Master Thesis

Significant improvement of Quark/Gluon separation with the ATLAS detector at the LHC

（LHC-ATLAS 実験におけるクォーク．グルーオンの識別の向上）

CHENG-YU LU

The University of Tokyo

Graduate School of Science

January 2022

Abstract

Higgs boson was discovered at the Large Hadron Collider (LHC) in 2012 and many measurements were performed for understanding its properties. The Higgs mechanism predicts the mass of elementary particles and provides the source of the electroweak symmetry breaking. The vector boson scattering with the same leading order term as Feynman diagrams gives an essential way to test the relation between the Higgs mechanism and electroweak symmetry breaking. In addition, beyond the standard model physics predicts other Higgs sectors, denoting another contribution to the vector boson scattering process.

The semileptonic vector boson scattering process provides more statistics than the full-leptonic process and fewer QCD backgrounds than the full-hadronic process. Hence, this thesis aims to improve the sensitivity of the semileptonic process. There are four quarks in the event, two from the vector boson decay, the other two from the spectator quarks, which originate from partonic quarks. The considerable gluons produced due to the QCD effect are the major backgrounds in this channel.

Machine-learning models to separate quark-induced hadron jets from gluon-induced ones (q/g tagging) are developed in this thesis. Several kinds of neural network q/g tagging models are studied and found to have significant improvement compared to the conventional boosted decision tree machine learning model. The best neural network q/g tagging model has improved the background rejection rate by approximately 10% than the boosted decision tree model. Developed q/g tagging is applied to the semileptonic vector-boson scattering analysis, and the sensitivity is improved by around 8.2%. There is still room to improve the sensitivity more in the future analysis, which is discussed in the last part of this thesis. As a result, the efficiency of searching signal became better but still have space to develop.

Acknowledgments

I would like to express my sincere thanks to everyone who helped me complete this thesis, especially the following people.

To begin with, I am deeply indebted to my supervisor, Prof. Shoji Asai for supporting me not only in the research but also in getting used to my student life abroad. He gave plenty of novel ideas and his expansive view of particle physics always provides another insight.

I would also like to extend my incredible gratitude to Dr. Takuya Nobe. His knowledge of theoretical and experimental particles physics was awe-inspiring. His precious advices have improved my study and analysis significantly during the discussions. I have learned a lot about experimental particle physics from his fantastic foresight.

Without the support and nurturing of Dr. Masahiko Saito, it is impossible to finish this study. He has a deep understanding of particle physics and machine learning, which fascinates me immensely. He has warmly helped my studies from the experimental particle physics knowledge to the technical details of the analysis.

I very much appreciate Prof. Junichi Tanaka, who invited me to participate in the seminar. I am grateful to have his support that gave me a stage to present my research and have a deeper discussion.

I would like to give special thanks to Prof. Koji Terashi, Prof. Yasuyuki Okumura and Dr. Tatsuya Masubuchi. The discussions in the seminar are always impressed me and have vastly improved the quality of my studies.

Finally, I must thank my parents for their support throughout my life. Despite the pandemic, they still encourage me to chase my dreams and always stay positive, giving me many strengths. Also, thanks to my little sister, who always brings joy to me.

Contents

Chapter 1

Introduction

The ultimate purpose of particle physics is to discover the equation of motion which is able to describe elementary particles in our universe. Although the standard model (SM) of elementary particle physics predicts numerous particle phenomena, there are still several unexplained issues. For instance, dark energy [1], dark matter [2], Higgs boson mass [3] and strong CP problem [4] are the mysteries needed to uncover. These new physics beyond SM (BSM) are believed to exist at the TeV-scale or higher energy region.

The Large Hadron Collider (LHC), which the center-of-mass energy $\sqrt{s} = 13$ TeV, is the largest and highest energy collider in the world. TeV-scale energy experiment gives us a chance to examine the properties of elementary particle physics. The ATLAS detector at the LHC is built to investigate the latest high-energy physics research. The ATLAS [5] and CMS [6] experiments at the LHC made a milestone for the discovery of the Higgs boson in 2012, which means the SM is finally completed. In order to understand the beyond SM (BSM) physics, LHC will be upgraded to High Luminosity LHC (HL-LHC), providing a center-of-mass energy of $\sqrt{s} = 14$ TeV with an integrated luminosity up to 3000 fb^{-1}.

Vector Boson Scattering (VBS) is a sensitive probe to examine the electroweak symmetry breaking (EWSM) in SM and BSM physics. If the coupling between the vector bosons and Higgs deviate from the SM prediction, the VBS cross-section increase when the center-of-mass energy rises to the higher energy scale. Moreover, the cross-section of the VBS process is expected to increase from new resonances, originating from an extended Higgs sector predicted from the BSM model [7]. SM examination and BSM investigation imply the VBS process's importance, motivating this thesis to improve the VBS signal searching.

The VBS process decays to two vector bosons (W or Z boson). Both W and Z bosons have hadronic and leptonic decay channels. W boson decays into qq and $\ell\nu$, hadronically and leptonically, respectively. Z boson decays hadronically and leptonically to qq and $\ell\ell$ or $\nu\nu$. ($\nu\nu$ is not considered in full-leptonic decay since the analysis strategy is different from $\ell\ell$.) The full-leptonic decay has the most significant signal-to-background ratio compared to other processes. The full-leptonic channel was observed during Run I at the LHC and confirmed by Run II. Besides the full-leptonic VBS process, full-hadronic and semileptonic are also measured by the ATLAS [8] and CMS collaborations [9]. The full-leptonic VBS process has a severe statistical problem because of the low branch ratio of leptonic decays in vector bosons. The full-hadronic VBS process research is highly affected by quantum chromodynamics (QCD), causing a large amount of background. Hence, this thesis focuses on the semileptonic decay process.

Since the gluons generated from the QCD are the primary background particles in the semileptonic VBS process, it is essential to develop a tool to separate the quarks and gluons, which is known as quark/gluon tagging (q/g tagging). The q/g tagging has been developed for a while since gluons usually represent an annoying background source. Furthermore, neural network models have played a pivotal role in machine learning in the last decades. Compared to conventional machine learning such as BDT, neural network models are able to make use of more low-level information, providing a way to avoid information loss. This thesis discusses several neural networks q/g tagging models and examines the improvement. The validation of q/g tagging models considered in this thesis is taken by the 1-lepton semileptonic VBS process for convenience.

This thesis consists of five chapters: Chapter 2 introduces some brief particle physics theory and the background knowledge for this study. The inner detector, calorimeters and some functions in the ATLAS are demonstrated in Chapter 3. The discussion of quark/gluon tagging models and their performance is shown in Chapter 4. Chapter 5 describes the definitions of the physics objects of the 1-lepton semileptonic VBS process. Then, the application and improvement of q/g tagging are examined by applying the q/g tagging models to the 1-lepton semileptonic VBS process. Chapter 6 gives a breif conclusion of entire thesis.

Chapter 2

Background Knowledge

Chapter 2 gives a brief introduction to the semileptonic VBS analysis. Section 2.1 describes the SM of elementary particle physics. Section 2.2 shows the importance of weak vector boson scattering and the mechanism of the process. Since jet phenomenology takes a vital role in the VBS scattering and the q/g tagging study in this thesis is also making use of jets, the introductions to the jet structure are explained in Section 2.3.

2.1 The Standard Model of particle physics

The success of the Standard Model in explaining the matter interactions governed by three fundamental forces is astounding. There are mainly three parts: the gauge interaction, the fermion fields, and the Higgs mechanism. In Section 2.1.1, the introduction of the gauge interaction is given to explain the interaction between fermion and gauge particles. The fermion fields due to the matters themselves are described in Section 2.1.2. Section 2.1.3 introduces the Higgs mechanism used to demonstrate the masses of the elementary particle due to the EWSB.

2.1.1 Gauge Interaction

One of the most remarkable insights in particle physics is that interactions are governed by the symmetry principle. The gauge interaction is one of symmetry mechanism and the gauge transformation is shown in Equation 2.1.

$$\psi(x) \rightarrow \psi'(x) = \hat{U}(x)\psi(x) = e^{iq\chi(x)}\psi(x), \tag{2.1}$$

where ψ is a complex field and qχ(x) here can be seen as a gauge parameter. The phase qχ(x) varies at all points in space-time. Other than $U(1)$

transformation, there are also $SU(2)$ transformation dictating electroweak interaction and $SU(3)$ transformation which describes strong force in the gauge transformation.
Using $U(1)$, $SU(2)$ and $SU(3)$ symmetry, the SM Lagrangian density is derivated as following equation,

$$\mathcal{L} = -\frac{1}{4}(\mathrm{F}^{\mathrm{a}}_{\mu\nu})^2 + \bar{\psi}(\mathrm{i}\gamma^{\mu}\mathrm{D}_{\mu}\psi) + \mathrm{y}_{\mathrm{ij}}\bar{\psi}_{\mathrm{i}}\psi_{\mathrm{j}}\phi + \mathrm{h.c.} + |\mathrm{D}_{\mu}\phi|^2 + \mu^2\phi^{\dagger}\phi - \lambda(\phi^{\dagger}\phi)^2, \tag{2.2}$$

where $\mathrm{F}^{\mathrm{a}}_{\mu\nu} = \partial_{\mu}\mathrm{A}^{\mathrm{a}}_{\nu} - \partial_{\nu}\mathrm{A}^{\mathrm{a}}_{\mu} + \mathrm{gf}^{\mathrm{abc}}\mathrm{A}^{\mathrm{b}}_{\mu}\mathrm{A}^{\mathrm{c}}_{\nu}$ and $\mathrm{D}_{\mu} = \partial_{\mu} - \mathrm{i}g\mathrm{A}^{\mathrm{a}}_{\mu}\mathrm{t}^{\mathrm{a}}_{\mathrm{r}}$ for gauge field A, fermion field ψ, scalar field ϕ, Yukawa coupling y_{ij}, Higgs self-coupling λ, representation matrices $\mathrm{t}^{\mathrm{a}}_{\mathrm{r}}$, gauge coupling g, structure constant $\mathrm{f}_{\mathrm{abc}}$, matrix representation of the Clifford algebra γ^{μ} and h.c. stands for Hermitian conjugate.
Investigating SM Lagrangian closer, fermions do not suppose to have masses in nature in Equation 2.2. However, physicists have measured the mass term for fermion particles, which is considered from fundamental symmetry broken. The Higgs mechanism explains the origin of masses for elementary particles by the EWSB.

2.1.2 Elementary Particle

The electron, the electron neutrino, the up quark and the down quark are known as the first generation particles. As higher energy goes in the high-energy collider, further complexity of elementary particles other than the first generation is discovered. Second generation and third generation particles have the same properties as first generation particles except for their masses. Besides, the gauge boson particles are also found at high-energy collision experiments. The gauge boson is regarded as the origin of three fundamental forces. With the discovery of the Higgs boson at LHC, all of the elementary particles predicted by the SM have now been observed.
The properties of all elementary particles are summarized in Table 2.1. The fermion particles such as leptons, neutrinos and quarks are dictated by the Fermi-Dirac statistics, indicating the proportional relation between their masses and Yukawa couplings. There are four kinds of gauge boson particles which are spin-1 particles and are described by Bose-Einstein statistics. The couplings of electroweak force are demonstrated by photon γ and $\mathrm{W}^{\pm}/\mathrm{Z}$ boson whose masses are approximately 80/91 GeV. On the other hand, the gauge boson for strong force is massless gluon g and it has eight color multiplets. The Higgs particle is also a boson particle but with spin-0.

Table 2.1: The essential quantum number of the elementary particles in the SM. J and Q stand for the spin and the electric charge.

Particle	J	Q
e,μ,τ	1/2	-1
$\nu_{\mathrm{e}},\nu_{\mu},\nu_{\tau}$	1/2	0
u,c,t	1/2	2/3
d,s,b	1/2	-1/3
γ	1	0
$\mathrm{W}^{\pm}$	1	±1
Z	1	0
g	1	0
h	0	0

2.1.3 Higgs Potential and Electronweak Symmetry

The Higgs mechanism represents an integral sector of the SM. Without the Higgs boson, the $\mathrm{W^+W^-} \rightarrow \mathrm{W^+W^-}$ scattering will violate unitarity at a center-of-mass energy of TeV scale. The unitarity violation comes from $\mathrm{W_L W_L} \rightarrow \mathrm{W_L W_L}$ scattering, where $\mathrm{W_L}$ indicates W boson is polarized longitudinally. As a result, the violation can be associated with the massive W bosons originating from the Higgs mechanism. The unitarity violation in WW scattering cancels out if the exchange of a scalar boson is included. This scalar boson is the Higgs boson which can solve the violation issue, as shown in Figure 2.1.
Consider a complex scalar field Equation as 2.3,

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + \mathrm{i}\phi_2), \tag{2.3}$$

which the Lagrangian can be written as following,

$$\mathcal{L} = (\partial_\mu \phi)^*(\partial^\mu \phi) - \mathrm{V}(\phi) \quad \text{with} \quad \mathrm{V}(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2, \tag{2.4}$$

then substituting the Equation 2.3 into the Equation 2.4,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_1)(\partial^\mu \phi_1) + \frac{1}{2}(\partial_\mu \phi_2)(\partial^\mu \phi_2) - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2. \tag{2.5}$$

For the convex shape potential to have a finite minimum, λ must be a positive number. Because of global $U(1)$ symmetry, Equation 2.4 is invariant after $\phi \rightarrow \phi' = e^{i\alpha}\phi$ transformation. The ground state occurs when $\mu^2 < 0$ is known as EWSB. Then, the potential can be defined by

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2, \tag{2.6}$$

where the v in Equation 2.6 is called vacuum expectation value. Consequently, $W^{\pm}/Z$ stands for three of degrees of freedom (The mixture of the $SU(2)\times U(1)$). The remaining one degree for Higgs potential is Higgs boson.

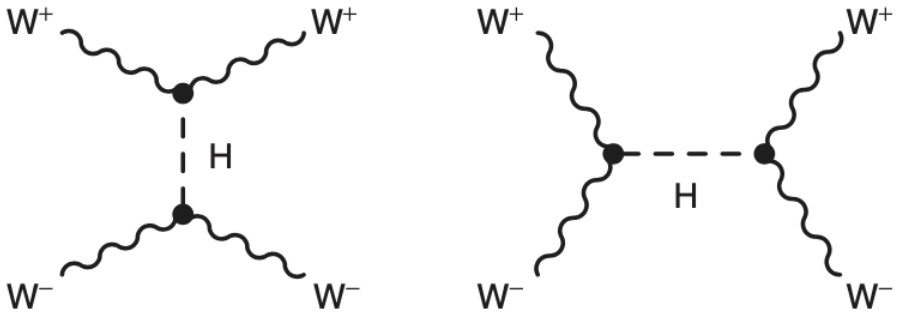

Figure 2.1: The mass of W boson originating from Higgs boson makes the local gauge invariance conserved in SM model.

2.2 Weak Vector Boson Scattering

The EWSB and Higgs boson properties are often examined by the VBS processes since the leading order Feynman diagrams are only related to electroweak bosons or Higgs bosons. VBS processes are also considered as a window to discover the portent of the BSM. For instance, BSM predicts that there are heavy particles that may modify the dynamics of electroweak interactions, which results in the increment of cross sections [10]. It can be measured in the VBS process. The Feynman diagram of VBS is shown in Figure 2.2. The details of the VBS decaying process will be discussed in Section 2.2.1.

2.2.1 VBS Processes

VBS processes consist of the Z bosons or W bosons scattering. The total channels there are $W^{\pm}W^{\pm}, W^{\pm}W^{\mp}, W^{\pm}Z$ and ZZ. The $W^{\pm}W^{\mp}$ scattering has a dominant contribution compared to the $W^{\pm}W^{\pm}$ scattering process and VZ process, where V represents the vector boson.
W boson decays into mainly $l\nu$ and qq. On the other hand, Z boson also has dominant qq decay which is approximately 70% and less chance decay to $\nu\nu$ and $\ell\ell$. The branching fractions for the W and Z bosons are given in Table

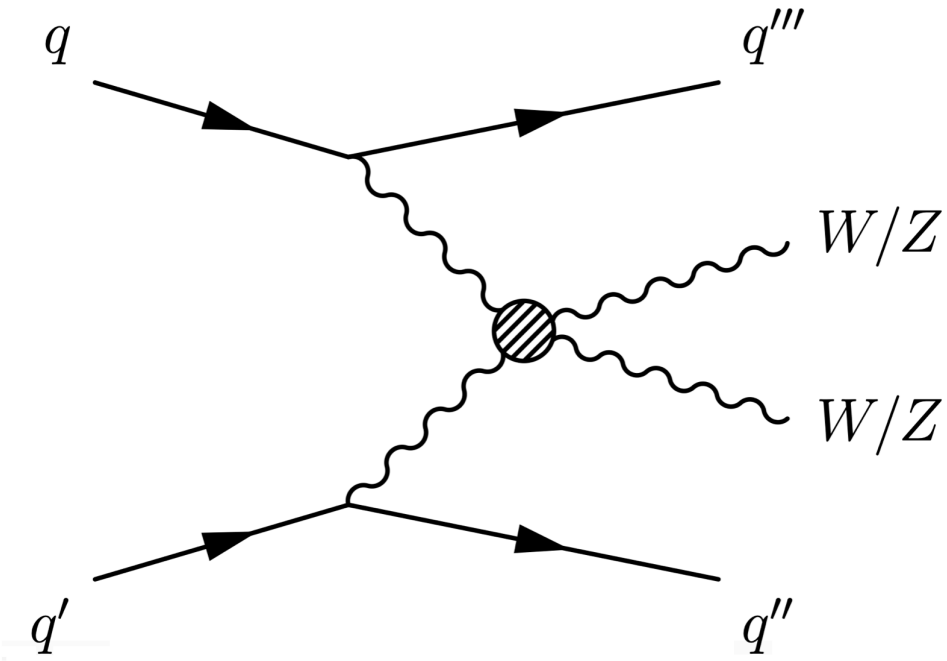

Figure 2.2: Two vector bosons are produced due to the proton-proton collision and then scattering each other.

2.2. Both Z and W bosons have leptonic and hadronic processes, implying full-leptonic, semileptonic and full-hadronic decay in the VBS final channel. ZZ $\rightarrow 4\nu$ final state is counted as full-invisible because neutrinos are invisible in the ATLAS detector. The decay branching ratio is shown at Table 2.3.

Table 2.2: The branching fractions of the W boson and Z boson are on the left hand and right hand sides. Values are obtained from Reference [11].

Particle		Decay	Branching Ratio
W	$\longrightarrow$	qq	67.41±0.27%
		eν	10.71±0.16%
		$\mu\nu$	10.63±0.15%
		$\tau\nu$	11.38±0.21%

Particle		Decay	Branching Ratio
Z	$\longrightarrow$	qq	69.91±0.056%
		ee	3.363±0.004%
		$\mu\mu$	3.366±0.007%
		$\tau\tau$	3.370±0.008%
		$\nu\nu$	20±0.055%

Table 2.3 indicates that full-hadronic channel has a dominant ratio overall final states. However, since there is an extensive QCD background in full-hadronic final states, estimating the cross sections in this background is harsh. Although the full-leptonic final states have the cleanest background to analyze, the limited branching ratio is fatal for the statistics. The sec-

Table 2.3: The branching ratio of final state for each VBS process. The full-hadronic represents WW/WZ/ZZ $\to 4q$. The full-leptonic stands for WW $\to \ell\nu\ell\nu$, WZ $\to \ell\nu\ell\ell$ and ZZ $\to \ell\ell\ell\ell/\ell\ell\nu\nu$, where ℓ is leptons e and μ, and ν represents three kinds of neutrino (ν_e, ν_μ and ν_τ). The including τs means there is one or more τs in final states. The semileptonic represents WW $\to \ell\nu qq$, WZ $\to \ell\nu qq/\ell\ell qq$, and ZZ $\to \ell\ell qq/\nu\nu qq$. This is because the analysis strategies of τs are different from the electrons and muons. The full-invisible consists of ZZ $\to 4\nu$.

VBS Process	Final States	Branching Ratio
WW	full-hadronic	45.7%
	full-leptonic	4.7%
	including τs	20.4%
	semileptonic	29.2%
WZ	full-hadronic	47.3%
	full-leptonic	5.8%
	including τs	13.8%
	semileptonic	33.2%
ZZ	full-hadronic	48.9%
	full-leptonic	3.1%
	including τs	6.6%
	semileptonic	37.4%
	full-invisible	4.0%

ond largest branching ratio goes to the semileptonic final state. Hence, the semileptonic VBS final state is a suitable process to investigate and worthwhile to improve.
In the semileptonic VBS final state, there are four quarks in the event, detected as four jets in the ATLAS. The background mainly comes from V+jets events, and approximately half of the selected jets are gluons. The VBS processes are studied with the luminosity of 36fb^{-1} in Reference [12] in the ATLAS. The VBS processes is measured to have observed (expected) significance of 2.7 (2.5) standard deviations.

2.3 Jet Struture

Semileptonic final states has 0-lepton, 1-lepton and 2-lepton types of final states, which are $qqVZ \to qqqq\nu\nu$, $qqVW \to qqqq\ell\nu$ and $qqVZ \to qqqq\ell\ell$. There are four quarks in each type of final state. The quarks and gluons are observed as jets in the ATLAS detector, reconstructed as two kinds of jets, respectively. One is Small-R Jet, another is Large-R Jet, where Small-R Jet primarily for quarks/gluons and Large-R Jet mostly for larger mass particles such as W/Z/top. The neural network tagger for q/g tagging used in this study takes Small-R Jet as the target. This section will give some brief introductions to the jet structure. The reconstruction of the jet will be described in Section 2.3.1. Since this study focuses on q/g tagging, the quark and gluon jets are demonstrated in Section 2.3.2. The jet discriminant is introduced in Section 2.3.3.

2.3.1 Reconstruction for Jet

There are several stages to reconstruct a jet: first defining the four-vectors of jet, and then defining jet algorithm and parameters, and finally either small-R jets or large-R jets will be reconstructed. First of all, the four-vectors is defined as the summation of all particles in the jet, shown in the Equation 2.7.

$$p_{\mathrm{J}} = \sum p_{\mathrm{i}} = (\sum_{\mathrm{i}} \mathrm{E_i}, \sum_{\mathrm{i}} \overrightarrow{p}_{\mathrm{i}}), \quad \mathrm{m_J} = \sqrt{(\sum_{\mathrm{i}} \mathrm{E_i})^2 - (\sum_{\mathrm{i}} \overrightarrow{p}_{\mathrm{i}})^2}, \qquad (2.7)$$

where $\mathrm{E_i}$ and $\overrightarrow{p}_{\mathrm{i}}$ represent the energy and momentum of i-th particle in jet. ATLAS primarily uses inner detector and topo-cluster to reconstruct the four-vectors of each jet, which the details will be given in Section 3.2 and 3.3.

The anti-k_t algorithm [13] with topo-cluster inputs is utilized for jet reconstruction at the ATLAS. Depending on the approaches of physics research, there are two types of jets often used, one is small-R (R=0.4) EM jets, the other one is large-R (R=1.0) LCM jets [14], where EM and LCW represent calibrated at either the electromagnetic (EM) scale or local cell weighting (LCW) scale. The jet consists of a group of particles forming a narrow cone of hadrons produced by the hadronization of quarks or gluons. Using the

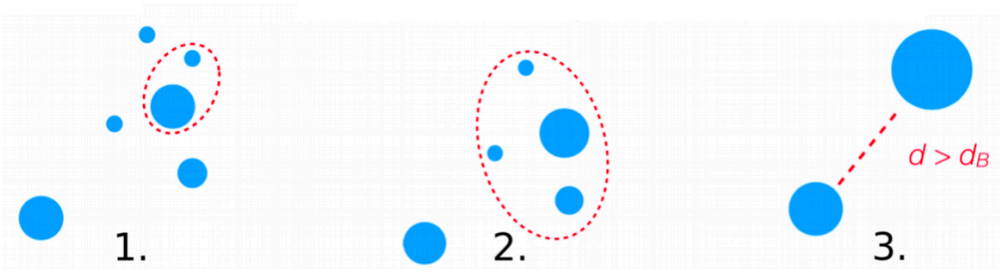

Figure 2.3: The simple illustration of the anti-k_t algorithm. Using the Equation 2.8 to accumulate the close particles to form the jet until the distance between the selected jet and other particles is far enough.

anti-k_t algorithm is the first step to differentiate a small-R jet from a large-R jet. Defining a distance named d_{ij} and calculating the distance of every two particles to find the minimum of d, seen in Equation 2.8.

$$d = \min(p_{T,i}^{-1}, p_{T,j}^{-1}) \frac{\Delta_{i,j}^2}{R^2}, \tag{2.8}$$

where $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and $p_{T,i}$ stands for the transverse momentum of i-th particle, and y_i and ϕ_i are the rapidity and azimuth of particle i. Radius parameter R is a parameter of choosing small-R jet (R=0.4). Figure 2.3 gives a brief image for the anti-k_t algorithm. The jets reconstructed by the anti-k_t algorithm are needed to be calibrated due to some effects, such as pile-up, data/MC diffs and non-compensation calorimeter response [15]. The steps of anti-k_t are shown as follows. In the first step, all reconstructed particles list is defined. Then, calculate the minimum of the d_{ij} combination over the list. Third, if $d_{ij} < d_jB$, then the i and j particles will be removed and the new particle $p_{new} = p_i + p_j$ will be added. In this step, if $d_{ij} > d_jB$, then the i particle will be considered as a jet. Repeat the step 1 to 3 until no particles are left. Finally, the jets can be defined from the reconstructed particles.

2.3.2 Quark Jet and Gluon Jet

Both quarks and gluons will be hadronized to a group of hadronic particles because of the QCD confinement, which only allows the colorless states. This group of particles is observed as a narrow cone and defined as a jet in the ATLAS. The key difference between a quark and gluon jets originates from their carrying color. Quark contains only one QCD color while gluon carries both color and anti-color. According to Altarelli-Parisi splitting functions [16], gluon radiation from a gluon has a factor of $C_A = N_C = 3$, and gluon radiation from a quark has a factor of $C_F = \frac{N_C^2-1}{2N_C} = \frac{4}{3}$, where the ratio equates $C_A/C_F = 9/4 = 2.25$. Hence, gluon jets tend to have more constituents than quark jets, causing gluon jets to have a broader radiation pattern than quark jets. ATLAS has developed a tagger based on n_{track}, trackwidth, trackC1 and also other variables to separate quark jets and gluon jets with these properties. Since there are more particles in gluon jets on average, n_{track}, trackwidth and trackC1 are also larger than quark jets, shown in Figure 2.4 and 2.5. The quark and gluon jets are defined by the labels generated in the Monte-Carlo Sample (MC sample). The d, u, s and g labels are used in Figure 2.4 and 2.5.

2.3.3 Jet Discriminant

There are several types of jet discriminant variables. Some of them are useful to have a jets identification. Some are usually utilized to remove pile-up jets. A representative variables are the number of tracks (n_{track}), trackwidth and trackC1 widely used for q/g tagging at the LHC analysis. This thesis also uses these three variables to identify quark and gluon. The ATLAS experiment measures the tracks with an inner detector (ID), including silicon micro-strips, and a transition radiation tracking detector in a 2 T axial field which is generated by a solenoid magnet [5]. More details will be introduced in Section 3.2. For the limitation of the inner detector, tracks are reconstructed at pseudorapidity of $|\eta| < 2.5$ and the transverse momentum $p_T > 500$ MeV. Both the distance from the primary vertex and the longitudinal length are required to be smaller than 1.5 mm. And at least 6 SCT hits are also needed. After the criterion above, the anti-k_T algorithm will select the tracks associated with the jet. n_{track} means the number of tracks associated with the jet observed by the inner detector. The distribution of n_{track} varies with p_T and different particles, and n_{track} also increases at higher p_T, shown in Figure 2.4.
Trackwidth is also found to be a useful variable for identifying the origin of

a jet, and the definition is seen as Equation 2.9.

$$\text{trackwidth} = \frac{\sum_{\text{i}}(\text{p}_{\text{T,i}}^{\text{trk}} \times \Delta\text{R}(\text{trk}_{\text{i}}, \text{Jet}))}{\sum_{\text{i}} \text{p}_{\text{T,i}}^{\text{trk}}}, \tag{2.9}$$

where $\Delta\text{R}(\text{x}, \text{y}) = \sqrt{(\eta_{\text{x}} - \eta_{\text{y}})^2 + (\phi_{\text{x}} - \phi_{\text{y}})^2}$, and $\text{p}_{\text{T,i}}^{\text{trk}}$ is the transverse momentum of i-th tracks in the selected jet.
TrackC1, is called "track-based-energy-correlation angularity", and shows the energy correlation among the tracks that are associated with the jet, shown in Equation 2.10.

$$\text{trackC1} = \frac{\sum_{\text{i}} \sum_{\text{j}} (\text{p}_{\text{T,i}}^{\text{trk}} \text{p}_{\text{T,j}}^{\text{trk}} \times \Delta\text{R}^{\beta}(\text{trk}_{\text{i}}, \text{trk}_{\text{j}}))}{(\sum_{\text{i}} \text{p}_{\text{T,i}}^{\text{trk}})^2}. \tag{2.10}$$

The indexes run over all the tracks in the jet with $\text{j} > \text{i}$, and the parameter β is tunable. $\beta = 0.2$ is utilized in this thesis.

