Mass-unspecific classifiers for mass-dependent searches

J.A. Aguilar Saavedra, **S. Rodríguez Benítez**
Instituto de Física Teórica, UAM/CSIC, Madrid

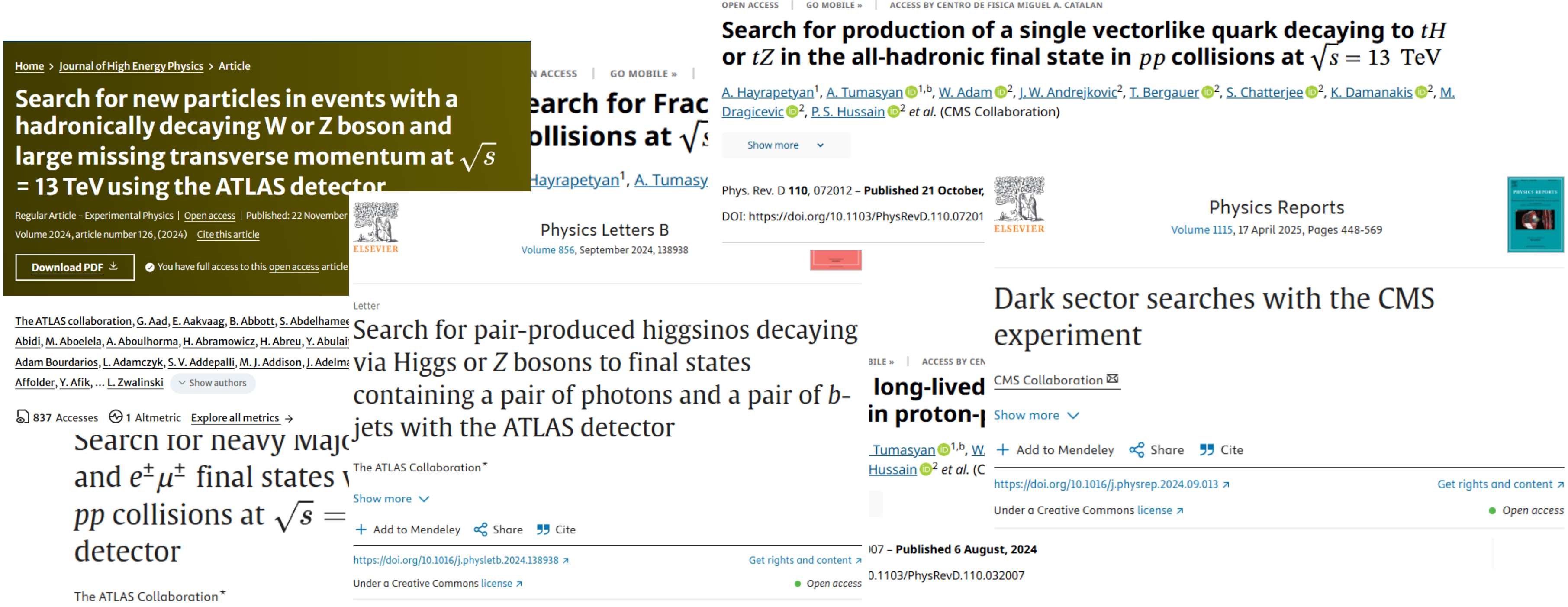
Based on: 2503.20926

projects PID2022-142545NB-C21, CEX2020-001007-S and grant PRE2022-101860 funded by MCIN/AEI/10.13039/501100011033

Motivation

Motivation

Collaborations have been extensively searching for **new physics** at colliders for years, continually improving their analysis techniques.

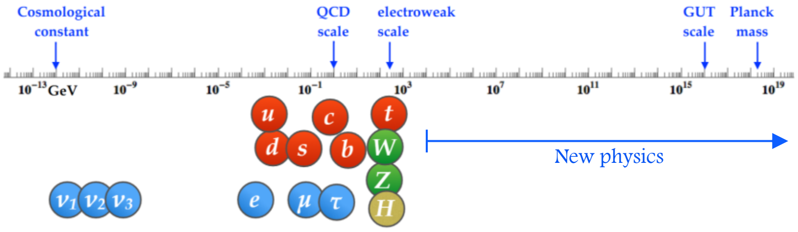


Motivation

Machine learning techniques play a critical role in experimental analyses. But how can we enhance existing multivariate methods to improve the sensitivity of **new physics searches**?

- ❑ Searches for new physics must be sensitive to the entire allowed parameter space.

- ❑ Fixed cuts cannot be simultaneously optimised for new particles over a wide mass range.

