



Pulling the Higgs and top needles from the jet stack with feature extended supervised tagging

J. A. Aguilar-Saavedra^a

Departamento de Física Teórica y del Cosmos, Universidad de Granada, 18071 Granada, Spain

Received: 18 April 2021 / Accepted: 6 August 2021 / Published online: 14 August 2021
© The Author(s) 2021

Abstract Jet tagging has become an essential tool for new physics searches at the high-energy frontier. For jets that contain energetic charged leptons we introduce Feature Extended Supervised Tagging (FEST) which, in addition to jet substructure, considers the features of the charged lepton within the jet. With this method we build dedicated taggers to discriminate among boosted $H \rightarrow \ell\nu q\bar{q}$, $t \rightarrow \ell\nu b$, and QCD jets (with ℓ an electron or muon). The taggers have an impressive performance, allowing for overall light jet rejection factors of $10^4 - 10^5$, for top quark/Higgs boson efficiencies of 0.5. The taggers are also excellent in the discrimination of Higgs bosons from top quarks and vice versa, for example rejecting top quarks by factors of 100–300 for Higgs boson efficiencies of 0.5. We demonstrate the potential of these taggers to improve the sensitivity to new physics by using as example a search for a new Z' boson decaying into ZH , in the fully-hadronic final state.

1 Introduction

From the last decade the Large Hadron Collider (LHC) is probing the high-energy frontier of particle interactions. With the high luminosity achieved, it has been possible to explore the multi-TeV scales not only in the search for new resonances, but also to test the SM production mechanisms at high energy, looking for possible deviations from the predictions of the Standard Model (SM). Being the two most massive SM particles, the Higgs boson and the top quark play a unique role in the search for physics beyond the SM, in particular to probe the electroweak symmetry breaking. The Higgs boson mainly decays hadronically or semileptonically (fully leptonic and diphoton decay modes are rare) while the top quark always produces a b quark in its decay. Therefore, when they are produced with a large boost, their decay products merge into a single jet J .

Jet tagging has witnessed a tremendous progress in the last decade [1–5] (see Ref. [6] for a review). The goal of the different tagging methods is to distinguish a ‘signal’ jet resulting from the hadronic decay of a boosted heavy particle, such as a weak W/Z boson, a Higgs boson, or a top quark, from a ‘background’ quark or gluon jet. The discrimination is done by the analysis of the jet substructure: while the former jets are multi-pronged (containing two or three quarks, depending on the decaying particle) the latter only have one prong. Jet tagging methods have been extensively used, for instance, in searches for new gauge bosons, scalar and spin-2 particles [7–16], vector-like quarks [17–20] and dark matter [21], as well as in SM measurements [22, 23].

Generic supervised taggers have also been developed [24–26] aiming to distinguish arbitrary multi-pronged jets from QCD jets. They have been found capable of separating jets containing ‘prompt’ (produced in the hard process) non-isolated leptons from QCD jets in which the leptons result from the decay of b , c quarks. However, to the best of our knowledge, no tagger has been specifically developed for jets containing such leptons. (Notice, however, that non-isolated leptons are routinely used as one of the ingredients for b -tagging of jets.) This paper aims to fill that gap. We build up on the previously introduced Mass Unspecific Supervised Tagging (MUST) [26] to develop neural network (NN) taggers which, in addition to jet mass, transverse momentum (p_T) and substructure variables, use as input the charged lepton energy fraction $z = E_\ell/E_J$ and the distance from the jet axis in the plane of pseudorapidity (η) and azimuthal angle (ϕ), $\Delta R = (\Delta\eta_{\ell J}^2 + \Delta\phi_{\ell J}^2)^{1/2}$. The method hereby introduced is dubbed as Feature Extended Supervised Tagging (FEST). We build dedicated taggers that can discriminate among $H \rightarrow \ell\nu q\bar{q}$, $t \rightarrow \ell\nu b$ and QCD jets, treating the $\ell = e, \mu$ cases separately. These two examples have the highest interest, since there are numerous measurements and searches by the ATLAS and CMS experiments involving top quarks or Higgs bosons in the boosted regime. We note that

^a e-mail: jaas@ugr.es (corresponding author)

early work [27, 28] pointed out the usefulness of non-isolated leptons for the identification of $t \rightarrow \mu\nu b$. The related lepton p_T fraction $\hat{z} = p_{T\ell}/p_{Tj}$ has been shown [29–31] very useful to discriminate boosted top quarks from QCD jets. A variant, using the lepton p_T fraction with respect to a sub-jet, was explored in Ref. [32], where a detailed study on lepton isolation was also performed. The electron energy fraction has also been indirectly used in Ref. [33].

2 Generating the event samples

The Monte Carlo samples used to train and test the NNs are obtained as follows. Boosted Higgs bosons are generated with MADGRAPH [34], in the SM process $pp \rightarrow ZH$, with $Z \rightarrow \nu\bar{\nu}$ and $H \rightarrow \ell\nu q\bar{q}$. For boosted top (anti-)quarks we use $pp \rightarrow Zt + Z\bar{t}$ mediated by a vector flavour-changing tcZ coupling [35], with $Z \rightarrow \nu\bar{\nu}$ and $t \rightarrow \ell\nu b$. For these processes the top flavour-changing neutral interactions are implemented in FEYNRULES [36] and interfaced to MADGRAPH5 using the universal Feynrules output [37]. QCD jets are generated in the inclusive process $pp \rightarrow jj$, with j a light jet (not including b quarks). A possible extension could include $b\bar{b}$ in the training too; however, the tagger trained on light jets has excellent performance for b jets, as it is explicitly seen in the example presented in Section 6.

Event samples are generated in 100 GeV bins of p_T starting at [300, 400] GeV, and up to to [2.1, 2.2] TeV in the case of QCD samples. For Higgs bosons and top quarks the jet p_T is actually smaller than the p_T of the decaying heavy particle, due to the missing neutrino. Therefore, we extend the generation up to the [2.9, 3.0] TeV and [3.4, 3.5] TeV bins, respectively. This guarantees coverage of the entire jet p_T range up to 2.2 TeV. Even though within each bin the events mainly populate the lower end of the interval, the bins are narrow enough to adequately parameterise the p_T dependence. For testing purposes, $b\bar{b}$ samples are generated using the same p_T binning.

