

# Identification of hadronically decaying $W$ bosons and top quarks using multivariate techniques at ATLAS

Tatsumi Nitta, On behalf of the ATLAS Collaboration

Waseda Research Institute for Science and Engineering, Waseda University, Tokyo, Japan

E-mail: [tatsumi.nitta@cern.ch](mailto:tatsumi.nitta@cern.ch)

**Abstract.** By colliding protons and examining particles emitted from the collisions, the Large Hadron Collider aims to study the interactions of quarks and gluons at the highest energy accessible in a controlled experimental way. In such collisions,  $W$  bosons or top quarks which have TeV scale momentum can be accessible. Reconstructing such boosted jets are becoming important. In particular, the ability to identify original particle that decays to quarks against normal QCD jets plays a central role in various searches at high energy scale. This is typically done by the use of a single physically motivated observable constructed from the constituents of the jet. In this contribution, multiple complementary observables are combined using boosted decision trees and neural networks to increase the ability to distinguish  $W$  bosons and top quarks from light quark jets in the ATLAS experiment.

Since hadronic decay ratios of  $W$  boson and top quark are higher than leptonic decay, these are essential probes for rare phase space, especially for heavy resonance search at the Large Hadron Collider (LHC). For instance,  $W$  and top quark jets are used in diboson resonance searches [1],  $t\bar{t}$  resonance search [2], and measurement of differential cross section of high momentum top quarks [3] in ATLAS [4]. In this regime, frequently  $W$  and top quark jets have much higher momenta than their rest mass, thus these can be reconstructed by a single large radius jet. Such a large radius jet tends to have similar features to QCD jets produced by the single quark or gluon at much higher rate at LHC. In order to take into account radiation pattern of the jets, physically motivated observables such as mass and the number of core energy deposition area inside the jets have been used to effectively tag  $W$  or top quark jets in ATLAS [5, 6]. Applying machine learning techniques with several high level features of the large radius jet is a natural way to expand their performance by analogy to a multivariate  $b$ -tagging technique [7]. In this study, two kinds of machine learning techniques, Deep Neural Network (DNN) implemented by Keras [8] and Boosted Decision Tree (BDT) implemented by TMVA [9] are studied.

**Table 1.** Summary of training and testing sample statistics for each signal and background processes.

Purpose Sample	Training (Top Quark)		Training ( $W$ Boson)		Testing		
	Signal	Background	Signal	Background	Top Signal	$W$ Signal	Background
Number of jets	$1 \times 10^6$	$1 \times 10^6$	$7 \times 10^5$	$7 \times 10^5$	$4 \times 10^5$	$3 \times 10^5$	$1 \times 10^6$

## 1. Training Sample

All simulation samples are made by PYTHIA8 [10],  $W$  bosons are collected from  $W' \rightarrow WZ$  process, top quarks are collected from  $Z' \rightarrow t\bar{t}$  and quark and gluon jets are collected from QCD dijets events. The number of events used in this study is summarized in table 1. To avoid any bias due to difference of  $p_T$  distribution, training samples are reweighted to obtain flat  $p_T$  spectrum. While testing samples are reweighted to background  $p_T$  spectrum. Input variable candidates for BDT and DNN are gathered from high level feature proposed by a theoretical community shown in table 2 [11].

**Table 2.** Summary of candidate variables. The definition of variables are discussed in Ref. [11].

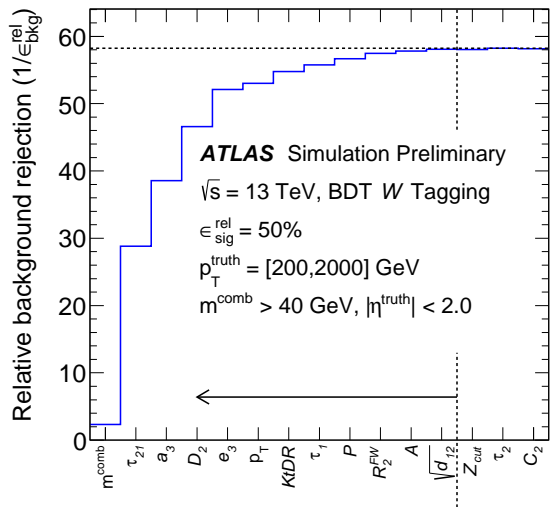
Observable	Variable	Used For
Jet mass	$m^{\text{comb}}$	top, $W$
Jet $p_T$	$p_T$	top, $W$
Energy Correlation Ratios	$e_3, C_2, D_2$	top, $W$
N-subjettiness	$\tau_1, \tau_2, \tau_3, \tau_{21}, \tau_{32}$	top, $W$
Center of Mass Observables	Fox Wolfram ( $R_2^{FW}$ )	$W$
Splitting Measures	$Z_{CUT}, \sqrt{d_{12}}, \sqrt{d_{23}}$	top, $W$
Planar Flow	$P$	$W$
Angularity	$a_3$	$W$
Aplanarity	$A$	$W$
KtDR	$KtDR$	$W$
Qw	$Q_w$	top

