SAPIENZA
UNIVERSITÀ DI ROMA

Identifying heavy flavour jets using calorimeter-only information at the ATLAS detector

Facoltà di Scienze Matematiche, Fisiche e Naturali
Laurea in Fisica (Classe L-30)

Saverio Capolongo
ID number 1913849

Advisor
Prof. Cesare Bini

Co-Advisor
Dr. Stefano Veneziano

Academic Year 2021/2022

The thesis project has been developed at the University of Geneva under the supervision of Prof. Anna Sfyrla and PhD Assistant Stefano Franchellucci in the ATLAS Group.

Identifying heavy flavour jets using calorimeter-only information at the ATLAS detector
Bachelor Thesis. Sapienza University of Rome

© Saverio Capolongo. All rights reserved

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: capolongosaverio@gmail.com

Remember - there's reality,
and then there's math
that is our best attempt
at describing how that reality works.

Ceci n'est pas une pipe.
It's a depiction of a pipe.
Remember to not confuse the two.

Abstract

One of the most abundant products of proton-proton collisions at high energies is given by the so called jets. Jets are streams of collimated particles initiated by a parton, either a quark or a gluon, directly produced in a pp collision. Among the jets, of special interest are those that are initiated by heavy-flavour quarks, charm (c) or beauty (b) that are also very abundant and related to several interesting physics processes. Heavy flavour jets are identified by fully exploiting the track information associated with the jets. Tracks, however, cannot be used at the first level of the trigger to decide wether to accept or reject the event, since the first level trigger relies only on calorimeter and muon detector informations.

This work, conducted within the ATLAS experiment at LHC, aims to investigate the possibility of identifying heavy flavour jets and distinguish them from light flavour jets using calorimeter-only information. This could become a powerful tool in the Phase-II ATLAS trigger selections, targeting final states with multiple such objects, such as di-Higgs production to four heavy-flavour jets. In the process, we will study in detail the differences between jets originating from different parton sources and we will employ machine-learning methods to distinguish the various types.

Contents

Chapter 1

Introduction

The goal of particle physics at high energy (also known as high energy physics) is to determine the most fundamental building blocks of matter and to understand the interactions between these particles. The underlying theoretical construct in particle physics is called the Standard Model (SM) and it contains (as shown in Figure 1.1) 6 quarks, 6 leptons, 4 gauge bosons, and one scalar boson (the Higgs boson); they interact through three interactions (strong force, weak force, and electromagnetism).

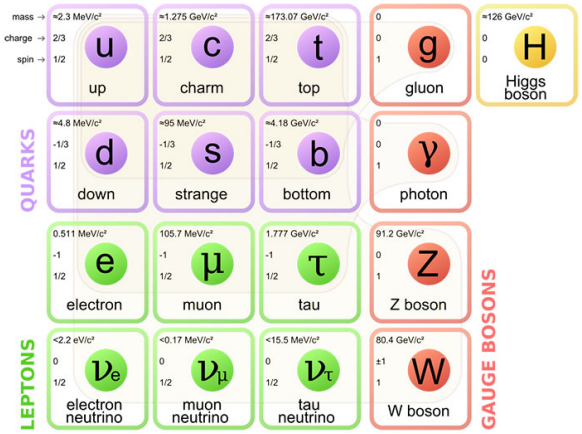

Figure 1.1. Chart showing the elementary particles of the SM with the respective properties: mass, charge, spin.

The measurement of hard scattering processes between protons that are bound states of three quarks, meaning those with energy scales of more than a few GeV, is the main method by which physics is being explored and extended by the experiments at the LHC[1]. Hard scattering processes probe distance scales far below the radius of the proton[2], and thus are best understood as collisions between the constituent quarks and gluons (generically called partons) of the proton.

[1] Large Hadron Collider, further explained later

[2] The size of a proton is of approximately $1fm = 10^{-15}$ m; compared to a size of an atom which is of around 1 $\mathring{A} = 10^{-10}$m.

Chapter 2

What is the LHC and what is ATLAS

The Large Hadron Collider (LHC shown in Figure 2.1) at CERN[1] is a two-ring proton-proton collider housed in a 27 km tunnel. Its main goal is that of exploring physics at energies above the Electroweak Simmetry Breaking Scale (EWSB), which is around 100 GeV and beyond. This energy region is special for several reasons. First of all, the mechanism for the EWSB is related to the Higgs field; indeed the mass of the Higgs boson[2] is $M_{Higgs} = 125.10 \pm 0.14$ GeV. At the same time, W, Z bosons and the top quark all have masses in this range, they are therefore copiously produced at the LHC. Moreover, through such searches, we are not only exploring a new energy range but also a qualitatively new region of physics. Ultimately, this is essential in order to establish whether our understanding of fundamental physics remains valid in this new region, or requires extension.

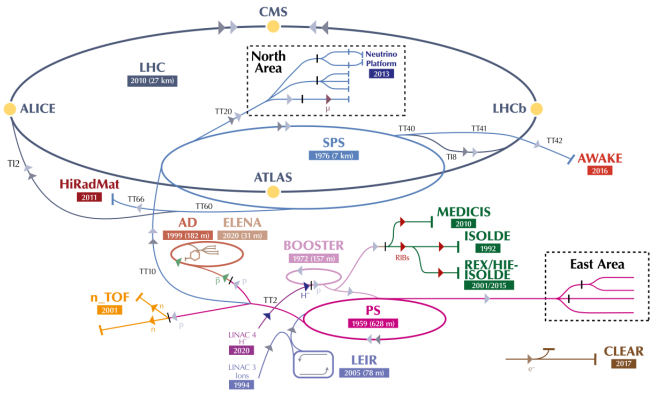

Figure 2.1. Schematic view of the CERN accelerator complex. It is made of a succession of accelerators which brings particles to an increasingly higher energy before injecting the beam into the next machine. The LHC is the last element of the chain.

[1] "European Organization for Nuclear Research" from the french "Conseil Européen pour la Recherche Nucléaire".

[2] The value is that of the the Particle Data Group (PDG), an international collaboration of particle physicists that compiles and reanalyzes published results related to the properties of particles and fundamental interactions.

The ATLAS detector is the largest volume particle detector ever constructed. It is 46 m long, 25 m high and 25 m wide and sits in a cavern 100 m below ground near the main CERN site, close to the village of Meyrin in Switzerland. It is one of two general-purpose detectors at the LHC. With its forward-backward symmetry, which provides almost full solid angle coverage around the proton-proton interaction point, it investigates a wide range of physics, from the search for the Higgs boson to extra dimensions and particles that could make up dark matter. Beams of particles from the LHC collide at the centre of the ATLAS detector making collision debris in the form of new particles, which fly out from the collision point in all directions.

A particle emerging from the interaction point is characterised by a set of coordinates, a brief summary of the coordinate system and nomenclature is given here. The interaction point is defined as the origin of the coordinate system. The z-axis is parallel to the beam and the x- and y- axes are perpendicular to the beam, forming a right-handed cartesian coordinate system where x points towards the centre of the LHC ring and y points upward. The x-y plane is called the transverse plane. Particles are often described by their transverse momentum p_T and transverse energy (projections in the transverse plane). The azimuthal angle ϕ is measured around the z-axis and the polar angle θ is measured from the z-axis. The pseudorapidity, as defined in Equation 2.1 and shown in Figure 2.2, is often preferable as a polar coordinate.

$$\eta = -\log(\tan(\theta)) \tag{2.1}$$

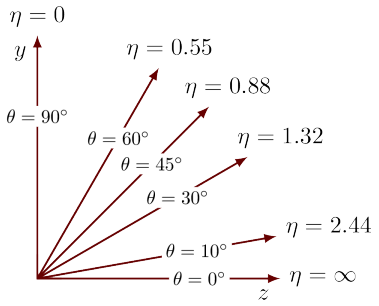

Figure 2.2. Graphical view of the relation between pseudorapidity and the polar angle θ as defined in Equation 2.1.

The distance ΔR between two particles in the $\eta - \phi$ space is defined as in Equation 2.2.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.2}$$

Six different detecting subsystems arranged in layers around the collision point record the paths, momentum, and energy of the particles, allowing them to be individually identified. A huge magnet system bends the paths of charged particles so that their momenta can be measured.

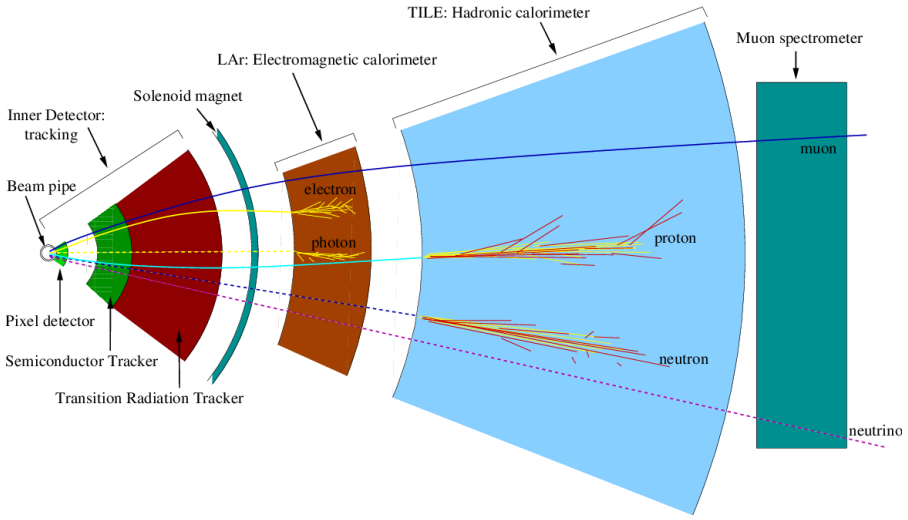

Figure 2.3. Section scheme of the ATLAS detector divided into its components; form inner to outer they are: Pixel detector, Semiconducor Tracker, Transition Radiation Tracker, Electromagnetic Calorimeter, Hadronic Calorimeter, Muon Spectrometer.

The ATLAS detector is composed of different sub-detectors, as shown in Figure 2.3. The sub-detectors are arranged in cylindrical layers around the interaction point. Closest to the beam pipe is the Inner Detector (ID), used to reconstruct the trajectory of charged particles. The ID is enclosed in a solenoid, which provides a strong magnetic field parallel to the beam, capable of bending the charged particles, emerging from the interaction, to measure their momenta and charge. The Electromagnetic (EM) Calorimeter surrounds the ID and is designed to precisely measure the energy of electrons and photons. Outside the EM Calorimeter there is the Hadronic (Had) Calorimeter, which measures the energy of hadronic particles. Finally, the calorimeters are surrounded by the Muon Spectrometer (MS), designed to reconstruct and identify muons, which usually escape the previous detector layers.

The ID provides track reconstruction within $|\eta| < 2.5$, employing a pixel detector (Pixel) close to the beam pipe, a silicon microstrip detector (SCT) at intermediate radii, and a transition radiation tracker (TRT) at outer radii (as shown in Figure 2.3). The calorimeter will be explained in details in the specific section. The MS consists of one barrel ($|\eta| < 1.05$) and two end-cap sections ($1.05 < |\eta| < 2.7$) and is based on alternate precision chambers and trigger chambers.

2.1 High Luminosity

Luminosity, which is the measure of the number of potential collisions per unit of surface and time, is an essential indicator of an accelerator's performance. Integrated luminosity is measured in inverse femtobarns (fb^{-1}). By the end of its first few years of operation at 13 TeV (at the end of 2018), the LHC produced 150 fb^{-1} of data. In the case of head-on collisions of bunches, the formula of the luminosity is displayed in Equation 2.3 where N_1 and N_2 are the number of particles per bunch travelling at a revolution frequency f, N_b is the number of bunches and σ_x and σ_y are the root means squared of the lenghts of the beam. It shows how the luminosity

depends on the number of particles per bunch and the beam sizes.

$$\mathcal{L}_{inst} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y} \tag{2.3}$$

We can now write the collision rate as in Equation 2.4, where Γ_{coll} is the collision rate, $\mathcal{L}_{inst}$ is the instantaneous luminosity as defined in Equation 2.3 and σ is the cross section which is a measure of the probability that a particular process will occur, measured in Barns[3].

$$\Gamma_{coll} = \sigma \cdot \mathcal{L}_{inst} \tag{2.4}$$

The High-Luminosity Large Hadron Collider (HL-LHC) project (whose time schedule is outlined in Figure 2.4) aims to crank up the performance of the LHC in order to increase the potential for discoveries. The objective is to increase the integrated luminosity by a factor of 10 beyond the LHC's design value. The High-Luminosity LHC, which should be operational from the beginning of 2029, will produce more than 250 fb^{-1} of data per year and will be capable of collecting up to 4000 fb^{-1}, thus at least 15 million Higgs bosons per year, compared to around three million from the LHC in 2017.

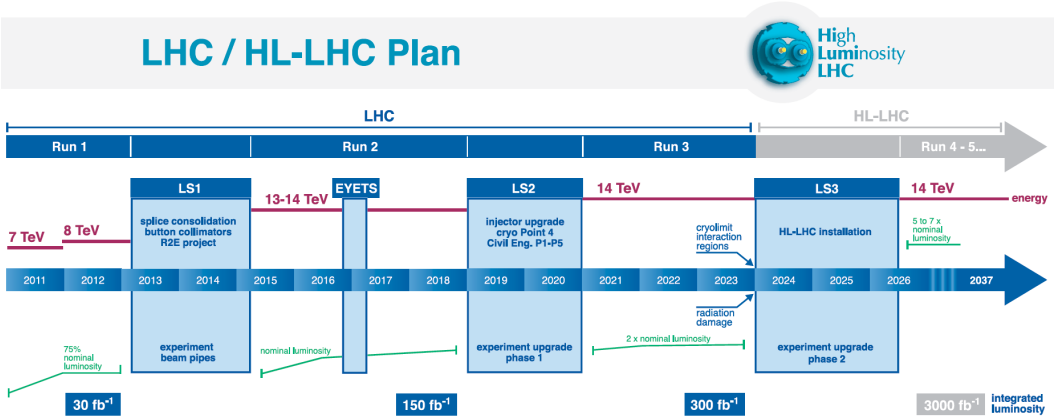

Figure 2.4. Upgrade program for the LHC accelerator complex. The status of the past years and the plans for the next two decades are shown.