The parton-level event samples so generated are passed through PYTHIA [38] for hadronisation and DELPHES [39] for a fast detector simulation, using the CMS card. Jets are reconstructed with FASTJET [40] applying the anti- k_T algorithm [41] with radius $R = 0.8$, and groomed with Recursive Soft Drop [42]. In the subsequent analysis we only keep jets with groomed mass $m_J \in [40, 170]$ GeV and $p_T \geq 400$ GeV. The chosen mass range encompasses the jet mass distributions for top quark and Higgs boson jets, and the latter cut is imposed in order to have a sufficient boost for top quarks, so that its decay products are contained within a $R = 0.8$ jet. We also ask that the jets contain a charged lepton with $p_T \geq 10$ GeV within a distance $\Delta R = 0.8$ of the jet axis. As discussed in the Appendix, the overall selection efficiencies for jet preselection plus tagging are quite independent

of this mild lower cut. For top and Higgs high- p_T jets, the leptons are already very energetic and the lepton p_T threshold has little influence. On the other hand, for QCD jets a higher threshold at preselection significantly lowers the efficiency. However, the NNs eventually learn that leptons are much softer for QCD jets, and a lower preselection efficiency is compensated by a higher mistag rate by the tagger.

We note that the requirement to contain a lepton, even with a threshold as low as $p_T \geq 10$ GeV, has a very low efficiency for the QCD jet samples. With our simulation we find that, for example, for the sample with $p_T \in [1, 1.1]$ TeV at the partonic level the efficiencies to find an electron or a muon above this threshold are 0.041 and 0.020, respectively. Therefore, huge samples of dijet events are needed to have sufficient statistics: 4×10^5 events per p_T bin for NN training and validation, and 6×10^5 for testing, totaling 19 million jj pairs.

3 Building the taggers

Jet substructure is characterised by the set of N -subjettiness variables proposed in [5],

$$\left\{ \tau_1^{(1/2)}, \tau_1^{(1)}, \tau_1^{(2)}, \dots, \tau_5^{(1/2)}, \tau_5^{(1)}, \tau_5^{(2)}, \tau_6^{(1)}, \tau_6^{(2)} \right\}, \quad (1)$$

computed for the ungroomed jets.¹ By means of a principal component analysis, it can be seen that the number of physically relevant combinations is actually smaller. Still, because the computational speed is not a serious issue, we keep the above set. As done in Ref. [26], we include as NN inputs the jet mass, but varying on a narrower range $m_J \in [40, 170]$ GeV, and the jet $p_T \in [0.4, 2.2]$ TeV. Moreover, as previously pointed out, for these taggers we also include the lepton energy fraction z and ΔR with respect to the jet. A standardisation of the 21 inputs, based on the SM background distributions, is performed to improve the NN learning.

Our goal is to simultaneously discriminate among jets corresponding to Higgs bosons, top quarks and light quarks / gluons. Therefore, we build NNs whose input are the aforementioned variables for jets corresponding to the three classes (H , t , j). The NN output is a list of three numbers (p_1 , p_2 , p_0), with $p_1 + p_2 + p_0 = 1$, giving the probabilities that a jet corresponds to the H , t or j class, respectively. The NNs contain two hidden layers of 512 and 64 nodes, with Rectified Linear Unit (ReLU) activation for the hidden layers and a softmax function for the outputs. The NNs are optimised with the categorical cross-entropy loss function, using the Adam [44] optimiser. Two independent NNs are built, for $\ell = e$

¹ We note that the performance might be improved by using low-level jet substructure variables. For top quarks decaying hadronically, it has been shown [43] that taggers using low-level variables achieve a background rejection ~ 1.4 times larger than taggers using N -subjettiness.

and $\ell = \mu$, using KERAS [45] with a TENSORFLOW backend [46]. The training sets for the e (μ) NN contain 6000 (5000) events from each class (H, t, j) and p_T slice, totaling around 3×10^5 training events. The validation sets used to monitor the NN performance have similar size and composition as the training ones.

For testing, we build additional two-class NNs to discriminate between (i) H and j ; (ii) t and j ; (iii) H and t , using the same architecture except for the loss function, for which we use the (binary) cross-entropy, and the output layer, which only contains one node with a sigmoid activation function. These NNs are trained only using the events corresponding to the two classes (H, j), (t, j) or (H, t), respectively. Furthermore, we also build NNs only using the jet mass and p_T , and the charged lepton z and ΔR as input, to investigate to which extent the jet substructure variables contribute to the discrimination.

Let us finally mention here some checks concerning the NN architecture. We have not found any performance improvement when duplicating the size of the first hidden layer. In previous work [26] we also verified that including higher-order $\tau_n^{(\beta)}$ does not improve the tagger discrimination. We also investigated the possibility of using unbalanced samples in the training, or other generalised loss functions such as the one proposed in [47], without noticeable improvements.

4 Tagger performance

We test the ability of our taggers to discriminate between different pairs of classes, marginalising over the third one. Figure 1 shows the receiver operating characteristic (ROC) curves for H versus j (top), t versus j (middle) and H versus t (bottom). In all plots, the horizontal axis gives the tagging efficiency ε for a given type of jet, and the vertical axis the tagging rejection ε^{-1} for another type of jet. In H versus t we consider t as ‘background’ because Higgs boson production is not usually a background for top quark measurements, but the discrimination can be performed in either way. The ROCs are shown for jets in four p_T intervals: [0.4, 0.6], [0.85, 1.15], [1.35, 1.65] and [1.8, 2] TeV.² The area under the ROC curve (AUC) is very high in all cases, reaching values around 0.998 for H versus j , 0.995 for t versus j and 0.98 for H versus t , for transverse momenta around 2 TeV.

Figure 2 shows the rejection factors ε^{-1} for fixed efficiencies of 0.7, as a function of the jet p_T . The efficiencies are evaluated within intervals of $p_T \in [\langle p_T \rangle - 200, \langle p_T \rangle + 200]$ GeV and plotted as a function of $\langle p_T \rangle$. We also include here lines corresponding to the discrimination against b -quark

² The test samples have a few tens of thousands of events, therefore for rejection factors above 5×10^3 the statistical fluctuations may become important, especially at high transverse momentum.