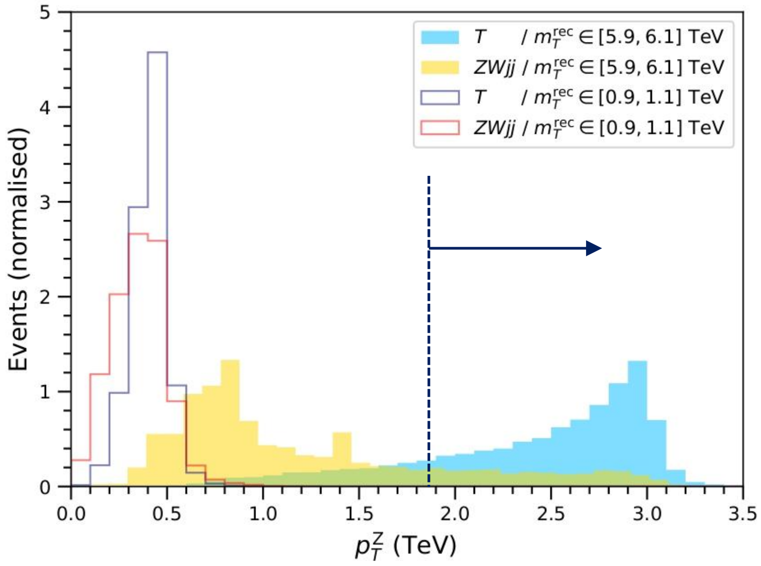


Motivation

Machine learning techniques play a critical role in experimental analyses. But how can we enhance existing multivariate methods to improve the sensitivity of **new physics searches**?

- Searches for new physics must be sensitive to the entire allowed parameter space.

- Fixed cuts cannot be simultaneously optimised for new particles over a wide mass range.

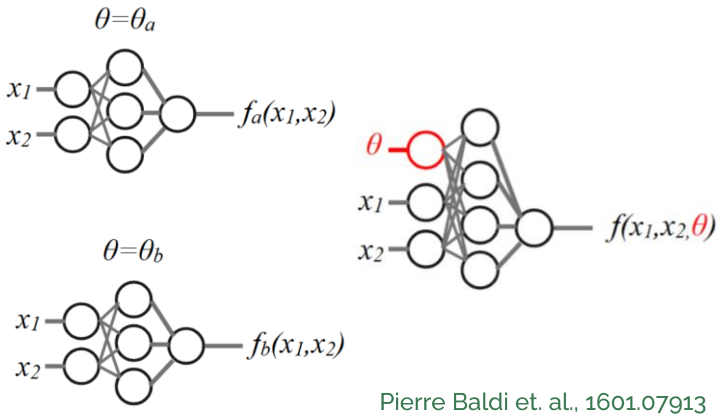


Motivation

Machine learning techniques play a critical role in experimental analyses. But how can we enhance existing multivariate methods to improve the sensitivity of **new physics searches**?

- Searches for new physics must be sensitive to the entire allowed parameter space.
- Fixed cuts cannot be simultaneously optimised for new particles over a wide mass range.

The ATLAS and CMS experiments use a **parameterised neural network** (pNN).

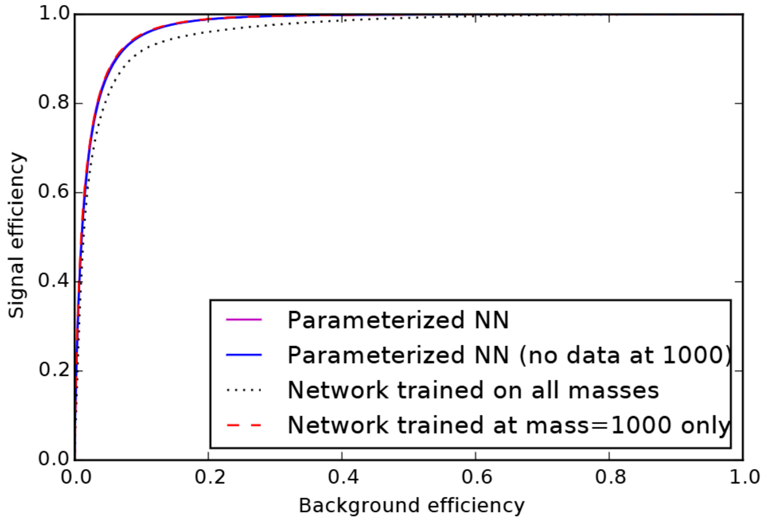


Pierre Baldi et. al., 1601.07913

Motivation

Machine learning techniques play a critical role in experimental analyses. But how can we enhance existing multivariate methods to improve the sensitivity of new physics searches?

- Searches for new physics must be sensitive to the entire allowed parameter space.
- Fixed cuts cannot be simultaneously optimised for new particles over a wide mass range.
- pNNs may not account for the correlations between the background and the mass scale.



