

# Construction of $b$ -jet origin neural network classifier in top quark pair production with additional $b$ -jets and electron-muon pair in the final state in proton-proton collision at $\sqrt{s} = 13$ TeV by the ATLAS detector

---

---

Student Name:  
Ning Ni

Name of University:  
Stony Brook University

University Supervisor:  
Prof. Dmitri Tsybychev

## Summary

This thesis uses TensorFlow framework to build two  $b$ -jet origin classifiers specifically for  $t\bar{t}$  production in association with additional  $b$ -jets and electron and a muon of the opposite charges in the final state in  $pp$  collisions at the LHC at a center-of-mass energy of 13 TeV using the ATLAS Run2 data corresponding to an integrated luminosity of 140  $fb^{-1}$ . After selecting events to suppress irrelevant background, Monte Carlo simulation of discriminative variables from Powheg + Pythia8 MC generated  $t\bar{t}$  events are used to train classifiers. We develop two classifiers separately, one for exactly three  $b$ -jets and one for at least 4  $b$ -jets in the final state. Besides loss and accuracy, we estimate the two neural networks by efficiencies of four categories of  $b$ -jets: leading and sub-leading  $p_T$   $b$ -jets from top quark quarks decays, and leading and sub-leading (if present)  $p_T$  additional  $b$ -jets.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	ATLAS detector . . . . .	5
1.2	$t\bar{t}$ production in proton-proton collisions . . . . .	5
<b>2</b>	<b>Monte Carlo Simulation and Event Selection</b>	<b>7</b>
2.1	Dataset and Monte Carlo Simulation . . . . .	7
2.2	Event Selection . . . . .	7
<b>3</b>	<b>Reconstructed discriminating Features</b>	<b>8</b>
<b>4</b>	<b>Neural Network based <math>b</math>-jets origin classifier</b>	<b>12</b>
4.1	Setup . . . . .	12
4.2	Transverse Momentum Bias . . . . .	15
4.3	Performance Comparison with BDT model . . . . .	19
<b>5</b>	<b>Domain-Adversarial Neural Network (DANN)</b>	<b>21</b>
5.1	Classifier Independent of Input MC Model . . . . .	21
5.2	Repress Transverse Momentum Bias . . . . .	26
<b>6</b>	<b>Conclusions</b>	<b>29</b>

## List of Figures

1	Mechanical structure of the ATLAS detector. . . . .	5
2	Feynman diagram of Higgs boson in final state of top quark production channel $t\bar{t} + H$ . . . . .	6
3	Feynman diagram of top quark pair production with additional $b$ -jet and electron-muon pair in the final state. . . . .	6
4	MC histogram modeling for the leading $p_T$ of $b$ -jets (left) and the number of $b$ -jets (right) in each event. . . . .	7
5	Radial distance between $b$ -jet and electron (left) or muon (right). . . . .	9
6	Radial distance between $b$ -jet and the lepton with leading $p_T$ (left) or sub-leading $p_T$ (right). . . . .	9
7	The close (left) or far (right) radial distance between $b$ -jet and leptons. . . . .	10
8	The closest (left) or farthest (right) radial distance between $b$ -jets. . . . .	10
9	Invariant mass between $b$ -jet and electron (left) or muon (right). . . . .	11
10	Invariant mass between $b$ -jet and the lepton with leading $p_T$ (left) or sub-leading $p_T$ (right). . . . .	11
11	Invariant mass between $b$ -jet and the lepton with close radial distance (left) or far radial distance (right). . . . .	12
12	The smallest (left) or the largest (right) invariant mass between $b$ -jets. . . . .	12
13	Training history and the final test AUC score (left). Neural Network score distribution (right). . . . .	15

14	Correlation matrix of input features and transverse momentum. . . . .	15
15	Distribution of transverse momentum of leading $b$ -jet from top quark in 3j3b (left) and 4j4b (right). . . . .	16
16	Distribution of transverse momentum of leading $b$ -jet from top quark in 3j3b (left) and 4j4b (right). . . . .	17
17	Distribution of transverse momentum of sub-leading $b$ -jet from top quark in 3j3b (left) and 4j4b (right). . . . .	17
18	Distribution of transverse momentum of leading additional $b$ -jet in 3j3b (left) and 4j4b (right). . . . .	18
19	Distribution of transverse momentum of sub-leading additional $b$ -jet in 4j4b. . . . .	19
20	Efficiency of four identities of $b$ -jets and both $b$ -jets from top quark in 3j3b (left) and 4j4b (right). . . . .	19
21	Efficiency of four identities of $b$ -jets and both $b$ -jets from top quark in 3j3b by NN (left) and BDT (right) with all features inputted. . . . .	20
22	Efficiency of four identities of $b$ -jets and both $b$ -jets from top quark in 4j4b by NN (left) and BDT (right) with all features inputted. . . . .	20
23	Efficiency of four identities of $b$ -jets and both $b$ -jets from top quark in 3j3b by NN (left) and BDT (right) with only radial distance features inputted. .	21
24	Efficiency of four identities of $b$ -jets and both $b$ -jets from top quark in 4j4b by NN (left) and BDT (right) with only radial distance features inputted. .	21
25	Domain-adversarial network as an alternative to reduce classification bias, adapted from [5]. . . . .	22
26	(Sherpa 2.2.10) Top: Signal and background prediction distribution by class label and domain label when $\lambda = 0$ (left) or $\lambda = 2$ (right). Bottom: The ratio of class label prediction to domain label prediction in each bin. .	23
27	(Sherpa 2.2.10) Top: Signal and background prediction distribution by class label and domain label when $\lambda = 6$ (left) or $\lambda = 50$ (right). Bottom: The ratio of class label prediction to domain label prediction in each bin. .	23
28	(Sherpa 2.2.10) The absolute difference value between class/domain ratio and 1, in dependence of the scale factor. . . . .	24
29	(aMC@NLO+Pythia8) Top: Signal and background prediction distribution by class label and domain label when $\lambda = 0$ (left) or $\lambda = 2$ (right). Bottom: The ratio of class label prediction to domain label prediction in each bin. .	25
30	(aMC@NLO+Pythia8) Top: Signal and background prediction distribution by class label and domain label when $\lambda = 6$ (left) or $\lambda = 50$ (right). Bottom: The ratio of class label prediction to domain label prediction in each bin. .	25
31	(aMC@NLO+Pythia8) The absolute difference value between class/domain ratio and 1, in dependence of the scale factor. . . . .	26
32	(DANN) Signal and background prediction distribution by class label and domain label when $\lambda = 0$ (left) or $\lambda = 0.5$ (right). . . . .	27
33	$p_T$ bias vs the scale factor value $\lambda$ from 0 to 15 for four categories of $b$ -jets: leading (1st FT) and sub-leading (2nd FT) $p_T$ $b$ -jets from top quark quarks decays, and leading (1st Add) and sub-leading (2nd Add) $p_T$ additional $b$ -jets. . . . .	28

34	$p_T$ bias vs the scale factor value $\lambda$ from 0 to 4 (left) and from 0.5 to 0.7 (right) for four categories. . . . .	29
----	--	----

## List of Tables

1	Use Heavy Flavor Filter Flag (HFFF) to avoid overlap of inclusive sample. . . . .	8
2	Reweighted event yield from each source after selection. . . . .	8
3	Setup of hyper-parameters. . . . .	13
4	Tuning with the layer number. . . . .	13
5	Decreasing node number by two times. . . . .	14
6	Tuning with the node number. . . . .	14

# <sup>1</sup> 1 Introduction

<sup>2</sup> This section introduces ATLAS detector at CERN and  $t\bar{t}$  pair production in association  
<sup>3</sup> with additional  $b$ -jets at the Large Hadron Collider (LHC).

## <sup>4</sup> 1.1 ATLAS detector

<sup>5</sup> The ATLAS (A Toroidal LHC ApparatuS) detector, as shown in Figure 1, is one of the  
<sup>6</sup> two general-purpose detectors at the LHC. It is a key tool in high energy physics by  
<sup>7</sup> providing comprehensive data on the vast amount of particles created in high energy col-  
<sup>8</sup> lisions. Its inner sensors and instruments are designed to identify particles and track their  
<sup>9</sup> observables with great precision, which enables physicists to reconstruct and analyze the  
<sup>10</sup> complex events resulting from the collisions at the smallest scale of nature. ATLAS plays  
<sup>11</sup> a central role in the LHC's research by allowing physicists to test theoretical predictions in  
<sup>12</sup> high energy physics, including the existence of predicted but undiscovered particles, and  
<sup>13</sup> to explore the fundamental interactions and materials that shapes our universe. Through  
<sup>14</sup> this extensive data collection and analysis, ATLAS contributes significantly to our under-  
<sup>15</sup> standing of the Standard Model and the search for new physics beyond it.

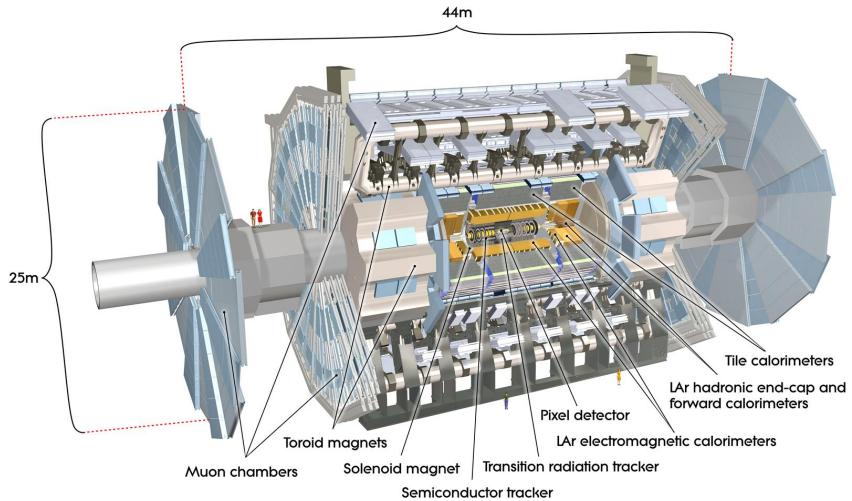


Figure 1: Mechanical structure of the ATLAS detector.

## <sup>16</sup> 1.2 $t\bar{t}$ production in proton-proton collisions

<sup>17</sup> This research focuses on the  $t\bar{t}$  production channel in proton-proton collisions, because  
<sup>18</sup> it is an important channel for various studies. After the discovery of the Higgs boson  
<sup>19</sup> in 2012 by the ATLAS and CMS Collaborations, many measurements of its properties  
<sup>20</sup> were performed [2]. The top quark is the heaviest elementary particle in the Standard  
<sup>21</sup> Model (SM) and has a strong connection to the SM Higgs boson. A probe of fundamental  
<sup>22</sup> interest to further explore the nature of the Higgs boson is its coupling to the top quark

<sup>23</sup> production,  $t\bar{t} + H$  (see Figure 2). The top quark production is also relevant to Higgs  
<sup>24</sup> boson by the rare process of four top quarks,  $t\bar{t}t\bar{t}$  [3].

