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

Universidad Autónoma de Madrid

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro

AGENCIA ESTATAL DE INVESTIGACIÓN

Motivation

Motivation

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

Home > Journal of High Energy Physics > Article

Search for new particles in events with a hadronically decaying W or Z boson and large missing transverse momentum at $\sqrt{s}$ = 13 TeV using the ATLAS detector

Regular Article – Experimental Physics | Open access | Published: 22 November

Volume 2024, article number 126, (2024) Cite this article

Download PDF

You have full access to this open access article

The ATLAS collaboration, G. Aad, E. Aakvaag, B. Abbott, S. Abdelhamee

Abidi, M. Aboelela, A. Aboulhorma, H. Abramowicz, H. Abreu, Y. Abulai

Adam Bourdarios, L. Adamczyk, S. V. Addepalli, M. J. Addison, J. Adelma

Affolder, Y. Afik, ... L. Zwalinski Show authors

837 Accesses 1 Altmetric Explore all metrics

Search for heavy Maj

and $e^{\pm}\mu^{\pm}$ final states v

pp collisions at $\sqrt{s}$ =

detector

The ATLAS Collaboration*

N ACCESS | GO MOBILE »

earch for Frac

ollisions at $\sqrt{s}$

Hayrapetyan[1], A. Tumasy

ELSEVIER

Physics Letters B

Volume 856, September 2024, 138938

Letter

Search for pair-produced higgsinos decaying via Higgs or Z bosons to final states containing a pair of photons and a pair of b-jets with the ATLAS detector

The ATLAS Collaboration*

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.physletb.2024.138938

Get rights and content

Under a Creative Commons license

Open access

OPEN ACCESS | GO MOBILE » | ACCESS BY CENTRO DE FISICA MIGUEL A. CATALAN

Search for production of a single vectorlike quark decaying to tH or tZ in the all-hadronic final state in pp collisions at $\sqrt{s} = 13$ TeV

A. Hayrapetyan[1], A. Tumasyan[1,b], W. Adam[2], J. W. Andrejkovic[2], T. Bergauer[2], S. Chatterjee[2], K. Damanakis[2], M. Dragicevic[2], P. S. Hussain[2] *et al.* (CMS Collaboration)

Show more

Phys. Rev. D **110**, 072012 – **Published 21 October,**

DOI: https://doi.org/10.1103/PhysRevD.110.07201

BILE » | ACCESS BY CEN

long-lived

in proton-

Tumasyan[1,b], W.

Hussain[2] *et al.* (C

)07 – **Published 6 August, 2024**

).1103/PhysRevD.110.032007

ELSEVIER

Physics Reports

Volume 1115, 17 April 2025, Pages 448-569

Dark sector searches with the CMS experiment

CMS Collaboration

Show more

Add to Mendeley Share Cite

https://doi.org/10.1016/j.physrep.2024.09.013

Get rights and content

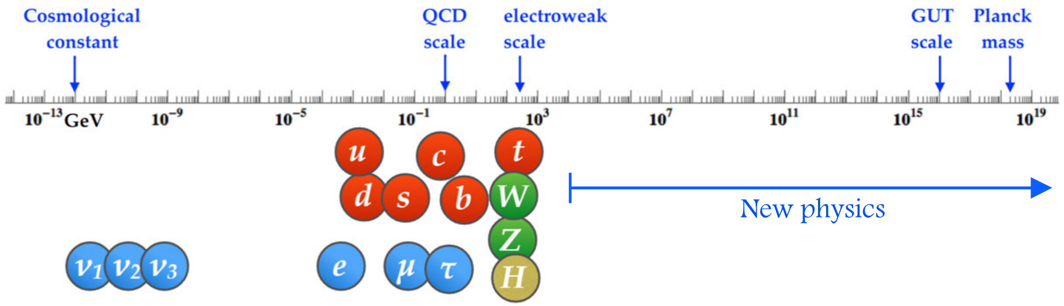
Under a Creative Commons license

Open access

Motivation

Machine learning techniques play a critical role in experimental analyses. How can ML help with **new physics searches**?

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

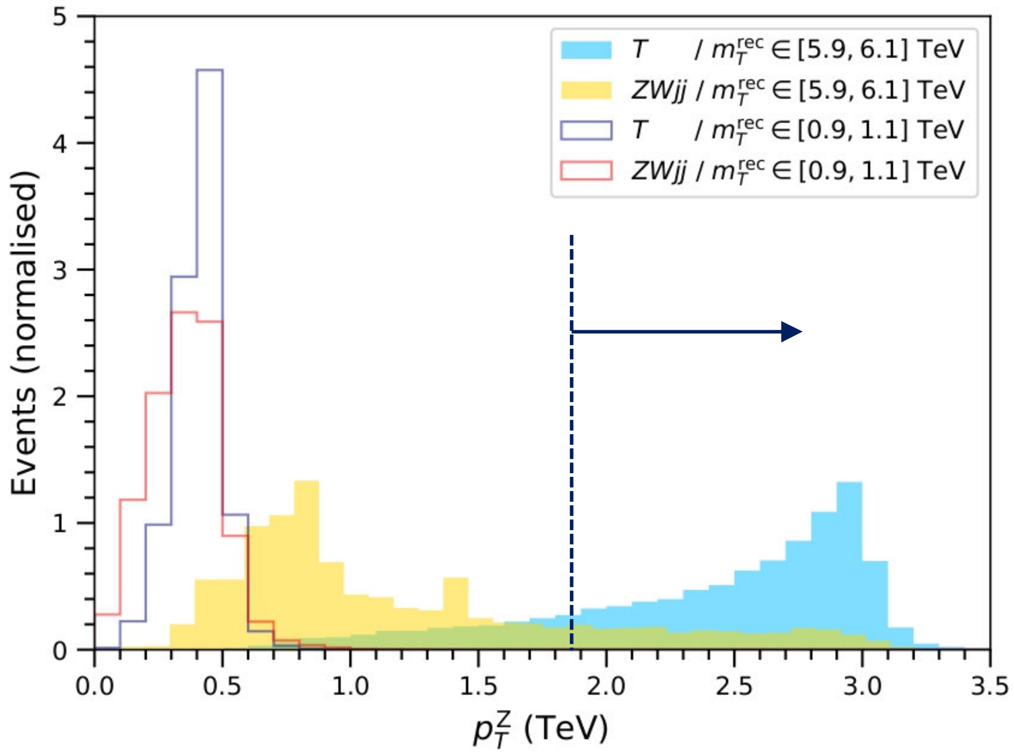


Motivation

Machine learning techniques play a critical role in experimental analyses. How can ML help with **new physics searches**?