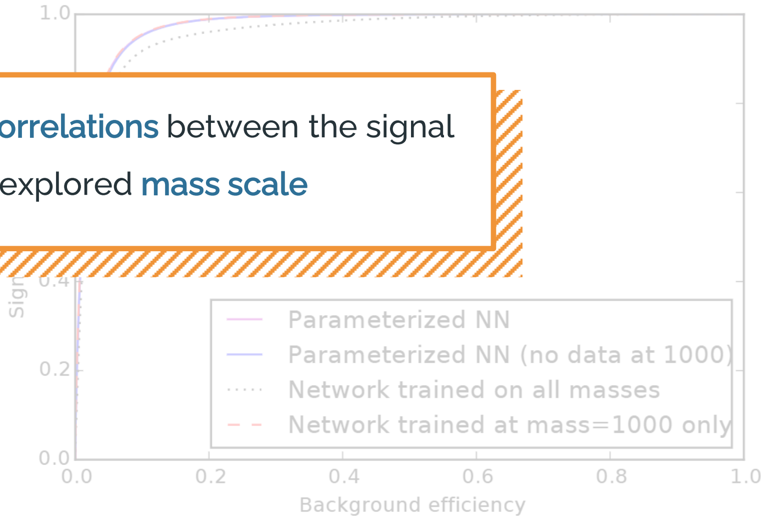
Pierre Baldi et. al., 1601.07913

Motivation

Machine learning techniques play a critical role in experimental analyses. But how can we enhance existing multivariate methods to improve the sensitivity of **new physics searches**?

- Searches fo w physics must be sensitive to the entire allowed para
- Fixed cuts ca particles ove
- pNNs may not account for the **correlations** between the background and the **mass scale**.

Make us of classifiers that fully exploit the **correlations** between the signal and background features and the explored **mass scale**

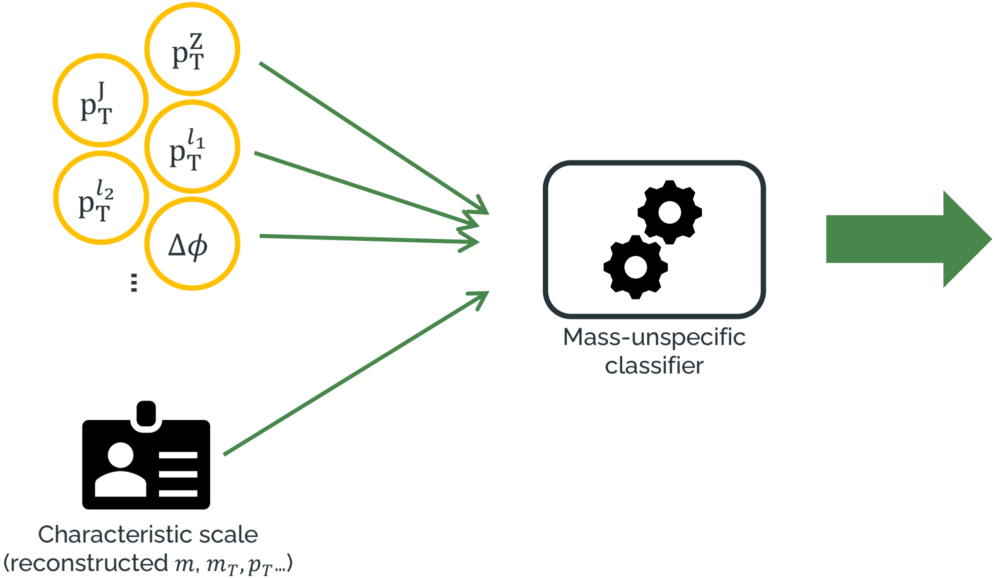


Pierre Baldi et. al., 1601.07913

Behind mass-unspecific classifiers

We make use of a multivariate method that fully exploits those **correlations** applying the mass unspecific supervised tagging (MUST) concept.

J.A. Aguilar-Saavedra et. al., 2008.12792



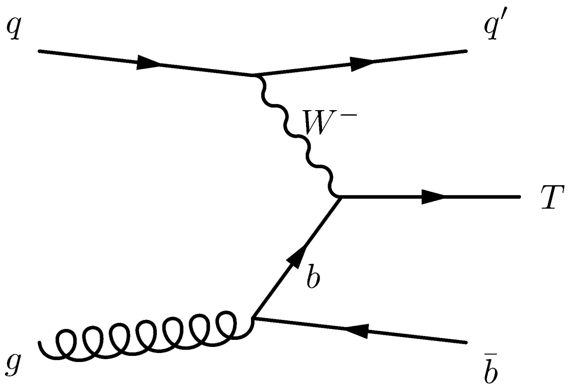
- The scale is meaningful for **both** signal and background.
- Captures the differences in the **background distributions** at different mass scales.

Event generation and reconstruction

Benchmark process

We used MadGraph, Pythia and Delphes to generate and simulate the events.

- **Signal:** Single production of a T quark (top partner), $pp \to T\bar{b}j$, with decay $T \to tZ \to l^+ \nu b l^+ l^-$, with $l = e, \mu$
- **Background:** We include the leading one, $pp \to ZW^+ jj$.