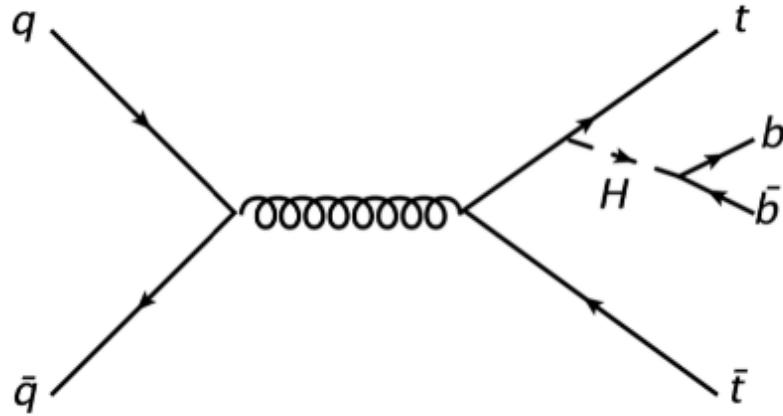


Figure 2: Feynman diagram of Higgs boson in final state of top quark production channel  $t\bar{t} + H$ .

<sup>25</sup> Besides testing predictions by SM, an improvement on our understanding on  $t\bar{t}$  produc-  
<sup>26</sup> tion could contribute to searches for physics beyond the standard model (BSM) such as  
<sup>27</sup> Supersymmetry (SUSY).  $t\bar{t}$  is the dominant background for many of the studies. Among  
<sup>28</sup> all the channels of  $t\bar{t}$ +jets, those generated heavy-flavor quark ( $t\bar{t} + b$ -jets + emu-OS) are  
<sup>29</sup> of special interest, since these suffer from poor precision in the theoretical predictions. By  
<sup>30</sup> construction of this  $b$ -jets origin classifier, we are able to improve the understanding and  
<sup>31</sup> measurements of differential cross-section of top quark production. This classifier aims to  
<sup>32</sup> identify  $b$ -jets origin as from top quarks decay or from initial state radiated gluons (see  
<sup>33</sup> Figure 3).

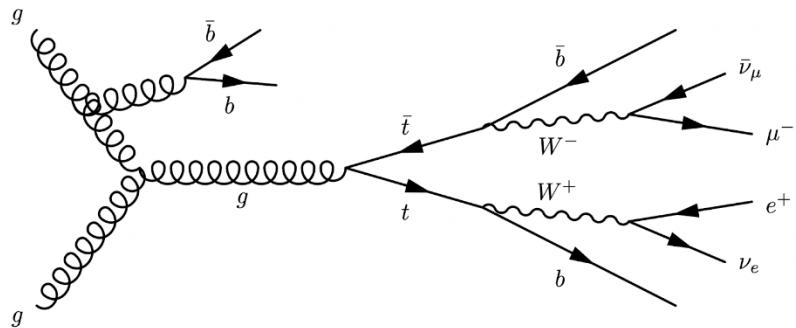


Figure 3: Feynman diagram of top quark pair production with additional  $b$ -jet and electron-muon pair in the final state.

## 34 2 Monte Carlo Simulation and Event Selection

35 This section describes ATLAS dataset and Monte Carlo simulation relevant to this analysis,  
 36 and how events are selected.

### 37 2.1 Dataset and Monte Carlo Simulation

38 The full dataset used in this research was collected by the ATLAS experiment during 2015–  
 39 2018 at  $\sqrt{s} = 13$  TeV, corresponding to an integrated luminosity of  $139\text{ fb}^{-1}$ . Monte Carlo  
 40 (MC) simulations are used to model the signal and background processes. The signal is  
 41  $t\bar{t}$  production in association with additional  $b$ -jets from a gluon and electron-muon pair in  
 42 the final state. The background includes processes of  $t\bar{t} + V$ ,  $t\bar{t} + H$ ,  $W/Z + \text{jets}$ , single  
 43 top, di-bosons and other small compositions (see Figure 4).

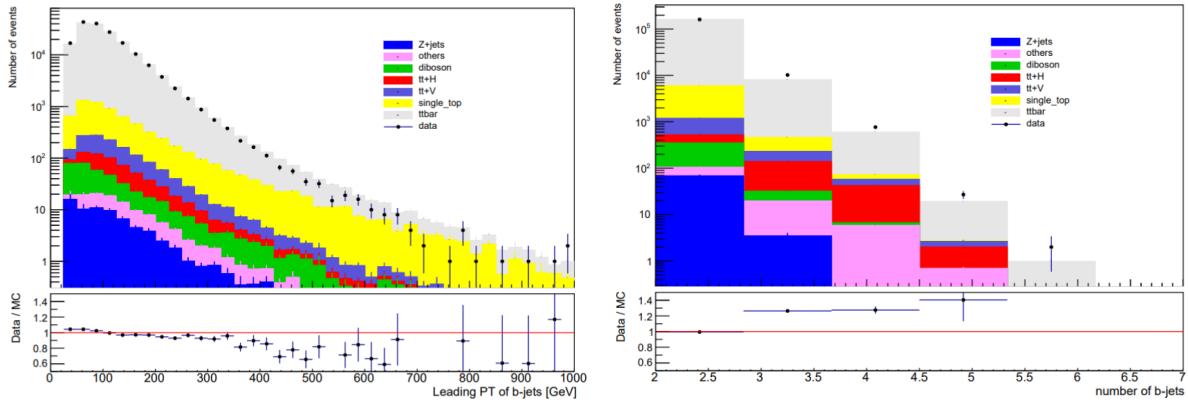


Figure 4: MC histogram modeling for the leading  $p_T$  of  $b$ -jets (left) and the number of  $b$ -jets (right) in each event.

### 44 2.2 Event Selection

45 This section describes how events used in this analysis are selected. The  $p_T$  (transverse  
 46 momentum, the momentum on the plane perpendicular to colliding protons) of leptons  
 47 are required to be at least 28 GeV to maximize the combined isolation efficiency which  
 48 describes event selection quality. The selection also requires exactly two leptons - electron  
 49 and muon - of the opposite charges. The requirement on different flavors suppresses  $Z + \text{jets}$   
 50 background ( $Z \rightarrow ll$  with both leptons of the same flavor). Tau leptons with small yield are  
 51 not considered in the analysis. The transverse momentum of jets are required to be at least  
 52 25 GeV. Jets are identified as  $b$ -jets by the means of a novel flavor tagging algorithm [1],  
 53 called DL1r, using a deep neural network, at 77% working point. The Powheg+Pythia8  
 54 MC simulation mixes the  $t\bar{t}$  inclusive sample with three exclusive:  $t\bar{t} + b\bar{b}$ ,  $t\bar{t} + b$ ,  $t\bar{t} + c$   
 55 (each of which has a higher number of raw events than a corresponding fraction in the  
 56 inclusive). To avoid event double counting, we apply the Heavy Flavor Filter Flag (HFFF)  
 57 to remove the overlap (see Table 1). We build neural networks for exactly three  $b$ -jets

<sup>58</sup> events (3j3b) (see Table 2) and at least four  $b$ -jets events (4j4b) separately, so different  
<sup>59</sup> selections on the number of  $b$ -jets are applied.

Sample	DID number	Only keep when HFFF is
$t\bar{t}$ inclusive	410472	0
$t\bar{t} + bb$	411076	1
$t\bar{t} + b$	411077	2
$t\bar{t} + c$	411078	3

Table 1: Use Heavy Flavor Filter Flag (HFFF) to avoid overlap of inclusive sample.

Source	2015+2016	2017	2018
Data	44615	55004	72485
Total MC	44586	55789	70277
$t\bar{t}$	42994	53556	67601
single top	1261	1627	2135
$t\bar{t} + V$	201	253	324
$t\bar{t} + H$	81	99	131
diboson	17	212	30
Z+jets	17	24	31
others	16	19	25

Table 2: Reweighted event yield from each source after selection.

### <sup>60</sup> 3 Reconstructed discriminating Features

<sup>61</sup> Additional  $b$ -jets are the ones originating from initial state radiation gluons through strong  
<sup>62</sup> interaction and other  $b$ -jets originate from top quarks pair decays through weak interac-  
<sup>63</sup> tion. The difference in the physical processes of  $b$ -jets origins results in a list of discrimi-  
<sup>64</sup> nating features associated with the  $b$ -jets and leptons, which allows us to build a  $b$ -jet origin  
<sup>65</sup> classifier. This section presents the discriminating features based on the radial distances  
<sup>66</sup> between and invariant masses of pairs of objects in an event.  $b$ -jets from top quarks are  
<sup>67</sup> expected to be:

- <sup>68</sup> 1. Close to leptons (Figure 5 - 7).
- <sup>69</sup> 2. far from each other (Figure 8).
- <sup>70</sup> 3. have lower invariant mass with leptons (Figure 9 - 11).
- <sup>71</sup> 4. have higher invariant mass with each other (Figure 12).

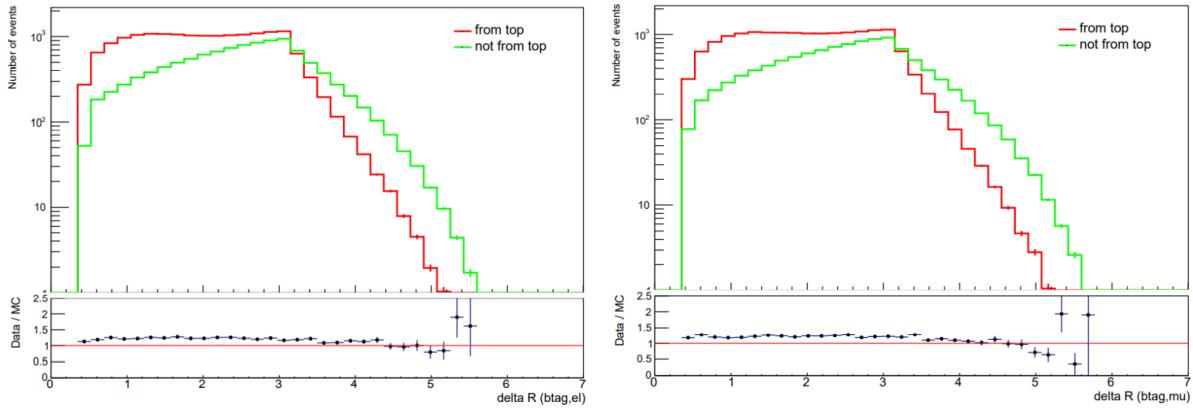


Figure 5: Radial distance between  $b$ -jet and electron (left) or muon (right).