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

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

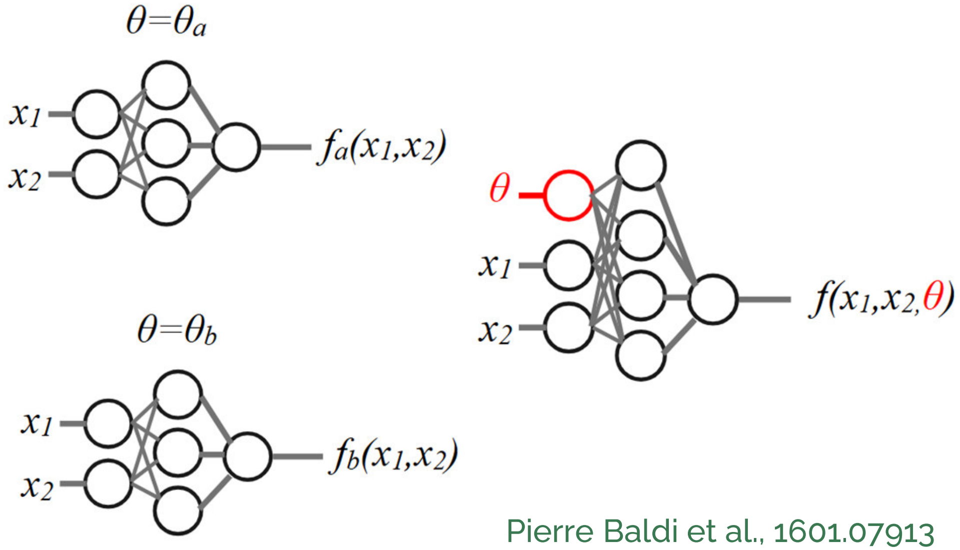


Motivation

Machine learning techniques play a critical role in experimental analyses. How can ML help with **new physics searches**?

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

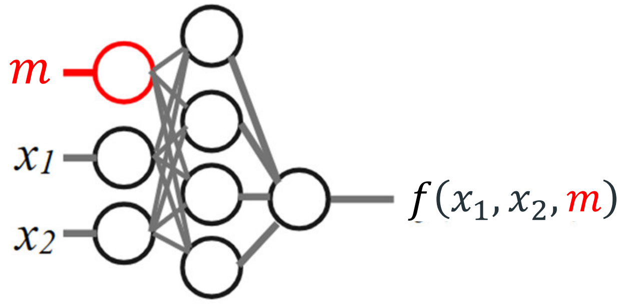
The ATLAS and CMS experiments use a parameterized neural network (pNN).



Pierre Baldi et al., 1601.07913

Motivation

PNN aims to learn the **conditional** signal distribution for any value of m



$$\boxed{f(x_1, x_2, m) \rightarrow p(x_1, x_2 | \, \mathrm{H}_1(m))}$$

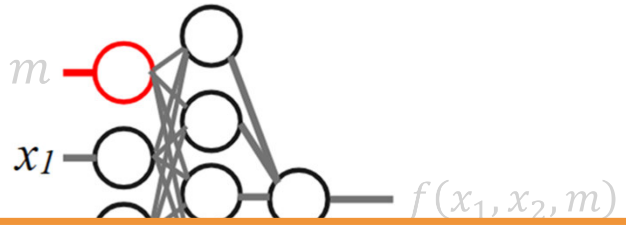
How are pNNs trained?

- The ***real mass*** is added as an input feature.
- For background events, a **random** mass value is assigned

→ Mass-background correlations are ignored

Motivation

PNN's goal is to learn the **conditional** signal distribution for any value of m



$$f(x_1, x_2, m) \rightarrow p(x_1, x_2 | \mathrm{H}_1(m))$$

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

How are pNNs

- The ***real mass*** is added as an input feature.
- For background events, a **random** value is assigned

