Machine Learning Techniques in 4 top final states

Cristina Giordano, Institute of High Energy Physics

Structure

- The Standard Model of particle physics
- Four top quark production
- State of the art
- Event Level MVA
- Top reconstruction
- Conclusions

The Standard Model (SM)

Three out of the four fundamental interactions

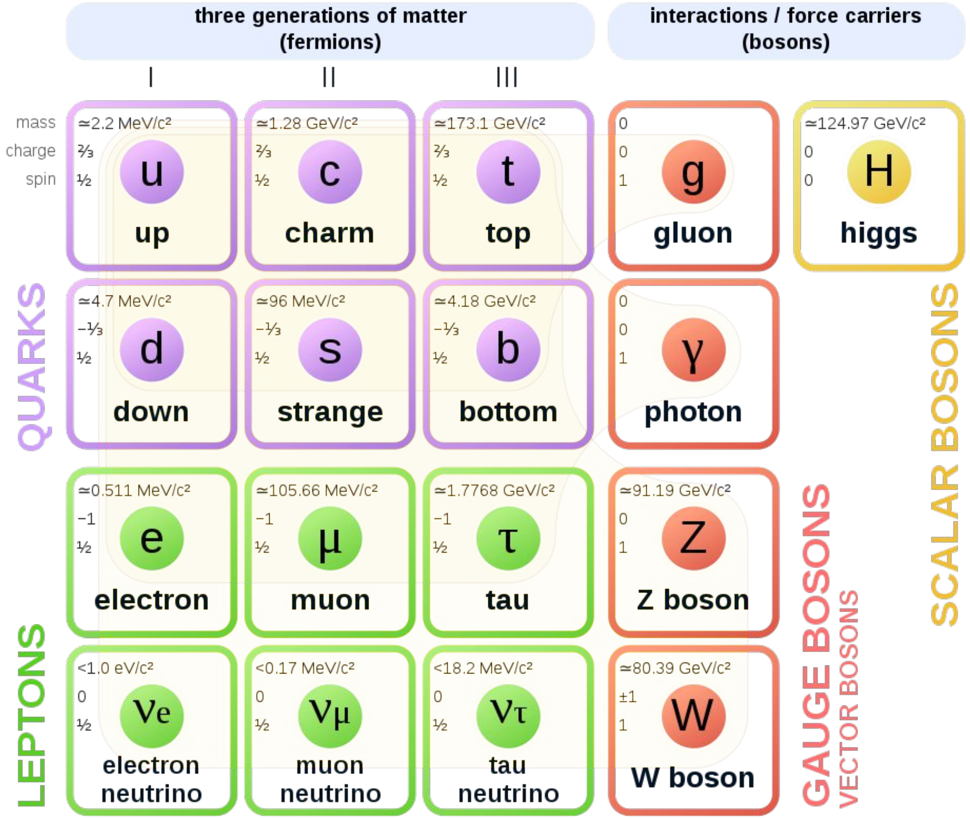
- ➢ Electromagnetic
- ➢ Strong
- ➢ Weak

Elementary particles

- ➢ Quarks
- ➢ Leptons
- ➢ Gauge bosons
- ➢ Higgs bosons

Not covered phenomena

- ➢ gravitational interaction
- ➢ dark Energy
- ➢ dark Matter
- ➢ matter-antimatter asymmetry



Why tttt?

arXiv:2305.13439

arXiv:2303.15061

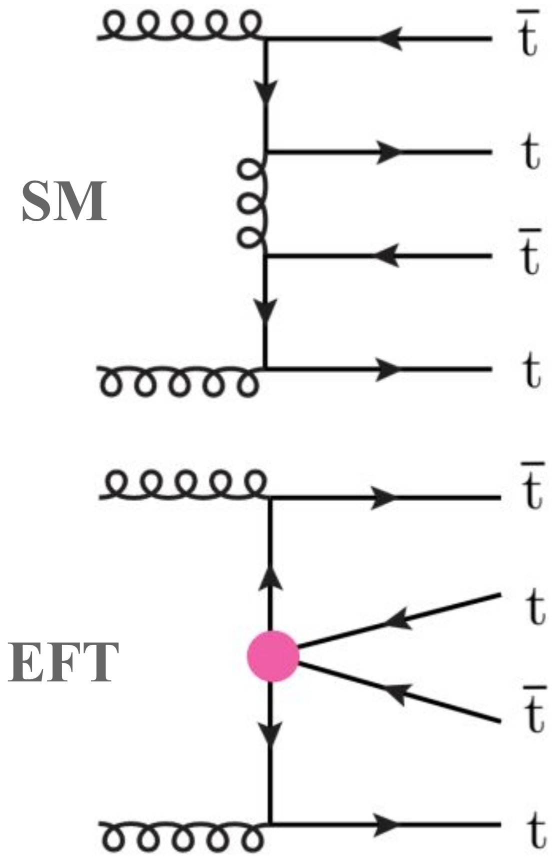
Important in the SM framework and as a probe for new **Physics Beyond SM**

4 top production:

- ➢ CMS: (expected 4.9) observed 5.5 σ
- ➢ ATLAS: (expected 4.3) observed 6.1 σ

Features:

- ➢ Statistically limited
- ➢ Distinctive signature
- ➢ Sensitivity to top-H Yukawa coupling and BSM physics
- ➢ Unique sensitivity for **4 fermions operators**



Why tttt?

arXiv:2305.13439

arXiv:2303.15061

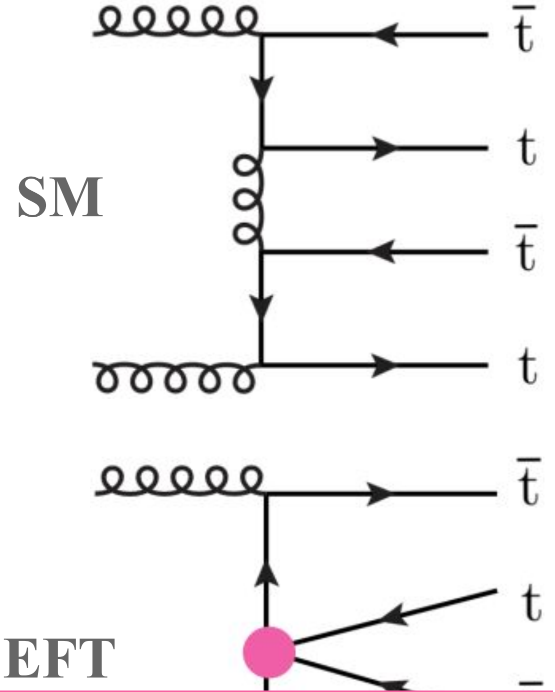
Important in the SM framework and as a probe for new **Physics Beyond SM**

4 top production:

- ➢ CMS: (expected 4.9) observed 5.5 σ
- ➢ ATLAS: (expected 4.3) observed 6.1 σ

Features:

- ➢ Statistically limited
- ➢ Distinctive signature
- ➢ Sensitivity to top-H Yukawa coupling and BSM physics
- ➢ Unique sensitivity for **4 fermions operators**



See Maryam's talk!

tttt decay channels

Multiple final states

1ℓ & 2ℓOS channels

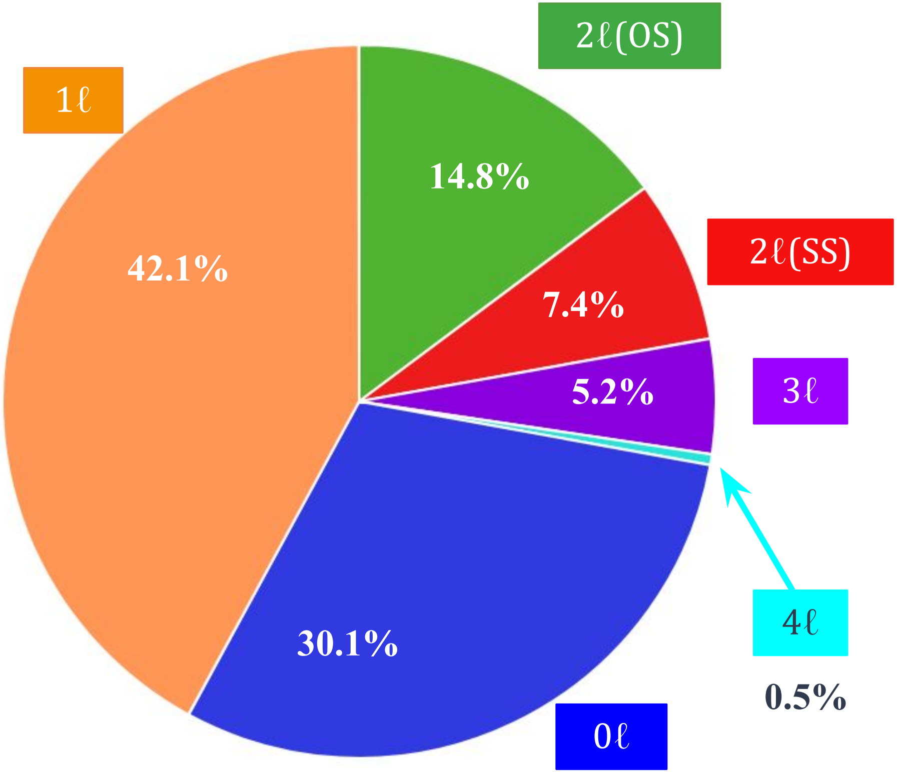
- ➢ **BR ~ 57%**
- ➢ large tt irreducible background

2ℓSS & multilepton

- ➢ **BR ~ 13%**
- ➢ highest sensitivity

0ℓ (hadronic)

- ➢ **BR ~ 31%**
- ➢ significant QCD & tt backgrounds



Machine Learning (ML) in tttt

Long history: single top quark production

- 2009 → observation @ Tevatron [arXiv:0906.0523]
- 2011 → ML aided measurement by CMS [PhysRevLett.107.091802] & ATLAS [arXiv:1205.3130v3]

ML in tttt

- Many searches through the years

Machine Learning (ML) in tttt

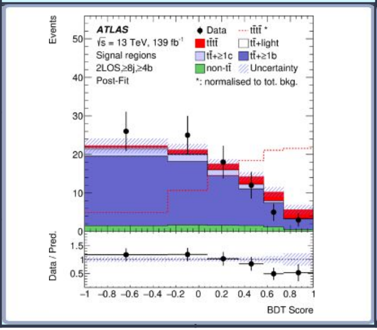
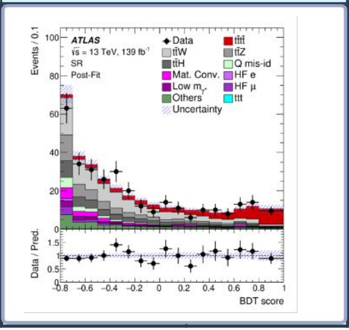
Long history: single top quark production

- ➢ 2009 → observation @ Tevatron [arXiv:0906.0523]
- ➢ 2011 → ML aided measurement by CMS [PhysRevLett.107.091802] & ATLAS [arXiv:1205.3130v3]

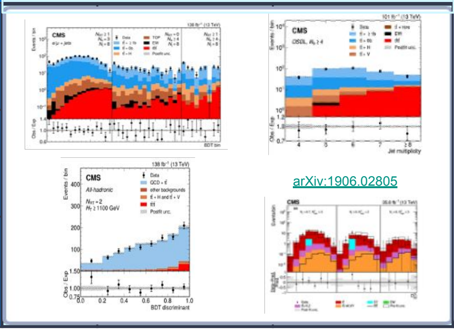
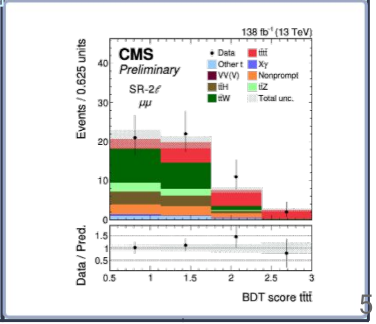
ML in tttt

- ➢ Many searches through the years

ATLAS EXPERIMENT

	0ℓ	1ℓ	2ℓOS	2ℓSS	3ℓ+
		arXiv:2106.11683		arXiv:2007.14858	
2018					
2017					
2016					

CMS

	0ℓ	1ℓ	2ℓOS	2ℓSS	3ℓ+
		arXiv:2303.03864		arXiv:2305.13439	
2018					
2017					
2016					

Machine Learning (ML) in tttt

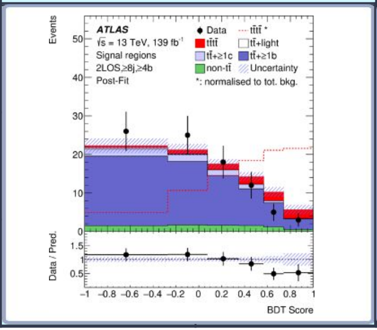
Long history: single top quark production

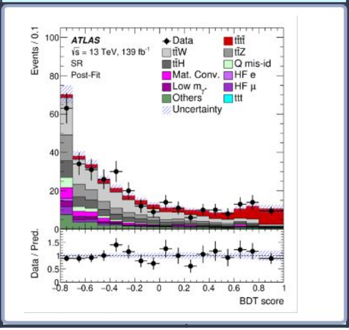
- ➢ 2009 → observation @ Tevatron [arXiv:0906.0523]
- ➢ 2011 → ML aided measurement by CMS [PhysRevLett.107.091802] & ATLAS [arXiv:1205.3130v3]

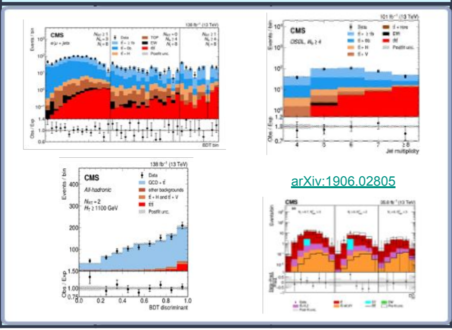
ML in tttt

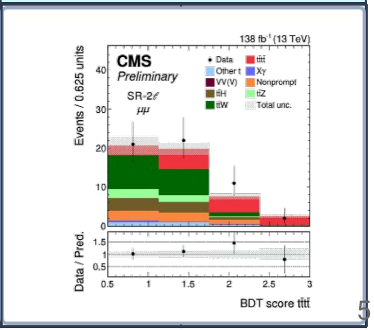
- ➢ Many searches through the years → **Boosted Decision Trees (BDTs) as the main tool**

ATLAS EXPERIMENT	0ℓ	1ℓ	2ℓOS	2ℓSS	3ℓ+	CMS	0ℓ	1ℓ	2ℓOS	2ℓSS	3ℓ+
		arXiv:2106.11683		arXiv:2007.14858			arXiv:2303.03864			arXiv:2305.13439	
2018						2018					
2017						2017					
2016						2016					









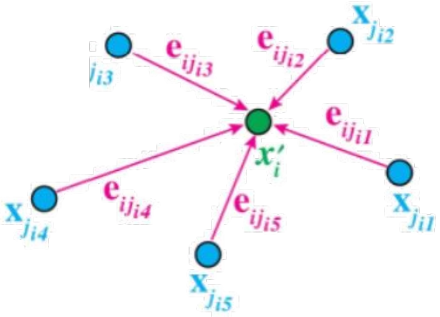
tttt @ ATLAS with Neural Networks (NNs)

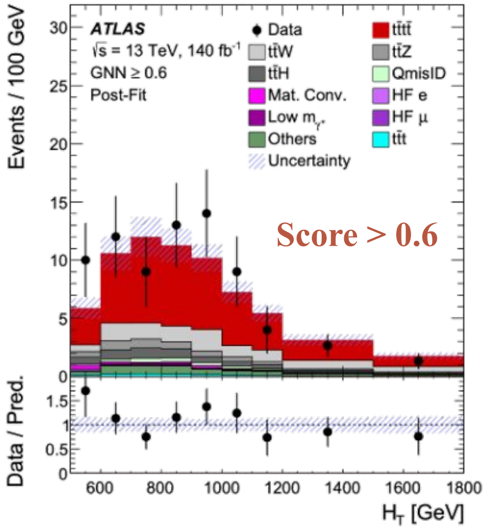
arXiv:2303.15061

2ℓSS & 3ℓ+ channels

- ➢ ttW BKG estimation procedure
- ➢ Separate treatment of 3t, tttW, tttq
- ➢ Lower jet and lepton p_T cuts

Edge convolution [1801.07829]



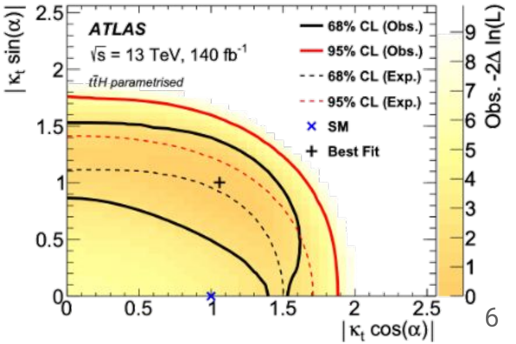


Graph NN (GNN) [GRAPH_NETS]

- ➢ Edge-Convolution layers (multi-jet correlation)
- ➢ **Each node**
  - ○ jets, leptons, E_T^{miss}
  - ○ 4-momenta, b-tagging scores, lepton charges, multiplicities
- ➢ **Edges:** 3 angular separation features
- ➢ Classifier training for xSec measurement