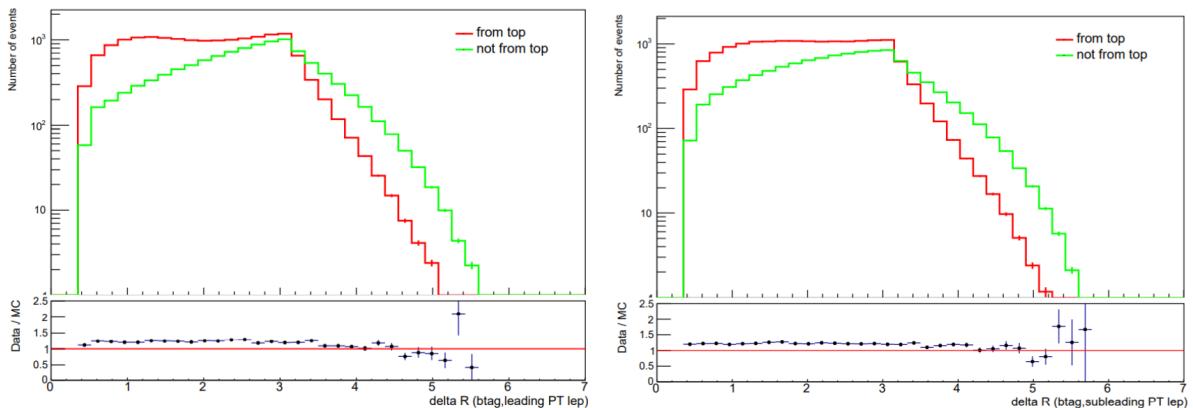


Figure 6: Radial distance between  $b$ -jet and the lepton with leading  $p_T$  (left) or sub-leading  $p_T$  (right).

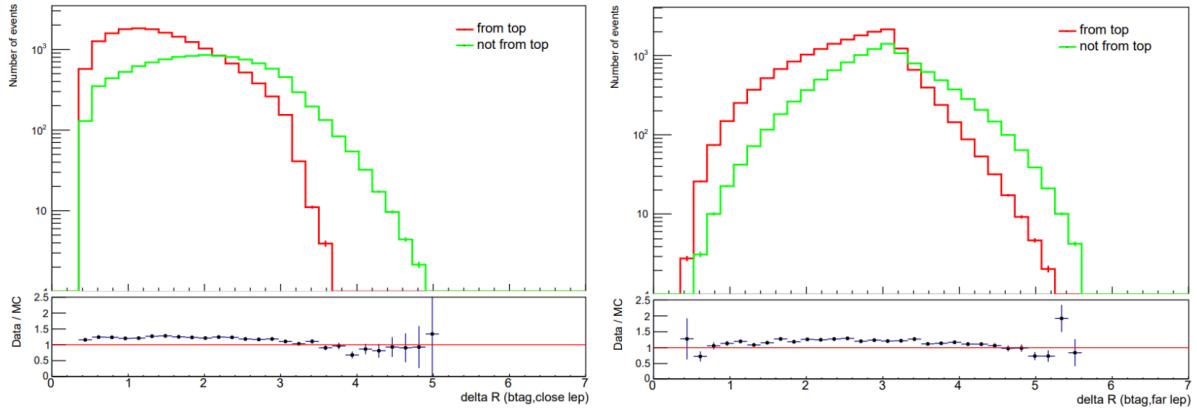


Figure 7: The close (left) or far (right) radial distance between  $b$ -jet and leptons.

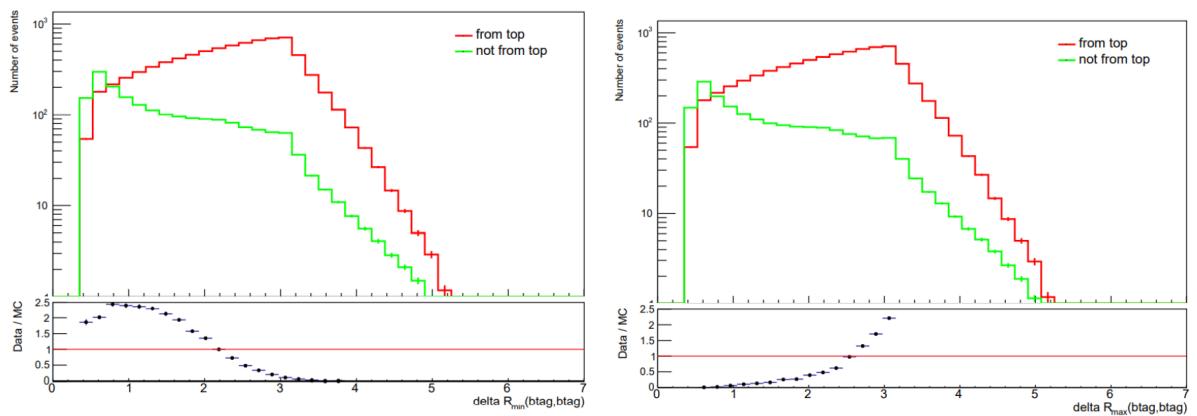


Figure 8: The closest (left) or farthest (right) radial distance between  $b$ -jets.

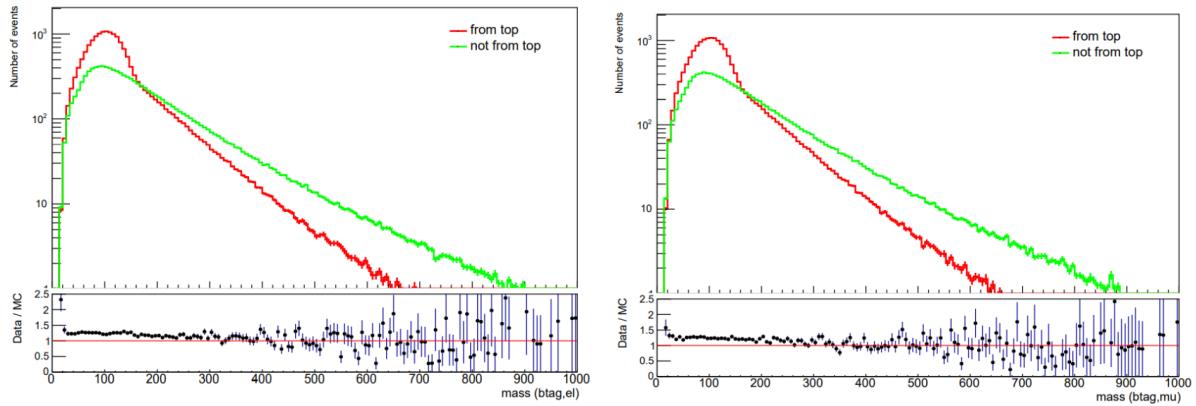


Figure 9: Invariant mass between  $b$ -jet and electron (left) or muon (right).

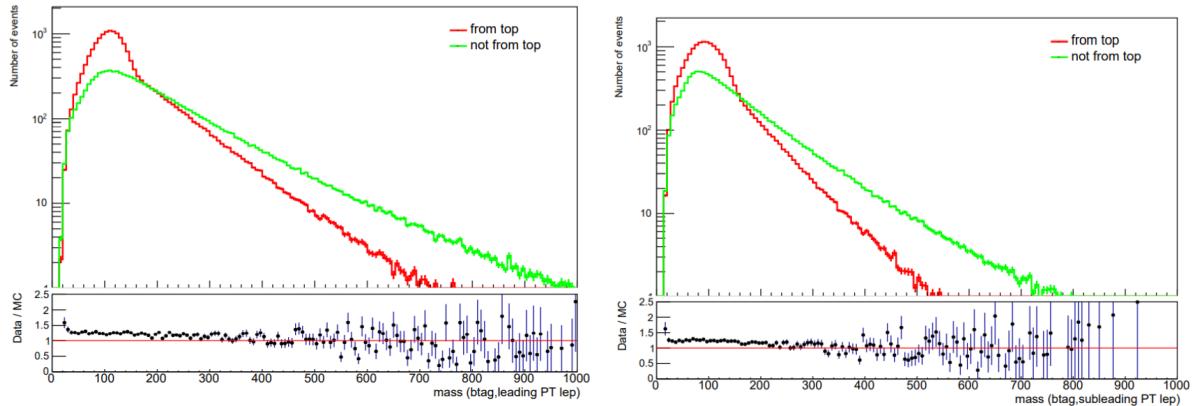


Figure 10: Invariant mass between  $b$ -jet and the lepton with leading  $p_T$  (left) or sub-leading  $p_T$  (right).

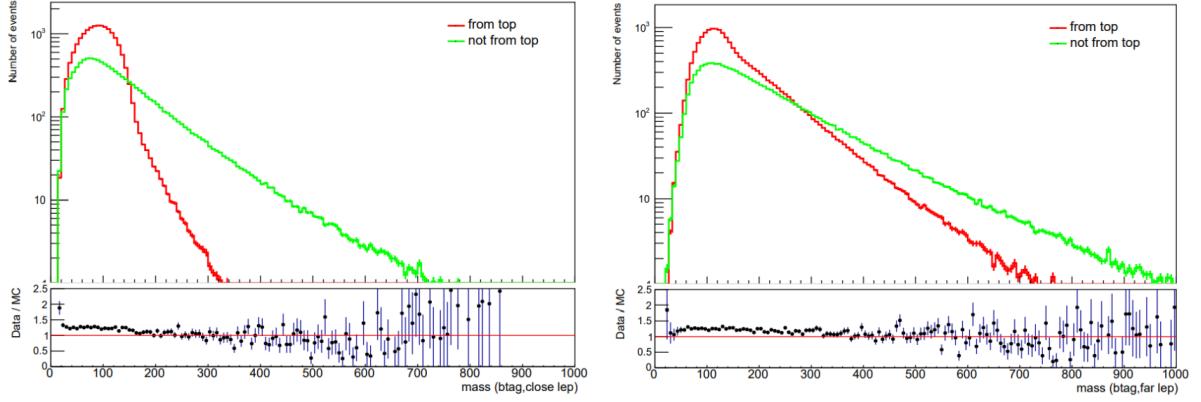


Figure 11: Invariant mass between  $b$ -jet and the lepton with close radial distance (left) or far radial distance (right).

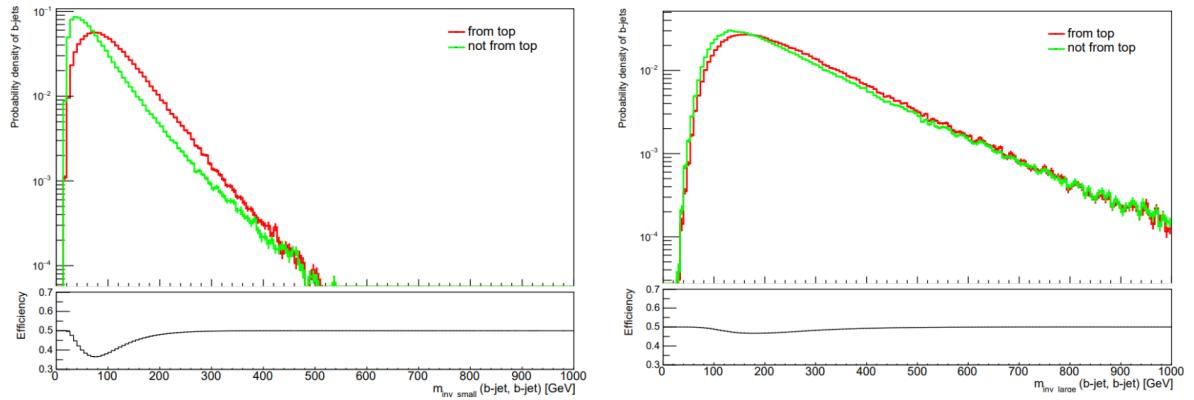


Figure 12: The smallest (left) or the largest (right) invariant mass between  $b$ -jets.