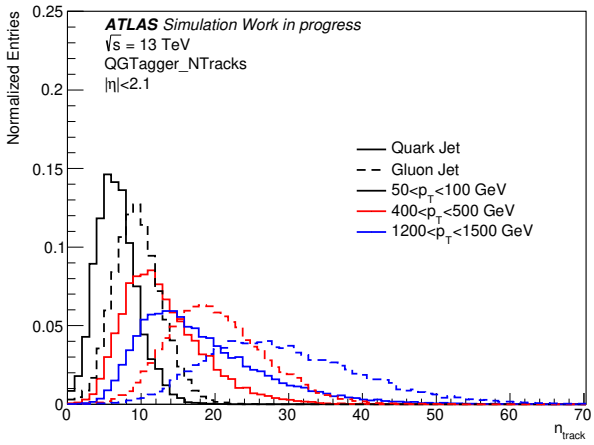

Figure 2.4: The distribution of n_{track}. The solid line stands for quark jets and the dotted line stands for gluon jets. Different colors indicate different p_{T} regions. As transverse momentum p_{T} goes to higher, n_{track} also becomes larger. Jets are required within the tracking acceptance which $|\eta| < 2.1$ and $\text{p}_{\text{T}} > 500$ MeV.

Figure 2.5 shows the distribution of trackwidth and trackC1. Trackwidth and trackC1 are also important discriminating variables. These three variables are also utilized in Boosted Decision Trees (BDT), a machine learning tool usually implemented in the LHC analysis for q/g tagging. This study makes use of n_{track}, trackC1, trackwidth for q/g tagging in Chapter 4.

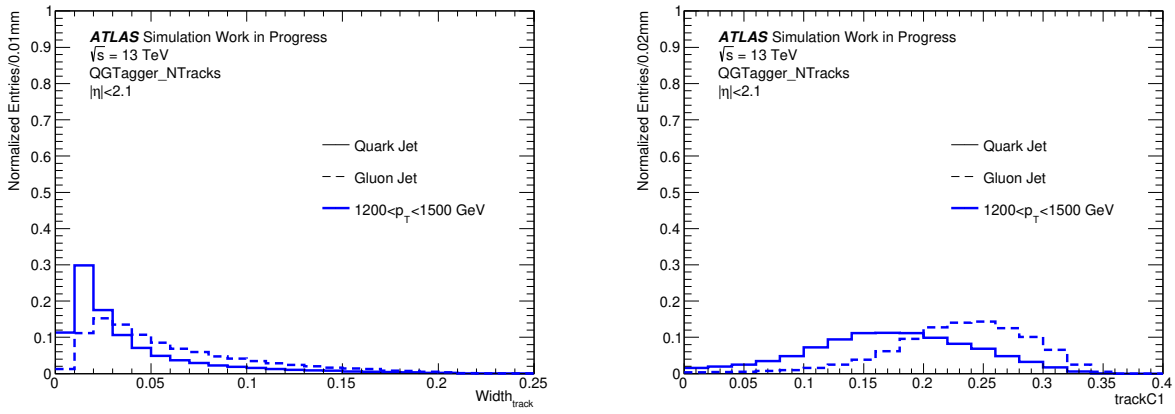

Figure 2.5: The distribution of trackwidth (left) and trackC1 (right) for gluon and quark at $1200 < p_T < 1500$ GeV. The distribution for gluon is slightly larger than the distribution for quark in general because of the difference of color factor, and more details are described in Section 2.3.3. Jets are required within the tracking acceptance which $|\eta| < 2.1$ and $p_T > 500$ MeV.

Apart from jet identification for different particles, pile-up jets from additional proton-proton interactions at the LHC gives a crucial issue for numerous physics analysis. To identify the jets colliding correctly against pile-up jets, jet-vertex-tagger (JVT) is a type of discriminaxnt to clean the pile-up jet constructed from R_{p_T} and corrJVF (correlated jet-vertex-fraction [17]). The JVT discriminant is based on a k-nearest neighbor (kNN) algorithm [18] by the MC simulations. The definition of R_{p_T} and corrJVF used in the JVT is described as follows.

$$R_{p_T} = \frac{\sum_k p_{T,k}^{track}(PV_0)}{p_T^{jet}}, \tag{2.11}$$

$$\text{corrJVF} = \frac{\sum_m p_{T,m}^{track}(PV_0)}{\sum_l p_{T,l}^{track}(PV_0) + \frac{\sum_{n\geq 1}\sum_l p_{T,l}^{track}(PV_n)}{k \cdot n_{track}^{PU}}}, \tag{2.12}$$

where PV_0 here means the hard-scatter vertex and PV_n stands for primary vertices originating from pile-up interactions and PV_n is pileup vertices. k is a constant parameter and n_{track}^{PU} represents the pileup tracks number in the event.

It is shown that using track and vertex information with JVT variables can remove pile-up jets effectively. According to reference [19], jets with $p_T < 50$ GeV and $|\eta| < 2.4$ required at the medium JVT working point accepts 92%

of hard-scatter jets and rejects 98% of pile-up jets.
In summary, this thesis uses the n_{track}, trackwidth and trackC1 as the inputs for the q/g tagger in both BDT and neural network models. On the other hand, in order to reduce the pile-up jets, all jets generated by the Monte Carlo method passed the medium JVT working point in this study.

Chapter 3

The ATLAS Experiment at the LHC

The Large Hadron Collider, the highest energy and most powerful particle accelerator, was built 100 m underground of CERN in Geneva. LHC collides proton to proton at center-of-mass equating to 13 TeV.
There are several purpose detectors in the LHC, which are ATLAS (A Toroidal LHC Apparatus), Compact Muon Solenoid (CMS), ALICE (A Large Ion Collider Experiment) and The Large Hadron Collider beauty (LHCb) experiments. ALICE specializes in heavy-ion physics and is designed to investigate strong interaction matter physics. LHCb is dedicated to studying the difference between matter and antimatter by investigating the b quark. ATLAS and CMS have similar scientific goals to study a wide range of physics but have different magnet-system designs and calorimeter material. The ATLAS detector [5] is the largest detector and is now used to find new physics beyond the SM.
This thesis focuses on the ATLAS and a brief introduction of the ATLAS detector is introduced in Section 3.1. The inner detector (ID) and Calorimeters used to detect the particles are described in Section 3.2 and 3.3. The track and topological cluster used for physics object reconstruction will be introduced in Section 3.4.

3.1 Overview of the ATLAS Detector

The ATLAS detector consists of several sub-detectors (inner tracker, calorimeter, muon spectrometer) built for various physics analysis at the TeV scale. The overall layout is shown in Figure 3.1. The inner detector's 2 T solenoidal field originates from four large superconducting magnets. The diameter and

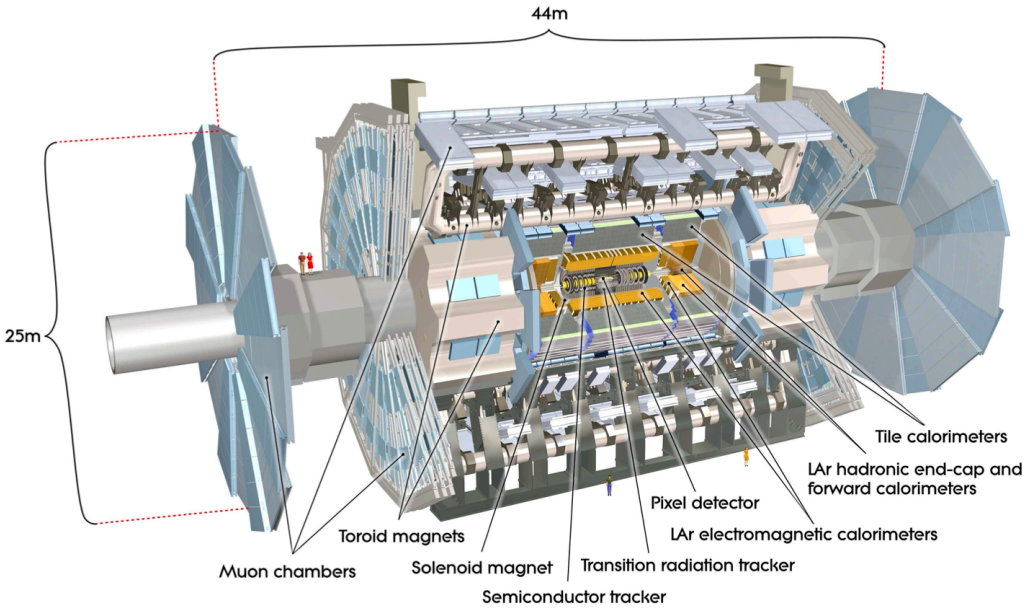

Figure 3.1: The schematic illustration for the ATLAS detector is taken from [5]. The detector is constructed in toroidal shape with 25 m in height and length in 44 m. The detector consists of an inner detector, solenoid magnet, calorimeters and muon spectrometer.

length of each magnet are approximately 2.56 m and 5.6 m, respectively. The magnet system provides a 2 T magnetic field in the detector.
The coordinate of the ATLAS detector is described as follows. The x-y plane is perpendicular to the beam direction while the z-axis goes to the collision direction.
There are several variables usually used in the LHC.

- Transverse particle momentum : $\mathrm{p_T} = \sqrt{\mathrm{p_x^2} + \mathrm{p_y^2}}$.
- Azimuthal angle : $\phi = \arctan \mathrm{p_y}/\mathrm{p_x}$, where is usually used for x-y plane.
- Polar angle : $\theta = \arctan \mathrm{p_z}/\mathrm{p_x}$.
- Pseudorapidity : $\eta = -\ln \tan \theta/2$, which describes the angle between the particle momentum and the beam axis.

3.2 Inner Dector

Inner Detector consists of three subsystems, designed to have significant momentum resolution and measure primary and secondary vertex for charged tracks above around 0.5 GeV with $|\eta| < 2.5$. A schematic overview of the inner detector is shown in Figure 3.2. For against the high radiation environment, enormous strategies are designed such as the material of the sensor and the temperature in the system. The inner detector provides the momentum of charged particles by calculating the curvature of the trajectory. It also identifies particle charge since opposite charged particles have opposite curves in the magnetic field. High resolution silicon pixel detector will be demonstrated in Section 3.2.1. Semiconductor Tracker (SCT) used a classic single-sided p-in-n technology as described in Section 3.2.2. Transition Radiation Tracker (TRT) consisting of polyimide straw tubes is introduced in Section 3.2.3.

3.2.1 Pixel Detector

Pixel detector is built at the most inside part of the inner detector. Therefore, pixels are designed with two dimension sensors with high resolution. There are nearly 400 thermistors installed within the pixel detector, utilized to make sure the cooling loops system works appropriately. The sensors are made of radiation-tolerant humidity sensors to reduce the effect of high radiation environment and moisture. Four barrel layers and three double-sided end-caps layers cover $|\eta| < 2.5$ by 1736 and 288 pixel sensors, respectively. The most inside one barrel layer named insertable b-layer (IBL) was installed in 2014. Preventing the leakage current, the entire system is operated at the temperature $\sim -10°$C. The sizes of each pixel are 50×250 μm^2 and 50×400 μm^2. The readout fibers have been improved recently to handle the data from high luminosity in the future upgrade. The pixel sensor uses oxygenated n-type wafers with readout on the n^+-implanted side of the detector. The n^+ implanted side allows the detector has the powerful charge-collection ability. The oxygenated materials provide good radiation tolerance while detecting the tracks of hadrons. These properties make the pixel detector have significant resolution and high granularity and good tolerance for the beginning area of detecting the charged particles.

3.2.2 Semiconductor Tracker

Semiconductor Tracker (SCT) detector is located at 30-50 cm from the beam. SCT is made of one dimension sensors named silicon microstrips. Two types

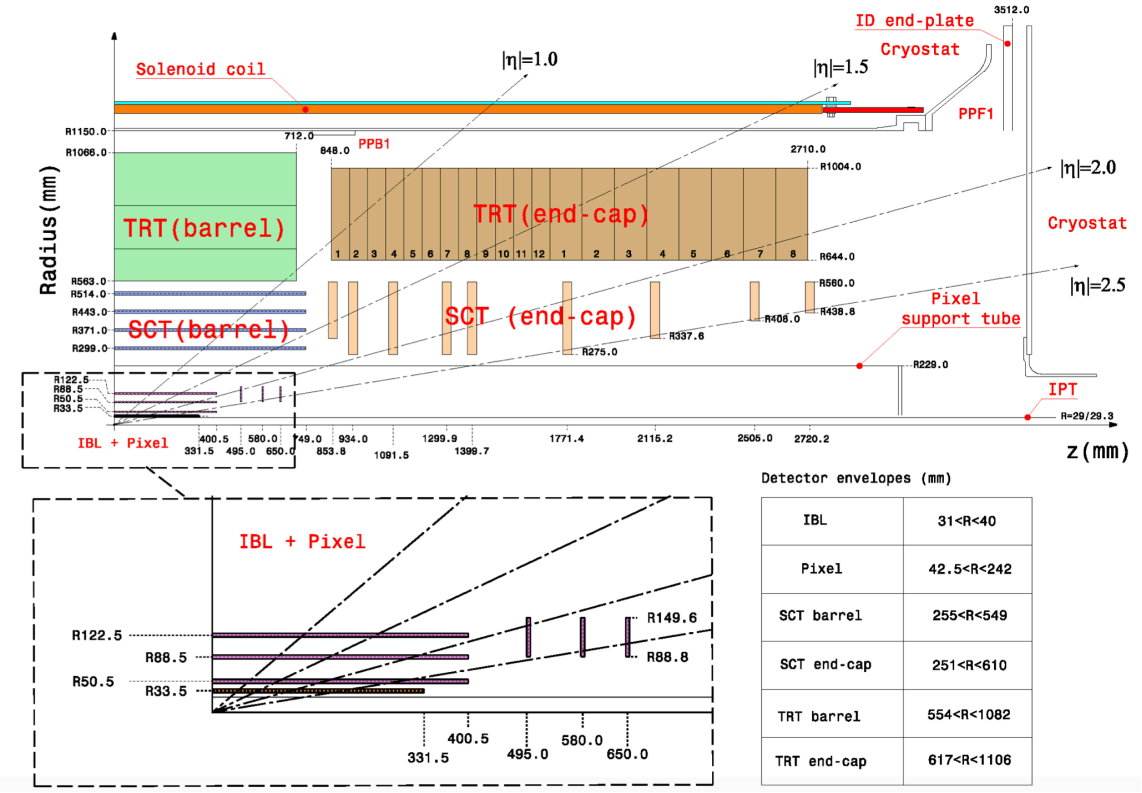

Figure 3.2: The illustration of inner detector taken from [20]. The radius of the detector is approximately 1 m. Pixel layer is placed at R = 33.5 - 122.5 mm, SCT is built at R = 299 - 514 mm, and TRT is located at R = 554 - 1082 mm. All of them consist of barrel layers and end-cap layers. The barrel layers cover at lower pseudorapidity while the end-cap extends to higher pseudorapidity.

of sensors are made for the SCT detector. The barrel sensor has 80 μm strips with two 6 cm long sensors. The end-cap sensors are designed with radial strips with a mean pitch of around 80 μm. Thermal pyrolytic graphite technology [21] is applied to give a high thermal conductivity between the sensors and cooling system. There are four sensors in each barrel layer, two on the top and two on the bottom sides. There are only two sensors in each end-cap layer, each sensor on the top and bottom sides. The top and bottom sides of the SCT sensors are rotated with a $\pm$20 mrad angle, allowing the strip-type sensors to measure the two-dimensional space observation. The heat is cooled down by evaporating C_3F_8 at around -25 °C, where cooling pipes are attached to each module.

3.2.3 Transition Radiation Tracker

The Transition Radiation Tracker is placed at the most outer part of the inner detector at 56.3 – 106.6 cm from the beamline. TRT detector uses tubes filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2 with roughly 10 mbar pressure. Different from the pixel and SCT detector, the TRT can measure the 1-dimensional space points for each tube. In order to avoid pollution from permeation, the tubes are enveloped with CO_2 gas. TRT tube has a concentric circle shape, where the gold wire is surrounded by the tubes. The gold wire inside and tube layer outside are anode and cathode, respectively. The voltage is operated at -1530 V toward the inside gold wire, giving the maximum electron collection time can be down to 48 ns. Due to the emission electron drift-time, TRT space resolution is around 130 μm.
When the charged particle passes the TRT tubes, it ionizes the gas companying with the radiated photons. The radiated photons interact with the gas molecules, producing more electrons moving toward the gold wire and generating the electronic signal.
All charged particles with $p_T > 0.5$ GeV and $|\eta| < 2.0$ will go through at least 36 straws, which can be used to reconstruct the trajectory of their tracks and transverse momentum.
The TRT detector can also distinguish different sorts of charged particles at an early stage. When a pion (electron) transverse the tubes, it radiates fewer (more) photons that bring fewer (more) electrons. These properties allow the device to separate charged particles by the magnitude of the electronic signal.
The TRT contains 73 layers of straws for a barrel module and 160 straws planes for an end-cap module. The barrel TRT is separated into three rings, each consisting of 32 modules. The TRT end-caps are made of two sets of straw wheels. One contains 12 wheels while the other one contains eight

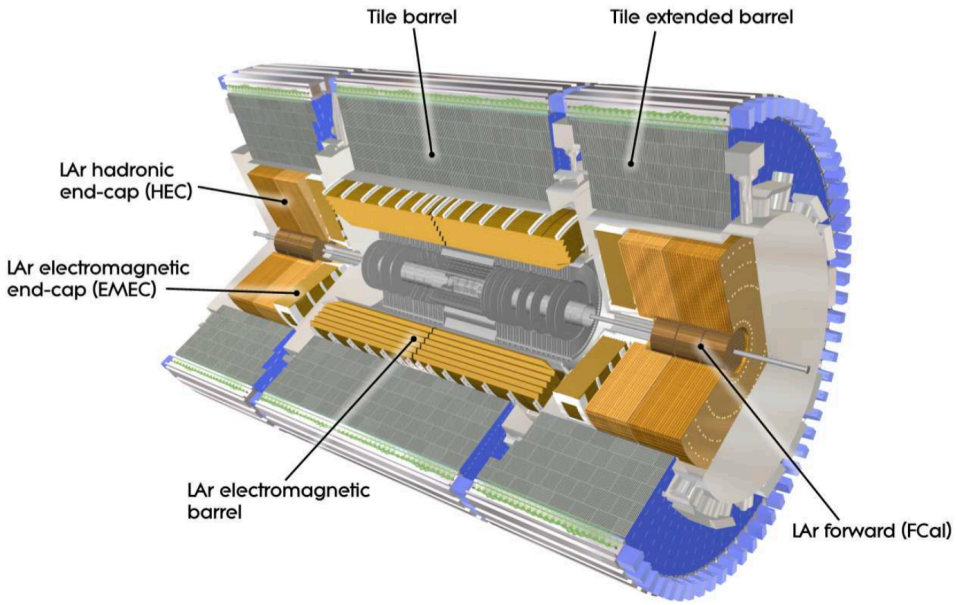

Figure 3.3: The schematic figure for the ATLAS calorimeter is taken from [5]. The calorimeter consists of LAr forward (FCal), LAr electromagnetic barrels, LAr electromagnetic end-cap (EMEC), LAr hadronic end-cap (HEC), Tile barrel and Tile extended barrel. Each system contains several layers which can determine the energy for particles.

wheels.

3.3 Calorimeters

Calorimeter systems stop the particles with electromagnetic and hadronic interaction to measure the energy. Therefore, it is located outside the inner detector. ATLAS calorimeter contains several regions, shown in Figure 3.3, and contains mainly electromagnetic and hadronic systems. Electromagnetic calorimeter samples the energy of photons and electrons with electromagnetic interaction. On the other hand, hadronic calorimeters measure the energy of hadronic particles such as protons and mesons by hadronic interaction. Combined with the Inner Detector and Muon Chamber [5], ATLAS detector can identify considerable different particles, as seen in Figure 3.4. Section 3.3.1 describes an Electromagnetic Calorimeter that mainly interacts with electrons and photons. Section 3.3.2 shows Hadronic Calorimeter which allows ATLAS detector to discriminate the hadrons from other particles.

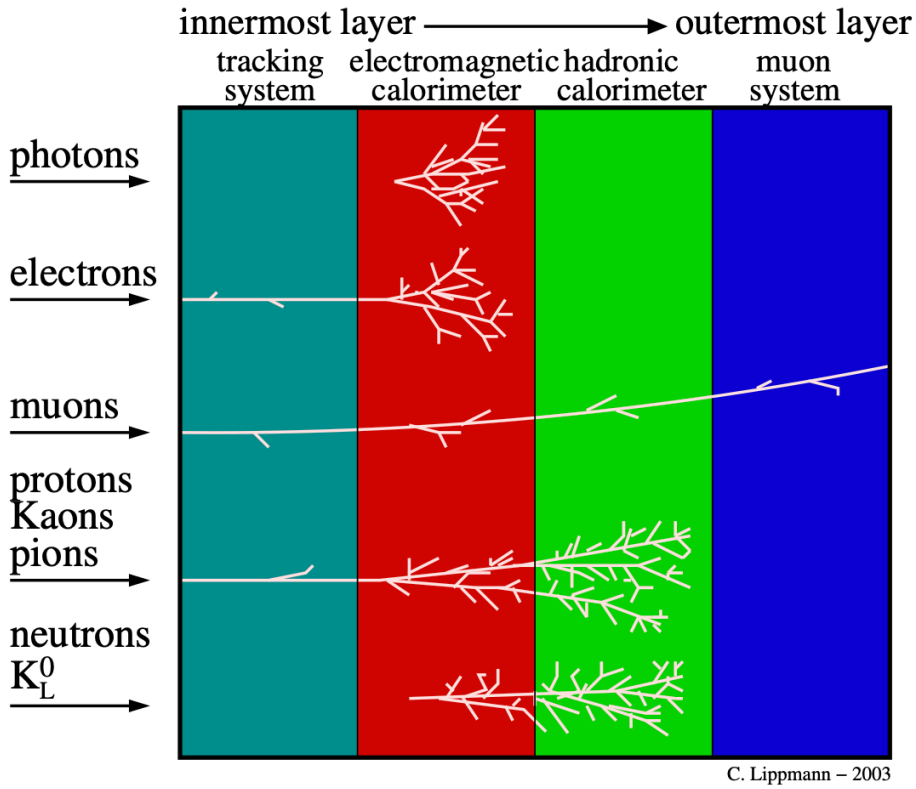

Figure 3.4: The particle identification schematic illustration is taken from [22]. Photons only interact with the matters at electromagnetic calorimeter. Electrons leave the tracks at the inner detector and stop at the electromagnetic calorimeter. Muons go through the tracking system and both calorimeters to the muon system. Charged hadrons are stopped at the hadronic calorimeter. Neural hadrons don't have tracks but also stop at the hadronic calorimeter. These properties allow the ATLAS detector to distinguish different sorts of particles.

3.3.1 Electromagnetic Calorimeter

Lead and Liquid Argon (LAr) are the main materials for the electromagnetic calorimeter. It is used to detect the particles having electromagnetic interaction such as electrons and photons. The electromagnetic calorimeter consists of barrel and end-caps ones which cover $|\eta| < 1.4$ and $1.4 < |\eta| < 3.2$, respectively.
Two half-barrels constitute the barrel electromagnetic calorimeter and surround the beamline to cover the entire azimuthal angle ϕ. The thickness of a module varies from 22 to 30 radiation lengths (X_0) between $|\eta| = 0$ and $|\eta| = 0.8$, and increases from 24 X_0 to 33 X_0 between $|\eta| = 0.8$ and $|\eta| = 1.3$ [5]. The end-cap calorimeters are made of two wheels used to measure the energy of particles in the area $1.375 < |\eta| < 3.2$. The end-cap calorimeter has thickness of 24 to 38 X_0 between $1.457 < |\eta| < 2.5$, and 26 to 36 X_0 between $2.5 < |\eta| < 3.2$. Therefore, the electromagnetic calorimeters cover a larger area than the inner detector at $|\eta| > 2.5$ region. The LAr calorimeter surrounding the inner detector stops the particles and converts them into an electromagnetic shower which contains a set of lower energy particles. The shower particles ionize lead and liquid argon layers and produce an electronic signal. It is designed as an accordion geometry in order to cover all ϕ regions without time dependence. LAr calorimeter is set at -184°C to keep the argon in liquid form. The electronmagnetic calorimeter energy resolution is $\frac{\sigma(\mathrm{E[GeV]})}{\mathrm{E[GeV]}} \approx \frac{10\sim17\%}{\sqrt{\mathrm{E[GeV]}}} \oplus 0.7\%$.

3.3.2 Hadronic Calorimeter

The hadronic calorimeter is built outside the electromagnetic calorimeter. It is subdivided into three types of calorimeters: one is Tile calorimeter, another one is LAr hadronic end-cap calorimeter (HEC) and the other is LAr Forward Calorimeter (FCal).
Since hadrons are predicted to go further at the calorimeter layers, hadronic calorimeters are located outside of electromagnetic ones. The hadrons pass through the electromagnetic calorimeters lose some energy and completely stop at the hadronic calorimeter.
When particles hit the steel layers, they are also converted to lower energy shower particles. The plastic scintillators generate photons and the photomultiplier tubes (PMT) amplify the electronic signal which will be sent to the readout system.

- **Tile Calorimeter**

The tile calorimeter is placed at $|\eta| < 1.0$ region, with two $0.8 < |\eta| < 1.7$ extended barrels. It uses scintillating tiles as an active material and steel as an absorber. The three-layer barrel has approximately 1.5, 4.1 and 1.8 interaction lengths. On the other hand, the extended barrel has 1.5, 2.6 and 3.3 interaction lengths, respectively.

- **LAr Hadronic End-Cap calorimeter**
  There are two wheels in the HEC, placed behind the end-cap electromagnetic calorimeter out to $|\eta| = 3.2$. The wheels of the HEC are made of copper plates to absorb the hadronic particles. The HEC has approximately 11 interaction lengths for overall thickness.

- **LAr Forward Calorimeter**
  The FCal is located at around 4.7 m from the interaction point. The tungsten is an absorber material while LAr is an active medium for hadronic interaction. And copper and LAr are used as absorbers and active materials for electromagnetic interaction. The FCal has roughly 10 interaction lengths.

The haronic jets energy resolution in the HEC and tile calorimeter is $\frac{\sigma(\mathrm{E[GeV]})}{\mathrm{E[GeV]}} \approx \frac{50\%}{\sqrt{\mathrm{E[GeV]}}} \oplus 3\%$. The FCal energy resolution is estimated as $\frac{\sigma(\mathrm{E[GeV]})}{\mathrm{E[GeV]}} \approx \frac{100\%}{\sqrt{\mathrm{E[GeV]}}} \oplus 10\%$.

3.4 Reconstruction

Track and topo-cluster play a vital role for the physics objects (such as electrons, muons and jets) reconstructions in the early step at the ATLAS. This section shows how the ATLAS reconstructs the track and cluster. Section 3.4.1 gives a brief introduction of tracks and vertices in the ATLAS. Section 3.4.2 describes the topological-clusters system.

