

Enhancing Jet Taggers with Mass Unspecific Supervised Tagging (MUST)

J. F. Seabra

Departamento de Física and CFTP, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal

Abstract

We introduce a new approach for training jet taggers based on multivariate methods, where the mass and transverse momentum are input variables, along with jet substructure observables, varying over wide ranges. Known as Mass Unspecific Supervised Tagging (MUST), this strategy allows the development of taggers that are sensitive to different types of signal and efficient across large kinematical regions. Additionally, it provides an optimal solution to the mass correlation problem that affects other supervised taggers. We build MUST-inspired generic taggers using neural networks which, when tested with various multi-pronged signals, perform much better than variables commonly used in experiments to discriminate signal from background. These taggers can also spot signals not used in training with a good efficiency. Taggers built upon MUST can be designed to identify the prongness of a jet, which could be extremely useful in a scenario where a new physics signal is discovered.

Keywords: jet tagging, MUST
DOI: 10.31526/ACP.BSM-2021.9

1. INTRODUCTION

As the Large Hadron Collider (LHC) continues to push the high energy frontier of particle physics, the development of new methods and tools for identifying boosted jets is of utmost importance. Consisting in collimated sprays of hadronised quarks and gluons, jets are abundantly produced at proton-proton colliders such as the LHC. While most of them arise from Quantum Chromodynamics (QCD) processes, some may be initiated by heavy Standard Model (SM) particles like the W , Z and Higgs bosons and the top quark, as well as by new particles that are predicted in a lot of theoretical frameworks beyond the SM. In fact, when these particles are produced with large boosts, their hadronic decay products become highly collimated, merging into single jets with a wider radius than QCD jets. This invaluable insight that jets can provide into the observed collision events lead to the creation of several techniques that aim to distinguish QCD jets (background) from those produced in collimated decays of heavy particles (signal). These techniques have been extensively used, not only in new physics searches [1, 2, 3, 4, 6, 7, 9, 8, 5, 10, 11, 12, 13, 14, 15], but also in SM measurements [16, 17].

The procedure for identifying jets comprises two tasks: The first one is to quantify their mass m_J , usually after applying some ‘grooming’ [18, 19, 20, 21] to remove soft collinear radiation; the other one, known as *tagging*, is to infer the number of quarks and gluons clustered inside them (prongs). To accomplish the second task, some tools rely on a single jet substructure variable like a N -subjettiness [22] or energy correlation function [23, 24], whereas others are built upon multivariate methods that receive a set of those variables [25] or jet images [26] as input. In general, the latter are based on machine learning algorithms that can be classified as supervised when their training data is obtained from Monte Carlo (MC) simulations of signal and background events, or unsupervised if taggers are trained directly on data. Despite the strengths and weaknesses of either approach, supervised taggers would certainly be needed to claim a discovery of a new physics signal, since the identification of its origin requires comparison with MC predictions.

When building a tagger, one must have in mind the issues raised by the correlation between jet substructure observables and kinematic variables such as m_J and the transverse momentum, p_T . Decorrelating those variables is crucial in many experimental searches, as it prevents artificial peaks from appearing in the m_J distribution of the SM background after imposing jet substructure constraints, and provides a way of improving its normalisation by imposing sidebands. Several methods to achieve mass decorrelation were already proposed [27, 28, 29, 30, 31], but so far, their implementation in taggers have been leaving a residual dependence on the m_J and p_T training ranges. As a result, the performance of taggers drop when applied to kinematical regions different from those used to train them. One possible solution for this issue could be to assemble an array of taggers in a two-dimensional grid of m_J and p_T to cover larger ranges of those variables. However, that solution is quite complex to develop and it could lead to potential problems with boundary effects.

Up to now, taggers relying on jet substructure have not taken m_J as input variable or have fixed it around some value suitable to tag a specific particle. We adopt a different strategy in this work, in which m_J and p_T become input variables varying over wide ranges. This new approach that we designate by Mass Unspecific Supervised Tagging (MUST) [33] provides a great performance across wide ranges of m_J and p_T , solving simultaneously the mass correlation problem in the best possible way by preserving the shape of the m_J distribution. Moreover, like other supervised taggers [31, 32], those based on MUST can be generic, in the sense that they can discriminate multiple types of signal from background. The most challenging aspect of these tools is generating signal

multi-pronged jets with continuous m_J and p_T distributions. As we will see in the next section, this is accomplished by means of a dedicated MC generator.

To fully demonstrate the benefits MUST can bring, we apply it to the training of NNs. Using these powerful multivariate algorithms, it is possible to design taggers for identifying multi-pronged jets, which outperform those based on single variables used frequently in experiments. The performance of MUST taggers discriminating signals for which they have not been trained is also tested. Finally, we build a prongness selection tagger inspired on the concept of MUST that could be used to identify new physics signals.

2. BUILDING THE TAGGERS

Generating background and signal events for training our supervised taggers is the first step we take¹. The background sample is composed of QCD jets generated with MADGRAPH [34], in the inclusive process $pp \rightarrow jj$. Its events cover the range [200, 2200] GeV for p_T , while $m_J \in [50, 250]$ GeV. Before proceeding, we should emphasise that the ranges we have chosen for the variables p_T and m_J are arbitrary and can be extended.