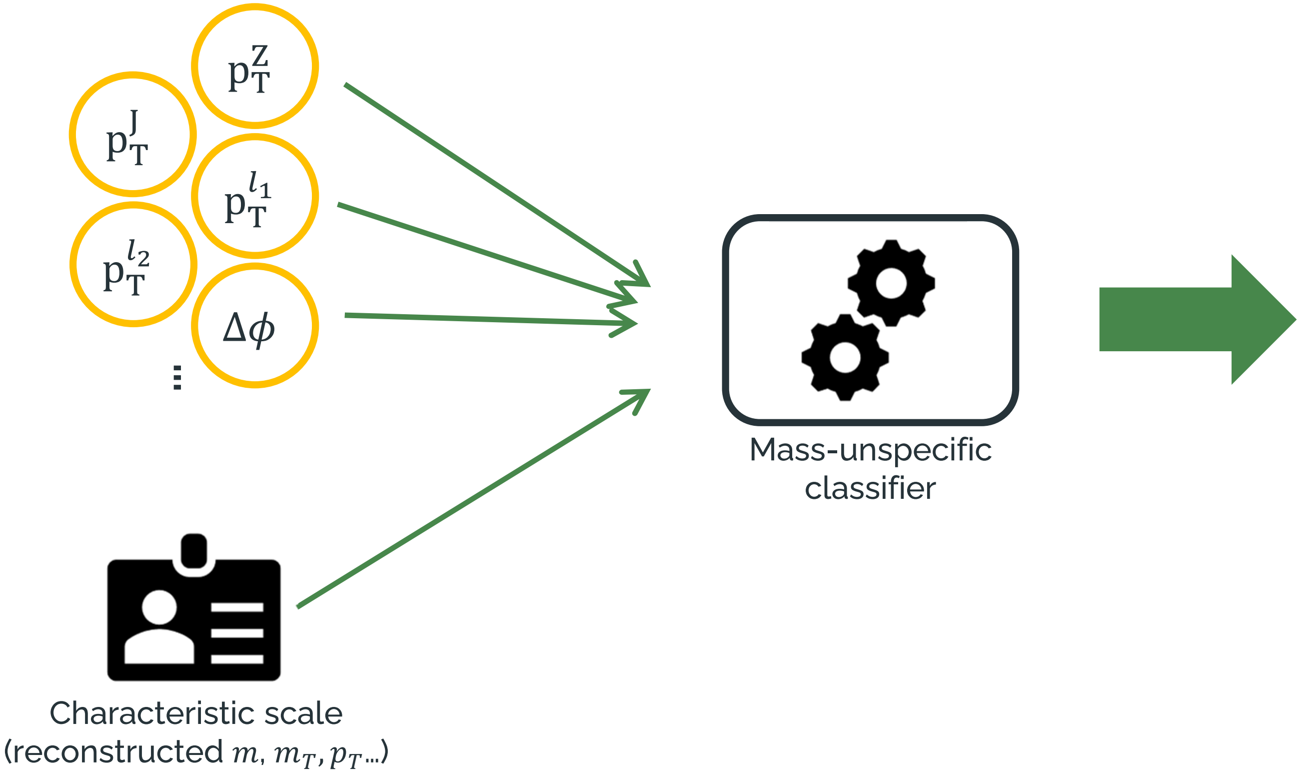
Mass scale-background correlations are ignored

Method overview

Method overview

The mass-unspecific classifiers implement two improvements:

1. Introduce a *new* characteristic **mass scale**

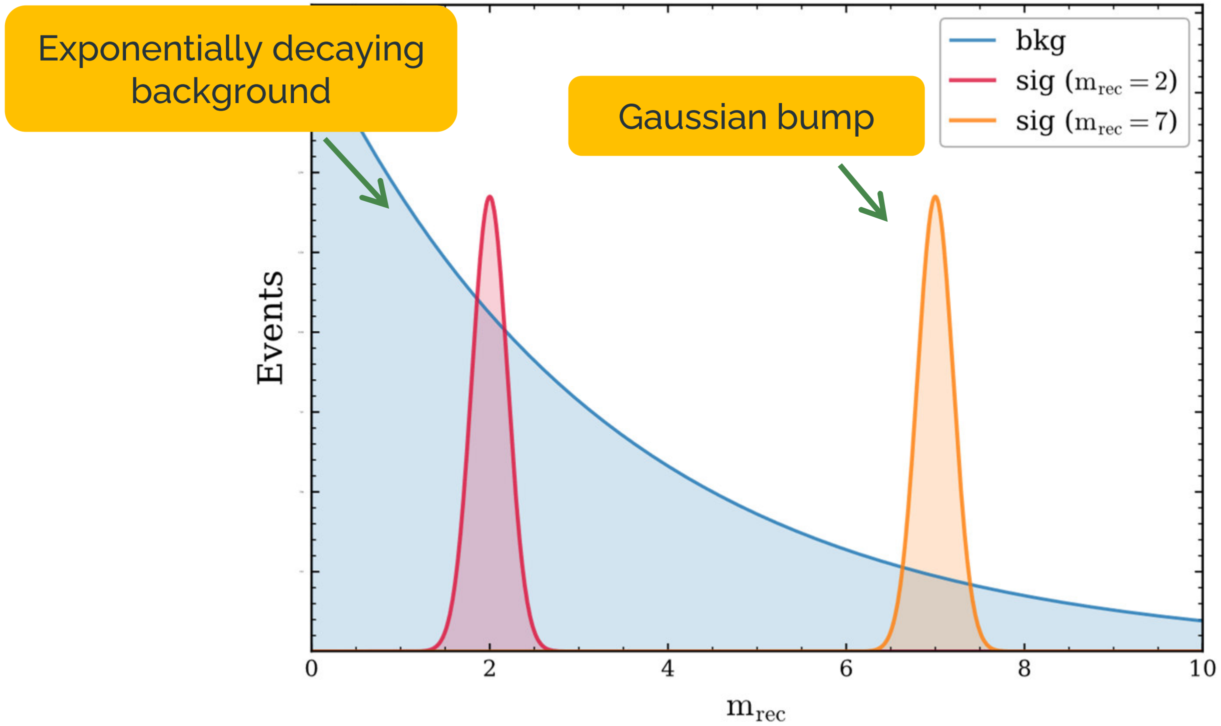


- ❑ The scale is meaningful for **both** signal and background.
- ❑ Captures the differences in the **background distributions** at different mass scales.