## 72 4 Neural Network based $b$ -jets origin classifier

73 This section presents the setup and performance evaluation of the neural network (NN)  
 74 classifier and comparison with previous boosted decision tree (BDT) classifier studied in  
 75 Run2  $t\bar{t}$  plus additional  $b$ -jets differential cross-section measurement with an electron and  
 76 a muon of the opposite signs in the final state.

### 77 4.1 Setup

78 The emergence of machine learning brought a new technical approach to analyze vast  
 79 amount of data provided by the ATLAS detector. This technique is able to learn behaviors  
 80 of observables in order to make predictions that we do not know from data. The previous  
 81 BDT classifier already showed better performance over classical approaches, and this  
 82 thesis uses Deep Neural Network, the other kind of machine learning technique, to try  
 83 to push the efficiency of the classifier even further and address weaknesses of the BDT

84 classifier. The BDT classifier suffers from  $p_T$  bias and impossibility to add matrix element  
85 and parton showering uncertainty treatment.

86 This Neural Network is constructed in TensorFlow and Keras framework.  $b$ -jets from  
87 Powheg+Pythia8 MC samples of  $t\bar{t}$  production are used to train the classifier. The training  
88 would stop automatically when the loss in test mode stops decreasing during the past ten  
89 epochs. Learning rate is adapted by exponential decay schedule:

$$L = L_0 * e^{-k*n}$$

90 where  $L$  is the current learning rate;  $L_0$  is the initial learning rate;  $k$  is a constant con-  
91 trolling the decay rate;  $n$  is the current epoch number. Setup of hyper-parameters are  
92 stated in Table 3.

Activation Function	Sigmoid
Loss Function	binary crossentropy
Hidden Layers	128, 128
Output Layer	1
Optimizer	Keras Adam
initial learning rate $L_0$	0.003
decay rate k	0.1

Table 3: Setup of hyper-parameters.

93 Multiple model architectures were studied and summarized in Table 4, 5 and 6.

Structure	Accuracy
2 x 128	0.7720
3 x 128	0.7717
4 x 128	0.7721
5 x 128	0.7716
9 x 128	0.7720
10 x 128	0.7647

Table 4: Tuning with the layer number.

Structure	Accuracy
32, 16, 8, 4	0.7707
64, 32, 16, 8, 4	0.7713
128, 64, 32, 16, 8, 4	0.7709
256, 128, 64, 32, 16, 8, 4	0.7709
512, 256, 128, 64, 32, 16, 8, 4	0.7714
1024, 512, 256, 128, 64, 32, 16, 8, 4	0.7716
2048, 1024, 512, 256, 128, 64, 32, 16, 8, 4	0.7709

Table 5: Decreasing node number by two times.

Structure	Accuracy
128, 128	0.7718
256, 256	0.7727
512, 512	0.7719
1024, 1024	0.7716
2048, 2048	0.7652

Table 6: Tuning with the node number.

94 Comparisons of the efficiencies show almost the same level of classifier performance for  
 95 most neural network structures with layer number from 2 to 9, and nodes number from  
 96 128 to 1024. One of optimal structures is represented by two layers with 128 nodes and  
 97 used in training (See Figure 13). The Area Under the Receiver Operating Characteristic  
 98 Curve (AUC score) in test mode after fully training is 0.832.

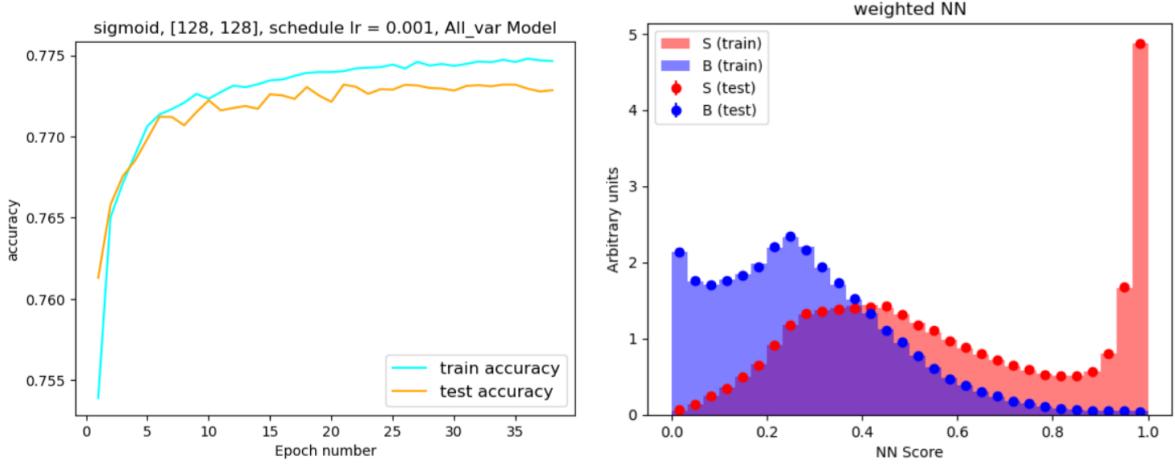


Figure 13: Training history and the final test AUC score (left). Neural Network score distribution (right).

## 99 4.2 Transverse Momentum Bias

100 One of the weaknesses of the BDT classifier mentioned in the beginning of the chapter  
 101 was bias to  $p_T$  of  $b$ -jets. Because  $p_T$  of  $b$ -jets assigned to top quarks and initial state  
 102 radiation gluons are measured in the differential cross-section analysis, such bias must be  
 103 minimized. As the classifier will be used for unfolding of leading, sub-leading from top  
 104 quark and additional  $b$ -jets  $p_T$  (4 variables in total), we aim that predictions of classifier  
 105 is independent of transverse momentum  $p_T$ . High correlation between input features and  
 106  $p_T$  could result in bias (See Figure 14).

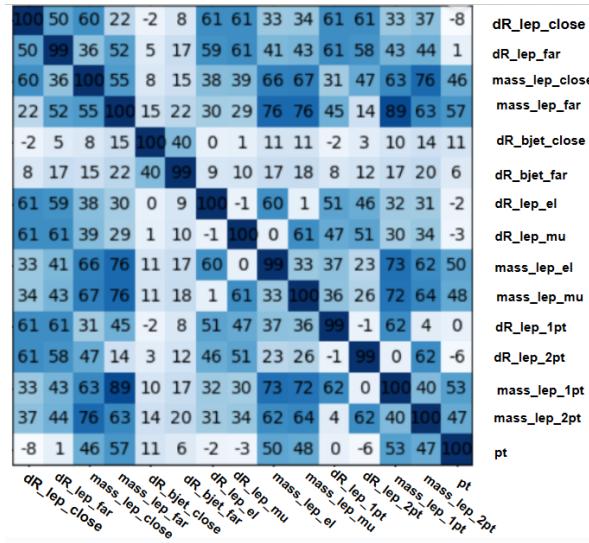


Figure 14: Correlation matrix of input features and transverse momentum.

107 As shown in the correlation matrix in figure 15, the invariant mass features have high  
 108 correlation which is larger than 40, as radial distance features have low correlation which

is almost smaller than 10. Results of previous  $b$ -jet origin classifier based on BDT showed a significant bias on  $p_T$  of  $b$ -jets to confirm the correlation matrix.

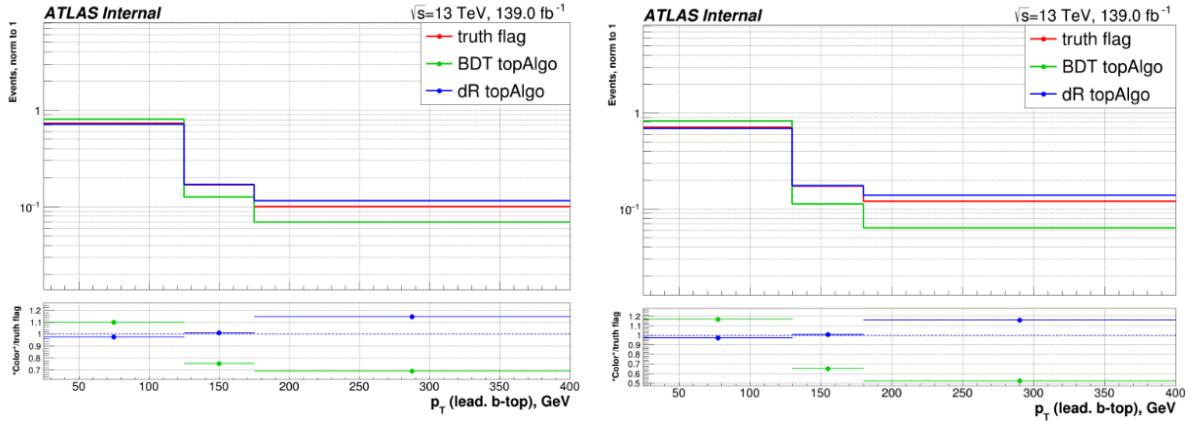


Figure 15: Distribution of transverse momentum of leading  $b$ -jet from top quark in 3j3b (left) and 4j4b (right).

BDT topAlgo (green), one classifier that is trained by all input features, preferred to identify  $b$ -jet with low  $p_T$  as the leading  $b$ -jet from top quark. The dR topAlgo (blue), the other classifier that is trained by only radial distance features, preferred to identify  $b$ -jet with high  $p_T$  as the leading  $b$ -jet from top quark. Compared with the truth label (red) by MC detector level, the two models have  $p_T$  bias in opposite directions, but the absolute value of ratio slope of dR topAlgo is smaller, which implies smaller  $p_T$  bias. It is consistent with the fact that radial distance features have low  $p_T$  correlation, as BDT topAlgo included invariant mass features with high  $p_T$  correlation.

However, the NN classifier shows different behaviors on  $p_T$  bias (see Figure 16 - 19). We train three models with different input features. Mass model (yellow) uses invariant mass features; dR model (blue) uses radial distance features; All var model (green) uses all input features. As mass model shows the largest  $p_T$  bias as expected, all var model shows even smaller bias than dR model does. For NN classifier, including invariant features with high  $p_T$  correlation decreases  $p_T$  bias. Reweighted bias considering width and height of bins also lead to the same conclusion. Although this phenomenon needs further research to discover reasons, it is beneficial to use a classifier with the best performance and the lowest  $p_T$  bias (See Figure 20). Definition of efficiency is described in next subsection.