The generation of signal is more demanding and it is done with a dedicated MC generator. Considering S a scalar particle, we implement in PROTOS [35] the process $pp \rightarrow ZS$ with $Z \rightarrow \nu\nu$. Depending on the type of multi-pronged events that are being generated, the scalar S may decay as follows:

$$\begin{aligned} 2\text{-pronged (2P)} : & \quad S \rightarrow u\bar{u}, S \rightarrow b\bar{b} \\ 3\text{-pronged (3P)} : & \quad S \rightarrow F\nu; \quad F \rightarrow udd, F \rightarrow udb \\ 4\text{-pronged (4P)} : & \quad S \rightarrow u\bar{u}u\bar{u}, S \rightarrow b\bar{b}b\bar{b}, \end{aligned} \tag{1}$$

with F being a colour-singlet fermion. Since we are interested in building generic taggers, it is important to make them as model-agnostic as possible. For that reason, the decays of S and F are implemented with a flat matrix element, so that the decay weight of different kinematical configurations correspond simply to the four, three or two-body phase space. Similarly to background events, all those belonging to signal samples have $p_T \in [200, 2200]$. The mass of S and F is randomly chosen for each event within the interval [30, 400] GeV. In order to guarantee that all decay products are contained in a jet of radius $R = 0.8$, an upper limit is set for the mass of S , namely $M_S \leq p_T R/2$. The parton-level event samples generated with MADGRAPH and PROTOS are passed through PYTHIA [36] for hadronisation and DELPHES [37] for a fast detector simulation, using the CMS card. Jets are reconstructed with FASTJET [38] applying the anti- k_T algorithm [39] with $R = 0.8$, and groomed with Recursive Soft Drop [40].

Given the training events described above, we develop the following supervised taggers:

- a fully generic tagger GenT , trained with all types of multi-pronged events labelled as signal;
- multi-pronged taggers GenT_{4P} , GenT_{3P} and GenT_{2P} , in which the events labelled as signal are just four-, three- and two-pronged jets, respectively;
- a prongness selection tagger, only trained with multi-pronged jet events. In contrast with the taggers mentioned above, the goal of this jet tagger is not to distinguish signal from background events. Instead, four-, three- and two-pronged jet events are assigned to different classes, allowing the tagger to learn to distinguish them.

As we said above, these taggers are built upon NNs where m_J and p_T are input variables, varying over wide ranges. Together with those two kinematical variables, we include 17 N -subjettiness observables,

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

which characterise jet substructure. They are computed for ungroomed jets and we have verified that including higher-order $\tau_n^{(\beta)}$ does not improve the performance of taggers. A standardisation of the 19 inputs is performed, taking into account the SM background distributions.

Regarding the architecture of the NNs, they all have two hidden layers with Rectified Linear Unit (ReLU) as activation function. In the cases of GenT and the prongness selection tagger (GenT_{2P} , GenT_{3P} and GenT_{4P}), those hidden layers have 2048 (1024) and 128 (64) nodes. No relevant performance improvements were found when using more hidden layers or layers with more nodes. All taggers where the events are either signal or background, i.e. GenT , GenT_{2P} , GenT_{3P} and GenT_{4P} , use the sigmoid activation function in the output layer and their optimisation relies on the binary cross-entropy loss function. Since the prongness selection tagger does not have a binary output, those settings cannot be applied to it. Instead, we use Softmax activation to its output layer and categorical cross-entropy loss function to optimise it.

¹In this paper we just present the most important properties of the datasets we use to train our taggers. For more details about their generation, refer to [33].

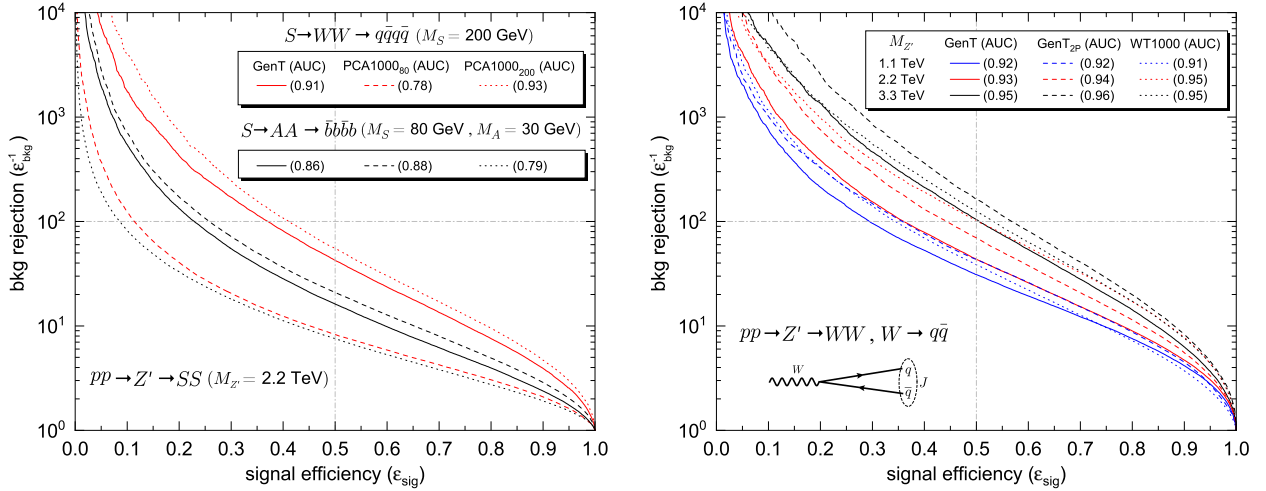


FIGURE 1: Left: comparison of the performances of GenT and taggers pca1000_{80} and pca1000_{200} trained on narrow jet mass and p_T intervals, for two examples of four-pronged jet signals. Right: comparison of the performances of GenT and GenT_{2p} with a tagger WT1000, specifically designed for W bosons.