The phenomena that physicists are looking for have a very low probability of occurring and this is why a very large amount of data is needed to detect them. Increasing luminosity produces more data, allowing physicists to study known mechanisms in greater detail and observe rare new phenomena that might reveal themselves.

Increasing the luminosity means increasing the number of collisions: at least 140 collisions will be produced at each bunch crossing, that is each time the particle bunches meet at the heart of the ATLAS and CMS[4] detectors, compared to around

[3]The Barn is a unit used in nuclear physics to indicate the probability of happening of a reaction between nuclear particles and atomic nuclei or even other nuclear particles. The dimension of the Barn is cross sectional area and 1 Barn is equal to $10^{-}28\ m^2$.

[4]CMS is the other general-purpose detectors at the LHC, visible in Figure 2.1.

40 collisions with the luminosity of 2015. To achieve this, the beam will need to be more intense and more focused than at present in the LHC, more intense as luminosity is directly proportional to N_1 and N_2 while more focused as luminosity is inversely proportional to σ_x and σ_y, as shown in Equation 2.3. New equipment will need to be installed over about 1.2 of the LHC's 27 kilometres.

This upgrade will open a number of new avenues for observing processes with very small cross sections. Major goals of HL-LHC thus belong to the following five categories: improved Standard Model measurements, searches for beyond the Standard Model (BSM) physics, flavour physics of heavy quarks and leptons, studies of the properties of the Higgs boson, and the studies of QCD[5] matter at high density and temperature.

However, as the rate of data will be increased, in order to keep the same level of selectivity with a higher background rate, a higher background rejection power is mandatory, in particular at trigger level as will be shown in the next section.

2.2 ATLAS Trigger: what it is and how it is structured

The large number of interactions in the ATLAS detectors create in principle an enormous flow of data. To digest the data, ATLAS uses an advanced "trigger" system to tell the detector which events to record and which to ignore. Complex data-acquisition and computing systems are then used to analyse the collision events recorded.

The Trigger and Data Acquisition (TDAQ) system shown in Figure 2.5 consists of a hardware-based first-level trigger (L1) and a software-based high-level trigger (HLT). The L1 trigger decision is formed by the Central Trigger Processor (CTP), which receives inputs from the calorimeter (L1Calo) and muon (L1Muon) detectors. When a pattern of detector elements is compatible with a particle coming from the interaction point (IP), a trigger signal is issued. If a L1 trigger signal is present, the event is buffered in the Read-Out System (ROS) and processed by the HLT. The HLT receives Region-of-Interest (RoI) information from L1, which can be used for regional reconstruction in the trigger algorithms. After the events are accepted by the HLT, they are transferred to local storage at the experimental site and exported to the Tier-0 facility[6] at CERN's computing centre for offline reconstruction.

[5]Quantum Chromodynamics (QCD) is the theory of strong interactions. It is formulated in terms of elementary fields (quarks and gluons), whose interactions obey the principles of a relativistic QFT, with a non-abelian gauge invariance SU(3).

[6]The Tier-0 facility at CERN is responsible for the first-pass processing of the raw data received from the ATLAS detector, for the archival of raw and derived data on the Tier-0 mass storage system, and for the distribution of the data from the Tier-0 to the Tier-1 centers around the world for further processing and analysis.

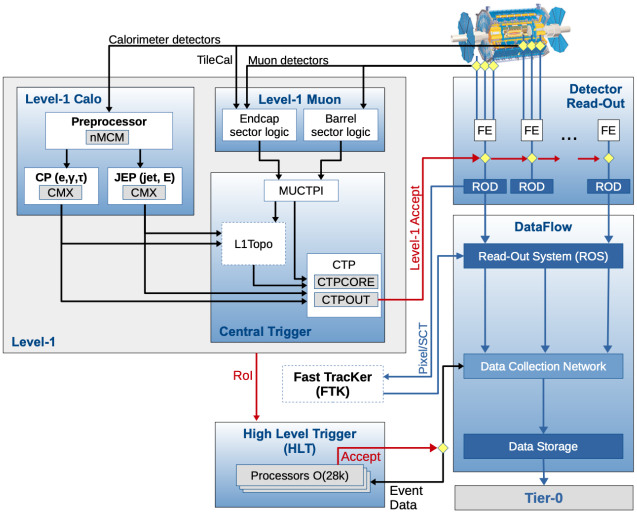

Figure 2.5. Scheme of the ATLAS. Showing its main division in the hardware-based first-level trigger (L1) and the software-based high-level trigger (HLT); it shows at the same time the subunits making up the trigger as a whole.

The trigger system is an essential component of any collider experiment as it is responsible for deciding whether or not to keep an event from a given bunch-crossing interaction for later study. During Run 1 (2009 to early 2013) of the LHC, the trigger system of the ATLAS experiment operated efficiently at instantaneous luminosities of up to $8 \cdot 10^{33}\ cm^{-2}\ s^{-1}$ and primarily at centre-of-mass energies, $\sqrt{(s)}$, of 7 TeV and 8 TeV. In Run 2 (since 2015) the increased centre-of-mass energy of 13 TeV and the higher luminosity ($2 \cdot 10^{34}\ cm^{-2}\ s^{-1}$) gave rise to an increased number of proton–proton interactions per bunch-crossing (pile-up).

Without upgrades of the trigger system, the trigger rates would have exceeded the maximum allowed rates when running with the trigger thresholds needed to satisfy the physics programme of the experiment. Any trigger condition is based on some threshold that defines above which value a trigger signal is issued. In general by incresing the threshold the rate is reduced but interesting physics events can be lost. A compromise between the need to retain intersting physics without incresing the rate to levels that could be difficult to manage is needed.

2.3 The Calorimeter

As shown in Figure 2.5, the L1 trigger only has informations from the Calorimeter and the Muon triggers, therefore it has no information about the reconstructed tracks of the particles. Therefore, as the aim of this study is that of implementing the trigger without using particle tracks, the calorimeter is for us of particular interest.

The calorimeter system covers the region $|\eta| < 4.9$, the forward region ($3.2 < |\eta| < 4.9$) being instrumented with a liquid-argon (LAr) calorimeter for electromagnetic and hadronic measurements. In the central region, a lead/LAr electromagnetic

calorimeter covers $|\eta| < 3.2$, while the hadronic calorimeter uses two different detector technologies, with steel/scintillator tiles ($|\eta| < 1.7$) or lead/LAr ($1.5 < |\eta| < 3.2$) as absorber/active material.

A series of reconstruction algorithms are used to convert signals from the calorimeter readout into objects, specifically cells and clusters, that then serve as input to the reconstruction of electron, photon, tau, and jet candidates and the reconstruction of the missing energy of the event (E_{miss}). These cells and clusters are also used in the determination of the shower shapes and the isolation properties of candidate particles (including muons), both of which are later used as discriminants for particle identification and the rejection of backgrounds. The reconstruction algorithms used in the HLT have access to full detector granularity and thus allow improved accuracy and precision in energy and position measurements with respect to L1.

Chapter 3

Jets in particle physics

By far the most common hard process in hadron collisions is the scattering of partons off each other. This is a consequence both of the high density in particular of gluons in the proton and the fact that the QCD coupling is significantly larger than the electroweak couplings.

Whenever in a hard p-p collision a parton is produced, a new phenomenon takes place (shown in Figure 3.1): the parton pulls new particles out of the vacuum and becomes a cloud of hadronic particles and photons, flying in roughly the same direction: this is a jet. Hadronic particles and photons deposit their energies mainly in the calorimeter system. In an attempt to resolve particles coming from the hard scatter, these energy deposits may be grouped into objects called reconstructed jets[1]; when working with particles reconstructed from energy deposits it is said to be working at reconstructed level. If physicists want to know how many quarks or gluons were emitted by an interaction, they have to disentangle the (sometimes overlapping) jets.

This phenomenon is unlike any other in particle physics. There are other processes that create particle-antiparticle pairs, but free quarks spontaneously generate more quarks until they are all bound, either in pairs forming mesons or triplets forming baryons. Isolated quarks cannot exist because the force between them doesn't fall off to zero with distance. Gravity, electric forces and the weak force all become negligible as objects separate, but the strong force, which governs the behaviour of quarks, increases as quarks separate, levelling off to a constant.

When a quark is far enough from its neighbours, the energy becomes large enough to become the mass of a new quark-antiquark pair. (Distance times force is energy, and energy can be converted into mass.) This is why an isolated quark doesn't stay isolated: it costs less energy to create more quarks than to be alone. For a highly energetic quark or gluon flying away from a collision, this process happens several times, resulting in a jet.

Jets are present in any interaction that produces quarks or gluons. Often they hide information about the primary interaction. For instance, more than half of Higgs bosons decay to a b quark and a b antiquark, but they appear in a particle detector as dozens of particles in two rough bundles.

[1] Jets can be also tought as a group of highly collimated hadrons.

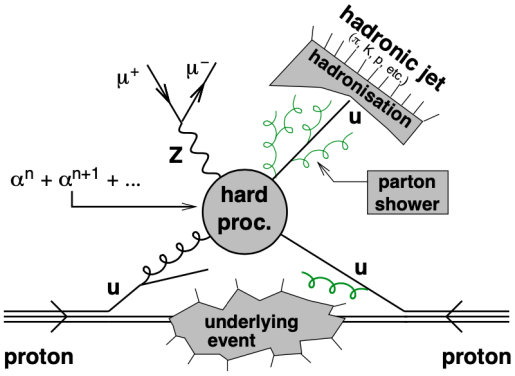

Figure 3.1. An example of hard scattering process, with parton undergoing hadronisation.

3.1 What are bjets and why are they interesting to study?

A substantial fraction of jets at the LHC are initiated by b-quarks. Bottom-quark-initiated jets are usually called b-jets. These are present in quark top decays and are present also in reactions predicted in several scenarios for physics beyond the SM.

The lifetime of b-hadrons is of the order of ps. This means that for b-hadrons with momenta of tens of GeV the typical decay length is of the order of few mm. If a b-hadron is produced in the "primary" collision vertex, it travels for few mm and then decays, giving rise to a "secondary" vertex displaced with respect to the first, and producing displaced tracks with respect to the primary vertex. The distance of closest approach of a displaced track to the primary vertex is defined as the transverse impact parameter, d0 (see Figure 3.2), and typically is large for tracks originating from the decay of b-hadrons. Conversely, jets initiated by light-flavour quarks or gluons will not exhibit these features and typically contain mostly prompt tracks originating from the primary vertex.

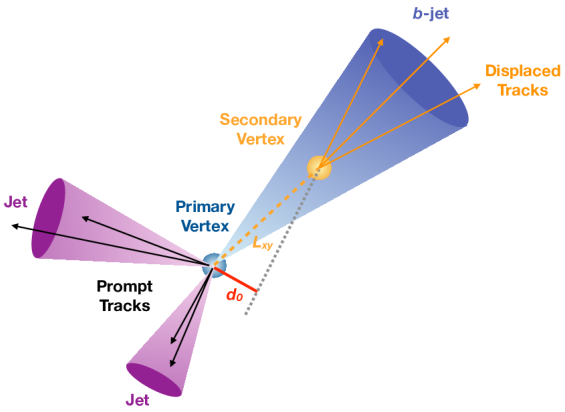

Figure 3.2. A schematic diagram of a section of the transverse plane (i.e. transverse to the beam) which shows an interaction producing two light-flavour jets and one bjet.

In its decay, the bottom quark transitions into a lighter quark, preferentially a charm quark and rarely an up quark, forming another known particle. The remaining energy is carried by a charged lepton: an electron, a muon or a tau, each accompanied by its associated neutrino; such a decay is thus called semi-leptonic one. A diagram of this decay is shown in Figure 3.3.

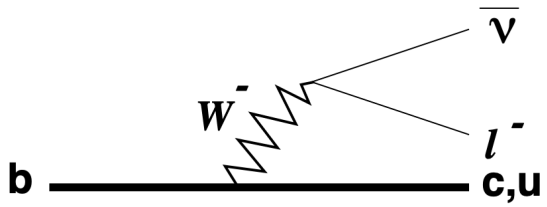

Figure 3.3. A diagram of the decay of a b quark transitioning to a c- or u- quark and consequently a W- which in turns decays to a lepton and the associated neutrino giving a semi-leptonic decay.

The presence of b-quarks is tagged with various algorithms, based on: the reconstruction of secondary vertices from the decays of hadrons containing b-quarks, the measurement of track impact parameters and the identification of muons with a sizable transverse momentum with respect to the jet axis.

3.2 Why studying di-Higgs (HH)?

As this study has the aim of testing a way of selecting b-jets, the natural applications are in b-jets related physics. The importance of the Higgs boson and the fact that it predominantly decays to b quarks makes this topic relevant for this discussion. Since the discovery of the long-sought Higgs boson, scientists have been working

to understand its properties and interactions with other particles – as well as its interaction with itself, its "self-coupling". One of the long-term goals of the LHC is to measure the Higgs-boson self-coupling, this can only be measured directly by studying the production of pairs of Higgs bosons (HH).

