ML-Based Top Taggers: Performance, Uncertainty and Impact of Tower & Tracker Data Integration

Rameswar Sahu[a, b], Kirtiman Ghosh[a, b]

[a] Institute of Physics, Bhubaneswar, India, [b] Homi Bhabha National Institute, Mumbai, India.

Abstract

Machine learning algorithms have the capacity to discern intricate features directly from raw data. We demonstrated the performance of top taggers built upon three machine learning architectures: a BDT that uses jet-level variables (high-level features, HLF) as input, while a CNN (miniature version of ResNet) trained on the jet image, and a GNN (LorentzNet) trained on the particle cloud representation of a jet utilizing the 4-momentum (low-level features, LLF) of the jet constituents as input. We found significant performance enhancement for all three classes of classifiers when trained on combined data from calorimeter towers and tracker detectors. The high resolution of the tracking data not only improved the classifier performance in the high transverse momentum region, but the information about the distribution and composition of charged and neutral constituents of the fat jets and subjets helped identify the quark/gluon origin of sub-jets and hence enhances top tagging efficiency. The LLF-based classifiers, such as CNN and GNN, exhibit significantly better performance when compared to HLF-based classifiers like BDT, especially in the high transverse momentum region. Nevertheless, the LLF-based classifiers trained on constituents' 4-momentum data exhibit substantial dependency on the jet modeling within Monte Carlo generators. The composite classifiers, formed by stacking a BDT on top of a GNN/CNN, not only enhance the performance of LLF-based classifiers but also mitigate the uncertainties stemming from the showering and hadronization model of the event generator. We have conducted a comprehensive study on the influence of the fat jet's reconstruction and labeling procedure on the efficiency of the classifiers.

Dataset

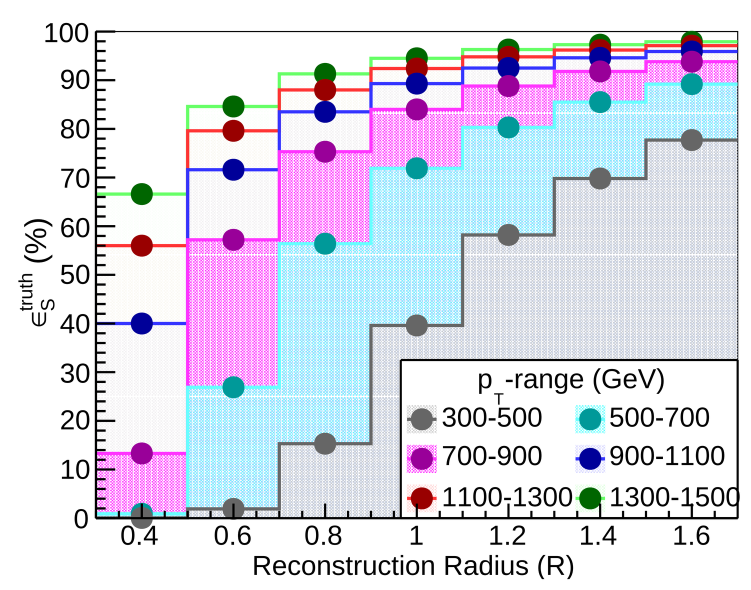
- Signal: Pair production of top quarks with both tops decaying hadronically.

$$pp \to t(\to bW^+(\to qq'))\ \ \bar{t}(\to \bar{b}W^-(\to qq'))$$

- Background: QCD Di-Jet events.

$$pp \to jj$$

- The Fat-Jets are generated in 6 different transverse momentum bins of 200 GeV covering the range 300 GeV to 1500 GeV
- Reconstructed top jets are matched with their partonic counterparts by demanding all three top decay products to lie within the cone of the fat jet. No such matching is performed for the QCD jets.
- The variation of truth level tagging efficiency with reconstruction radius for top jets in different p_T bins :

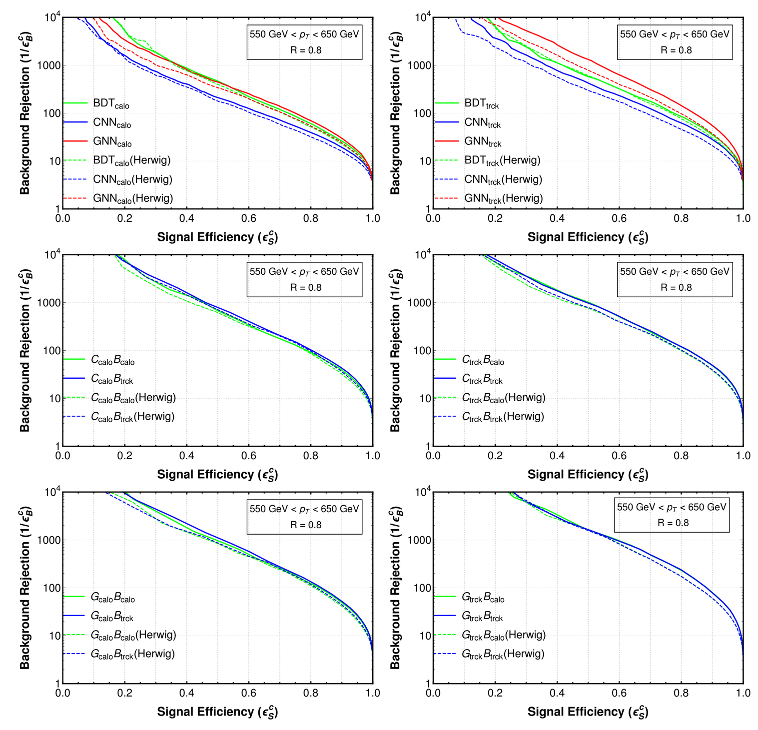


ML - Algorithms

- BDT_{calo} : M, N-subjettiness, b-tag
- BDT_{trck} : BDT_{calo} + additional track based variables.
- CNN_{calo} : Single layered images based on calorimeter energy diposits.
- CNN_{trck} : Two layered images based on calorimeter energy diposits + tracking information.
- GNN_{calo} : Uses jet constituents originating from Calorimeter.
- GNN_{trck} : Uses jet constituents originating from Calorimeter + tracker.
- Composite Classifiers : Uses the score of a CNN/GNN as input variable in a BDT.
- CNN : A 10-layered ResNet, GNN : LorentzNet

Effect of Tracking Information

- The ROC curves of the different classifiers for top and QCD fat jets in the p_T range 550 GeV - 650 GeV :



- We observe a almost 100% improvement in performance in going from $CNN_{calo} \to CNN_{trck}$, $GNN_{calo} \to GNN_{trck}$.
- For BDT_{calo} the improvement is less as variables like M and N-subjettiness already incorporate the tracking information.
- We observe a significant improvement in going from $CNN_{calo} \to C_{calo}B_{calo}$, $CNN_{trck} \to C_{trck}B_{calo}$, $GNN_{calo} \to G_{calo}B_{calo}$, $GNN_{trck} \to G_{trck}B_{calo}$ because of the inclusion of additional HLFs.
- $C_{calo}B_{calo} \to C_{calo}B_{trck}$ and $G_{calo}B_{calo} \to G_{calo}B_{trck}$ show additional 20-30 % improvement in performance due to the inclusion of track based observables.
- $C_{trck}B_{calo} \to C_{trck}B_{trck}$ and $G_{trck}B_{calo} \to G_{trck}B_{trck}$ show no such improvement as the tracking information are alredy present in C_{trck} and G_{trck}.

Dependance on MC generator

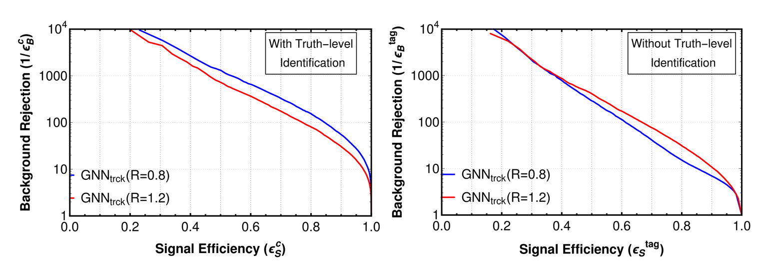
- The background rejection at 70% and 50% signal efficiency for Pythia-generated (Herwig-generated) datasets :

Classifier	$1/\epsilon_B^c(\epsilon_S^c = 0.7)$	$1/\epsilon_B^c(\epsilon_S^c = 0.5)$
BDT_{calo}	119(105)	467(398)
CNN_{calo}	70(57)	211(178)
GNN_{calo}	139(106)	444(341)
BDT_{trck}	175(159)	579(610)
CNN_{trck}	124(90)	423(299)
GNN_{trck}	311(214)	1322(789)
$C_{calo}B_{calo}$	176(175)	682(619)
$C_{calo}B_{trck}$	208(204)	811(737)
$C_{trck}B_{calo}$	249(218)	1023(768)
$C_{trck}B_{trck}$	257(221)	995(799)
$G_{calo}B_{calo}$	260(241)	969(842)
$G_{calo}B_{trck}$	278(256)	1141(894)
$G_{trck}B_{calo}$	489(397)	1641(1604)
$G_{trck}B_{trck}$	493(399)	1736(1666)

- Pythia and Herwig utilize different showering and hadronization models, therefore classifiers like CNN_{trck} and GNN_{trck} that utilize low level information like the four-momentum of jet constituents for training depend strongly on the MC generator.
- However this dependence reduce significantly in composite classifiers like $G_{trck}B_{calo}$ and $G_{trck}B_{trck}$ with the inclusion of additional high level features.

Effect of Reconstruction Radius

- Jets with high p_T are collimated, so a larger reconstruction radius will pick large contribution from background events.
- At the same time jets with low p_T require a larger radius for efficient reconstruction.
- The ROC curves for GNn_{trck} for fat-jets in the 55 GeV - 650 GeV p_T bin reconstructed using R=0.8 and R=1.2 anti-kT jets and matched with(left) and without(right) their partonic counterparts.

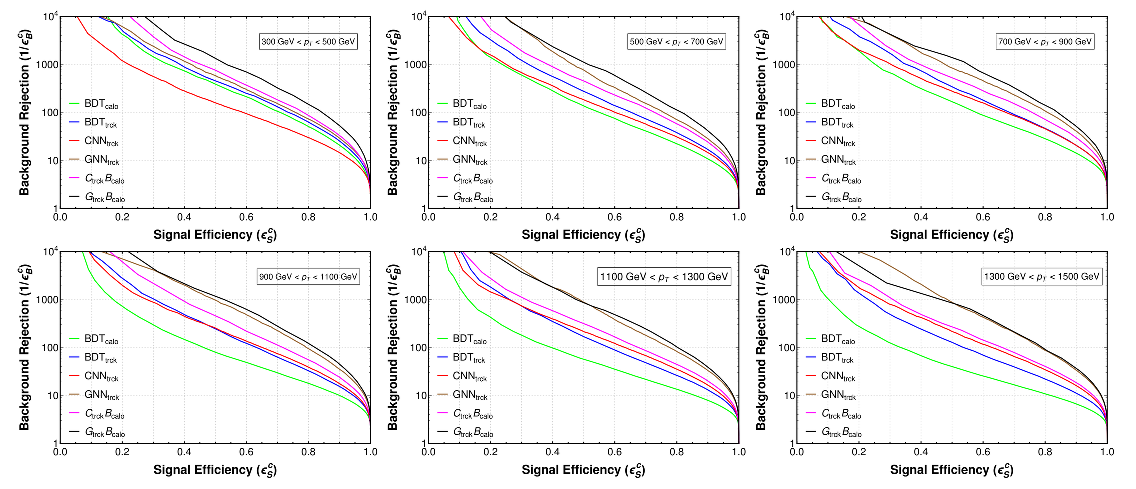


Variable	$1/\epsilon_B^c$ ($\epsilon_s^c = 50\%$)	$1/\epsilon_B^{tag}$ ($\epsilon_s^{tag} = 50\%$)
$R = 0.8$	1298	274
$R = 1.2$	711	424

- Clearly R=0.8 jets show better performance at the classifier level but when used in an actual analysis the performance degrades due to the prencence of a large fraction of fat-jets that are not properly reconstructed.

Final Results

- The ROC corves of the classifiers for the six p_T bins considered in our analysis :



- Note that the fat jets in the p_T range [300, 500] GeV and [500, 700] GeV have different R-parameters ($R = 1.2$) and hence different truth-level identification efficiency than those in the remaining p_T bins where fat jets are constructed with a RR of $R = 0.8$. Therefore, comparing the classifier's performances for fat jets belonging to these two groups is unsuitable.

With Truth Level Tagging

- The Background rejection at 50% signal efficiency of the classifiers for the six p_T bins considered in our analysis :

p_T [GeV]	BDT_{calo}	BDT_{trck}	CNN_{trck}	GNN_{trck}	C_TB_C	G_TB_C
300-500	388	456	159	587	762	1413
500-700	136	276	184	765	455	1178
700-900	168	345	278	845	538	1409
900-1100	79	247	256	971	466	1175
1100-1300	56	167	214	882	318	872
1300-1500	39	127	217	877	273	850

- The invariant mass of the QCD jets scales with p_T and resembles more with that of the top jets resulting in a gradual reductin in performance for BDT_{calo} and BDT_{trck}.
- With CNN_{trck} we see a slight reduction in performance for the last four p_T bins as the top jet images gets more and more collimated with p_T and resemble that of QCD jet images.
- For GNN_{trck}, we see comparable performance in the last few p_T bins.
- In case of CNN_{trck} and GNN_{trck}, the [300, 500] GeV p_T jets have a smaller $1/\epsilon_B^c$ than the [500, 700] GeV p_T jets. This is because an R-parameter 1.2 is inefficient in capturing all the constituents of the [300, 500] GeV fat jets and reduces the performance.
- $CNN_{trck}BDTcalo$ and $GNN_{trck}BDT_{calo}$ show substancial improvement in performance compared to CNN_{trck} and GNN_{trck}.However, this improvement gradually decreases with increasing p_T as the performance of the BDT decreases.

Without Truth Level Tagging

- The Background rejection at 50% signal efficiency of the classifiers for the six p_T bins considered in our analysis (without truth level matching of the test sample):

p_T [GeV]	BDT_{calo}	BDT_{trck}	CNN_{trck}	GNN_{trck}	C_TB_C	G_TB_C
300-500	95	119	54	121	157	250
500-700	83	152	110	303	243	581
700-900	84	166	147	421	258	582
900-1100	57	148	168	534	279	789
1100-1300	45	124	157	540	234	651
1300-1500	34	101	167	609	217	662

- The performance falls substancially compared to the previous case and the fall in performance is proportional to the truth level tagging efficiency.

Summary

- We found a significant increase in the classifier's performance due to including the jet constituents' tracking data for charged constituents in the training and testing process.
- This performance enhancement can be attributed to the fact that jets initiated by light quarks or gluons exhibit distinct differences in the distribution and composition of charged and neutral hadrons. Consequently, information about the charged and neutral constituents of a jet in the form of tracking and tower data helps identify the quark/gluon origin of sub-jets within a fat jet and hence enhances top tagging efficiency.
- It is important to note that despite their high performance, LLF-based classifiers like GNN_{trck} have a significant drawback: they are heavily dependent on the jet modeling provided by the Monte Carlo simulator, such as Pythia or Herwig, which introduces substantial systematic uncertainties.
- Strict reconstruction and identification criteria increase the purity of the sample, simultaneously decreasing truth level identification efficiency (ϵ_S^{truth}). A classifier trained on such pure samples is biased, and the performance cannot be efficiently generalized to new unseen data. We showed that properly selecting the reconstruction radius can improve the ϵ_S^{truth} and help mitigate this issue.

References

- (1) S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, An efficient Lorentz equivariant graph neural network for jet tagging, JHEP 07, 030, arXiv:2201.08187 [hep-ph].
- (2) K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition 10.1109/CVPR.2016.90 (2015), arXiv:1512.03385 [cs.CV]..
- (3) A. Butter et al., The Machine Learning landscape of top taggers, SciPost Phys. 7, 014 (2019), arXiv:1902.09914 [hep-ph]..

ICHEPAP 2023, SINP Kolkata (Dec 11 to Dec 15, 2023)