3. PERFORMANCE OF GENERIC TAGGERS ON MULTI-PRONGED JETS

Throughout this section, our MUST taggers are tested with multi-pronged jet signals produced by W bosons, top quarks and new scalars with various masses. Namely,

- two-pronged, with $W \rightarrow q\bar{q}$; new scalars $A \rightarrow b\bar{b}$ with $M_A = 80$ GeV and $A \rightarrow u\bar{u}$ with $M_A = 200$ GeV;
- three-pronged, with $t \rightarrow Wb \rightarrow q\bar{q}b$;
- four-pronged, with new scalars $S \rightarrow AA \rightarrow b\bar{b}b\bar{b} / u\bar{u}u\bar{u}$ with $M_S = 80$ GeV and $M_A = 30$ GeV [41, 42]; $S \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ with $M_S = 200$ GeV.

All the particles mentioned above are assumed to be produced with a high boost from the decay of a heavy resonance Z' , for which we choose the representative masses $M_{Z'} = 1.1, 2.2, 3.3$ TeV. The background sample used in our tests is composed of quark and gluon jets generated in $pp \rightarrow Zq, pp \rightarrow Zg$, with $Z \rightarrow \nu\nu$, in a 1:1 ratio. All these processes are generated in MADGRAPH and passed through the simulation and reconstruction chain described in the previous section. The tagger performances are evaluated by comparing the efficiencies for signal (ϵ_{sig}) and background (ϵ_{bkg}) within a narrow m_J interval and with a lower cut on p_T , in order to isolate the jet substructure discrimination power from the one obtained with any other variable, such as m_J and p_T . More specifically,

- For signals with $M_{Z'} = 1.1, 2.2, 3.3$ TeV, we set $p_T \geq 0.5, 1.0, 1.5$ TeV, respectively, for both signal and background;
- For W bosons and new scalars decaying as $A \rightarrow b\bar{b} / u\bar{u}$ and $S \rightarrow AA \rightarrow b\bar{b}b\bar{b} / u\bar{u}u\bar{u}$ with $M_{A,S} = 80$ GeV, we use $m_J \in [60, 100]$ GeV; for top quarks, we consider $m_J \in [150, 200]$ GeV; when $A \rightarrow b\bar{b}$ and $S \rightarrow q\bar{q}q\bar{q}$ with $M_{A,S} = 200$ GeV, we take $m_J \in [160, 240]$ GeV.

To analyse the discriminant power for each signal, we show its receiver operating characteristic (ROC) curves and we calculate the area under the ROC curve in the $(\epsilon_{\text{sig}}, \epsilon_{\text{bkg}})$ plane. Vertical lines at signal efficiency of 0.5 and horizontal lines at background rejection of 100 are drawn in all our plots to visualise better the intersection of the ROC curves with these reference values.

To confirm that our taggers discriminate well multi-pronged signals from background across large intervals of m_J and p_T , we compare their performances with those that supervised taggers trained in narrow kinematical regions achieve. For one of these tests, we use the four-pronged signals $S \rightarrow AA \rightarrow b\bar{b}b\bar{b}$ ($M_S = 80$ GeV and $M_A = 30$ GeV) and $S \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ ($M_S = 200$ GeV). We consider in both cases that the resonance Z' initiating those decays has $M_{Z'} = 2.2$ TeV. Following [31], we build the PCA-decorrelated taggers pca1000_{80} and pca1000_{200} , for $p_T \geq 1$ TeV and $m_J \in [60, 100]$ GeV, $m_J \in [160, 240]$ GeV, respectively. The corresponding ROC curves are plotted in Fig. 1 on the left, together with those obtained for the same signals using GenT. By looking at the AUCs shown in the plot, we conclude that taggers trained on a narrow interval of mass close to the signal mass perform slightly better than GenT, but are much less efficient when applied to signals with m_J outside of the training region.

Another comparison we make involves GenT, GenT_{2p} and a PCA-decorrelated tagger we specifically design for identifying W bosons, named WT1000. For this purpose, we generate a sample of $Z' \rightarrow WW$ with $M_{Z'} = 2.2$ TeV and the W bosons decaying hadronically. We train WT1000 using this sample as signal and QCD jets as background, with $p_T \geq 1$ TeV and $m_J \in [60, 100]$ GeV. The NN architecture is the same as the one used for the taggers in [31]. The three taggers used in this study are then applied to $Z' \rightarrow WW$ ($W \rightarrow q\bar{q}$) signals with $M_{Z'} = 1.1, 2.2, 3.3$ TeV. The results are shown in Fig. 1 on the right panel. As expected, they reveal that