3.4.1 Tracks and Vertices

The track and vertex reconstruction algorithms in the ALTAS have proven significant in considerable studies. The Inner Detectors (ID) introduced in Section 3.2 gives the data of hits from particles in a three-dimensional position. These multi-hits in the ID can be reconstructed into tracks for charged particles. First, the seed track starts from the inside layers of the pixel detector to the SCT detector and extends to the TRT detector. Then Kalman Filter [23] provides an excellent estimate for early measurements. The algorithm calculates the energy loss in the ID for an ionization particle and

bremsstrahlung by the track parameter. Secondly, a track search is taken from segments in TRT and reconstructs the trajectory with the pixel and SCT hits. On the other hand, the tracks with TRT segments but without silicon detectors are counted as TRT-standalone tracks. According to [12], track reconstruction efficiencies are shown as follow, 82% $\sim$ 94% at $0.006 < \Delta R(\text{jet, track}) < 0.01$ for light flavor jets.
For the reasons such as pile-up and proton-proton collisions, the ALTAS experiment is in the harch environment to distinguish the signal from a large amount of background. A vertex reconstructed from tracks with $p_T > 400$ MeV is utilized for finding the signal. Vertex Finding algorithm [24] is used to identify the vertex position for tracks. After all, vertices are reconstructed, the highest $\sum p_T^2$ vertex is defined as a primary vertex. The details of finding primary vertex are written in [25]. The hard-scatter jets satisfy that an MC sample truth-particle jet with $p_T > 10$ GeV and is found within $\Delta R < 0.3$.
Tracks are necessary for jets reconstruction. Jet consists of several hadrons which leave numerous tracks. The tracks in the associated jet are utilized to reconstruct its four-vector. The variables for discriminating described in Section 2.3.2 (n_{track}, trackwidth and trackC1) are also calculated from the four-vector of tracks.
The tracks of jets in the ATLAS are reconstructed from the hits in the inner detectors. The azimuthal range and pseudorapidity of jet tracks are selected in the range of $0 < \phi < 2\pi$ and $|\eta| < 2.5$. Tight quality criteria [26] is applied to reduce the effect from pile-up background and reject fake tracks from the noise or charged particles. The following describes several additional cuts for the jets in this thesis. Each track in the associated jet satisfies $p_T > 0.5$ GeV and comes from the hard-scatter primary vertex. Track-to-vertex matching is implemented to define the primary vertices by the vertex reconstruction. Track excluded in vertex reconstruction are implanted the nearest vertex strategy with the distance $|\Delta z \times \sin\theta| < 3$ mm, where Δz represents the longitude distance between the object and defined vertex. This study requires $|\eta| < 2.1$ for jets in order to obtain full tracker acceptance in the ID. Jet $p_T > 25$ GeV is also required for reducing the soft emission. The working point of JVT is set at Medium working point [19] and energy calibration [27] is applied.

3.4.2 Topological-Clusters

A topological-cluster (topo-cluster) is to assemble a group of particles into clusters from the electromagnetic and hadron calorimeter measurement. The topo-cluster is usually exploited for reconstructing jets from calorimeter cells. The reconstruction of the topo-cluster is shown as follows. The seed finding

in the first step is searching the calorimeter cells with $E^{EM}_{cell}/\sigma^{EM}_{noise,cell} > 4$. Here E^{EM}_{cell} represents energy corrected by EM-scale (The scales for e γ in the ATLAS electromagnetic calorimeters). The reconstructed seeds are collected as a proto-cluster from the primary vertex. Secondly, the adjacent cells in the given region overlapping in the (η, ϕ) with $E^{EM}_{cell}/\sigma^{EM}_{noise,cell} > 2$ are added to the proto-cluster. Finally, the $E^{EM}_{cell}/\sigma^{EM}_{noise,cell} > 0$ cells nearby each proto-cluster will be merged to one cluster.

Chapter 4

Quark/Gluon Tagging

Quark/Gluon tagging is necessary for many physics analysis in high-energy physics. Especially, the LHC experiment with proton and proton collision causes the immense QCD effect. The QCD interaction generates many background gluons, which are seen as jets in the ATLAS detector. Therefore, the separation of quark and gluon helps physicists discriminate the signal from the background. This chapter introduces several kinds of q/g tagging models and compares their performances. Section 4.1 shows the MC sample used when the q/g tagging models are trained. Section 4.2 gives the details of each q/g tagging model, including conventional BDT model and neural network models. Section 4.3 describes the performance of the models and compares their p_T and η dependence, respectively, and introduces the training and testing results by the different MC generators.

4.1 MC sample

The dijet process@leading order (multijet) MC samples have been used to train the q/g tagging models. The jets are evolved with Pythia 8. The pileup and noise are also considered in order to approach the real data. The generator Pythia 8 [28] utilizes the NNPDF2.3 PDF set [29], the A14 tune [30], and run all simulation in the ATLAS detector. For generating heavy flavor decays, Pythia 8 is interfaced with EvtGent v1.2.0.
The quark and gluon jets are defined by the labels generated in the MC sample. There are three kinds of labels (d, u, s) and one label (g) for quark and gluon jets, respectively.
To have better learning for the q/g tagging model, other steps are taken. The p_T of quark jet and gluon jets is flatted to the same distribution and

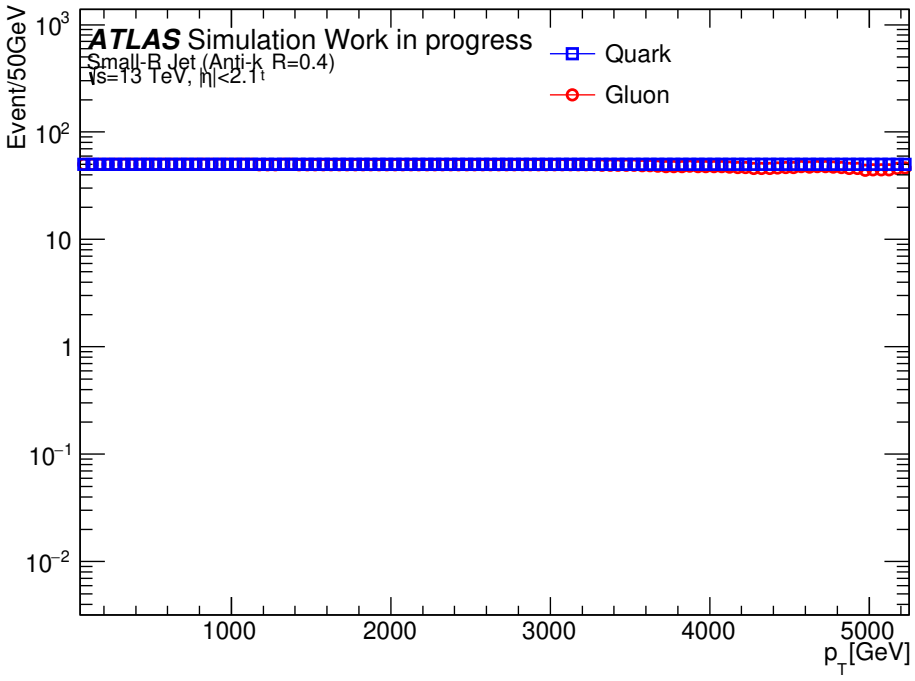

Figure 4.1: The p_T of quark and gluon jet in multijet MC sample is flatted to one in order not to affect by p_T bias.

normalized to one to make sure models are not affected by p_T bias, shown in Figure 4.1. The quark and gluon inputs are collected in the same numbers since the training performs better with the same amount of quark and gluon inputs in neural network models.
Training is performed using a single NVIDIA Tesla T4 GPU with 500000 quark jet and 500000 gluon jet while 250000 quark jet and 250000 gluon jet for testing. The testing here means that the trained q/g tagging models are applied to other samples to check that the overtraining does not occur. Different models require different times to train. The BDT model is the fastest one with around half an hour while Deep Sets and Energy Flow models need the most time with approximately 6 hours.

4.2 Quark/Gluon Tagging Models

This study uses several q/g tagging models to predict whether it is quark or gluon. This section has documented the details of q/g tagging models.

The conventional q/g tagging model usually used is the BDT model while neural network models are usually used in recent decades. This is because compared to the BDT model which can only use high-level input such as n_{track}, trackwidth and trackC1, the neural network models are able to utilize more low-level inputs such as the p_T of each track in the associated jet and the energy from each calorimeter layer. The performance of q/g separation usually increases in the neural network since low-level input represents more information used for inference.
The BDT model and its conventional inputs (n_{track}, trackwidth, trackC1) are described in Section 4.2.1. BDT model is used as a baseline to make the comparison with other neural network models. The neural models are introduced from Section 4.2.2 to 4.2.5. MultiLayer Perceptron (MLP) model, the simplest neural network model (Since the MLP only uses the dense layers, the most straightforward unit in the neural network models, it can only input the high-level information.), is shown in Section 4.2.2. Convolutional Neural Network (CNN) model is demonstrated in Section 4.2.3. CNN model is a neural network model using image input. Hence, CNN can use the position and p_T of each track information that represents low-level input. MLP and CNN models are combined into the MLP-CNN model demonstrated in 4.2.4. Pointwise models are introduced in Section 4.2.5 which there are two types of models, the Deep Sets model and the Energy Flow model. These two models have a similar structure while energy flow multiplies track p_T with other terms such as ϕ, η in the model.

4.2.1 BDT Model

The BDT machine algorithms are implemented by the Toolkit called Multivariate Data Analysis (TMVA [18]) on ROOT [31] based analysis tool. The principle of the BDT model is having a different cut on different variables to obtain the best classification. First of all, give a cut on one of the variables and calculate the remaining signal and background percentage. Secondly, choose the better one and give another cut on another variable. This second cut also provides the remaining signal and background sample percentage. The step above keeps performing until it is stopped which usually depends on the MaxDepth of the BDT. The schematic illustration is shown in Figure 4.2. The signal background ratios are calculated with the different decision trees until the best selection is determined. Boosting here means that the trees depend on the prior trees. This ensures that the final trained model is more stable and has better accuracy. Quark jets are input as the signal sample while gluon jets are input as background in the BDT model. The setup of BDT model is listed in Table 4.1.

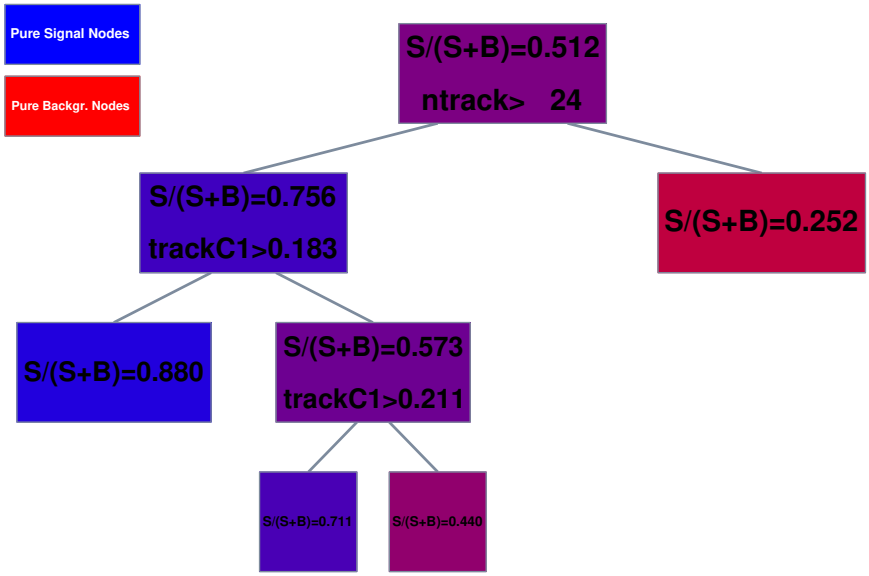

Figure 4.2: The illustrating figure for the decision tree. In this example, the first threshold cut is the number of tracks larger than 24. The sample satisfying the threshold goes left, otherwise right. The additional selection thresholds are performed depending on the MaxDepth setup in the BDT.

The BDT model learns the MC samples with n_{track}, trackwidth and trackC1 variables, shown in Figure 4.3. The brief descriptions of the BDT setup option are explained as follows.

- AnalysisType : The learning samples have binary labels, so classification analysis is used.
- NTrees : Number of trees in the BDT model.
- MinNodeSize : Minimum fraction of training events
- MaxDepth : Max depth of the decision tree.
- AdaBoostBeta : The parameter in the AdaBoost boosting technique.
- AdaBoost : An algorithm that can learn the past error sequentially from the previous training.
- BaggedSampleFraction : Relative size of bagged event sample used in the MC sample.

Table 4.1: The setup of BDT model.

AnalysisType=Classification
NTrees=850
MinNodeSize=2.5%
MaxDepth=3
AdaBoostBeta=0.5
BaggedSampleFraction=0.5
SeperationType=GiniIndex
nCuts=20

- Bagged event sample : The sample selected from the bagging method is an algorithm to reduce the variance of the sample to avoid overfitting.
- SeparationType : The separation criterion for node splitting.
- node splitting : A decision tree makes the decision and divides a node into multiple sub-nodes to determine the better decision.
- nCuts : Number of grid points in the variable region used in determining the optimal cut in the node splitting.

4.2.2 MLP Model

MLP model is the simplest neural network model that connects several dense layers. The structure is displayed in Figure 4.4. The dense layer consists of several neurons with different kernel functions and weights. The weight (The arrow mark in Figure 4.4.) is upgraded by the gradient descent method, shown as follows.

$$\Delta\omega_{\mathrm{ji}}(\mathrm{n}) = -\lambda\frac{\partial\varepsilon(\mathrm{n})}{\partial\nu_{\mathrm{j}}(\mathrm{n})}\mathrm{y_i}(\mathrm{n}), \tag{4.1}$$

where λ is the learning rate, $\mathrm{y_i}$ is the output of the previous neuron. ε is the square sum of error (truth value − prediction). ν is the variable in the kernel function. Each neuron is connected to all neurons in the previous dense layer or all inputs. The input values with weights will be calculated in the kernel function and output a value to the following dense layer.
There are 31 inputs for each small-R jet. The inputs are connected to two iterations of the dense layer with 64 neurons and finally output the quark and gluon jet classification. The 31 inputs are described as follows, $\mathrm{n_{track}}$, trackwidth, trackC1, electromagnetic energy fraction and the energy of the calorimeter layers. The calorimeter layers used in the MLP model are shown in Table 4.2. Rectified Linear Unit (ReLU) [32] activation is utilized for all

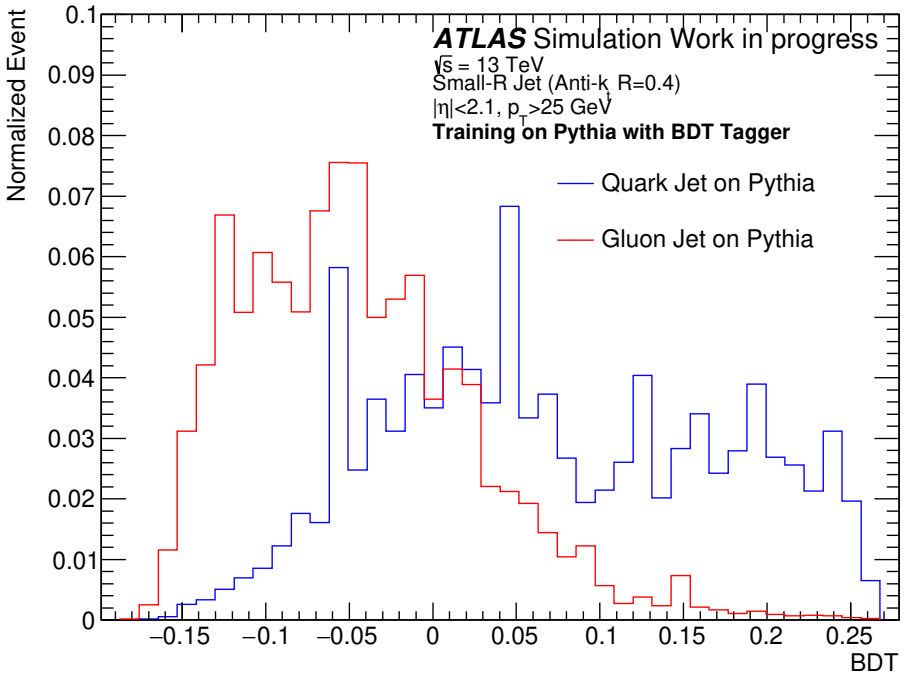

Figure 4.3: The BDT response of quark jet and gluon jet with the pythia multijet MC test sample.

dense layers. Dropout=0.3 [33] is also applied to all dense layers in order to prevent overtraining. Additionally, an L2 regularization [33] with strength 10^{-8} is used in all layers. Also, earlystopping is applied to patience=15 to prevent overfitting and epochs = 100. The optimizer is Adam and its learning rate equates to 10^{-5}. The loss function is set as SparseCategorical-Crossentropy[1] [33]. The MLP output is a softmax function [33] of size two which provides the probability of quark and gluon relatively. The option is summarized in Table 4.3.

4.2.3 CNN Model

The MLP and BDT can only utilize the high-level inputs, causing information loss during the calculations. The image inputs are considered as a new input that includes low-level information. 2D Convolutional Neural Network (CNN) is a deep learning model used to learn visual images. Therefore,

[1]A format of true label inputs when training the models.

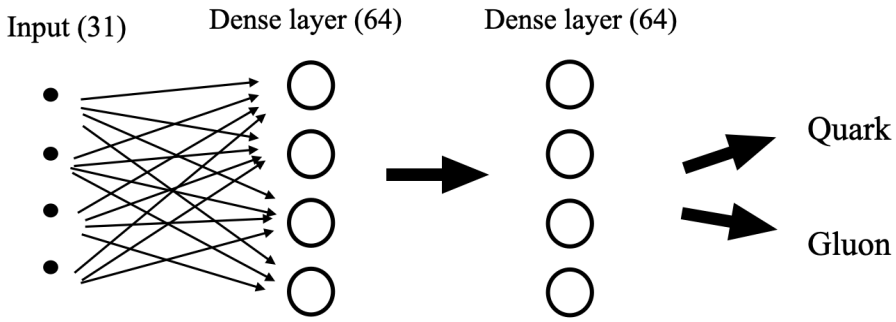

Figure 4.4: The structure of the MLP model used in this study. There are 31 inputs with one input jet, passing through two 64-neurons dense layers. Finally, the input jet is classified as a quark or gluon jet.

Table 4.2: The energy inputs of calorimeter layers used in the MLP model.

LAr barrel	LAr EM endcap	Hadronic endcap	Tile barrel
PreSamplerB	PreSamplerE	HEC0	TileBar0
EMB1	EME1	HEC1	TileBar1
EMB2	EME2	HEC2	TileBar2
EMB3	EME3	HEC3	
Tile gap	Tile extended barrel	Forward EM endcap	Mini FCAL
TileGap1	TileExt0	FCAL0	MINIFCAL0
TileGap2	TileExt1	FCAL1	MINIFCAL1
TileGap3	TileExt2	FCAL2	MINIFCAL2
			MINIFCAL3

converting the track and topo-cluster information into a visual image is necessary.

The CNN model and its inputs in this study follow the previous study [34]. All constituents in a jet are rotated to $\phi_{\text{jet}} = \eta_{\text{jet}} = 0$ based on the center of the associated jet to form a jet image. And then, the p_T of each track and cluster in the associated jet is used to fill into a fixed grid of size 16×16 at $|\eta| < 0.4$ and $|\phi| < 0.4$ region. After the transformation, the center of the jet is at the center of the image and the pixel size is $\eta \times \phi = 0.05 \times 0.05$. Next, the image is normalized to $\sum_i p_{T,i} = 1$. The normalization must drop some information, but according to [34], the impact is small enough to ignore. Another advantage of normalization is that it will work better in neural network training. The topo-cluster and track constituents of a quark jet image are displayed in Figure 4.5. The average of millions of quark and gluon

Table 4.3: The setup of neural network models in this thesis.

Activation Function=ReLU
Dropout=0.3
Regularization=L2
Optimizer=Adam
Learning Rate=10^{-5}
Output=Softmax function
Loss function=SparseCategoricalCrossentropy

topo-cluster images and their image difference are displayed in Figure 4.6. Reminding that the trackwidth distribution of gluon jet is more significant than quark jet. This property can also be seen in Figure 4.6. The middle image in Figure 4.6 shows that the average of quark jets concentrates more than gluon jets which is a good discriminant in the CNN model to separate quark and gluon particles.

The jet images reconstructed above are implemented as inputs into the CNN model. The CNN model is a feed-forward artificial neural network that calculates the weights linearly on a m $\times$ n size input image. The input image is divided into several pixels and sent to the Convolutional filter. In the convolutional filters, the output of each filter is calculated by dot-product with different weights, shown in the following equation.

$$O = MI + B, \tag{4.2}$$

where, O and I represent the output and input of the convolutional filter. M is the weighting matrix and B is a constant matrix. Tuning the weighting matrix can change the image size of output and input.

After the operation, a new output image is reconstructed for the next filter. Max-pooling is another procedure for down-sampling the image. The output of Max-pooling is the maximum value in the defined image. The non-overlapping sets of pixels are taken from convolutional filter input and find the maximum value in the sets of pixels as output. The process in the Max-pooling is implemented to obtain the characteristic feature in the image. After the convolutional filter or Max-pooling, the 2-dimension image is flat into a 1-dimension dense layer. Then, the same thing done in the MLP model performs again to obtain the probability of quark jet and gluon jet. The illustration of the CNN model is shown in Figure 4.7. Track and topo-cluster normalized p_T images are used as one input in which the input shape is (16,16,2) in the CNN model. The CNN model consists of three iterations of a convolutional layer and is then connected with one Max-pooling layer. The Max-pooling layer is flatted and connected to two 1-dimension dense

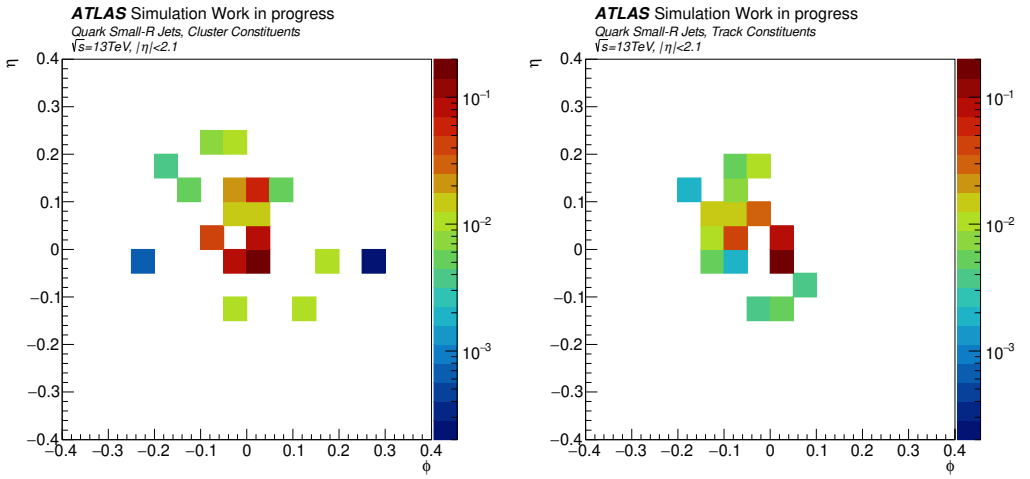

Figure 4.5: The CNN input image of quark jet. The topo-cluster constituents image is on the left hand side while the track constituents image is on the left hand side. Both of them have 16×16 pixels at $|\eta$ & $\phi| < 0.4$ region. The constituents in the associated jets are rotated to $\eta_{\text{jet}} = \phi_{\text{jet}} = 0$ and their p_T are normalized to $\sum p_T = 1$.

layers followed by a softmax function of size two, giving the probability for quark and gluon jet. Three convolutional filters have respectively 256, 128, 64 filters, which the sizes of filters are 7×7, 5×5, 3×3. The Max-pooling layers are operated with 2×2 downsampling. Two dense layers have 64 neurons, followed by softmax function. The setup of the CNN neural network model is the same as the MLP model, shown in Table 4.3.

4.2.4 MLP-CNN Model

MLP-CNN model consists of MLP model and CNN model described in Section 4.2.2 and 4.2.3. For the MLP part, there are 31 inputs including n_{track}, trackwidth, trackC1, EM energy fraction and energy of the calorimeter layers shown as Table 4.2. For the CNN part, topo-cluster and track images, same with the CNN model in Section 4.2.3. Combining MLP and CNN models means both high-level and low-level inputs are able to be used for more information of jets. There are two dense layers with 64 neurons for DNN parts. Same as CNN model in Section 4.2.3, three convolutional filters of 256, 128, 64 filters with size of 7×7, 5×5 and 3×3 respectively. Then the convolutional filters are followed by a Max-pooling layer with a size of (2,2). Finally, the output images are flatted to a 128-neuron dense layer. In order to combine MLP and CNN models, concatenate function is operated

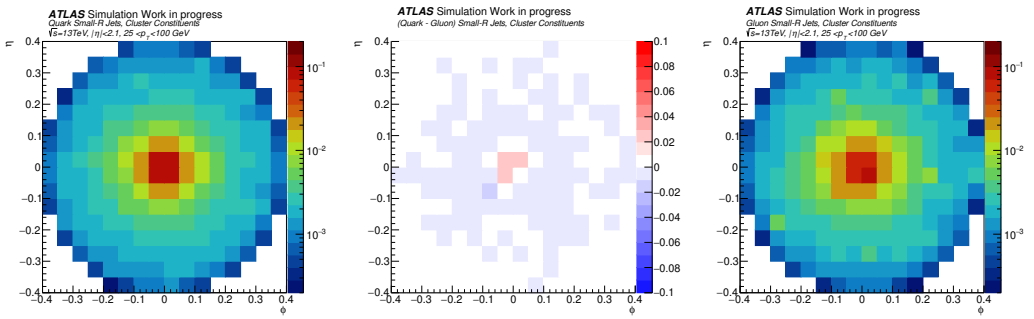

Figure 4.6: The average intensity of topo-cluster quark and gluon jets. The left image shows the average of topo-cluster quark jets while the right image shows the topo-cluster gluon jets. The middle image is drawn with the difference between the average of quark jets and gluon jets.

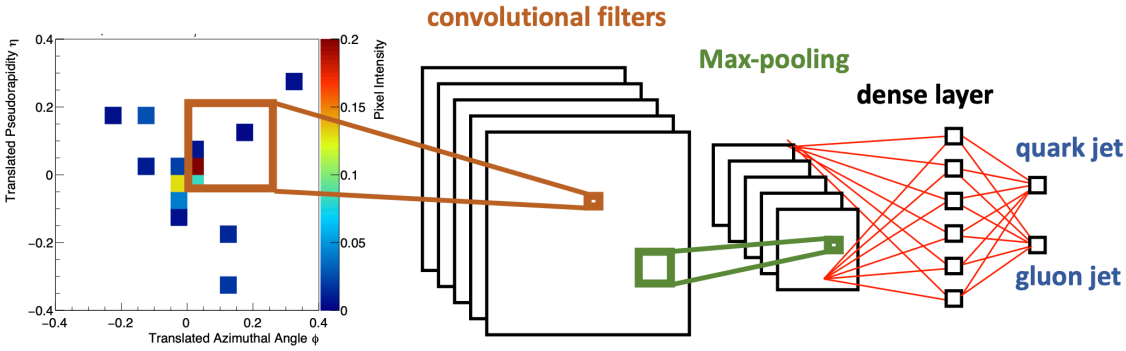

Figure 4.7: The schematic illustration of CNN architecture is taken from [34]. Track and topo-cluster normalized p_T images are used as input in this thesis.

to connect the 64-neuron dense layer in the MLP model with the 128-neuron dense layer in the CNN model. After combining two dense layers from MLP and CNN, respectively, it is followed by a 192-neuron dense layer and then output the softmax function with the size of two neurons which predict the quark-like and gluon-like probability. The setup of the MLP-CNN neural network model is listed in Table 4.3.

4.2.5 Pointwise Model

The CNN model is needed to have an image input, making the learning inconvenient. The Pointwise model is introduced as a model that can directly use input variables. This section has two Pointwise models: The Deep Sets

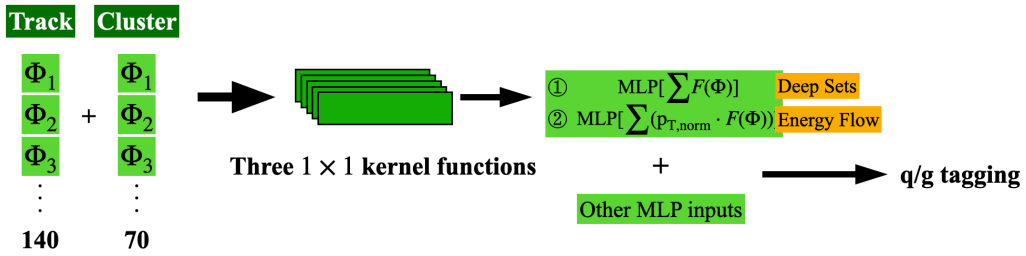

Figure 4.8: The schematic image of the Deep Sets and Energy Flow. For the Deep Set, the p_T, η, ϕ of topo-cluster or track is sent to input and the summation is operated to output the result followed by several dense layers. On the other hand, for Energy Flow, only η and ϕ are input and they are weighted by the associated p_T before the summation. Then after the summation, they are also sent to several dense layers. The dense layers above are connected to the MLP input dense layers to perform the quark/gluon separating.

model and the Energy Flow model. The Deep Sets and Energy Flow models have been recently developed well. For example, in ref [35], the Deep Sets model is used to generalize the permutation-invariant functions of variable-length in point clouds. Using the Deep Sets model, it is possible to solve the problems such as red-shift estimation of galaxy clusters. The critical operation in Deep Sets is the summation since symmetry usually exists in many physics fields. The summation operation can encapsulate the symmetric effect in the latent space. Consider a physics observable O is a symmetric function. Then the Deep Sets model can be written as follow.

$$O(x_1, ..., x_M) = F(\sum_{i=1}^{M} \Psi(x_i)) \tag{4.3}$$

Here x_i is (p_T, η, ϕ) of i-track or i-topo-cluster for Deep Sets model. And Ψ represents the operation before summation.
Next, let's consider a homogenous physics observable O again. With a modification of Equation 4.3, infrared-safe and collinear-safe (IRC-safe) observables can be added, shown as the following equation.

$$O(x_1, ..., x_M) = F(\sum_{i=1}^{M} z_i \Psi(x_i)) \tag{4.4}$$

Here x_i represents (η, ϕ) of i-track or i-topo-cluster and z_i is the p_T of each track in the Energy Flow model. Ψ is the operation before summation. The

summation operation implies that the infrared safety is still satisfied. The p_T weighting indicates that although the particles split into two particles, they can be reconstructed as if a vector with p_T weight. Hence, the Energy Flow model satisfies infrared safety and collinear safety which is named IRC-safe. The definition of IRC-safe is shown as follows.