Signal samples generated for a range of T masses:

- 1.5×10^5 events for T masses from 1 to 6.4 TeV

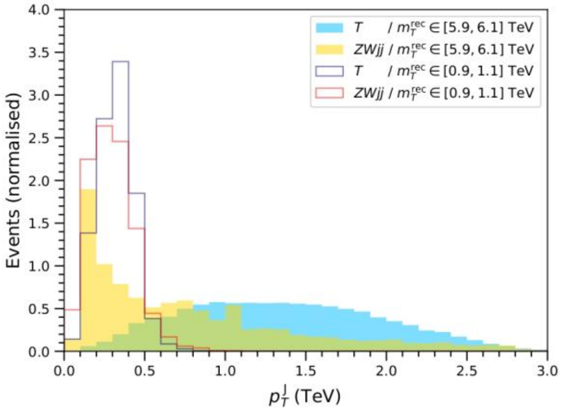
Inclusive and **binned** generation of the background:

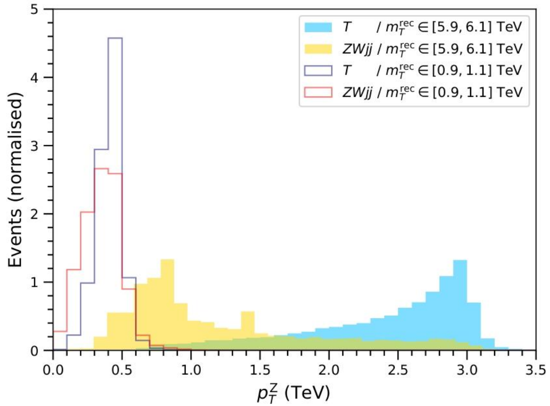
- 6×10^6 events, with $m_{ZWj} > 800$ GeV
- 4×10^5 events in 200 GeV intervals of m_{ZWj}

Discriminating variables

Kinematical distributions of interest in this work:

- Reconstructed T mass m_T^{rec}.
- Transverse momenta of Z leptons, $p_T^{l_1}$ and $p_T^{l_2}$.
- Reconstructed top quark mass, m_t^{rec}.
- Azimuthal angle between the reconstructed Z and the third lepton, $\Delta\phi(Z, l_3)$.
- Forward $R = 0.4$ jet multiplicity $(2.5 \leq |\eta| \leq 5)$.





Correlations with the mass scale help discriminate signal from background.

mass-unspecific classifiers

Shaping Data for Training

Training

The main difference between our mass-unspecific classifiers and the parameterised NN (pNN) lies in the training strategy:

μNN & μBDT

Training dataset:

- **Samples:** Uses binned generation for background.
- **Selection**: 5000 background + signal events per interval of m_T^{rec}.

Features:

- m_T^{rec} included as a feature.
- Background has a **physical** value of m_T^{rec}.

pNN

Training dataset:

- **Samples:** Uses inclusive background events.
- **Selection:** 5000 signal events for each interval of m_T^{rec}.

Features:

- Include m_T as a feature.
- Background m_T value is assigned **randomly.**

Training

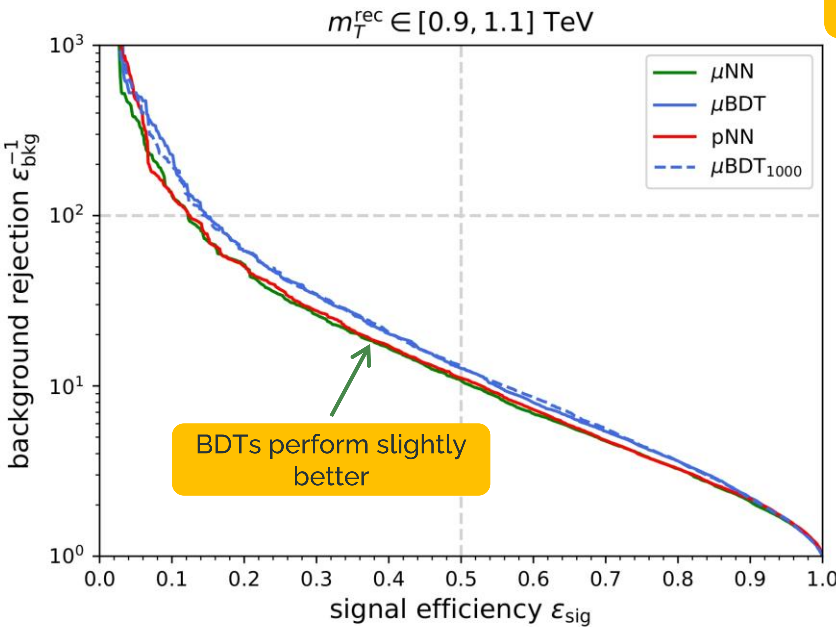
For comparison, additionally to our mass-unspecific classifiers we train a weighted NN (wNN), different parameterised NN (pNN) and a mass-specific BDT (μBDT $_{m_T}$):

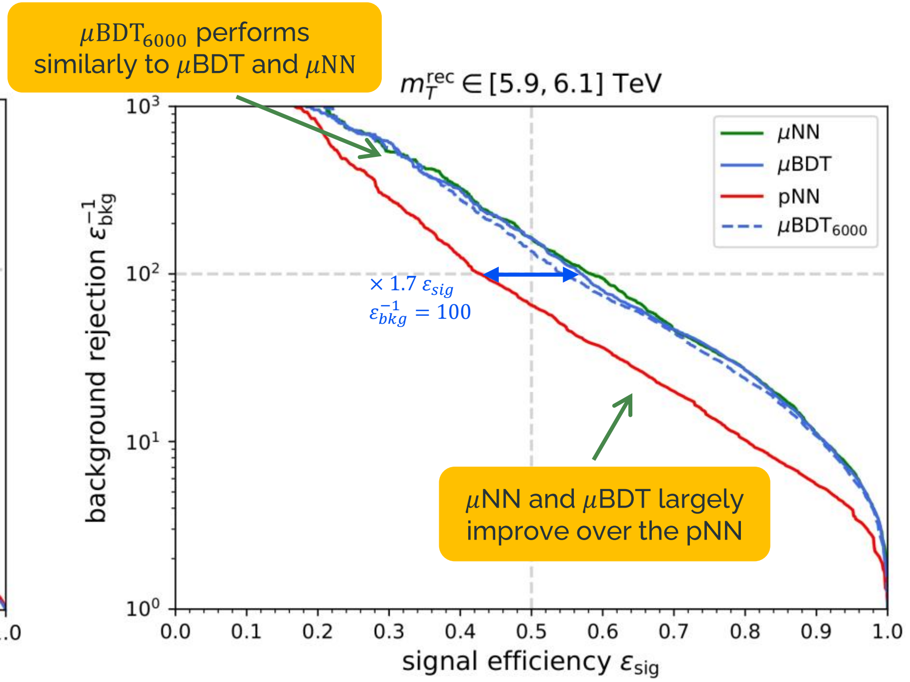
	Training range	Signal gen.	Background gen.	Mass label
μNN,μBDT	[0.9,6.5] TeV	0.2 TeV steps	0.2 TeV bins	m_T^{rec}
wBDT	[0.9,6.5] TeV	0.2 TeV steps	inclusive, weighted	m_T^{rec}
pNN	[0.9,6.5] TeV	0.2 TeV steps	inclusive, unweighted	m_T/ random
pNN $_B$	[0.9,6.5] TeV	0.2 TeV steps	0.2 TeV bins	m_T/ random
pNN $_X$	[0.9,6.5] TeV	0.2 TeV steps	inclusive, unweighted	(m_T/ random), m_T^{rec}
wNN	[0.9,6.5] TeV	0.2 TeV steps	inclusive, weighted	m_T^{rec}
μBDT $_{1000}$	[0.9,1.1] TeV	$m_T = 1$ TeV	[0.9,1.1] TeV	m_T^{rec}
μBDT $_{6000}$	[5.9,6.1] TeV	$m_T = 6$ TeV	[5.9,6.1] TeV	m_T^{rec}

Results

Performance

Receiver operating characteristic (ROC) curves comparing the different discriminators for different values of the T **mass**:





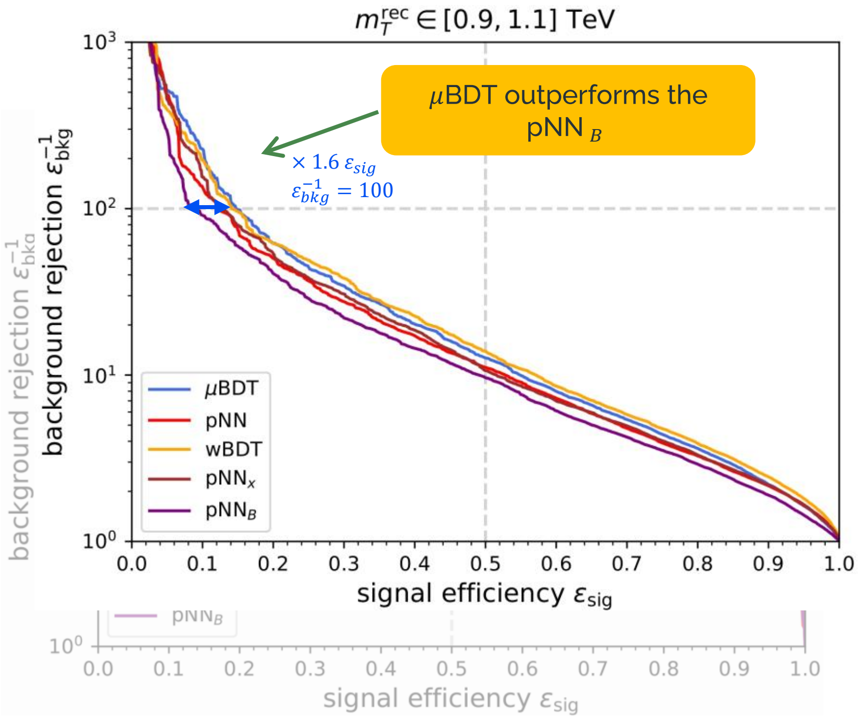
Performance

Receiver operating characteristic (ROC) curves comparing the different discriminators for different values of the T **mass**:

Why the μNN and μBDT perform better?