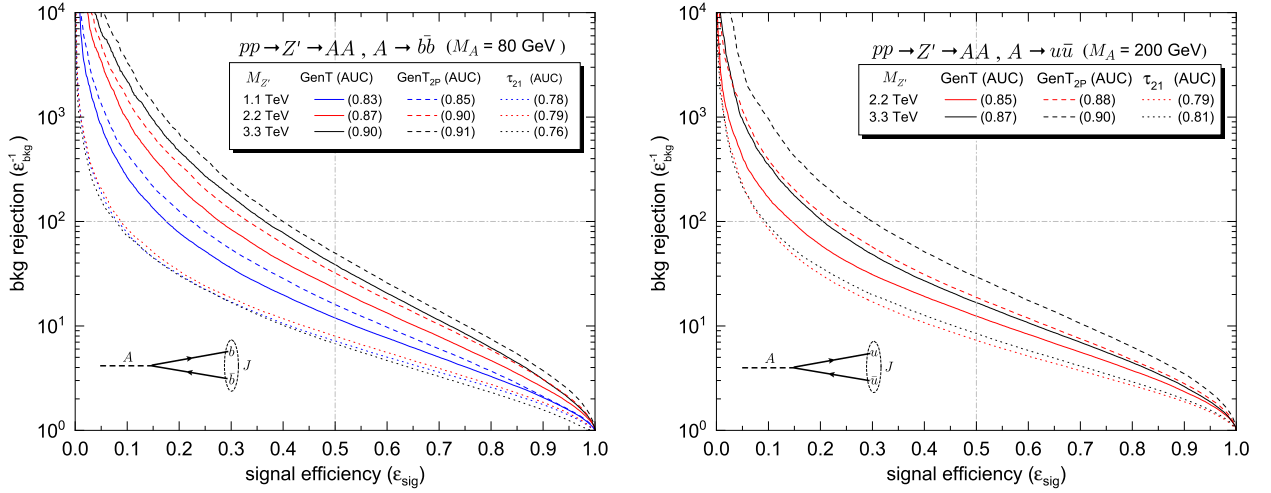


FIGURE 2: ROC curves for two-pronged jet signals (see the self-explanatory legends and the main text for more details).

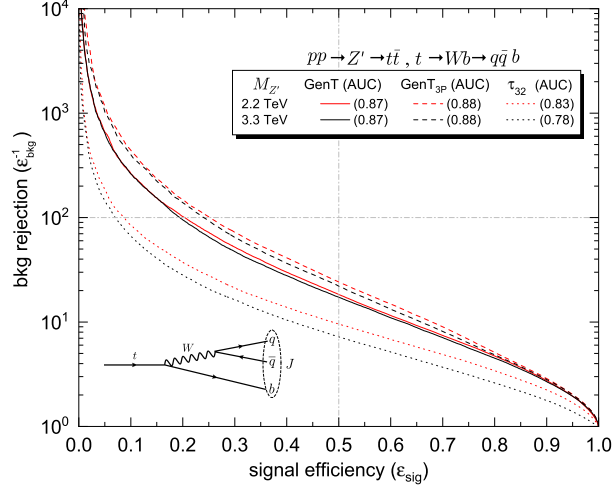


FIGURE 3: ROC curves for top quark jet signals (see the self-explanatory legends and the main text for more details).

- (i) WT1000 performs slightly better than GenT and GenT_{2P} in the kinematical region where it is trained, namely $p_T \geq 1$ TeV and $m_J \in [60, 100]$ GeV;
- (ii) WT1000 performs slightly worse than GenT_{2P} for other momenta, $p_T \geq 0.5$ TeV and $p_T \geq 1.5$ TeV. In the former region of p_T , it also performs slightly worse than GenT.

At this point, we should note that a W tagger could also be built upon MUST to cover W jets in a wider range of p_T . However, we expect little differences between such tagger and GenT_{2P}. According to the results observed in Fig. 1, a dedicated W tagger would have a more uniform presence across all p_T ranges, at the expense of some overall degradation.

Our next goal is to check whether our taggers produce better results than those obtained with N -subjettiness ratios, $\tau_{mn} \equiv \tau_m / \tau_n$, commonly used in experiments. Starting by the analysis of two-pronged events, Fig. 2 shows the performance of GenT and GenT_{2P} for two signals: on the left panel, $A \rightarrow b\bar{b}$ with $M_A = 80$ GeV, and on the right one, $A \rightarrow u\bar{u}$ with $M_A = 200$ GeV. The ROC curves for τ_{21} , used as a discriminator by the CMS Collaboration in [11, 8, 2], are also included in these plots. The results we get with MUST taggers are very good, outperforming τ_{21} in the discrimination of both signals from background. Another interesting conclusion we can take from Fig. 2 (and also from the right plot in Fig. 1) is that the performance of our taggers improve with jet momenta p_T , i.e. with increasing $M_{Z'}$.

The results for the top quark signal are shown in Fig. 3 for a couple of Z' masses, and compared with the ratio τ_{32} , often used as discriminant [1, 12, 13, 14]. GenT and GenT_{3P} perform well on top quark jets, even though fully-dedicated taggers [43] achieve better results. Nevertheless, generic searches using either GenT or GenT_{3P} would not miss top quark signals, which is the point we want to verify here.

Finally, in Fig. 4, some results involving four-pronged signals are shown. For comparison, we show the ROC curves for τ_{42} , since it is the ratio τ_{4n} with highest AUC. The left (right) panel corresponds to the signal $S \rightarrow AA \rightarrow u\bar{u}u\bar{u}$ ($b\bar{b}b\bar{b}$) with $M_S = 80$ GeV

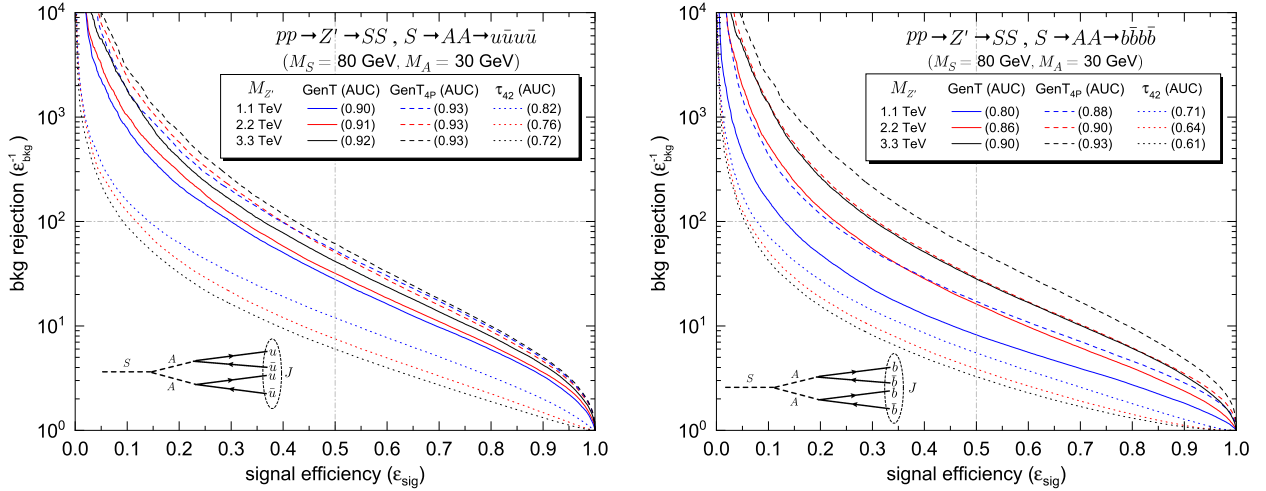


FIGURE 4: ROC curves for four-pronged jet signals (see the self-explanatory legends and the main text for more details).

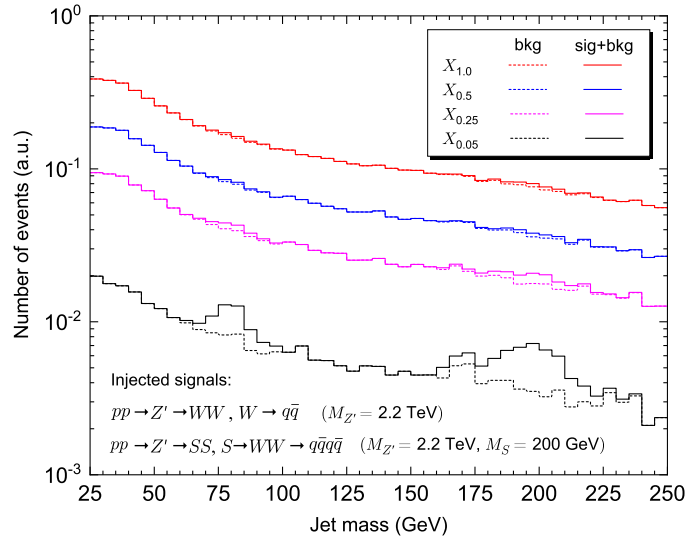


FIGURE 5: Jet mass distribution for the SM background and two injected signals, after applying increasingly tighter cuts on GenT's NN score.

and $M_A = 30$ GeV. The performance of GenT and GenT_{4P} is great, showing (again) better results than the N -subjettiness ratio we have chosen to compare. In contrast with τ_{42} , the performance of MUST taggers improves with increasing $M_{Z'}$, as we have also observed in Fig. 2 with two-pronged jets.

All the tests we have shown in this section reveal, as expected, that multi-pronged taggers (GenT_{2P}, GenT_{3P} and GenT_{4P}) provide a higher discrimination power than the more generic one (GenT) for their corresponding multi-pronged signals. However, it is worth to note that their scope of usability is not as broad as the one that GenT has.

4. MASS DECORRELATION

We have shown throughout the previous section that MUST taggers are sensitive to multi-pronged jet signals across the whole range of m_J and p_T . This means that the SM background shape can be preserved using the varying-threshold scheme employed, for example, by the CMS Collaboration in [7]. We will now demonstrate this with an example, in which we take into account the NN score of GenT for background, as well as for a two-pronged (W) and a four-pronged (S) signal. Considering the variable $\rho = 2 \log m_J / p_T$, we build a two-dimensional grid (ρ, p_T) with $\rho \in [-9, 0]$ in bins of width 0.2 and $p_T \in [0.25, 2.2]$ TeV in bins of 50 GeV. Within each bin, we calculate the 5%, 25% and 50% percentiles of the NN score X , labelled as $X_{0.05}$, $X_{0.25}$ and $X_{0.50}$, respectively. Fig. 5 shows the jet-mass distribution of the SM background for $p_T \geq 1$ TeV plus the two and four-pronged signals that were injected, after applying event selections $X \leq X_{0.05}$, $X_{0.25}$, $X_{0.50}$. The uncut distribution is also shown in the plot and it is represented by $X_{1.0}$. This method keeps the background distribution after each selection and the injected signals show up when the cut is sufficiently tight. Then, we may conclude that our generic taggers provide a perfect solution to the mass correlation problem.

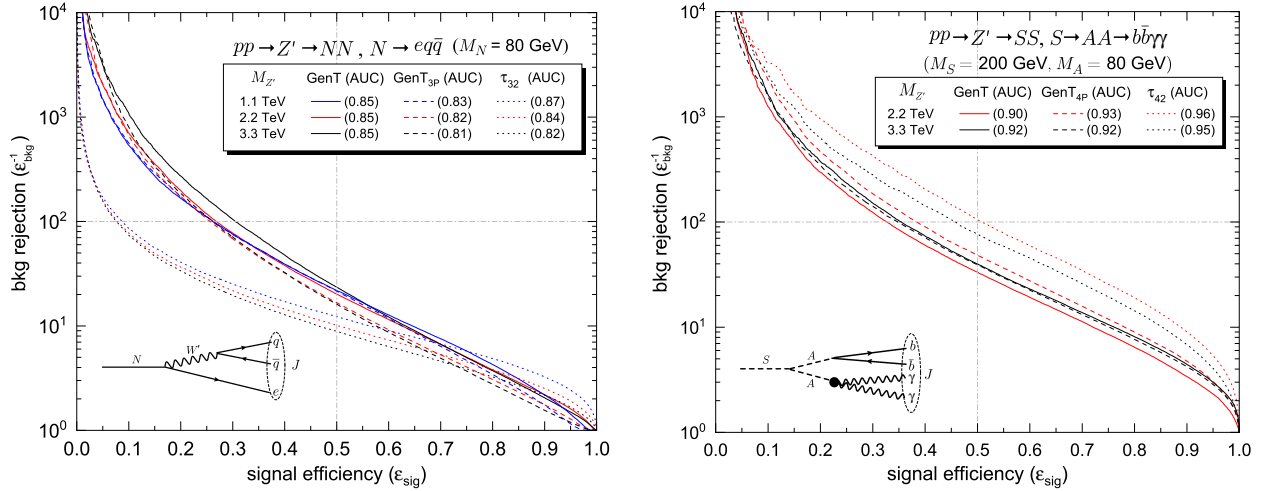


FIGURE 6: ROC curves for complex jet signals not used to train MUST taggers (see the self-explanatory legends and the main text for more details).

5. TAGGER PERFORMANCE FOR JETS NOT USED IN TRAINING

At this point, it is worth to remind that our taggers were trained with multi-pronged jet events containing two or more quarks. All signals used so far to evaluate their performance have that same property, so one might wonder if MUST taggers are capable of discriminating other types of complex signals not used in their training.

In order to address that question, we perform two tests using the following signals:

- jets containing two quarks and a hard electron, resulting from a heavy neutrino decay $N \rightarrow eq\bar{q}$ mediated by an off-shell W' boson, with $M_N = 80$ GeV;
- jets containing two quarks and two hard photons [44], resulting from the decay $S \rightarrow AA \rightarrow b\bar{b}\gamma\gamma$ with $M_S = 200$ GeV and $M_A = 80$ GeV.

In Fig. 6, we present on the left plot the ROC curves for GenT and GenT_{3P}, as well as the ratio τ_{32} for neutrino jets. MUST taggers perform quite well for this kind of signal and their inclusion in experimental searches would improve their sensitivity to new physics. It should also be noted that in this case, AUC is not a good metric for evaluating the performance of taggers. Although τ_{32} performs considerably worse in the region where the signal efficiency is lower, its AUC is still bigger than those we get for GenT and GenT_{3P} because it performs marginally better in the region where $\epsilon_{\text{sig}} \sim 1$. The plot on the right in Fig. 6 presents the ROC curves that correspond to GenT, GenT_{4P} and τ_{42} for jets with two quarks and two hard photons. The results show that MUST taggers can detect those ‘unseen’ signals with a good efficiency. However, simpler multivariate methods like logistic regression [32] may perform better for this kind of jets and taggers trained with them would provide even better discrimination from the background. It is also interesting to observe that the performance of τ_{42} for these complex jets is quite good, in contrast with the results shown in Fig. 4 for true four-pronged jets.

6. PRONGNESS SELECTION TAGGER PERFORMANCE

Assuming that a new physics signal is found and it is related to a boosted particle with multi-pronged jet signature, it might be difficult to discriminate it if the jet was produced by a purely hadronic decay. In this context, a tagger that is capable of identifying the prongness of a jet might be extremely useful. This is the goal of the prongness selection tagger we have built, which is designed to distinguish four, three, and two-pronged jets.

For simplicity, and since we have already shown that MUST taggers can separate multi-pronged jets from background very well, our prongness selection tagger is only trained with the three types of multi-pronged signals that we have also used to train GenT. However, each type of signal belongs now to a different class and consequently, the NN output for each event are the probabilities P_{2P} , P_{3P} and P_{4P} of the jet being two, three, and four-pronged, respectively. The event is naturally assigned to the class with highest probability, $P_{\text{sel}} = \max(P_{2P}, P_{3P}, P_{4P})$. By requiring that P_{sel} is larger than a threshold P_{min} , one may improve the accuracy of the tagger. Thus, in our tests, we make $P_{\text{min}} = 0.5$ and the events where $P_{\text{sel}} < P_{\text{min}}$ are labelled as *Undefined*.

To evaluate the performance of the prongness selection tagger, we observe the following benchmarks:

1. Four-pronged jets produced by a boosted 80 GeV particle: $Z' \rightarrow SS$, $S \rightarrow AA \rightarrow b\bar{b}b\bar{b}$, with $M_{Z'} = 2.2$ TeV, $M_S = 80$ GeV, $M_A = 30$ GeV;
2. Two-pronged jets produced by a boosted 80 GeV particle: $Z' \rightarrow AA$, $A \rightarrow b\bar{b}$, with $M_{Z'} = 2.2$ TeV, $M_A = 80$ GeV;
3. Four-pronged jets produced by a boosted 200 GeV particle: $Z' \rightarrow SS$, $S \rightarrow WW \rightarrow q\bar{q}q\bar{q}$, with $M_{Z'} = 3.3$ TeV, $M_S = 200$ GeV;

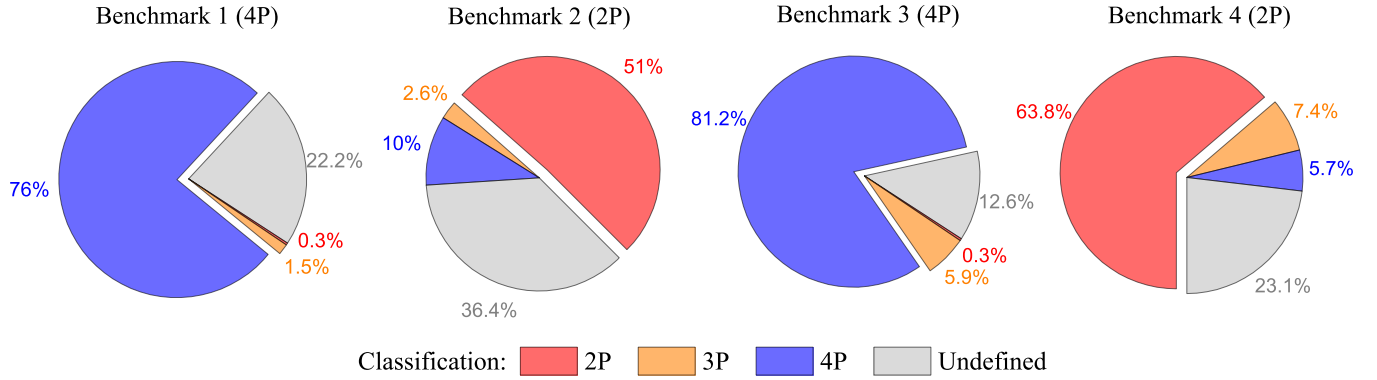


FIGURE 7: Fraction of jets that are classified as 2P, 3P and 4P, or remain undefined, for the four benchmark examples described in the text.

4. Two-pronged jets produced by a boosted 200 GeV particle: $Z' \rightarrow AA$, $A \rightarrow u\bar{u}$, with $M_{Z'} = 3.3$ TeV, $M_A = 200$ GeV;

The pie charts drawn in Fig. 7 show the percentage of events classified as 2P, 3P and 4P, as well the percentage of those that remain undefined. The identification of four-pronged jets is excellent, with an overwhelming fraction of jets correctly assigned and a minor fraction of jets for which $P_{\text{sel}} \geq P_{\text{min}}$ is not verified. The identification of two-pronged jets is quite good as well, with the fraction of correctly identified jets being still several times larger than that of misidentified ones. We also checked that in all benchmarks, it is possible to reduce the mistag rates even further by increasing the value of the threshold P_{min} .

7. CONCLUSIONS

In this work we introduced the method of Mass Unspecific Supervised Tagging (MUST) for identifying multi-pronged signals. Since taggers built upon this method achieve great performances across wide ranges of m_J and p_T , we no longer have the need to build multiple taggers for narrower kinematical regions. They can also spare us from developing taggers that are sensitive to very specific types of jets. In fact, a single generic tagger GenT is able to discriminate four, three and two-pronged jets over the whole range of m_J and p_T , outperforming variables that are often used in experiments. Another great advantage of MUST to keep in mind is the simple solution that it provides to the mass correlation problem. Using the varying-threshold method, it is possible to implement mass decorrelation in MUST-inspired taggers.

We have also verified that our taggers keep a good performance on signals for which they have not been trained, namely jets initiated by heavy neutrinos and jets with energetic photons in the final state. As the possible manifestations of new physics at colliders remain unknown, tools with such a broad sensitivity are extremely valuable. Furthermore, their performance for those signals can be further improved by simply including them in the training set.

Finally, we have built a prongness selection tagger that is able to determine if a signal jet is two, three or four-pronged, with a large likelihood. In a scenario where a new physics signal is found, having a tool with this ability can be crucial to unveil its origin. It is unlikely that discrimination tasks like these may be performed using unsupervised methods. Then, in this respect, MUST taggers could (and hopefully will) play a leading role.

ACKNOWLEDGEMENTS

This work is supported by Fundação para a Ciência e a Tecnologia (FCT, Portugal) through the projects UIDB/00777/2020, UIDP/00777/2020, CERN/FIS-PAR/0004/2019, and PTDC/FIS-PAR/29436/2017. The author's work is supported by the FCT grant SFRH/BD/143891/2019.

References

- [1] A. M. Sirunyan *et al.* [CMS], JHEP **08**, 029 (2017) [arXiv:1706.04260 [hep-ex]].
- [2] A. M. Sirunyan *et al.* [CMS], JHEP **09**, 148 (2018) [arXiv:1712.03143 [hep-ex]].
- [3] M. Aaboud *et al.* [ATLAS], Phys. Lett. B **781**, 327-348 (2018) [arXiv:1801.07893 [hep-ex]].
- [4] M. Aaboud *et al.* [ATLAS], Phys. Lett. B **788**, 316-335 (2019) [arXiv:1801.08769 [hep-ex]].
- [5] M. Aaboud *et al.* [ATLAS], Phys. Lett. B **783**, 392-414 (2018) [arXiv:1804.01126 [hep-ex]].
- [6] M. Aaboud *et al.* [ATLAS], Phys. Rev. D **98**, no.3, 032015 (2018) [arXiv:1805.01908 [hep-ex]].
- [7] A. M. Sirunyan *et al.* [CMS], Phys. Rev. D **99**, no.1, 012005 (2019) [arXiv:1810.11822 [hep-ex]].
- [8] A. M. Sirunyan *et al.* [CMS], Eur. Phys. J. C **80**, no.3, 237 (2020) [arXiv:1906.05977 [hep-ex]].
- [9] A. M. Sirunyan *et al.* [CMS], Phys. Rev. D **100**, no.11, 112007 (2019) [arXiv:1909.04114 [hep-ex]].
- [10] G. Aad *et al.* [ATLAS], [arXiv:2004.14636 [hep-ex]].
- [11] A. M. Sirunyan *et al.* [CMS], Phys. Lett. B **781**, 574-600 (2018) [arXiv:1708.01062 [hep-ex]].
- [12] A. M. Sirunyan *et al.* [CMS], Eur. Phys. J. C **79**, 90 (2019) [arXiv:1809.08597 [hep-ex]].