However, Higgs-boson-pair production in the Standard Model is very rare – a factor of thousand less likely than for the production of a single Higgs boson. While this is currently beyond the reach of the present statistics collected at LHC[2], multiple extensions of the Standard Model predict enhancements to the HH production rate. The resonant production of HH could occur through the decay of an intermediate, high-mass particle – such as a graviton or an additional Higgs boson – which would be observed as a peak of excess events in a narrow kinematic range. Alternatively, they could appear in non-resonant production, where such a narrow peak is not expected.

Searches of Higgs make use of its wide variety of decay modes with large branching ratios[3] or clean signatures. The HH branching ratios are determined from the known branching ratios of the SM Higgs boson. A summary of the branching ratios of the most important HH decay modes is shown in Figure 3.4. Each final state has advantages and disadvantages that determine its sensitivity across kinematic regimes. For a universally optimal search, a combination of many final states is required.

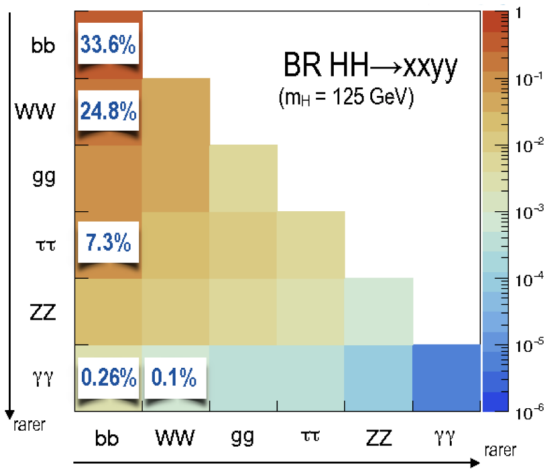

Figure 3.4. Branching ratios for the most important HH decay modes assuming SM couplings, calculated at NLO (Next to Leading Order).

The Standard Model Higgs boson decays predominantly to $b\bar{b}$ with a branching ratio of BR(H $\rightarrow b\bar{b}$) $\approx$ 58%. As a result, HH $\rightarrow b\bar{b}b\bar{b}$ is the dominant HH decay mode with a branching ratio of BR(HH $\rightarrow b\bar{b}b\bar{b}$) $\approx$ 33% (as visible in Figure 3.4). This large branching ratio makes $b\bar{b}b\bar{b}$ an important search channel, especially for

[2]However, with HL-LHC this could be observed.

[3]The branching ratio (branching fraction) is the fraction of events for a chosen particle measured to decay in a certain way. The sum of branching ratios for a particle is one. The branching ratio is defined as the partial decay width divided by the total width.

resonance masses above 1 TeV. The purely hadronic signature results in a large multijet background that can be reduced through the use of b-tagging methods to identify hadronic jets consistent with the decay of B-hadrons.

3.3 Top quarks

As the study will be developed on data where the generated event is the production of quarks top anti-top ($t\bar{t}$) it is important to give few informations about the top quark and the ways it can decay. To begin, the measurement of top quark production and the determination of top quark properties is interesting for several reasons: first of all the top is the heaviest quark, with a mass close to the EW scale. This suggests that the top quark might play a special role in EWSB; in many models for physics beyond the SM, the new states couple preferentially to top quarks and might manifest themselves, e.g., via decays into tt pairs; because of its very short lifetime ($5 \cdot 10^{-25}$ s), the top quark decays before it has a chance to hadronise, this gives a rather unique access to its properties. The top quark decays almost exclusively to a W boson and a bottom quark, which gives rises to a set of final states that are classified according to the W decays. As shown in Figure 3.5 the W boson can decay leptonically or hadronically. If both W bosons (from both the top and anti-top) decay leptonically we will have a dilepton signature accompanied by the b-quark jet and large missing energy due to the two neutrinos from the W decays. If one of the Ws decays hadronically, we will have a lepton plus jets signature. Finally, we can have both W bosons decaying hadronically and therefore leaving an all-hadronic signature.

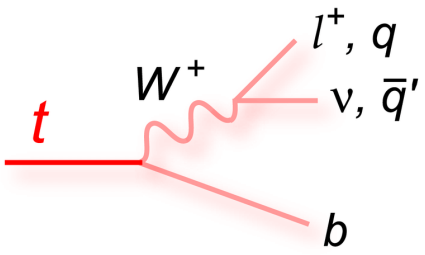

Figure 3.5. Decay of a top quark to a W boson and a b-quark; also the possible decay channels of the W boson are showed: either to a lepton and the relative neutrino or the quark and its antiquark.

Chapter 4

Characterisation of Jets

As already mentioned, b-jets are usually identified by fully exploiting the tracking information associated with the jets. Tracks, however, are not reconstructable at the first level of the trigger on ATLAS, neither now nor in its Phase-II upgrade. The aim of this study is therefore to investigate the possibility of identifying b-jets and distinguish them from light flavour jets using calorimeter-only information as a way of incrementing background rejection. In order to do so, a Neural Network (NN) has been developed and tuned. As a preliminary step, different samples of Monte Carlo Truth Jets have been examined with the goal of finding handles to use for the NN. All the studies of the samples have been made through python scripts written specifically for this project.

Multiple datasets have been used, each one consisting of Monte Carlo data with about 10^5 jets. For each jet, the complete set of available variables are displayed in the appendix. The most important ones are: for the jet: p_T, η, energy and mass; for each costituent: p_T, η, ϕ, energy mass and dr.

Plots of the variables at constituent level are available in the Appendix.

It is important to specify that the jets were divided in 4 classes: Light, Charm, Bottom, Tau. It is important to note the way jets were labelled. They are labelled as b-jets if they are matched to at least one b-hadron having $p_T \geq 5$ GeV within ΔR(b-hadron, jet)< 0.3 of the jet axis. If this condition is not satisfied, then c-hadrons and then τ leptons are searched for, with similar selection criteria. If a jet is matched to a c-hadron (τ-lepton), it is labelled a c-jet (τ-jet). A jet that does not meet any of these conditions is called a light-flavour jet.

4.1 The Datasets

The first dataset contained only a restricted set of particles, not including neutrinos and muons and was at first limited to a maximum number of 40 components per jet which was then extended to 60. We then studied the number of components per jets through the comparison between the 4 different jet types, both with the maximum number of 40 and 60 components. The results are shown in Figure 4.1. It is important to underline that the whole study has been developed at truth level. This means at the generator-level, thus with informations about raw particles or jets before they get passed on to the detector. Therefore, the particles themselves

are the truth particles which make up the jets.

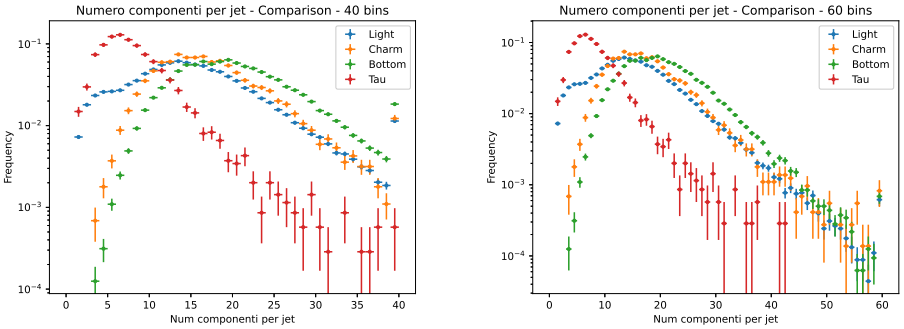

Figure 4.1. Distributions of the number of components per jet, for 40 and 60 max components.

The extension was made in order to have more precise results. This is what was then found though, as expected, it did not significantly changed them. In fact, what we can see from the plots is that the majority of the jets have less than 40 components, therefore what was added with the extension were just the tails of the distributions. However, the extension was kept for the small improvement it made. What is clearly visible from these plots is that the 4 distributions differ a lot for both different peaks and overall distribution, with Taus being completely separated from the others. This will probably be a good handle for the NN.

4.2 Final Dataset

Finally, it was chosen to go for a last dataset with a different dressing, including now neutrinos and muons; this particles were included for the study as it was thought they could have been useful for the following developement of the network. At first we studied the abundance of the jets and the average number of components per type of jet. The results are shown in Figures 4.2 and 4.3 and in Table 4.1.

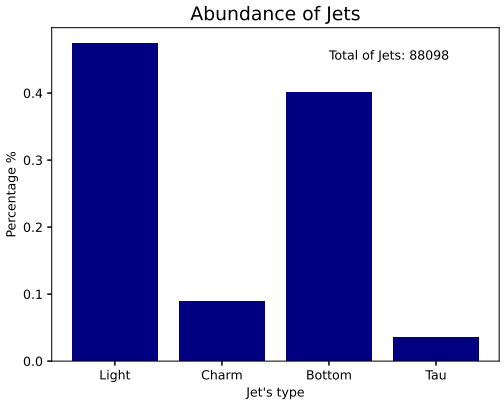

Figure 4.2. Abundance of jet.

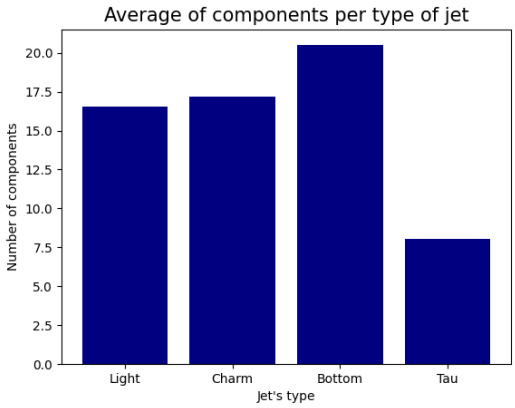

Figure 4.3. Average of components per type of jet.

Jet's type	Average value
Light	0.47
Charm	0.09
Bottom	0.40
Tau	0.04

Jet's type	Average value
Light	16.52
Charm	17.18
Bottom	20.49
Tau	8.03

Table 4.1. Table displaying the the average values found for the number of components per type of jet.

As expected, the average number of components for b-jets is higher as the mass of a bottom is higher than that of the others. Moreover, the biggest portion of jets was taken by b-jets and light-jets, this means that the biggest background for our study are Light jets. The composition in particles of the different type of jets was then studied and is displayed in Figure 4.4.

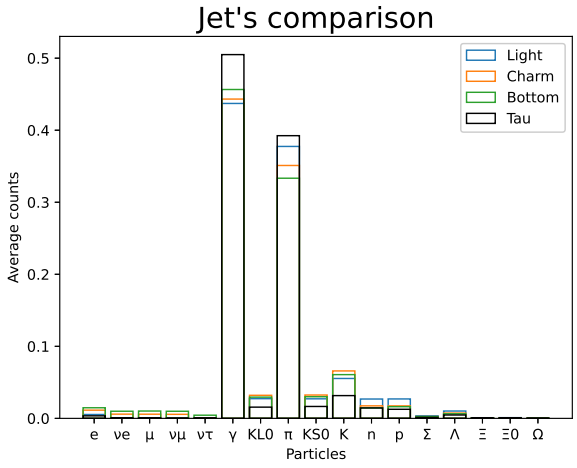

Figure 4.4. Composition in particles of the different type of jets in linear y-scale; KL0 and KS0 are respctively the long-lived neutral Kaon (K-Long) and the short-lived neutral kaon (K-short).

The compositions are quite similar, therefore to see the small differences a second logarithmic plot has been made. The resulting plot is shown in Figure 4.5.

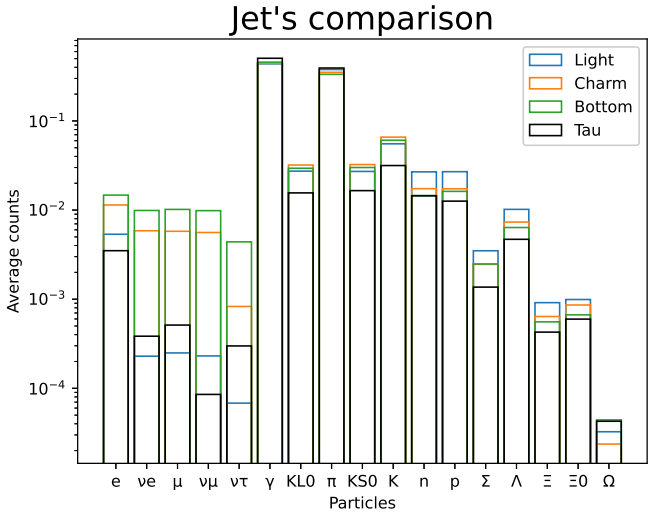

Figure 4.5. Composition in particles of the different type of jets in log y-scale; KL0 and KS0 are respctively the long-lived neutral Kaon (K-Long) and the short-lived neutral kaon (K-short).

The most significant differences have been found for neutrinos and muons. The possibility of knowing whether a particle is a muon or not is a possible input variable for the network, therefore this difference has been further studied. Combining the abundance of muons (shown in Figure 4.4) for every jet type with the average number of particle (shown in Table 4.1), the values displayed in Table 4.2 and Figure 4.6 were found.

Flavour	Count of μ	Number of constituents	Num. of μ in 100 jets
Bottom:	10^{-2}	20	20
Charm:	$0.4 \cdot 10^{-2}$	17	7
Tau:	$5 \cdot 10^{-4}$	8	0.4
Light:	$2 \cdot 10^{-4}$	17	0.3

Table 4.2. Average number of muons in 100 jets for the 4 flavours where the count of muons is for the single jet and the number of constituents is the one obtained on average for a jet.