Method overview

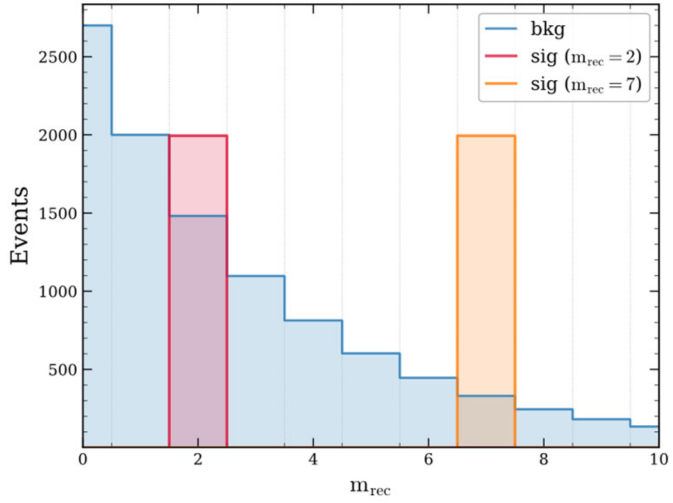
The mass-unspecific classifiers implement two improvements:

2. Use a **balanced set** for training

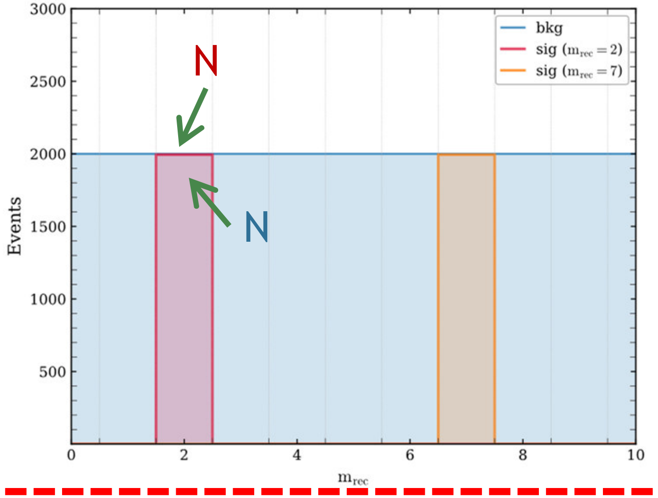


Binning choice

Inclusive generation



Binned generation



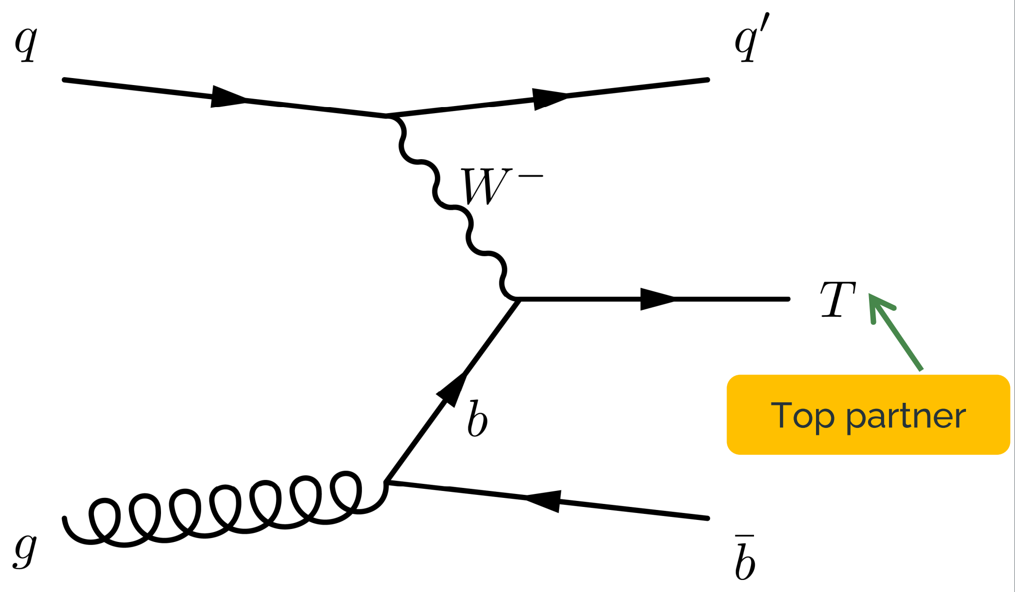
Benchmark signal

Benchmark process

The cross-section is proportional to the square of the mixing angle and 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-dependent search**



Training

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

μNN & μBDT

Training dataset:

- **Samples:** Use 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}.

VS

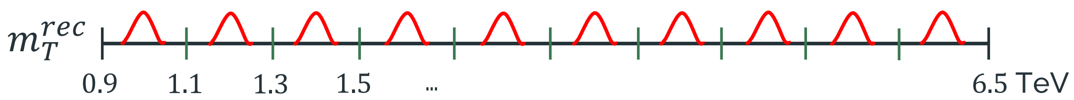
pNN

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

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

μNN & μBDT

Training dataset:

- **Samples:** Use 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}.

