PROCEEDINGS OF SCIENCE

Advances in machine learning tools, software, and calibration for jet-flavor identification in CMS

D. Troiano[a,b,*] **on behalf of the CMS collaboration**

[a] *Department of Physics, University of Bari, Bari, Italy*
[b] *Istituto Nazionale di Fisica Nucleare (INFN)*

E-mail: donato.troiano@cern.ch

Identification of hadronic jets originating from heavy-flavor quarks is essential to several physics analyses in high energy physics, such as studies of the properties of the top quark and the Higgs boson and searches for new physics. Recent algorithms used in the CMS experiment are developed using state-of-the-art machine-learning techniques to distinguish jets emerging from the decay of heavy flavor (charm and bottom) quarks from those arising from light-flavor (udsg) ones. Increasingly complex deep neural network architectures, such as graphs and transformers, have helped achieve unprecedented accuracies in jet tagging. Furthermore, the models are extended also to identify jets originating from hadronic tau leptons and conduct a flavor-aware jet momentum regression. Along with these advances, we present new calibration methods using flavor-enriched selections of proton-proton collision events, which allow us to measure flavor tagging performances in Run 3 of the LHC. We also present modern software and data analysis tools, which allow for a fast and comprehensive development of machine-learning models and the analysis of ever-increasing volumes of data.

The European Physical Society Conference on High Energy Physics (EPS-HEP2025)
7-11 July 2025
Marseille, France

*Speaker

PoS(EPS-HEP2025)643

1. Jet flavor-tagging at CMS

The Compact Muon Solenoid (CMS) experiment [1] clusters hadronic decay products, reconstructed as high-quality Particle-Flow (PF) [2] constituents, with the anti-k_T algorithm [3]. Hadronization processes originated from quark and gluon are reconstructed as jet with a cone radius (R) of 0.4 (AK4 jet). Instead, collimated decays of resonances (W, Z, H) can be identified with a jet of radius 0.8 (AK8 jet). Identification of jets originating from b/c quarks (heavy-flavor tagging) is performed exploiting their long lifetime and looking at displaced charged tracks and secondary vertices (SVs), in addition to all the possible features of the jet constituents, such as the charge and the distance from the jet axis. Modern taggers combine tracking, vertexing and substructure inputs — often via deep learning — to deliver powerful background rejection and enhance sensitivity to both standard model (SM) measurements and beyond SM searches. The most recent taggers developed by the CMS Collaboration are ParticleNet (PNet) [4] and ParticleTransformer (ParT) [5]

2. ParticleNet

PNet is a Dynamic Graph Convolutional Neural Network which represents jets as particles clouds (an unordered set of particles). In particular, the jet is seen as a graph where the nodes correspond to the PF jet constituents and the SVs inside the jet. The core of PNet is EdgeConv [6], a convolution-like operation for point clouds. It builds an edge feature, the distance in the pseudorapidity (η) - azimuthal angle (ϕ) plane to link each node to its nearest neighbors. For each node it looks at the local neighborhood, compares the node own features to its neighbors features, passes those pairwise comparisons through a convolutional neural network, and then pools those neighbor-wise outputs into a single updated feature for the node. The result is a learned, locally-aware representation. Nodes gain information about the pattern of features around them, not just their own raw values.

PNet is defined for both AK4 and AK8 jets. For AK4 jets, PNet outputs a probability vector assigning the jet to one of the following flavour classes: b-jet, c-jet, light-quark jet (uds), gluon jet or hadronic τ-jet. AK8 jets are tagged by ParticleNet-MD (PNet-MD), an algorithm with the PNet architecture trained to be independent on the mass of the jet (mass-decorrelated) [7]. Mass decorrelation is used so that a tagger's output is not strongly correlated with the jet mass. If a classifier is mass-dependent, applying a cut sculpts the QCD background mass distribution and can produce fake bumps in the mass spectrum. PNet-MD outputs a probability vector describing how compatible an AK8 jet is with different signal hypotheses, such as resonances decaying to $b\bar{b}$, $c\bar{c}$, light-flavor (udsg) pairs, or to hadronic and semi-leptonic $\tau\tau$ final states.