- The **mass label** is meaningful for the **background.** (pNN $_B$ vs μBDT)
- The model is trained using **balanced sets** within each m_T^{rec} bin. (pNN vs pNN $_B$)

The mass-unspecific discriminator is able to learn the **correlations** between the **background** distributions and the **mass scale.**



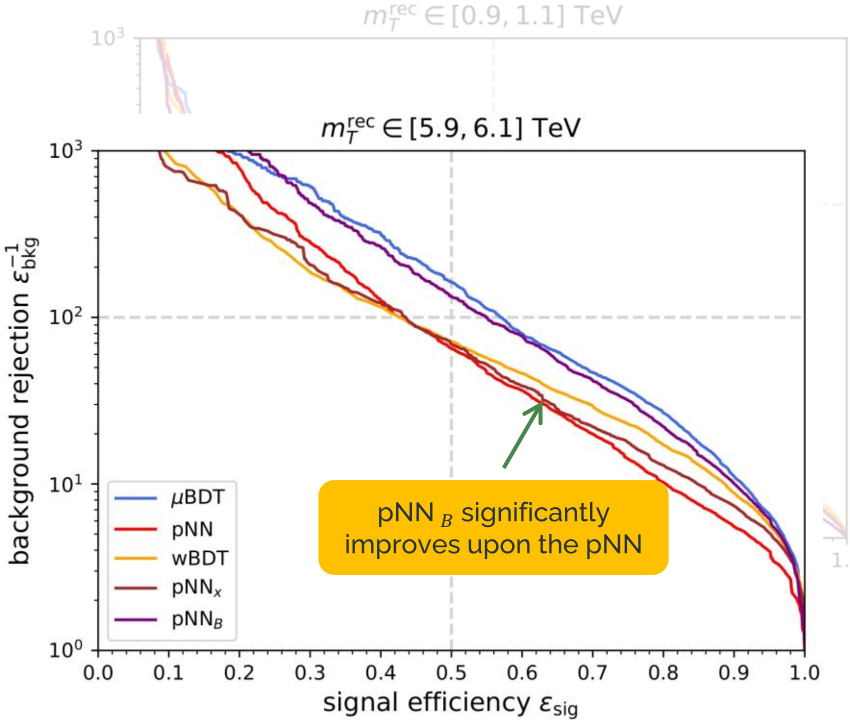
Performance

Receiver operating characteristic (ROC) curves comparing the different discriminators for different values of the T **mass**:

Why the μNN and μBDT perform better?

- The **mass label** is meaningful for the **background.** (pNN $_B$ vs μBDT)
- The model is trained using **balanced sets** within each m_T^{rec} bin. (pNN vs pNN $_B$)

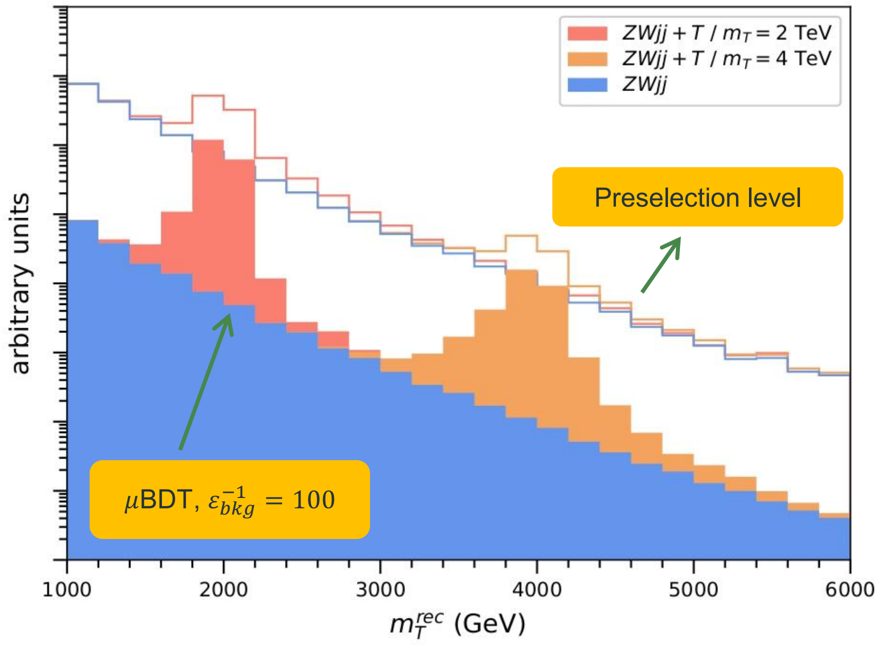
The mass-unspecific discriminator is able to learn the **correlations** between the **background** distributions and the **mass scale.**



Shape preservation

It is possible to preserve the **background and signal shape** after the application of mass-unspecific classifiers by **varying the threshold.**

Despite being trained in bins of m_T^{rec}, the μBDT produces a **continuous output** across bins.