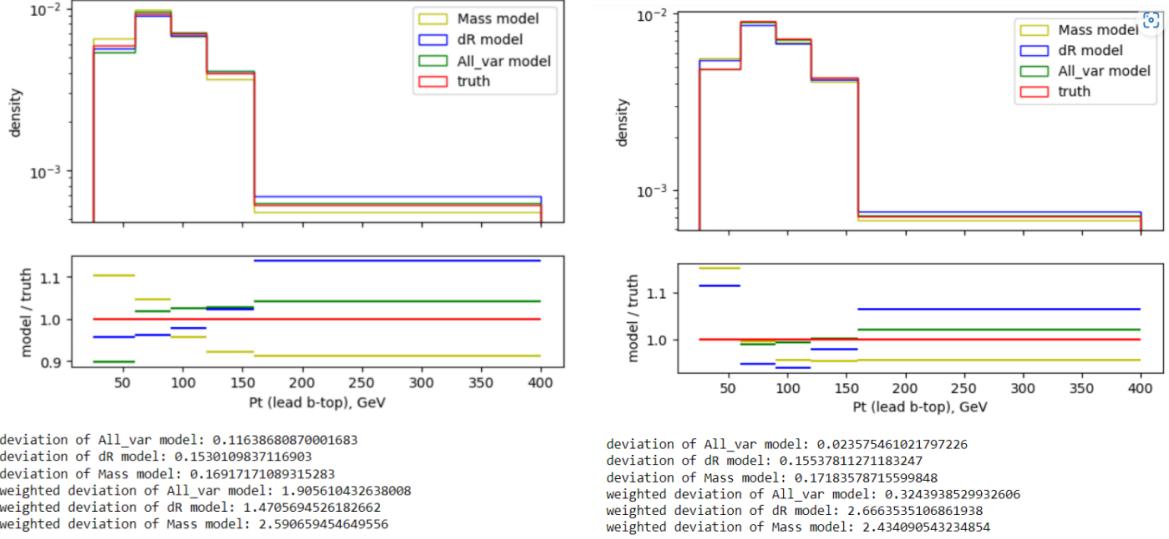


Figure 16: Distribution of transverse momentum of leading  $b$ -jet from top quark in 3j3b (left) and 4j4b (right).

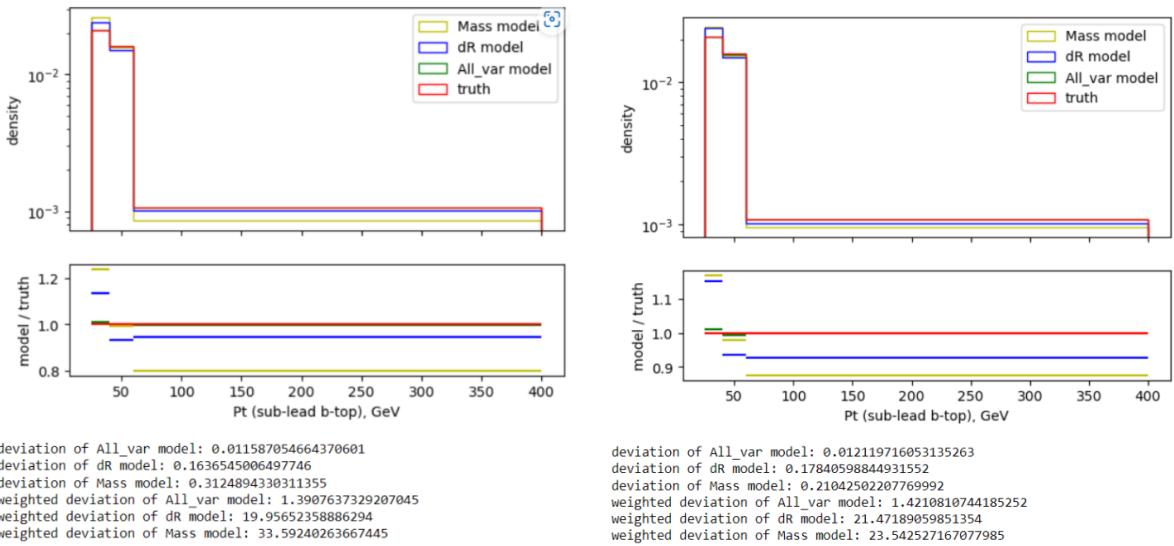


Figure 17: Distribution of transverse momentum of sub-leading  $b$ -jet from top quark in 3j3b (left) and 4j4b (right).

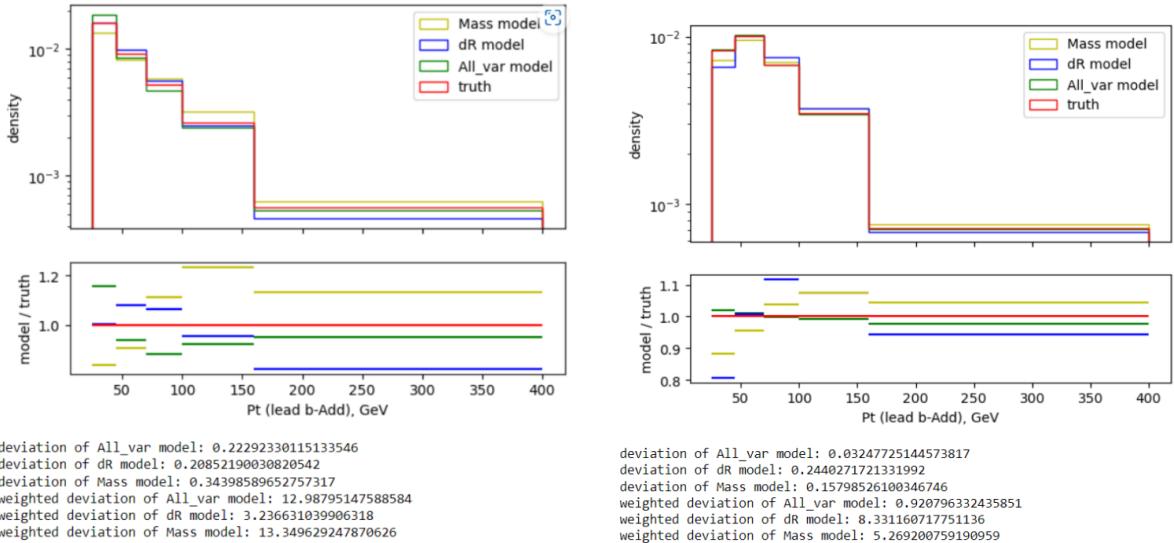


Figure 18: Distribution of transverse momentum of leading additional  $b$ -jet in 3j3b (left) and 4j4b (right).

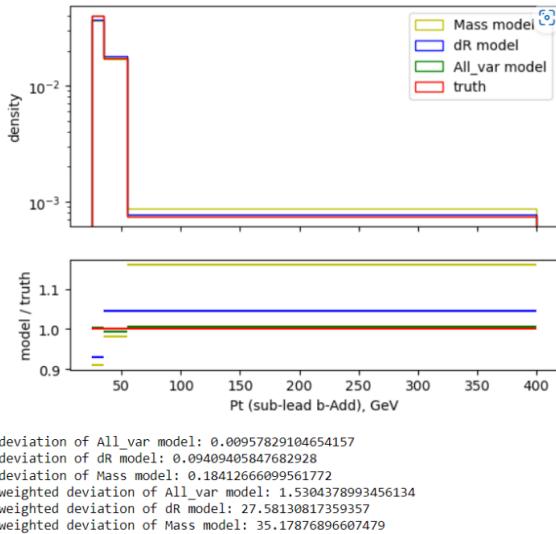


Figure 19: Distribution of transverse momentum of sub-leading additional  $b$ -jet in 4j4b.

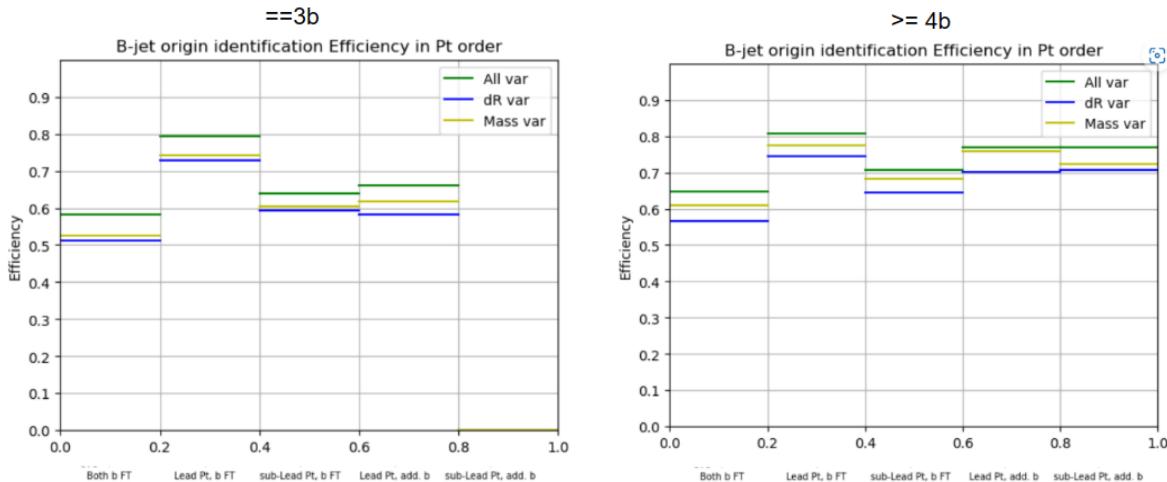


Figure 20: Efficiency of four identities of  $b$ -jets and both  $b$ -jets from top quark in 3j3b (left) and 4j4b (right).

### 128 4.3 Performance Comparison with BDT model

129 The standard Keras accuracy for sequential models is not the best approach to quantify  
130 performance of classifiers, as it uses a fixed threshold for all events, which is not an optimal  
131 approach for the  $b$ -jets origin classifier. Instead we describe performance of classifiers by  
132 efficiency without fixing at any threshold. In each event, we assign two  $b$ -jets with the lowest  
133 two NN scores to top quarks, and the rest of  $b$ -jets as additional. We then distinguish  
134 the  $b$ -jets from tops as the leading and sub-leading  $p_T$  from tops. The additional  $b$ -jets,  
135 if there is more than one, are sorted from the highest to the lowest  $p_T$  as well. Efficiency

for each category of  $b$ -jet is expressed as ratio of the number of correctly assigned  $b$ -jets by the classifier to the truth number of  $b$ -jets in one category. For each  $b$ -jet, we use this assignment process to predict its category, so the total assignment number is same as the  $b$ -jet number for each category.