1. Infrared safety : The observable is not affected by the low-energy and long distance physics, such as soft emission.

2. Collinear safety : Collinear splittings of particles don't affect the result.

The Pointwise model consists of several 1×1 kernel convolution layers. The calculations of the Pointwise model and the CNN model are similar. They are both operated by matrices to obtain the outputs. The output of the CNN model is an image. However, the output of the Pointwise model is one dimension layer. There are 140 sets of inputs for the tracks in the Deep Sets case. Each set of inputs will be operated with a matrix and output into a (64,1) one dimension layer. Therefore, there are 140 sets of one dimension layers after the operation.
Unlike the CNN model that only tackles the image inputs, Pointwise models are more flexible to have other kinds of inputs. Another advantage of the Pointwise model over the CNN model is that it is more convenient to input all p_T, η, ϕ information of associated tracks and topo-clusters. In the CNN model, the pixel size of the image is always a troublesome issue. If the pixel size is too large, the resolution and accuracy decrease. On the other hand, computer memory and training time become a fatal problem if the pixel size is made too tiny. This leads to a fundamental defect for the CNN model. In the Pointwise model, it can take p_T, η and ϕ as a set of input. Hence, as there are, says fifteen tracks, only fifteen inputs are needed in the associated jet. This saves the computer memory and time, and can use the jet information directly (the CNN model needs to transform the jet information into an image, so some information is lost actually.).
Since the Deep Sets and Energy Flow models are similar, the entire architecture will be introduced first, followed by the precise structure. The inputs are separated into two parts, MLP inputs and CNN inputs. For the MLP inputs part, it is the same as the MLP model, n_{track}, trackwidth, trackC1 and calorimeter layer energy; a total of 31 inputs are taken. For CNN parts, the input shape is different for track and topo-cluster. The shape of track and topo-cluster is (140, Φ) and (70, Φ), respectively. Here Φ is (p_T, η,ϕ) for the Deep Sets model and (η,ϕ) for the Energy Flow model. The number of 140 and 70 is set because the maximum of n_{track} and $n_{cluster}$ is found fewer than these two numbers. This ensures that all track and topo-cluster

information is taken as inputs. If the inputs don't approach the upper limit, say 140 or 70 for track and topo-cluster inputs, respectively, then the inputs are set to zero in order not to affect the training result. Then all inputs above are connected with several 1×1 convolutional layers followed by some operations and several dense layers. (1×1 kernel convolutional layer is called by Conv1D function in TensorFlow [36].) The following describes three input parts, track input, topo-cluster input and MLP input for Deep Sets and Energy Flow models.

- Track and topo-cluster inputs
  (140/70, [p_T, η, ϕ]) and (140/70, [η, ϕ]) are inputs in Deep Sets and Energy Flow for track/topo-cluster, respectively. Then there are three iterations of 1×1 kernel convolutional layers with the size of (64, 1)/(32,1). After passing three 1×1 convolutional layers, there is a small difference between Deep sets and Energy Flow models. For Deep Sets one, the output of the last 1×1 convolutional layer is multiplied with the initial input (p_T, η, ϕ). The reason for multiplying them is that there are some zero terms in the 140/70 inputs initially because the track number varies in different jets. The multiplying operation initializes the 1×1 convolutional layers by multiplying with the initial inputs. For example, there are five tracks in a jet. Therefore, the 140 1×1 convolutional layers multiplying with the initial inputs make only the first five 1×1 convolutional layers have value and the others are initialized to zero. Next, the summation is implemented to satisfy the Infrared Safety. For the Energy Flow model, the output of the 1×1 convolutional layer is multiplied with (p_T) in the initial input. Then the summation is performed again. Multiplying p_T and summing operation satisfy the IRC-safe for the Energy Flow model. Finally, two dense layers with 128/64 neurons follow the summing operation output in both cases.

- MLP input
  In the MLP input section, Deep Sets and Energy Flow models have the same structures. There are 31 inputs connected to three iterations of a 1×1 convolutional layer with a size of (16, 1) followed by two 64-neuron dense layers.

Then the outputs from MLP, track and topo-cluster are connected and sent to a 192-neuron dense layer followed by a softmax function of size two to perform the q/g discriminant.
Same as the models mentioned before, the setup of Pointwise models is described in Table 4.3.

Table 4.4: A summary of the input variables in each model.

Model	Input variables	
	n_{track}	
BDT	trackwidth trackC1	
MLP	n_{track} trackwidth trackC1	calorimeter layer energy
CNN	η, ϕ	$\frac{p_T(\phi,\eta)}{\text{all } p_T} = p_{T,norm}$
Pointwise	$\eta, \phi, p_{T,norm}$ calorimeter layer energy	n_{track} trackwidth trackC1

4.3 Performance of q/g Tagging Models

This section demonstrates the performance of the BDT, MLP, CNN, MLP-CNN, Deep Sets and Energy Flow model and the performance between the different testing samples. The input variables are summarized in the Table 4.4. Section 4.3.1 shows the Receiver Operating Characteristic Curve (ROC) curve of each model. Also, it describes the dependence of both p_T and η of associated jets and explains the difference among all models. Section 4.3.2 compares two kinds of MC samples generated Pythia and Herwig. The Pythia MC sample is a leading-order generator, and the Herwig MC sample is a next-leading-order generator. The difference between the Pythia and Herwig generators is the hadronization model. The hadrons are generated based on the string generating and cluster model in the Pythia and Herwig, respectively. Comparing these two MC samples gives brief research of studying MC sample generators for future works.
Since the earlystopping is applied to all neural models to avoid overtraining, the loss function convergence is examined and the overtraining didn't happen.

4.3.1 Receiver Operating Characteristic Curve Result

The receiver operating characteristic curve (ROC curve) shows the discrimination ability of a binary classifier system. Since the q/g tagger in this model is a binary tagger, it is convenient to compare the performance among different tagging models. There are many different types of ROC curves. In this section, the x axis represents the quark efficiency rate which means the per-

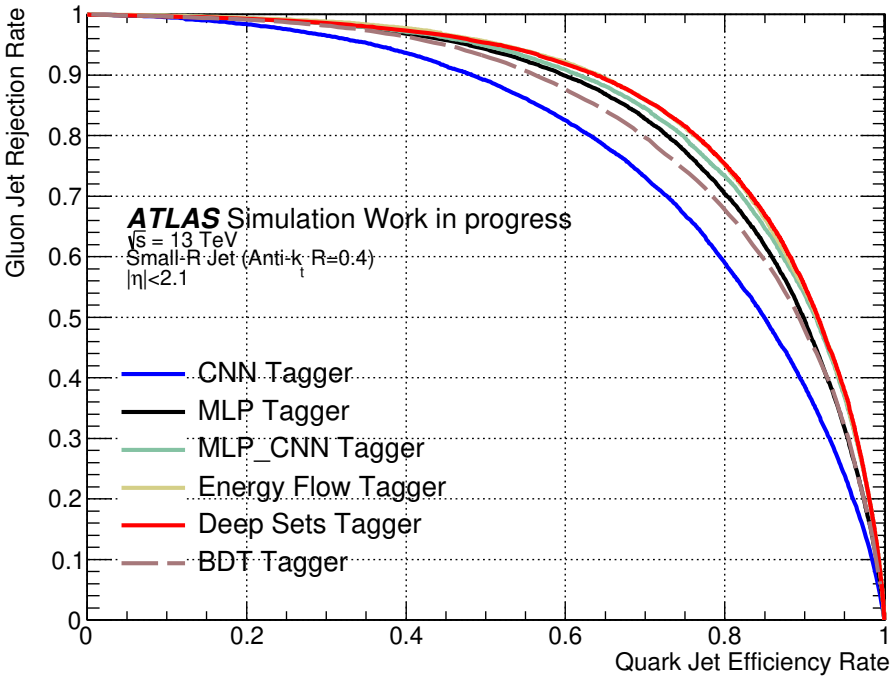

Figure 4.9: The ROC curves of all q/g taggers are described in Section 4.2. Quark Efficiency/Gluon Rejection (x and y axis) represents the percentage of remaining quark/rejected gluon after the threshold cut on quark/gluon discrimination output.

centage of remaining quark after the discrimination threshold. On the other hand, the y axis represents the gluon rejection rate which means the percentage of rejected gluon after the discrimination threshold. The definition is shown in the following equations.

$$\text{Quark Efficiency} = \frac{\mathrm{N}_{\text{true postive}}}{\mathrm{N}_{\text{positive}}}, \quad \text{Gluon Rejection} = 1 - \frac{\mathrm{N}_{\text{true negative}}}{\mathrm{N}_{\text{negative}}}, \quad (4.5)$$

where $\mathrm{N}_{\text{positive}}$ and $\mathrm{N}_{\text{negative}}$ are the numbers of particles with a quark and a gluon labels generated by the MC generator (it is the numbers of quarks and gluons, respectively.). True positive means it is a quark and predicted as a quark by the q/g tagging models. True negative means it is a gluon and recognized as a gluon by the q/g tagging models. Using this definition, the ROC curve of q/g taggers in this study are shown as concave curves and the ROC curve with up and right curvature has better performance.
Figure 4.9 shows the ROC curve of all q/g tagger introduced in Section 4.2.

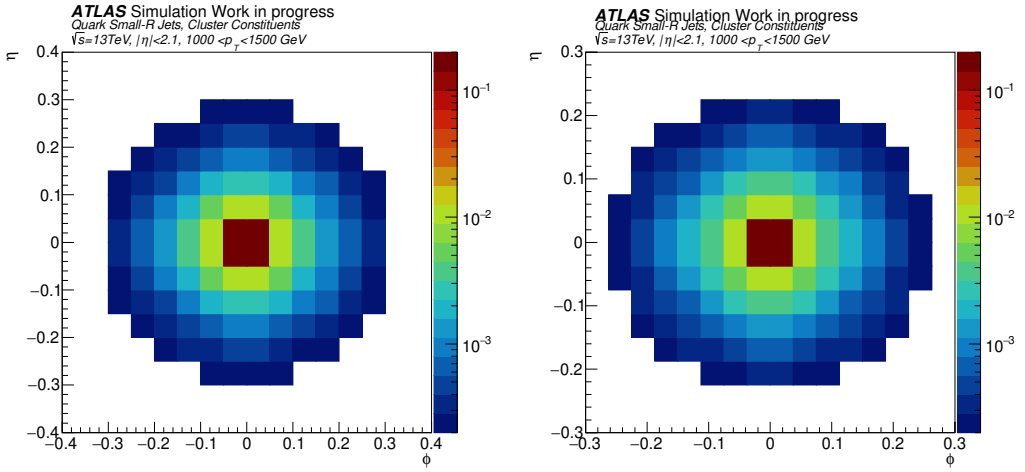

Figure 4.10: The topological cluster image in $1000 < p_T < 1500$ GeV region. Left image has the pixel size of 0.05 at $|\phi,\ \eta| < 0.4$ while right image has the pixel size of 0.0375 at $|\phi,\ \eta| < 0.3$.

Deep Sets (red curve) and Energy Flow (yellow curve) models have almost the same performance and achieve the best performance among all models. The inputs of the Pointwise model (Deep Sets and Energy Flow) are the same as MLP-CNN (green curve) model but better than it. This implies that compared to the MLP-CNN model, the Pointwise model utilizes the information of inputs more efficiently. The pointwise model utilizes the track and topo-cluster information better than the MLP-CNN model since CNN input is needed to consider pixel size limitations.
Ideally, the CNN (blue curve) model should have performed better than MLP (black curve) and BDT (dotted brown curve) tagger because the CNN model uses more low-level inputs with more information than the MLP and BDT with only high-level inputs. However, the ROC curve shows that the CNN model is worse. This is because pixel size optimization is needed to maximize the performance of the CNN model. To check how pixel size optimization problems affect the performance, the pixel is tuned to a smaller size, shown in Figure 4.10. The η and ϕ range are changed from smaller than 0.4 to smaller than 0.3, and the pixel size becomes from 0.05 to 0.0375. This smaller pixel size allows one to see the track and topo-cluster image more precisely. The result is shown in Figure 4.11. The pixel size is related to the amount of input information. Therefore, the performance of smaller pixel sizes better than larger pixel size emphasizes how significant the input variables are. As the pixel size becomes smaller, the performance of the CNN model gets better as

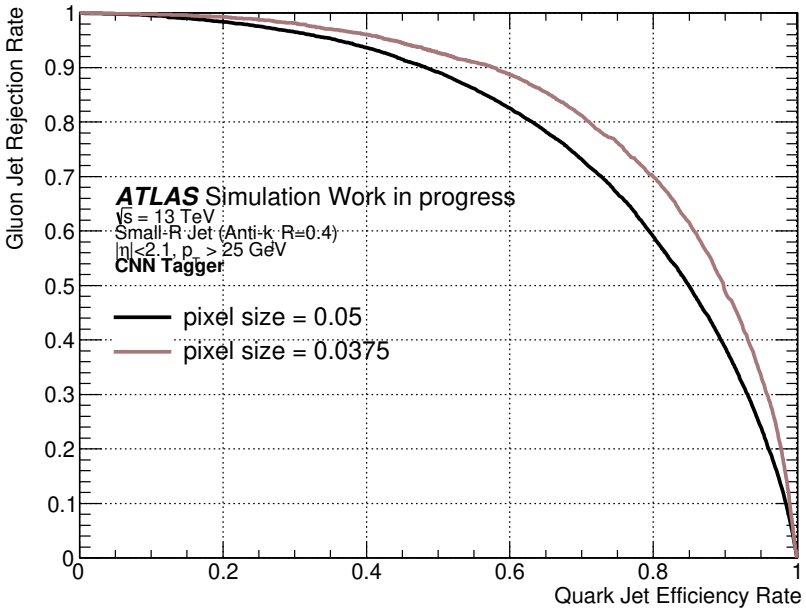

Figure 4.11: The comparison between different pixel size CNN tagger. A smaller pixel size CNN model (brown curve) is apparently having a better larger pixel one (black curve). This indicates that ideally reducing the pixel size gives the same performance as the Pointwise model.

expected. Nevertheless, the smaller pixel size in the same η, ϕ range implies more pixel numbers. This requires more computer memory and takes more time to train the models, which is impractical.

Although the performance of the MLP-CNN model is worse than the Pointwise model, it is better than any others. The MLP-CNN inputs usage gives a better result than MLP, BDT, and CNN models that use partially or less information.

The MLP model is better than the BDT model because MLP utilized the calorimeter layer energy as additional inputs. Again, these things imply the importance of the input. For comparing the difference between different architecture, the MLP using only n_{track}, trackwidth, trackC1 variables are also input. The result is shown in Figure 4.12 and the difference between them is tiny. From this, it is known that the tagger performance depends more on the input variables and the way to tackle the inputs, and not so much on the model architecture.

Finally, compared to the conventional tagger (BDT model), considered as the baseline of machine learning, the gluon rejection rate of the Pointwise

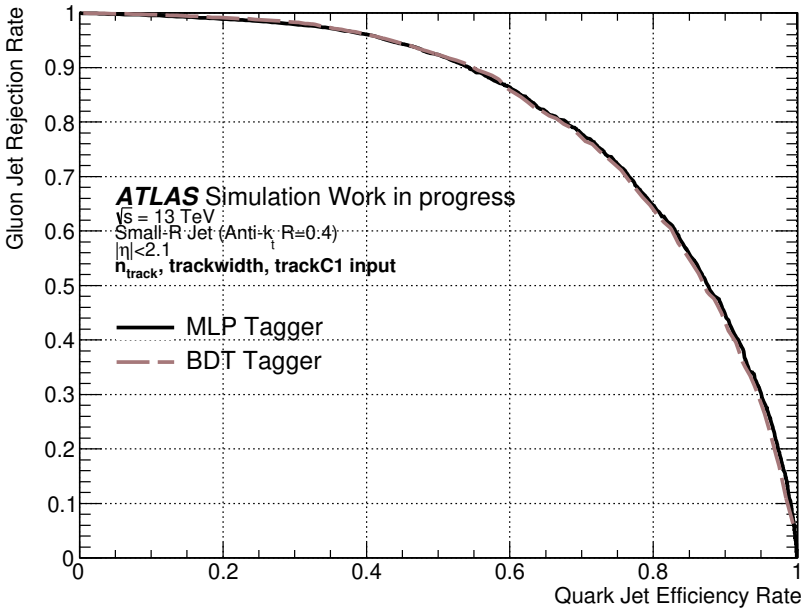

Figure 4.12: Compare MLP and BDT tagger with same inputs (n_{track}, trackwidth, trackC1). With the same inputs, the performance of the two models is similar, indicating that input information is more important than the model architecture.

models improved approximately 10% better at the 80% quark efficiency rate, where the gluon rejection rates of Pointwise and BDT models are around 75% and 68%, respectively. This shows that neural network models using low-level inputs can improve the q/g tagging significantly.
Besides the performance at all ranges, the dependence on p_T and η is also essential. Different transverse momentum and pseudorapidity range indicates the different fields of physics analysis. Hence, different strategies are needed to consider in the associated region. Figure 4.13 displays the p_T and η dependance for all models. They are plotted at 80% quark efficiency which means 80% quark remains as gluon jet rejection rate is obtained.
In the p_T dependance plot (top), the bin is set to 100 GeV below 1000 GeV such as 0-100 GeV, 100-200 GeV, etc. Above 1000 GeV, 1000-1500, 1500-2000, 2000-3000, 3000-4000, 4000-5000 GeV bins are selected. It comes to be better performance as p_T increases in the plot. Worse performance at low p_T region results from the similar track and topo-cluster distribution of low-energy quark/gluon. The high-energy gluon jet turns into generating more particles, making it easier to implement the q/g tagging. At $p_T < 100$ GeV

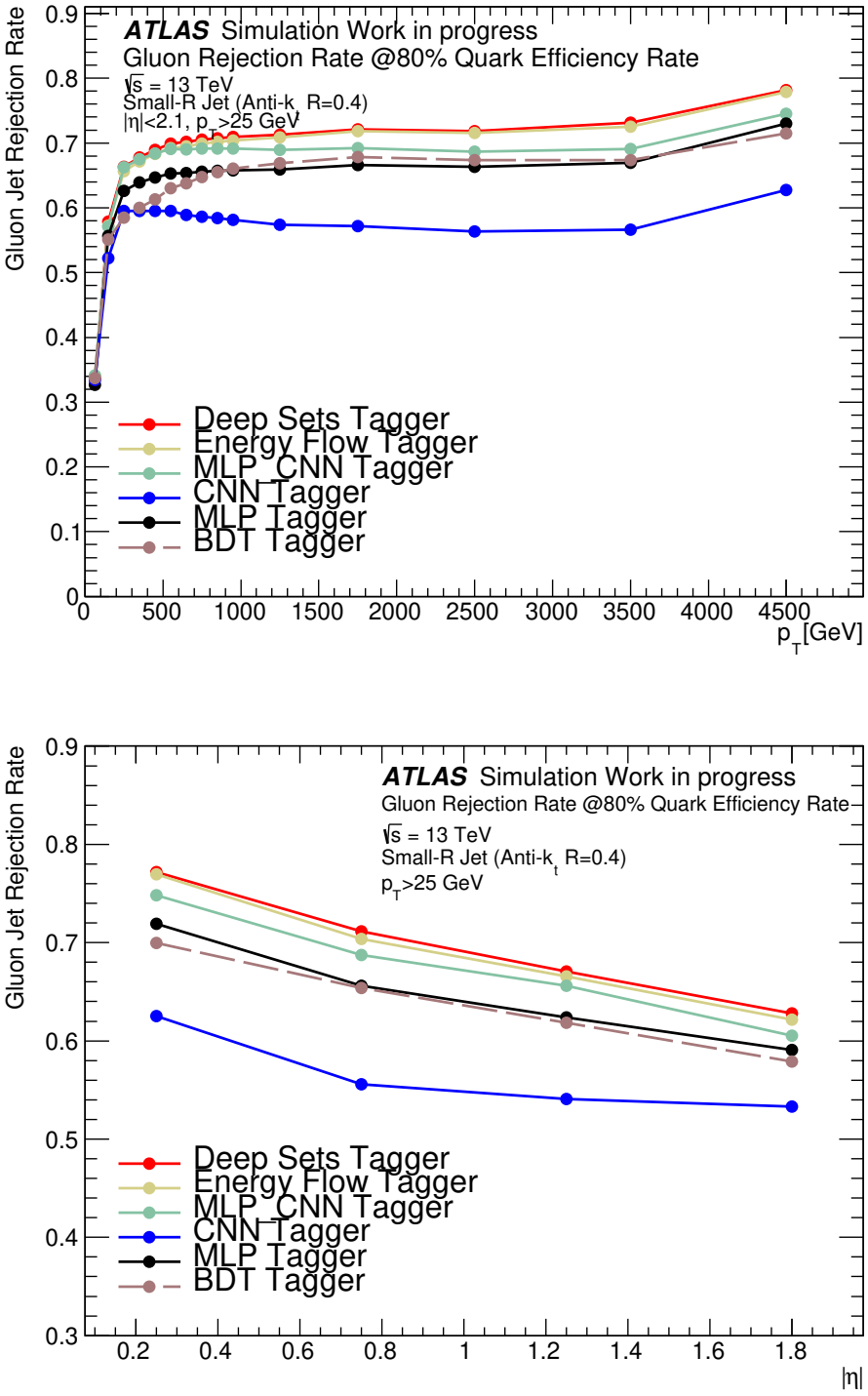

Figure 4.13: The transverse momentum (top) and pseudorapidity (bottom) dependence of all models introduced in Section 4.2. The y axis in both plots is gluon jet rejection rate at 80% quark efficiency rate.

region, all model has similar performance. At $100 < p_T < 200$ GeV region, the Deep Sets, Energy Flow and MLP-CNN tagger have better performance. In the η dependance plot (bottom), the $|\eta|$ bin is selected as 0-0.5, 0.5-1, 1-1.5, 1.5-2.1. It performs better at the low η region because of the inner detector in the ATLAS. At lower η, the number of associated jet tracks increases and gives more information. More information brings better accuracy and better performance. Same as the p_T dependence plot, Deep Sets and Energy Flow taggers have the best performance among all models in all regions. It can be concluded that neural network models can use more information and better performance at all p_T and η regions.
In conclusion, it is worthwhile to use different types of flexible neural network models to include more jets information to operate the q/g tagging. For example, besides the p_T intensity of track and topo-cluster used in Pointwise models, the charge of particles inside the jets can be added to the Deep Sets model as another input information.

4.3.2 Difference Between Pythia and Herwig MC sample

The MC sample used in Section 4.3.1 is generated by Pythia 8 generator. To compare the difference between other MC generators, this Section introduces another MC sample based on Herwig 7 generator [37]. The multijet MC sample generated by Herwig 7 matches to next-to-leading order (NLO) calculations developed by Matchbox module [38]. Predictions of the NLO calculations in the perturbative effects of QCD in the strong coupling give significant accuracy in the LHC experiments. Moreover, the MC sample generated by Herwig 7 in this section exploits the angular-ordered and dipole showers for the multijet merging algorithm [37, 38].
The multijet MC samples generated by Pythia 8 and Herwig 7 are learned in Deep Sets and Energy Flow models, respectively. And then the models trained on different MC samples are used to test Pythia 8 and Herwig 7 generated Samples. There are four different categories. Trained on Pythia and tested on Pythia, trained on Pythia and tested on Herwig, trained on Herwig and tested on Herwig, trained on Herwig and tested on Pythia. Comparing these four categories with Deep Sets and Energy Flow to investigate the properties of different models and MC samples. The ROC curves for Deep Sets and Energy Flow models are plotted in Figure 4.14. From the ROC curves, it is clear that no matter what samples are used to train, the performance of testing on Pythia Sample is always better than testing on Herwig Sample.

Another thing is that the Energy Flow satisfies IRC-safety, and it is expected to have several advantages. First, the IRC-safe satisfying models less affected by the soft emission indicate they can perform better at the lower energy scale, which is extremely important for the quark/gluon discriminant. Second, IRC-safe models correspond to the robustness of the collinear splitting, implying the consistency of distinguishing quarks and gluons' ability. Third, the models have more generality on both Pythia and Herwig Samples, even when based on different hadronization generating methods.
From difference panels in Figure 4.14, the difference between Pythia and Herwig testing on Energy Flow is smaller than on the Deep Sets model. The results above give several brief conclusions. Either trained on Pythia MC sample or Herwig Sample, testing on the same MC sample has similar results. This can be seen by comparing the dotted lines or solid lines. Then, the MC sample generated by the Pythia seems more accessible to separate quarks and gluons by q/g tagging than the MC sample generated by the Herwig generator. Finally, the Energy Flow model has more similar performance between the different MC samples, but further research is still needed to investigate whether this is from IRC-safe ability. These might result from the fundamental difference of the generating strategy between Pythia and Herwig generators.
The difference of n_{track}, trackwidth and trackC1 between Pythia and Herwig MC samples are described in Appendix A. This section gave a brief work for studying these two MC samples with Deep Sets and Energy Flow models. It is also attractive to analyze the Pythia and Herwig MC samples with not only Deep Sets and Energy Flow but also other neural network models that can utilize more low-level information for future works.

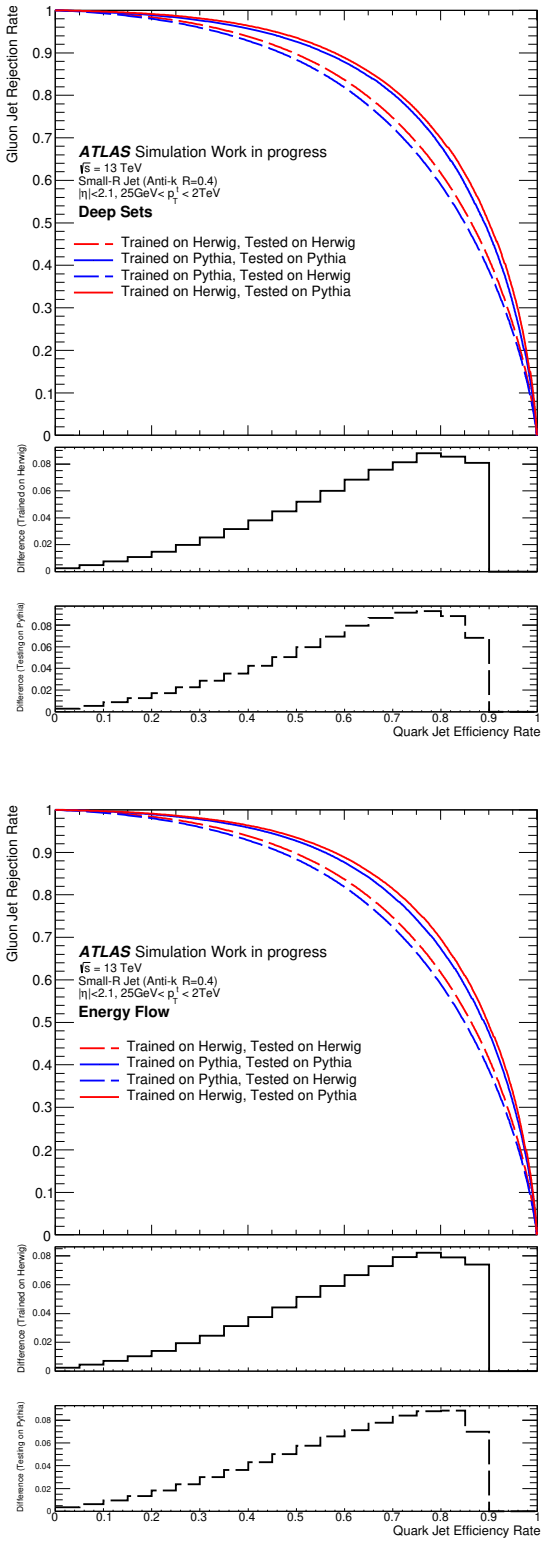

Figure 4.14: The ROC curves trained on Herwig or Pythia and tested on Herwig or Pythia with Deep Sets/Energy Flow models are shown in the top/bottom figure. Solid/Dotted curves represent testing on Pythia/Herwig Sample. Blue/Red curves represent training on Pythia/Herwig Sample. The black solid/dotted line stands for the difference between the models training on the different samples and testing on the same samples.

Chapter 5

Quark/Gluon tagging Application in 1-lepton Semileptonic VBS

The q/g tagger should improve the sensitivity to searches and measurements of electroweak processes in the hadronic final state, characterized by quark-induced jets in the events. Usually, analyses using hadronic final states are complex at the LHC due to the huge amount of multijet background. However, if the q/g tagger can select quark-induced jets while suppressing the gluon-induced jets, it can strongly suppress the multijet background. As a benchmark electroweak process at the LHC, VBS has been studied in this chapter. Since the Energy Flow model has the best performance among all models, it is used to separate the VBS signal and background samples for validation.

The measurement of the semileptonic VBS processes was performed by ATLAS using a 36fb^{-1} dataset [12]. This thesis only focuses on the 1-lepton resolved channel for convenience. In the 1-lepton resolved channel, aiming for WW and WZ scatterings at the lower effective center-of-mass energy, the signal candidate events were selected by four hadron jets, in addition to exactly one lepton in the event. For signal events, all four jets are originated from quarks. The main sources of the background process are W+jets and $t\bar{t}$. About 32%, 12%, and 1% of W+4jets events have 2, 3 and 4 gluon-induced jets in the selected jets. In this thesis, the analysis in Reference [12] is tried to be reproduced, and the impact of the q/g tagger is studied.

The MC samples used for Application and the setup for their generators are demonstrated in Section 5.1. The Event Selection for 1-lepton semileptonic VBS and background Samples are shown in Section 5.2. There are mainly three q/g tagging strategies for selecting purer VBS signal, also introduced

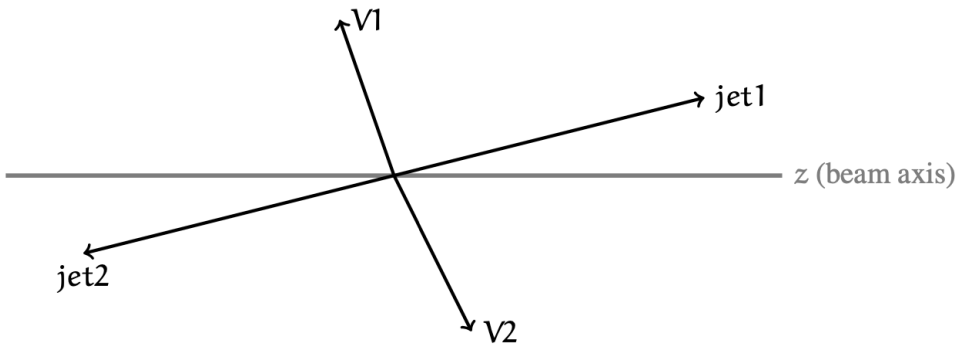

Figure 5.1: An illustrating diagram of a VBS topology taken from [12].

in Section 5.2. The performance and comparison of different strategies are described in Section 5.3.

5.1 Introduction of MC samples

The signal sample used for q/g tagging application is the semileptonic VBS MC sample. For the brief application, only two kinds of background samples are used in this Chapter. One is W+jet and the other one is $t\bar{t}$ MC sample. In the VBS process, two forward jets originate from spectator quarks. This forward jet pair is required with the highest invariant mass (referred to m_{jj}^{tag}) of all small-R jet pairs in each event. There are also two Vector bosons (Z or W bosons). In this thesis, one decays into two quarks and the other decays through the leptonic process, which is the so-called semileptonic process. The schematic diagram is shown in Figure 5.1. Section 5.1.1 gives a brief introduction for Semileptonic VBS Sample and Background Samples in the study. Section 5.1.2 explains the Object Definition such as lepton, jets, etc.

5.1.1 Signal Sample and Background Samples

The ATLAS experiment collects the luminosity 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV from 2015 to 2018 during Run II of the LHC. Therefore, the signal and background samples used in this Chapter are scaled to the luminosity of 139 fb^{-1}.
This Chapter focuses on the 1-lepton channel of the Semileptonic Vector Boson Scattering MC sample. There are two channels considered, which are $WW \to \ell\nu qq$ and $WZ \to \ell\nu qq$. The semileptonic VBS samples are generated by MadGraph (v.2.6.6)[39]+Pythia8[28]+EvtGen(v.1.7.0)[40]. The genera-

Table 5.1: The MC samples used in Chapter 5.

Process	Generator	Cross section[pb]
VWjj→ $\ell\nu qq$+jj	MadGraph + Pythia8 +EvtGen	2.25
W+jet → $\ell\nu$ + jet	SHERPA 2.2.1	6.16×10^4
$t\bar{t} \rightarrow \ell\nu qq$bb	POWHEG + Pythia8 + EveGen	396.89

tor uses the A14[30] and NNPDF23LO PDF[41] as a set of tuned parameters. The VBS MC samples are generated with two on-shell vector bosons (WW or ZW), which one decays leptonically (W → $\ell\nu$, where $\ell = e, \mu, \tau$) and the other one decays hadronically (V → qq). The details of the MC sample are shown in Table 5.1.
There are two background MC samples used in this application. One is W+jet and the other one is $t\bar{t}$ sample. The W+jet Sample is generated by the SHERPA 2.2.1 event generator[42]. The W+jet event generator uses the parameter set of NNPDF30LO PDF [41]. The W boson is enforced to decay to a lepton and a neutrino in the W+jet MC samples.
$t\bar{t}$ → WbWb → $\ell\nu$bqqb/$\ell\nu$b$\ell\nu$b process is considered in the $t\bar{t}$ MC sample. It is generated by the event generator POWHEG[43] + Pythia 8[28] + EvtGen (v1.6.0). The POWHEG event generator is the abbreviation of POsitive Weight Harrest Emission Generator. As its name, only the positive event weight is generated. The POWHEG event generator is developed to tackle the negative event weight problem and is usually utilized in a hadronic process.
This thesis aims to simulate the jjjj$\ell\nu$ final states. Electromagnetic particles and hadrons are simulated at the inner detector and calorimeters in the ATLAS detector. Muons are simulated from the inner detector to the muon spectrometer.
The backgrounds from the additional proton-proton collisions are referred to as pile-up, presenting a serious issue to the analyses in the ATLAS. Thus, pile-up background events are considered to study more data-like samples. The Monte Carlo simulation method[44] is applied to generate the pile-up events and add them with other events. Figure 5.2 shows the distribution of interactions mean number utilized to run the simulation events.

5.1.2 Physics Object

Physics Object is defined from the particles representing the observed characteristics and used to reconstruct the events. In this section, the lepton(e, μ), Jet and E_T^{miss} will be described for the later analysis.

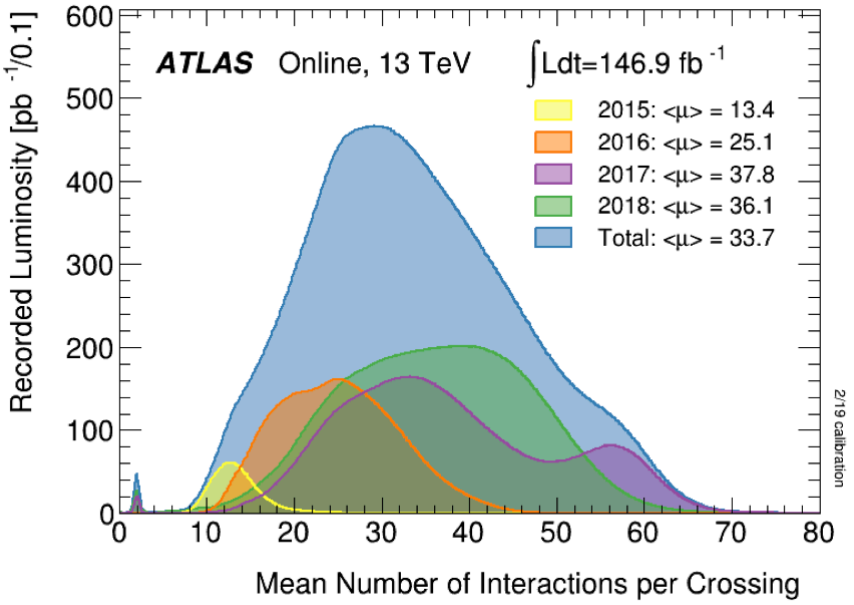

Figure 5.2: Mean number of interactions per crossing from 2015 to 2018.

- Electron
  In this analysis, the electron p_T threshold is set at 30 GeV and $|\eta| < 2.47$ is required for the reason of inner detector cover range.
  There are three working points (Loose/Medium/Tight) for Electron ID[45], using the Likelihood-based (LH) algorithm. The electron object in this chapter uses the TightLH working points. The details is described in Table 5.2.
  d_0 is the minimum distance between the primary vertex and the track. When d_0 is calculated, the minimum distance point will be found on the track. z_0 represents the projecting distance between the primary vertex and this minimum distance point on the z axis. The BL means that the parameters are relative to the beamline and σ is its uncertainty.
  The electron object energy is calibrated by the BDT to reduce the effect from the calorimeters and their neighboring materials. The BDT is trained by the MC simulations and tested by the MC samples and real data detected in the ATLAS[46].

- Muon
  Muons are minimum ionizing particles having long decay life. They are expected to go through all devices in the ATLAS detector, from

Table 5.2: The electron object definition used in this study.

Cut	Selection
p_T	$p_T > 30$ GeV
η	$\|\eta\| < 2.47$
Track to Vertex Association	$\|d_0^{BL}(\sigma)\| < 5$, $\|\Delta z_0^{BL} \sin\theta\| < 0.5$ mm
Identification	ElectronID = TightLH

Inner Detector, calorimeter, to Muon Spectrometer[5]. Therefore, this analysis requires that muons leave the tracks in these three regions. The definition of the signal muon is shown in Table 5.3.
$Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events are exploited to calibrate muon energy scale and resolution. The calibration is imposed by fitting the invariant mass of muon pairs of these two processes on the MC simulation to the observed data.

Table 5.3: The Muon object definition used for the MC sample in this chapter.

Cut	Selection
p_T	$p_T > 30$ GeV
η	$\|\eta\| < 2.5$
Track to Vertex Association	$\|d_0^{BL}(\sigma)\| < 3$, $\|\Delta z_0^{BL} \sin\theta\| < 0.5$ mm
Identification	MuonQuality = Medium

- Small-R Jet
  Jet is one of the essential physics objects in LHC experiments. The JVT cut is utilized to suppress the object from the pile-up jet. The introductions of jet reconstruction are already shown in Section 2.3.1. This application study focuses on the small-R jet investigating the quark/gluon tagging. Table 5.4 shows the details of the definition for Small-R Jet Object.
  Small-R jet calibration consists of several serial steps: the origin correction, the jet are a based pile-up correction, the MC-based calibration, and the global sequential calibration.
  The origin correction performed in the first step is reorienting a small-R jet to the primary vertex. Secondly, the jet area-based pile-up correction is imposed to reduce the impact of the pile-up contamination. In the third step, the MC-calibration uses the MC simulation sample to determine the energy response function of $|\eta|$. In the global sequential calibration step, the residual dependence on the jet substructure

is corrected by the fraction of the jet energy measured in the hadronic calorimeter.

Table 5.4: The Small-R Jet object definition used for the MC sample in this chapter. EMPFlow represents the particle flows (small-R jets here) are reconstructed with electromagnetic scale topo-cluster.

Cut	Selection
Algorithm	anti-k_T (R=0.4)
Input Constituent	EMPFlow
p_T	p_T >20 GeV
η	$\|\eta\| < 4.5$
JVT	> 0.95 for $p_T < 120$ GeV and $\|\eta\| < 2.4$ > 0.11 for $p_T < 120$ GeV and $2.4 < \|\eta\| < 2.5$ (Medium Working Point)

- Missing E_T
  Since neutrinos don't interact with anything in the ATLAS detector, observable physics objects' energy and momentum vector are not conserved. The missing transverse momentum vector in the detector is defined as E_T^{miss}. The calculation can be written as Equation 5.1 easily.

$$\begin{aligned}\sum E_T = &\sum E_T^e + \sum {E_T}^\gamma + \sum {E_T}^\tau \\ &+ \sum E_T^{jet} + \sum E_T^\mu \\ &+ \sum E_T^{\text{soft object}} + E_T^{miss}\end{aligned} \tag{5.1}$$

  $E_T^{e,\gamma,\tau,jet}$ represents the transverse energy E_T of electron photon, τ lepton and jet. E_T^μ represents the transverse energy of muon. $E_T^{\text{soft object}}$ represents the object is observable but cannot be defined as an object with the selection, such as the soft particles not reconstructed into jets.
  The missing E_T calibration is performed with $Z \to ee$ and $Z \to \mu\mu$ events. Since these events do not have neutrinos, the transverse momentum is used to calibrate the missing E_T by measuring the deviations from the ee, $\mu\mu$ events.

5.2 1-lepton Event Selection

The selections are performed to enhance the signal purity. For an easy discussion, this section only describes the 1-lepton event. Section 5.2.1 demonstrates the preselection first. Section 5.2.2 will introduce the BDT discriminant and its variables. Section 5.2.3 shows that different strategies by q/g tagging.

5.2.1 Preselection

According to the jet selection criteria, there are mainly two categories in the semileptonic VBS process. One or more large-R jet in events is categorized as merged events. Otherwise, events without large-R jets are categorized as resolved events. For validating q/g tagging, this paper only discusses 1-lepton resolved events.
For 1-lepton resolved events, the candidates include one lepton and one neutrino, a pair of tagging jets[1]. And a pair of signal jets originating from the signal vector boson. Therefore, the preselection is imposed to obtain the pure 1-lepton resolved events.
The preselection cut is shown as follows.
The lepton trigger is applied to the MC samples in this Chapter. Common small-R jet selection is applied to both tagging and signal jets, $p_T > 20$ GeV for $|\eta| < 2.1$ and $p_T > 30$ for $2.1 < |\eta| < 4.5$. The cut of $|\eta| < 2.1$ is because the q/g tagging models uses the input variables at $|\eta| < 2.1$ region only.
For $W \to \ell\nu$ samples, only one tight lepton, and no loose lepton are required. The tight lepton and loose lepton stand for the lepton passing through the ElectronID and MuonQuality cut with a tight or loose working point. The lepton (e and μ) transverse momentum threshold is $p_T > 30$ GeV. The missing E_T cut is also imposed with $E_T^{miss} > 80$ GeV because of the existence of the neutrino.
For tagging jets selections, tagging jets are selected by choosing the largest invariant mass(m_{jj}^{tag}) of a small-R jets pair each event. Both leading and subleading tag jet transverse momentum is required as $p_T > 30$ GeV. The invariant mass of tagging jets pair satisfies with $m_{jj}^{tag} > 400$ GeV. The m_{jj}^{tag} distribution is displayed in Figure 5.3.
For signal jets selection, other than tagging jets, the number of small-R jets needs to be greater than or equal to two jets. To acquire more signal-like events, leading signal jets transverse momentums are required as $p_T > 40$ GeV.

[1]A jets pair directly coming from partonic quarks inside the collision protons.

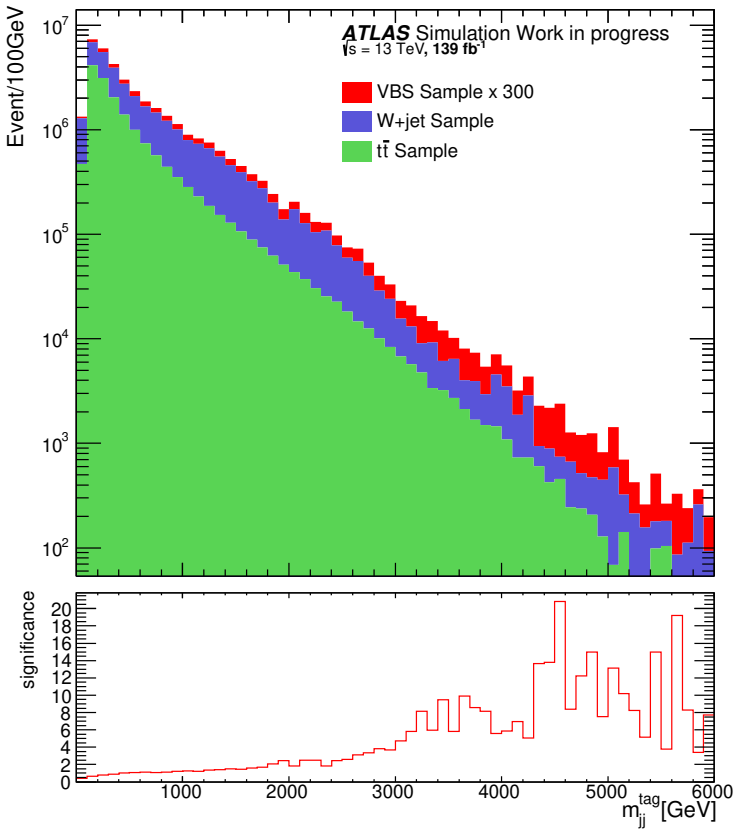

Figure 5.3: The m_{jj}^{tag} distribution of 1-lepton resolved process. The events are selected with the 1-lepton trigger and have exactly one lepton. Missing E_T cut and jet p_T cut are also required as shown in Section 5.2.1.

Additional two preselection cuts are applied to enhance the purity. The number of b-tagged jets equal to zero is needed. The invariant mass of two signal jets and a specific jet is required as $m_{jjj} > 220$ GeV. The specific jet is selected by calculating the closest invariant mass m_{jjj} from the top quark mass[2]. This selection is also imposed to reduce the background from the $t\bar{t} \to \ell\nu qqbb$ process since top quark mainly decays through the process $t \to Wb \to qqb$. After the selections above, two different signal jets pair section strategies are considered in this analysis. One is 2leading and the other one is minmass, shown as follows.

- 2leading : The signal jets pair is selected by choosing the two highest

[2] $m_t = 172.76$ is used in this study.

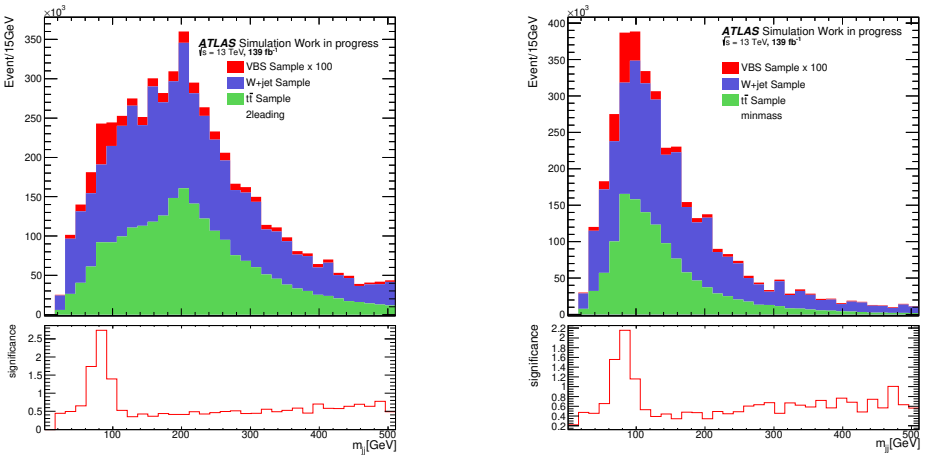

Figure 5.4: The distribution of signal jets pairs invariant mass. All preselection cuts except signal region cut are applied to both plots. The 2leading and minmass strategies plots are shown on left and right, respectively. The 2leading has better significance because the background is less than the minmass strategy. However, the minmass strategy has more signal and background events, implying it might have more space to improve by q/g tagging.

p_T small-R jets other than tagging jets.

- minmass : The signal jets pair is selected by choosing the two small-R jets other than tagging jets that have the closest invariant mass from W/Z boson mass[3].

The 2leading strategy is a new selection strategy introduced in this thesis. On the other hand, the minmass strategy was applied in the Reference [12]. Both 2leading and minmass strategies have different advantages and disadvantages. 2leading strategy has a larger probability of selecting the jets from $t\bar{t}$ samples, but the significance is better than the minmass strategy. The invariant mass selection makes sure that the minmass strategy obtains more jets decaying from W/Z vector bosons. However, the m_{jj} distribution tends to concentrate at 80-90 GeV region leading to the decrement of significance. The m_{jj} distributions of 2leading and minmass strategies are shown in Figure 5.4.
The signal region (SR) selection is defined as the invariant mass of signal jets pair is within the range of W/Z boson mass. $64 < m_{jj} < 106$ GeV is imposed to obtain a pure VBS signal. The summary of preselection is listed in Table

[3]This analysis utilizes $m_W = 80.385$ GeV and $m_Z = 91.187$ GeV.

Table 5.5: A list of preselection for 1-lepton semileptonic VBS signal search in this analysis. The jet m_{jj}^{tag} and p_T are shown in Figure 5.3 and 5.6, respectively. Other distributions are described in Figure Appendix B

Objects	Cuts	threshold
$W \to \ell\nu$	Number of Tight leptons	1
	Number of Loose leptons	0
	E_T^{miss}	> 80 GeV
	$p_T(\ell)$	> 30 GeV
tagging jets pair	tagging jets p_T	> 30 GeV
	m_{jj}^{tag}	> 400 GeV
Signal jets pair	Number of signal jets	≥ 2
	p_T(signal jet)	> 20 GeV for $\|\eta\| < 2.1$ > 30 GeV for $2.1 < \|\eta\| < 4.5$
	Leading jet p_T	> 40 GeV
Others	Signal Region	$64 < m_{jj} < 106$ GeV
	Number of additional b-tagged jets	0
	m_{jjj}	> 220 GeV

5.5.
The BDT discriminant in Section 5.2.2 and q/g tagging model in Section 5.2.3 will be applied to both two strategies. And then the performance of them will be discussed in Section 5.3.

5.2.2 BDT Discriminant

The SR ($64 < m_{jj} < 106$ GeV) defined in the previous section is not enough to observe the VBS signal process with the required significance. Hence, the BDT machine learning tool is performed to enhance the SR purity. The BDT is used to learn the signal sample (VBS sample) and background samples (W+jet and $t\bar{t}$ samples), respectively. The W+jet and $t\bar{t}$ MC samples are applied the preselection individually. The input variable set and hyperparameters of the BDT in this analysis follow the example in Reference [12]. The chosen hyperparameters are shown in Table 5.6. The hyperparameters are considered to be against overtraining effect.
Input variables are selected to have higher discriminant between signal and background samples. Most variables are related to kinematics such as transverse momentum. Furthermore, the quark/gluon separating variables such as the number of tracks and trackwidth is also considered since the backgrounds most come from QCD interactions which generate numerous gluons.

Table 5.6: A summary of hyperparameters in BDT Discriminant model.

BDT hyperparameters	1-lepton
NTrees	800
MaxDepth	4
Shrinkage	0.3
MinNodeSize	5%
nCuts	20
UseBaggedBoost	True
BaggedSampleFraction	0.5
SeparationType	GiniIndex

There are totally 23 input variables for the BDT discriminant training. The definition of each variable is described as follows.

- m_{jetjet} : Invariant mass of jets pair.
- $\Delta\eta_{jj}^{jetjet}$: η difference between two jets.
- p_T^{jet1} : Transverse momentum of leading jet.
- p_T^{jet2} : Transverse momentum of subleading jet.
- w^{jet} : Trackwidth of jet.
- n_{tracks}^{jet} : Number of tracks within jet.
- n_j : The number of jets having track constituents.
- $n_{j,extr}$: The number of additional jets other than the jets from the vector boson.
- $\Delta R(\ell, \nu)$: Angular separation between lepton and neutrino.
- ς_V : Boson centrality, a topological variable.
- m_{VV} : Invariant mass of two vector bosons from signal jets pair and lepton-neutrino pair.
- m_{VVjj} : Invariant mass of vector bosons pair and tagging jets pair.

Boson centrality is defined as $\varsigma_V = \min(\Delta\eta_-, \Delta\eta_+)$,
where $\Delta\eta_+ = \max[\eta_{tag,j_1}, \eta_{tag,j_2}] - \max[\eta(V_{had}), \eta(V_{lep})]$
and $\Delta\eta_- = \min[\eta(V_{had}), \eta(V_{lep})] - \min[\eta_{tag,j_1}, \eta_{tag,j_2}]$. This variable is sensitive when the tagging jet has a significant separation in η.

The W boson mass constraint determines the momentum of the neutrino on the lepton because of $W \to \ell\nu$ decay. The entire calculation is shown as following equations.

$$\begin{aligned} {m_W}^2 &= (E_\ell + \sqrt{p_{T,\nu}^2 + p_{z,\nu}^2})^2 - ((\overrightarrow{p}_{T,\nu} + \overrightarrow{p}_{T,\ell}) + (p_{z,\nu} + p_{z,\ell})), \\ &= m_\ell^2 + 2E_\ell\sqrt{p_{T,\nu}^2 + p_{T,\ell}} - 2\overrightarrow{p}_{T,\nu} \cdot \overrightarrow{p}_{T,\ell} - 2p_{z,\nu}p_{z,\ell}, \end{aligned} \tag{5.2}$$

where the mass of neutrino is set to zero. The $p_{T,\nu}$ is assumed to be missing E_T and $p_{x,\nu}$ and $p_{y,\nu}$ are obtain from known $\phi_{E_T^{miss}}$. Therefore the remaining unknown variable is $p_{z,\nu}$ only and it can be calculated by a quadratic equation.

$$0 = 4p_{T,\ell}^2 \cdot p_{z,\nu}^2 - 4(m_W^2 + 2\overrightarrow{p}_{T,\nu} \cdot \overrightarrow{p}_{T,\ell})]p_{z,\ell} \cdot p_{z,\nu} - (m_W^2 + 2\overrightarrow{p}_{T,\nu} \cdot \overrightarrow{p}_{T,\ell})^2 + 4E_\ell^2 p_{T,\nu}^2, \tag{5.3}$$

where $p_{z,\nu}$ is acquired by solving the root of the quadratic formula. The smaller real component of the solutions is selected since it is close to the real physics in this analysis. Eventually, four-momentum of neutrino is obtainable by collecting E_T^{miss}, $\phi_{E_T^{miss}}$, $p_{z,\ell\nu}$ and massless property.

To avoid overtraining and use all samples, cross-validation is used in BDT. Cross-validation is a method that uses different portions of samples to train and test on different steps. For example, the samples are divided into two portions A and B in this analysis. In the first step, A sample is trained and tested on B sample. And then the model trained with B sample is utilized to test on A sample. Cross-validation is powerful to use better the training and testing samples, which exploits much more information. The input variables and their linear correlation coefficients are evaluated and described in Appendix C. Smaller linear correlation coefficients represent the weak correlation worth inputting them to train. Generally, all input variables are found to be good training inputs into BDT.

After training and testing in the BDT, the BDT score is output as a discriminant between signal and background samples. A suitable BDT score threshold is required to separate signal and background, which will be introduced in Chapter 5.3. As a result, the BDT score threshold cuts down the ratio of signal and background significantly. The BDT score distributions of VBS signal sample, W+jet and $t\bar{t}$ background sample are shown, respectively in Figure 5.5.

5.2.3 Quark/Gluon Tagging Enhance

Quark/Gluon tagging models are studied in the previous Chapter. This section will introduce the strategies of exploiting the Energy Flow tagger trained

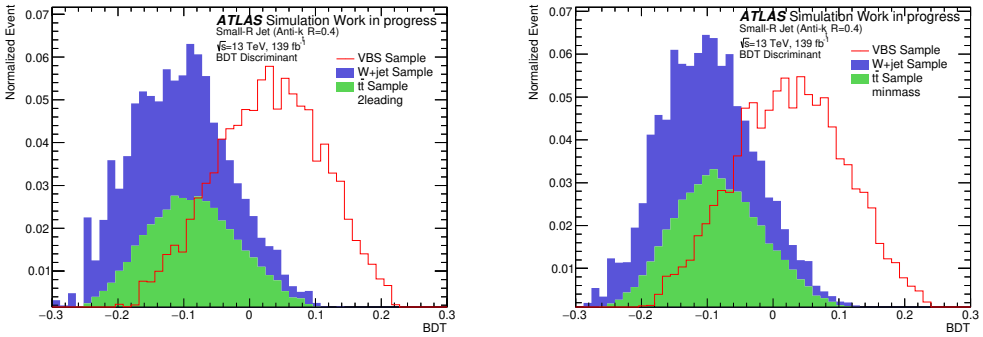

Figure 5.5: The BDT discriminant output of the 2leading (left) and the minmass (right) strategies. From the comparison, the 2leading strategy has a better separation between signal and background samples. Note that the signal and background events are normalized to 1 in both figures.

Table 5.7: A summary of different selection strategies applied in this thesis.

	2leading with BDT Discriminant	minmass with BDT Discriminant
Energy Flow q/g tagging	Neural 2leading	Neural minmass
BDT q/g tagging	BDT 2leading	BDT minmass

in Chapter 4 to improve the signal significance. Moreover, the BDT q/g tagger is also discussed to implement a quick comparison between the neural network model and the BDT model. Including Energy Flow and BDT models, there are four types of strategies, described in Table 5.7. Basically, the signal purity improved outstandingly after preselection and BDT discriminant enhance. Hence, the q/g tagging is based on the selection above as well. Energy Flow and BDT q/g tagger output is utilized as a threshold cut after 2leading and minmass strategies mentioned in the previous section. Because both Energy Flow and BDT q/g taggers are trained at $|\eta| < 2.1$, the output threshold cut is only imposed for the jets that $|\eta|$ <2.1. The ratios of $|\eta| < 2.1$ quarks in the 2leading and minmass strategies are 60165/2137700 $\approx$ 28.1% and 926571/3462640 $\approx$ 26.8%, respectively. The 2leading and minmass strategies followed by the Energy Flow model are referred to as neural 2leading and neural minmass strategies. In addition, 2leading and minmass followed by BDT q/g tagger are referred to as BDT 2leading and BDT minmass strategies. Selecting signal jets pair with the order of Energy Flow or BDT q/g tagger output is considered as another strategy. However, the performance declined a lot compared to 2leading and minmass strategies. This order strategy is shown in Appendix E.

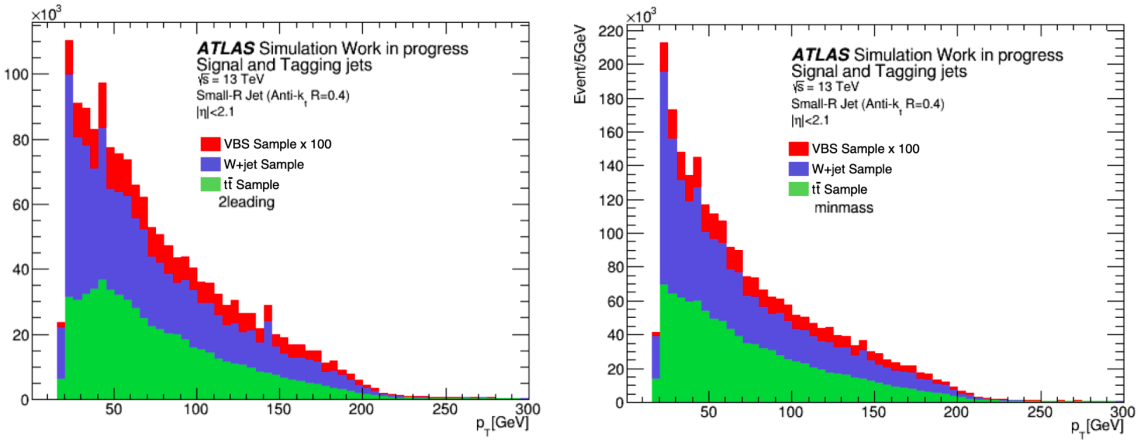

Figure 5.6: The transverse momentum distribution of the 2leading (left) and minmass (right) jets at 139 fb^{-1}. Both signal and tagging jets that $|\eta| < 2.1$ are chosen because they are the targets for implementing q/g taggers.

Although the Energy Flow tagger is shown much better than BDT tagger in previous, their performance is comparable in this analysis since the jets of the MC samples range at lower transverse momentum (approximately smaller than 200 GeV, shown in Figure 5.6). This indicates that it is worthwhile comparing these two different q/g taggers. The outputs of Energy Flow and BDT q/g taggers are shown in Figure 5.7 and Figure 5.8, respectively.

5.3 Application Results

Both the 2eading and minmass strategies with BDT discriminant are well studied in the previous research and improved a lot. Unlike the BDT discriminant that uses high-level input variables, Energy Flow q/g tagging with low-level inputs is expected to improve much more than the previous methods. Therefore, the different strategies introduced above will be compared in this section. Section 5.3.1 introduces a criterion, significance, for the performance judgment. Section 5.3.2 describes the signal-backgrounds separation performance of the BDT discriminant as a baseline analysis of the following q/g tagging models. Section 5.3.3 will show the q/g tagging ability and its improving limit. Section 5.3.4 demonstrates the significance of all strategies and shows the improvement after applying Energy Flow q/g tagging.

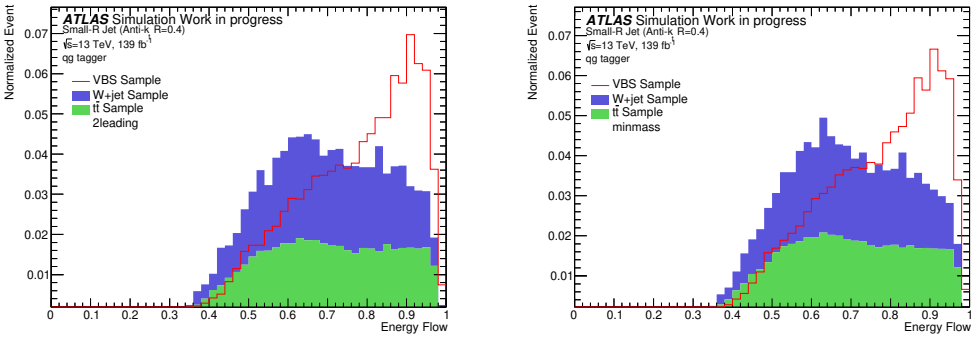

Figure 5.7: The Energy Flow q/g output distribution of 2leading (left) and minmass (right) jets. Both signal and tagging jets that $|\eta| < 2.1$ are chosen because they are the targets for implementing q/g taggers. The energy flow output is the likelihood probability of quark. Therefore, it ranges between 0 to 1.

5.3.1 Discovery significance and Uncertainty

Discovery significance is an estimating value to describe the opportunity that the observation is not from the background fluctuation and is believable. Significance can also be related to the p-value which provides the probability of different observations occurring. The smaller p-value implies that the test hypothesis is false. The relation of significance Z and p-value can be described as Equation 5.4.

$$\mathrm{Z}(\mathrm{N_s}, \mathrm{N_b}, \delta_\mathrm{b}) = \sqrt{2}\mathrm{erf}^{-1}(1 - 2\mathrm{p}), \tag{5.4}$$

where $\mathrm{N_s}$, $\mathrm{N_b}$ and δ_b represents the number of signal events, the number of background events and the uncertainty of background events. erf represents the error function. And the definition of p-value is shown in the equation below.

$$\mathrm{p} = \int_0^\infty \mathrm{db}\ \mathrm{N}(\mathrm{b}; \mathrm{N_b}, \delta_\mathrm{b}\mathrm{N_b}) \sum_{\mathrm{i}=\mathrm{N_s}+\mathrm{N_b}}^{\infty} \mathrm{P}(\mathrm{i}; \mathrm{b}), \tag{5.5}$$

where N and P are Gaussian and Poisson distribution, respectively.
The significance in this study was calculated by **BinomialExpZ** in **RooStats**. To compare different situations, the background systematic uncertainty is set to 10%, 20% and 30%, respectively.

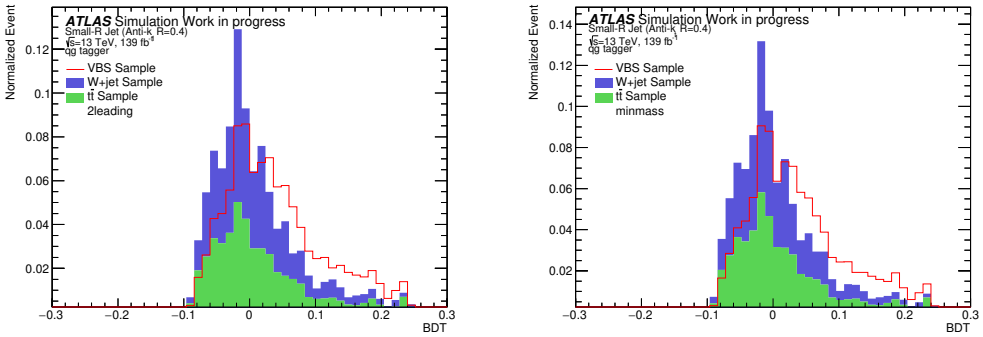

Figure 5.8: The BDT q/g output distribution of 2leading (left) and minmass (right) jets. Both signal and tagging jets that $|\eta| < 2.1$ are chosen because they are the targets for implementing q/g taggers.

5.3.2 Sensitivity of the Baseline Analysis

The 2leading and minmass with BDT discriminant strategies are usually used for the VBS signal searching studies. For example, Reference [12] exploits the minmass with BDT discriminant selection to find the VBS signals. Therefore, it is helpful to calculate the significance of the 2leading and minmass with BDT discriminant strategies as the performance baseline of VBS searching compared to the additional q/g tagging selection discussed in the later sections.

There are 1059 VBS signal events, 297845 W+jet events, and 235519 $t\bar{t}$ events, respectively for 2leading strategy selections.

For minmass strategy, there are 1419 VBS signal events, 454939 W+jet events, and 409285 $t\bar{t}$ events.

The results are shown in the first column in Table 5.8, where the best significances at 10% background systematic uncertainty are 0.78 and 0.73 for the 2leading and minmass strategies, respectively. The significance drops when the systematic uncertainty increases as expected. Comparing the 2leading and minmass with BDT discriminant, it is found that the 2leading strategy is always better than the minmass no matter what the systematic uncertainty is.

5.3.3 q/g tagging Improving Limit

The BDT discriminant is a powerful tool for separating VBS signal events from background events. However, the high-level input variables limit its improvement to another level. Using the q/g tagging model, especially neu-

ral network q/g tagging, provides another way to enhance the signals by inputting the lower-level information in the ATLAS.
To define the improving limit of q/g tagging, it is necessary to investigate how many gluons and quarks are in each event when the 2leading and minmass selections are imposed. Figure 5.9 shows the number of $|\eta| < 2.1$ gluons for 2leading and minmass selections. $|\eta| < 2.1$ gluons are counted since the q/g tagging models in this thesis can only be utilized at $|\eta| < 2.1$ region. There are four gluons at most from two signal small-R jets and two tagging small-R jets.
In order to estimate the limit of the q/g tagging in this study, suppose that quarks and gluons can be separated perfectly. In other words, only zero $|\eta| < 2.1$ gluon events are selected in this situation. Then, applying the BDT discriminant to only none gluon events to estimate the q/g tagging improving limit. The significance estimation of perfect q/g tagger is shown in Table 5.8 when the q/g tagging differentiates quarks and gluons with 100% efficiency.
Applying perfect q/g tagging to the 2leading and minmass with BDT discriminant, the significance improved obviously no matter how large the systematic uncertainty is. This indeed demonstrates the potential of q/g tagging. On average, the 2leading and minmass with BDT discriminant have 56.9% and 76.4% improvement by perfect q/g tagging, respectively.
On the other hand, q/g tagging seems to work better on the minmass strategy than the 2leading strategy. Although the 2leading is better than the minmass with only BDT discriminant, the minmass strategy has more remarkable improvement and larger significance after the perfect q/g tagging is imposed.

5.3.4 q/g tagging Improvement

This section will show the improvement of 1-lepton VBS searching after applying the Energy Flow and BDT q/g models trained in Chapter 4.
The threshold cuts of q/g tagger output and BDT discriminant are calculated separably. In the first step, the BDT discriminant is applied to the 2leading and minmass strategies to obtain the outputs. Secondly, Energy Flow and BDT q/g taggers are applied to the 2leading and minmass strategies to get the output distribution. The output distribution of the BDT discriminant is displayed in Figure 5.5. And the output distribution of Energy Flow and BDT q/g tagging models are shown in Figure 5.7 and Figure 5.8, respectively. Finally, drawing a two dimensional significance plot, which x axis is BDT discriminant output and y axis is q/g tagger output. The two dimensional significance plots of Energy Flow and BDT taggers with 10%, 20% and 30%

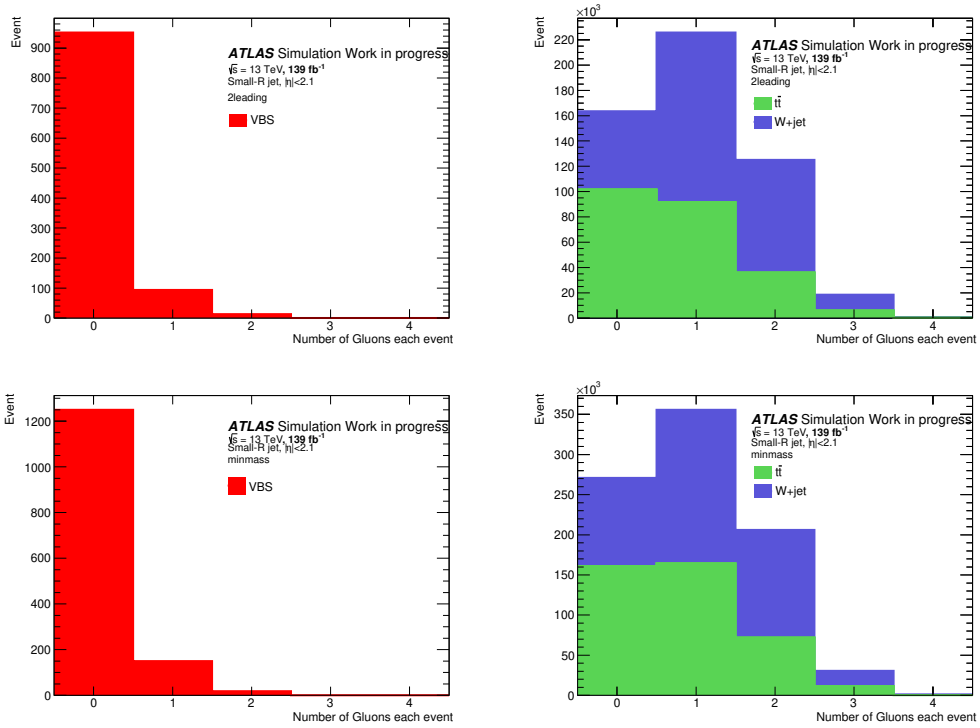

Figure 5.9: The number of gluons each event for the VBS (left) and the background (right) MC samples for 2leading (top) and minmass (bottom) strategies. The VBS signal sample has the most number for non-gluon events, while the W+jet and $t\bar{t}$ samples dominate at 1-gluon event. Therefore, it is useful to separate signals and backgrounds by applying the q/g tagging.

systematic uncertainty are shown in Appendix D. The tremendous significance is obtained by these two dimensional significance plots, summarized in Table 5.8.

Compared to the baseline significance calculated in Section 5.3.2, it is found that all strategies improved after the q/g tagging models were applied. The significances of the neural 2leading and neural minmass strategies are improving by approximately 3.9% and 8.2%, respectively on the average of all systematic uncertainty assumptions. On the other hand, the BDT 2leading and BDT minmass strategies improve by about 0.6% and 5.1%.

Taking 10% background systematic uncertainty for example, the significances of the neural 2leading and neural minmass are 0.81 and 0.77, respectively, which are better than the baseline significance discussed in the previous section. The significances of the BDT 2leading and BDT minmass are 0.78 and 0.73. From the comparison between the significance of Energy Flow and BDT

q/g tagger, the Energy Flow model performs better than BDT q/g tagging model. The ability to input low-level variables gives the neural network q/g tagging models having more space for improvement.
The improvement rate shows that the improving amplitude of the minmass strategy is more significant than the 2leading strategy, which is consistent with the perfect q/g tagging discussion. This indicates that the minmass strategy using q/g tagging has a larger potential for better VBS signal searching.

Table 5.8: A summary of all strategies performance discussed in Section 5.3.2 (baseline), 5.3.3 (perfect q/g tagger) and 5.3.4 (Energy Flow and BDT models). The uncertainty of W+jet and $t\bar{t}$ includes both systematic and statistical uncertainties. All strategies are applied with the BDT discriminant. The significance is calculated with both the statistical and systematic error.

Selection stratgy	Bkg systematic uncertainty	Signal	W+jet	$t\bar{t}$	Significance
2leading	10%	44	42±14	128±38	0.78
	20%	44	42±14	128±38	0.58
	30%	37	30±14	99±43	0.40
minmass	10%	45	57±18	103±39	0.73
	30%	45	57±20	103±43	0.57
	30%	45	57±24	103±49	0.39
2leading with perfect q/g tagger	10%	65	14±6	134±42	1.11
	20%	65	14±7	134±48	0.91
	30%	60	13±7	114±51	0.70
minmass with perfect q/g tagger	10%	71	13±7	168±42	1.19
	20%	53	3±4	99±37	0.97
	30%	53	3±4	99±43	0.76
neural 2leading	10%	48	46±13	148±41	0.81
	20%	48	46±15	148±48	0.60
	30%	34	18±10	89±40	0.42
neural minmass	10%	65	83±22	242±58	0.77
	20%	35	33±15	79±34	0.59
	30%	22	1±1	40±23	0.44
BDT 2leading	10%	75	88±18	366±70	0.79
	20%	44	42±15	128±44	0.58
	30%	37	30±14	99±43	0.40
BDT minmass	10%	44	51±17	83±36	0.73
	20%	31	32±15	59±27	0.58
	30%	28	26±14	50±27	0.44

Chapter 6

Conclusions

The Vector Boson Scattering (VBS) is the highly critical process to probe the properties of the electroweak interaction in the Standard Model. The cross-section of the VBS process is highly sensitive to electroweak symmetry breaking[47] which is still one of the mysteries in physics. Additionally, the VBS process has the same leading order term of Feynman diagrams with Higgs bosons, indicating that studying the VBS process is an indirect way to investigate the Higgs boson properties. With LHC which can operate the particle collisions at center-of-mass energy $\sqrt{s} = 13$ TeV, the physicists can examine the vector boson scattering and electroweak symmetry breaking issues. 1-lepton semileptonic VBS process is studied to validate q/g tagging trained in this thesis. The MC samples are scaled to the integrated luminosity of 139 fb^{-1} corresponding to the Run II data from 2015 to 2018 at the LHC.

1-lepton semileptonic VBS process decays through WZ and WW channels. The W boson decays into lepton and neutrino in the first and second channels, while the other vector boson decays into two quarks. Combing with the tagging quark originating from the colliding protons, there are totally four quarks. Hence, it is essential to utilize the q/g tagging to discriminate the quarks and gluons which are the main backgrounds.

The different properties between quarks and gluons are helpful to separate them. For example, n_{track}, trackwidth and trackC1 are the usually used variables in the Boosted Tree Decision (BDT) q/g tagging in the past. However, the BDT q/g tagger limitation implies that the input variables will lose some information during the calculation from the low-level variables such as η, p_T, etc. Neural network q/g tagging models are considered to solve this kind of issue in this study. Three types of neural network models are compared in

Chapter 4, which are MLP, CNN and Pointwise q/g tagging models. As a result, the Energy Flow and Deep Sets models belonging to Pointwise models have the best performance among all q/g tagging models. The gluon rejection rate of the Energy Flow and Deep Sets models improved approximately 10% than the conventional BDT q/g tagging model at the 80% quark efficiency rate. Moreover, both Energy Flow and Deeps Sets models are applied to different multijet MC to examine the IRC-safe. The result is that the gluon rejection rate differences between trained on the different MC samples with Energy Flow are smaller than Deep Sets. Also, it is found that no matter the models are trained in the Pythia or Herwig MC samples, the separation of quarks and gluons always works better in the Pythia MC samples. This might indicate that the difference between these two MC samples is more prominent, which is a vital issue for future works.

The Energy Flow and BDT q/g tagging models are examined in the 1-lepton semileptonic VBS process. There are two selection strategies for the signal quarks considered for the 1-lepton semileptonic VBS process, the 2leading and minmass selections. The BDT discriminant discussed in Section 5.2.2 enhances both strategies and creates a baseline for the previous research. The improving limit of q/g tagging is investigated. The 2leading and minmass strategies with BDT discriminant with perfect q/g tagging can have the maximum improvement of 56.9% and 76.4% compared to the baseline significance calculated from the non-q/g tagging cases. This is exciting since the q/g tagging still has the space to develop. The improving amplitude of the neural 2leading and neural minmass is approximately 3.9% and 8.2%, respectively, indicating that the q/g tagger seems to perform more potent in the neural minmass strategy. The best performance obtained in this thesis is neural 2leading, which its significance is 0.81. Another thing found is that the backgrounds seriously affect the significance and fewer backgrounds give a better performance. This results in better performance occurring at the strict BDT discriminant and q/g tagger threshold cuts that lower the number of background events.

Significant improvement of quark/gluon separation with the q/g tagging models is examined by the 1-lepton VBS process at the ATLAS in this paper. Since the improvement of q/g tagging models in this study still has space from the improving limit, it is worth researching different q/g tagging models, especially the neural network models. Furthermore, improving the q/g tagging models with additional low-level input variables and tuning the hyperparameter is perspective research for future works.

Appendix A

Input and output of Quark/Gluon Tagging Models

Figure A.1, A.2 and A.3 shows the input variables in the BDT q/g tagging model.

Figure A.4 compares the difference between the testing on Pythia and Herwig MC samples.

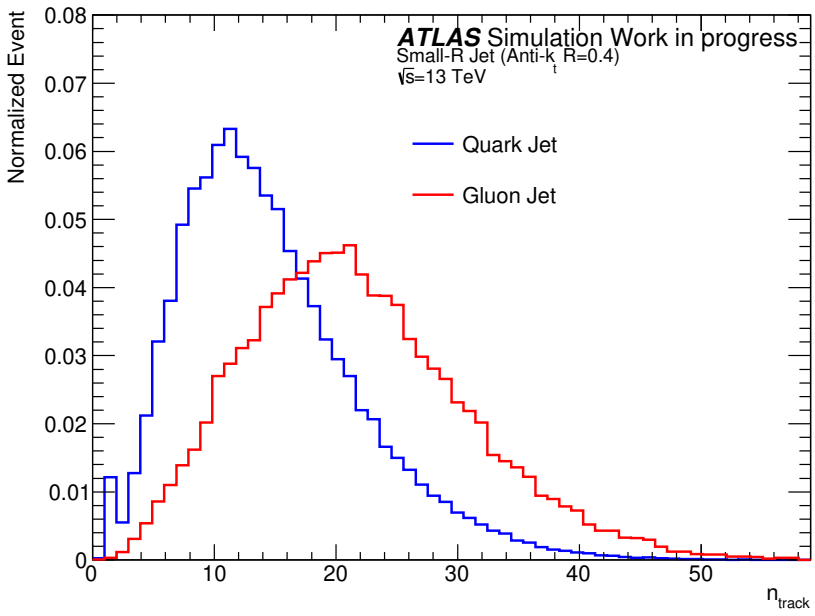

Figure A.1:

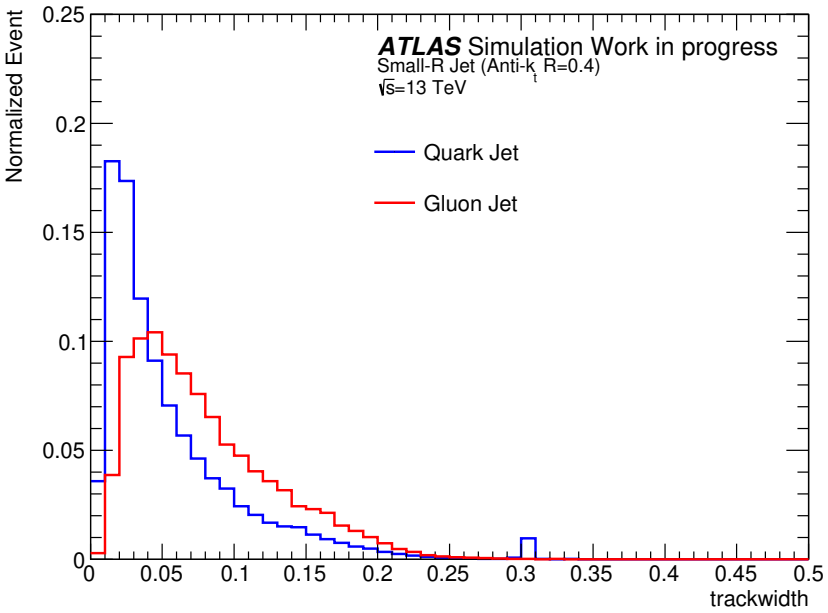

Figure A.2:

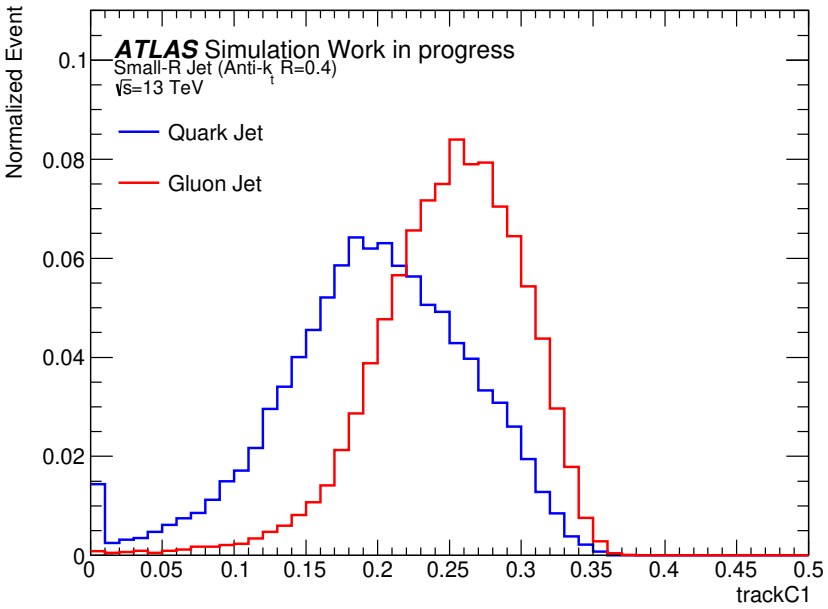

Figure A.3:

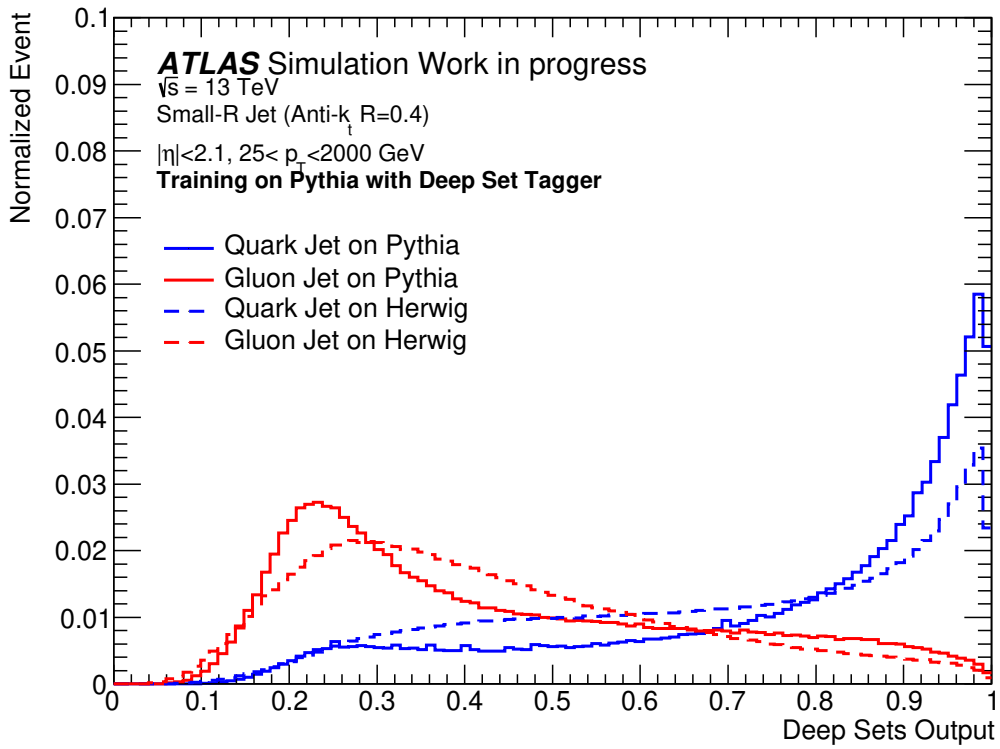
ATLAS Simulation Work in progress
√s = 13 TeV
Small-R Jet (Anti-k_t R=0.4)
|η|<2.1, 25< p_T <2000 GeV
Training on Pythia with Deep Set Tagger
Quark Jet on Pythia
Gluon Jet on Pythia
Quark Jet on Herwig
Gluon Jet on Herwig
Normalized Event
Deep Sets Output
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure A.4:

Appendix B

The distributions for 1-lepton semileptonic preselection

Figure B.1 and B.2 show the distributions of the E_T^{miss} and the lepton p_T. Figure B.3 and B.4 show the distributions of the m_{jjj} of 2leading and minmass strategies.

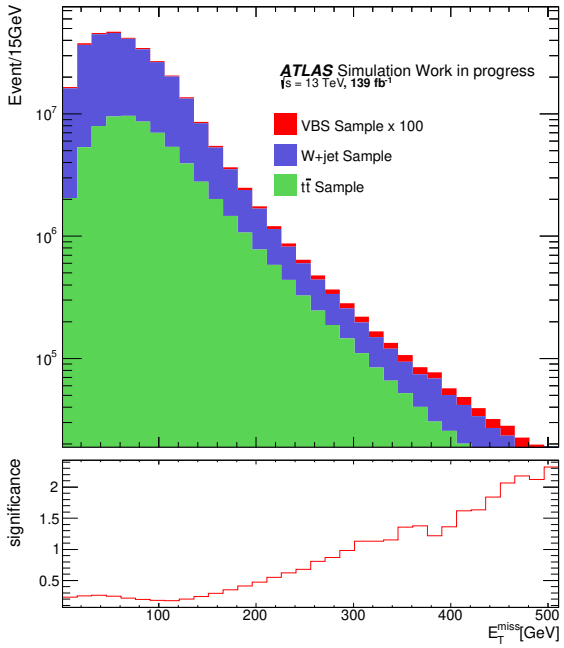

Figure B.1:

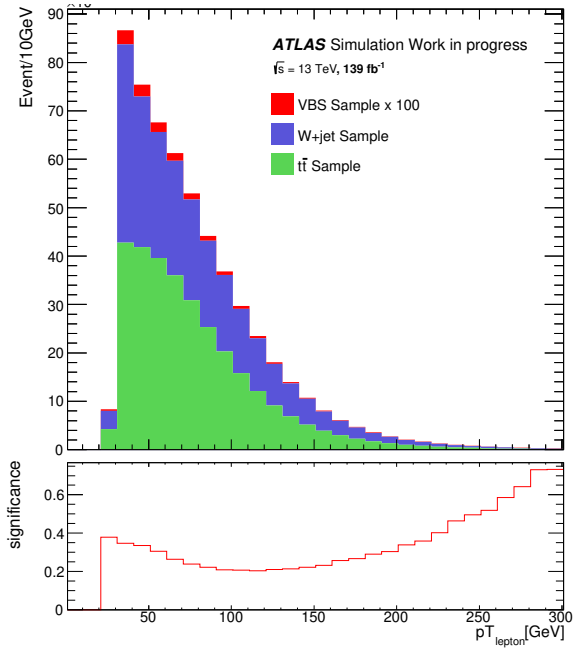

Figure B.2:

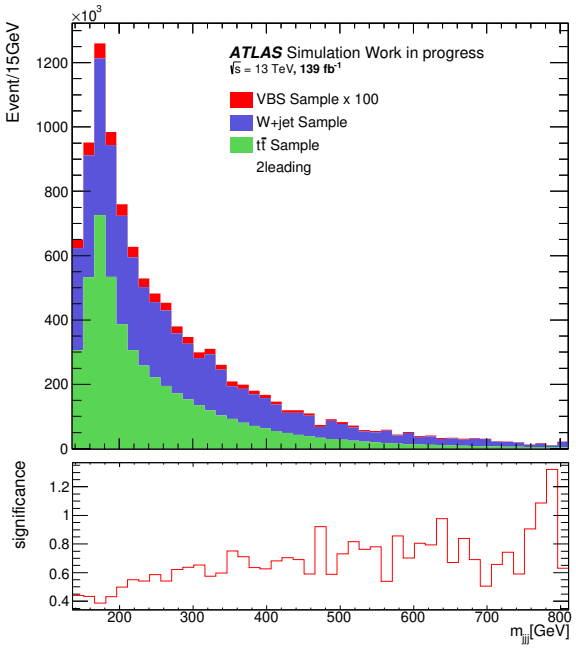

Figure B.3:

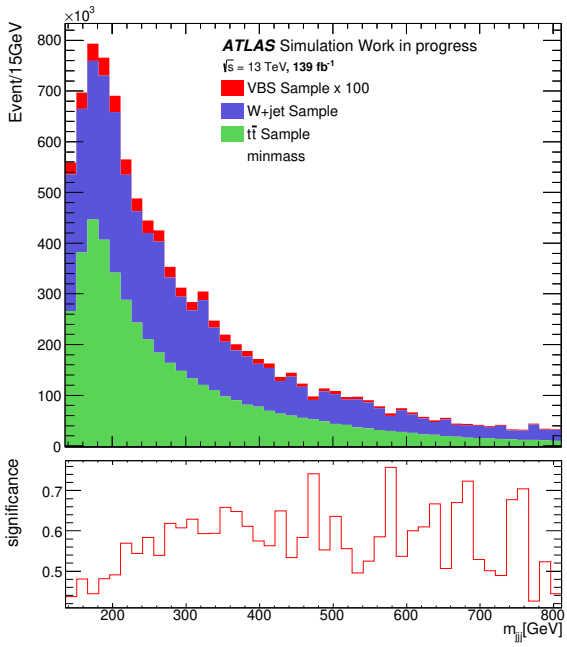
×10^3
Event/15GeV
ATLAS Simulation Work in progress
√s = 13 TeV, 139 fb^-1
VBS Sample x 100
W+jet Sample
tt̄ Sample
minmass
significance
m_jjj [GeV]

Figure B.4:

Appendix C

Input Variable of BDT Discriminant

Figure C.1, C.2, C.3 and C.4 displays the BDT discriminant inputs with the 2leading strategy selection.

Figure C.5, C.6, C.7 and C.8 shows the BDT discriminant inputs with the minmass strategy selection.

C.1 2leading input variables

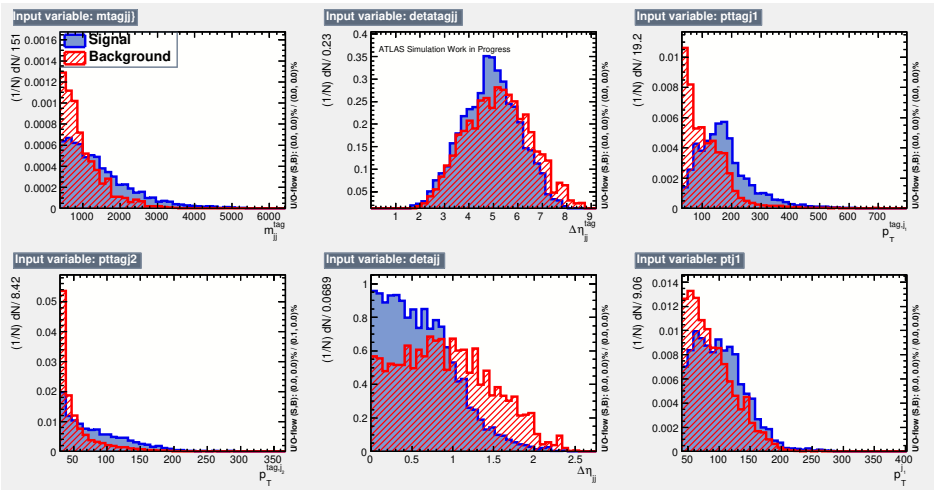

Figure C.1:

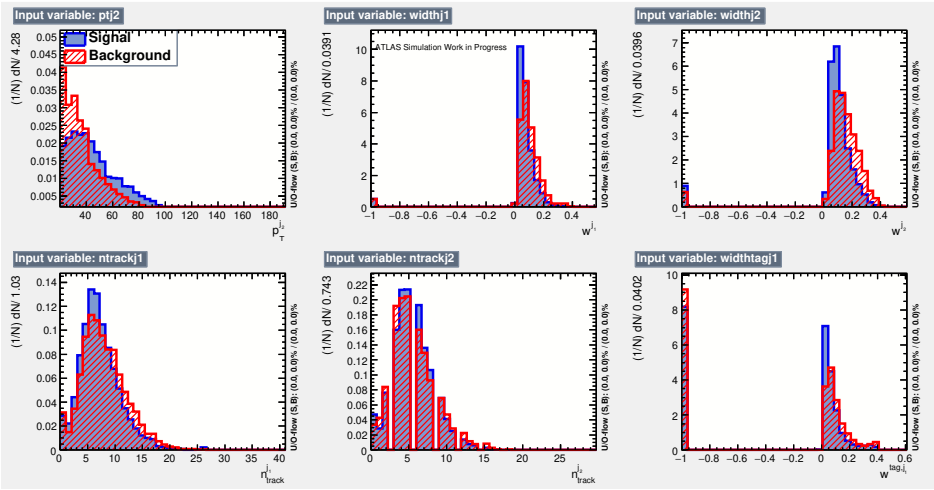

Figure C.2:

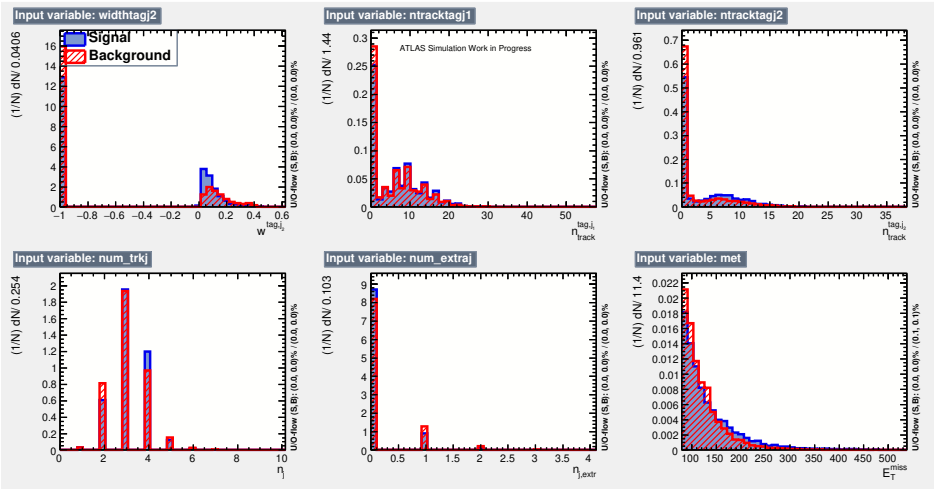

Figure C.3:

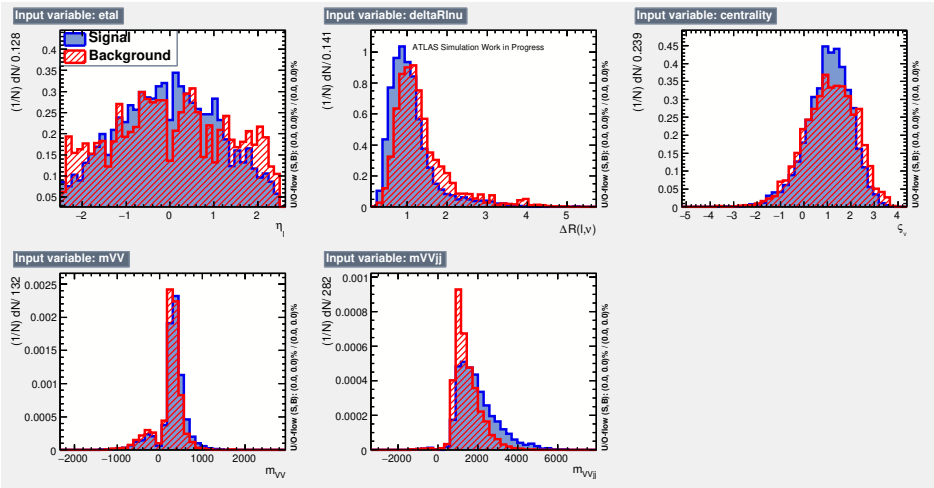
Input variable: etal
Signal
Background
(1/N) dN / 0.128
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Input variable: deltaRlnu
ATLAS Simulation Work in Progress
(1/N) dN / 0.141
ΔR(l,ν)
Input variable: centrality
(1/N) dN / 0.239
Input variable: mVV
(1/N) dN / 132
m_VV
Input variable: mVVjj
(1/N) dN / 282
m_VVjj

Figure C.4:

C.2 minmass input variables

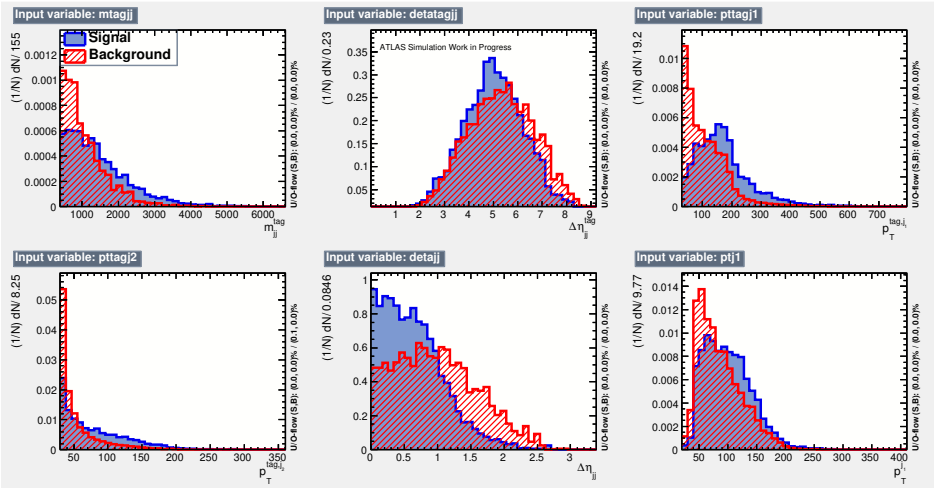

Figure C.5:

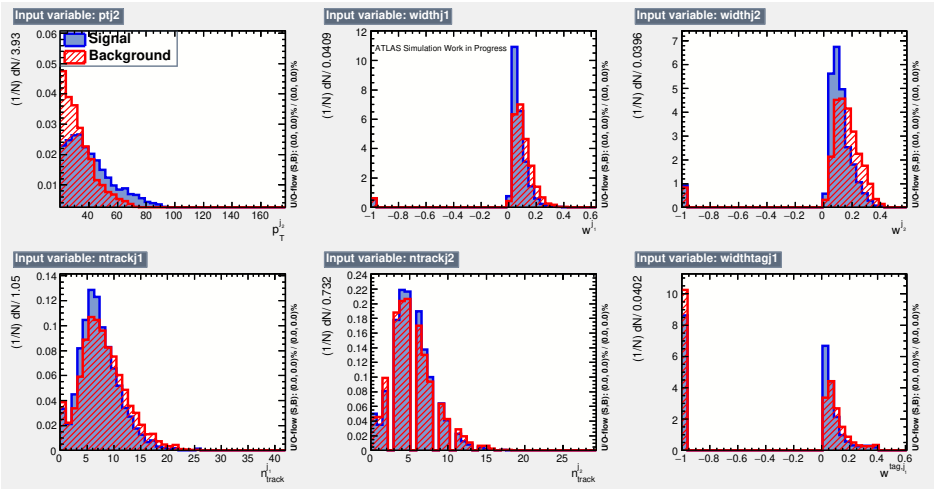

Figure C.6:

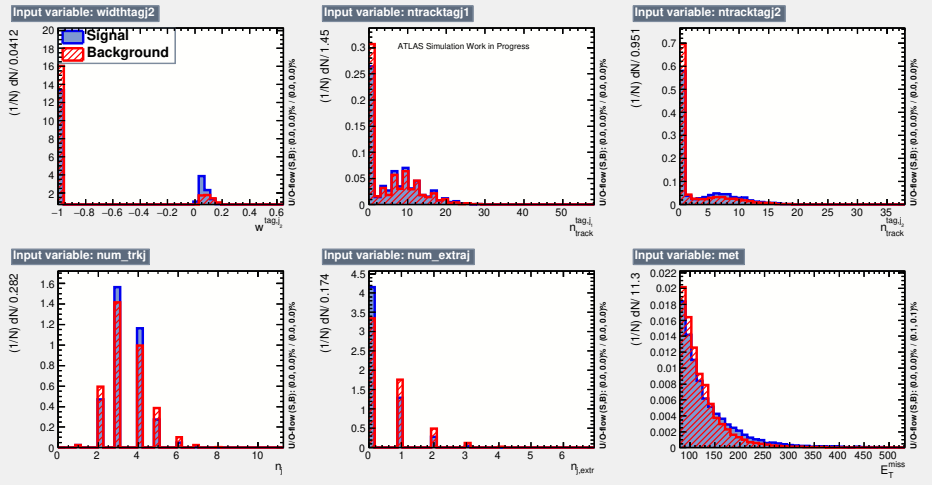

Figure C.7:

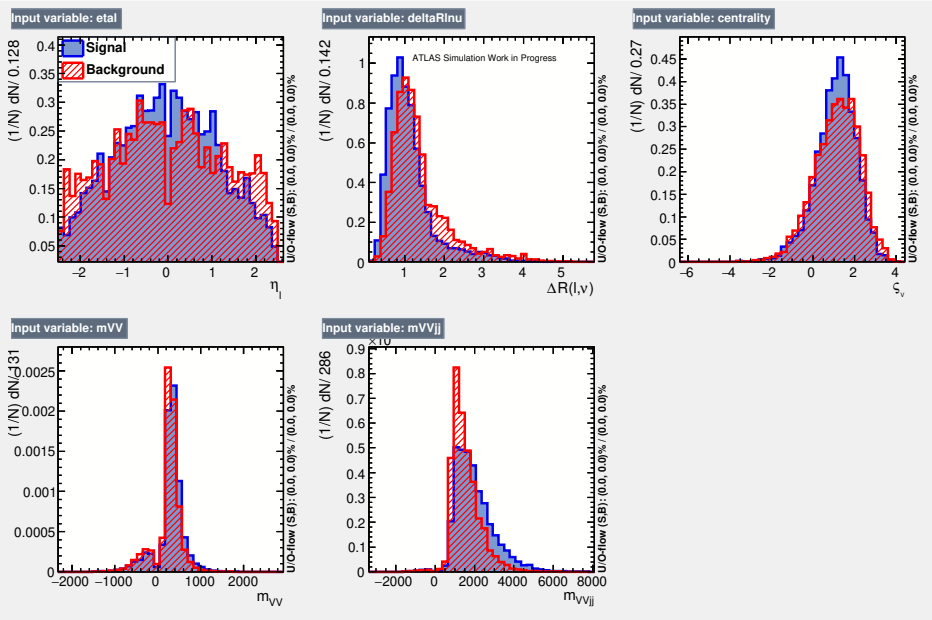

Figure C.8:

Appendix D

Tuning BDT discriminant and q/g tagger

The z axis represents the significance after the thresholds are applied. D.1 and D.2 shows the neural 2leading and BDT 2leading two dimensional significance plots. D.3 and D.4 shows the neural minmass and BDT minmass two dimensional significance plots, respectively. There are some gaps in the 2D plots, such as the place of BDT $= 0.13$ and EnergyFlow $= 0 - 0.8$ in Figure D.1. The reason for these gaps is considered from the deficiency of statistics.

D.1 Neural 2leading

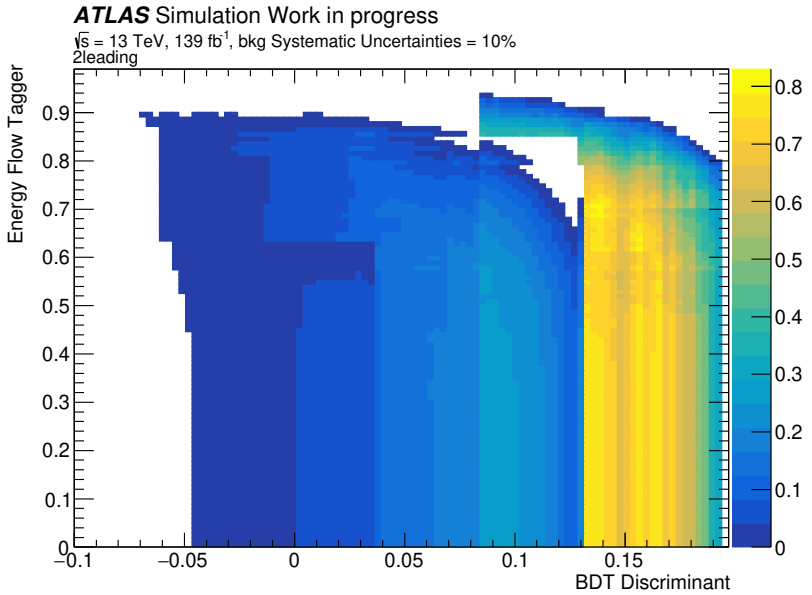

Figure D.1:

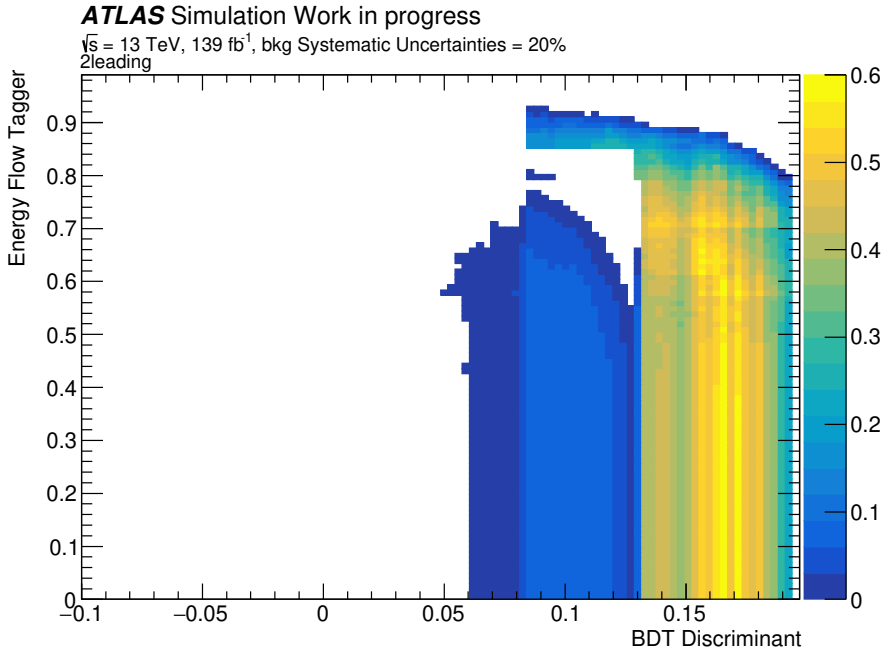

Figure D.2:

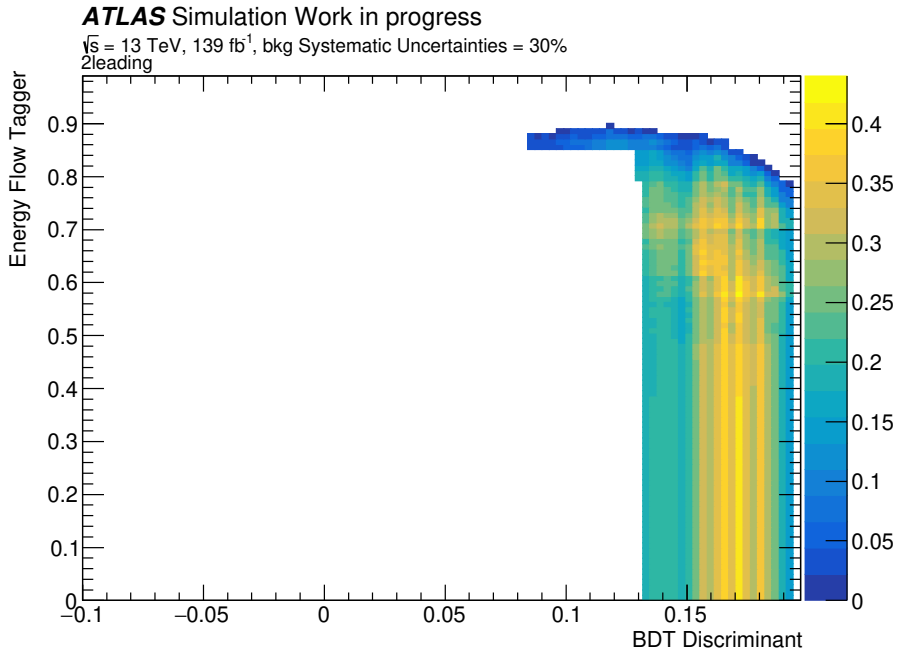
ATLAS Simulation Work in progress
√s = 13 TeV, 139 fb-1, bkg Systematic Uncertainties = 30%
2leading
Energy Flow Tagger
BDT Discriminant
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
−0.1
−0.05
0
0.05
0.1
0.15
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0

Figure D.3:

D.2 BDT 2leading

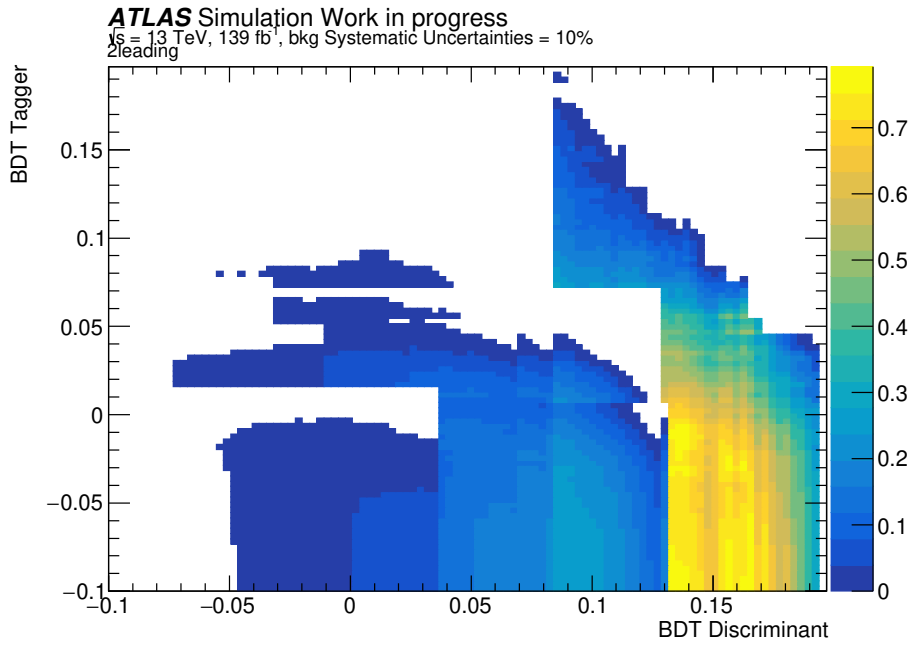

Figure D.4:

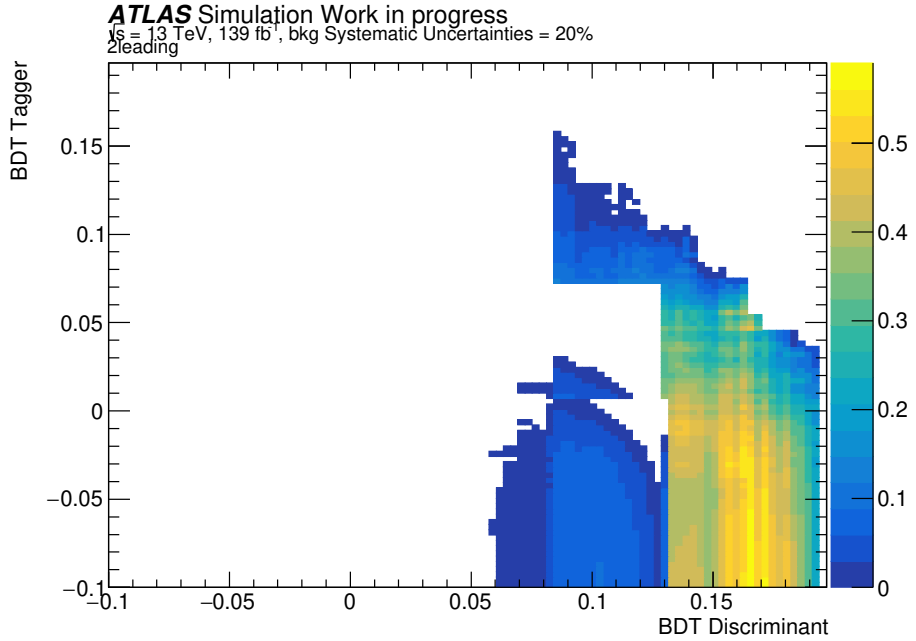

Figure D.5:

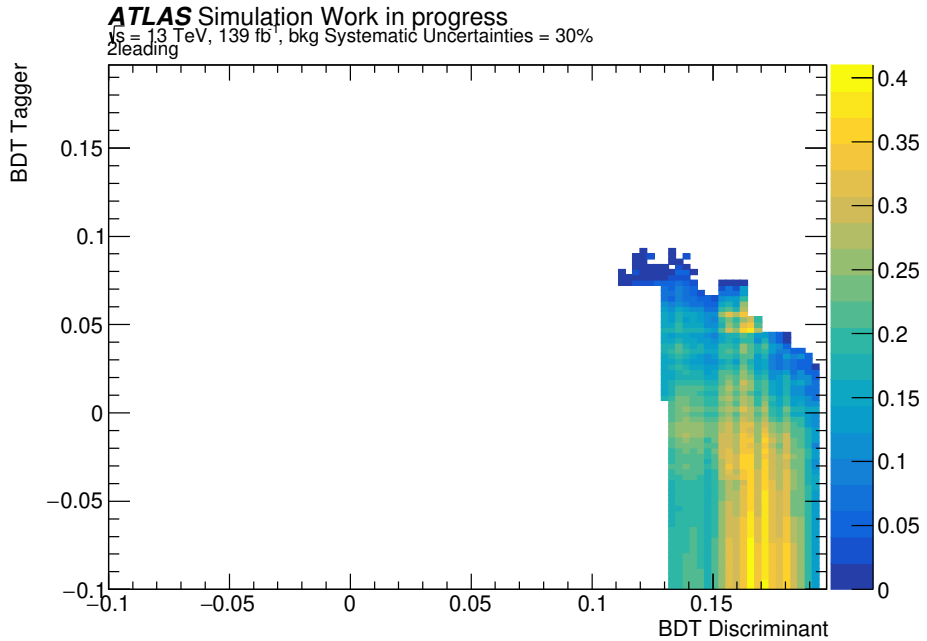

Figure D.6:

D.3 Neural minmass

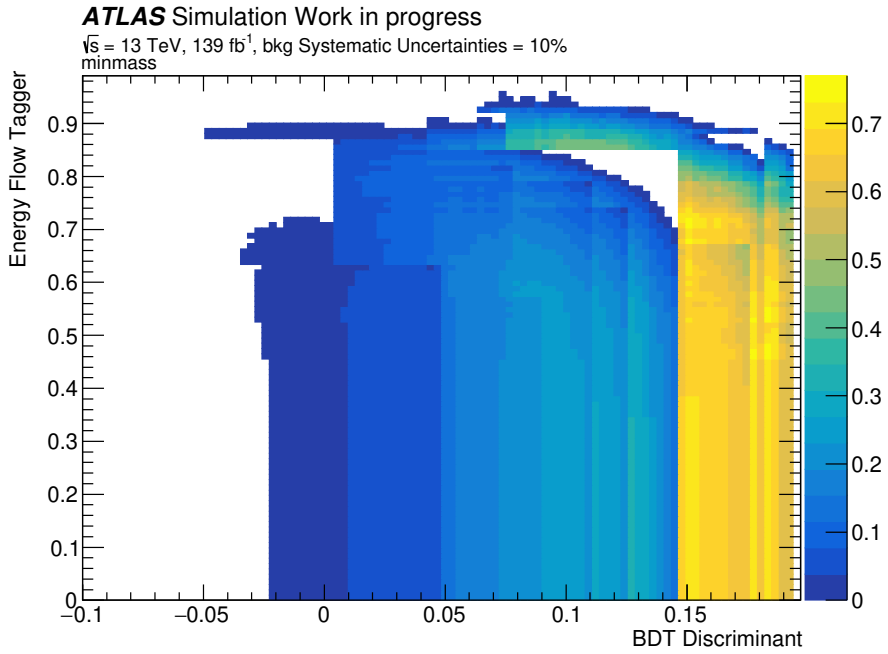

Figure D.7:

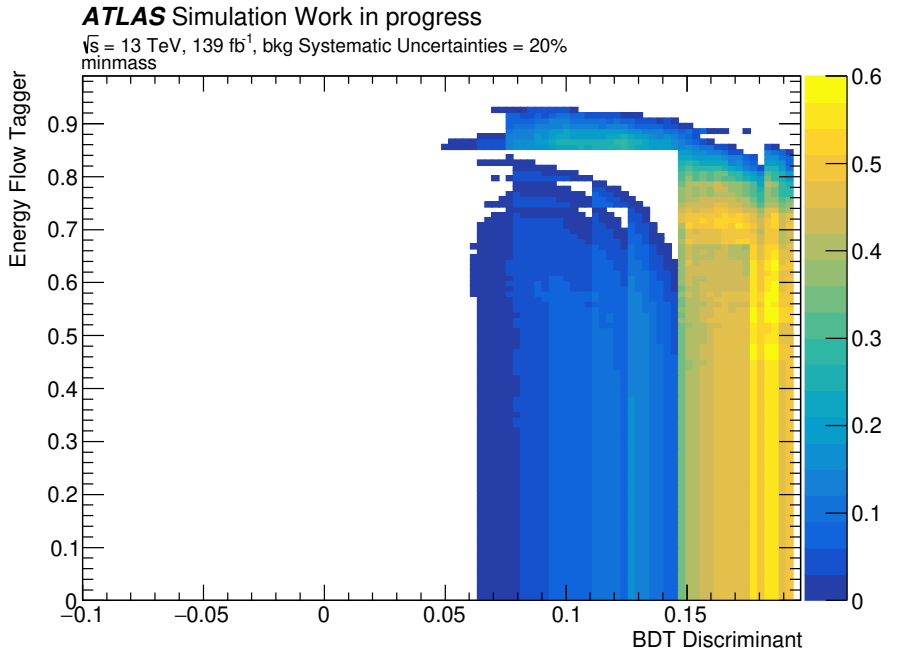

Figure D.8:

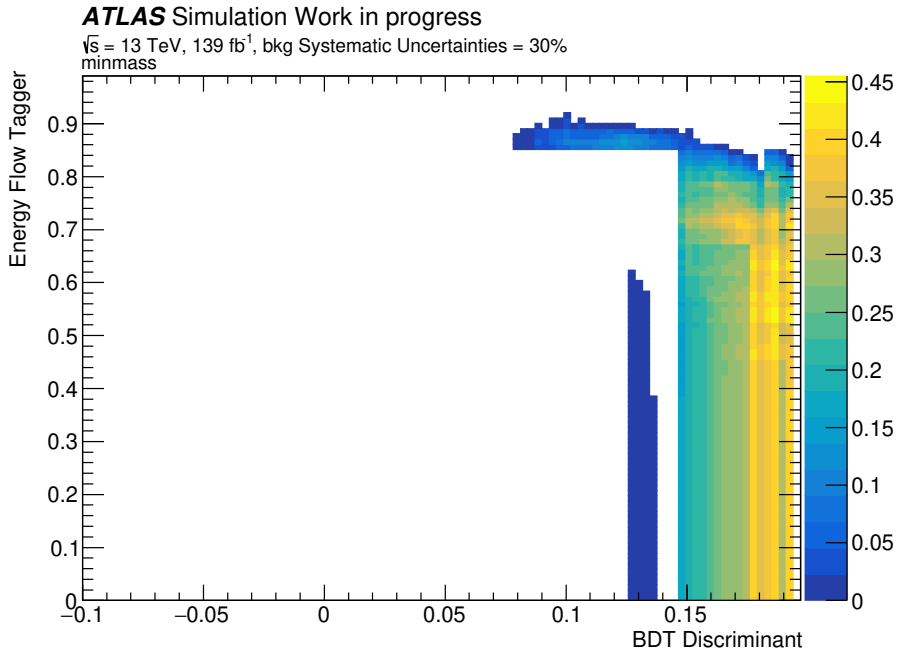

Figure D.9:

D.4 BDT minmass

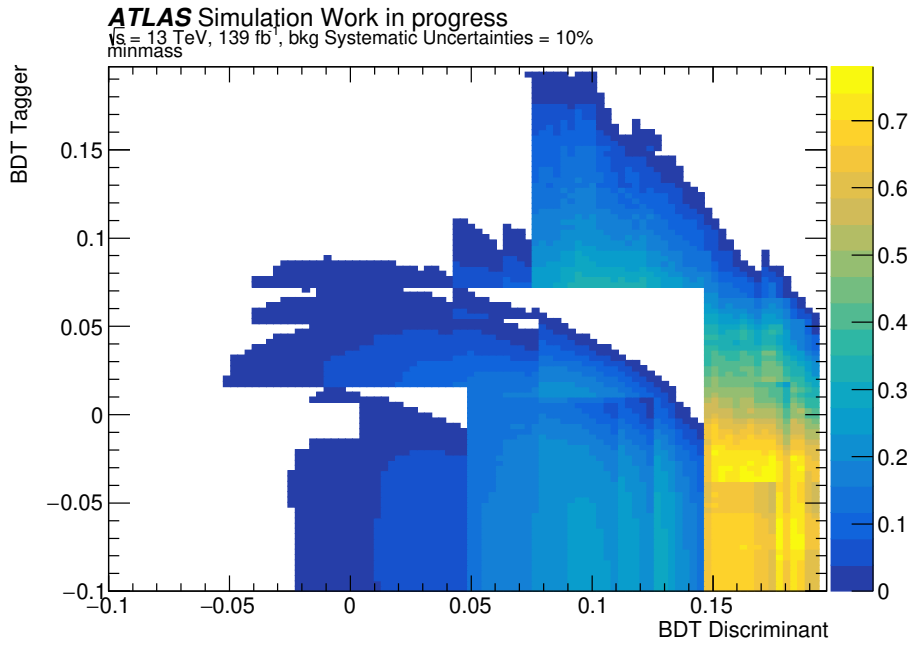

Figure D.10:

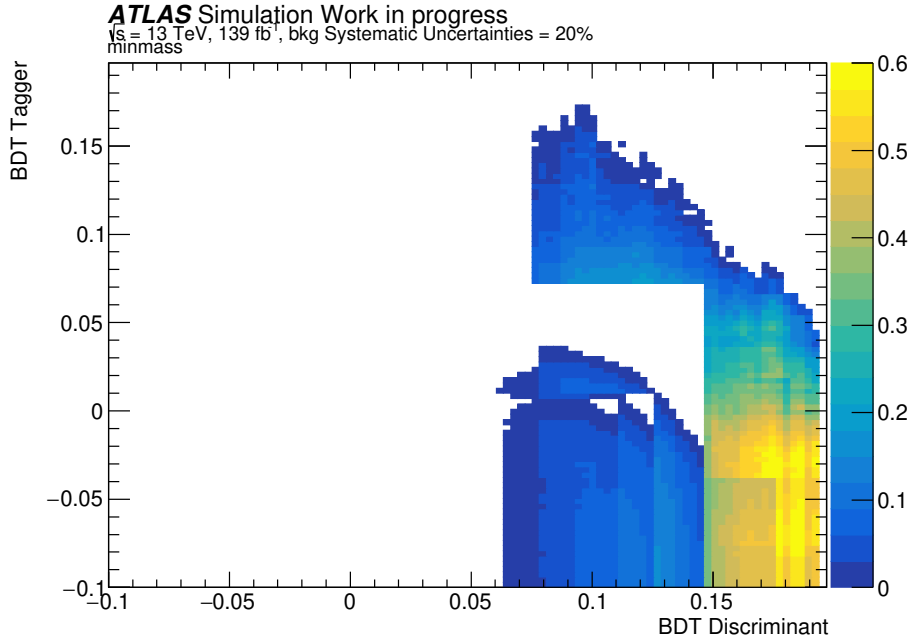

Figure D.11:

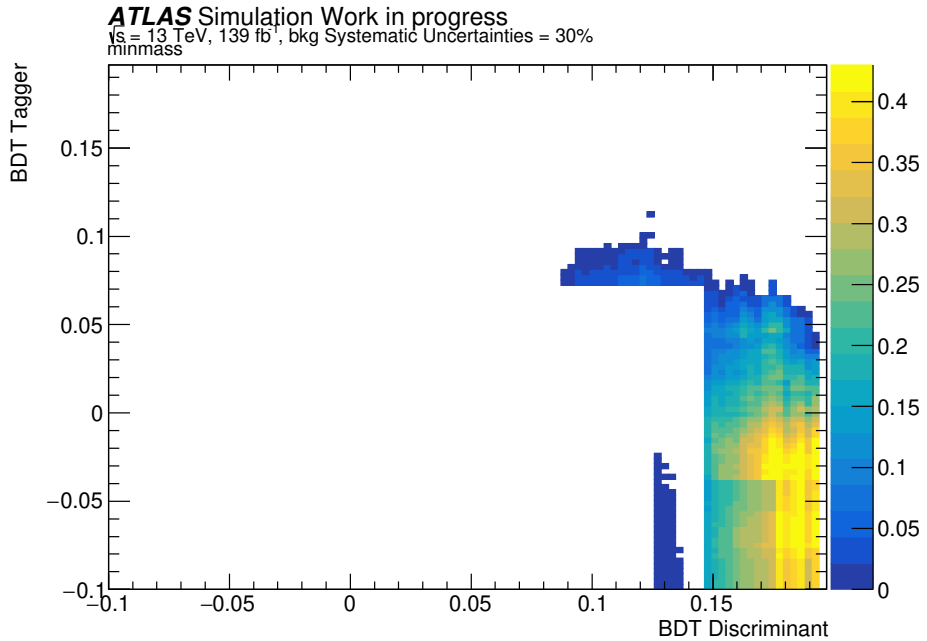

Figure D.12:

Appendix E

Order Stragety

Unlike 2leading and minmass selection strategies, order strategy selects tagging jet and signal jet pairs by the magnitude of Energy Flow tagger output. In the first step, choose the largest four Energy Flow output jets. And then, select the largest invariant mass jet pair as a tagging jet pair, and the remaining two jets are chosen as a signal pair. After picking up the tagging and signal jet pairs, the preselection is performed as Table 5.5.
To obtain the best significance, A threshold of the Energy Flow and BDT discriminant outputs is investigated, same as neural 2leading and neural minmass strategies. The systematic uncertainty is determined with 10%, 20% and 30% to study the different situations.
The significance is shown in Table E.1 and the two dimensional plot of Energy Flow and BDT discriminant output is displayed in Figure E.1, E.2 and E.3 with 10%, 20% and 30% systematic uncertainty. Compared to the significance baseline in Section 5.3.2, it is found that the neural order selection strategy discussed in this Appendix performs worse than the 2leading with only the BDT discriminant strategy. This indicates that using q/g tagger output to select tagging and signal jet pairs is still not a suitable method and shows the worse efficiency compared to 2leading and minmass selection strategies.

Table E.1: A summary of the order strategy significance.

Selection Stratgy	Bkg Systematic Uncertainty	Signal	W+jet	$t\bar{t}$	Significance
order	10%	59	68±16	208±51	0.754
	20%	59	68±16	208±51	0.529
	30%	24	8±6	59±24	0.363

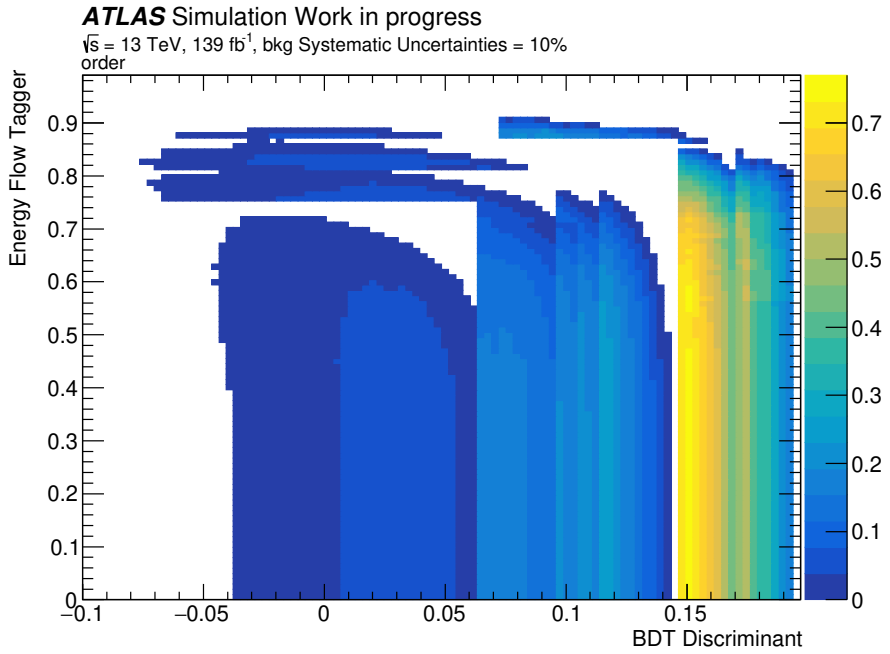

Figure E.1:

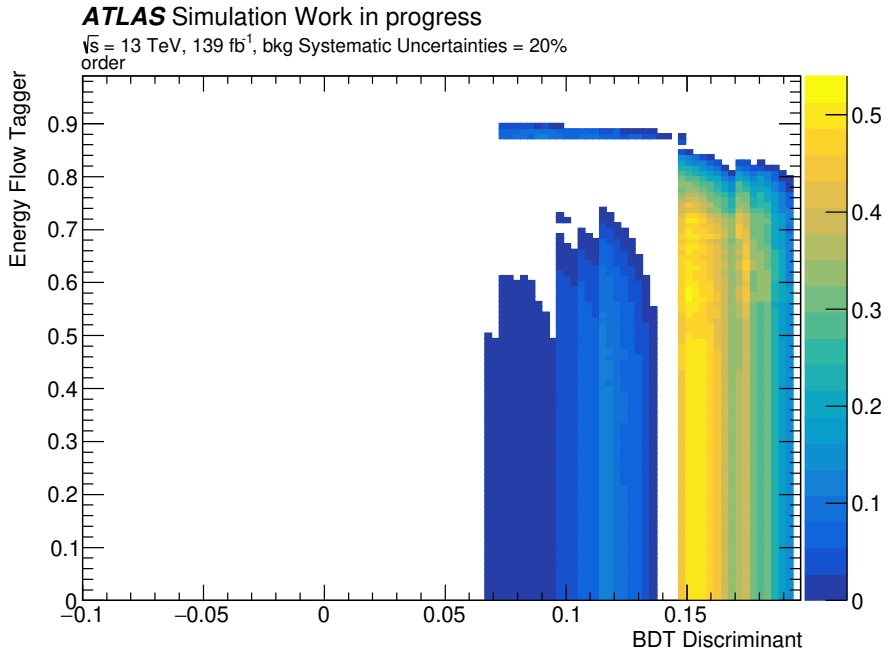

Figure E.2:

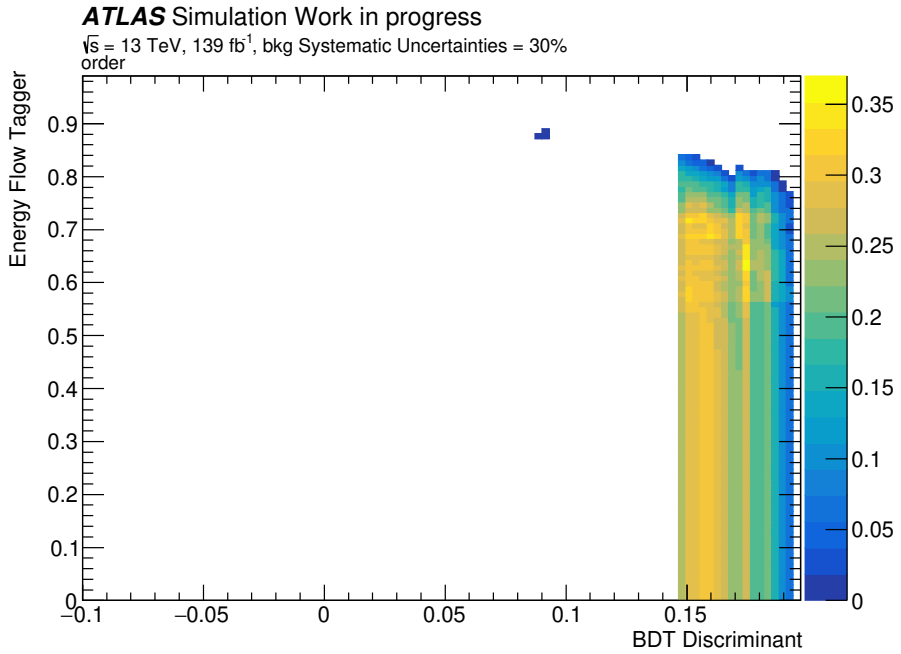

Figure E.3:

Bibliography

[1] J. Zuntz et al. Dark Energy Survey Year 1 Results: Weak Lensing Shape Catalogues. *In:Mon.Not.Roy Astron.Soc.481.1*, pages 1149–1182, 2018. DOI : 10.1093/mnras/sty2219. arXiv: 1708.01533[astro − ph.CO].

[2] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *In:Helvetica Physica Acta 6*, pages 110–127, 1933. arXiv: 1711.01693[astro − ph.IM].

[3] M. Aaboud et al. Measurement of the Higgs boson mass in the H→ZZ*→4l and H→ $\gamma\gamma$ channels with $\sqrt{s}$=13 Tev pp collisions using the ATLAS detector. *Physics Letters B*, 784:345–366, 2018. DOI : 10.1016/j.physletb.2018.07.050. arXiv: 1806.00242[hep − ex].

[4] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Pseudoparticles. *In: Phys. Rev. Lett. 38*, pages 1440–1443, 1977. DOI : 10.1103/PhysRevLett.38.1440.

[5] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Jinst*, 3:S08003, 2008. DOI : 10.1088/1748 − 0221/3/08/S08003.

[6] CMS Collaboration. The CMS experiment at the CERN LHC. *Jinst*, 803:S08004, 2008. DOI : 10.1088/1748 − 0221/3/08/S08004.

[7] Michele Gallinaro et al. Beyond the Standard Model in vector boson scattering signatures. 2020. arXiv: 2005.09889[hep − ph].

[8] Shu Li. Latest VBS measurements in ATLAS. 2020. URL : https : //cds.cern.ch/record/2710588/.

[9] Nadezda Chernyavskaya. Latest results on VBF and VBS processes from CMS experiment. *LHCP2018*, 321, 2018. DOI : 10.22323/1.321.0291.

[10] A Dedes et al. SM EFT effects in Vector-Boson Scattering at the LHC. *Phys. Rev. D*, 104(1):013003, 2021.
DOI : 10.1103/PhysRevD.104.013003. arXiv: 2011.07367[hep − ph].

[11] M. Tanabashi et al. *Review of Particle Physics*, volume 98. 2018.
DOI : 10.1103/PhysRevD.98.030001.

[12] Tatsumi Nitta. Search for the weak vector boson scattering in semileptonic final states in pp collisions at $\sqrt{s} = 13$ Tev with the ATLAS detector. 2020. URL : https : //cds.cern.ch/record/2754187/.

[13] Matteo Cacciari and Gavin P. Salam. The anti-k_t jet clustering algorithm. *JHEP 08*, 2008(04):063, 2008.
DOI : 10.1088/1126 − 6708/2008/04/063. arXiv: 0802.1189[hep − ph].

[14] Steven Schramm. ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects. *Physical Review D*, 182:02113, 2018.
URL : https : //cds.cern.ch/record/2291608.

[15] ATLAS Collaboration. Pile-up subtraction and suppressio for jet in ATLAS. 2013. URL : http : //cds.cern.ch/record/1570994.

[16] G.Parisi G.Altarelli. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298–318, 1977.
DOI : 10.1016/0550 − 3213(77)90384 − 4.

[17] KG Tomiwa. Performance of Jet Vertex Tagger in suppression of pileup jets and E_T^{miss} in ATLAS detector. *J. Phys.: Conf. Ser.*, 802(1):012012, 2017. DOI : 10.1088/1742 − 6596/802/1/012012.

[18] A. Hoecker et al. TMVA - Toolkit for Multivariate Data Analysis. 2007.
DOI : 10.1063/1.4771869. arXiv: physics/0703039[physics.data − an].

[19] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96(7):072002, 2017.
DOI : 10.1103/PhysRevD.96.072002. arXiv: 1703.09665[hep − ex].

[20] B. Abbott et al. Production and integration of the ATLAS Insertable B-Layer. *Journal of Instrumentation*, 13(05):T05008, 2018.
DOI : 10.1088/1748 − 0221/13/05/t05008.
arXiv: 1803.00844[physics.ins − det].

[21] R.J. Apsimon et al. Application of advanced thermal management technologies to the ATLAS SCT barrel module baseboards. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565(2):561–571, 2006. DOI : 10.1016/j.nima.2006.06.058.

[22] Christian Lippmann. Particle identification. *Nucl. Instrum. Methods Phys. Res., A*, 666:148–172, 2012. DOI : 10.1016/j.nima.2011.03.009. arXiv: 1101.3276[hep − ex].

[23] NA Thacker and A Lacey. Tutorial: The Kalman Filter. *Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester*, page 61, 1998.
URL : http : //citeseerx.ist.psu.edu/viewdoc/download?doi = 10.1.1 .408.5594&rep = rep1&type = pdf.

[24] Atlas Collaboration. Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. 2012. URL : https : //cds.cern.ch/record/1435196.

[25] G Piacquadio, K Prokofiev, and A Wildauer. Primary vertex reconstruction in the ATLAS experiment at LHC. *J. Phys.: Conf. Ser.*, 119(3):032033, 2008. DOI : 10.1088/1742 − 6596/119/3/032033.

[26] ATLAS collaboration. Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV. 2015.
URL : https : //cds.cern.ch/record/2110140.

[27] Atlas Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C*, 73(2304), 2013. DOI : 10.1140/epjc/s10052 − 013 − 2304 − 2
arXiv: 1112.6426[hep − ex].

[28] Peter Skands Torbjorn Sjostrand, Stephen Mrenna. PYTHIA 6.4 Physics and Manual. *JHEP*, 2006(05):026, 2006.
DOI : 10.1088/1126 − 6708/2006/05/026. arXiv: hep − ph/0603175.

[29] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 2015(4):1–148, 2015. DOI : 10.1007/JHEP04(2015)040.
arXiv: 1410.8849[hep − ph].

[30] Atlas Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. 2014.
URL : https : //cds.cern.ch/record/1966419?ln = ja.

[31] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 389(1-2):81–86, 1997.
DOI : 10.1016/S0168 − 9002(97)00048 − X.

[32] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep Sparse Rectifier Neural Networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 15:315–323, 2011.
URL : https : //proceedings.mlr.press/v15/glorot11a.html.

[33] Y. Bengio I. Goodfellow and A. Courville. *DeepLearning*. 2016.
URL : https : //www.deeplearningbook.org.

[34] ATLAS Collaboration. Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector. 2017.
URL : https : //cds.cern.ch/record/2275641.

[35] Manzil Zaheer et al. Deep Sets. 2017. arXiv: 1703.06114[cs.LG].

[36] Martín Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. 2016. arXiv: 1603.04467[cs.DC].

[37] J. Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):1–8, 2016. DOI : 10.1140/epjc/s10052 − 016 − 4018 − 8.
arXiv: 1512.01178[hep − ph].

[38] Stefan Gieseke Simon Platzer. Dipole Showers and Automated NLO Matching in Herwig++. *Eur. Phys. J. C*, 72(11):1–19, 2012.
DOI : 10.1140/epjc/s10052 − 012 − 2187 − 7.
arXiv: 1109.6256[hep − ph].

[39] Tim Stelzer Fabio Maltoni. MadEvent: Automatic event generation with MadGraph. *JHEP*, 2003(02):027, 2003.
DOI : 10.1088/1126 − 6708/2003/02/027. arXiv hep − ph/0208156.

[40] David J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1-2):152–155, 2001. DOI : 10.1016/S0168 − 9002(01)00089 − 4.

[41] Richard D. Ball et al. Parton distributions with LHC data. *Nuclear Physics B*, 867(2):244–289, 2013.
DOI : 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303[hep − ph].

[42] T Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 2009(02):007, 2009.
DOI : 10.1088/1126 − 6708/2009/02/007. arXiv: 0811.4622[hep − ph].

[43] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. A Positive-Weight Next-to-Leading-Order Monte Carlo for Heavy Flavour Hadroproduction. *JHEP*, 2007(09):126, 2007.
DOI : 10.1088/1126 − 6708/2007/09/126. arXiv: 0707.3088[hep − ph].

[44] Zachary Marshall et al. Simulation of Pile-up in the ATLAS Experiment. *J. Phys.: Conf. Ser.*, 513(2):022024, 2014.
DOI : 10.1088/1742 − 6596/513/2/022024.

[45] 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. *Eur. Phys. J. C*, 79(639), 2019.
DOI : 10.1140/epjc/s10052 − 019 − 7140 − 6
arXiv: 1902.04655[physics.ins − det].

[46] Atlas Collaboration. Electron and photon energy calibration with the ATLAS detector using the 2015-2016 LHC proton-proton collision data. *JINST*, 14(P03017), 2019.
DOI : 10.1088/1748 − 0221/14/03/P03017
arXiv: 1812.03848[hep − ex].

[47] Kaustubh Agashe and Roberto Contino. The minimal composite Higgs model and electroweak precision tests. *Nuclear physics B*, 742(1-3):59–85, 2006. DOI : 10.1016/j.nuclphysb.2006.02.011. arXiv: hep − ph/0510164.