**Table 3.** Top-quark tagging inputs groups for DNN as used in figure 2.

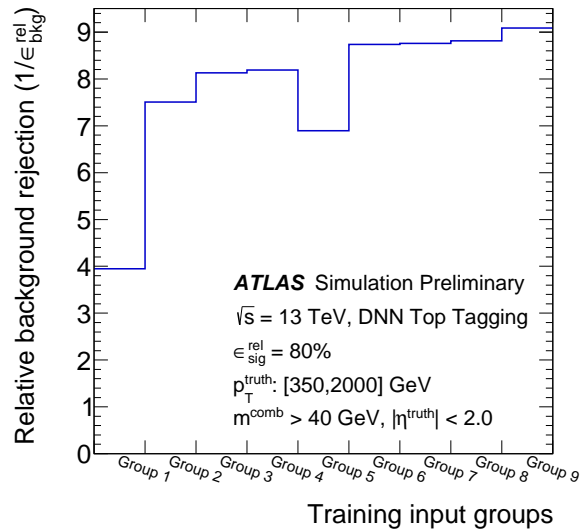
Group	variables
1	$C_2, D_2, \tau_{21}, \tau_{32}$
2	1, $m^{\text{comb}}$
3	2, $p_T$
4	$\tau_1, \tau_2, \tau_3, e_3, m^{\text{comb}}, p_T$
5	1, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
6	2, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
7	4, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
8	3, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$
9	1, 4, $\sqrt{d_{12}}, \sqrt{d_{23}}, Q_W$

## 2. Optimization

Input variables are selected from the candidate variables for each BDT and DNN independently to achieve the best performance for each technique. To find the optimal set of BDT input variables, the single input variable that gives the largest discrimination power is sequentially added to the BDT input list. Similar to the BDT training, the DNN is trained on different sets of input variables in order to find the optimal set of input variables. Unlike the BDT, sets of input variables are not defined by successively adding variables but are defined by grouping the input variables related to the corresponding signal shown in table 3. Input variables selection is shown in figure 1 and 2.



**Figure 1.** BDT background rejections for each variable set at the 50%  $W$  tagging efficiency. 12 variables which are at left side of dashed line are chosen for BDT  $W$ -tagging.



**Figure 2.** DNN background rejections for each variable group at 80% top-quark tagging efficiency. Input variable groups are described in table 3. Group 9 is chosen as a input set for DNN tagger.

Hyper-parameters for each BDT and DNN have been optimized by a grid-scan approach and the detail is described in elsewhere [11].

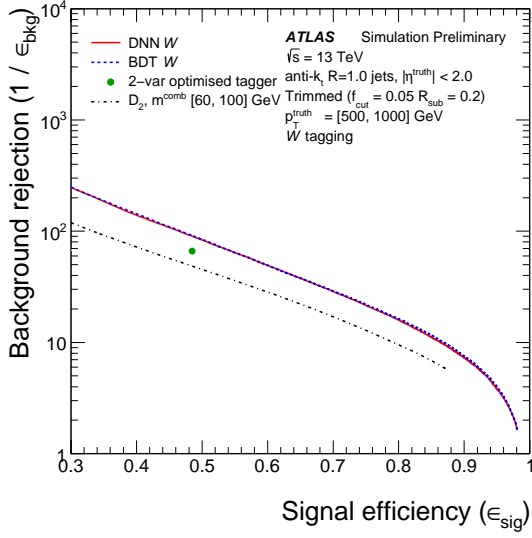
### 3. Result

The discrimination power of the taggers are quantitatively evaluated by receiver operating characteristic (ROC) curves as shown in figure 3 and 4. Both BDT and DNN taggers improve background rejection by 20% (200%) at 50%  $W$  (80% top) jet efficiency compared to 2-variables optimized tagger currently used in ATLAS [6].

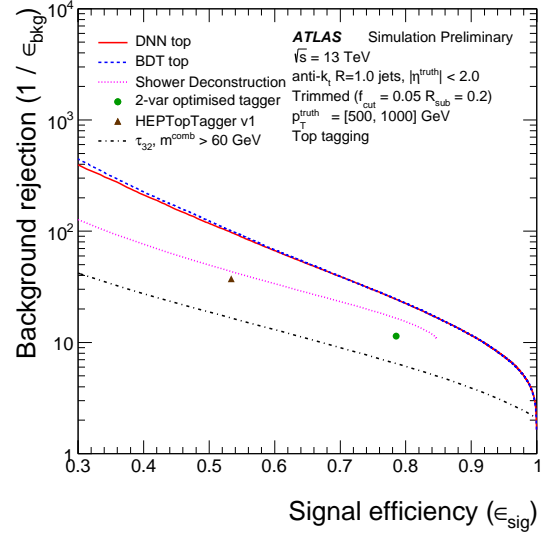
These taggers are validated by real data where signal and background are separately enriched by analysis-level event selections. The data used were taken in 2015 and 2016 at a center of mass energy of  $\sqrt{s} = 13$  TeV corresponding to integrated luminosities between 36.1 and 36.7 fb<sup>-1</sup>. These data are ensured that all sub-detector systems and trigger system were properly functional. From this dataset, three different samples are analyzed to study the performance of  $W$ -boson and top-quark tagging algorithms as follows:

- $t\bar{t}$  signal sample, enriched in hadronically decaying top quarks,
- quark or gluon jet from  $\gamma + \text{jet}$  background sample, covering the  $p_T$  range from 200 GeV to about 2 TeV,
- multijet background sample, covering the  $p_T$  range from 500 GeV to about 3.5 TeV.

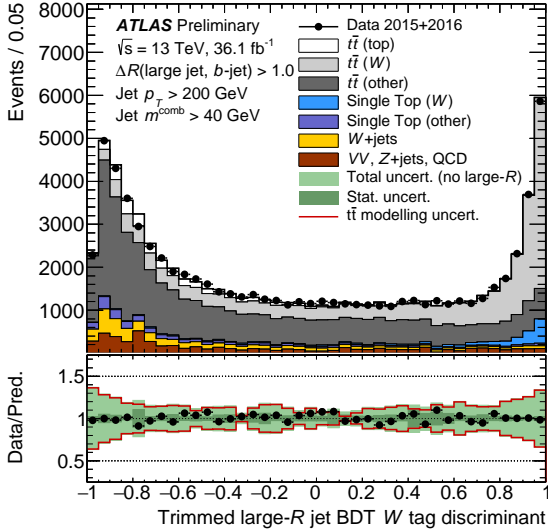
Signal modelings for each BDT- $W$  tagger and DNN-Top tagger are shown in figure 5 and 6. Background rejections are compared between data and MC simulation by PYTHIA8 [10] and SHERPA [12] as shown in figure 7 and 8. Differences between simulation and data are less than 10% within statistical uncertainties.



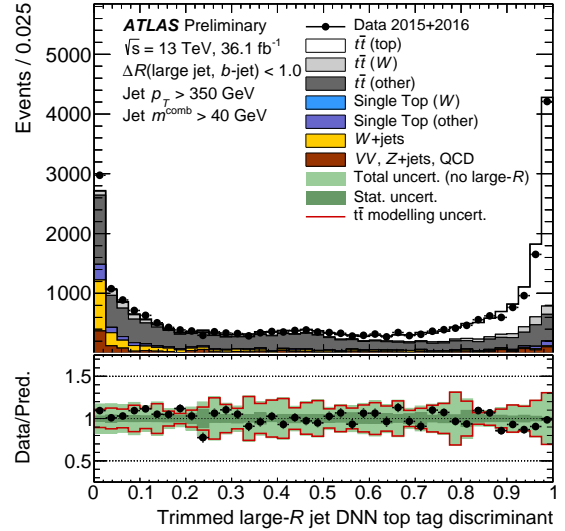
**Figure 3.** The performance comparison of the  $W$  tagger in a  $p_T = [500, 1000]$  GeV as a function of signal efficiency.



**Figure 4.** The performance comparison of the top tagger in a  $p_T = [500, 1000]$  GeV as a function of signal efficiency.



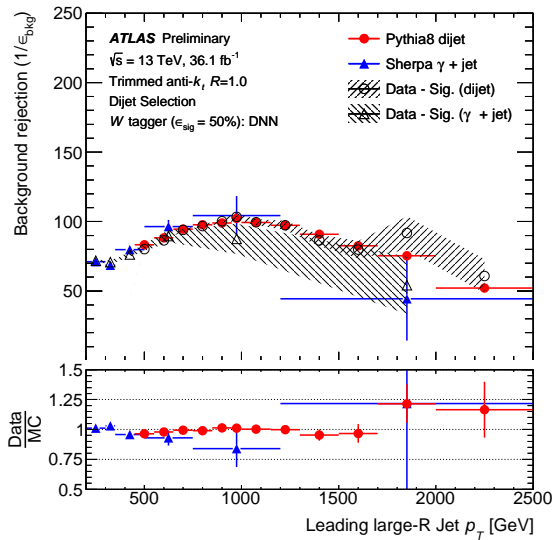
**Figure 5.** Comparison of the observed data and predicted MC distributions of the BDT discriminants for  $W$ -boson tagging. The  $t\bar{t}$  MC sample is decomposed using generator level information. The systematics uncertainties for these distributions do not include large radius jet uncertainties.



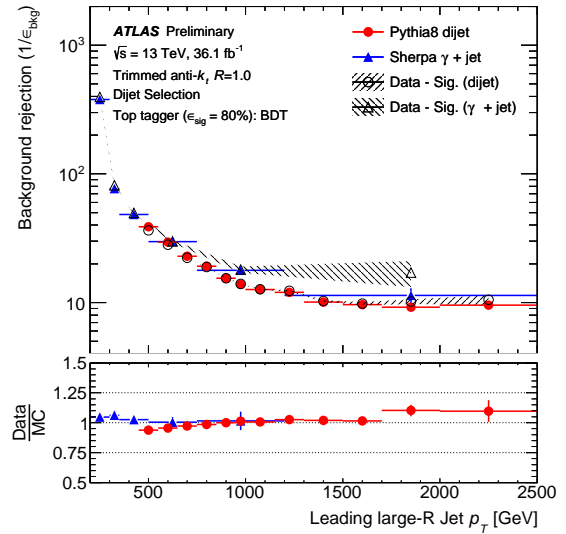
**Figure 6.** Comparison of the observed data and predicted MC distributions of the DNN discriminant for top-quark tagging. The  $t\bar{t}$  MC sample is decomposed using generator level information. The systematics uncertainties for these distributions do not include large radius jet uncertainties.

#### 4. Conclusion

$W$  and top quark jet taggers with BDT and DNN are constructed in the ATLAS experiment. Both BDT and DNN taggers improve background rejections by 20% (200%) at 50%  $W$  (80%



**Figure 7.** Background rejection as a function of leading jet  $p_T$  for the DNN  $W$ -boson taggers.  $\gamma + \text{jet}$  simulation (data) are shown as solid blue (hollow black) triangles and dijet simulation (data) are shown as red (hollow black) circles. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. Only the statistical uncertainties are shown.



**Figure 8.** Background rejection as a function of leading jet  $p_T$  for the BDT top-quark taggers.  $\gamma + \text{jet}$  simulation (data) are shown as solid blue (hollow black) triangles and dijet simulation (data) are shown as red (hollow black) circles. For each bin in data, the predicted signal contribution is subtracted prior to computing the background rejection. Only the statistical uncertainties are shown. Last point of  $\gamma + \text{jets}$  in ratio panel is not shown since it is outside of scale.

top) jet efficiency compared to 2-variables optimized tagger currently used in ATLAS. There is no difference between BDT and DNN taggers in terms of signal efficiency and background rejection power. Signal and background modelings are studied with  $t\bar{t}$ ,  $\gamma + \text{jet}$  and dijet events. The differences between data and simulation are within statistical uncertainties.

## References

- [1] ATLAS Collaboration 2016 *JHEP* **09** 173 (*Preprint* 1606.04833)
- [2] ATLAS Collaboration 2016 Search for heavy particles decaying to pairs of highly-boosted top quarks using lepton-plus-jets events in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector ATLAS-CONF-2016-014 URL <https://cds.cern.ch/record/2141001>
- [3] ATLAS Collaboration 2016 *Phys. Lett. B* **756** 52 (*Preprint* 1512.06092)
- [4] ATLAS Collaboration 2008 *JINST* **3** S08003
- [5] ATLAS Collaboration 2016 *Eur. Phys. J. C* **76** 154 (*Preprint* 1510.05821)
- [6] ATLAS Collaboration 2017 Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS ATLAS-CONF-2017-064 URL <https://cds.cern.ch/record/2281054>
- [7] ATLAS Collaboration 2015 Expected performance of the ATLAS  $b$ -tagging algorithms in Run-2 ATL-PHYS-PUB-2015-022 URL <https://cds.cern.ch/record/2037697>
- [8] F Chollet 2015 Keras URL <https://github.com/fchollet/keras>
- [9] A Hoecker et al 2007 TMVA: Toolkit for Multivariate Data Analysis PoS ACAT (2007) 040
- [10] Sjostrand T, Mrenna S and Skands P Z 2008 *Comput. Phys. Commun.* **178** 852–867 (*Preprint* 0710.3820)
- [11] ATLAS Collaboration 2017 Identification of Hadronically-Decaying  $W$  Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at  $\sqrt{s} = 13$  TeV ATL-PHYS-PUB-2017-004 URL <https://cds.cern.ch/record/2259646>
- [12] Gleisberg T, Hoeche S, Krauss F, Schonherr M, Schumann S et al. 2009 *JHEP* **0902** 007 (*Preprint* 0811.4622)