- [13] M. Aaboud *et al.* [ATLAS], JHEP **05**, 041 (2019) [arXiv:1812.09743 [hep-ex]].
- [14] A. M. Sirunyan *et al.* [CMS], JHEP **01**, 036 (2020) [arXiv:1909.04721 [hep-ex]].
- [15] A. M. Sirunyan *et al.* [CMS], Eur. Phys. J. C **79**, no.3, 280 (2019) [arXiv:1811.06562 [hep-ex]].
- [16] A. M. Sirunyan *et al.* [CMS], Phys. Rev. Lett. **120**, no.7, 071802 (2018) [arXiv:1709.05543 [hep-ex]].
- [17] A. M. Sirunyan *et al.* [CMS], [arXiv:2006.13251 [hep-ex]].
- [18] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].
- [19] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010) [arXiv:0912.0033 [hep-ph]].
- [20] D. Krohn, J. Thaler and L. T. Wang, JHEP **02**, 084 (2010) [arXiv:0912.1342 [hep-ph]].
- [21] For a review see e.g. A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, JHEP **05**, 146 (2014) [arXiv:1402.2657 [hep-ph]].
- [22] J. Thaler and K. Van Tilburg, JHEP **03**, 015 (2011) [arXiv:1011.2268 [hep-ph]].
- [23] A. J. Larkoski, I. Moult and D. Neill, JHEP **12**, 009 (2014) [arXiv:1409.6298 [hep-ph]].
- [24] I. Moult, L. Necib and J. Thaler, JHEP **12**, 153 (2016) [arXiv:1609.07483 [hep-ph]].
- [25] K. Datta and A. Larkoski, JHEP **06**, 073 (2017) [arXiv:1704.08249 [hep-ph]].
- [26] A. J. Larkoski, I. Moult and B. Nachman, Phys. Rept. **841**, 1-63 (2020) [arXiv:1709.04464 [hep-ph]].
- [27] J. Dolen, P. Harris, S. Marzani, S. Rappoccio and N. Tran, JHEP **05**, 156 (2016) [arXiv:1603.00027 [hep-ph]].
- [28] C. Shimmin, P. Sadowski, P. Baldi, E. Weik, D. Whiteson, E. Goul and A. Søgaard, Phys. Rev. D **96**, no.7, 074034 (2017) [arXiv:1703.03507 [hep-ex]].
- [29] S. Chang, T. Cohen and B. Ostdiek, Phys. Rev. D **97**, no.5, 056009 (2018) [arXiv:1709.10106 [hep-ph]].
- [30] T. Dorigo and P. De Castro Manzano, [arXiv:2007.09121 [stat.ML]].
- [31] J. A. Aguilar-Saavedra, J. H. Collins and R. K. Mishra, JHEP **11**, 163 (2017) [arXiv:1709.01087 [hep-ph]].
- [32] J. A. Aguilar-Saavedra and B. Zaldívar, Eur. Phys. J. C **80**, no.6, 530 (2020) [arXiv:2002.12320 [hep-ph]].
- [33] J. A. Aguilar-Saavedra, F. R. Joaquim and J. F. Seabra, JHEP **03**, 012 (2021) [arXiv:2008.12792 [hep-ph]].
- [34] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, JHEP **07**, 079 (2014) [arXiv:1405.0301 [hep-ph]].
- [35] J.A. Aguilar-Saavedra, Protos, a PROgram for TOp Simulations, <http://jaguilar.web.cern.ch/jaguilar/protos/>
- [36] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178**, 852-867 (2008) [arXiv:0710.3820 [hep-ph]].
- [37] J. de Favereau *et al.* [DELPHES 3], JHEP **02**, 057 (2014) [arXiv:1307.6346 [hep-ex]].
- [38] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) [arXiv:1111.6097 [hep-ph]].
- [39] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04**, 063 (2008) [arXiv:0802.1189 [hep-ph]].
- [40] F. A. Dreyer, L. Necib, G. Soyez and J. Thaler, JHEP **06**, 093 (2018) [arXiv:1804.03657 [hep-ph]].
- [41] J. A. Aguilar-Saavedra, Eur. Phys. J. C **77**, no.10, 703 (2017) [arXiv:1705.07885 [hep-ph]].
- [42] J. A. Aguilar-Saavedra and F. R. Joaquim, JHEP **10**, 237 (2019) [arXiv:1905.12651 [hep-ph]].
- [43] S. Macaluso and D. Shih, JHEP **10**, 121 (2018) [arXiv:1803.00107 [hep-ph]].
- [44] J. A. Aguilar-Saavedra and F. R. Joaquim, Eur. Phys. J. C **80**, no.5, 403 (2020) [arXiv:2002.07697 [hep-ph]].