Figure 21 - 24 present efficiencies of the NN and BDT classifiers for each category of  $b$ -jets. The efficiencies are similar between the two classifiers.

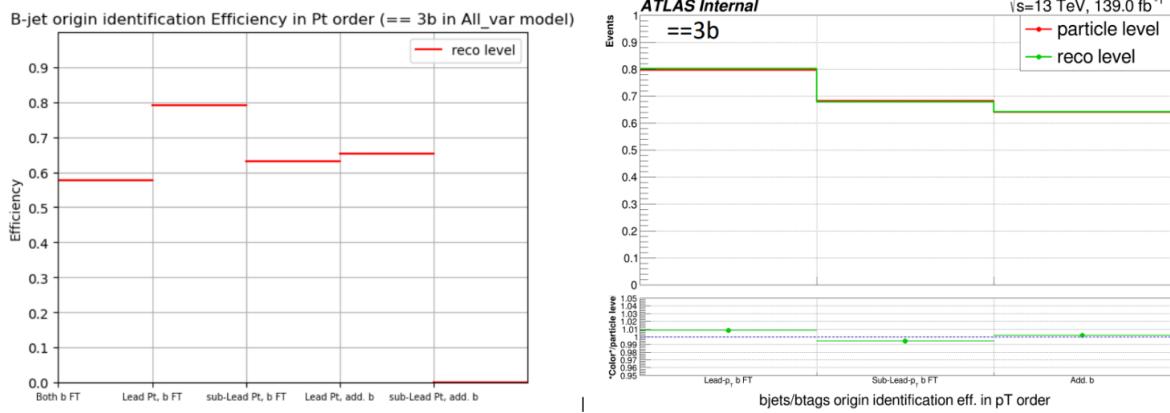


Figure 21: Efficiency of four identities of  $b$ -jets and both  $b$ -jets from top quark in 3j3b by NN (left) and BDT (right) with all features inputted.

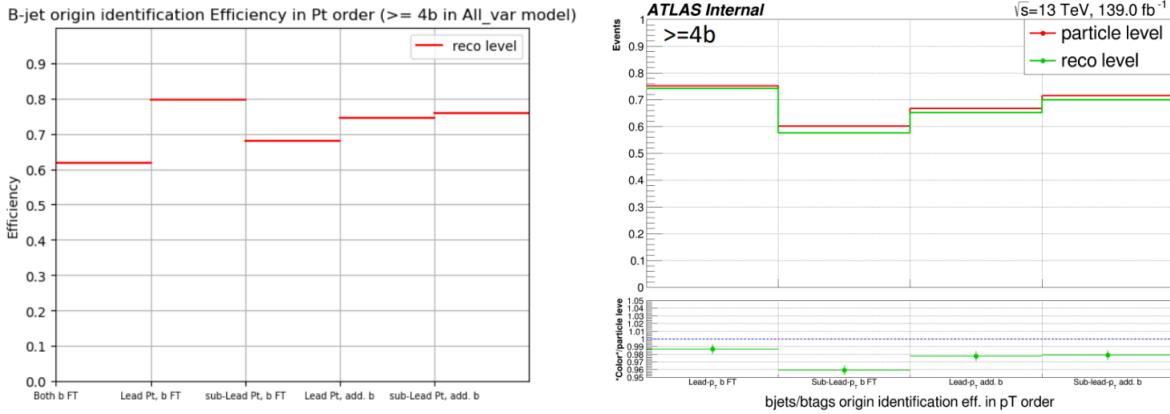


Figure 22: Efficiency of four identities of  $b$ -jets and both  $b$ -jets from top quark in 4j4b by NN (left) and BDT (right) with all features inputted.

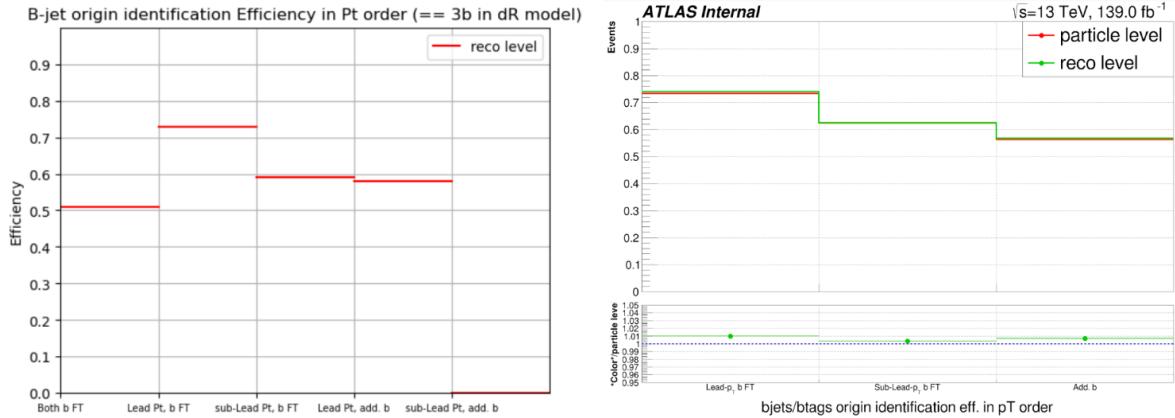


Figure 23: Efficiency of four identities of  $b$ -jets and both  $b$ -jets from top quark in 3j3b by NN (left) and BDT (right) with only radial distance features inputted.

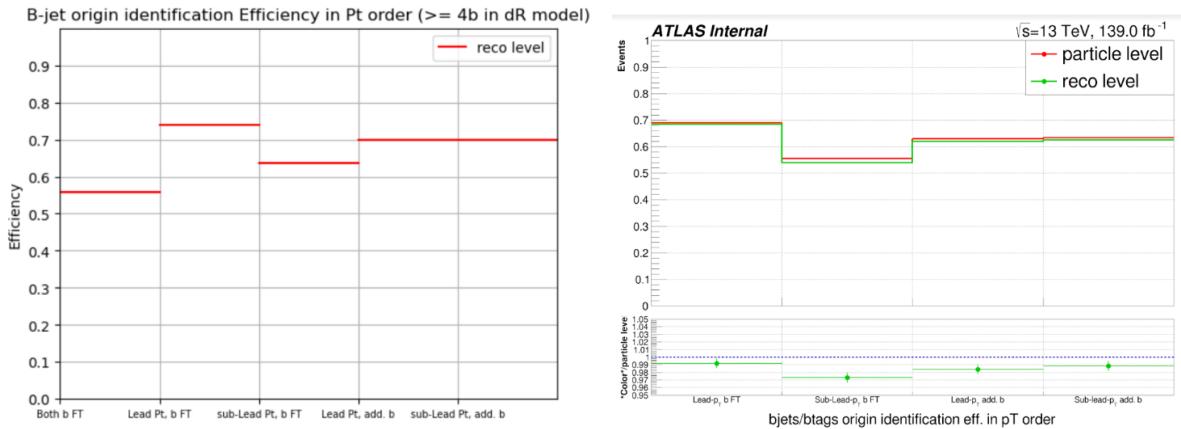


Figure 24: Efficiency of four identities of  $b$ -jets and both  $b$ -jets from top quark in 4j4b by NN (left) and BDT (right) with only radial distance features inputted.

## 142 5 Domain-Adversarial Neural Network (DANN)

143 We apply adversarial domain adaptation to our neural network model, in order to build  
 144 independence of MC model and decrease of  $p_T$  bias.

### 145 5.1 Classifier Independent of Input MC Model

146 As stated before, the results of our neural network are predicted by a classifier which  
 147 is trained by nominal MC (Powheg+Pythia8) generator. The classifier therefore suffers  
 148 from a generator bias when used on the real data to identify  $b$ -jet origin. To solve this  
 149 problem, we adopt domain adversarial approach to build a classifier which is independent  
 150 of input MC model.

151 We make use of a neural network containing event and domain classifier with a gradi-  
 152 ent reversal layer to simultaneously enable signal versus background events classifica-  
 153 tion on the one hand, while on the other hand minimising the difference in response of the  
 154 network to background samples originating from different MC models via adversarial do-  
 155 main classification loss (see Figure 32) [4]. The gradient reversal layer enables the model  
 156 to minimize the loss  $L_y - \lambda L_d$  by finding an optimal point in two directions of each la-  
 157 bel classifier. The factor  $\lambda$  thus scales effect of domain label. When  $\lambda = 0$ , the DANN  
 158 becomes equivalent to the regular DNN.

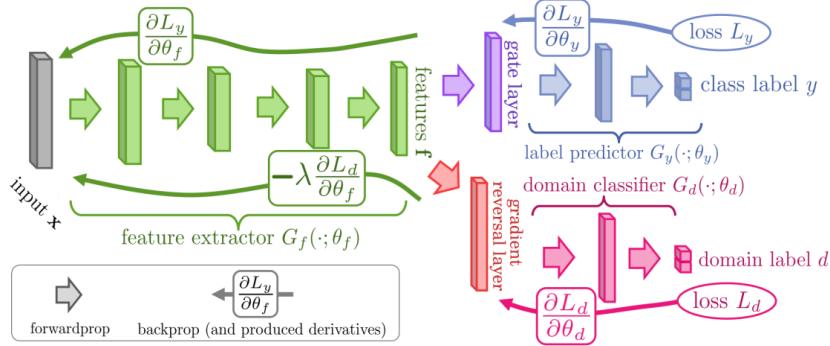


Figure 25: Domain-adversarial network as an alternative to reduce classification bias, adapted from [5].

159 As implementing the nominal MC (Powheg+Pythia8) production into class label, we  
 160 use Sherpa 2.2.10 for both the matrix element and the parton showering as an alternative  
 161 MC production on domain label. Firstly four values of scale factors are tried during  
 162 training of DANN, in order to show the signal and background distributions by class  
 163 label and domain label respectively. (see Figure 26 and 27). Then we make a scan of the  
 164 scale factor from 0 to 50 with step 2 to find the optimal value (see Figure 28). The DANN  
 165 effect is quantified as ratio deviation, the absolute difference value between class/domain  
 166 ratio and 1, ratio = 1 means a perfect generalization between two labels. We also use an  
 167 alternative for the matrix element (aMC@NLO) and the nominal (Pythia8) for the parton  
 168 showering (see Figure 29, 30 and 31). The results show that difference of background label  
 169 ( $b$ -jets from top quark) decreases, but difference of signal label increases as the scale factor  
 170 increases, which is not expected to be. Further improvements on DANN algorithm are  
 171 necessary to generalize MC input model better.