VS

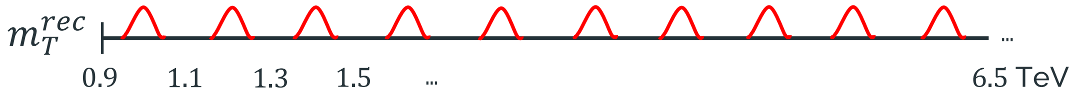
pNN

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

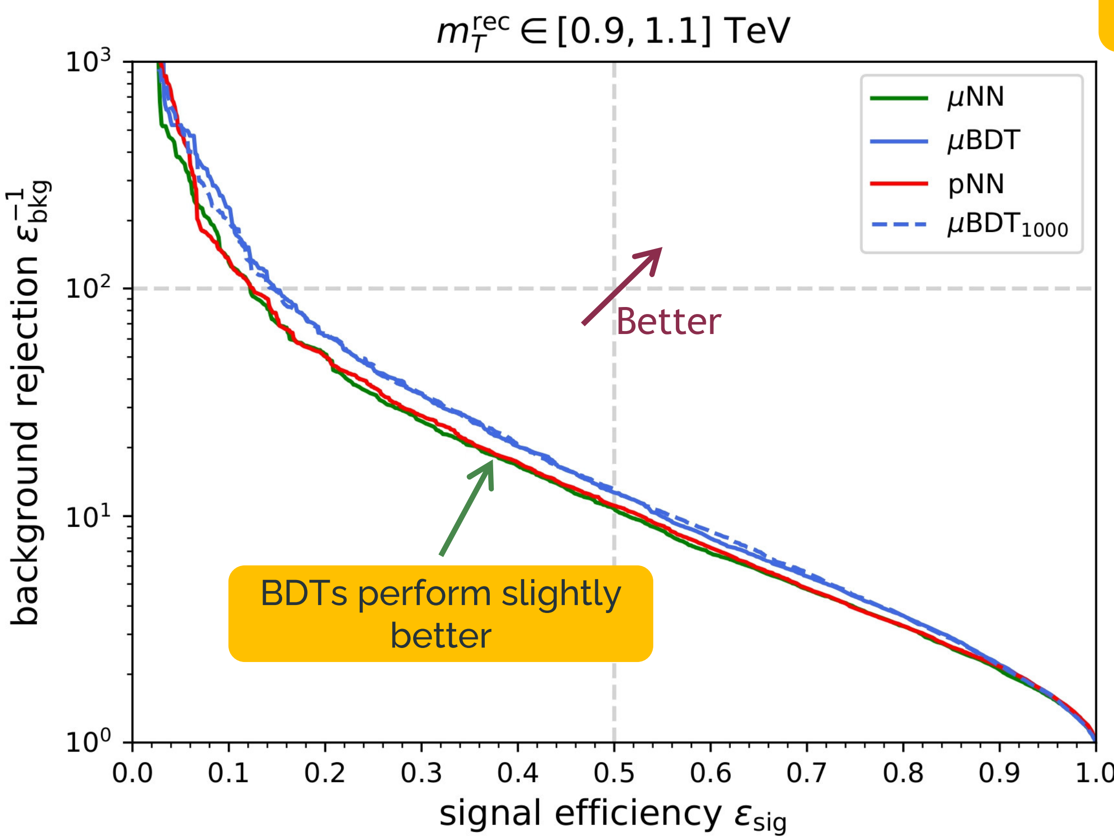
For comparison, in addition to our mass-unspecific classifiers we train a weighted NN (**wNN**), different parameterized NN (**pNN**) and a mass-specific BDT (**μBDT $_{m_T}$**):

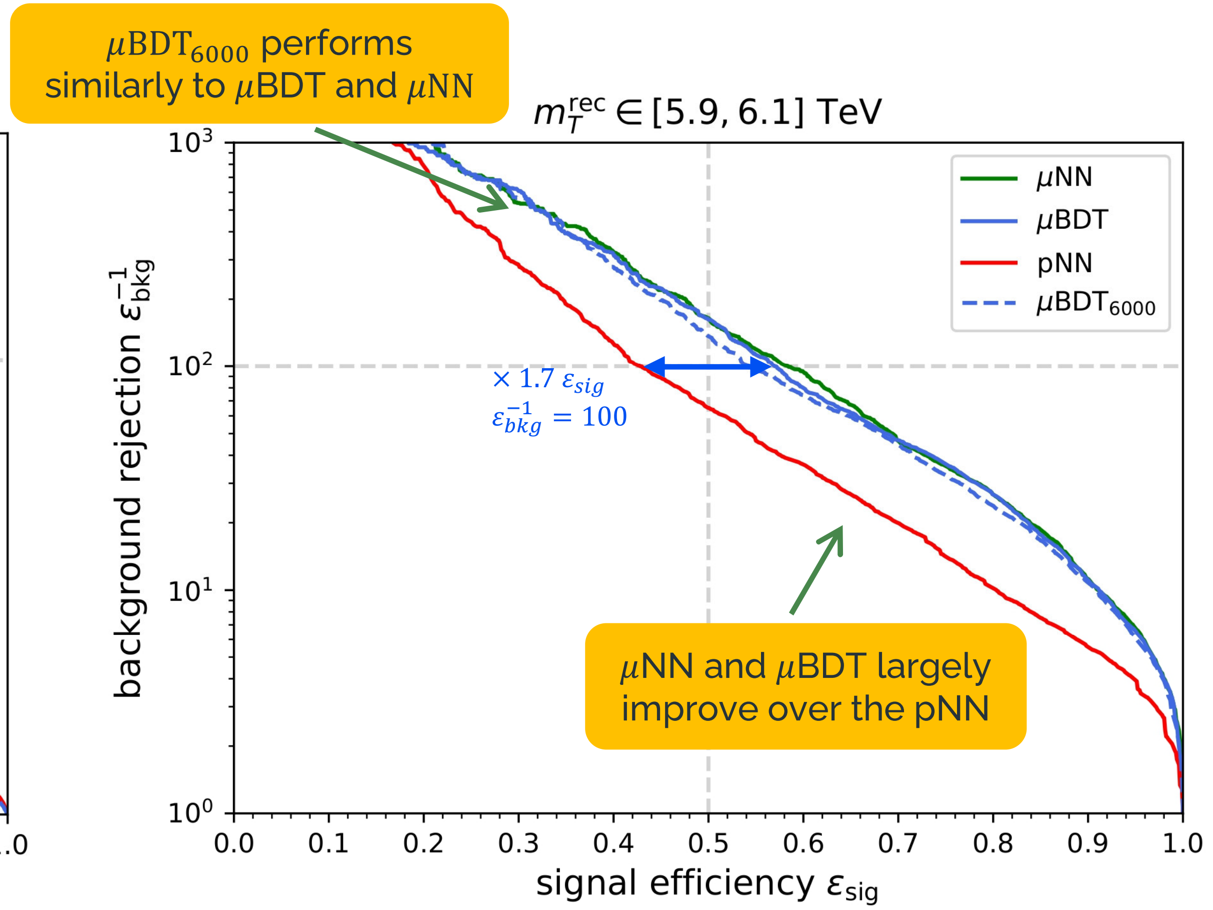
	Training range	Signal gen.	Background gen.	Mass label
μNN,μBDT	[0.9,6.5] TeV	0.2 TeV steps	0.2 TeV bins	m_T^{rec}
wBDT	[0.9,6.5] TeV	0.2 TeV steps	inclusive, weighted	m_T^{rec}
pNN	[0.9,6.5] TeV	0.2 TeV steps	inclusive, unweighted	m_T/ random
pNN $_B$	[0.9,6.5] TeV	0.2 TeV steps	0.2 TeV bins	m_T/ random
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 different discriminators for several T **mass** values:



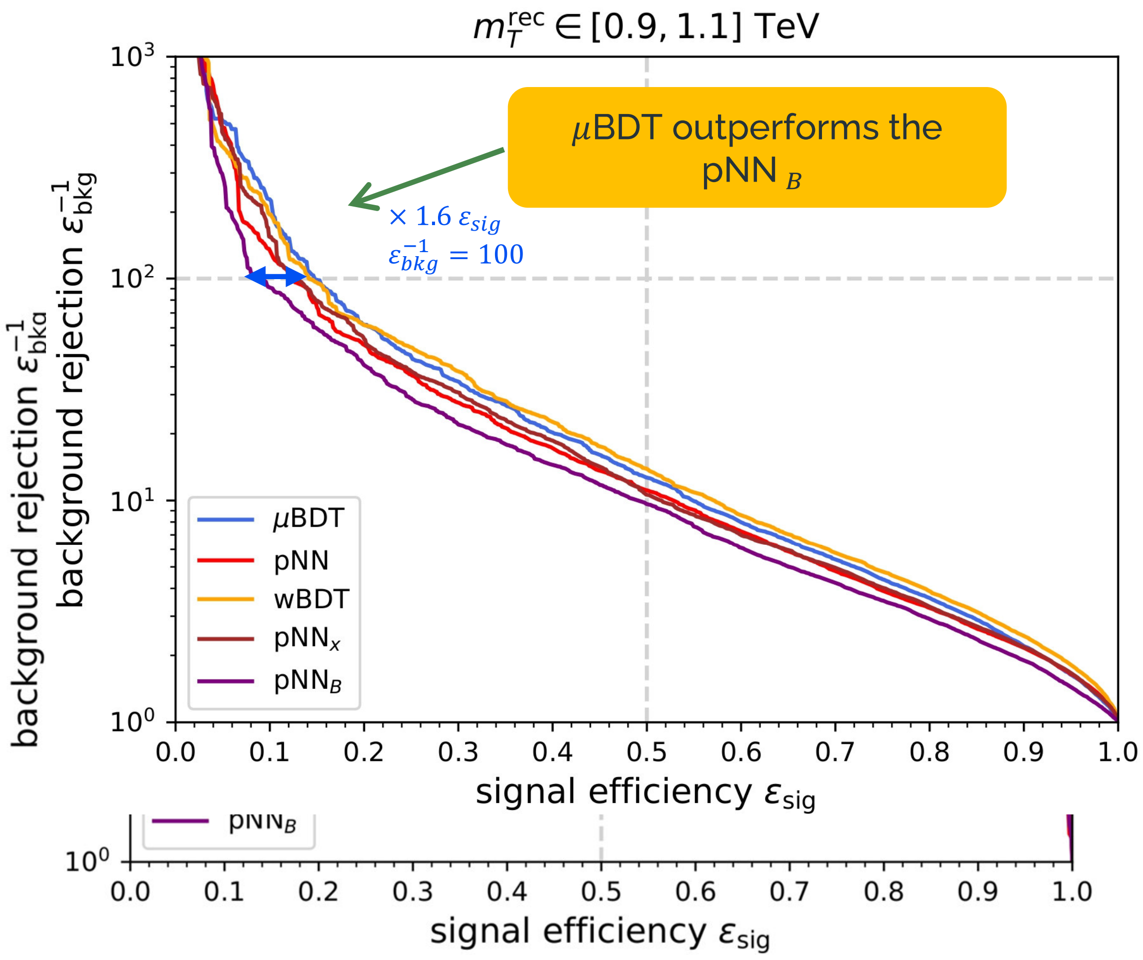


Performance

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

Why do 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.**

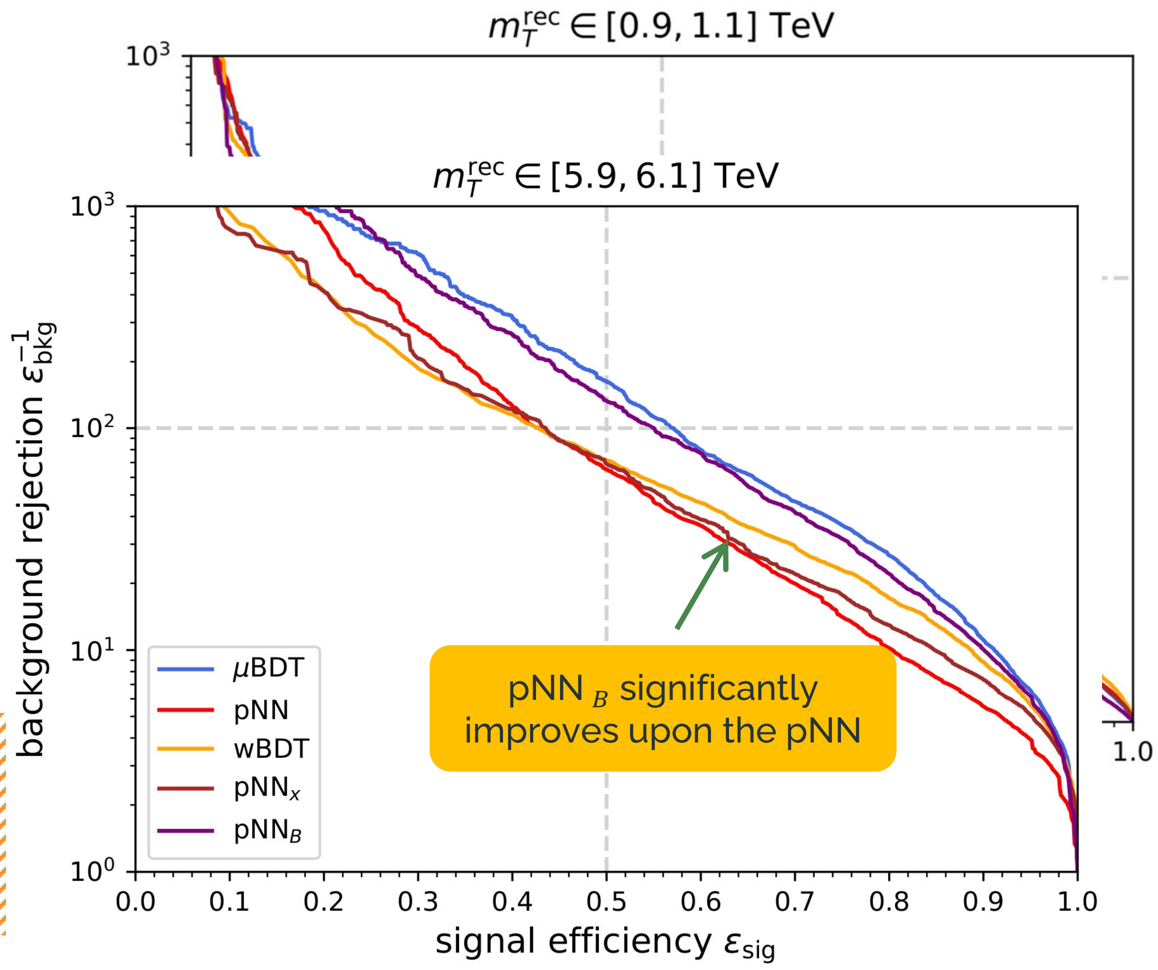
Performance

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

Why do 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 $_B$)

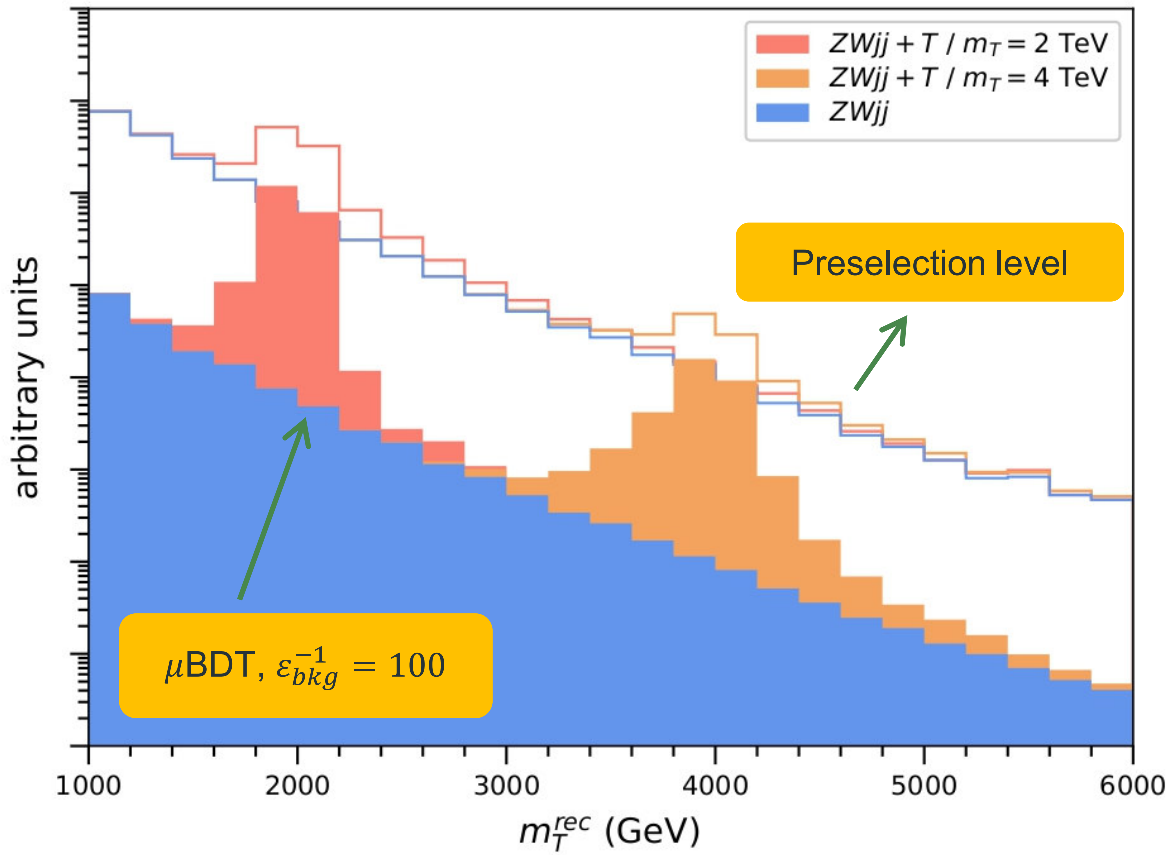
The mass-unspecific discriminator is able to learn the **correlations** between the **background** distributions and the **mass scale.**



Shape preservation

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

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



Conclusions

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

- ☑ The μNN and μBDT perform as well as 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.

