). URL: <https://arxiv.org/abs/1904.09237>.