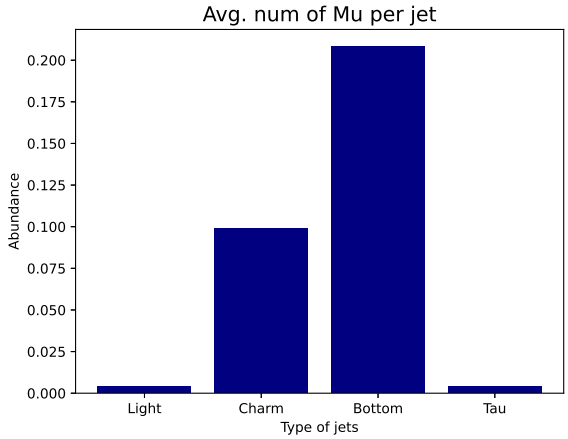

Figure 4.6. Average number of muons per type of jet.

What has been found gives already a difference in abundance of muons between b quarks and c quarks which gets even bigger with respect to the category flagged as 'Light'. From these results we can say that this will surely be a good handle for the Neural Network.

Another feature of interest was the distribution of the pt of the components. A first plot was made but, as the majority of the data were all in the first bin, it was then chosen to make a second x-logarithmic plot in order to have a better understanding of that bin. The plots are shown in Figure 4.7 and in Figure 4.8.

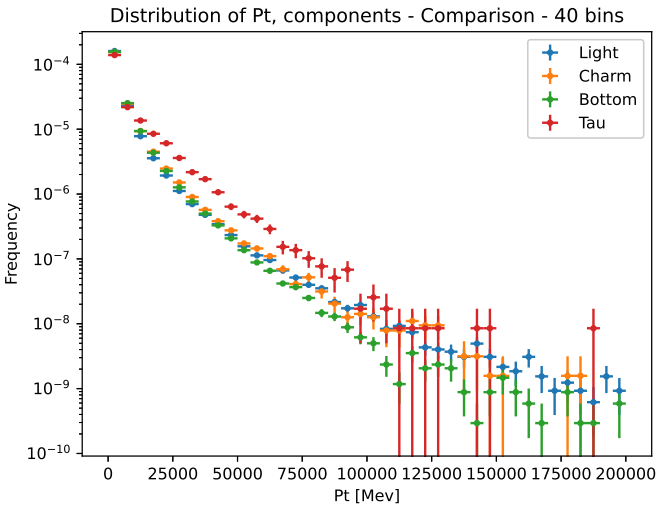

Figure 4.7. Distribution of the p_T of the components; the plot is in linear x-scale.

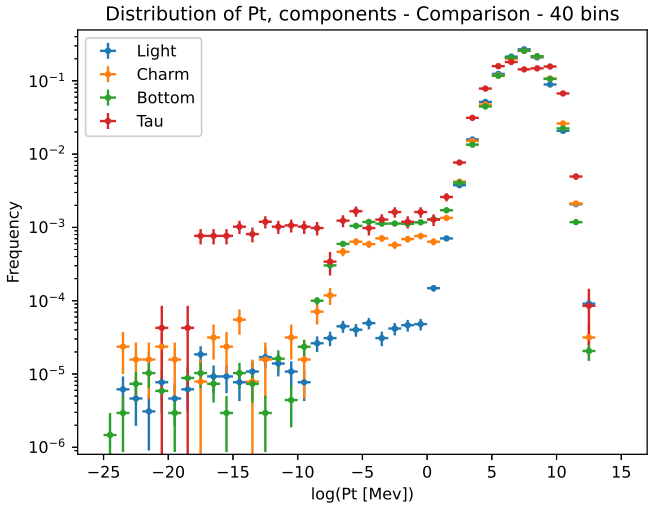

Figure 4.8. Distribution pf the pt of the components; the plot is in logarithmic x-scale.

From both plots, we can easily see that there are small differences that could be exploited in a study in order to discriminate, but what anew stands out is the outstanding difference between taus and all the other flavours.

What has been then calculated was the maximum dr[1] of a component for every jet. Thus, the distance of each component from the center of the the cone of each jet was taken and, by taking the maximum value for each jet the distribution displayed in Figure 4.9 was found.

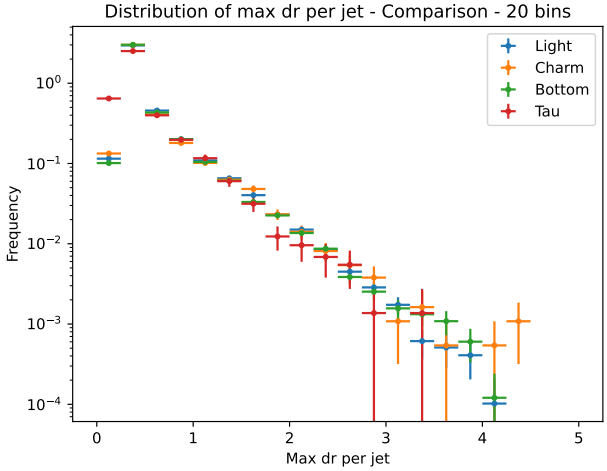

Figure 4.9. Distribution of the maximum dr per jet in linear x-scale.

As the majority of the data are packed up in the bins with smaller dr, it was chosen to study the distribution in logarithmic x-scale. The result is displayed in Figure 4.10.

[1] As defined in the introductory sections.

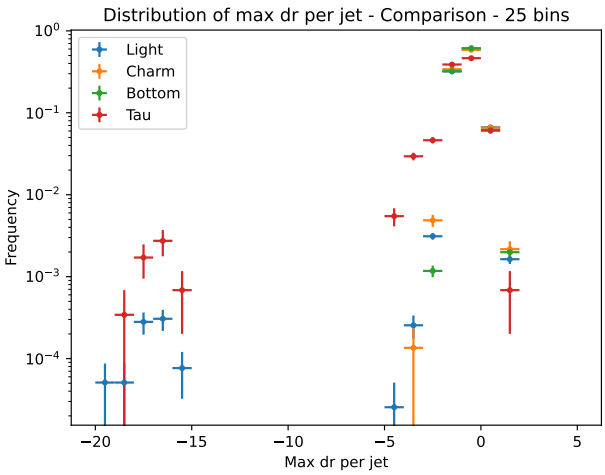

Figure 4.10. Distribution of the maximum dr per jet in logarithmic x-scale.

Studying the values in logarithmic scale, it is clear that they divide in two clusters. The first one, containing the great majority, goes from log values of -5 to around 2. A further zoom was then made in the central area, -2 to 1.5, with a thinner binning. What was thought it could have been found was a diminishment in the dr of the jets, mainly for the heavier b-jets, with an increasing pt of the jet. The plots in Figure 4.11 were then made in different pt regions, but not much of a difference was found.

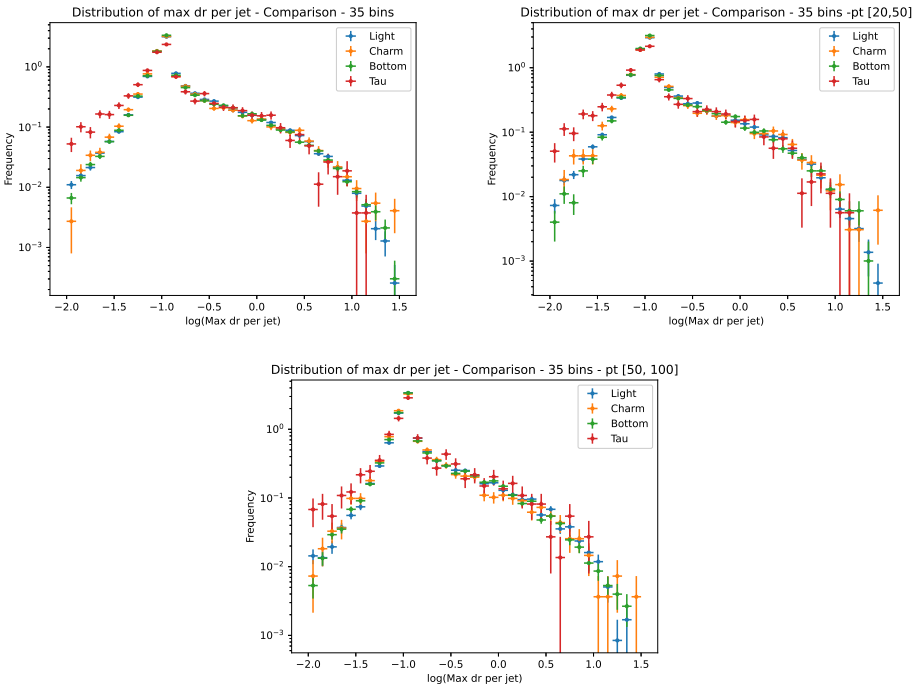

Figure 4.11. Distribution of maximum dr per jet with selections applied to the pt.

For what concerns the smaller cluster, it is situated at big negative logarithmic values and the only flavours found were Taus and Lights. This was then explained by looking at the distribution of the number of components per jet in Figure 4.12, this time made on the data from the final dataset, thus including also muons and neutrinos. What is visible is that the first bin is only filled with Taus and Lights. This means that the so low values in the second cluster are meant to show a value of dr equal to zero; thus, they are the only jets that show to have a single component, as shown in the distribution in Figure 4.12 .

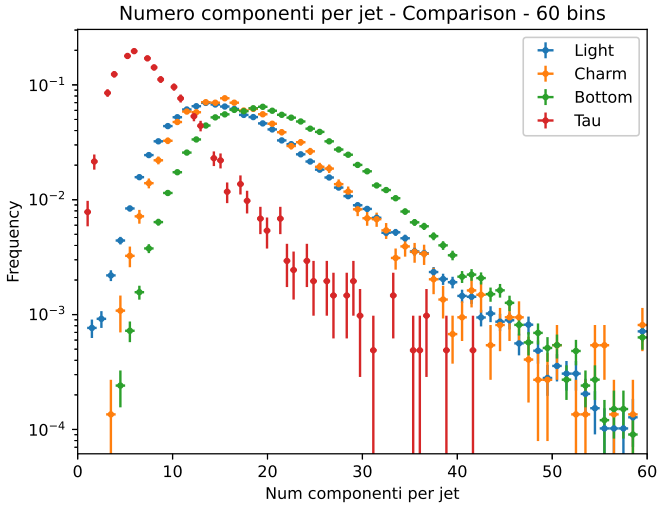

Figure 4.12. Distribution of the number of components per jet.

From this study, what has been found is that muons, together with the number of components per type of jet, can give a good hint for recognising b-jets. What is more, the other inspected variables seem less powerful but still underlying correlation might be picked-up from the NN.

Chapter 5

Elements of Neural Networks

Neural networks are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

Neural Networks (NN) are comprised of node layers, containing an input layer, one or more hidden layers, and an output layer, an exemplification of a NN is shown in Figure 5.1. Each node, or artificial neuron, connects to one another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

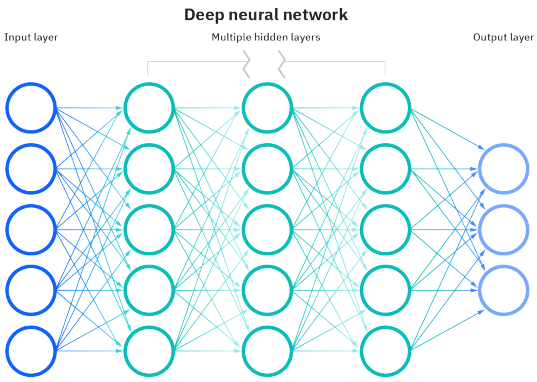

Figure 5.1. Sketch of a Neural Network.

Neural networks rely on training data to learn and improve their accuracy over time. However, once these learning algorithms are fine-tuned for accuracy, they are powerful tools in computer science and artificial intelligence, allowing us to classify and cluster data at a high velocity.

5.1 How do neural networks work?

Each individual node can be seen as its own linear regression model[1], composed of input data (x_i, up to the n^{th} input node), weights (w_i), a bias[2] and an output. An example of the possible formulae is displayed in Equation 5.1 and in Equation 5.2.

$$\Sigma_i w_i x_i + bias = w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n + bias \tag{5.1}$$

$$\text{output} = f(x) = \begin{cases} 1 & \text{if} \quad \Sigma w_1 x_1 + b >= 0; \\ 0 & \text{if} \quad \Sigma w_1 x_1 + b < 0 \end{cases} \tag{5.2}$$

Once the nodes making up the input layer are determined[3], weights are assigned. These weights help determine the importance of any given variable, with larger ones contributing more significantly to the output compared to other inputs. All inputs are then multiplied by their respective weights and then summed. Afterward, the result of the sum is passed through an activation function (in our example this is given in Eq. 5.1). If the output of the activation function exceeds a given threshold (this is set by choosing the bias), it "fires" (or activates) the node (this is the setting of f(x) to 1), passing data to the next layer in the network. This results in the output of one node becoming the input of the next node. This process of passing data from one layer to the next layer defines this neural network as a feedforward network.

As we start to think about more practical use cases for neural networks, like image recognition or classification, supervised learning or labeled datasets will be used to train the algorithm. As we train the model, we'll want to evaluate its accuracy using a cost (or loss) function. This is also commonly referred to as the mean squared error (MSE). This is defined as in Equation 5.3, where: i represents the index of the sample[4], $\hat{y}$ is the predicted outcome, y is the actual value, and m is the number of samples.

$$\text{Cost Function} = MSE = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y_i)^2 \tag{5.3}$$

The ultimate goal is to minimize the cost function to ensure correctness of fit for any given observation. As the model adjusts its weights and bias, it uses the cost function and reinforcement learning to reach the point of convergence, or the local minimum (as shown in Figure 5.2). The process in which the algorithm adjusts its weights is through gradient descent, allowing the model to determine the direction to take to reduce errors (or minimize the cost function). With each training example, the parameters of the model adjust to gradually converge at the minimum.

[1] A linear regression model is a linear approach for modelling the relationship between a dependent variable, y, and one or more independent variables, X.

[2] This is the threshold.

[3] The number of neurons in the input layer is equal to the number of features.

[4] In the case of classification this might be the single jet in the dataset. The output of the NN for the single jet will then be compared to the value of the known label in the dataset.

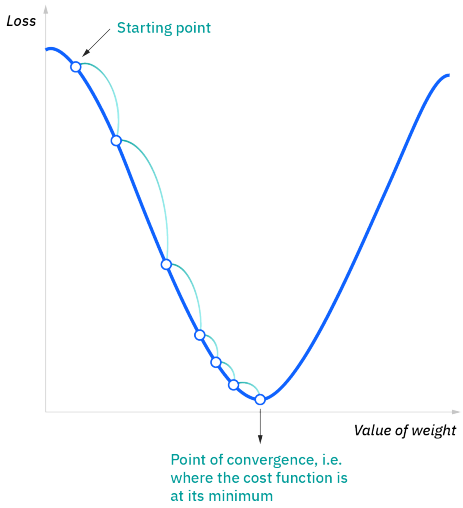

Figure 5.2. Minimization of a Loss Function to a point of convergence.

5.2 Some important vocabulary

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.

Among the most important hyperparameters we find the Batch size. This is the one that defines the number of samples to work through before updating the internal model parameters. It is one of the crucial steps to making sure your models hit peak performance.

The conventional wisdom states that an increasing batch size drops the learners' ability to generalize. It is also claimed that it is because Large Batch methods tend to result in models that get stuck in local minima. The idea is that smaller batches are more likely to push out local minima and find the Global Minima.

Another important hyperparameter is the Learning rate. It refers to the rate at which an algorithm converges to a solution. As it is one of the most important hyperparameters for training neural networks it is very important to set up its value as close to the optimal as possible.

Also, the learning rate doesn't have to have a fixed value. For example, we might define a rule that the learning rate will decrease as epochs[5] for training increase. Besides that, some adaptive learning rate optimization methods modify the learning rate during the training.

To end, it is important to state what signal efficiency and background rejections are. The first term is usually defined as the fraction of the desired events that you actually get[6], while the second one as the fraction of events that don't belong in the sample that are excluded from it[7].

[5] An epoch is one single pass over the entire training set to the network.
[6] It goes down as the false negative rate goes up.
[7] It goes down as the false positive rate goes up.

5.3 DIPS

DIPS is an architecture for Flavour Tagging based on Deep Sets, which models the jet as a set of tracks, in order to identify the experimental signatures of jets containing heavy flavour hadrons using the impact parameters and kinematics of the tracks.

An important point of DIPS is that it treats the tracks in the jet as an unordered, variable-sized set rather than as a sequence, avoiding the need to specify a sequence ordering. Given that the b-hadron decay products do not exhibit any intrinsic sequential ordering, the Deep Sets architecture is also better physically motivated.

This architecture was first employed in particle physics in a phenomenological study on the identification of different types of jets. If p_i is the vector representing the inputs associated with the i^{th} track in the jet, then the Deep Sets architecture applies a neural network (NN) Φ to each track, sums over the tracks, and then applies additional processing on the summed representation with a feed forward NN F, as described in Equation 5.4, where $O(p_1, ..., p_n)$ represents the b-, c-, and light-flavour class probabilities derived from the inputs for the n tracks in the jet[8].

$$O(p_1, ..., p_n) = F(\Sigma_{i=1}^{n}(p_i)) \tag{5.4}$$

The architecture bifurcates the problem into operations over inputs and operations over sets, where the track-network Φ extracts the relevant track features, and the jet-network F accounts for the correlations between the tracks. The permutation invariance of the set is encoded with the permutation invariant sum operation. The presence of this aggregation layer in the architecture encodes information about track multiplicity inside the jet, which is a useful information for identifying b-jets.

In addition, Deep Sets offers the major additional advantage that the operation of processing the tracks in the jet with the Φ network can be easily parallelised. This allows training and evaluation to make significantly more efficient use of GPUs over the non-parallelisable iterative processing.

A simplified scheme of the DIPS architecture is shown in Figure 5.3.

[8]The number n is used to refer to a generic number.

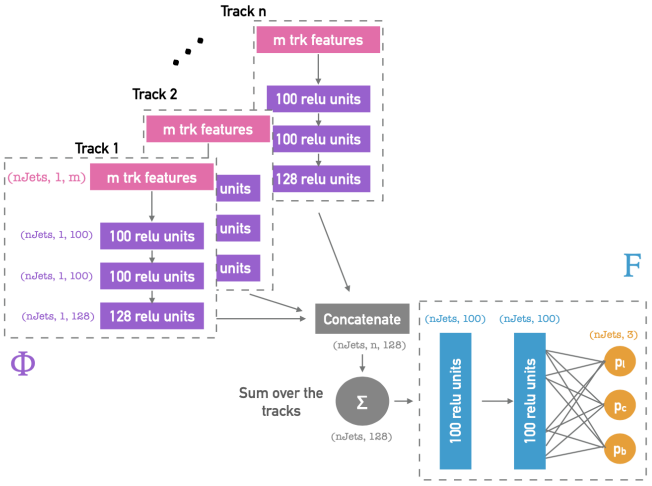

Figure 5.3. Architecture for the DIPS algorithm. From " Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS"

The network has been exploited not giving features about the tracks (as the aim of this study is exactly not using tracks) while giving as inputs the features studied in the previous section, both at constituent and at jet levels.

Chapter 6

The Neural Network

The umami framework[1] has been then used for training DIPS. At first, the variables used for the development of the neural network were: ϕ, η, p_T, dr and energy at constituent level and of the p_T at jet level.

The data where Monte Carlo (MC) samples where the generated event was the production of a pair of quarks top anti-top ($t\bar{t}$) from proton-proton collisions at a centre of mass energy of 13 TeV. The available decay channels where the ones providing at least one lepton in the final state. The system was then evolved considering interactions with the ATLAS detector. In the end, the study has been developed on jets at truth level.

In the preparation of the data for the preprocessing and training selections such as jets with a $p_T > 20$ GeV and $|\eta| < 2.5$ were applied.

6.1 The Preprocessing

The very first thing that has been done with the network was the preprocessing of the data. The motivation for preprocessing the training samples results from the fact that the input datasets are highly imbalanced in their flavour composition[2]. Indeed, there is a large quantity of light jets with respect to the other flavours, they are therefore the main source of background for our search for b-jets. Four steps have been done: preparation, resampling, scaling and writing of the sample. For the first one the different flavours were extracted from the files and separated into flavour-specific files. Also the split in training/validation/evaluation is done at this step. The sample has been divided as explained in Table 6.1.

Training Data :	$\frac{4}{6}$
Validation Data:	$\frac{1}{6}$
Test Data:	$\frac{1}{6}$

Table 6.1. Division of the sample in training/validation/test.

[1] Umami is a framework which can be used for training (most) machine-learning-based taggers used in ATLAS. The link to the whole documentation can be found in the Bibliography.

[2] Such statement has been discussed in Section 5.

The second step is resampling, which is a widely adopted technique for dealing with highly unbalanced datasets. In this specific case there were classes such as the one of the light jets who contained more than half of the data while the charm class less than 10%. Therefore, the resampling consistes in removing samples from the majority class (doing therefore an under-sampling) and / or adding more examples from the minority class (this is called over-sampling); the two possibilities are graphically explained in Figure 6.1. In under-sampling, the simplest technique involves removing random records from the majority class, which can cause loss of information. Another approach can be to tell the network how important samples from each class are.

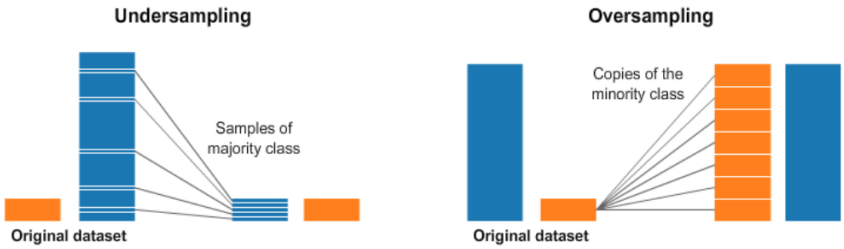

Figure 6.1. Resampling through under- or over- sampling. Picture from Umami documentation.

After the jets are resampled and the training set creation is completed, we need to scale and shift the training jet variables to normalise the range of the independent variables. Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization. For example, many classifiers calculate the distance between two points by the Euclidean distance. If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance. Another reason why feature scaling is applied is that gradient descent converges much faster with feature scaling than without it.

In the final step of the preprocessing, training jets are scaled/shifted and then written to disk, in a format that can be used for training[3].

The result of the preprocessing is visible in Figure 6.2.

[3]In this last step, the collections of jets and track variables are encoded and flattened so that they can be loaded/used for training.

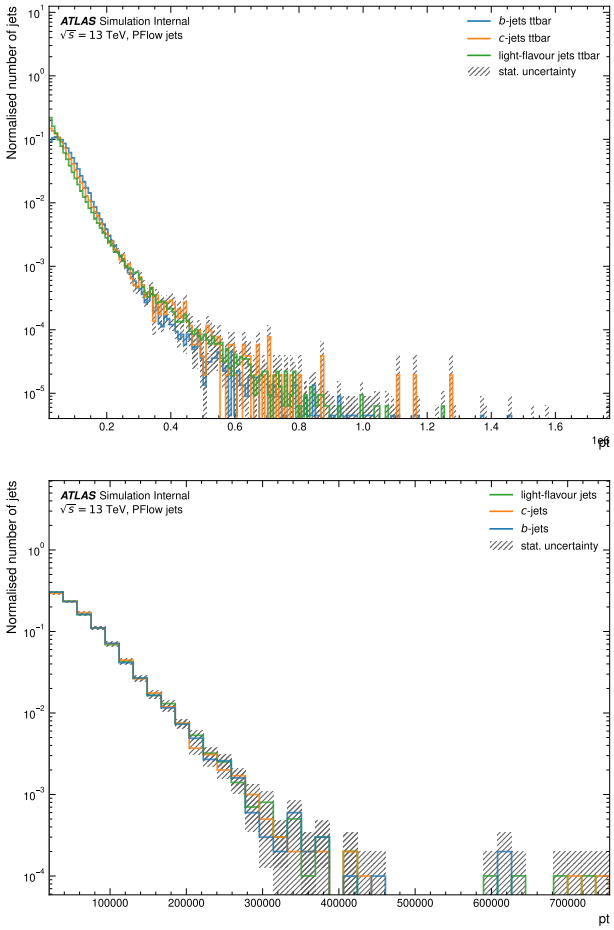

Figure 6.2. Results of Preprocessing: p_T before and after preprocessing.

In the first upper plot is visible the distribution of the p_T before the preprocessing, while in the lower one is displayed the distribution once the preprocessing was applied.

6.2 The Trainings

After the preprocessing, the network is trained using Tensorflow[4] with the Keras[5] backend. The model was able to use GPU resources which drastically shortened the time needed for training.

For the structure of the networks it was opted for: 3 layers of respectively 100, 100 and 128 nodes for the Φ network while 4 layers of 100, 100, 100 and 30 nodes for the network after summing up the outputs of the previous network. Moreover, the option for the Learning Rate reducer was activated and the dropout was fixed to zero. Dropout has been later tested with poor results, it has been therefore not

[4]TensorFlow is an open-source library used for deep learning applications.

[5]Keras is a deep learning API (Application Programming Interface) written in Python, running on top of the machine learning platform TensorFlow.

used for the best models.

The very first trainings have been made with the sample studied in Section 4 and then with a broader sample of 920.000 jets[6]. Due to the relative small sample it was then opted to use the oversampling method instead of using a undersampling as this would have lead to a loss in the available informations. Some first trainings were made but it was clear that bigger samples were needed. With this in mind, a final broader sample which resulted, after the already discussed preprocessing, in a set of $16.5 \cdot 10^6$ jets with an equal abundance of bottom, charm and light flavoured quarks. It was obvious that at this point more computational power was needed. The development of the network has been therefore continued on a remote cluster.

When the first trainings were done on the bigger sample, the obvious choice was to disable the oversampling mode, thanks to the big amount of data available. The accuracy, the loss and the rejection plots were made and what displayed in Figure 6.3. These three variables are important indicators of how good the NN is working and developing. The accuracy used in the following plots is one way of measuring how often the algorithm classifies a data point correctly. Accuracy is the number of correctly predicted data points out of all the data points. More formally, it is defined as the number of true positives and true negatives divided by the number of true positives, true negatives, false positives, and false negatives. A true positive or true negative is a data point that the algorithm correctly classified as true or false, respectively. A false positive or false negative, on the other hand, is a data point that the algorithm incorrectly classified. For example, if the algorithm classified a false data point as true, it would be a false positive. The loss is the function defined in Equation 5.3. While the last variable is the rejection of background (uninteresting) events correctly identified as such. In this last plot there is a double y-scale for both light and c- flavoured jets. All the plots are displayed as the number of epochs increases (displayed on the x-axis).

[6]Up to this point, all the studies were made on a local computer.

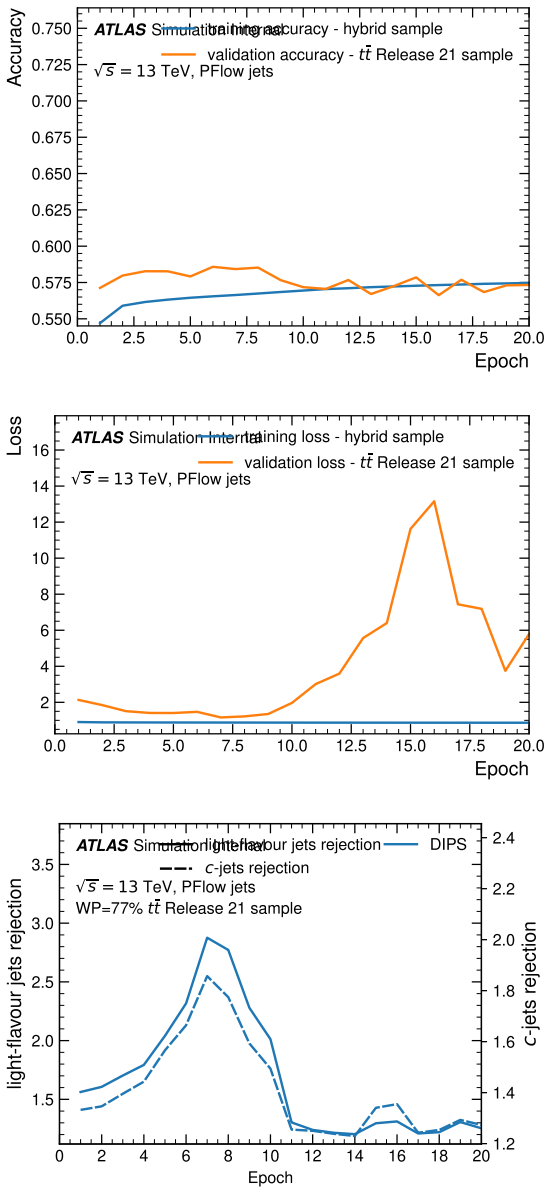

Figure 6.3. Accuracy, Loss and rejection plots from the training of the Full Sample with the downsampling set.

What is clear from this plots is that the model was not having a good performance. In fact, as the number of epochs increases what was found was: a peak of the Loss function, while the aim is to have it going down, a smaller rejection for both light and c- jets and a general unstable trend. Some more training were therefore done with the upsampling mode. The results of the same plots as before are in Figure 6.4.

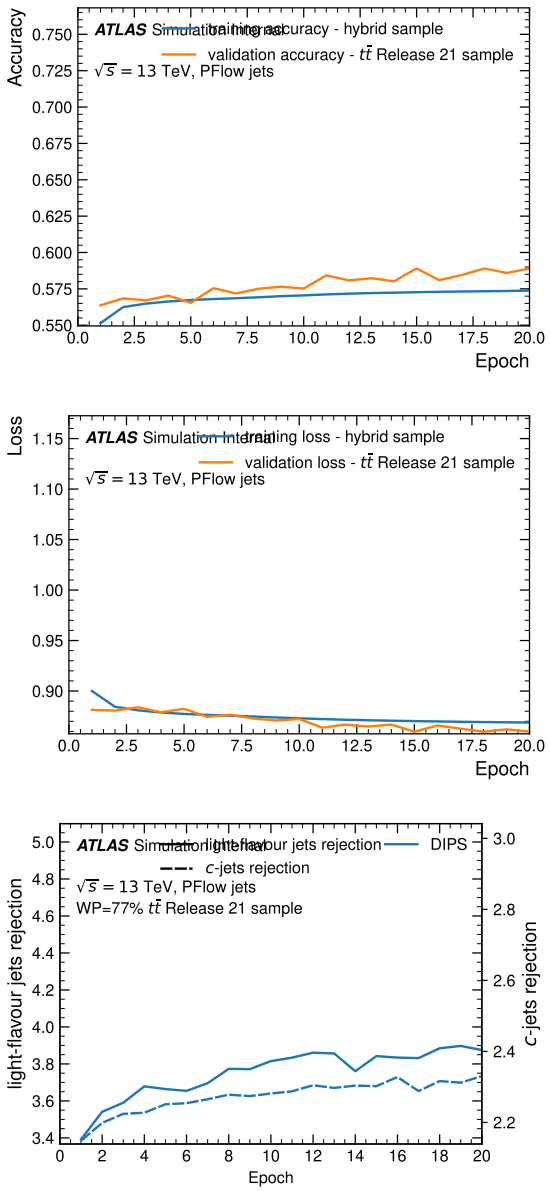

Figure 6.4. Accuracy, Loss and rejection plots from the training of the Full Sample with the upsampling set.

It is clearly visible that this is what is looked after in a good model. Indeed, as the epochs sum up, the loss function goes to zero and both the accuracy and the rejections, on average, have a good tendency to increase. Having obtained such results, it was from this point on preferred to continue with the upsampling mode.

In order to optimize the NN a tuning of the hyper parameters has been done. A various range of values has been studied for: learning rates (LR) between 0.01 and 0.0001, batch sizes (BS) of powers of two between 1024 and 65536 all with a standard number of 20 epochs and at a working point of 0.77. It is important to underline that the fixing of the number of epochs to 20 has been decided as this

could already give a discrimination power to do the tuning nonetheless saving time as the trainings are long due to the big samples. The results are shown in the table in Appendix, from which the following parallel coordinates plot has been made.

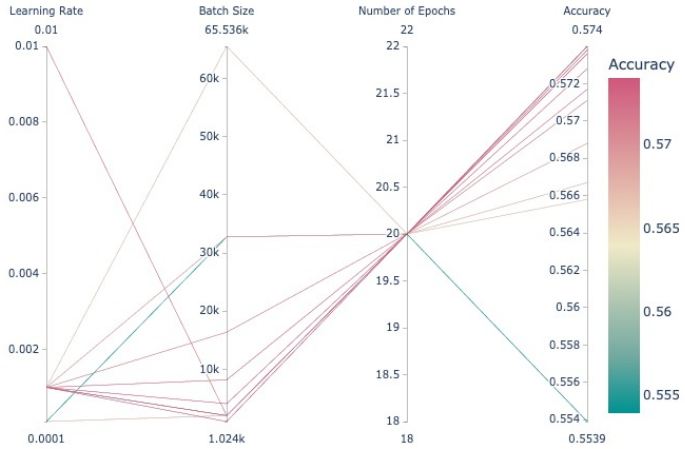

Figure 6.5. Tuning of the NN for $0.01 \leq LR \leq 0.0001$, $1024 \leq BS \leq 65536$.

The result of this study is the finding of the best tuned model, which is found to have the following characteristics (as shown in Figure 6.5): Learning Rate = 0.001, Batch Size = 2048, Working Point = 0.77 and with an Accuracy of 0.5740 after 20 epochs.

The comparison of the ROC curves[7] before and after the tuning has been made by plotting the two curves together. The results are shown in Figure 6.6. When displaying a ROC curve, on the x-axis we find the b-jet efficiency while on the y-axis we get the correspondent background rejection. In the following plots both the ROC curves for the light and the charm jets will be displayed, the first one always traced with a straight line, the second one with a line-dot line. When in a plot there is more than one line of the same type (dotted/continous) this is because multiple models are being compared; different models will be displayed in different colours. The two lower plots are always the ratio between the ROC curves of the upper plot, respectively for the light and for the charm jets. In the comparison, if a curve is above the other, it means that the corresponding model (result of the NN) is performing better as the rejection of the background is higher at every value of b-jet efficiency.

[7]A ROC curve (Receiver Operating Characteristic curve) is a graph showing the performance of a classification model at all classification thresholds.

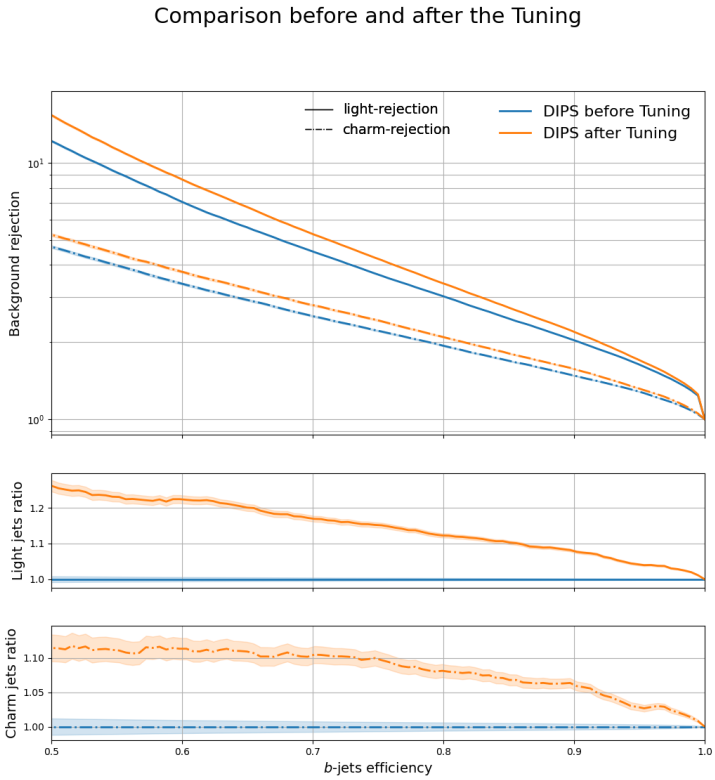

Figure 6.6. Comparison ROC curves before and after tuning.

We find a visible improvement in the ROC curve after the tuning. In the upper plot we see that it has a better background rejection for every b-jet efficiency. In the second and third plot, showing the ratios between the two curves for both light and charm jets, the improvement is percentualized with an increment of around 12% and 8% respectively for the first and the latter.

As discussed in Section 4, the information of whether a particle is a muon or not used as an input variable it is thought to be a good handle for the NN to discriminate the b-jets. The following step has been therefore the natural one of giving this information as an input to the network and study the resulting impact. A great improvement was already signaled by the first plots about accuracy, loss and rejections. The three plots are available in the Appendix. The three of them all show an unexpected good result as in both the two first plots show a validation accuracy and validation loss which is even better than the training ones, while the rejection rates after just 20 epochs go up to 4.50 and 2.8 for light-flavour and c-jets with respect to the previously discussed model with rates of respectively 3.8 and 2.4. Likewise, this substantial improvement is even further supported by the comparison to the ROC curve of the model already improved through tuning. What arises is a decisive gain in the rejection of the light jets, which is the main source of background for this study, with an increment of the percentage of around 25% at a b-jet efficiency of 80%. A not significant improvement was found for the rejection of the charm flavoured jets but they remain stable for b-jet efficiencies above 70%

and in everycase the aim of this study is to preserve at best the signal (b-jets) while reducing to the minimum the main backroung (light-jets). All this is shown in Figure 6.7.

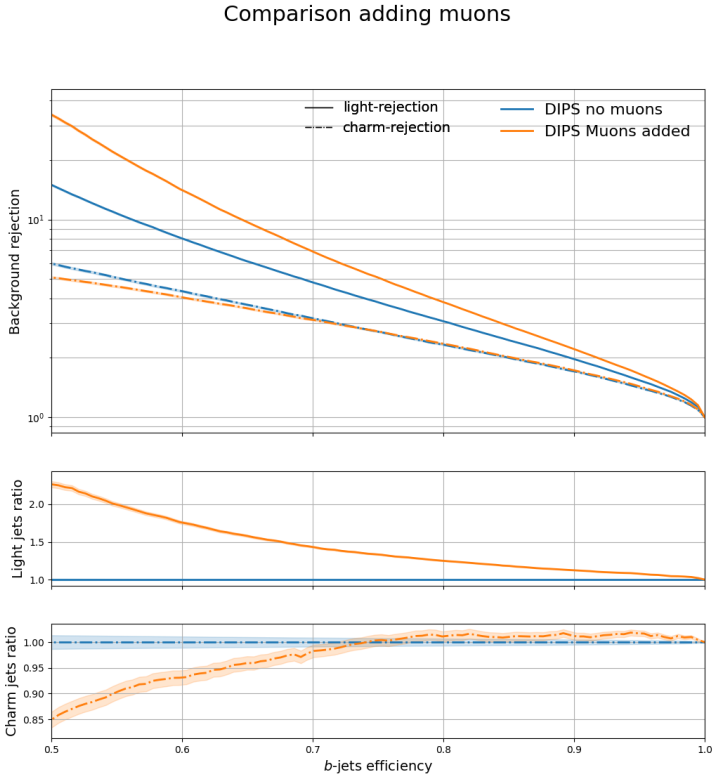

Figure 6.7. ROC curves with the adding of the Mu variable as input.

The next step has been that of trying to train the sample by feeding it with a sample to which a cut on the p_T was applied. This was done as the physics to which we are interested is that of higher energy. Therefore the cuts were made at: 30, 40 and 50 GeV. All the preliminary plots are available in the Appendix. The same trends and statements discussed for the previous model are applicable also to this three. All the model have been then analized at their peak performance at the 20th epoch and the plot in Figure 6.8 has been made.

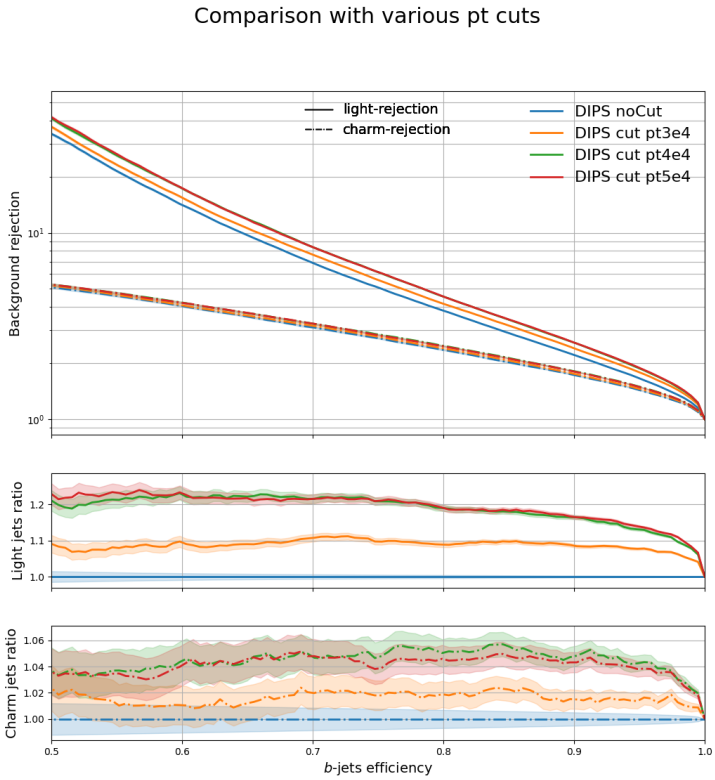

Figure 6.8. Comparison of ROC curves with increasing p_T cuts

What results from this comparison is that with a first cut on the p_T at 30 GeV there is a gain on the rejection of light jets of around 10% for b-jet efficiencies up to 90% and a small improvement also on the rejection of charm flavoured oned. This gained is doubled with the second cut for p_T greater than 40 GeV, which translates in another improvement of 20% in the rejection of the main backgound. However the third cut of 50 GeV seems to be almost aligned to the second one, giving no significant contribution.

In the end, the two best model developed have been studied after a total of 100 epochs. For what concerns the accuracy and the loss what results is a stable improvement until an almost constant value is reached. To better study the rejection rates, the ROC curves of the two models have been compared to the best results reached until now, at the 20^{th} epoch. The comparison is shown in Figure 6.9.

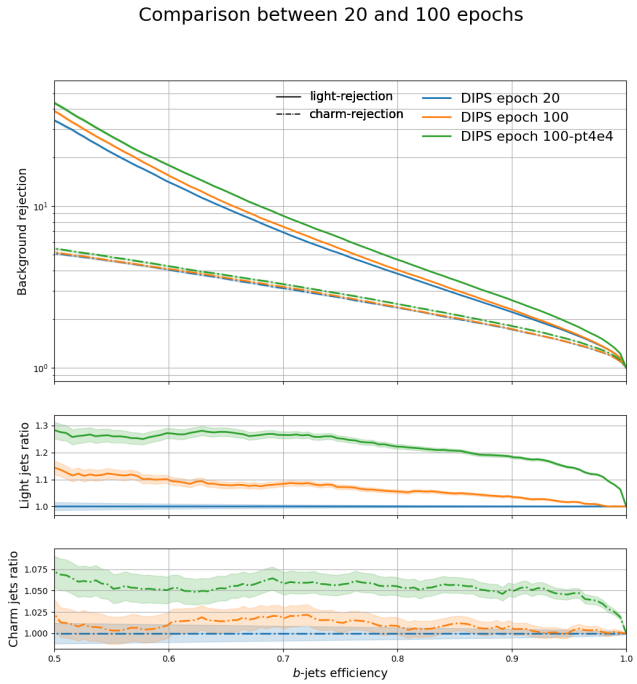

Figure 6.9. Comparison of ROC curves with a maximum of 100 epochs.

Also in this case, we get an improvement for both rejections with a greater one for the one of most interest for this study. At a b-jet efficiency of 80% we get an increase in percentage of around 8% for the model at the 100^{th} epoch while around 22% for the model with a cut at a minimum p_T of 40 GeV. This last is the best model obtained. Its ROC curve is shown in Figure 6.10.

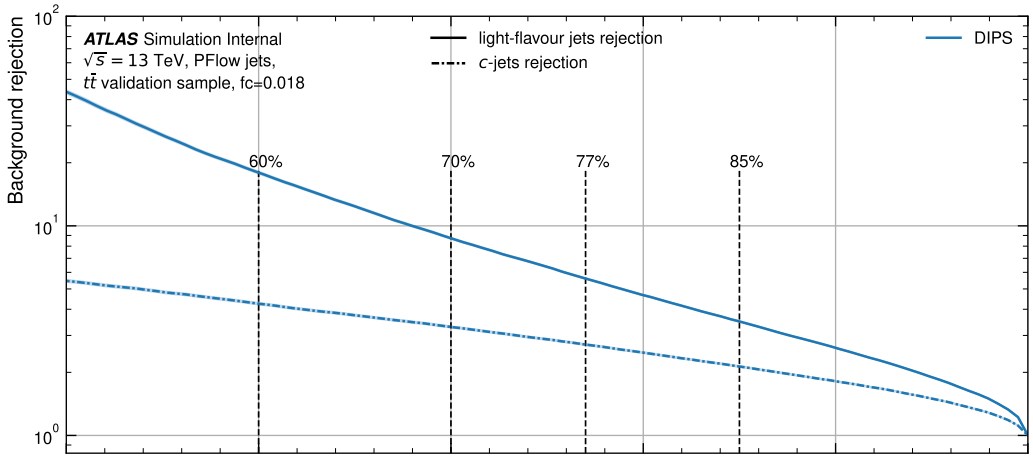

Figure 6.10. ROC curves of the best model after 100 epochs.

6.3 Study of signal and background event

Finally, specific events of 4 jets have been studied with the aim of discerning when the network performs best. Aiming at studying the di-Higgs decay, a signal of 4 b-jets have been studied. In such a sense, the efficiency on the signal events (4 b-jets) have been calculated computationally with respect to the efficiency of b-tagging of a single jet. The results are displayed in Figure 6.11 .

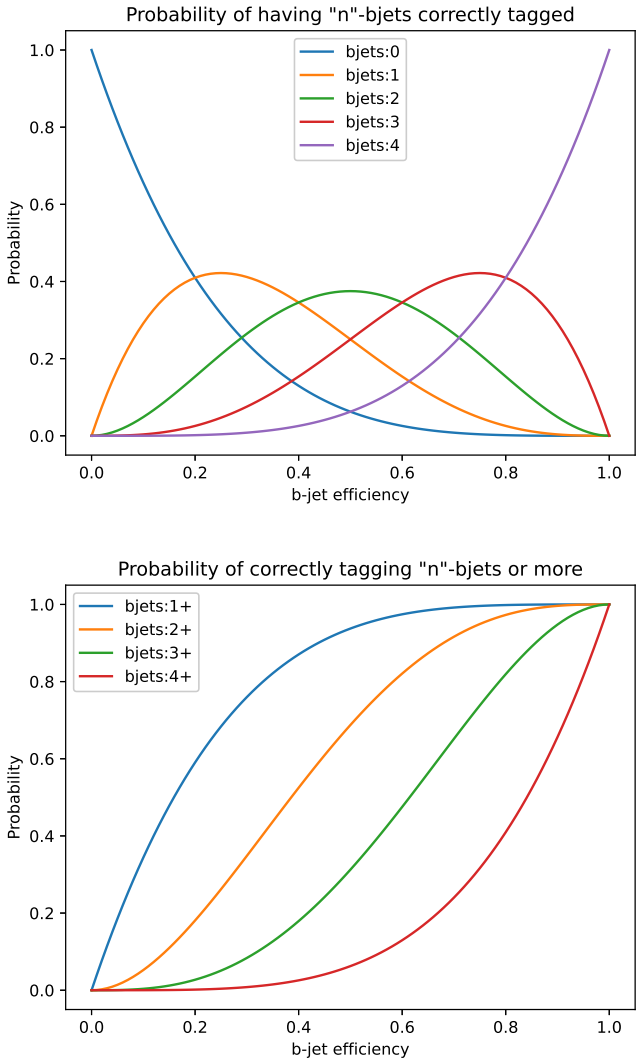

Figure 6.11. Probability of having n (or more) b-jets correctly tagged.

In the first plot, each curve states the probability of having a specific number of b-jets correctly labelled as such, in a range of 0 (minimum number of possible jets) to 4 (the total number of jets in the 4-jet event). From the information of the first plot, the second plot has been made. It shows the cumulative probability for different selections of b-tagged jets (for 1 or more, 2 or more and so on correctly labelled b -jets) as the efficiency on the single b-jet varies.

A similar plot has been made, this time with the aim of studying the fraction of background events, made of 4 light-jets, that passes the selection with the respect to both b-jet efficiency and number of b-jets tagged. However, this time the tagged b-jets are not due to a correct tag but to the a mis-tag. Indeed, the data for this plots are taken from the results of the best trained model obtained at the end of the previous study. The input data have therefore been the background rejection with respect to the b-jet efficiency. Analogous plots have been then made and are shown in Figure 6.12.

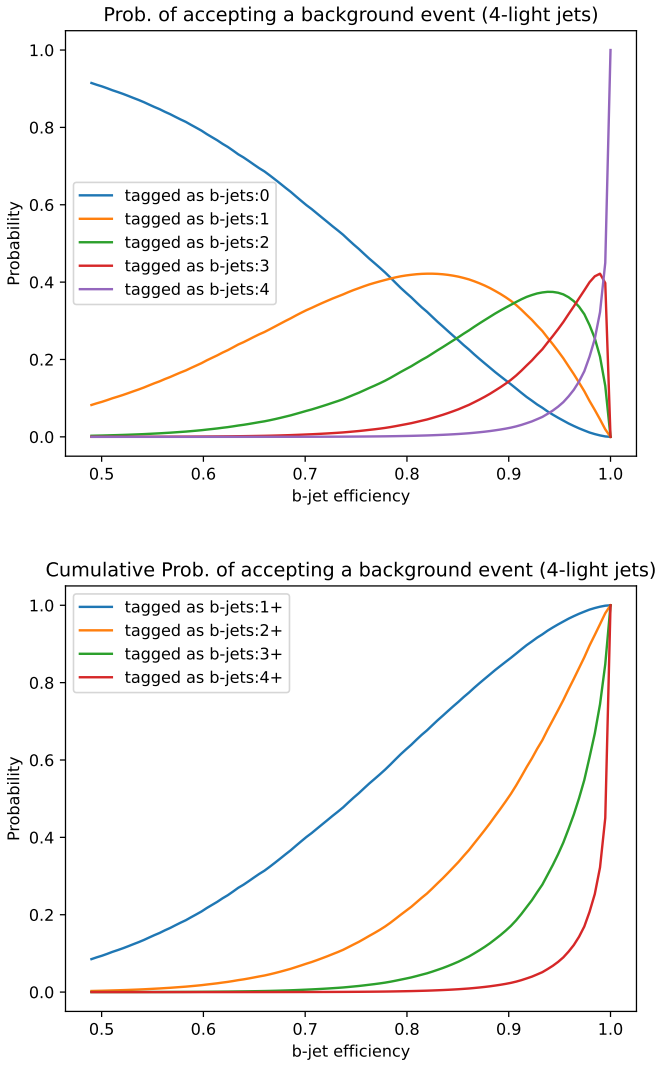

Figure 6.12. Probability of accepting a background event by mistagging n (or more) light jets as b-jets.

The first of the two plots displays the probability of accepting a background event (event with 4 light-jets) with a number of light-jets mis-tagged as b-jets which varies between a minimum of zero and a maximum of 4. The second plot, similarly to the just discussed signal case, is a consequence showing the cumulative probability

of the study.

From what obtained, we see that the harder we work to prevent one class of errors (e.g. b-jets) the more of the other class we get (e.g. light jets). There is therefore a trade off. To end, from this last plots we can get the characteristics of specific working points. As an example, if set for at least 2 b-tagged jets with an efficiency of 85% of b-tagging for a single jet we get: $\sim 99\%$ of the 4 b-jets events (signal) which pass the selection while only $\sim 25\%$ of the events containing 4 light jets (background) are selected. We have therefore reduced the aimed background of a factor 4 while kept almost all the physics we are interested in. Otherwise, we can decide for a minimum of 3 b-tagged jets with an efficiency of b-tagging of a single jet of 80% and what we get is that $\sim 82\%$ of the signal events are kept while only around 5% of the background pass the selection. Our rejection factor is therefore 20.

Chapter 7

Conclusions

As the upgrade of the LHC will highly increase the luminosity, in order to keep the same level of selectivity, a higher background rejection power is needed. The aim of the study was to to investigate the possibility of identifying heavy flavour jets and distinguish them from light flavoured ones using calorimeter-only information in pursuance of testing whether this avenue could be a good one to decrease the uninteresting background. The NN developed had the main aim of tagging b-jets while rejecting other classes. What has been found is that after an appropriate study of the samples and tuning, intelligent use of the input variables and design of the network more than satisfying results can be obtained as it is shown in Figure 6.10. For example for a fixed value of b-jet efficiency of 80%, the associated background rejection is of 2.5 for c-jets and around 5 for light-jets, which is the main source of background as they are the most aboundand class as shown in Section 4. As already discusseds in Section 6.3, which focuses on 4-jets events having di-Higgs in mind, what has been found for an inclusive selction are really promising results which still depend on the working point (WP) which can be varied with respect to the aim of the selection. For example for a b-tagging efficiency of 85% with at least 2 b-tagged jets we keep almost all the signal we are interested about (99%) and at the same time we reduced the unwanted background by a factor 4. At all events, the WP must be chosen in dependance of the aim. To conclude, as the study has been developed at truth level and has led to really encouraging results, we can state that this research has opened the way to hopefully promising further studies on the same path. A first one might be the studying and comparing of the performance of a developed network working at reconstructed level.