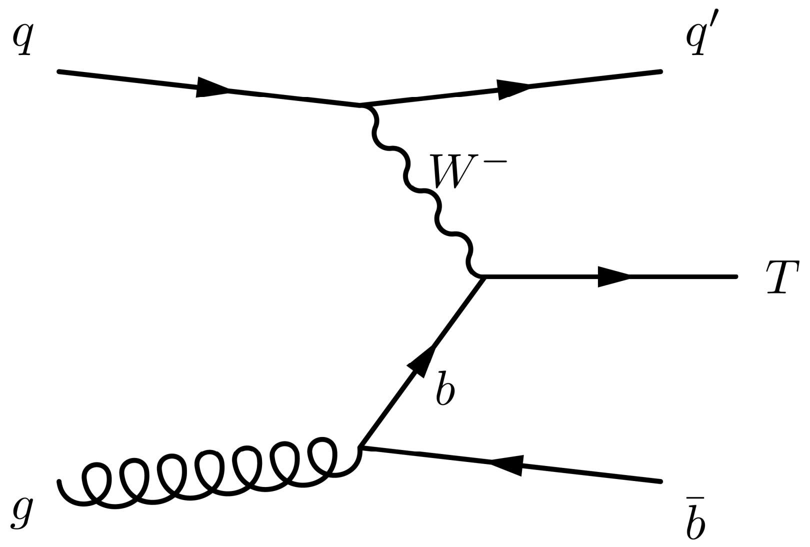
- ☑ Despite the use of a benchmark of single T production at the HL-LHC, our conclusions **extend to other new physics processes and colliders.**

Backup

Benchmark process

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

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



Signal events 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}

Preselection cuts

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

1. ≥ 3 leptons, which we are required 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

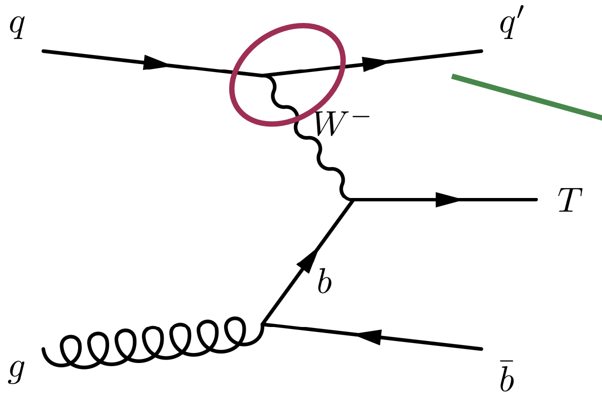
- Neutrino: we assumed it is produced in the W decay. $(p_\nu)_{x,y} = (\not{E}_T)_{x,y}$ and $\left(p_{l_3} + p_\nu\right)^2 = M_W^2$ provides two solutions for $(p_\nu)_z$.
- Slim jet: $R = 0.4$ light jet.

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

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.

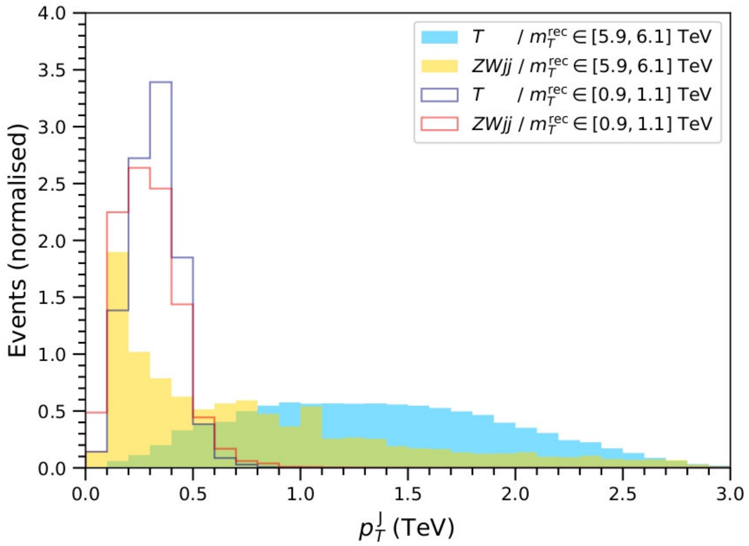
The T signal produces a peak in the reconstructed T mass
$m_T^{rec} = [(p_J \ + p_\nu \ + p_Z\ )^2]^{1/2}$

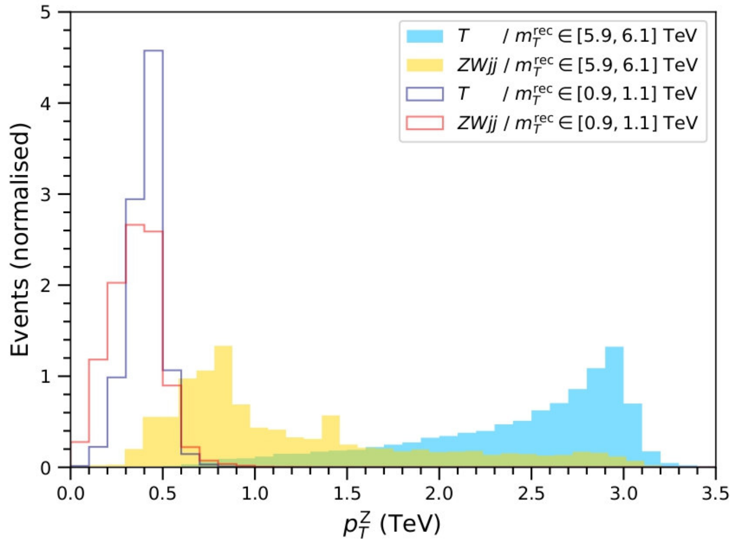
m_T^{rec} is our [ID] for the classifier

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

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.
- $\boldsymbol{\mu}\mathbf{BDT}_{\boldsymbol{m_T}}$: 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 loss function, optimization 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: