Frascati Physics Series Vol. 76 (2024)
8th young researchers' workshop 'Physics Challenges in the LHC Era'
May 13 and 16, 2024

MACHINE LEARNING TECHNIQUES FOR FOUR TOP FINAL STATES EXTRACTION

Cristina Giordano
Institute of High Energy Physics (HEPHY), Austrian Academy Of Sciences (ÖAW),
Technical University of Vienna (TUW), Vienna, Austria

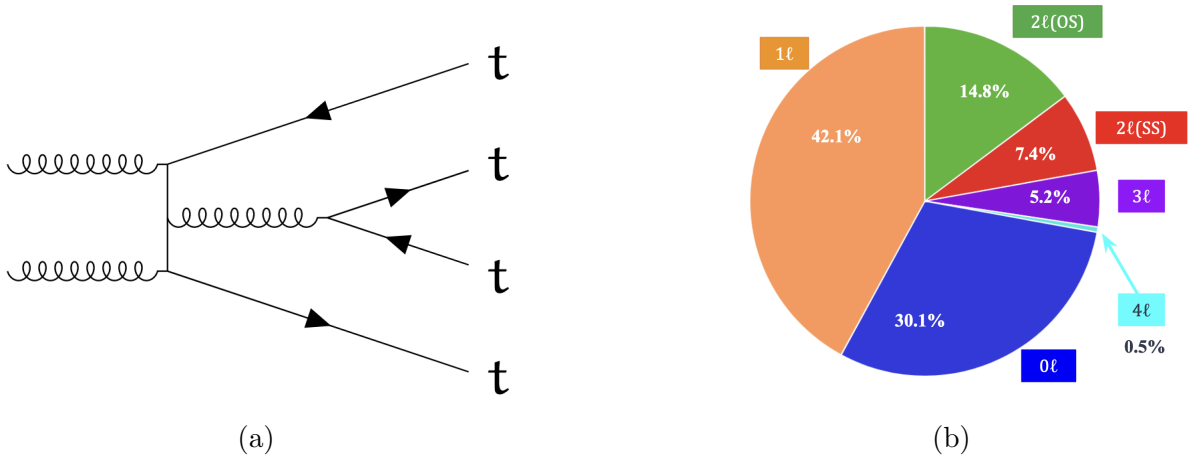
Abstract

The four top discovery reported by both ATLAS and CMS in 2023 represents another confirmation of the Standard Model of particle physics and a possible gateway to physics beyond Standard Model. Because of the high top multiplicity, this rare phenomenon branches out in many final states; one of the channels not included in the observation papers was the final state denoted by the case in which a top and an anti-top decay leptonically, resulting in two leptons having opposite signs (2LOS). The main hardship of this channel is due to the inextricable relation to its main backgrounds, in particular the simultaneous production of a top-antitop pair, commonly referred as $t\bar{t}$. The strategy and development status for a top-tagging approach targeting four top final state in 2LOS is presented.

1 Introduction

The observation of the production of four top quarks, claimed by both ATLAS [1)] and CMS [2)] in 2023, was an important milestone in understanding the Standard Model (SM) of particle physics, the current best description of three out of the four known fundamental interactions, namely the electromagnetic, the strong and the weak forces.

The SM has a high level of predictivity, and it has been extensively experimentally verified with high precision, and all particles included in its framework have been observed, the last one being the Higgs boson whose discovery was reported by both ATLAS and CMS at the Large Hadron Collider (LHC) in 2012. However, the SM is incomplete, for it does not explain many phenomena, like the existence of Dark Matter and Dark Energy, neutrino masses, the matter-antimatter asymmetry and so on. Many theories Beyond the SM (BSM) have been formulated in order to overcome these shortcomings, and the four top final states is considered in many of these as a promising signature to probe them: this measurement of this production process is statistically limited, it has a distinctive signature, it is sensitive to the top-Higgs



(a) (b)

Figure 1: On the left, an example of a Feynman diagram contributing to four quark production. On the right, pie chart representing the possible final states for four top production: all hadronic decay ($0L$); with one, three or four charged leptons ($1L$, $3L$, $4L$) with two charged leptons of opposite ($2LOS$) or same signs ($2LSS$).

Yukawa coupling and many top-philic heavy resonances, and holds a unique sensitivity to four fermions operators in the context of SM Effective Field Theory (SMEFT). Because of the large number of top quarks being produced, it has many final states with very different branching ratios, as shown in Figure 1. Over the years, many searches across many channels have been performed by both ATLAS [3, 4] and CMS [5, 8]. The $2LOS$ final state is one of the most challenging, because it is heavily affected by its main irreducibile background, the production of a top-antitop pair, commonly referred as $t\bar{t}$; in particular, when additional jets are present in the final state, this background process can easily emulate the four top quarks signal, as shown in Figure 2. In the following sections, a Machine Learning approach to

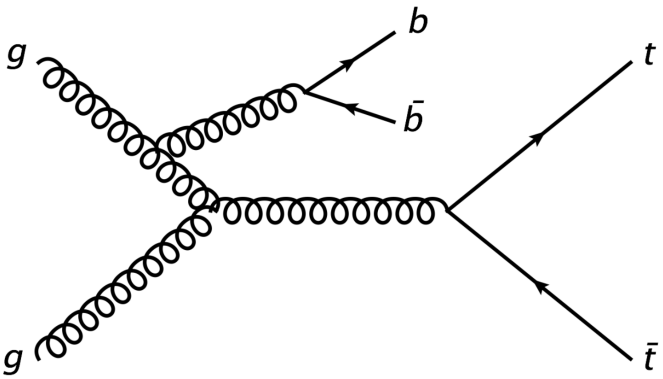


Figure 2: Feynman diagram of the $t\bar{t} + jets$ process, in which the additional jets are coming from the hadronization of 2 bottom quarks.

extract the signal is presented. Two different techniques are described: a Neural Network (NN) trained on event-level information, and an object-level top-tagging approach.

2 Machine Learning approach

This section concerns the use of ML performed to separate the four top signal from the backgrounds in the $2LOS$ final state. NNs are currently being used in two ways: an event-level Multivariate Analysis algorithm (MVA) in order to discriminate the four top signal from the $t\bar{t}$ background by evaluating the information extracted by numerous input features at once; a resolved hadronic top tagger used to gain information on the hadronic component of top quarks present in the signal.

2.1 Event-level MVA

As main background sample, a $t\bar{t} + jets$ is used, and it is separated into different categories according to the flavour of the jets, that is the type of quark that initiated the jet. The NN architecture consists of an input layer, a normalization layer, and 2 dense hidden layers. The output layer is tailored to the regression task, so that the number of outputs corresponds to the number of classes in the training data, namely $t\bar{t}t\bar{t}$, $t\bar{t}b\bar{b}$, $t\bar{t}c\bar{c}$, $t\bar{t}light$. The aim of the network is to classify input data into one of these categories, and to predict the probability of each class. A feedforward NN is built using the Keras [6] package; the number of neurons of the first layer is determined by the number of input features. The first hidden layer has 94 neurons, and the second has 53, in accordance to one of the standard design choices for NNs ($N_{input} * 2/N_{input} + 5$). The two hidden layers and the output layer of the NN use the sigmoid activation function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

In order to minimize the categorical cross-entropy loss function:

$$CE = -\frac{1}{n} \sum_{i} y_i \log(\hat{y}_i(x; \theta)) \tag{2}$$

in which x represents the input feature vector, y_i the target values, $\hat{y}$ the predicted values, and θ the model parameters, the Adam optimizer is used [7]. The training is performed by using 80% of the total dataset, while the remaining 20% is used for the testing process. Table 1 shows the number of entries used during the training. The input features comprehend the kinematic properties of the jets, b-jets, and leptons, the b-tagging scores of the jets, the separations between leptons and/or jets, H_T, and invariant masses and transverse masses of pairs of leptons/jets coming from top quark decays. The model has an early stopping mechanism monitoring the loss over the training; if the loss does not improve for 3 consecutive epochs the training is stopped in order to prevent overfitting and save computational resources. Fig. 3

Category	**Count**
$t\bar{t}t\bar{t}$	1292052
$t\bar{t}light$	890844
$t\bar{t}c\bar{c}$	255187
$t\bar{t}b\bar{b}$	97361

Table 1: Number of entries for each sample.

shows Receiver Operating Characteristic (ROC) curves for each output and their relative area under the ROC curve (AUC) for an inclusive opposite-sign dilepton "signal-like" selection, namely off the Z peak, 4+ jets and 2+ b-tagged jets and a 500 GeV H_T cut.

Furthermore, a Long Short Term Memory layer (LSTM) [9] was added as a strategy to improve the discriminating power of the algorithm. This layer feeds jet features input during the training, in order

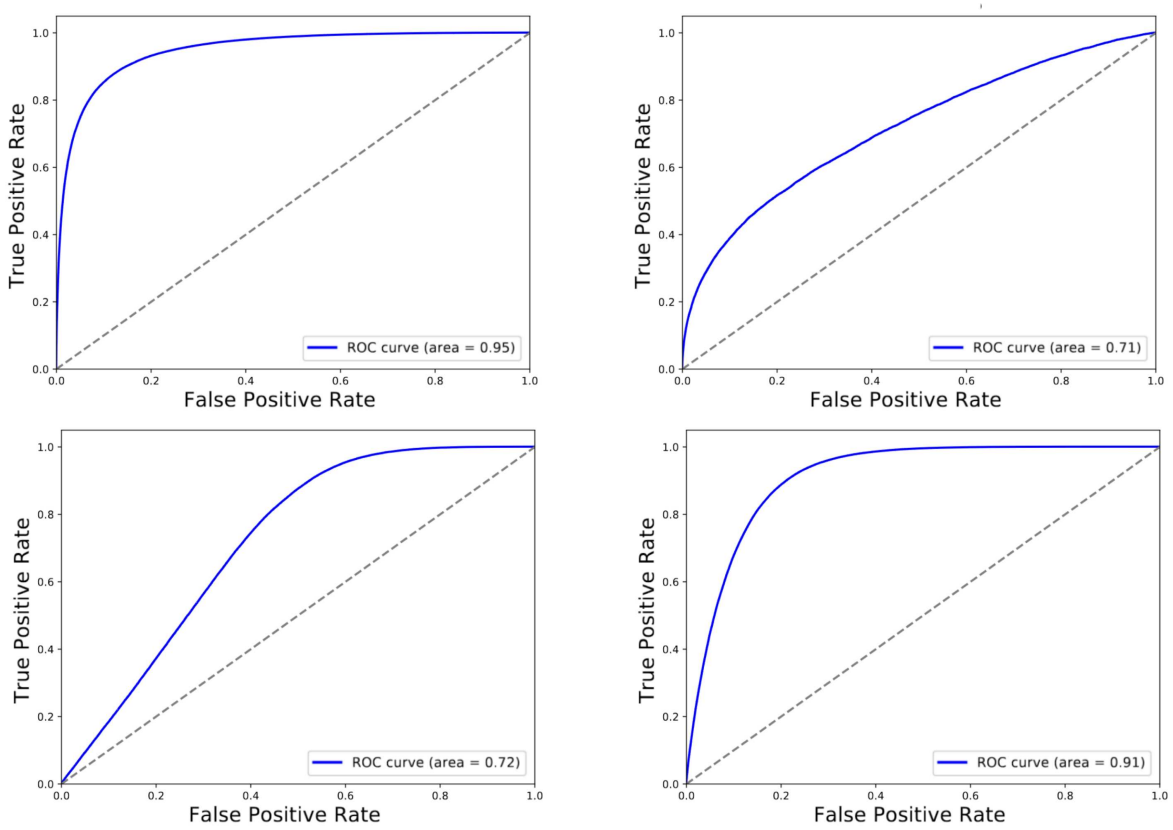


Figure 3: ROC curves for the 4 sample categories; as expected, the score is higher for those categories having a higher number of entries (tttt and ttlight), while the score is relatively low/ biased for the other two

to learn long-term dependencies between the steps; this procedure addresses the recurring problem of vanishing gradients in traditional NNs. The improvement obtained can be seen in Fig.5 . However, in order to extract more information on the hadronic component of the four top process, historically more challenging, a tagging-like strategy is being developed.

2.2 Resolved top tagger

The hadronic top reconstruction was performed for the *resolved* configuration, in which each of the jets (1 b-jet, 2 light jets) emerging from the top quark is matched (or not) to generator level particles. Each triplet of jets that has a match for each and ever one of its components is flagged as a *True* top quark (aka Class 1); on the other hand, those triplets where at least one of the jets is not matched to generator level quarks are flagged as *False* (Class 0). A jet is considered *matched* if the angular distance between the RECO jet and the GenPart quark is below 0.4, namely $\Delta R(j, q) < 0.4$. Two of the most relevant input features to the classifier are shown in Figure 7.

The architecture of the network consists of 2 dense layers with 64 and 32 neurons respectively, that use the ReLU (Rectified Linear Unit) activation function, separated by a Dropout layer, that sets 20% of units to 0 in order to prevent overfitting, and an output layer that uses the Adam optimizer to minimize the binary cross entropy. The input features are standardised using the Robust Scaler from scikit-learn [?];

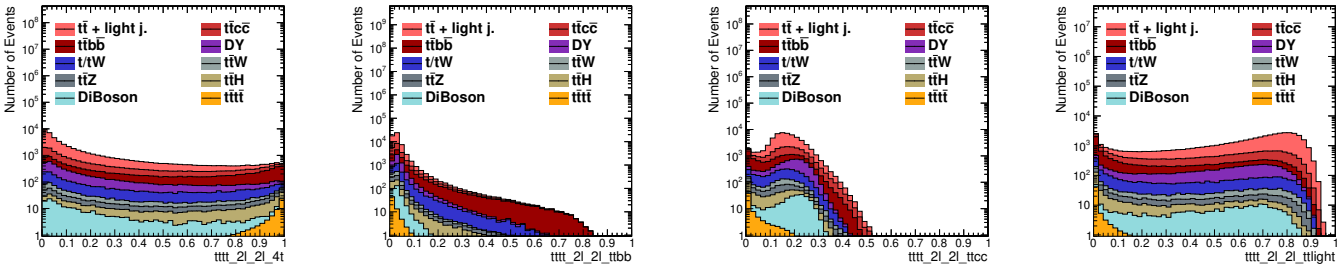


Figure 4: Scores of the event-level MVA for the four classes.

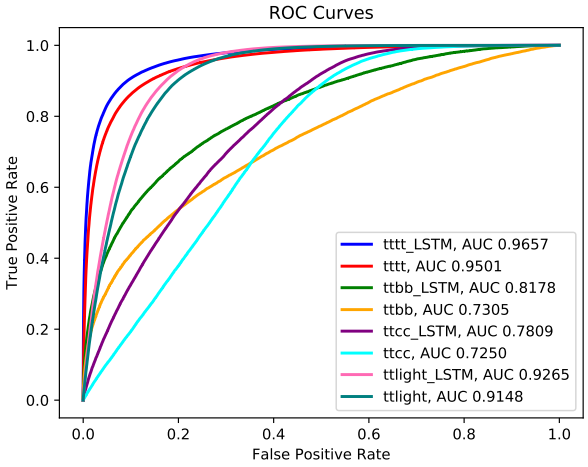


Figure 5: ROCs and AUCs comparisons of the two models (baseline and with the addition of an LSTM layer) for all the categories. The improvement due to the LSTM can be mainly see in the ttbb category.

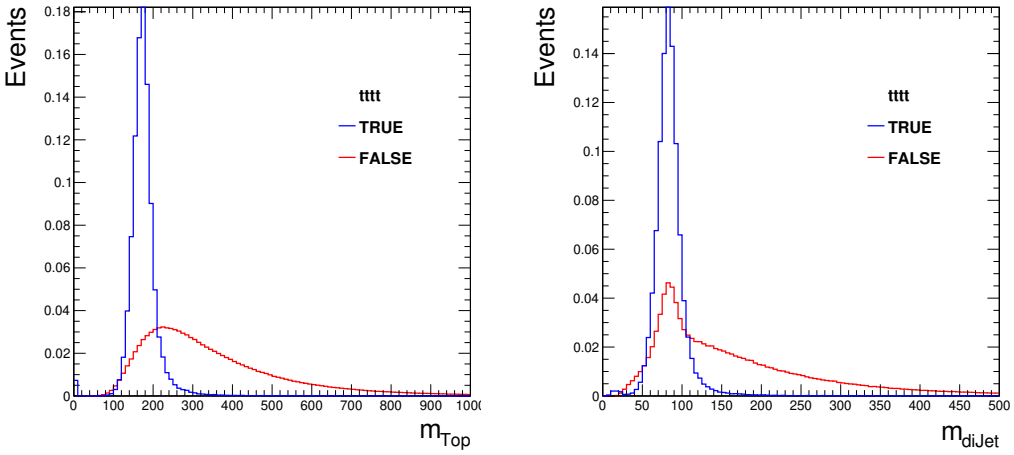
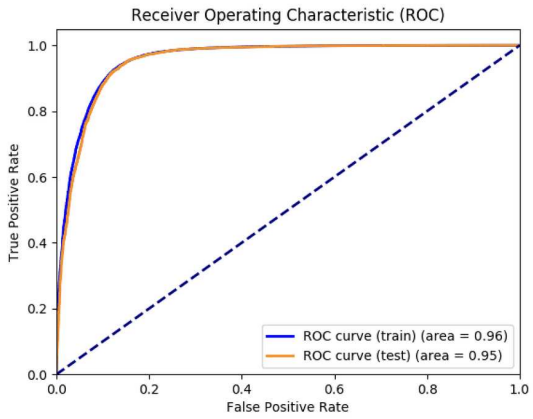
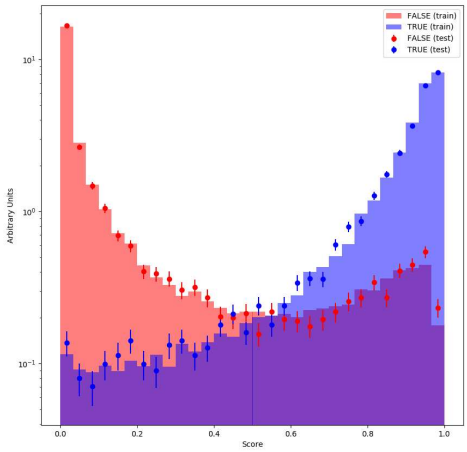


Figure 6: Some input features for the True and Fake top quark categories.

these include the most important kinematic variables for the reconstructed triplets, the W-boson system (i.e. the subsystem reconstructed with the two light jets), the single sub-jets included in the triplets, invariant masses of the many dijet systems, angular separations, and the transverse momentum of the three jets in the top quark centre-of-mass frame. The training was performed with a dataset of 260K entries, equally split between signal and background, obtained from a 2017 four top sample.



(a) ROC curve for the binary classification of true vs false top quarks



(b) NN score for the signal and background categories

Figure 7: On the left, the ROC AUC scores for the training and test sets, both showing a good behaviour in terms of binary classification; on the right, the score distribution for the two classes; blue represents the signal, while represents the combinatorial background from non-matched triplets.

These studies on this resolved objects seem promising, and the efforts in order to understand how to perform the extraction from the $t\bar{t}$ background are ongoing.

References

1. ATLAS Collaboration, Observation of four-top-quark production in the multilepton final state with the ATLAS detector, Eur. Phys. J. C 83 (2023) 496
2. CMS Collaboration, Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Lett. B 847 (2023) 138290
3. ATLAS Collaboration, Measurement of the $\mathrm{tt}^{-}\mathrm{tt}^{-}$ production cross section in pp collisions at $\sqrt{s} =$ 13 TeV with the ATLAS detector, JHEP 11 (2021) 118
4. ATLAS Collaboration, Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, Eur. Phys. J. C 80 (2020) 1085
5. CMS Collaboration, Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Lett. B 844 (2023) 138076
6. Chollet F. et al, Keras, https://github.com/fchollet/keras

7. D. P. Kingma et al., Adam: A Method for Stochastic Optimization, arXiv:1412.6980

8. CMS Collaboration, Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, JHEP 11 (2019) 082

9. Hochreiter, S. and Schmidhuber, J.,Long short-term memory, Neural Computing, Vol.9, 1997