Constraints on $\varkappa_t$ and CP mixing angle α

$|\kappa_t| < 1.8\ (1.6)$

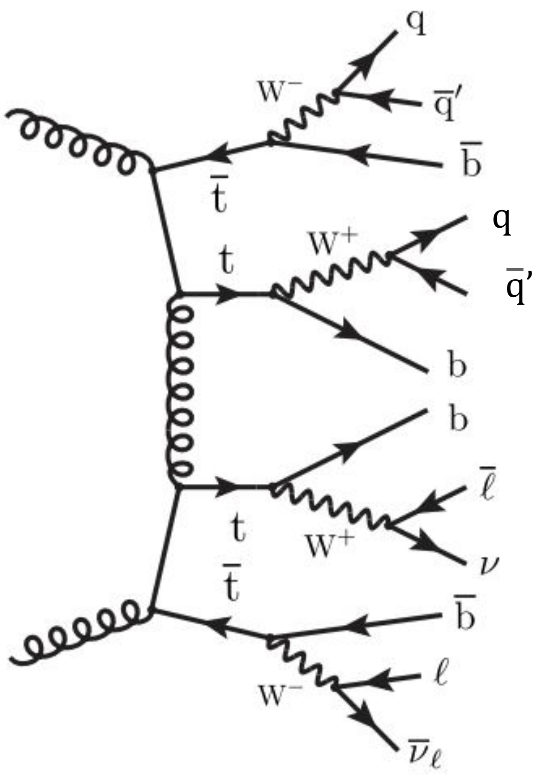


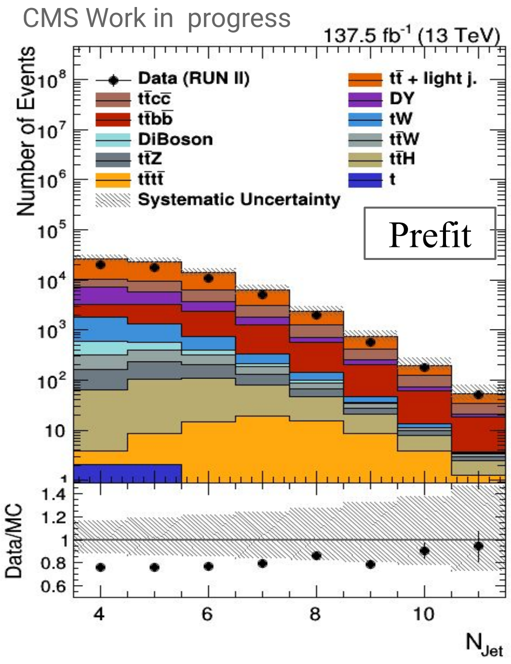
6.1σ (4.3σ) expected, consistent with SM

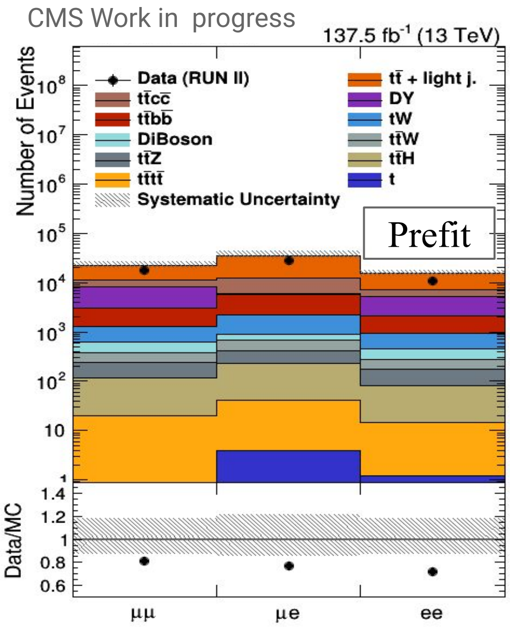
2ℓOS channel

arXiv:2303.03864

3rd Highest Branching Ratio



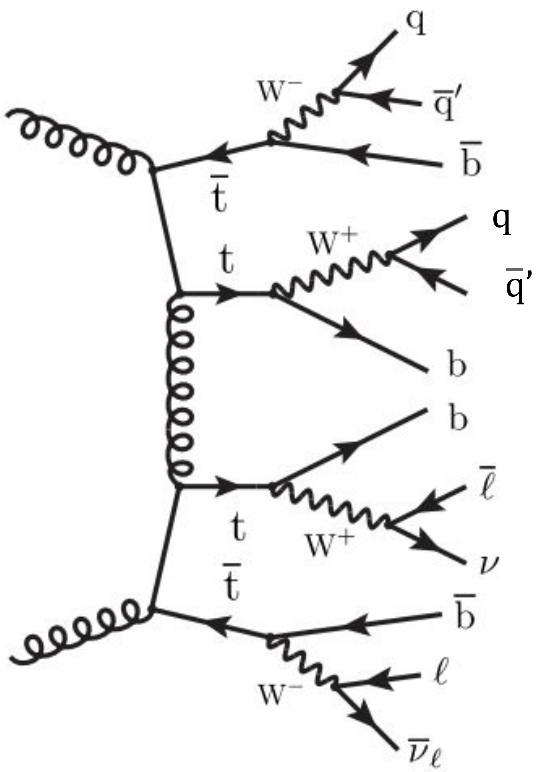




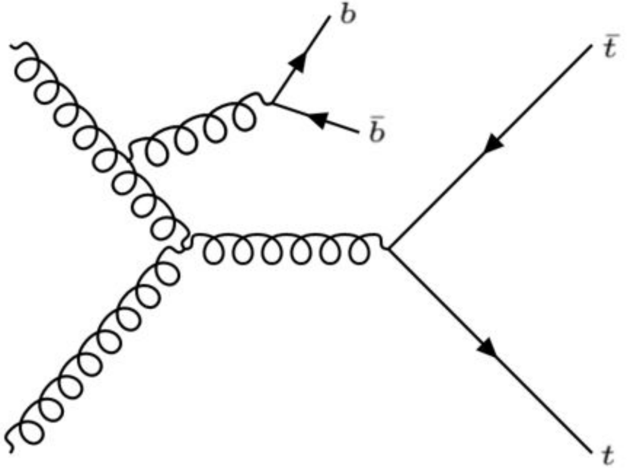
2ℓOS channel

arXiv:2303.03864

3rd Highest Branching Ratio



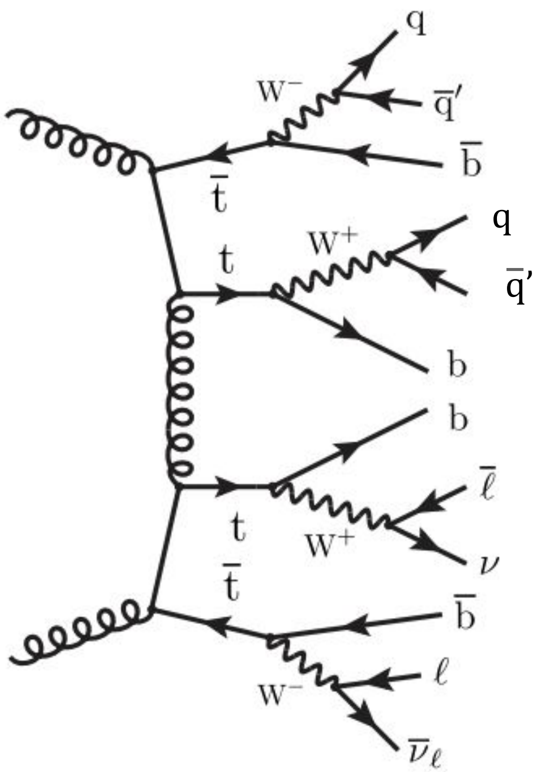
Main irreducible background



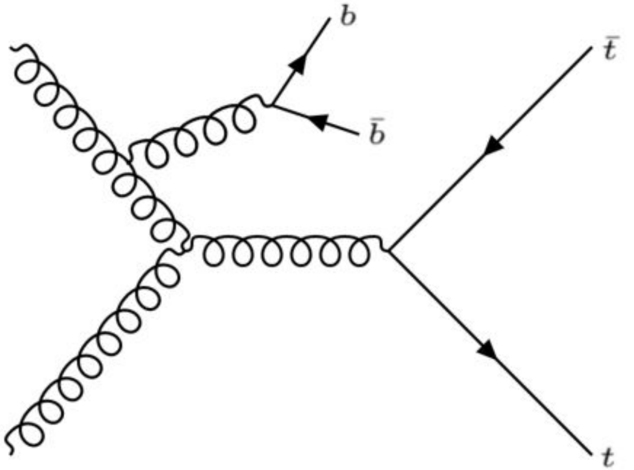
2ℓOS channel

arXiv:2303.03864

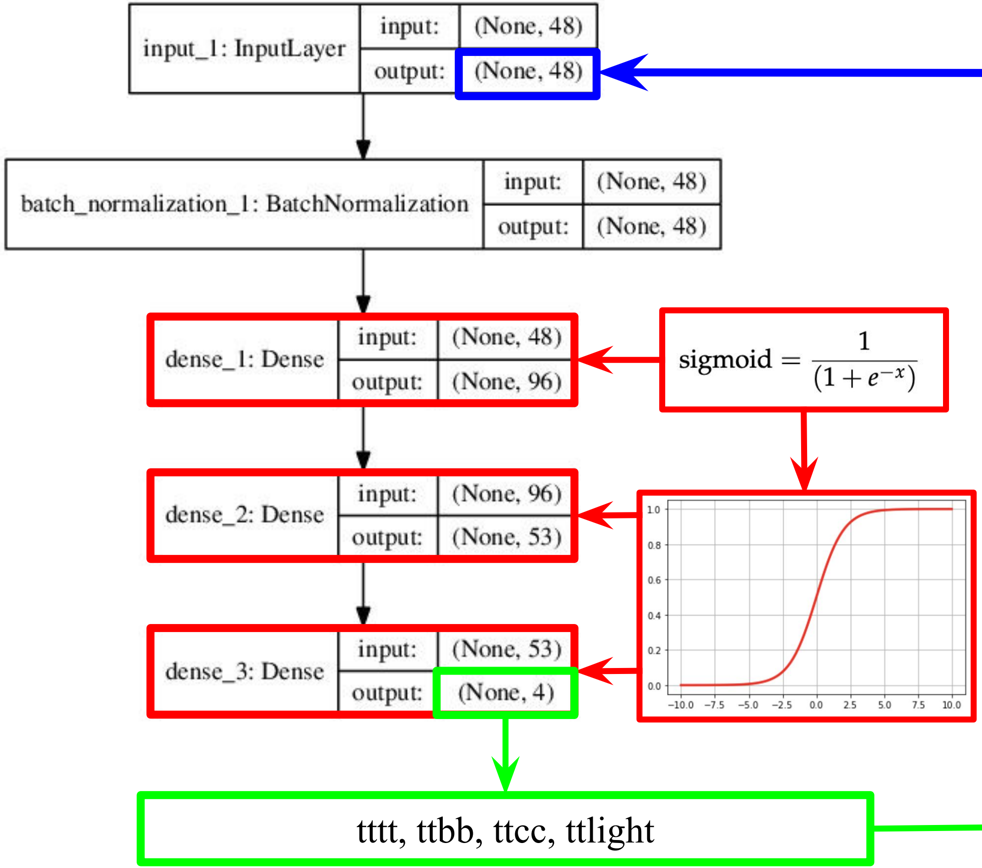
3rd Highest Branching Ratio



Main irreducible background



ML approach



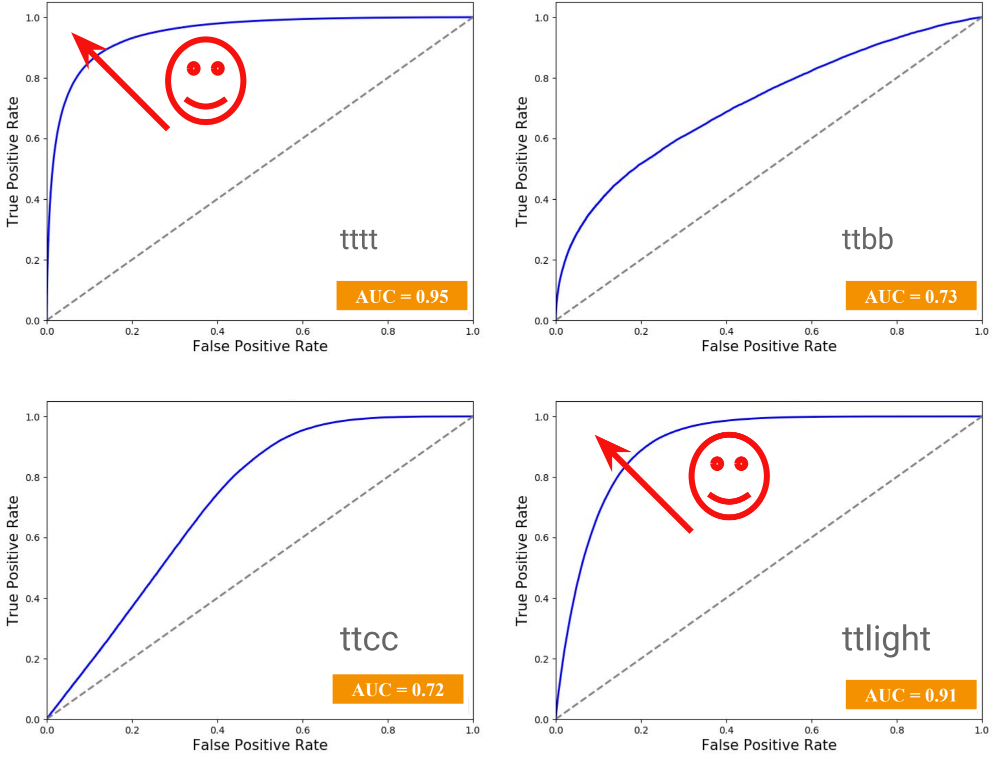
Input features:

- N_{jets}
- m_T of leptons, jets, lepton/jet-system
- p_T, η, ϕ of (sub)leading lepton and first 3 leading jets
- p_T of first 8 jets
- H_T, $H_{T,b}$, ΔH_T
- Δϕ. Δη of leptons and jet system
- ΔR between b-jets, b-jets & leptons
- m_{4b}
- m^2_T of leptons and b-jets

Category	Count
TTTT	1292052
TTLight	890844
TTCC	255187
TTBB	97361

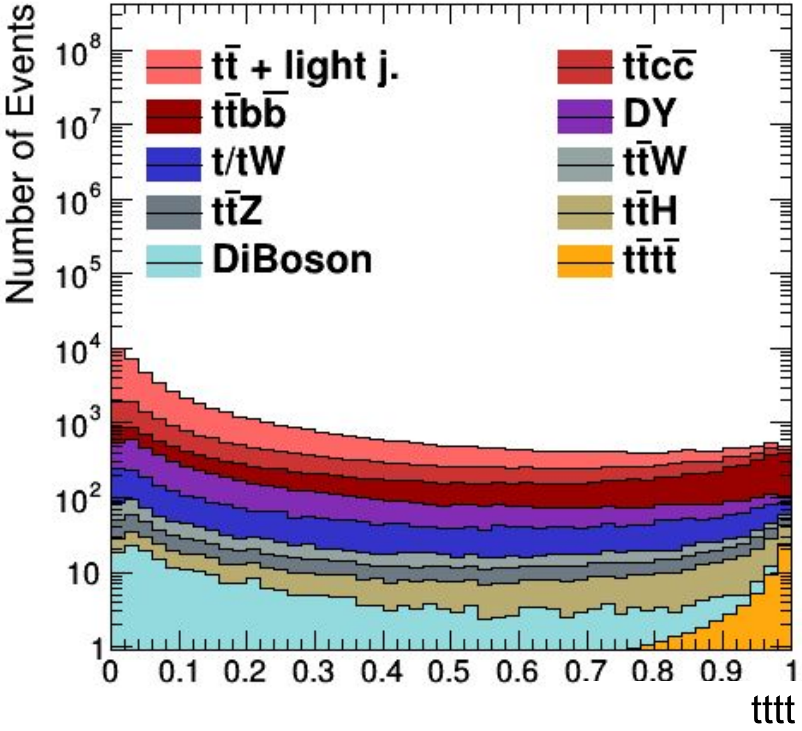
Event level Neural Network (NN)

MVA Evaluation



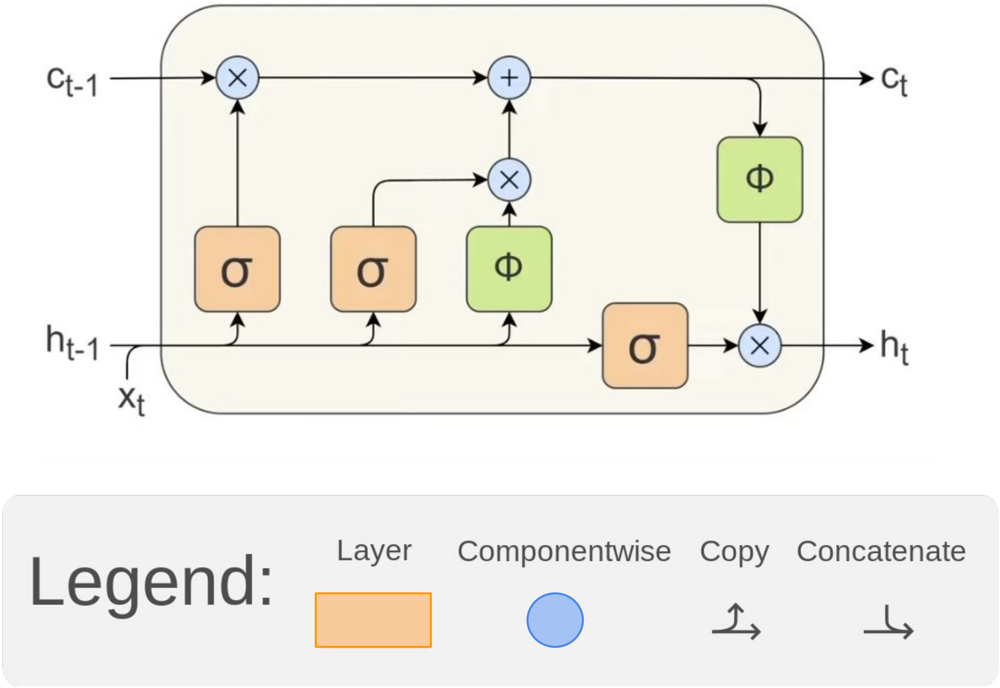
- ➢ Better performance for tttt and ttlight

MVA Evaluation



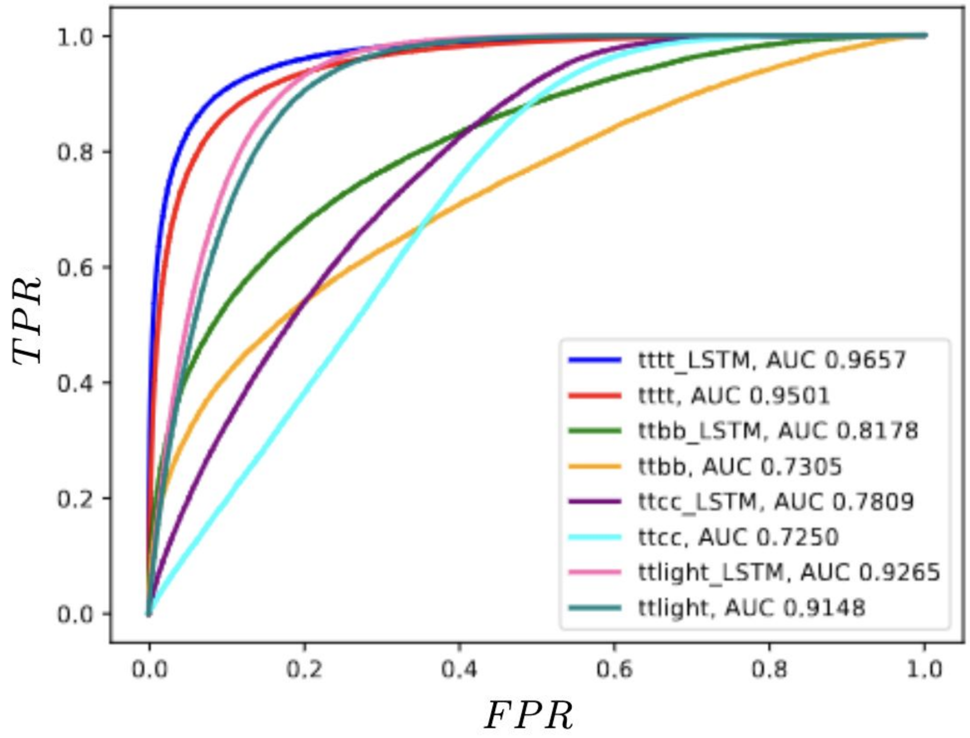
- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?

MVA Evaluation



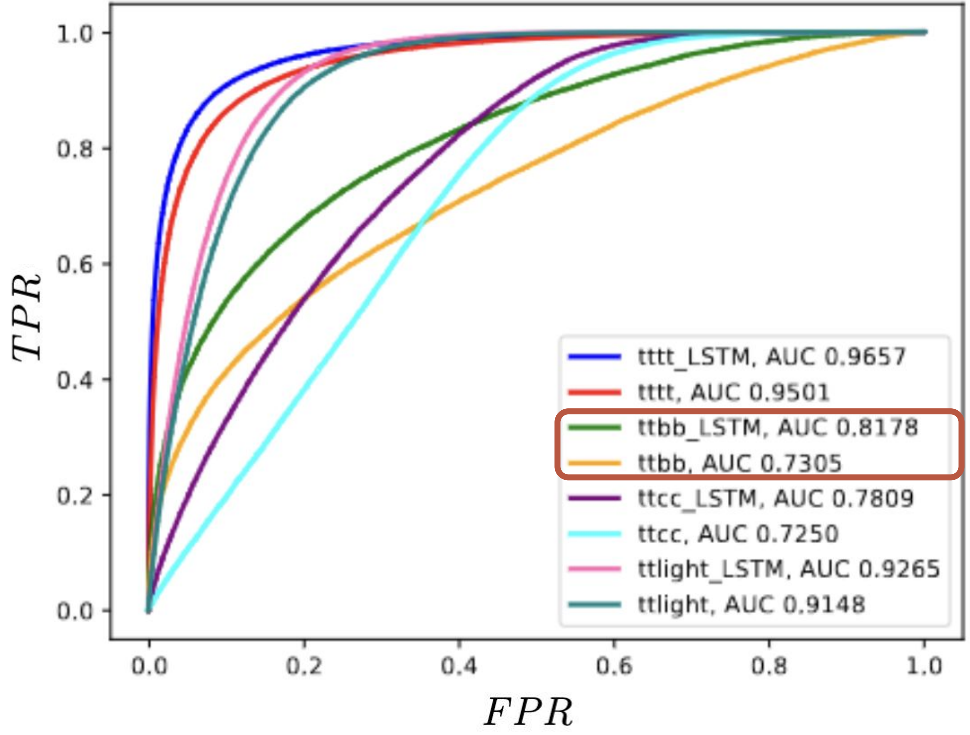
- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer

MVA Evaluation



- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer

MVA Evaluation

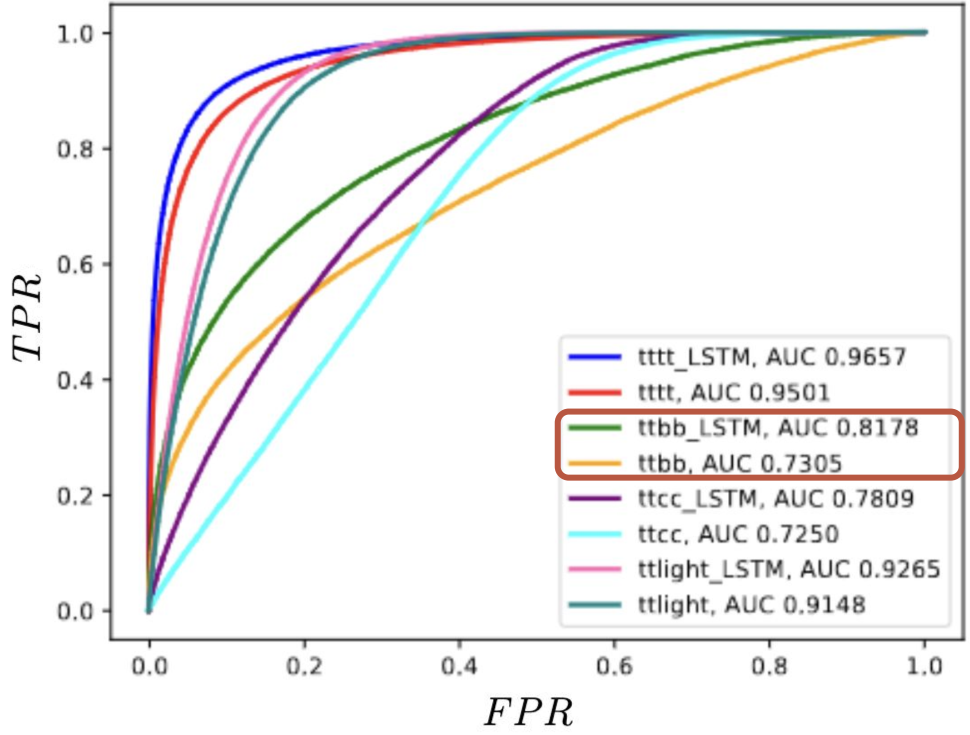


- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer

1% increase for tttt, ttcc, ttlight

3% for ttbb

MVA Evaluation

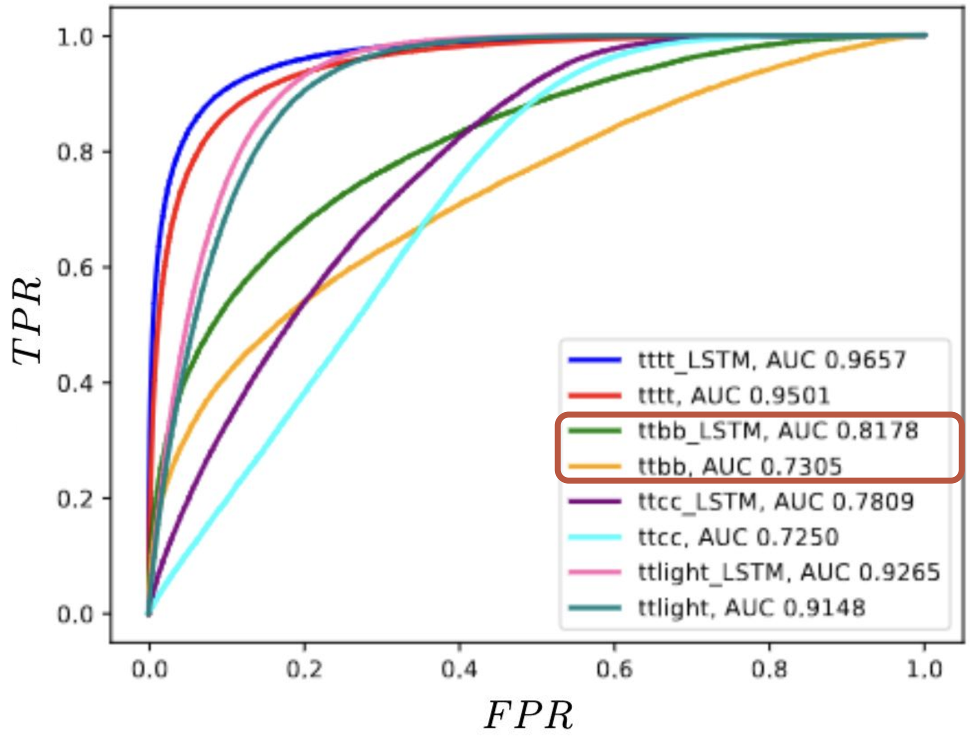


- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer
  - ○ Adding topology variables + LSTM

$$M^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |p_i|^2} \longrightarrow S = \frac{3}{2}(\lambda_2 + \lambda_3)$$

$$C = \frac{H_T}{\sum_i E_i}$$

MVA Evaluation



- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer
  - ○ Adding topology variables + LSTM

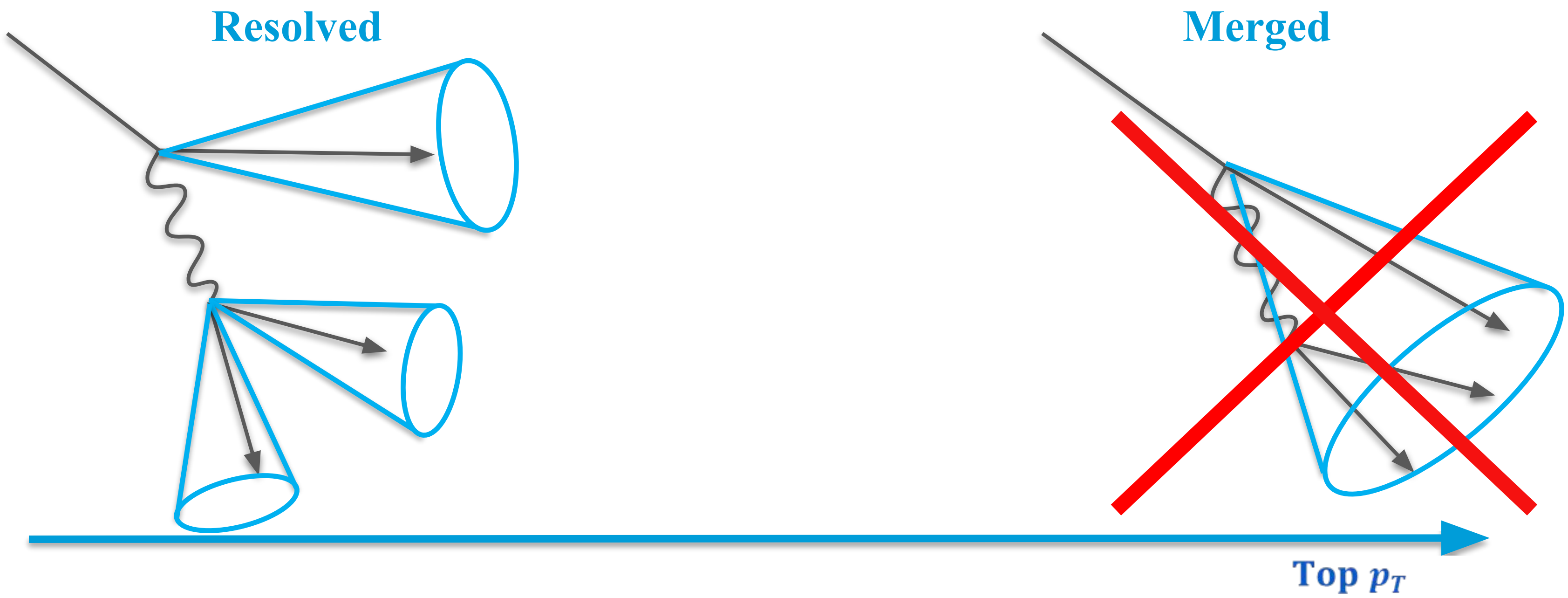
$$\boxed{M^{\alpha\beta} = \frac{\Sigma_i p_i^{\alpha} p_i^{\beta}}{\Sigma_i |p_i|^2} \longrightarrow S = \frac{3}{2}(\lambda_2 + \lambda_3)}$$

$$\boxed{C = \frac{H_T}{\Sigma_i E_i}}$$

- ➢ Tagging hadronic component

Top quark configurations

2 possible configurations based on the angular superposition between quarks and b-jet

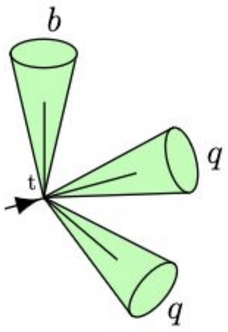


Category definition

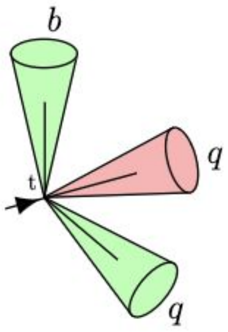
tttt

Adapting the strategy from a previous charged Higgs search

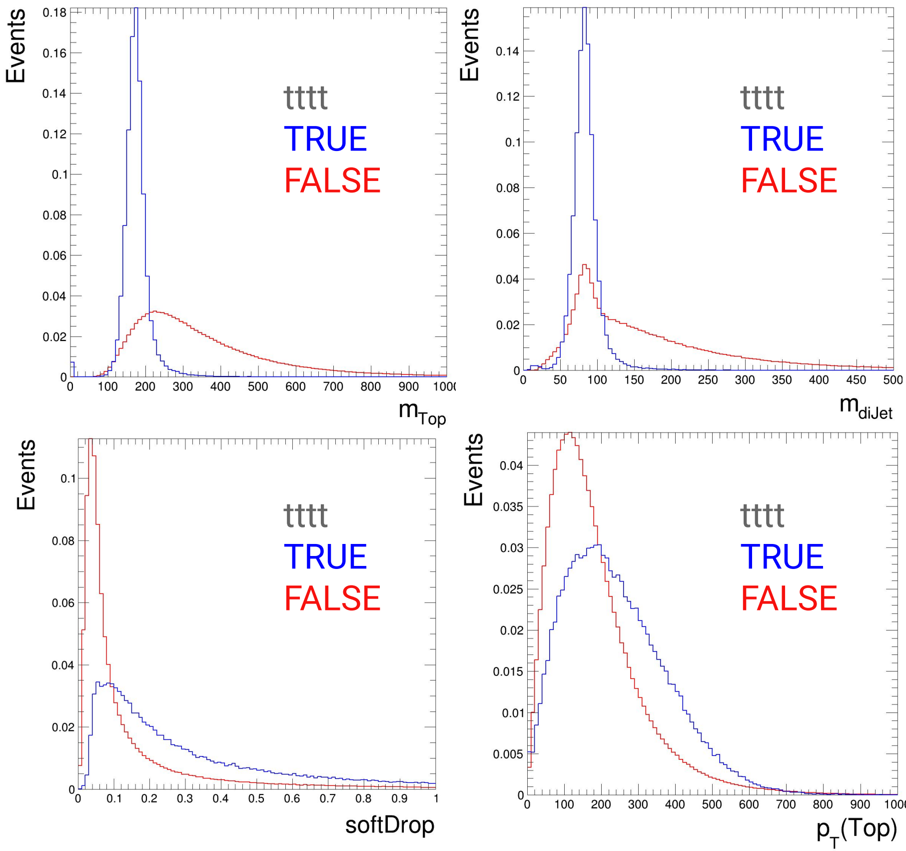
- **Signal** truth-matched trijets
- **Background** at least one non-matched jet
- $\Delta R(q, j) < 0.4$



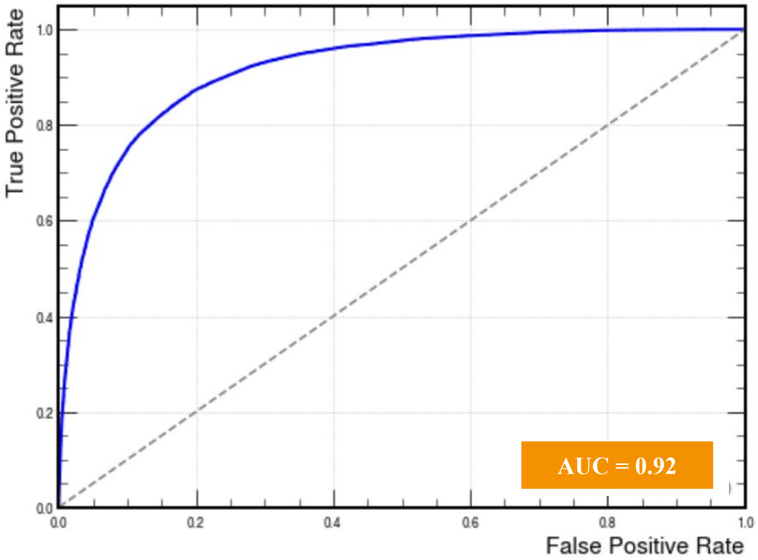
Signal



Background



Classifier performances

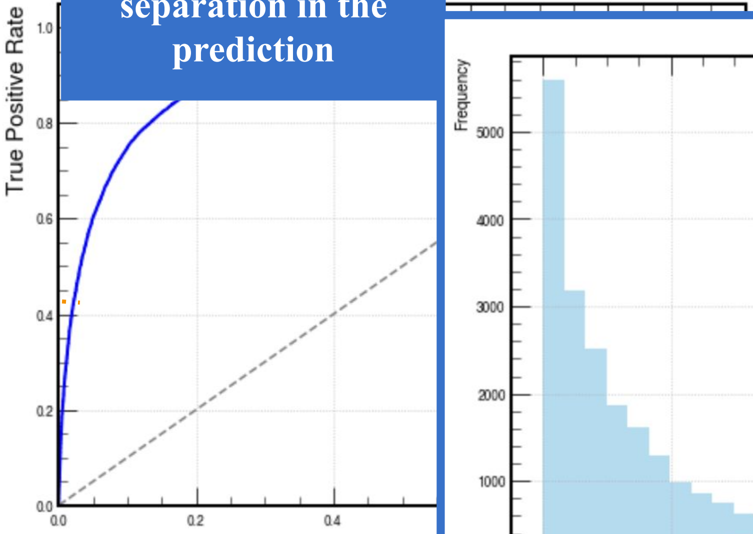


Good performance with a Keras DNN with 2 Dense Layers (64/32), a Dropout Layer (20%), and a ~300k balanced dataset normalized with a robust scaling technique.

Define 4 WPs →

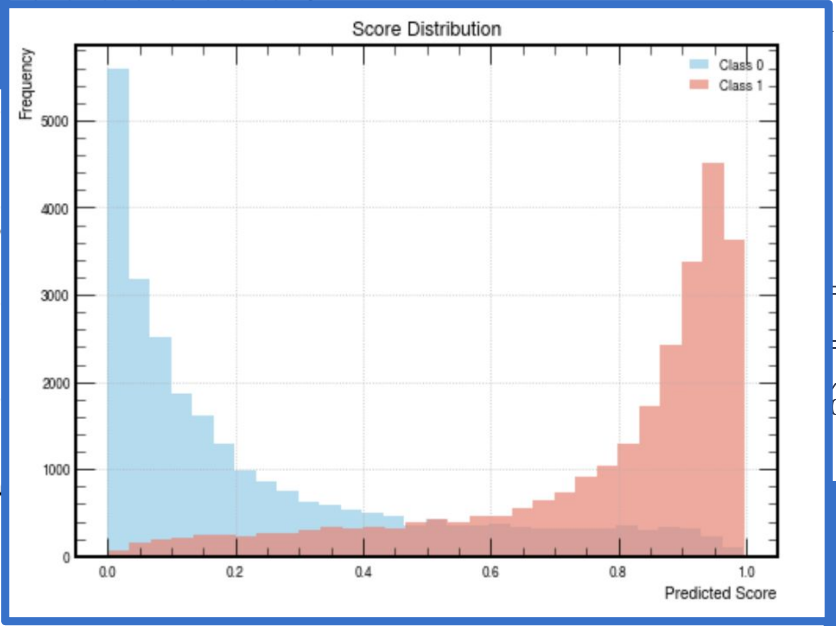
- FPR @ 10% = 0.675 → L
- FPR @ 5% = 0.599 → M
- FPR @ 1% = 0.936 → T
- FPR @ 0.1% = 0.977 → VT

Classifier performances



Good level of separation in the prediction

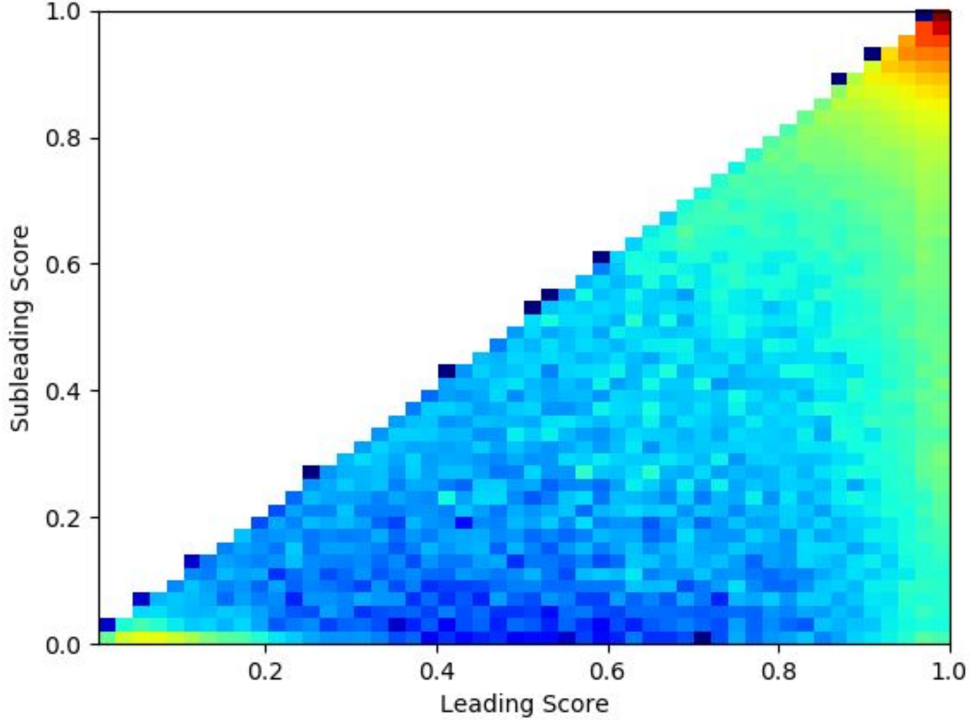
Good performance with a Keras DNN with 2 Dense Layers (64/32), a Dropout Layer (20%), and a ~300k balanced dataset
st scaling technique.



= 0.675 → L
= 0.599 → M
= 0.936 → T
= 0.977 → VT

Residual misclassification due to low statistics

Classifier performances



- ➢ Most signal events have two high scores
- ➢ Tail to low values for the subleading
- ➢ Very few events have no signal-like hadronic top candidates

Summary

- ➢ tttt provides input for future searches
- ➢ BDTs as leading classifiers
- ➢ 2lOS channel not included in the observation
- ➢ Attempt to increase sensitivity via ML techniques
- ➢ Event level MVA shows limited classification capability
- ➢ Extract information on hadronic top quarks

Summary

- ➢ tttt provides input for future searches
- ➢ BDTs as leading classifiers
- ➢ 2lOS channel not included in the observation
- ➢ Attempt to increase sensitivity via ML techniques
- ➢ Event level MVA shows limited classification capability
- ➢ Extract information on hadronic top quarks

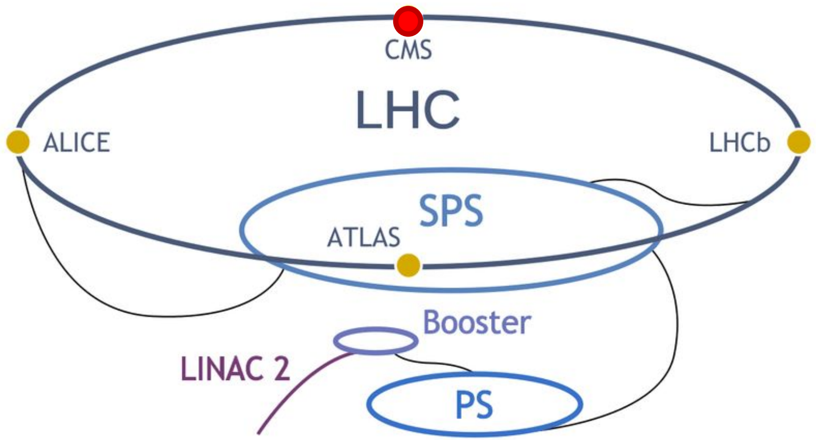
Thank you

Backup

References

- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (Year). "Dynamic Graph CNN for Learning on Point Clouds." Massachusetts Institute of Technology, UC Berkeley / ICSI, Imperial College London / USI Lugano. arXiv:1801.07829v2 [cs.CV], 11 Jun 2019
- GitHub repository: Google DeepMind. "graph_nets." https://github.com/google-deepmind/graph_nets
- Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-term Memory." Neural computation, 9(8), 1735-80. DOI: 10.1162/neco.1997.9.8.1735
- Hayrapetyan, A. et al. "Observation of four top quark production in proton-proton collisions at √s = 13 TeV." *Physics Letters B* 847 (2023). DOI: 10.1016/j.physletb.2023.138290
- Aad, G. et al. "Observation of four-top-quark production in the multilepton final state with the ATLAS detector." The European Physical Journal C 83, no. 6 (2023). DOI: 10.1140/epjc/s10052-023-11573-0\
- Gillberg, D. (for the D0 and CDF Collaborations). "Discovery of Single Top Quark Production." Proceedings for Moriond QCD and High Energy Interactions, March 14th - March 21st 2009. arXiv:0906.0523 [hep-ex]. DOI: 10.48550/arXiv.0906.0523.
- Chatrchyan, S. et al. (CMS Collaboration). "Measurement of the t-Channel Single Top Quark Production Cross Section in pp Collisions at √s=7 TeV." Phys. Rev. Lett. 107, 091802 (2011). DOI: 10.1103/PhysRevLett.107.091802
- ATLAS Collaboration. "Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at √s=13 TeV with the ATLAS detector." Journal of High Energy Physics 11 (2021) 118. DOI: 10.1007/JHEP11%282021%29118.
- CMS Collaboration. "Evidence for four-top quark production in proton-proton collisions at √s = 13 TeV." Phys. Lett. B 844 (2023) 138076. DOI: 10.1016/j.physletb.2023.138076. arXiv:2303.03864 .
- CMS Collaboration. "Observation of four top quark production in proton-proton collisions at √s = 13 TeV." Phys. Lett. B 847 (2023) 138290. DOI: 10.1016/j.physletb.2023.138290. arXiv:2305.13439.
- ATLAS Collaboration. "Observation of four-top-quark production in the multilepton final state with the ATLAS detector." Eur. Phys. J. C 83 (2023) 496. DOI: 10.1140/epjc/s10052-023-11573-0. arXiv:2303.15061.

Large Hadron Collider (LHC)



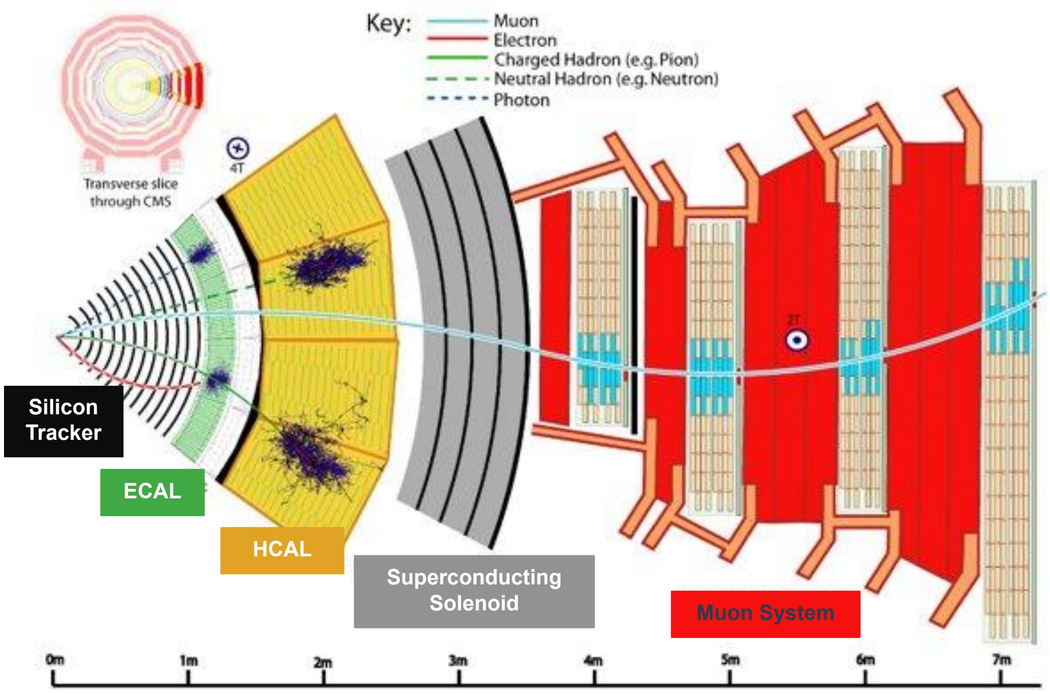
LINAC 2	Linear Accelerator 2
PS	Proton Synchrotron
SPS	Super Proton Synchrotron
LHC	Large Hadron Collider

- **Proton-proton collision with (Run-II 2016-2018)**
  - Centre of mass energy
    $\sqrt{s} = 13\ TeV$
  - Peak instantanous luminosity
    $L = 2 \times 10^{34}\ cm^{-2}s^{-1}$
- **4 main experiments**
  - ALICE
  - ATLAS
  - LHCb
  - **CMS**

Compact Muon Solenoid (CMS)

Cylindrical detector made of:

- ➢ **Silicon Tracker;**
- ➢ **Electromagnetic Calorimeter (ECAL);**
- ➢ **Hadronic Calorimeter (HCAL);**
- ➢ **Superconducting Solenoid;**
- ➢ **Muon System.**

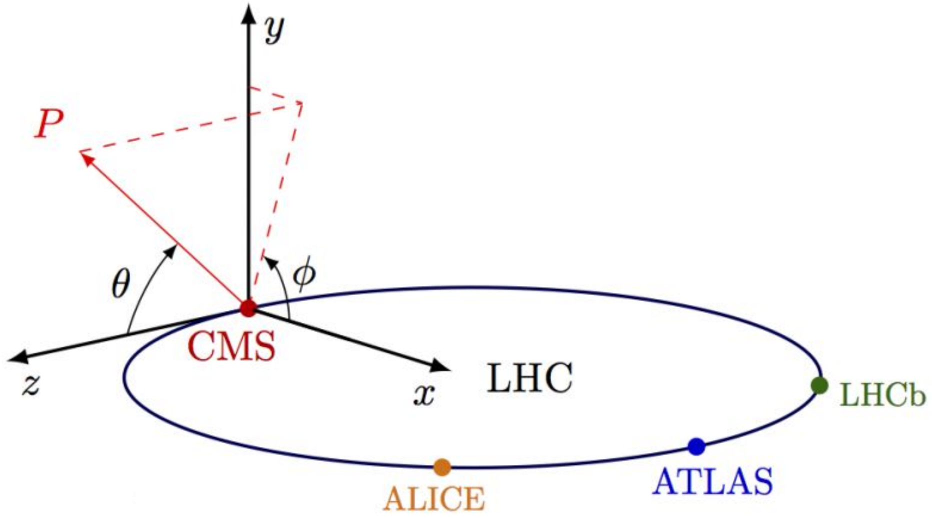


CMS reference frame

- **r :** radial distance from z-axis;
- Φ : angle on (x-y) plane;
- **η :** pseudorapidity, defined as:

$$\eta = -\ ln\left(tan\frac{\theta}{2}\right)$$

In which θ is the polar angle



Angular distances between objects using Φ and

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$$

tttt @ LHC

Many SM and BSM searches through the years

SM

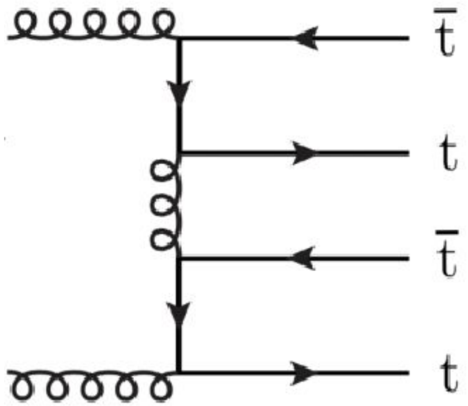
CMS
2ℓSS & multilepton: [TOP-22-013]
All-hadronic & 1ℓ & 2ℓOS: [TOP-21-005]

ATLAS
2ℓSS & multilepton: [CERN-EP-2023-055]
2ℓSS & multilepton: [Eur. Phys. J. C 80 (2020) 1085]

BSM

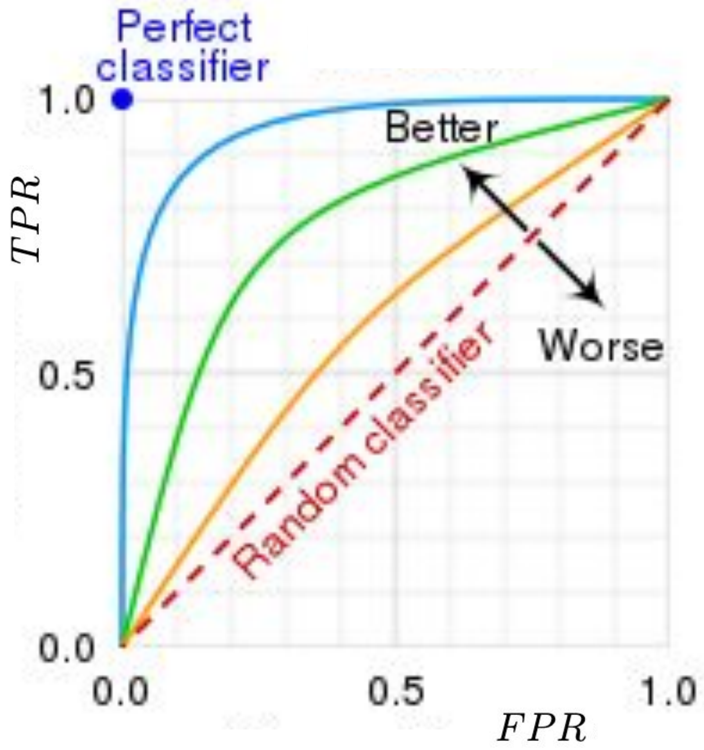
CMS
tttt + MET SUSY: [SUS-21-007]
type II HDM: [TOP-18-003]

ATLAS
top-philic resonances : [CERN-EP-2023-048]
Heavy Higgs: [ATLAS-CONF-2022-039] [CERN-EP-2022-170]
R-parity violating SUSY: [Eur. Phys. J. C 81 (2021) 102]
Gluinos + bjets: [ATLAS-CONF-2018-041]



MVA Evaluation

Model performance given in terms of ROC Area Under Score (AUC)



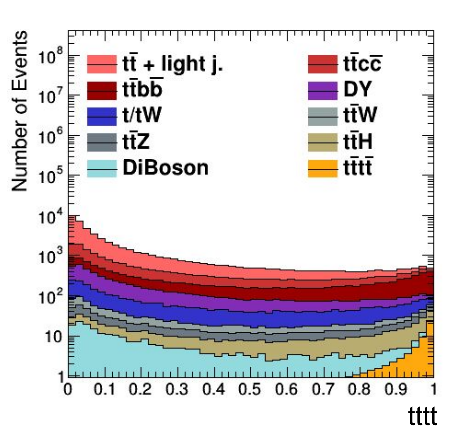
$$TPR = \frac{TP}{TP + FN}$$

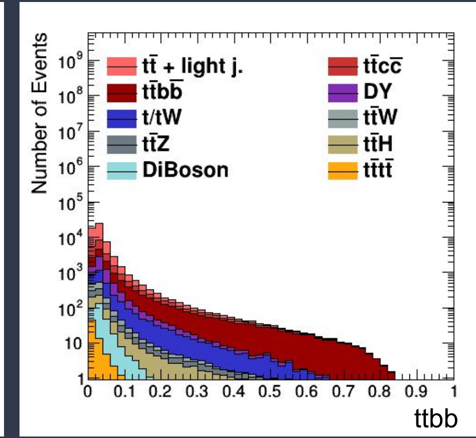
$$FPR = \frac{FP}{FP + TN}$$

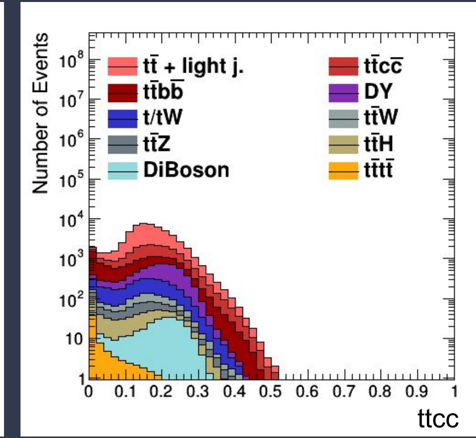
Where

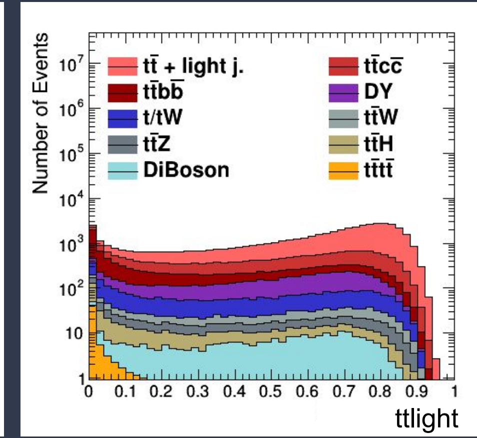
- ➢ TP = True Positive
- ➢ FP = False Positive
- ➢ TN = True Negative
- ➢ FN = False Negative

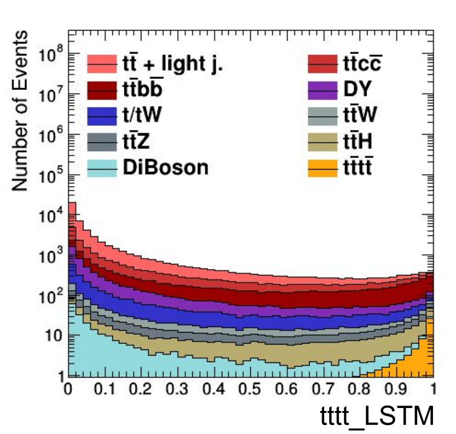
Model performances

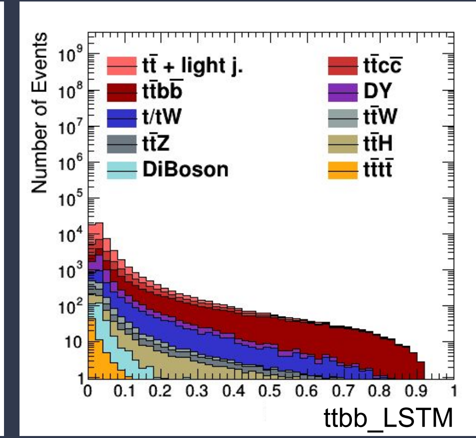


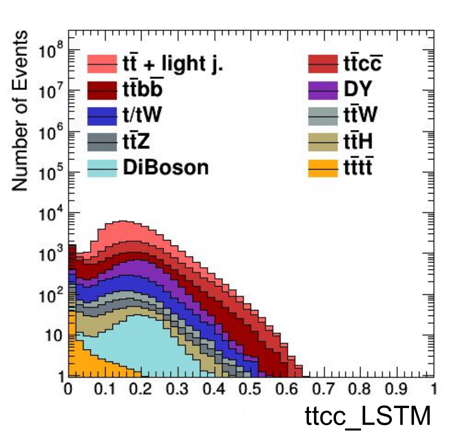


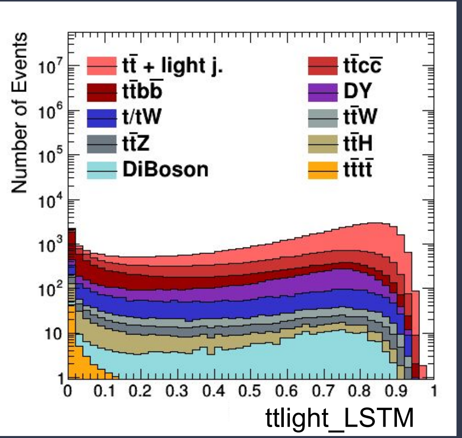




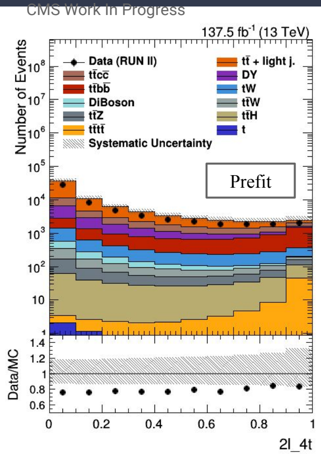


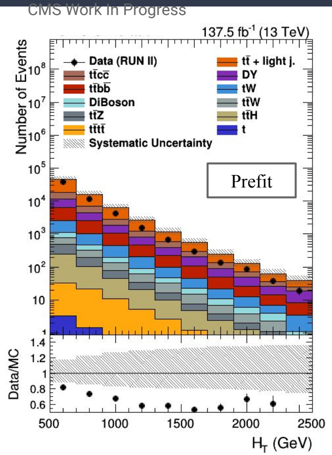


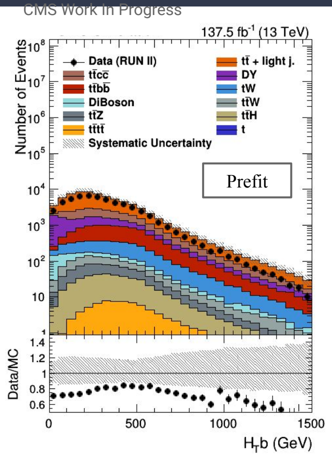


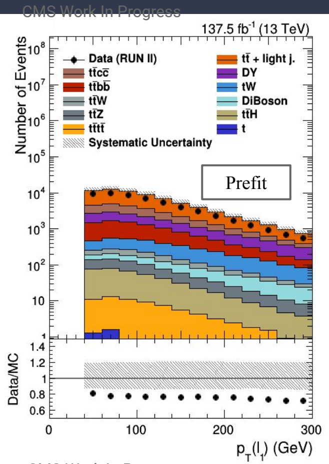


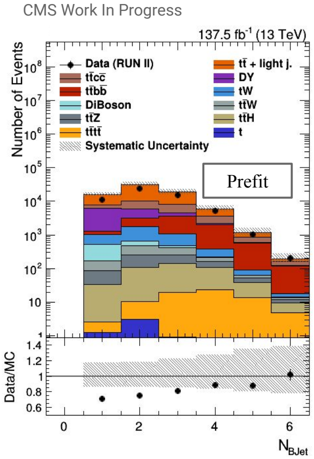
2lOS inclusive plots

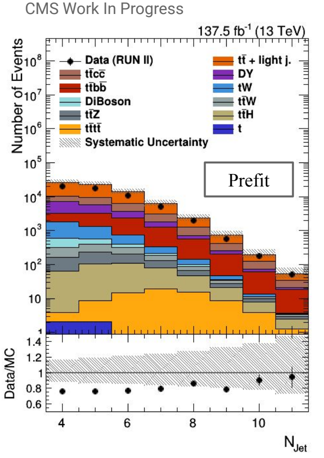


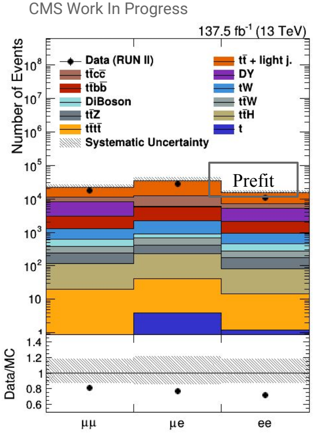


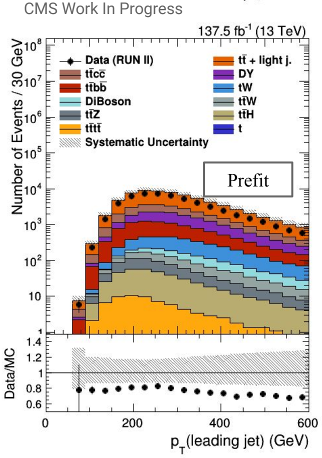












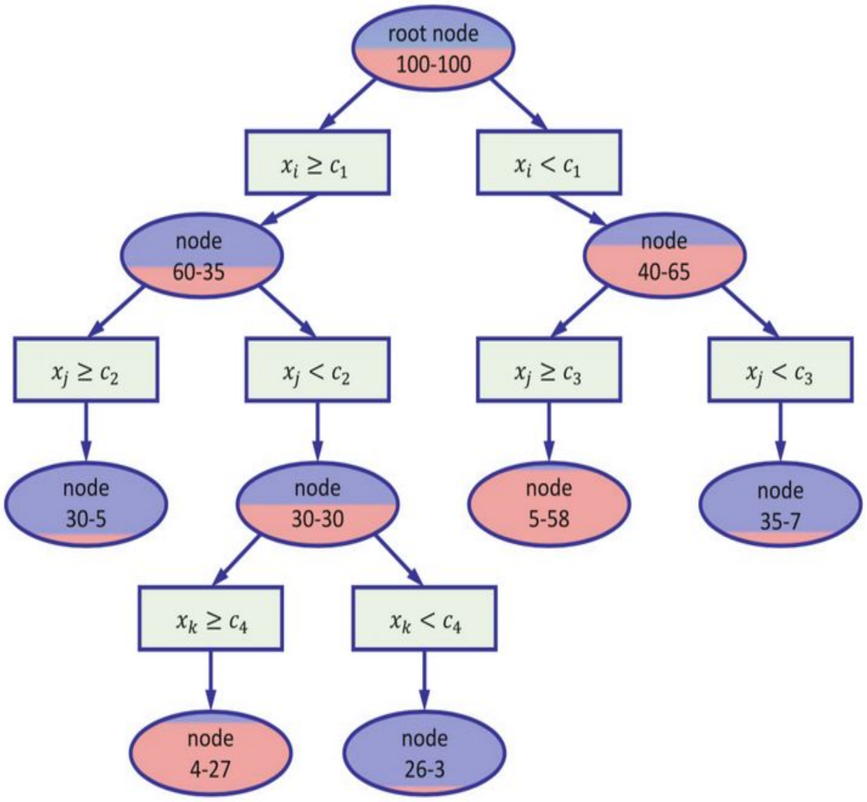
Boosted Decision Trees (BDTs)

Supervised Learning: classification known from the outset

- **Input:** kinematics and high level variables
- **Output:** score ∈ [0,1]

Many benefits…

- **Efficiency:** training and predictions are very fast
- **Not scale sensitive:** mix of continuous and discrete variables
- **Non linear decision boundaries:** modelling complex non linear relationships
- **Easy interpretability:** importance of each feature known



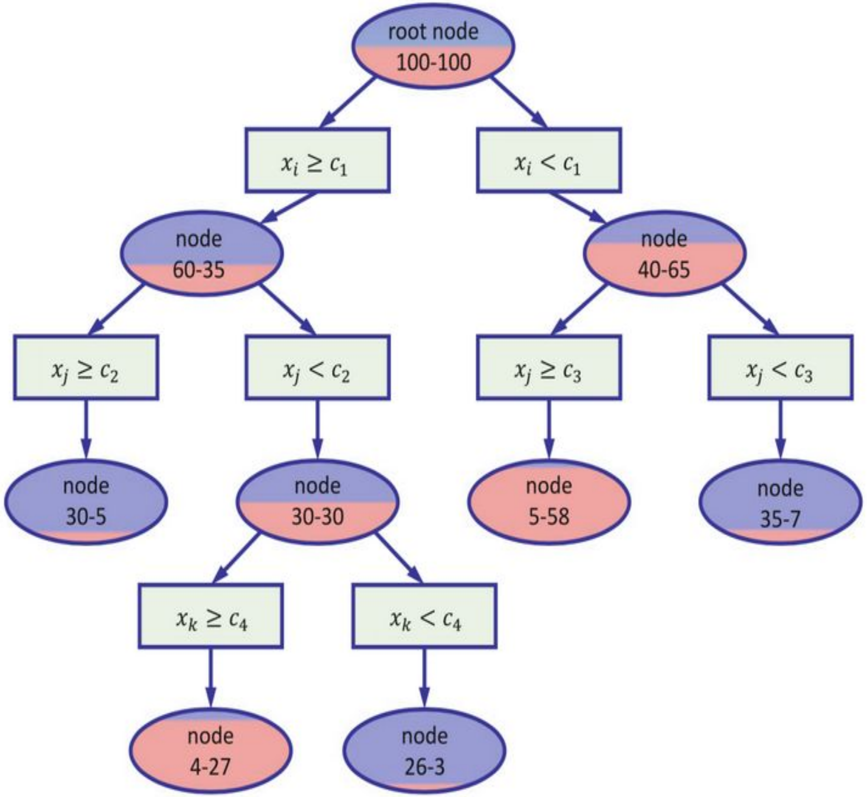
Boosted Decision Trees (BDTs)

Supervised Learning: classification known from the outset

- **Input:** kinematics and high level variables
- **Output:** score ∈ [0,1]

… and many limitations

- **Sensitive to overfitting:** always cross validate!
- **Limited generalization:** small or noisy datasets affect prediction
- **Difficulty with large datasets:** can easily become computationally expensive



tttt @ CMS with BDTs

arXiv:2305.13439

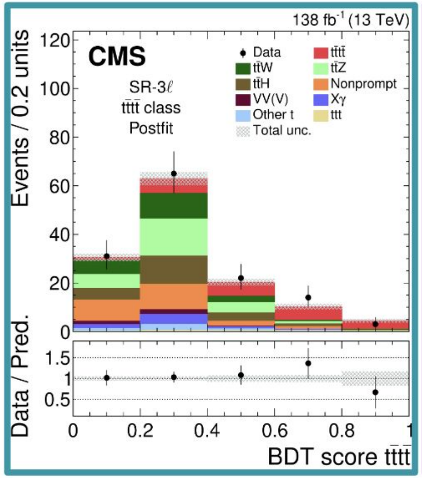
2ℓSS & 3ℓ & 4ℓ

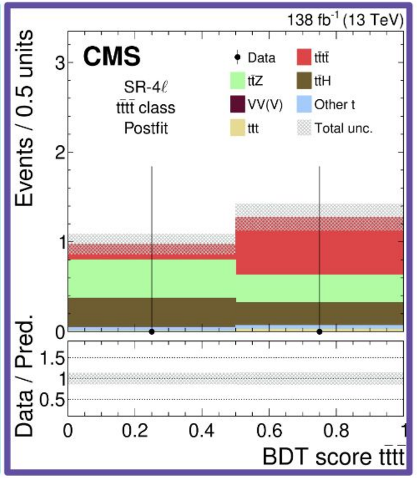
- ➢ ttW modelling: NLO QCD MC
- ➢ Additional large uncertainty on tt̄W + jets
- ➢ Improved lepton ID, b-tagging, SR selection

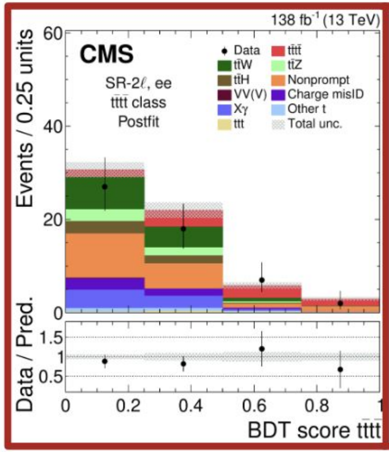
Event level BDT

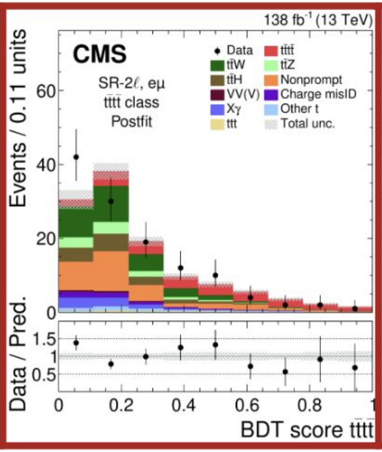
- ➢ Multivariate analysis: tttt, ttX, tt
- ➢ 31 input features
- ➢ Different BDTs trained for different channels to account for kinematic differences
- ➢ Thorough tuning and cross validation

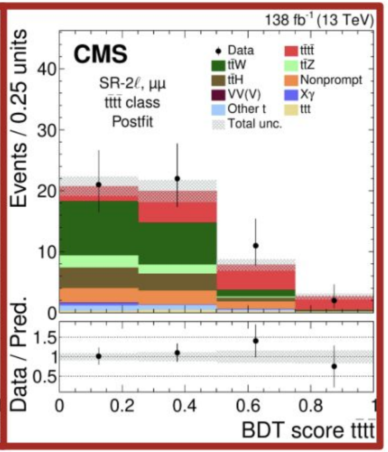
Fit optimized by flavour splitting the SR
5.5σ (4.9σ) expected, consistent with SM







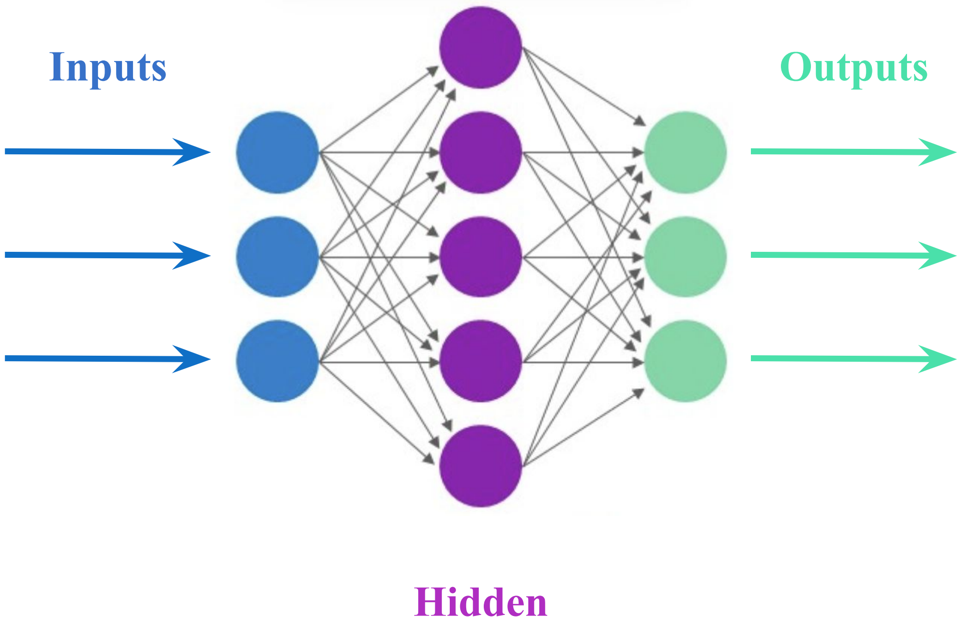




Neural Networks (NNs)

Layers of interconnected nodes (neurons) that convert weighted inputs to outputs;
during the learning process these weights are updated. **Recognises patterns in the data.**

Many uses in HEP: tracking, Fast Triggering. detector calibration, background rejection, anomaly detection, jets classification, event selection…



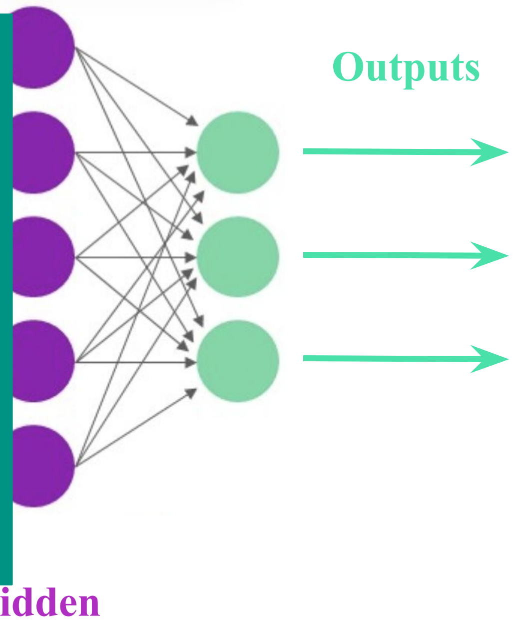
Neural Networks (NNs)

Layers of interconnected nodes (neurons) that convert weighted inputs to outputs; during the learning process these weights are updated. **Recognises patterns in the data.**

Many uses in HEP: tracking, Fast Triggering. detector calibration, background rejection, anomaly detection, jets classification, event selection…

Pros

- ➢ **Pattern Recognition:** recognizing complex patterns, aiding tasks like particle identification and event reconstruction.
- ➢ **Data-driven Insights:** valuable insights from data, revealing correlations not immediately apparent.
- ➢ **Adaptability:** ML adapts to changing conditions, handling diverse datasets effectively.
- ➢ **Improved Sensitivity:** ML techniques enhance experiment sensitivity by improving signal discrimination.



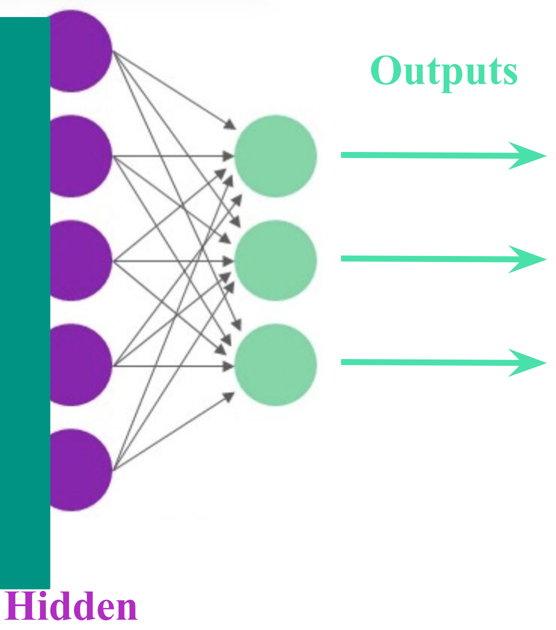
Neural Networks (NNs)

Layers of interconnected nodes (neurons) that convert weighted inputs to outputs; during the learning process these weights are updated. **Recognises patterns in the data.**

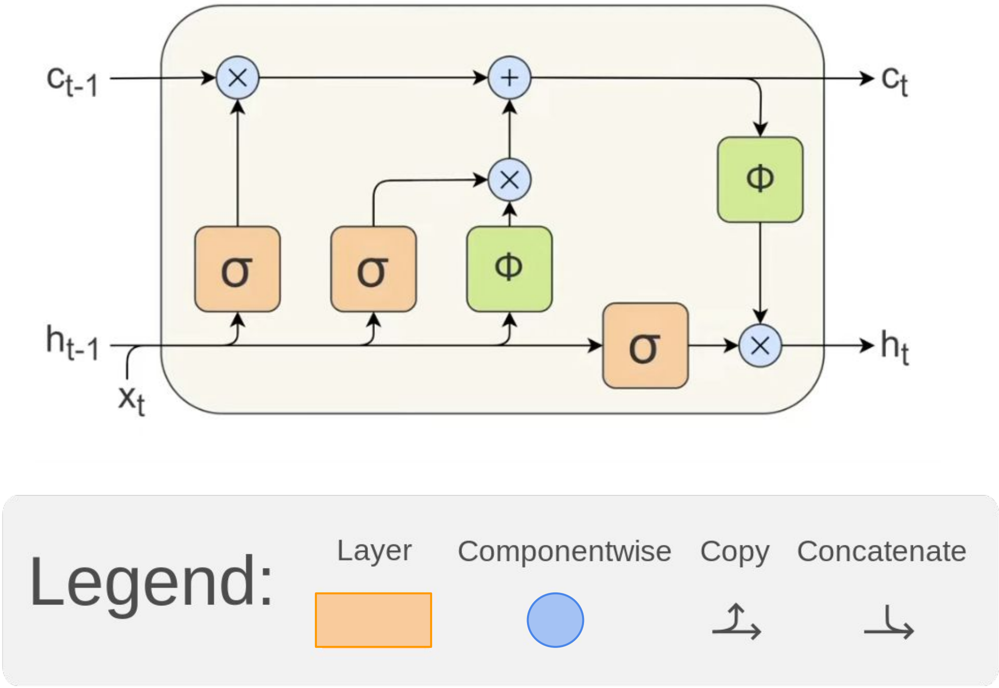
Many uses in HEP: tracking, Fast Triggering. detector calibration, background rejection, anomaly detection, jets classification, event selection…

Cons

- ➢ **Interpretability:** may be difficult to pick up the logic behind the learning
- ➢ **Overfitting:** may capture noise or spurious correlations instead of genuine physics signals
- ➢ **Computational Complexity:** computationally intensive, requiring significant resources
- ➢ **Model Validation and Uncertainty Estimation:** calidating predictions and estimating uncertainty is challenging, requiring rigorous methods

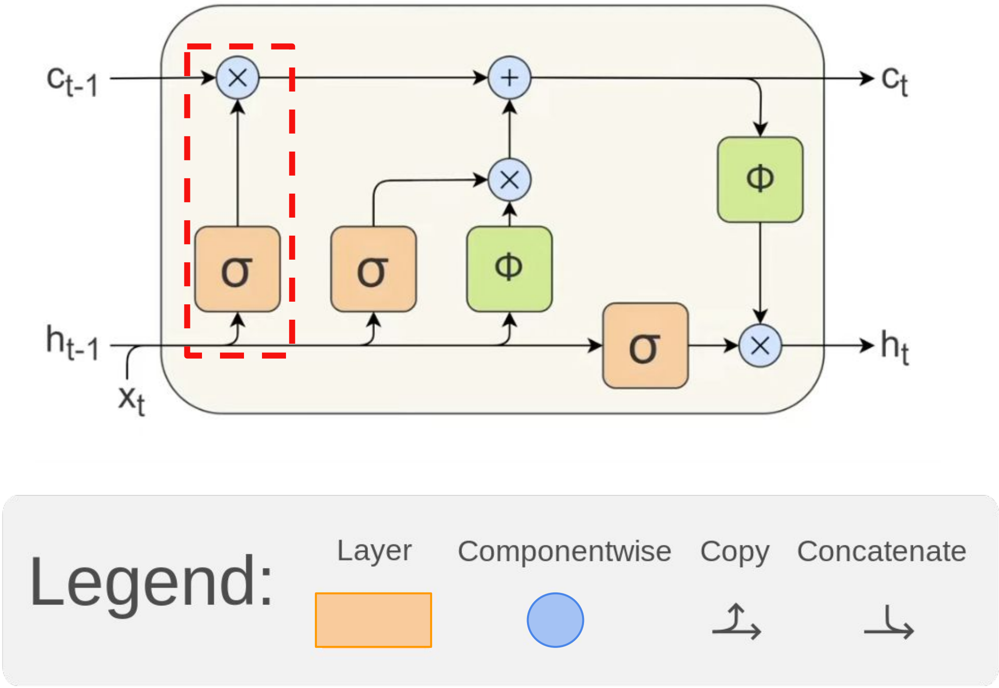


LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$
$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$
$$h_t = o_t \odot \phi(c_t)$$

LSTM Layer



$$\boxed{f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)}$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

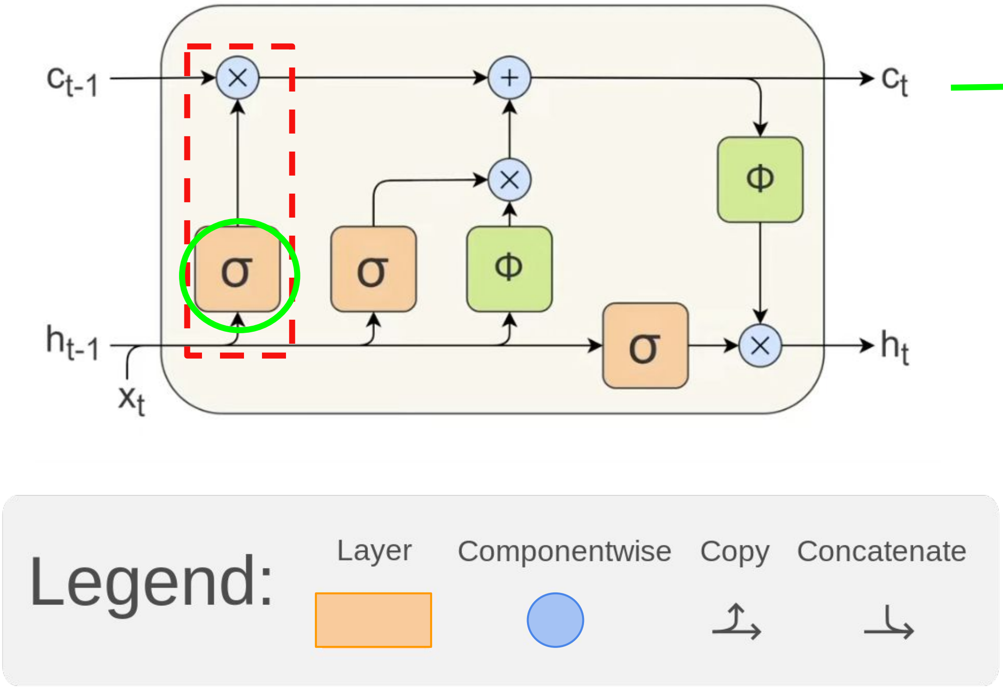
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = \boxed{f_t \odot c_{t-1}} + i_t \odot g_t$$

$$h_t = o_t \odot \phi(c_t)$$

LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

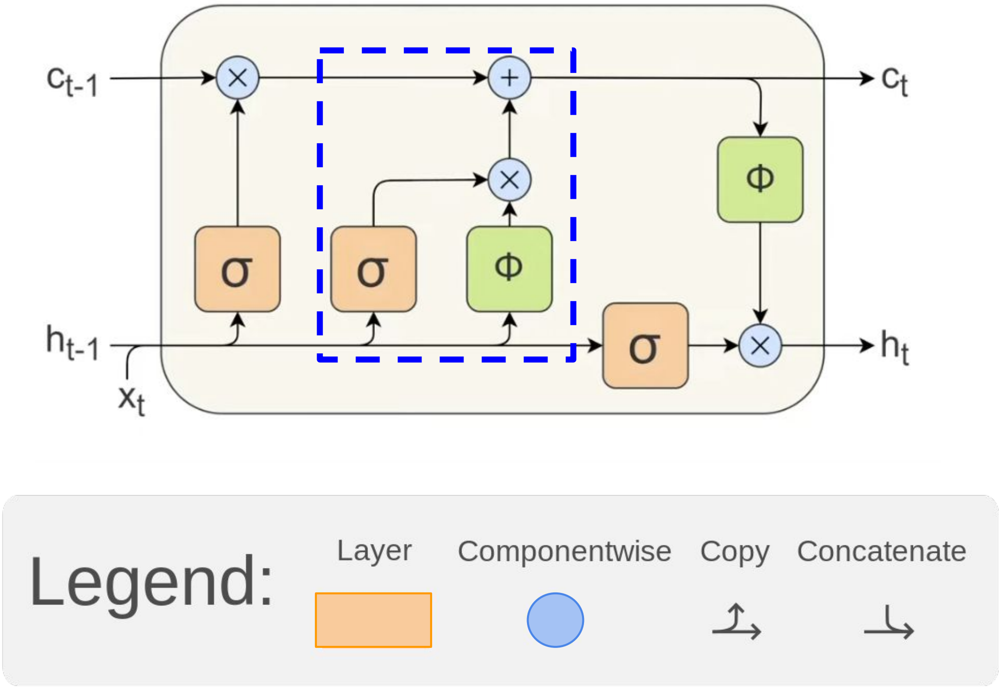
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$h_t = o_t \odot \phi(c_t)$$

LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

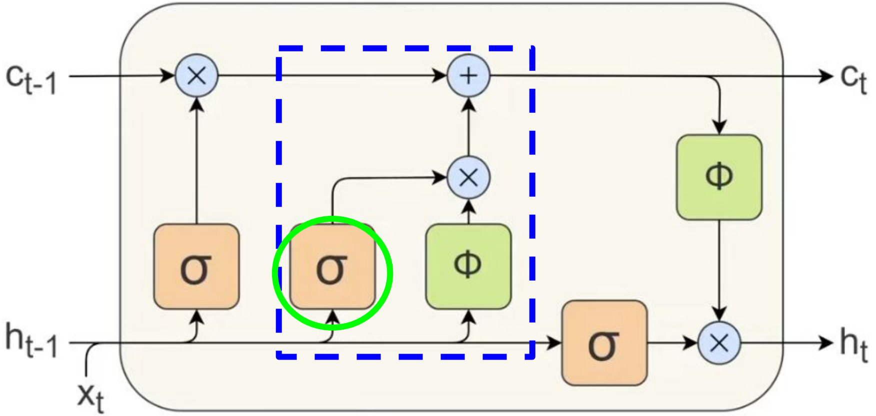
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

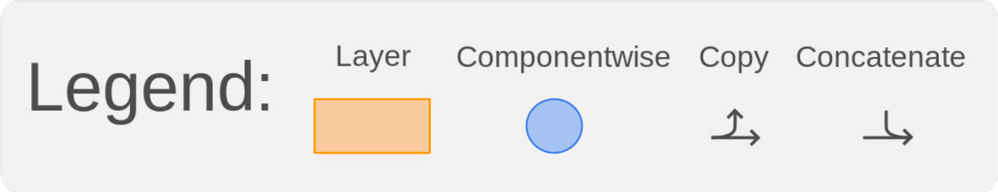
$$c_t = f_t \odot c_{t-1} + \boxed{i_t \odot g_t}$$

$$h_t = o_t \odot \phi(c_t)$$

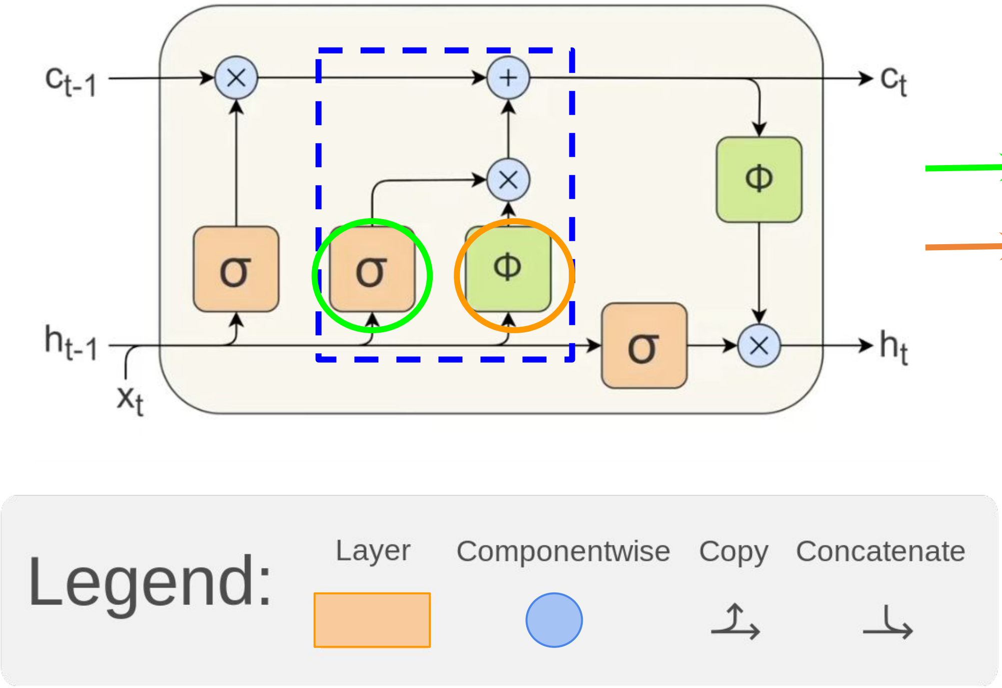
LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$
$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$
$$c_t = f_t \odot c_{t-1} + \boxed{i_t \odot g_t}$$
$$h_t = o_t \odot \phi(c_t)$$



LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

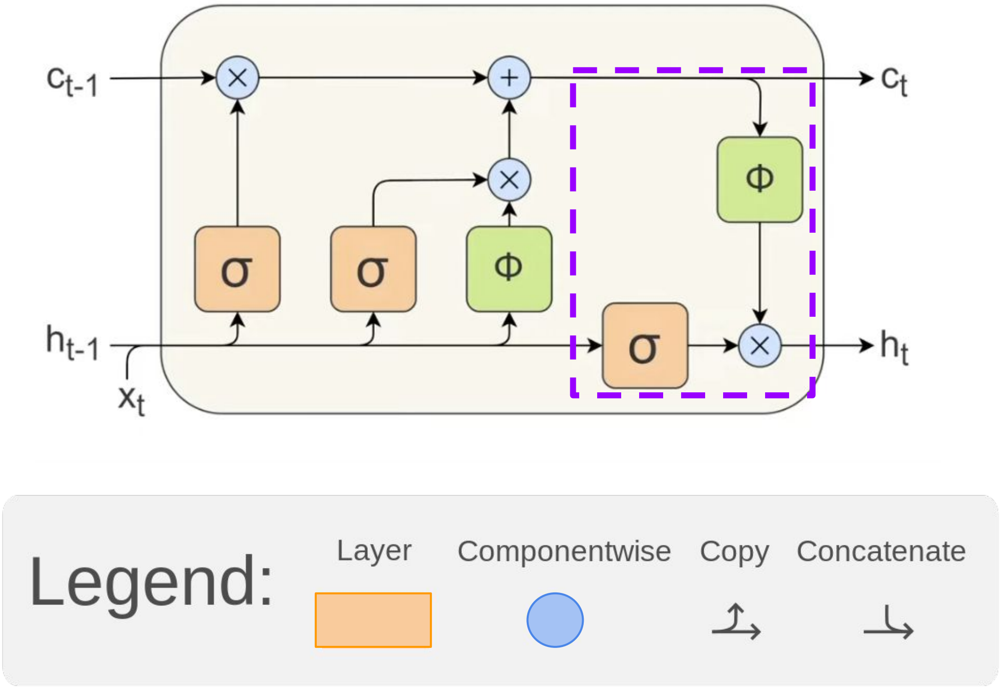
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = f_t \odot c_{t-1} + \boxed{i_t \odot g_t}$$

$$h_t = o_t \odot \phi(c_t)$$

LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

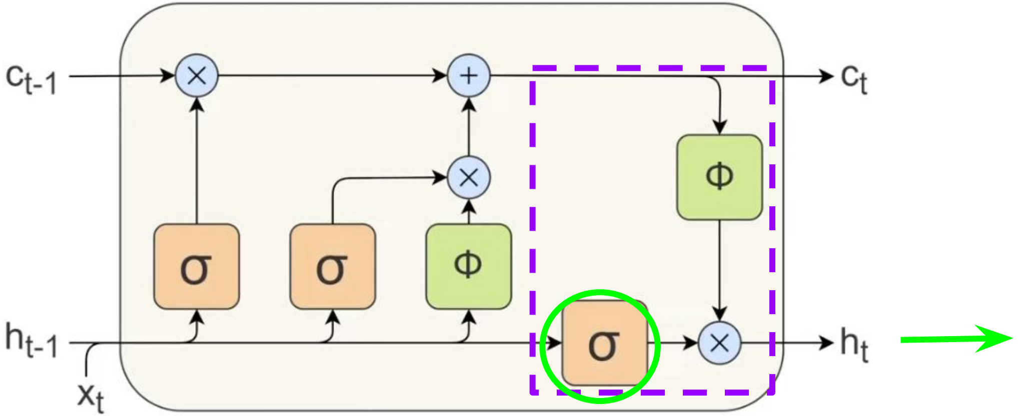
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

$$\boxed{o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$\boxed{h_t = o_t \odot \phi(c_t)}$$

LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

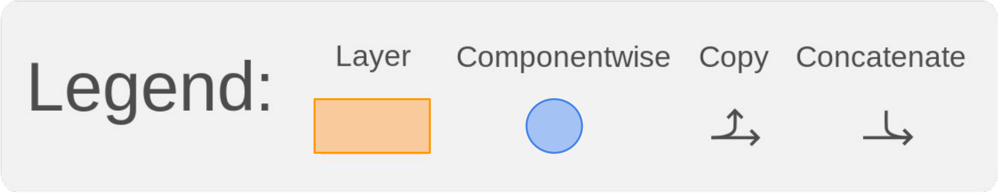
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

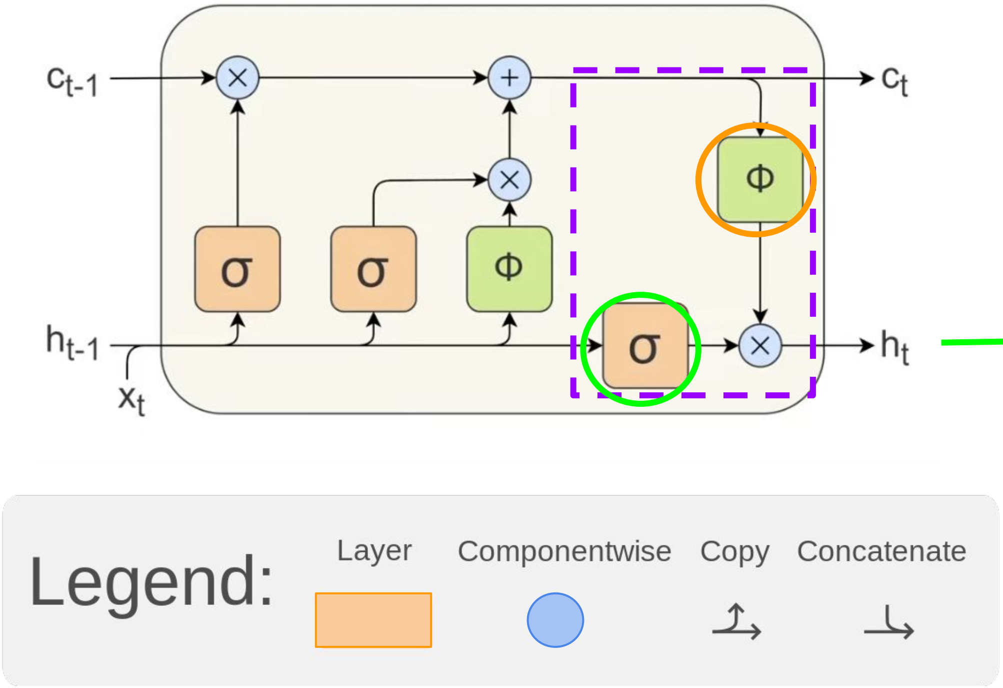
$$o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$h_t = o_t \odot \phi(c_t)$$



LSTM Layer



$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

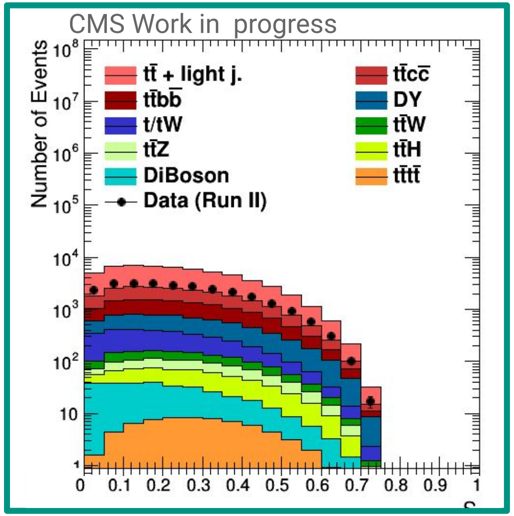
$$g_t = \phi(W_{xg}x_t + W_{hg}h_{t-1} + b_g)$$

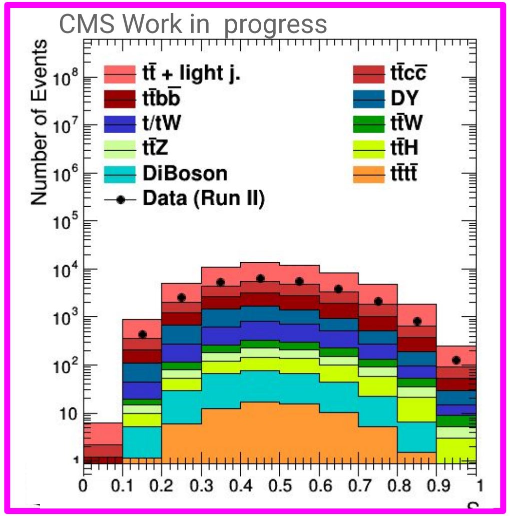
$$\boxed{o_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t$$

$$\boxed{h_t = o_t \odot \phi(c_t)}$$

MVA Evaluation





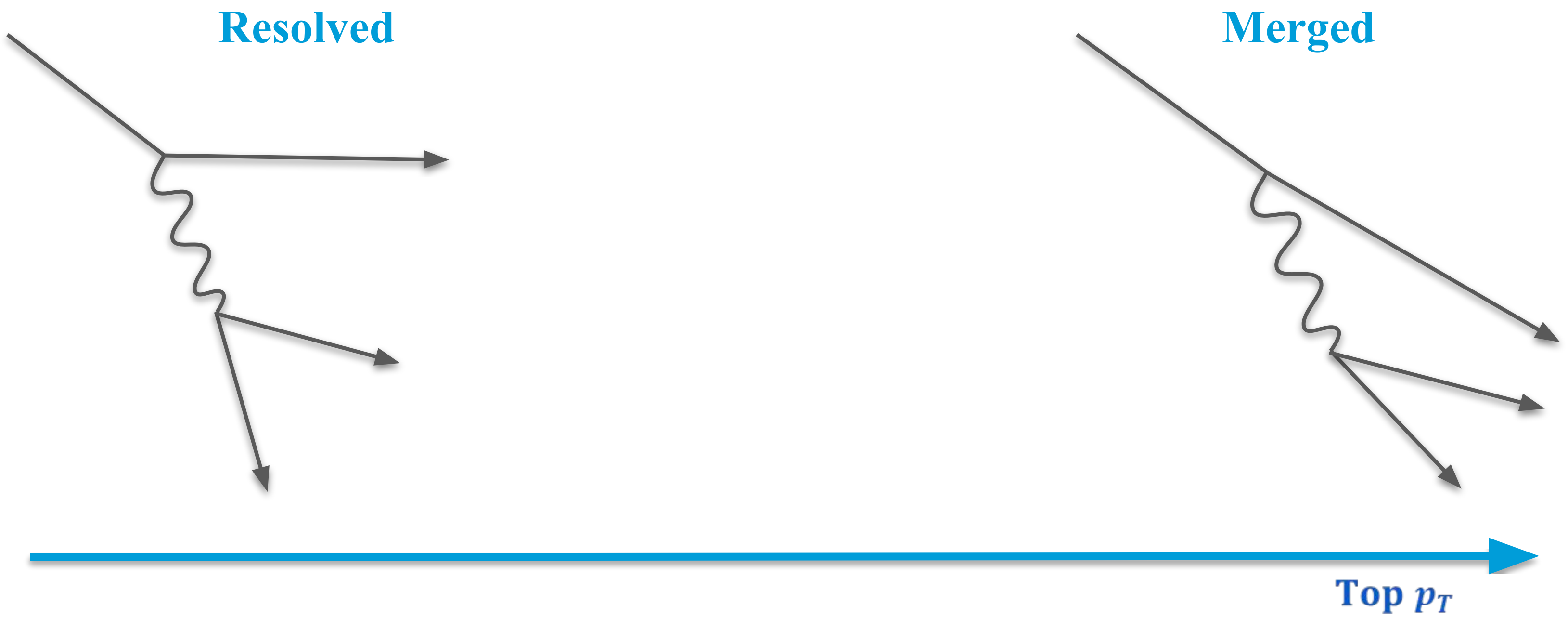
- ➢ Better performance for tttt and ttlight
- ➢ How to improve the score?
  - ○ Long Short Term Memory layer
  - ○ Adding topology variables + LSTM

$$M^{\alpha\beta} = \frac{\Sigma_i p_i^{\alpha} p_i^{\beta}}{\Sigma_i |p_i|^2} \longrightarrow S = \frac{3}{2}(\lambda_2 + \lambda_3)$$

$$C = \frac{H_T}{\Sigma_i E_i}$$

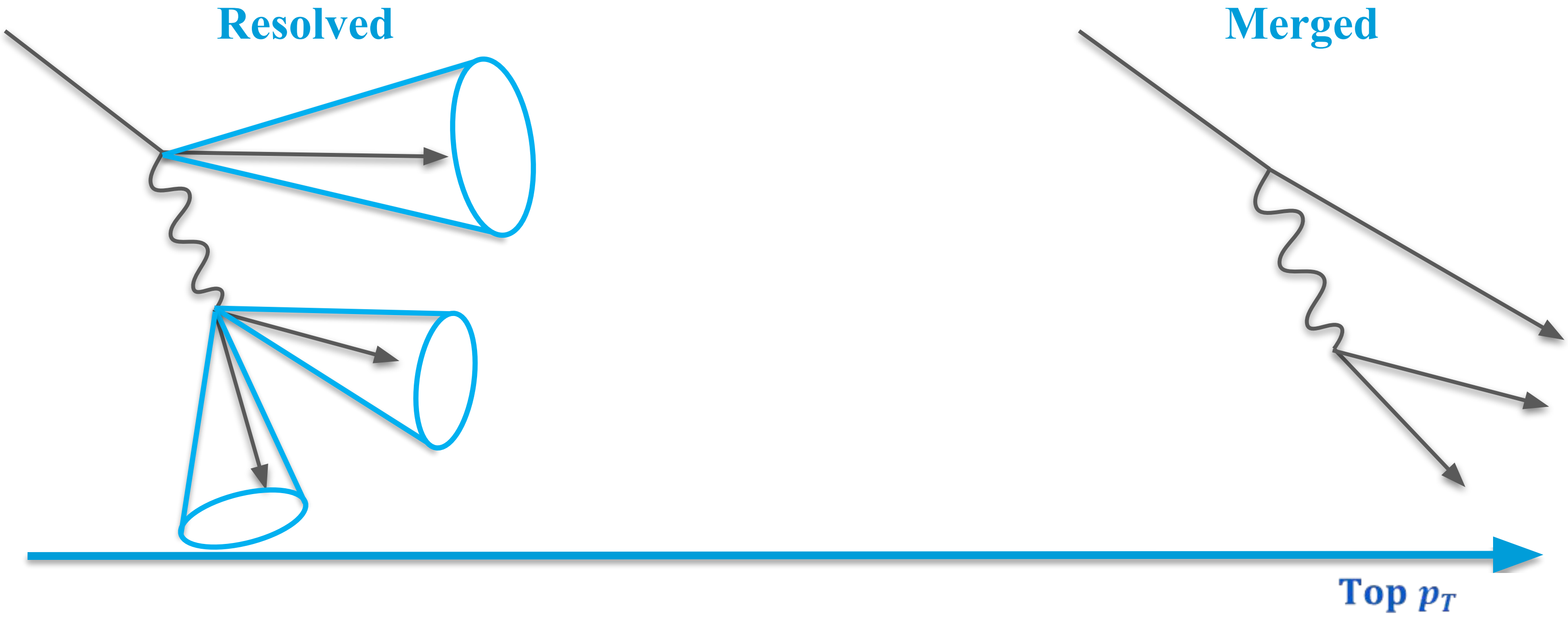
Top quark configurations

2 possible configurations based on the angular superposition between quarks and b-jet



Top quark configurations

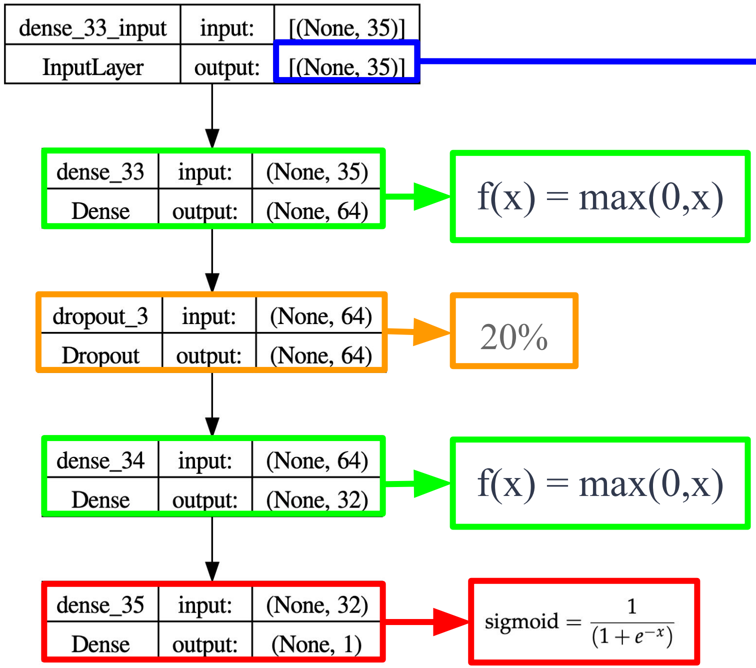
2 possible configurations based on the angular superposition between quarks and b-jet



Top quark configurations

2 possible configurations based on the angular superposition between quarks and b-jet

Object Level NN



Kinematic variables of:

- ➢ Single subjets;
- ➢ W boson system;
- ➢ Top quark system;

Angular separations;

Jets flavour scores;

Soft Drop defined as:

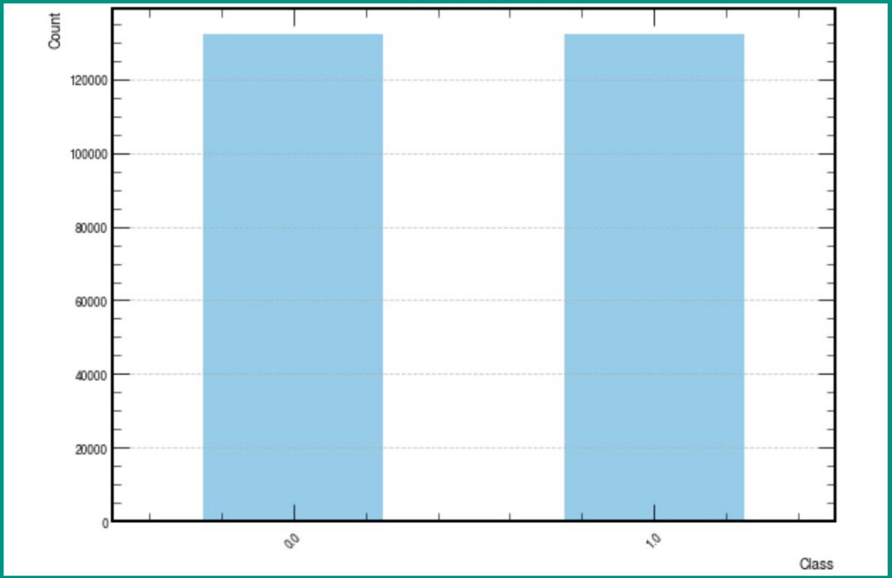
$$\frac{min\left(p_T^{LgdJet}, p_T^{SubldgJet}\right)}{p_T^{LgdJet} + p_T^{SubldgJet}} \cdot \Delta R(p_T^{LgdJet},\ p_T^{SubldgJet})^2$$

Object Level NN

dense_33_input	input:	[(None, 35)]

- 260k entries in 2lOS baseline selection
  - 50% signal (1)
  - 50% background (0)
- Robust Scaler for data preprocessing;

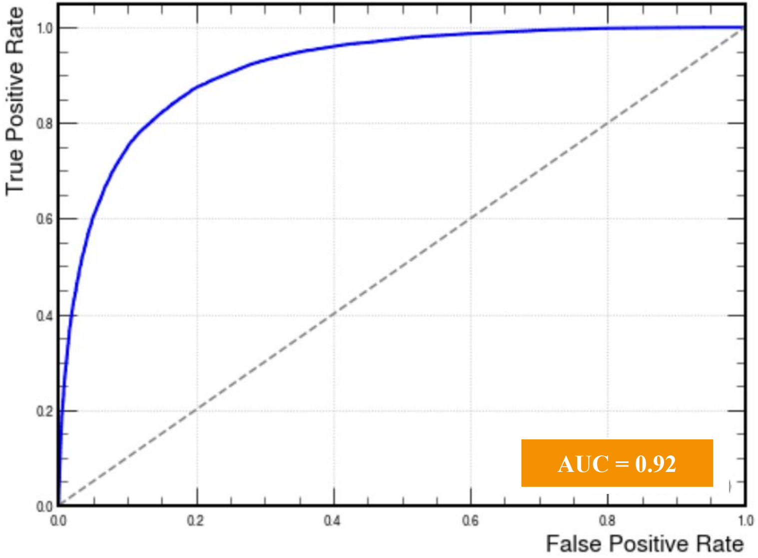
$$X_{\text{scale}} = \frac{x_i - x_{\text{med}}}{x_{75} - x_{25}}$$



dense_35	input:	(None, 32)
Dense	output:	(None, 1)

$$\text{sigmoid} = \frac{1}{(1+e^{-x})}$$

Classifier performances

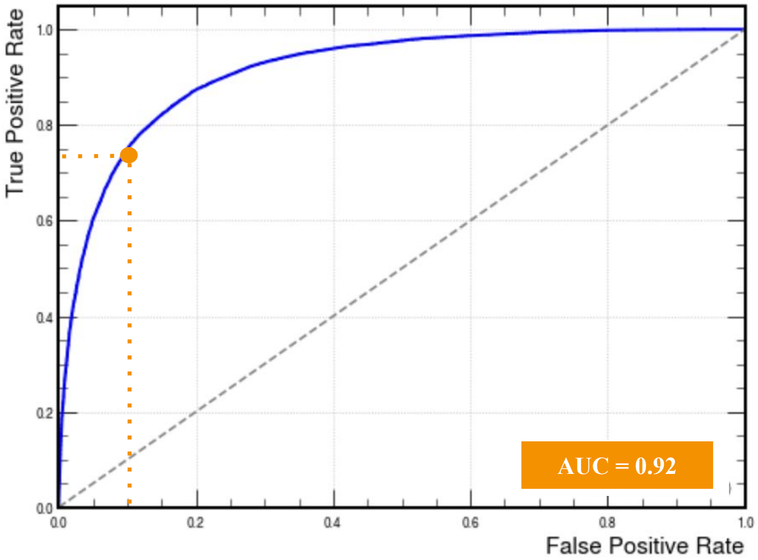


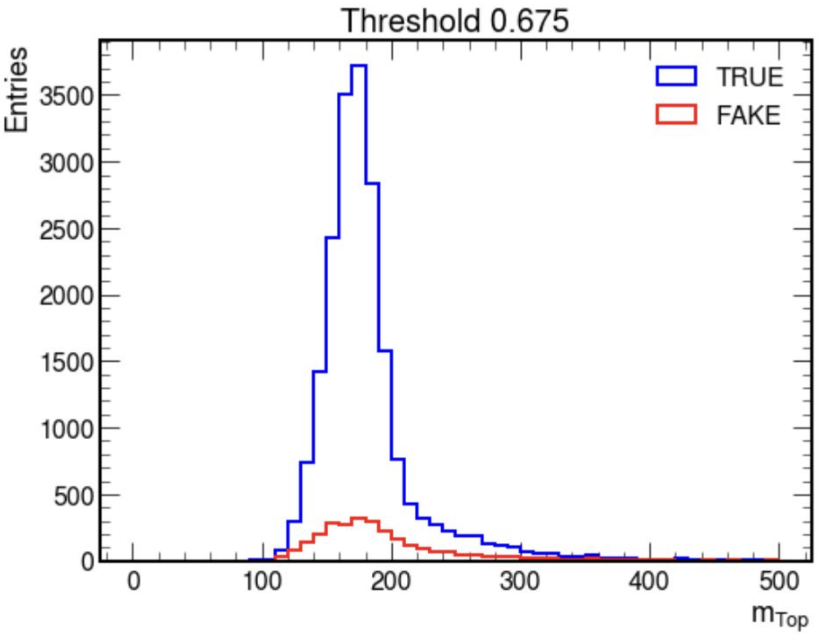
Good performance with a Keras DNN with 2 Dense Layers (64/32), a Dropout Layer (20%), and a ~300k balanced dataset normalized with a robust scaling technique.

Define 4 WPs →

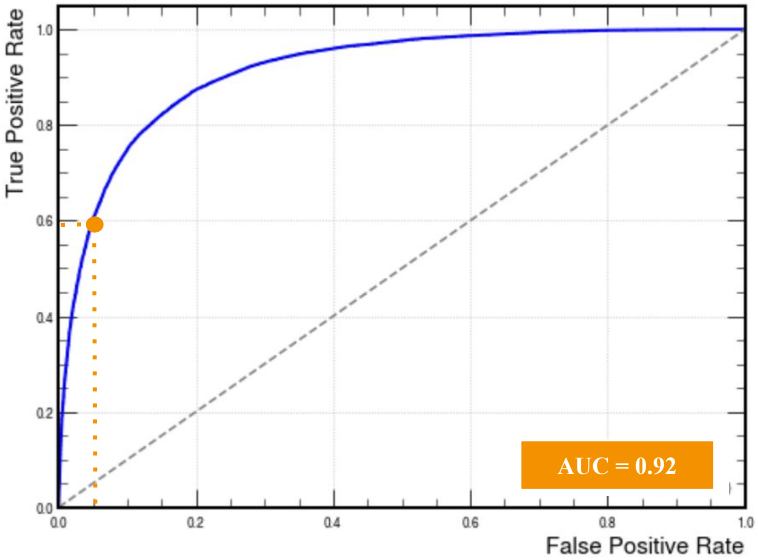
- FPR @ 10% = 0.675 → L
- FPR @ 5% = 0.599 → M
- FPR @ 1% = 0.936 → T
- FPR @ 0.1% = 0.977 → VT

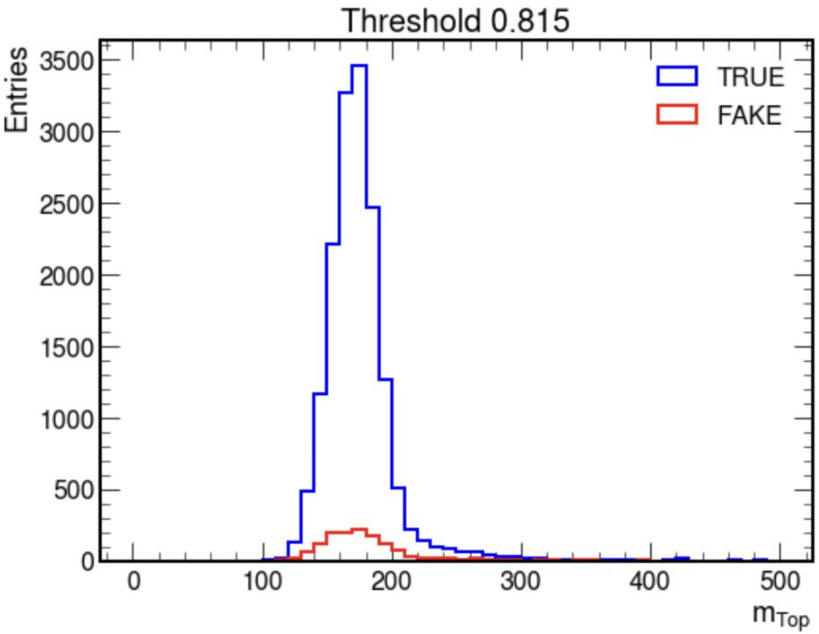
Classifier performances



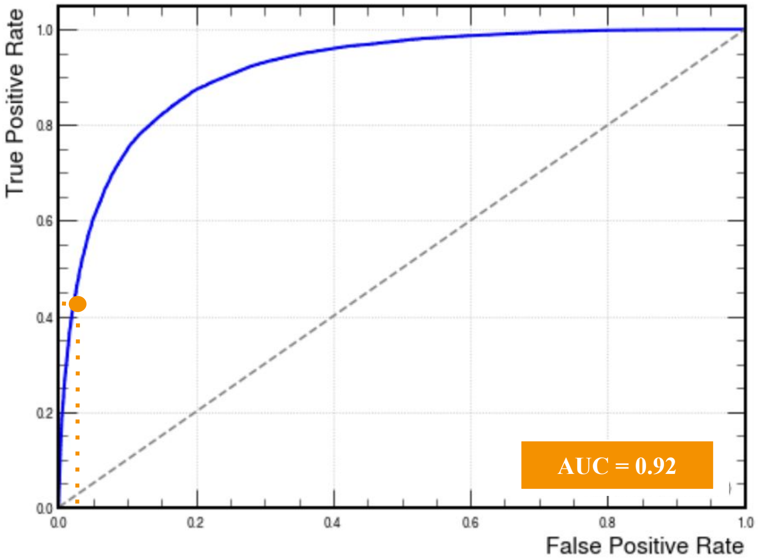


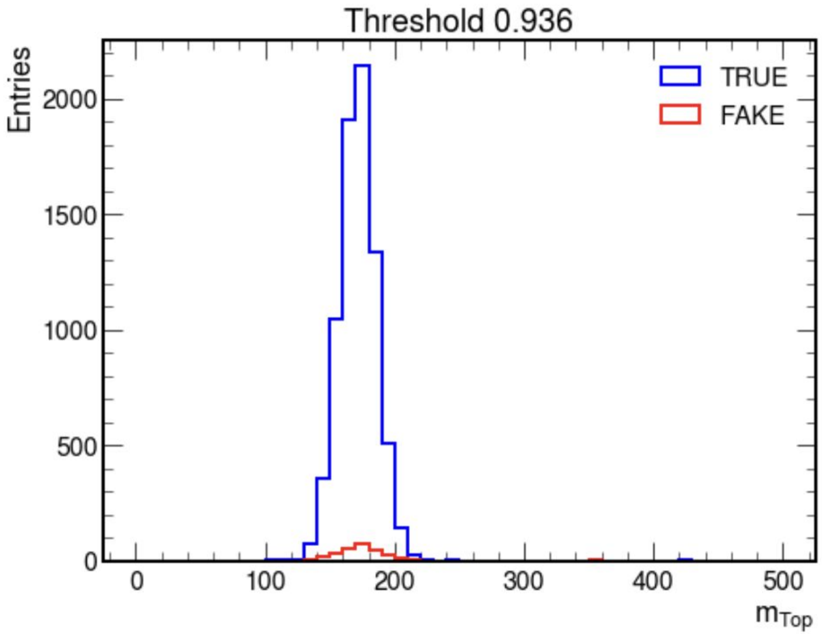
Classifier performances





Classifier performances





Classifier performances