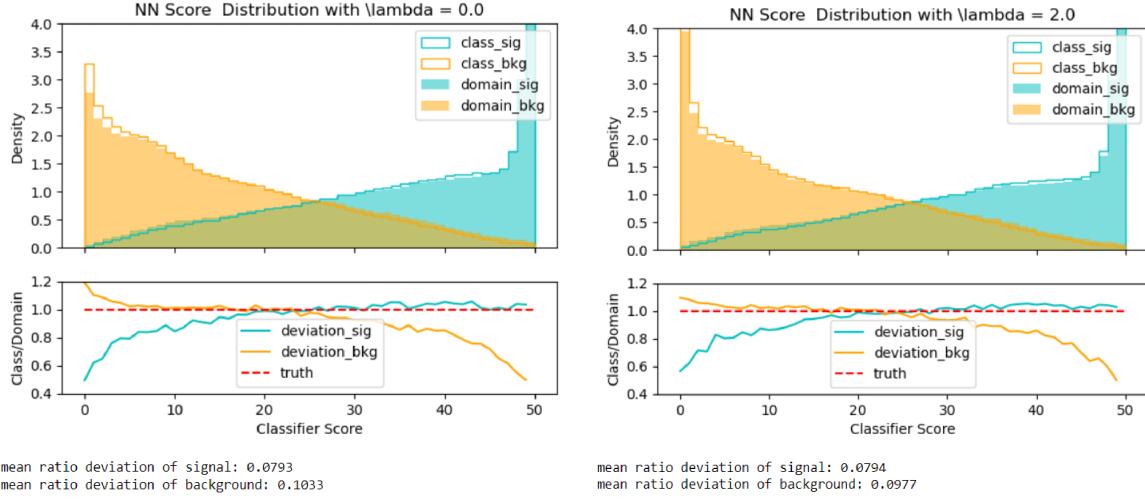


Figure 26: (Sherpa 2.2.10) Top: Signal and background prediction distribution by class label and domain label when  $\lambda = 0$  (left) or  $\lambda = 2$  (right). Bottom: The ratio of class label prediction to domain label prediction in each bin.

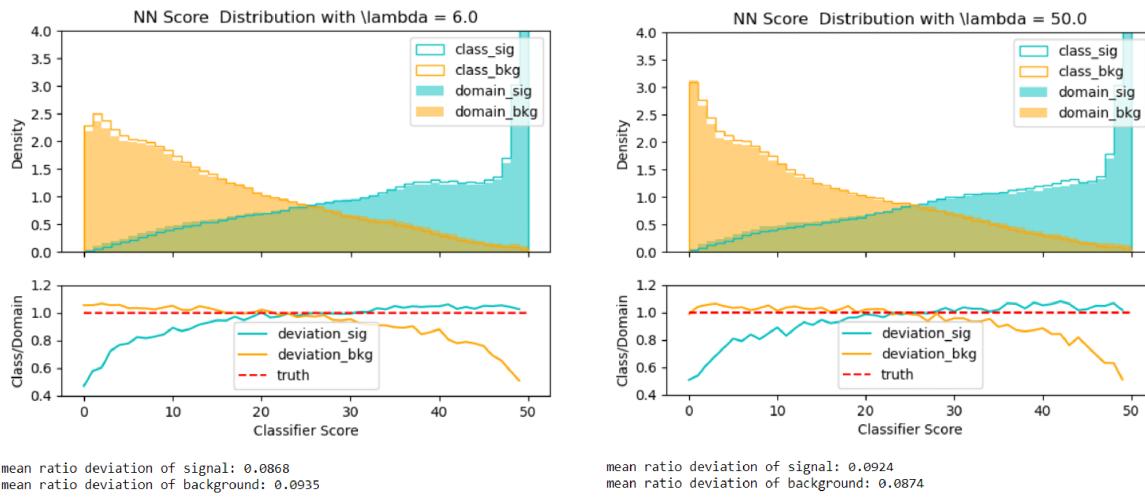


Figure 27: (Sherpa 2.2.10) Top: Signal and background prediction distribution by class label and domain label when  $\lambda = 6$  (left) or  $\lambda = 50$  (right). Bottom: The ratio of class label prediction to domain label prediction in each bin.

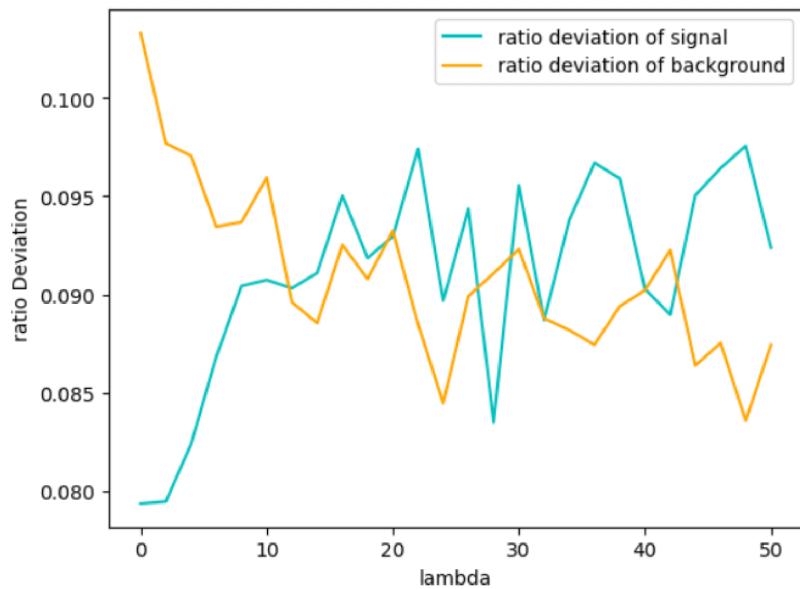


Figure 28: (Sherpa 2.2.10) The absolute difference value between class/domain ratio and 1, in dependence of the scale factor.

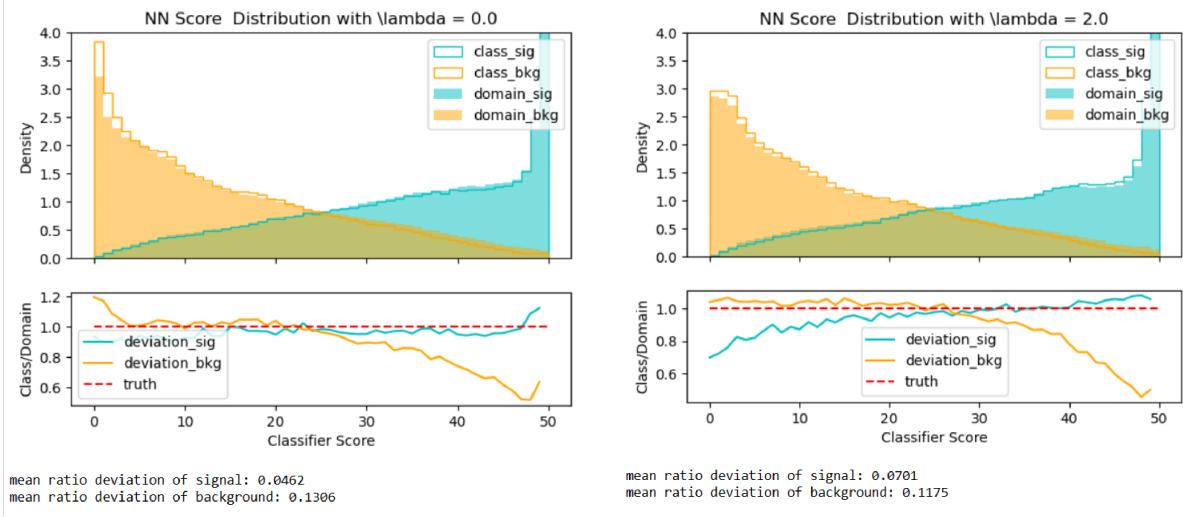


Figure 29: (aMC@NLO+Pythia8) Top: Signal and background prediction distribution by class label and domain label when  $\lambda = 0$  (left) or  $\lambda = 2$  (right). Bottom: The ratio of class label prediction to domain label prediction in each bin.

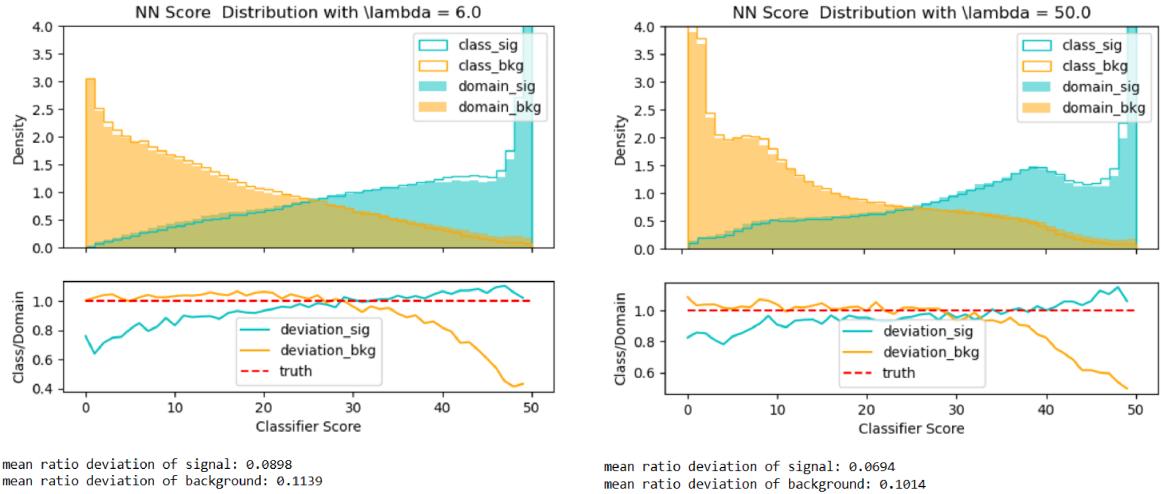


Figure 30: (aMC@NLO+Pythia8) Top: Signal and background prediction distribution by class label and domain label when  $\lambda = 6$  (left) or  $\lambda = 50$  (right). Bottom: The ratio of class label prediction to domain label prediction in each bin.

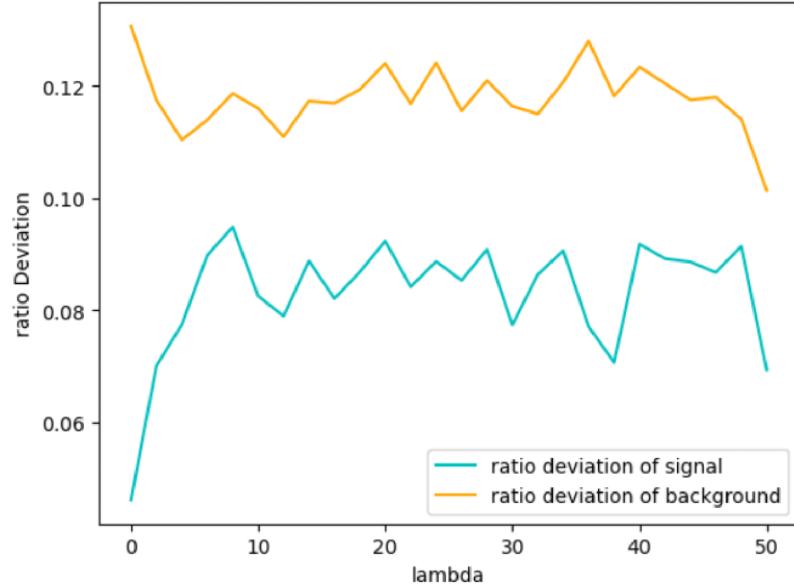


Figure 31: (aMC@NLO+Pythia8) The absolute difference value between class/domain ratio and 1, in dependence of the scale factor.