Conclusions

- The performance of the μNN and μBDT is quite close to the one of the mass-specific classifiers.

- The μNN and μBDT perform as well or better than a pNN. There are two key factors:
  - The background mass scale is correlated with the actual background shape.
  - The training sample sets equal weight for high and low scales, allowing the classifier to learn those differences.

- Despite the use of a benchmark of single T production at the HL-LHC, our conclusions extend to other new physics processes and colliders.

Backup

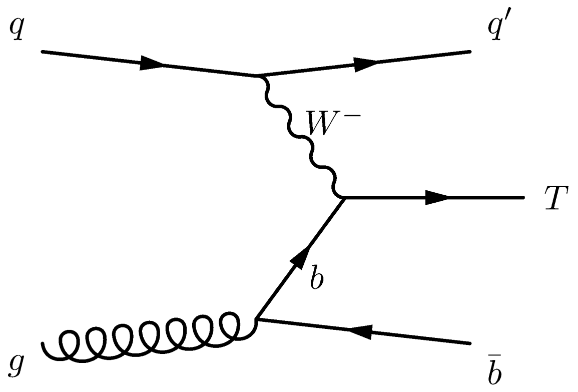
Benchmark process

The benchmark process is the single production of a **vector-like quark singlet** T with charge 2/3 at the HL-LHC. Why?

The cross-section is proportional to the square of the mixing angle, and it also depends on the quark mass. The process can probe:

$m_T \sim$ several TeV with $V_{Tb} \sim 1$ ⟷ $m_T \sim 1$ TeV with small V_{Tb}

➡ **Suited to test the mass-unspecific classifiers for a mass-dependant search**



Preselection cuts

Minimal **preselection cuts** are required to efficiently reconstruct the signal and background events.

1. ≥ 3 leptons, which we require to have $p_T \geq 25$ GeV, $|\eta| < 2.5$ and $p_T^{leading} > 40$ GeV.
2. We require the Z boson to have $p_T^Z > 60$ GeV.
3. ≥ 1 fat jet ($R = 0.8$), with the Z leptons outside the radius, and the third lepton inside; with $p_T > 80$ GeV and $|\eta| < 2.5$.
4. ≥ 1 light slim jet ($R = 0.4$) without any lepton inside the radius, $p_T > 25$ GeV and $|\eta| < 5$.
5. $\Delta R(Z, J) \geq 1$, with $\Delta R = [(\Delta\eta)^2 + (\Delta\phi)^2]^{1/2}$.

Mass	**Cut 1**	**Cut 2**	**Cut 3**	**Cut 4**	**Cut 5**	**Cut 6**
$m_T = 1000$ GeV	0.73304	0.68898	0.58860	0.30207	0.29762	0.29734
$m_T = 2000$ GeV	0.76763	0.73699	0.64231	0.50801	0.50137	0.50107
$m_T = 6400$ GeV	0.78919	0.7692	0.66545	0.60937	0.59738	0.59704

Reconstruction

After applying a minimal set of preselection cuts, the reconstruction is done:

- Z **boson:** opposite-sign same–flavour pair of leptons, $p_Z = p_{l_1} + p_{l_2}$. If more than one candidate, the pair with invariant mass closest to the Z boson mass is chosen.
- **Fat jet:** $R = 0.8$ fat jet containing the third lepton l_3.

Boosted top quark from the T decay

- **Neutrino:** we assumed it is produced in the W decay. $(p_\nu)_{x,y} = (\not{E}_T)_{x,y}$ and $\left(p_{l_3} + p_\nu\right)^2 = M_W^2$ provides two solutions for $(p_\nu)_z$.

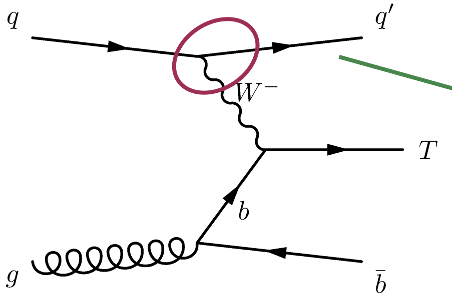
Take the one such that $[(p_J + p_\nu)^2]^{1/2} \rightarrow M_t$

- **Slim jet:** $R = 0.4$ light jet.

Reconstruction

After applying a minimal set of preselection cuts, the reconstruction is done:

- Z **boson:** opposite-sign same–flavour pair of leptons, $p_Z = p_{l_1} + p_{l_2}$. If more than one candidate, the pair with invariant mass closest to the Z boson mass is chosen.
- **Fat jet:** $R = 0.8$ fat jet containing the third lepton l_3.
- **Neutrino:** we assumed it is produced in the W decay. $(p_\nu)_{x,y} = (\not{E}_T)_{x,y}$ and $\left(p_{l_3} + p_\nu\right)^2 = M_W^2$ provides two solutions for $(p_\nu)_z$.
- **Slim jet:** $R = 0.4$ light jet.



Light **forward** jet

Reconstruction

After applying a minimal set of preselection cuts, the reconstruction is done:

- Z **boson:** opposite-sign same–flavour pair of leptons, $p_Z = p_{l_1} + p_{l_2}$. If more than one candidate, the pair with invariant mass closest to the Z boson mass is chosen.
- **Fat jet:** $R = 0.8$ fat jet containing the third lepton l_3.
- **Neutrino:** we assumed it is produced in the W decay. $(p_\nu)_{x,y} = (\not{E}_T)_{x,y}$ and $\left(p_{l_3} + p_\nu\right)^2 = M_W^2$ provides two solutions for $(p_\nu)_z$.
- **Slim jet:** $R = 0.4$ light jet.

The T signal produces a peak in the reconstructed T mass
$m_T^{rec} = [(p_J \; + p_\nu \; + p_Z \;)^2]^{1/2}$

m_T^{rec} is our for the classifier

Test and validation samples

In order to prevent overtraining, we use validation samples generated using the same procedure as the training data:

- μNN and μBDT: 3500 signal and background events per m_T^{rec} bin.
- wNN and pNN: 3500 signal events for each T mass. 9.4×10^4 background events from the inclusively generated sample.
- $\mu\mathbf{BDT}_{m_T}$: 3500 signal and background events.

For the test we used the number of events that remain after removing the ones used for the training and validation:

$m_T^{rec} \in [0.9, 1.1]$ TeV $\longrightarrow$ 51428 signal and 20894 background events

$m_T^{rec} \in [5.9, 6.1]$ TeV $\longrightarrow$ 15456 signal and 15752 background events

Model architectures

A detailed description of all the models we trained in this work:

NNs

- Implemented using Keras with a Tensorflow backend.
- [64,64] with ReLU activation for the hidden layers and a sigmoid function for the output one.
- Binary cross entropy los function, optimisation using the Adam algorithm.
- From 5 trainings with different initial seeds, we select the one that gives the best AUC.

BDTs

- Implemented using XGBoost.
- A maximum of 500 boosting trees and a depth of 5.
- Learning rate of 0.15.
- From 5 trainings with different initial seeds, we select the one that gives the best AUC.

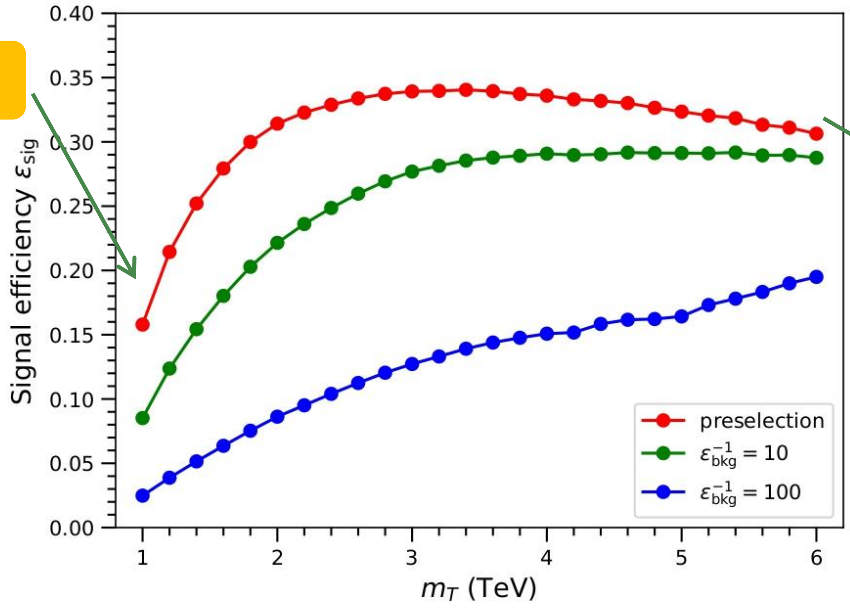
D. P. Kingma et. al., 1412.6980

T. Chen et. al., 1603.02754

Signal efficiency

Signal efficiencies of the μBDT as a function of the T quark mass:

Top quark not sufficiently boosted



b-tagging efficiency