Chapter 8

Appendix

8.1 Characterisation of Jets: appendix

The complete set of available variables used in the study of the characterisation of jets were: at jet level

['pt', 'eta', 'energy', 'mass', 'HadronConeExclTruthLabelID', 'PartonTruthLabelID', 'jetPtRank', 'mcEventWeight', 'eventNumber', 'averageInteractionsPerCrossing', 'actualInteractionsPerCrossing', 'nPrimaryVertices']

and at constituent level

['pt', 'eta', 'phi', 'E', 'mass', 'dr', 'Lxy', 'charge', 'flavour', 'pdgId', 'barcode', 'valid']

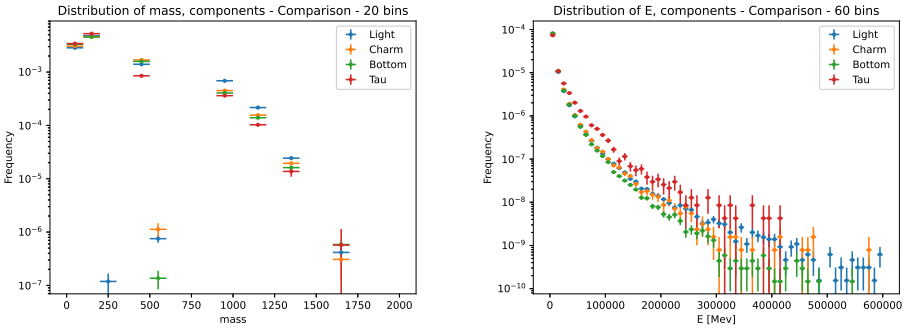

Figure 8.1. Distribution of the masses and energies of the components.

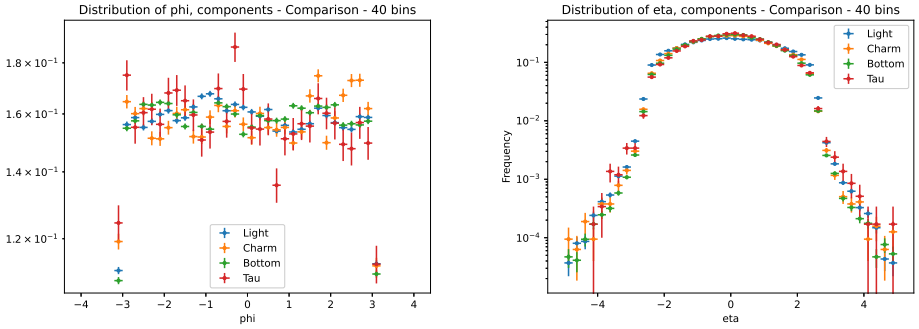

Figure 8.2. Distribution of the values of phi and eta of the components.

8.2 The Neural Network: appendix

Learning Rate	Batch Size	Epochs	Accuracy
0.0001	32768	20	0.5539
0.001	16384	20	0.5711
0.001	32768	20	0.5688
0.001	8192	20	0.5728
0.001	65536	20	0.5658
0.001	4096	20	0.5738
0.001	2048	20	0.5740
0.001	1024	20	0.5736
0.001	2048	20	0.5740
0.01	2048	20	0.5717
0.0001	2048	20	0.5667

Table 8.1. Table of tuning parameters for the NN.

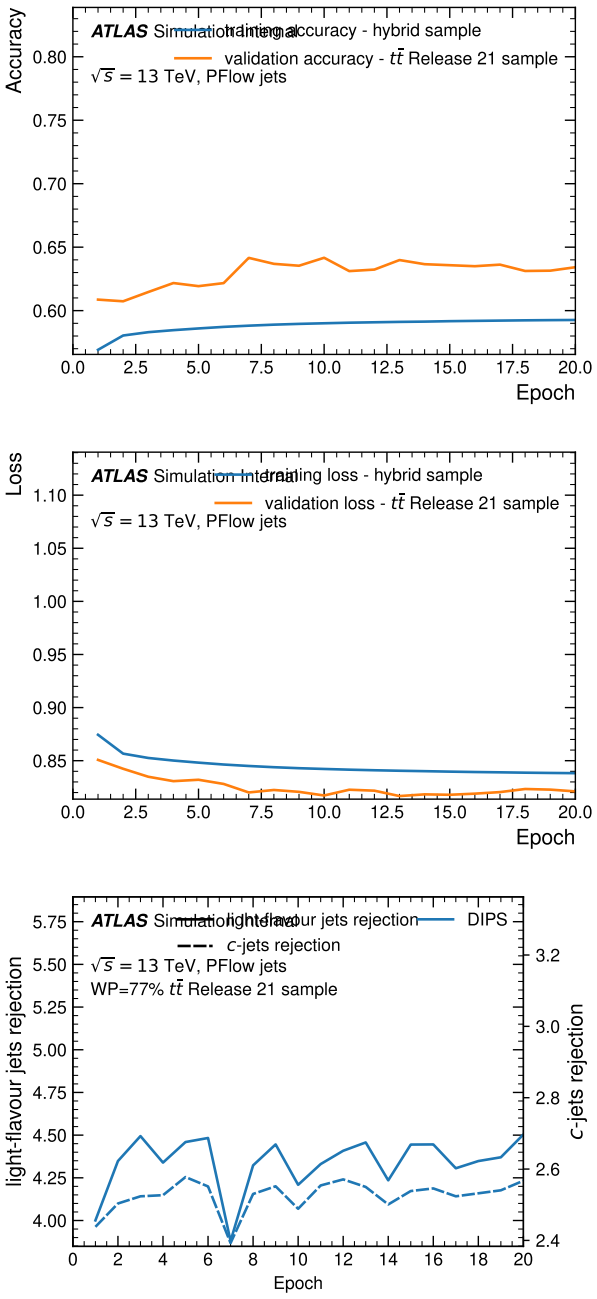

Figure 8.3. Accuracy, Loss and rejection plots from the training of the Full Sample with Mu variable as input.

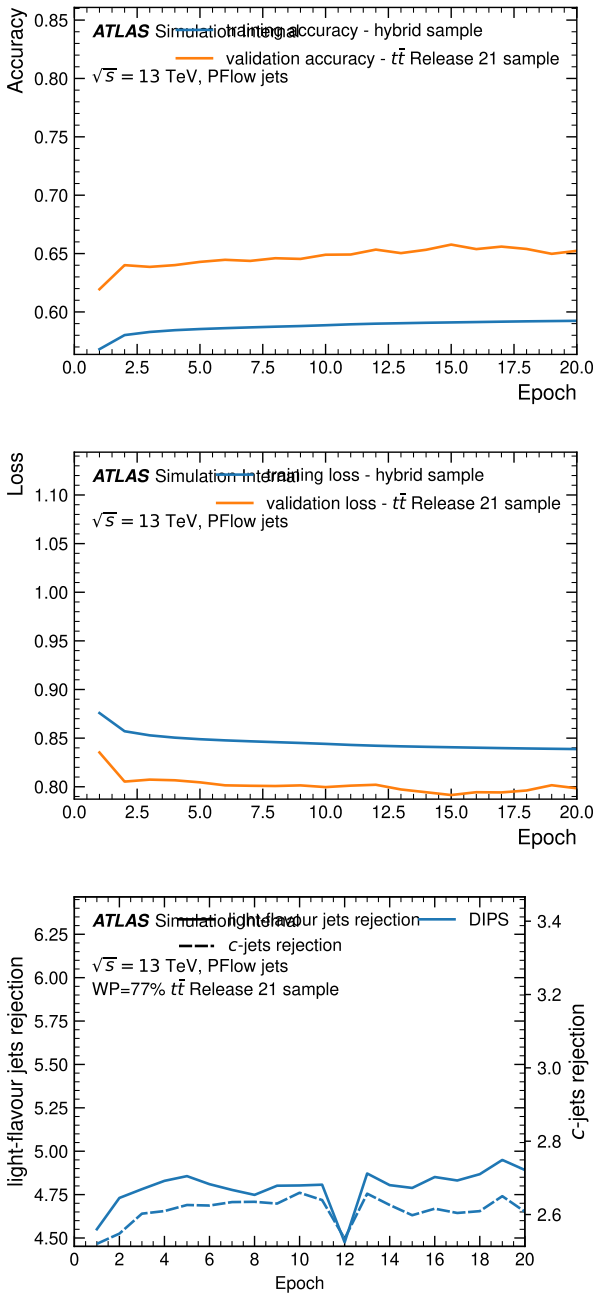

Figure 8.4. Accuracy, Loss and rejection plots from the training of the Full Sample with pt-cut of 30 GeV.

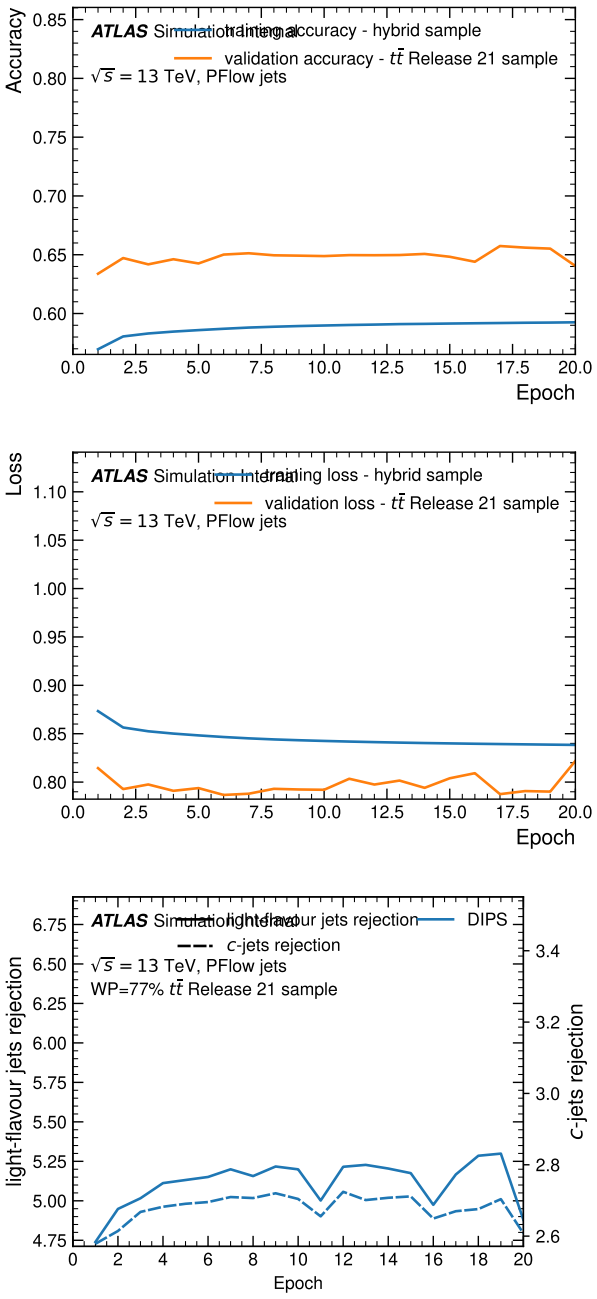

Figure 8.5. Accuracy, Loss and rejection plots from the training of the Full Sample with pt-cut of 40 GeV.

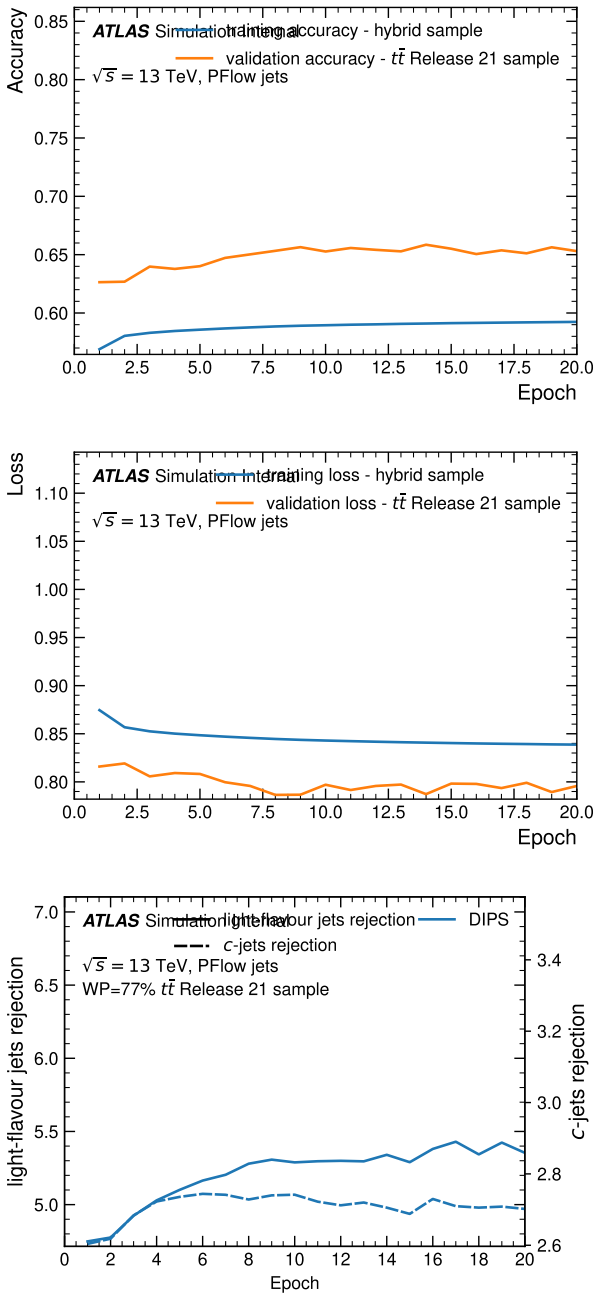

Figure 8.6. Accuracy, Loss and rejection plots from the training of the Full Sample with pt-cut of 50 GeV.

Bibliography

[1] P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020)

[2] Veatch, Jason. (2022). Searches for Resonant Scalar Boson Pair Production Using Run 2 LHC Proton-Proton Collision Data. Symmetry. 14. 260. 10.3390/sym14020260.

[3] arXiv:2106.03584

[4] Michelangelo L. Mangano; CERN, TH Division, Geneva, Switzerland; INTRODUCTION TO QCD

[5] ATLAS Experiment © 2022 CERN

[6] Cern Site at "https://home.cern"

[7] Umami Documentation at "https://umami-docs.web.cern.ch"

[8] Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS - The ATLAS Collaboration