## 172 5.2 Repress Transverse Momentum Bias

173 DANN approach have a potential to decrease  $p_T$  bias. On the class label, we input all  
 174 discriminating variables, including invariant mass features and radial distance features.  
 175 On the domain label, besides features input into the class label,  $p_T$  of  $b$ -jets is added in.  
 176 The DANN model is thus expected to learn the behavior of  $b$ -jets and predict their origins  
 177 generally with and without  $p_T$  information. The generalization effect by the nominal MC  
 178 (Powheg+Pythia8) 4j4b sample is shown in Figure 32.

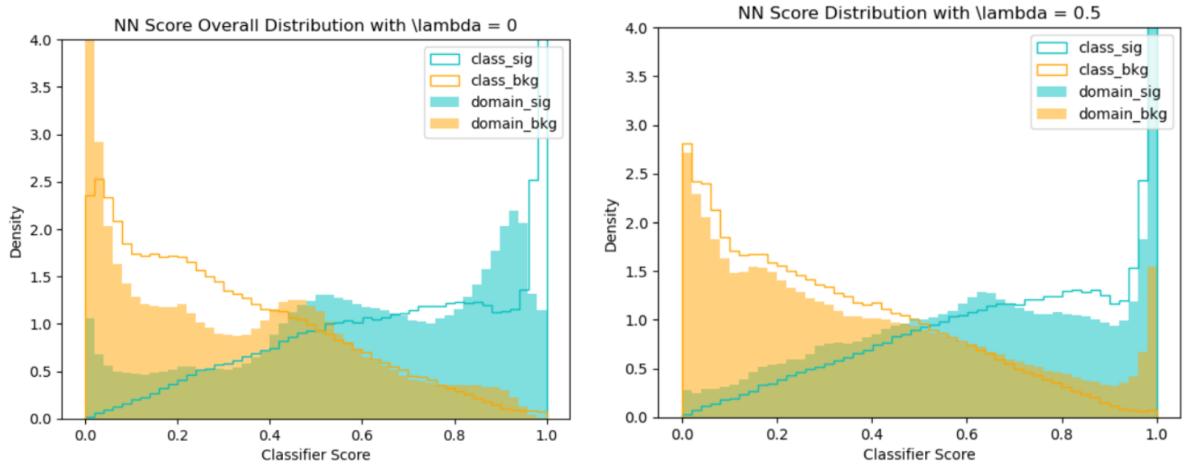


Figure 32: (DANN) Signal and background prediction distribution by class label and domain label when  $\lambda = 0$  (left) or  $\lambda = 0.5$  (right).

179 To find the optimal value of  $\lambda$  to minimize  $p_T$  bias, models with  $\lambda$  from 0 to 15 were  
 180 scanned (see Figure 33). Although it is not clear to tell the minimum of  $p_T$  bias due to  
 181 large fluctuation, the optimal value of  $\lambda$  is the most likely to be in domain from 0.5 to  
 182 2.5.

Class: M\_inv, dR  
 Domain: b-jet pT, M\_inv, dR

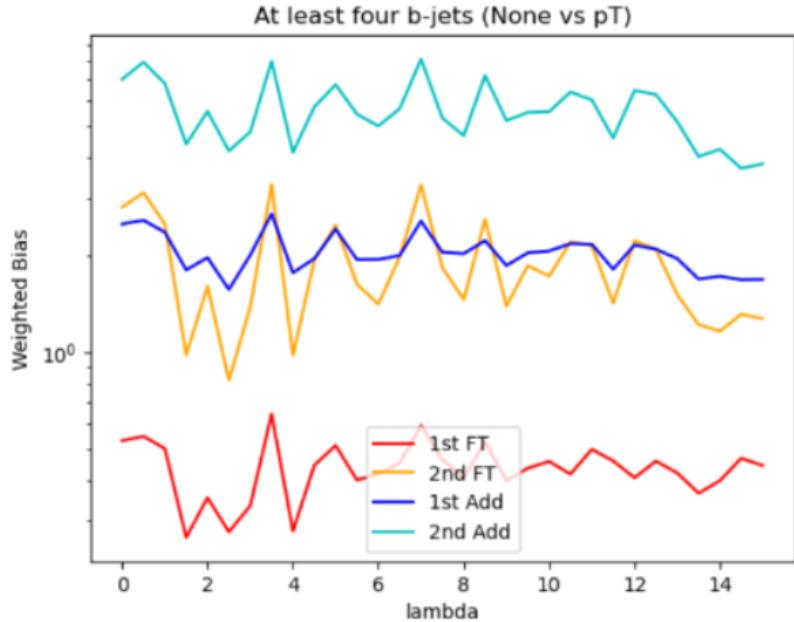


Figure 33:  $p_T$  bias vs the scale factor value  $\lambda$  from 0 to 15 for four categories of  $b$ -jets: leading (1st FT) and sub-leading (2nd FT)  $p_T$   $b$ -jets from top quark quarks decays, and leading (1st Add) and sub-leading (2nd Add)  $p_T$  additional  $b$ -jets.

183 Then two narrower scans but with shorter steps are operated by the DANN model.  
 184 Both of their results show a minimum when  $\lambda = 0.6$  (see Figure 34). However, the  
 185 decrease is relatively small compared with the large fluctuation. The conclusion of 0.6  
 186 could change under model training with different randomness seeds. Further research is  
 187 necessary to improve the function of decreasing  $p_T$  bias by DANN.

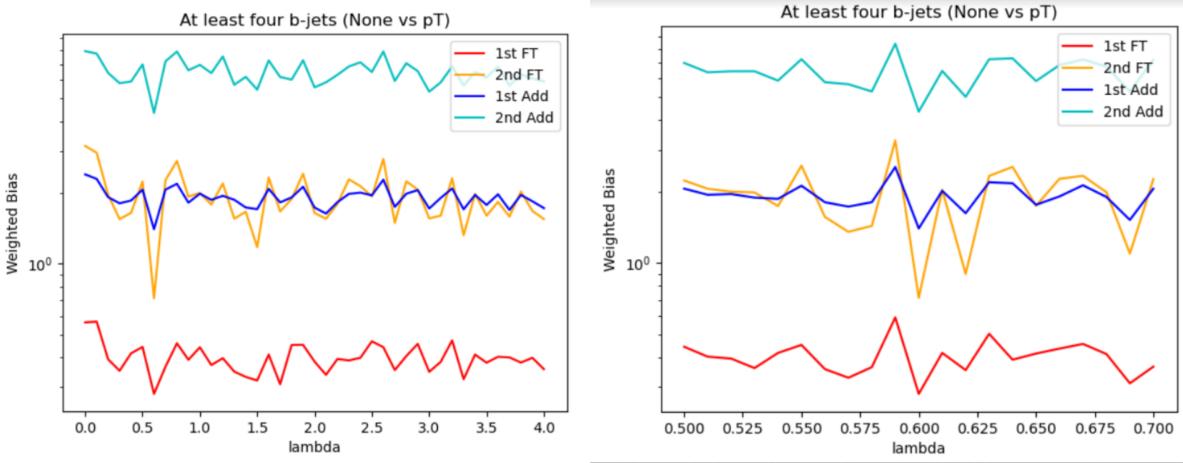


Figure 34:  $p_T$  bias vs the scale factor value  $\lambda$  from 0 to 4 (left) and from 0.5 to 0.7 (right) for four categories.

## 188 6 Conclusions

189 This thesis applies deep neural network (DNN) to build two  $b$ -jet origin classifiers specific-  
 190 ally for  $t\bar{t}$  production in association with additional  $b$ -jets and an electron and a muon of  
 191 the opposite charges in the final state of 3j3b and 4j4b events respectively. In 3j3b events,  
 192 this DNN behaves slightly worse than the previous BDT on prediction of sub-leading  $p_T$   
 193  $b$ -jet from top quarks, but better on  $b$ -jets in 4j4b events. Different behaviors on  $p_T$  bias  
 194 are showed, which implies complicated relation between  $p_T$  correlation of input features  
 195 and  $p_T$  bias of the classifier. DNN presents much smaller  $p_T$  bias than BDT does, as DNN  
 196 is trained by all input features, but BDT has to exclude invariant mass features.

197 The implement of DANN shows a potential to build a MC sample independent clas-  
 198 sifier. The result presents generalization improvement on predictions of  $b$ -jets from top  
 199 quarks, but make it worse for predictions of additional  $b$ -jets. It is probably because the  
 200 two MC sample from different generators has been so similar that the current general-  
 201 ization power by DANN cannot generalize them even more. On the other hand, DANN  
 202 for decreasing  $p_T$  bias shows better generalization effect, as the inputs to class label and  
 203 domain label differ from each other more. We operate three scans over the scale factor  
 204  $\lambda$  with narrower range and shorter step than the last scan, in order to save computation  
 205 time. We find out the optimal value of  $\lambda$  at 0.6, but the decrease of weighted bias is too  
 206 small to exclude influence from random fluctuation. The problems above need further  
 207 research to improve the current DANN application and even add the second or more do-  
 208 main labels, in order to combine the purposes of decreasing  $p_T$  bias and generalizing MC  
 209 input samples into one DANN classifier.

## 210 References

- 211 [1] Optimisation and performance studies of the ATLAS  $b$ -tagging algorithms for the  
 212 2017-18 LHC run. 7 2017.

- 213 [2] Morad Aaboud, Georges Aad, Brad Abbott, B Abeloos, DK Abhayasinghe, SH Abidi,  
214 OS AbouZeid, NL Abraham, H Abramowicz, H Abreu, et al. Observation of higgs  
215 boson production in association with a top quark pair at the lhc with the atlas detector.  
216 *Physics Letters B*, 784:173–191, 2018.
- 217 [3] Georges Aad, B Abbott, Kira Abeling, Nils Julius Abicht, SH Abidi, Asmaa Aboul-  
218 horma, Halina Abramowicz, Henso Abreu, Yiming Abulaiti, AC Abusleme Hoffman,  
219 et al. Observation of four-top-quark production in the multilepton final state with the  
220 atlas detector. *The European Physical Journal C*, 83(6):496, 2023.
- 221 [4] Jose M. Clavijo, Paul Glaysher, Judith M. Katzy, and Jenia Jitsev. Adversarial domain  
222 adaptation to reduce sample bias of a high energy physics classifier. 2021.
- 223 [5] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle,  
224 François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial  
225 training of neural networks. 2016.