2.1 PNet at HLT

Figure 1 shows the performance of PNet and PNet-MD at the High Level Trigger (HLT) level [8] for Run 3 in the identification respectively of b-jets (b-tagging) and resonances decaying into a $b\bar{b}$ pair (bb-tagging). PNet BvsAll = $\frac{p(b)}{p(b)+p(c)+p(uds)+p(g)}$ [1] and PNet-MD HbbVsQCD = $\frac{p(X\to b\bar{b})}{p(X\to b\bar{b})+p(QCD)}$ [2] are respectively the scores used for b-tagging and bb-tagging. Both PNet and PNet-MD were retrained using the 2023 Monte Carlo (MC) simulations. Thanks to the fact that the

[1] p(b), p(c), p(uds) and p(g) are respectively the probabilities that the AK4 jet comes from the hadronization of a b-, c-, light- quark or a gluon.

[2] p($X \to b\bar{b}$) + p(QCD) are respectively the probabilities that the AK8 jet comes from the hadronization of a resonance decaying in a $b\bar{b}$ pair or from a QCD event.

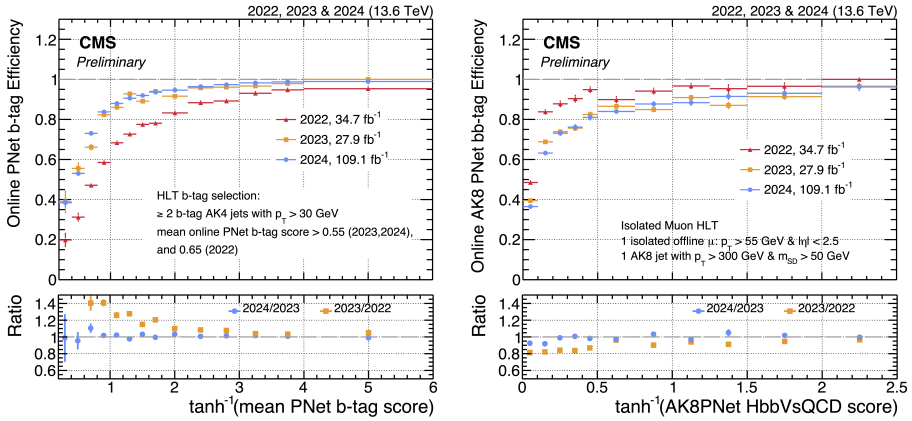

Figure 1: Efficiencies of the online ParticleNet@HLT b-tagging algorithm (left) and the online AK8 ParticleNet@HLT bb-tagging algorithm (right) [10].

models have been adapted to 2023 conditions [9], and by modifying the SV definition for AK8 jets, the b-tagging efficiency improved from 2022 to 2023 up to 40% while bb-tagging misidentification rate was reduced by a factor∼3 from 2022 to 2023, resulting in a reduction in bb-tagging efficiency. Both PNet and PNet-MD provided stable performance between 2023 and 2024.

3. ParticleTransformer

In the ParT representation, a jet is seen as a fully connected graph where all nodes (PF jet constituents and SVs) are connected via the following multiple physics inspired pairwise information (edges):

- $\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}$
- $k_T = min(p_{T,a}, p_{T,b})\Delta$
- $z = min(p_{T,a}, p_{T,b})/(p_{T,a} + p_{T,b})$
- $m^2 = (E_a + E_b)^2 - ||\vec{p}_a + \vec{p}_b||^2$

where y is the rapidity and p_T is the transverse momentum of the constituents. Multiple relationships between nodes are simultaneously tested by a multi-head attention mechanism [11] in order to have a better comprehension of the jet internal structure. ParticleTransformerAK4 [12], RobustParT [14] and UParT [14] are the transformer-based AK4 taggers developed by CMS. ParticleTransformerAK4 is the first one developed and use the ParT architecture to produce the same PNet output probabilities, RobustParT introduced adversarial training [15], described better in the next subsection, to improve robustness, UParT provides a unified multi-task network for simultaneous flavor classification and auxiliary regressions. For what concerns AK8, GloParT [16] is a transformer developed by the CMS collaboration which extends the probability output scores.

3.1 UParT

UParT is the latest CMS jet tagger based on the ParT architecture specifically designed for AK4 jets. It integrates heavy-flavor identification with the jet energy regression and jet energy resolution estimation independent from the flavor of the jet. A principal innovation of UParT lies in its adoption

of a novel adversarial training strategy, enabling the model to effectively learn from perturbations in the input features and thereby enhance its robustness against potential mismodeling in simulations. UParT retains interpretability by using adversarial gradients as a probe of the model sensitivity to inputs. Concretely, the gradient of the loss/logit with respect to the input is computed and analyzed to find out which input features most influence the classifier output by testing how small changes to each input affect the scores. These adversarial gradient maps highlight which constituents and kinematic observables the model relies on for classification, enabling compact feature-importance summaries and helping to link learned representations back to physically meaningful observables. This approach ensures a high heavy-flavor tagging performance, ensuring mismodeling robustness leading to better data/MC agreement. Figure 2 shows the improvement in heavy-flavor tagging and in the jet energy response (ratio between the reconstructed energy and the generated energy) with respect to previous CMS AK4 taggers. In particular UParT reduces PNet b-tagging misidentification by a factor ~2 and improves in the median of the regressed jet energy response ~2% with respect to PNet [13]. Additionally, UParT introduces at the CMS experiment a probability output score that the jet is originating from the hadronization of a s-quark.

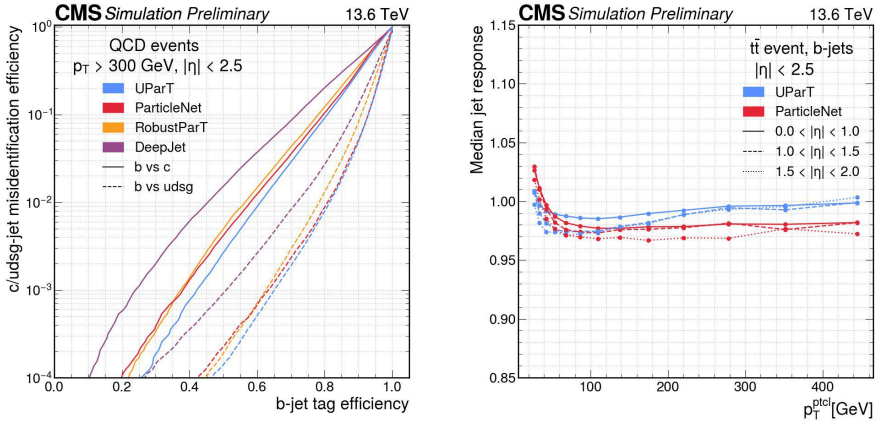

Figure 2: ROC curves for b-tagging (left). UParT shows an improvement both in c- and light-flavor jet misidentification. Median of the raw regressed jet energy response as a function of the generator level jet transverse momentum (p_T^{ptcl}) (right). UParT shows a constant improvement compared to PNet [14].

3.2 GloParT

GloParT is a mass-decorrelated jet classifier derived from the ParT architecture, adapted to identify AK8 jets. GloParT outputs a probability vector over many jet-level hypotheses since it was trained on an expanded set of targets that now includes, for the first time in CMS studies, both fully-hadronic and semi-leptonic $H \to VV$ topologies ($VV \to q\bar{q}q\bar{q}$ or $VV \to q\bar{q}\ell\nu$) alongside the usual hadronic and semi-leptonic top-quark decay modes. This broad coverage allows a single model to perform unified boosted-Higgs and top tagging while sharing representations and reducing the need for separate task-specific networks. GloParT has been deployed by CMS in recent boosted Higgs searches [16] as tagger for $H \to VV$ identification.

4. Calibration validation

Due to imperfect modeling of input variables, discrepancies between data and simulations on the final score are usually observed. These differences appear as altered trends in tagger-score

distributions. Scale factors (SFs) are therefore derived and applied to adjust simulations to data; such corrections are indispensable to preserve the reliability of both inclusive and flavor-aware tagging algorithms. The SFs are evaluated as the ratio between of the tagging efficiency (the efficiency of a jet of flavor f to be tagged, defined as the ratio between the number of tagged and total jets of flavor f) in data and the tagging efficiency in simulation. CMS has developed separate methods for evaluating SFs for AK4 and AK8 jets.

4.1 AK4 calibration

There are two principal calibration strategies for the AK4 SFs evaluation: working-point (WP) and shape-based. The WP method computes a single correction factor for each predefined discriminator threshold (*i.e.* each operating point). Shape-based calibration, by contrast, modifies the entire distribution of the discriminant (for example BvsAll) so that simulation and data are brought into agreement across the full range of values. A comparison between pre- and post- shape-based calibration distributions is shown in Figure 3 where a clear enhancement is observed for the data-MC agreement.

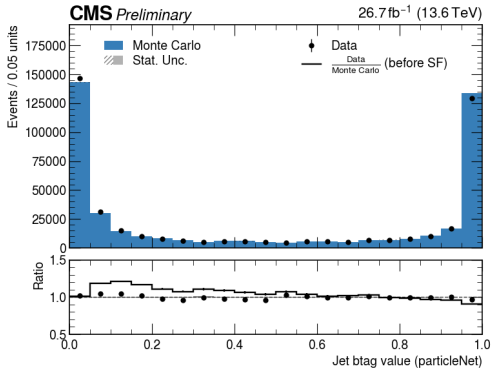

Figure 3: The jets distribution of the PNet BvsAll score for data recorded in late 2022 period, after applying the shape-based SFs. In the lower panel the data-to-simulation ratios before after the SFs application are shown. After the application of SFs, a better agreement is observed [17].

4.2 AK8 calibration

During Run 2 (2016-2018) CMS developed three independent AK8 calibration methods for $X \to b\bar{b}$ and $X \to c\bar{c}$ taggers, to evaluate the SFs in three different p_T ranges [18]. SFs are measured in data control samples that serve as substitutes for $H \to b\bar{b}(c\bar{c})$ jets — *i.e.* samples selected from data that closely resemble the signal jets in kinematics and composition. In the sfBDT method, a boosted decision tree (BDT) selects signal-like $g \to b\bar{b}(c\bar{c})$ jets from QCD multijet events. In the μ-tagged method, jets are obtained selecting QCD jets containing a reconstructed low-p_T muon, since b- and c- quarks have respectively the 20% and the 10% probability of generating a muon or an electron. The boosted $Z \to b\bar{b}$ method selects jets with a high PNet-MD HbbVsQCD score and mass near $m_Z \approx 91.2$ GeV; the observed Z mass peak is then used to extract tagging efficiencies in a kinematic region closely matching the signal. The three measurements are largely orthogonal and were combined to produce the final AK8 tagger scale factors and their uncertainties.

4.3 Run 3 ParticleNet-MD score validation

Given the updated detector conditions, PNet-MD was retrained for Run 3. The PNet-MD HbbVsQCD score was validated using Z→ q$\bar{\text{q}}$ +jets–like events in 2022 data [19]. The samples were

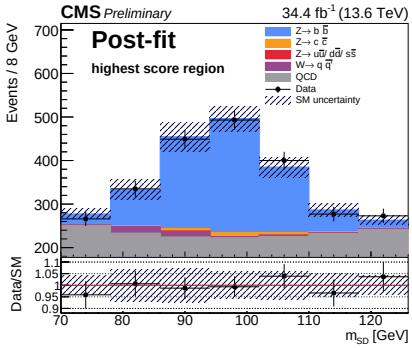

Figure 4: Post-fit data (black markers) to prediction (stacked histograms) comparison for PNet-MD Hbb-VsQCD > 0.988. The lower plot shows the ratio between data and prediction [19].

divided into PNet-MD HbbVsQCD regions and a simultaneous likelihood fit of the soft-drop mass m_{SD} [20] distributions was performed to check the data–simulation agreement. The fit includes the SM Z→ qq̄ signal and W→ qq̄' background from simulation, and the dominant QCD multijet contribution estimated from data. The Z boson mass peak is more distinct as the score increases, indicating that the model correctly discriminates Z→ bb̄ from QCD, as shown in Figure 4 where there is a clear peak.

5. Summary

Modern deep-learning taggers (PNet and ParT families) significantly improve heavy-flavor identification at CMS, including deployment at the HLT with up to ∼40% b-tagging efficiency gains and a ∼3 reduction in bb misidentification between 2022 and 2023, with stable behavior through 2024. Systematic AK4 calibration and complementary AK8 calibration methods (sfBDT, μ-tagged, boosted $Z \to b\bar{b}$) yield reliable SFs. These developments materially enhance CMS reach in analyses relying on heavy-flavor tagging.

After the 2025 EPS conference, CMS provided a comparison between hadronic τ taggers [21] and a new version of UParT (UParT v2) [22]. With respect to UParT, UParT v2 is trained on a substantially larger dataset composed of the same physics processes as UParT, and benefits from architectural refinements, yielding improved accuracy. This enabled UParT v2 to improve b-tagging performance, in particular UParT v2 improves UParT b-tagging misidentification rejection by a factor ∼2. Concerning hadronic τ identification, PNet provides the best rejection of misidentified quark and gluon jets, while UParT v2 provides the best rejection of misidentified muons and electrons.

References

[1] The CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004

[2] The CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003

[3] Matteo Cacciari *et al.*, *The anti-k_t jet clustering algorithm*, *JHEP* **63** (2008) 0804

[4] Huilin Qu and Loukas Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019

[5] Huilin Qu and Congqiao Li and Sitian Qian, *Particle Transformer for Jet Tagging*, *PMLR* **162** (2022) 18281-18292

[6] Yue Wang *et al.*, *Dynamic graph CNN for learning on point clouds*, *ACM Trans. Graph* **38** (2019) 146

[7] Chase Shimmin *et al.*, *Decorrelated Jet Substructure Tagging using Adversarial Neural Networks*, *Phys. Rev. D 96, 074034 (2017)*

[8] Cittolin, Sergio and Rácz, Attila and Sphicas, Paris, *CMS The TriDAS Project*, *Technical design report. CMS* (2002)

[9] The CMS Collaboration, *Novel strategy targeting HH and HHH production at High Level Trigger in Run 3, CMS-DP-2023-050*, CERN, Geneva (2023) https://cds.cern.ch/record/2868787

[10] The CMS Collaboration, *Performance of the ParticleNet b and bb-tagging algorithms in the CMS High-Level Trigger in Run 3, CMS-DP-2025-009*, CERN, Geneva (2024) https://cds.cern.ch/record/2926357

[11] Ashish Vaswani *et al.*, *Attention Is All You Need*, 2017 https://doi.org/10.48550/arXiv.1706.03762

[12] The CMS Collaboration, *Transformer models for heavy flavor jet identification, CMS-DP-2022-050*, CERN, Geneva (2022) https://cds.cern.ch/record/2839920/

[13] The CMS Collaboration, *Jet energy scale and resolution of jets with ParticleNet p_{T} regression using Run3 data collected by the CMS experiment in 2022 and 2023 at 13.6 TeV, CMS-DP-2024-064*, CERN, Geneva (2024) https://cds.cern.ch/record/2904700

[14] The CMS Collaboration, *A unified approach for jet tagging in Run 3 at $\sqrt{s}$=13.6 TeV in CMS, CMS-DP-2024-066*, CERN, Geneva (2024) https://cds.cern.ch/record/2904702

[15] A. Stein, X. Coubez, S. Mondal, A. Novak, A. Schmidt, *Improving Robustness of Jet Tagging Algorithms with Adversarial Training*, Comput Softw Big Sci 6 (2022) 15

[16] The CMS Collaboration, *Search for highly energetic double Higgs boson production in the two bottom quark and two vector boson all-hadronic final state*, CERN, Geneva (2024) https://cds.cern.ch/record/2904879

[17] The CMS Collaboration, *Performance summary of AK4 jet b tagging with data from 2022 proton-proton collisions at 13.6 TeV with the CMS detector, CMS-DP-2024-025*, CERN, Geneva (2024) https://cds.cern.ch/record/2898464

[18] The CMS Collaboration, *Performance of heavy-flavour jet identification in boosted topologies in proton-proton collisions at* $\sqrt{s}$ = 13 TeV*, CMS-PAS-BTV-22-001*, CERN, Geneva (2023) https://cds.cern.ch/record/2866276

[19] The CMS Collaboration, *Performance of boosted* $b\bar{b}$ *jet tagging at* $\sqrt{s}$ *= 13.6 TeV with Run 3 CMS data, CMS-DP-2024-055*, CERN, Geneva (2024) https://cds.cern.ch/record/2904691

[20] A. J. Larkosk *et al.*, *Soft drop*, *JHEP* **05** (2014) 146

[21] The CMS Collaboration, *Comparison of the performance of tau reconstruction and identification algorithms in Run 3, CMS-DP-2025-073*, CERN, Geneva (2025) https://cds.cern.ch/record/2946445

[22] The CMS Collaboration, *Flavour tagging performance of the updated Unified Particle Transformer algorithm with the CMS experiment at* $\sqrt{s}$*=13.6 TeV, CMS-DP-2025-081*, CERN, Geneva (2025) https://cds.cern.ch/record/2948917