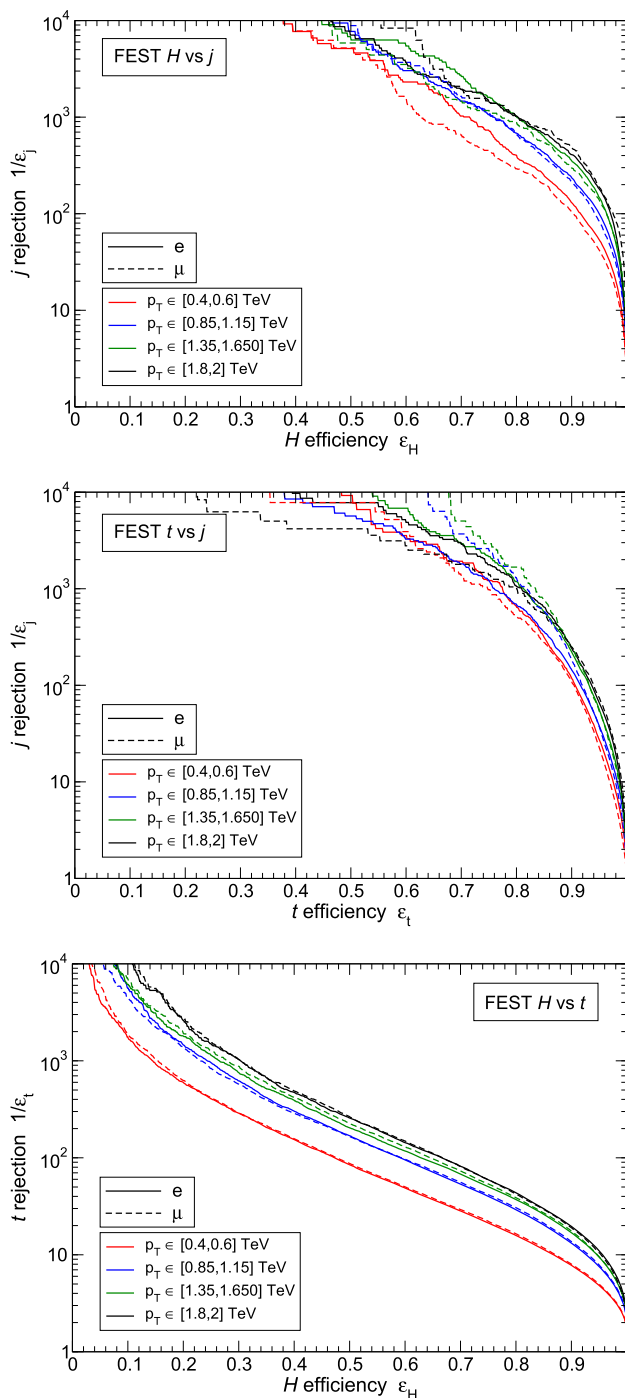


Fig. 1 ROC curves for the separation between different jet classes (for details see the legends and the main text)

jets, which have not been used in the NN training. As it can be readily seen, the discrimination of both H and t jets from b jets is excellent, and likely sufficient to reject backgrounds involving b quarks.

The tagger rejection for QCD jets is impressive. Furthermore, let us remind the reader that the test samples, for which the ROC curves in Fig. 1 and rejection factors in Fig. 2 are

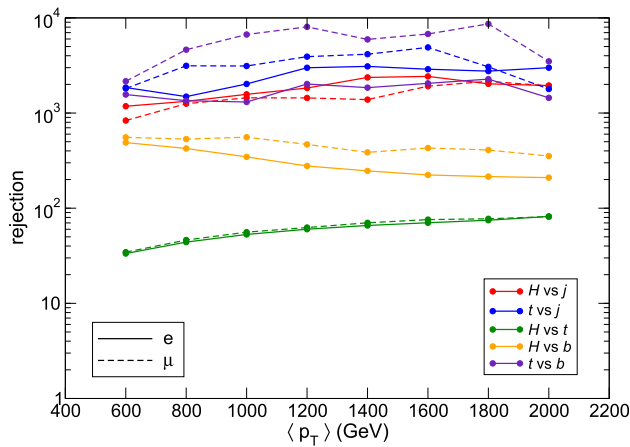


Fig. 2 Rejection factors for H versus j (red), t versus j (blue) and H versus t (green) as a function of transverse momentum for fixed tagging efficiencies of 0.7 (see the text for details)

computed, are composed of jets that already pass the preselection requirement of a charged lepton with $p_T \geq 10$ GeV. And for QCD jets, the efficiency of this lepton requirement is quite small (see the Appendix). For a given overall H efficiency $\bar{\varepsilon}_H$, the overall QCD jet rejection $\bar{\varepsilon}_j^{-1}$ is straightforwardly calculated as follows³:

- By dividing the selected overall efficiency $\bar{\varepsilon}_H$ by the preselection efficiency (either for electrons or for muons) we get a H tagging efficiency ε_H , to which corresponds a j rejection ε_j^{-1} .
- Then, dividing ε_j^{-1} by the preselection efficiency for QCD jets (either for electrons or muons), we obtain the overall QCD jet rejection factor $\bar{\varepsilon}_j^{-1}$.

For example, the preselection efficiencies for $H \rightarrow \ell\nu q\bar{q}$ jets with $p_T \in [1, 1.1]$ TeV are 0.61 and 0.91 in the electron and muon channel, respectively. For QCD jets, they are 0.041 and 0.020. Therefore, considering jets with $p_T \sim 1$ TeV, for an overall efficiency $\bar{\varepsilon}_H = 0.5$, the corresponding light jet rejection factors are

$$e : \varepsilon_H = 0.61 \rightarrow \varepsilon_j^{-1} = 3000 \rightarrow \bar{\varepsilon}_j^{-1} = 7.4 \times 10^4$$

$$\mu : \varepsilon_H = 0.55 \rightarrow \varepsilon_j^{-1} = 4400 \rightarrow \bar{\varepsilon}_j^{-1} = 2.2 \times 10^5$$

These overall rejection factors of the order of 10^5 for QCD jets make the tagger quite useful, even if the decays $H \rightarrow \ell\nu q\bar{q}$ are subdominant.

³ We use a bar to distinguish the overall efficiencies, including preselection, from the tagger efficiencies ε . The overall efficiency for H and t jets is defined relative to the full $H \rightarrow \ell\nu q\bar{q}$ and $t \rightarrow \ell\nu b$ samples (within some p_T range) before preselection, not summing over lepton flavours. Likewise, the overall efficiency for QCD jets is computed relative to the full sample within some p_T range.

Similar comments can be made regarding the top jet discrimination from QCD jets. The preselection efficiencies for t jets with $p_T \in [1, 1.1]$ TeV are 0.73 and 0.80 in the electron and muon channel, respectively. Therefore, for an overall t efficiency $\bar{\varepsilon}_t = 0.5$, the corresponding light jet rejection factors $\bar{\varepsilon}_j^{-1}$ are

$$e : \varepsilon_t = 0.68 \rightarrow \varepsilon_j^{-1} = 2000 \rightarrow \bar{\varepsilon}_j^{-1} = 4.8 \times 10^4$$

$$\mu : \varepsilon_t = 0.62 \rightarrow \varepsilon_j^{-1} = 11,000 \rightarrow \bar{\varepsilon}_j^{-1} = 5.5 \times 10^5$$

As expected, the QCD jet rejection is much larger than for the top fully-hadronic decay. For reference, NN taggers for the hadronic top quark decay mode have a light jet rejection factor of 500 for a top tagging efficiency of 0.5, working in the same p_T range [48, 49]. (Note that neither of these taggers, nor the FEST tagger presented here, use b tagging to identify top quarks.) Of course, the figures are not comparable because they refer to different decay modes. A meaningful comparison can be made considering the improvement on the S/\sqrt{B} ratio (with S standing for signal and B for background) brought by the different taggers, also taking into account the branching ratio for the hadronic and leptonic modes,

$$t \rightarrow q\bar{q}b : \quad \text{Br}(t \rightarrow q\bar{q}b) \frac{\varepsilon_t}{\sqrt{\varepsilon_j}} = 7.5,$$

$$t \rightarrow e\nu b : \quad \text{Br}(t \rightarrow e\nu b) \frac{\varepsilon_t}{\sqrt{\varepsilon_j}} = 12,$$

$$t \rightarrow \mu\nu b : \quad \text{Br}(t \rightarrow \mu\nu b) \frac{\varepsilon_t}{\sqrt{\varepsilon_j}} = 40. \quad (2)$$

With this figure of merit, one can see that tagging the top semileptonic decays with FEST offers much better prospects to probe for new physics.

The discrimination between H and t jets is also excellent, as seen in the lower panel of Fig. 1, and this is of high importance because top quark production may constitute a background to Higgs boson measurements, as will be seen in the $Z' \rightarrow ZH$ example presented in the following.

5 Comparison with two-class taggers

We restrict ourselves to the electron channel and the test interval $p_T \in [0.85, 1.15]$ TeV to compare the three-class tagger discriminating among H , t and j , with less general two-class taggers. The results are shown in Fig. 3. Interestingly, the discrimination power is the same for the three-class and the two-class taggers, with minor differences that may well have a statistical nature. This fact shows that the discrimination power between two given classes is not degraded when building a tagger that simultaneously tries to distinguish among H , t and j .

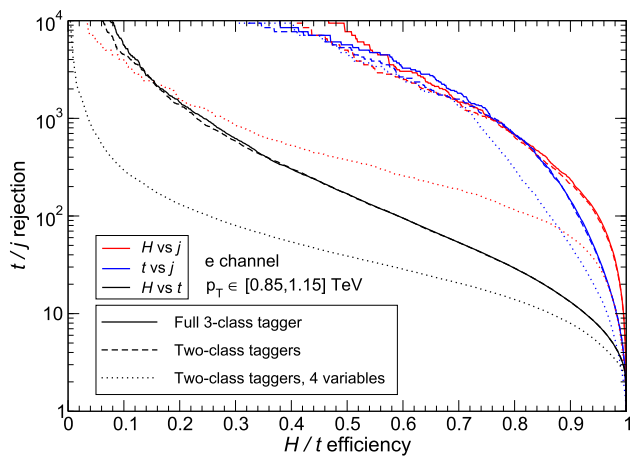


Fig. 3 Comparison of ROC curves for different taggers (for details see the legends)

Because the lepton energy (or p_T) fraction has previously been used as a simple discriminating variable between top quarks decaying semileptonically and QCD jets [27–33], it is worth exploring to which extent the jet substructure variables add to the discrimination. With this purpose, we build two-class taggers that only use as input the jet mass and p_T , as well as z and ΔR . As expected, for $H \rightarrow \ell\nu q\bar{q}$ (with two quarks) the jet substructure significantly enhances the discrimination with respect to light jets. For $t \rightarrow \ell\nu b$, jet substructure variables help but are less important. For H versus t discrimination the analysis of the jet substructure is crucial, as expected, because the former jets have two quarks and the latter only one.

Conversely, as seen in Refs. [25,26], generic taggers only using substructure variables have a poorer discrimination between jets with leptons and QCD jets. The tests in those references are performed using as signal jets from boosted heavy neutrinos decaying $N \rightarrow e\bar{q}\bar{q}$, but the conclusion is expected to be general.

6 Example: $Z' \rightarrow ZH$

We investigate here the usefulness of the taggers here introduced to improve the sensitivity of LHC measurements. Tagging of boosted $H \rightarrow b\bar{b}$ is performed both by the ATLAS and CMS Collaborations by looking at b -tagged subjets of a large-radius jet containing the $H \rightarrow b\bar{b}$ decay products. Namely, the ATLAS Collaboration uses $R = 0.2$ subjets in earlier searches [50] and variable radius jets in the most recent one [51] with the full Run 2 dataset. The CMS Collaboration uses subjets of $R = 0.4$ [52]. Requiring one or two b -tagged subjets significantly suppresses the QCD background, especially in the latter case. The ATLAS Collaboration has considered the decay $H \rightarrow \ell\nu q\bar{q}$ in a search for HH reso-

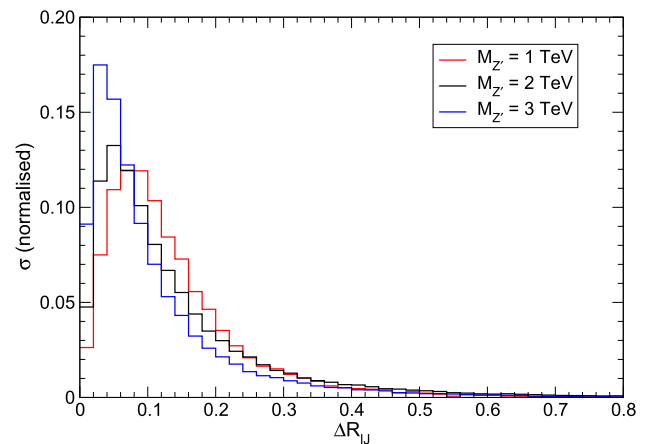


Fig. 4 ΔR separation between the charged lepton and the Higgs jet produced in the $Z' \rightarrow ZH$ decay

nances [53] in the resolved case, where this decay produces two narrow $R = 0.4$ jets and a charged lepton that can be independently reconstructed. As the Higgs bosons are more boosted, the efficiency of the resolved final state decreases and the final state where all H decay products are merged into a single jet becomes more sensitive. This can be seen in Fig. 4, where we show the ΔR separation between the charged lepton and the axis of the jet containing the H decay products in $Z' \rightarrow ZH$, $H \rightarrow \ell\nu q\bar{q}$. We select three different Z' masses to illustrate the dependence on the heavy resonance mass. Because the lepton isolation criterion requires the absence of significant energy in a cone of radius $\Delta R \sim 0.1$ around the charged lepton, the resolved channel is disfavoured for resonances beyond the TeV scale. Future studies are required to compare the sensitivity of the resolved and merged final states for boosted $H \rightarrow \ell\nu q\bar{q}$.

Our goal here is to evaluate the potential sensitivity of new physics searches targeting the $H \rightarrow \ell\nu q\bar{q}$ decay in the merged final state, tagged using FEST. The branching ratio $\text{Br}(H \rightarrow \ell\nu q\bar{q}) = 0.13$ (summing over $\ell = e, \mu$ and lepton charges) is much smaller than $\text{Br}(H \rightarrow b\bar{b}) = 0.58$ [54] but the excellent performance of the FEST tagger makes the decay mode competitive for large luminosities, and especially in final states where the background is large. Otherwise, the large background rejection achieved by FEST is less useful.

We investigate the sensitivity of ZH resonance searches in the decay modes $Z \rightarrow q\bar{q}$, $H \rightarrow \ell\nu q\bar{q}$. This fully-hadronic final state also allows to show the usefulness of the tagger to simultaneously suppress backgrounds with light jets and top quarks – at the end the latter turn out to be the dominant ones. We take as our reference for comparison the search for ZH resonances in the fully-hadronic channel by the ATLAS Collaboration with the full Run 2 dataset [51], focusing on the $Z \rightarrow q\bar{q}$, $H \rightarrow b\bar{b}$ decay modes. Because our results

are obtained with fast simulation, the comparison with the sensitivity achieved in Ref. [51] has the caveat of a possible degradation of the tagger performance in the environment of a real experiment, therefore the comparison has to be taken with a grain of salt.

We perform a simulation including the backgrounds from jj , $t\bar{t}$, Wjj and tW production. Potential backgrounds with fake leptons cannot be handled with the fast simulation, but we expect them not to be dominant. In any case, in an experimental analysis they must be included. The dijet sample is the same one used to test the NN performance, and $t\bar{t}$, Wjj and tW samples are also generated in the same 100 GeV slices of p_T . Samples with $p_T \geq 2.2$ TeV are also generated, and the different samples are combined with weight proportional to the cross section. A 2 TeV $Z' \rightarrow ZH$ signal is generated with $Z \rightarrow q\bar{q}$, $H \rightarrow \ell\nu q\bar{q}$. For $M_{Z'} = 2$ TeV, the 95% confidence level upper limit on the production cross section times decay branching ratio from Ref. [51] is $\sigma(pp \rightarrow Z' \rightarrow ZH) \leq 5.3$ fb. We use this cross section as reference for comparison between the two H decay channels. Events are passed through the simulation chain described before. In addition to $R = 0.8$ jets, we use a collection of ‘track jets’ of radius $R = 0.2$, reconstructed using only tracks. A jet is considered as b -tagged if a b -tagged track jet (using the 70% efficiency working point) within the $R = 0.8$ jet is found.

As event preselection we require two jets with $m_J \geq 40$ GeV, $p_T \geq 400$ GeV and $|\eta| \leq 2.5$. At least one of them is required to have a charged lepton inside the jet. That jet is labeled as the ‘ H ’ jet; if both jets have charged leptons, the one having the lepton with highest z is selected. The remaining jet is labeled as ‘ Z ’. As a proxy for the Z' mass we use the invariant mass of the two jets plus the neutrino, $m_{JJ\nu}$. The neutrino three-momentum is taken parallel to the one of the charged lepton, with its transverse component equal to the missing energy in the event.⁴ The $m_{JJ\nu}$ distribution for the background (overwhelmingly jj) at preselection is shown in Fig. 5, normalised to a luminosity of 139 fb^{-1} .

Before jet tagging, we require a separation $|\Delta\eta| \leq 1.5$ among the two jets, jet masses $m_J \leq 110$ GeV, and perform a b -tag veto on the H jet. These simple cuts reduce the background (which still is dominated by jj production) by a factor of 10 – 100, as shown in Fig. 5.

Finally, tagging of both jets is performed. For the H jet we require probabilities $p_0 \leq 0.01$, $p_2 \leq 0.9$ that the jet corresponds to the j and t class, respectively. For the Z jet we

⁴ We have also explored an alternative neutrino momentum reconstruction, with the longitudinal component and energy determined by requiring that the invariant mass of the neutrino and the H jet equal the Higgs boson mass. This constraint yields a second degree equation; among the two solutions we choose the one that gives smaller longitudinal momentum. The results with this alternative reconstruction are slightly worse.

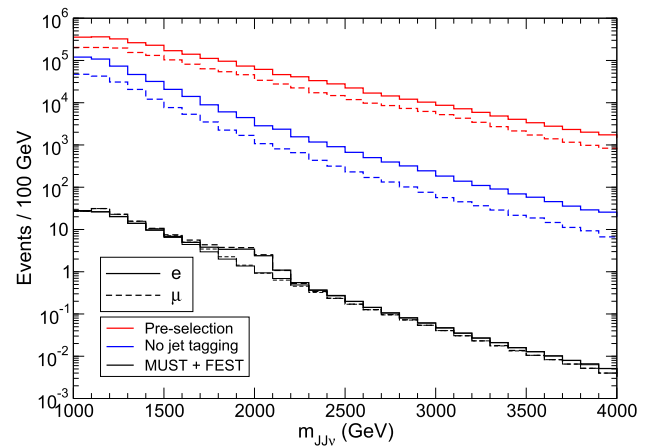


Fig. 5 Reconstructed Z' mass at different levels of event selection, for the background plus an injected signal

use the two-pronged MUST-based tagger T_{2P} developed in Ref. [26], requiring a NN score (quantifying the probability that the jet is two-pronged) $X \geq 0.8$. Tagging the H jet reduces the dijet background by a factor of 2.8×10^{-3} , and tagging the Z jet reduces it by an additional factor of 0.04. Thus, the tagging reduces the background by 3 – 4 orders of magnitude, as shown in Fig. 5, and allows the injected Z' signal to be seen as a bump in the falling $m_{JJ\nu}$ distribution. For clarity, the background-only distributions after tagging are shown as thin lines.

After tagging, the expected number of events for the signal and the different backgrounds near 2 TeV is given in Table 1. Other backgrounds from Zj and Wj production, with Z/W hadronic decay, are less important, and $b\bar{b}$ is even smaller. At the region near 2 TeV, the former two amount to 1/7 and 1/3 of the jj background in the electron and muon channel, respectively, and the latter to 1/20 and 1/9, with the final event selection.

The expected significance of the Z' signal can be computed by performing likelihood tests for the presence of narrow resonances over the expected background, using the CLs method [55] with the asymptotic approximation of Ref. [56]. The local significance at $m_{JJ\nu} = 1.95$ TeV is of 2.2σ in the e channel and 2.4σ in the μ channel, neglecting systematic uncertainties.⁵ Combining both, the local significance reaches 3.2σ . Therefore, even having in mind that the comparison with full simulation is not fair, it seems likely that the sensitivity to $Z' \rightarrow ZH$ may be improved, or at least matched, by the $H \rightarrow \ell\nu q\bar{q}$ decay mode.

⁵ Because the background after event selection at the signal region amounts to a handful of events, we expect background systematic uncertainties to be much smaller than the statistical uncertainty itself. On the other hand, for the signal it is in principle possible to calibrate the tagging efficiency in samples involving boosted Higgs bosons.

Table 1 Expected number of events for the signal and backgrounds in the bins with $m_{JJ\nu} \in [1.9, 2.1]$ TeV, for a luminosity of 139 fb^{-1}

	e	μ
Z'	3.5	3.9
$t\bar{t}$	0.75	1.12
Wjj	0.87	0.87
jj	0.51	0.17
tW	0.18	0.19

7 Concluding remarks

We have developed a three-class tagger to discriminate among boosted $H \rightarrow \ell\nu q\bar{q}$, $t \rightarrow \ell\nu b$, and light jets, with an impressive rejection rate for the latter, and excellent discrimination between top quarks and Higgs bosons. For top quarks, its possible applications are numerous, because the huge rejection factor for light jets overly compensates the smaller semileptonic decay branching ratio. Using as figure of merit the branching ratio times significance improvement, c.f. (2), tagging top quarks in the electron and muon channels improves over the hadronic decay mode previously considered by factors of 1.6 and 5, respectively. For Higgs boson the prospects are quite good too, despite the smaller branching fraction for $H \rightarrow \ell\nu q\bar{q}$.

Our tagger has been built to work on a very wide range of jet $p_T \in [0.4, 2.2]$ TeV. (In contrast, several hadronic top taggers in the literature [48, 49] are trained with jets within a narrow p_T range.) This interval is sufficiently large so as to demonstrate that the tagger can correctly learn to distinguish the differences in jet substructure arising from different p_T regimes and from different jet prongness. Moreover, it has been shown in Ref. [26] that the performance of a tagger trained on wide intervals of jet mass and p_T nearly matches the performance of a tagger trained on narrow intervals. Therefore, the arbitrarily chosen range $p_T \in [0.4, 2.2]$ TeV can be further extended and we do not expect a performance drop.

One possible caveat to the practical application of the tagger is the possible difficulty and uncertainties in the measurement of z and ΔR for electrons embedded within jets, and the possible appearance of fakes. Reference [32] performed a detailed study regarding electron isolation, and there are good prospects that the measurements will be feasible. But even in a worst-case scenario that measurements in the electron channel could not be performed – which, we stress again, seems unlikely – the sensitivity in the muon channel alone is better than in hadronic top decays, c.f. (2), and likewise is expected for Higgs decays, as shown in the previous section.

Generally, one expects that $H \rightarrow \ell\nu q\bar{q}$ and $t \rightarrow \ell\nu b$ with the taggers here introduced will provide the best sensitivity for boosted Higgs boson and top quark measurements,

except at the kinematical end of the spectrum where the background is already quite small. Therefore, for large integrated luminosities, and especially at the high-luminosity upgrade of the LHC, tagging these decay modes may provide the best sensitivity for boosted H , t measurements across a very wide kinematical range.

Finally, let us comment that more generic taggers for jets containing leptons can also be built, which could be sensitive for example to boosted heavy neutrinos decaying $N \rightarrow \ell q\bar{q}$, and may be presented elsewhere.

Acknowledgements I thank J. Aguilar Saavedra for the use of computing resources, and F.R. Joaquim and J. Seabra for previous collaboration in the MUST development. This work has been supported by MICINN project PID2019-110058GB-C21 and by FCT project CERN/FIS-PAR/0004/2019.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: There are no further comments, simply there are no associated data.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³.

Appendix A: Overall performance and event preselection

The tagger is built based on a sample of jets that already contain a charged lepton, with a minimum transverse momentum $p_T \geq 10$ GeV. As it has been argued, the overall performance should have little dependence on this choice, within reasonable limits. In this appendix we explicitly test this, by restricting ourselves to the electron channel and using jet samples that contain electrons with $p_T \geq 20$ GeV. The preselection efficiencies for jets of the three classes are collected in Table 2. The same procedure is followed to train the NN, and the results are compared in Fig. 6 with the results previ-

Table 2 Preselection efficiencies for Higgs (H), top (t) and QCD (j) jets with $p_T \in [1, 1.1]$ TeV, with the requirement to contain an electron with p_T above the given threshold

	H	t	j
$p_T \geq 10$ GeV	0.81	0.73	0.041
$p_T \geq 20$ GeV	0.80	0.72	0.023

ously obtained. We denote by PT10 and PT20 the taggers built using electron thresholds $p_T \geq 10$ GeV, $p_T \geq 20$ GeV, respectively. As expected, the performance in H versus j and t versus j jets in the ROC plots is degraded, since the higher preselection threshold already makes part of the work of the tagger in separating H and t (with energetic electrons) from

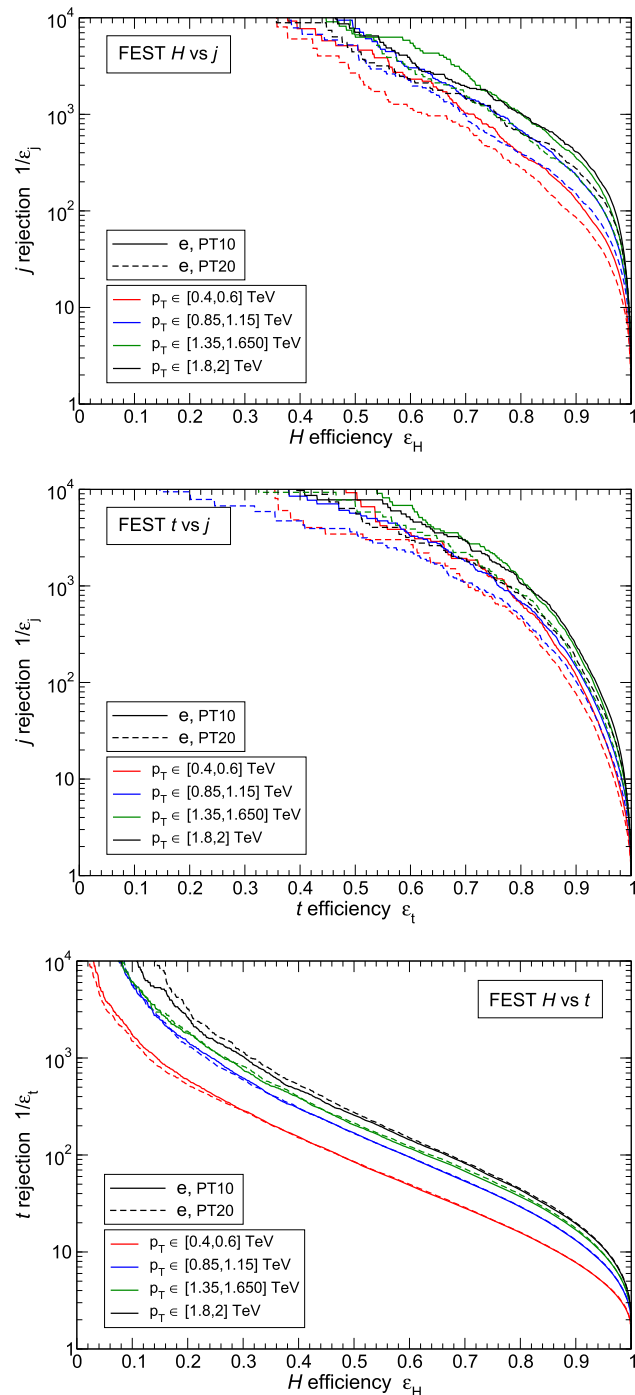


Fig. 6 Comparison of the ROC curves for the separation between different jet classes, for the samples with event preselection $p_T \geq 10$ GeV (as used throughout the paper) and $p_T \geq 20$ GeV

j . Also as expected, the discrimination between H and t is practically unaltered, up to small differences arising from the use of different NNs.

Still, as argued in Sect. 2, the overall performance of the tagger is nearly independent of the lepton p_T threshold. Let us calculate for example the j rejection for jets with $p_T \sim 1$ TeV, for an H overall efficiency $\bar{\epsilon}_H = 0.5$, as done in Sect. 4. For the two taggers, we have

$$\begin{aligned} \text{PT10} : \epsilon_H = 0.61 &\rightarrow \epsilon_j^{-1} = 3000 \rightarrow \bar{\epsilon}_j^{-1} = 7.4 \times 10^4 \\ \text{PT20} : \epsilon_H = 0.63 &\rightarrow \epsilon_j^{-1} = 1900 \rightarrow \bar{\epsilon}_j^{-1} = 8.3 \times 10^4 \end{aligned}$$

The $O(10\%)$ difference in the overall light jet rejection factor is due to statistical fluctuations in the jet samples, caused by the high value of ϵ_j^{-1} . Likewise, can test the light jet rejection for an overall t efficiency $\bar{\epsilon}_t = 0.5$,

$$\begin{aligned} \text{PT10} : \epsilon_t = 0.68 &\rightarrow \epsilon_j^{-1} = 2000 \rightarrow \bar{\epsilon}_j^{-1} = 4.8 \times 10^4 \\ \text{PT10} : \epsilon_t = 0.69 &\rightarrow \epsilon_j^{-1} = 1100 \rightarrow \bar{\epsilon}_j^{-1} = 4.9 \times 10^4 \end{aligned}$$

and in this case the j rejection is nearly the same when using either preselection threshold.

References

1. J.M. Butterworth, A.R. Davison, M. Rubin, G.P. Salam, Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.* **100**, 242001 (2008). [arXiv:0802.2470](#) [hep-ph]
2. J. Thaler, K. Van Tilburg, Identifying boosted objects with N-subjettiness. *JHEP* **03**, 015 (2011). [arXiv:1011.2268](#) [hep-ph]
3. A.J. Larkoski, I. Moult, D. Neill, Power counting to better jet observables. *JHEP* **12**, 009 (2014). [arXiv:1409.6298](#) [hep-ph]
4. I. Moult, L. Necib, J. Thaler, New angles on energy correlation functions. *JHEP* **12**, 153 (2016). [arXiv:1609.07483](#) [hep-ph]
5. K. Datta, A. Larkoski, How much information is in a jet? *JHEP* **06**, 073 (2017). [arXiv:1704.08249](#) [hep-ph]
6. A.J. Larkoski, I. Moult, B. Nachman, Jet substructure at the large hadron collider: a review of recent advances in theory and machine learning. *Phys. Rep.* **841**, 1–63 (2020). [arXiv:1709.04464](#) [hep-ph]
7. A.M. Sirunyan et al., [CMS Collaboration], Searches for W' bosons decaying to a top quark and a bottom quark in proton–proton collisions at 13 TeV. *JHEP* **08**, 029 (2017). [arXiv:1706.04260](#) [hep-ex]
8. A.M. Sirunyan et al., [CMS Collaboration], Search for $Z\gamma$ resonances using leptonic and hadronic final states in proton–proton collisions at $\sqrt{s} = 13$ TeV. *JHEP* **09**, 148 (2018). [arXiv:1712.03143](#) [hep-ex]
9. M. Aaboud et al., [ATLAS Collaboration], Search for $W' \rightarrow tb$ decays in the hadronic final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B* **781**, 327–348 (2018). [arXiv:1801.07893](#) [hep-ex]
10. M. Aaboud et al., [ATLAS Collaboration], Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B* **788**, 316–335 (2019). [arXiv:1801.08769](#) [hep-ex]
11. M. Aaboud et al., [ATLAS Collaboration], Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B* **783**, 392–414 (2018). [arXiv:1804.01126](#) [hep-ex]

12. M. Aaboud et al., [ATLAS Collaboration], Search for heavy resonances decaying to a photon and a hadronically decaying $Z/W/H$ boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D* **98**(3), 032015 (2018). [arXiv:1805.01908](#) [hep-ex]
13. A.M. Sirunyan et al., [CMS Collaboration], Search for low-mass resonances decaying into bottom quark–antiquark pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D* **99**(1), 012005 (2019). [arXiv:1810.11822](#) [hep-ex]
14. A.M. Sirunyan et al. [CMS Collaboration], A multi-dimensional search for new heavy resonances decaying to boosted WW , WZ , or ZZ boson pairs in the dijet final state at 13 TeV. *Eur. Phys. J. C* **80**(3), 237 (2020). [arXiv:1906.05977](#) [hep-ex]
15. A.M. Sirunyan et al., [CMS Collaboration], Search for low mass vector resonances decaying into quark–antiquark pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D* **100**(11), 112007 (2019). [arXiv:1909.04114](#) [hep-ex]
16. G. Aad et al. [ATLAS Collaboration], Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. [arXiv:2004.14636](#) [hep-ex]
17. A.M. Sirunyan et al., [CMS Collaboration], Search for single production of a vector-like T quark decaying to a Z boson and a top quark in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B* **781**, 574–600 (2018). [arXiv:1708.01062](#) [hep-ex]
18. A.M. Sirunyan et al., [CMS Collaboration], Search for single production of vector-like quarks decaying to a top quark and a W boson in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **79**, 90 (2019). [arXiv:1809.08597](#) [hep-ex]
19. M. Aaboud et al., [ATLAS Collaboration], Search for large missing transverse momentum in association with one top-quark in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP* **05**, 041 (2019). [arXiv:1812.09743](#) [hep-ex]
20. A.M. Sirunyan et al., [CMS Collaboration], Search for electroweak production of a vector-like T quark using fully hadronic final states. *JHEP* **01**, 036 (2020). [arXiv:1909.04721](#) [hep-ex]
21. A.M. Sirunyan et al., [CMS Collaboration], Search for dark matter produced in association with a Higgs boson decaying to a pair of bottom quarks in proton–proton collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79**(3), 280 (2019). [arXiv:1811.06562](#) [hep-ex]
22. A.M. Sirunyan et al. [CMS Collaboration], Inclusive search for a highly boosted Higgs boson decaying to a bottom quark–antiquark pair. *Phys. Rev. Lett.* **120**(7), 071802 (2018). [arXiv:1709.05543](#) [hep-ex]
23. A.M. Sirunyan et al. [CMS Collaboration], Inclusive search for highly boosted Higgs bosons decaying to bottom quark–antiquark pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV. [arXiv:2006.13251](#) [hep-ex]
24. J.A. Aguilar-Saavedra, J.H. Collins, R.K. Mishra, A generic anti-QCD jet tagger. *JHEP* **11**, 163 (2017). [arXiv:1709.01087](#) [hep-ph]
25. J. A. Aguilar-Saavedra, B. Zaldivar, Jet tagging made easy. *Eur. Phys. J. C* **80**(6), 530 (2020). [arXiv:2002.12320](#) [hep-ph]
26. J.A. Aguilar-Saavedra, F.R. Joaquim, J.F. Seabra, Mass unspecific supervised tagging (MUST) for boosted jets. *JHEP* **03**, 012 (2021). [arXiv:2008.12792](#) [hep-ph]
27. J. Thaler, L.T. Wang, Strategies to Identify Boosted Tops. *JHEP* **07**, 092 (2008). [arXiv:0806.0023](#) [hep-ph]
28. K. Rehermann, B. Tweedie, Efficient identification of boosted semileptonic top quarks at the LHC. *JHEP* **03**, 059 (2011). [arXiv:1007.2221](#) [hep-ph]
29. J.A. Aguilar-Saavedra, B. Fuks, M.L. Mangano, Pinning down top dipole moments with ultra-boosted tops. *Phys. Rev. D* **91**, 094021 (2015). [arXiv:1412.6654](#) [hep-ph]
30. J. A. Aguilar-Saavedra, Ultraboosted Zt and γt production at the HL-LHC and FCC-hh. *Eur. Phys. J. C* **77**(11), 769 (2017). [arXiv:1709.03975](#) [hep-ph]
31. J. A. Aguilar-Saavedra, M. L. Mangano, New physics with boosted single top production at the LHC and future colliders. *Eur. Phys. J. C* **80**(1), 5 (2020). [arXiv:1910.09788](#) [hep-ph]
32. C. Brust, P. Maksimovic, A. Sady, P. Saraswat, M.T. Walters, Y. Xin, Identifying boosted new physics with non-isolated leptons. *JHEP* **04**, 079 (2015). [arXiv:1410.0362](#) [hep-ph]
33. S. Chatterjee, R. Godbole, T.S. Roy, Jets with electrons from boosted top quarks. *JHEP* **01**, 170 (2020). [arXiv:1909.11041](#) [hep-ph]
34. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP* **07**, 079 (2014). [arXiv:1405.0301](#) [hep-ph]
35. F. del Aguila, J. A. Aguilar-Saavedra, L. Ametller, Zt and γt production via top flavor changing neutral couplings at the Fermilab Tevatron. *Phys. Lett. B* **462**, 310–318 (1999). [arXiv:hep-ph/9906462](#)
36. A. Alloul, N.D. Christensen, C. Degrande, C. Duhr, B. Fuks, FeynRules 2.0: a complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.* **185**, 2250 (2014). [arXiv:1310.1921](#) [hep-ph]
37. C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, T. Reiter, UFO: the universal FeynRules output. *Comput. Phys. Commun.* **183**, 1201 (2012). [arXiv:1108.2040](#) [hep-ph]
38. T. Sjostrand, S. Mrenna, P.Z. Skands, A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.* **178**, 852–867 (2008). [arXiv:0710.3820](#) [hep-ph]
39. J. de Favereau et al., [DELPHES 3], DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *JHEP* **02**, 057 (2014). [arXiv:1307.6346](#) [hep-ex]
40. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. *Eur. Phys. J. C* **72**, 1896 (2012). [arXiv:1111.6097](#) [hep-ph]
41. M. Cacciari, G.P. Salam, G. Soyez, The anti- k_r jet clustering algorithm. *JHEP* **04**, 063 (2008). [arXiv:0802.1189](#) [hep-ph]
42. F.A. Dreyer, L. Necib, G. Soyez, J. Thaler, Recursive soft drop. *JHEP* **06**, 093 (2018). [arXiv:1804.03657](#) [hep-ph]
43. G. Kasieczka et al., The machine learning landscape of top taggers. *SciPost Phys.* **7**, 014 (2019). [arXiv:1902.09914](#) [hep-ph]
44. D.P. Kingma, J.B. Adam: a method for stochastic optimization. [arXiv:1412.6980](#) [cs.LG]
45. F. Chollet, Keras: deep learning for python (2015). <https://github.com/fchollet/keras>
46. M. Abadi et al., TensorFlow: large-scale machine learning on heterogeneous systems (2015). <http://tensorflow.org/>
47. C.W. Murphy, Class imbalance techniques for high energy physics. *SciPost Phys.* **7**(6), 076 (2019). [arXiv:1905.00339](#) [hep-ph]
48. G. Kasieczka, T. Plehn, M. Russell, T. Schell, Deep-learning top taggers or the end of QCD? *JHEP* **05**, 006 (2017). [arXiv:1701.08784](#) [hep-ph]
49. S. Macaluso, D. Shih, Pulling out all the tops with computer vision and deep learning. *JHEP* **10**, 121 (2018). [arXiv:1803.00107](#) [hep-ph]
50. M. Aaboud et al., [ATLAS Collaboration], Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B* **774**, 494–515 (2017). [arXiv:1707.06958](#) [hep-ex]
51. G. Aad et al., [ATLAS Collaboration], Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D* **102**, 112008 (2020). [arXiv:2007.05293](#) [hep-ex]

52. A.M. Sirunyan et al. [CMS Collaboration], Search for a heavy vector resonance decaying to a Z boson and a Higgs boson in proton–proton collisions at $\sqrt{s} = 13$ TeV. [arXiv:2102.08198](#) [hep-ex]
53. M. Aaboud et al., [ATLAS], Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP* **04**, 092 (2019). [arXiv:1811.04671](#) [hep-ex]
54. D. de Florian et al., [LHC Higgs Cross Section Working Group], Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector. [arXiv:1610.07922](#) [hep-ph]
55. A.L. Read, Presentation of search results: the CL(s) technique. *J. Phys. G* **28**, 2693–2704 (2002)
56. G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71**, 1554 (2011). [arXiv:1007.1727](#) [physics.data-an] erratum: *Eur. Phys. J. C* **73**, 2501 (2013)