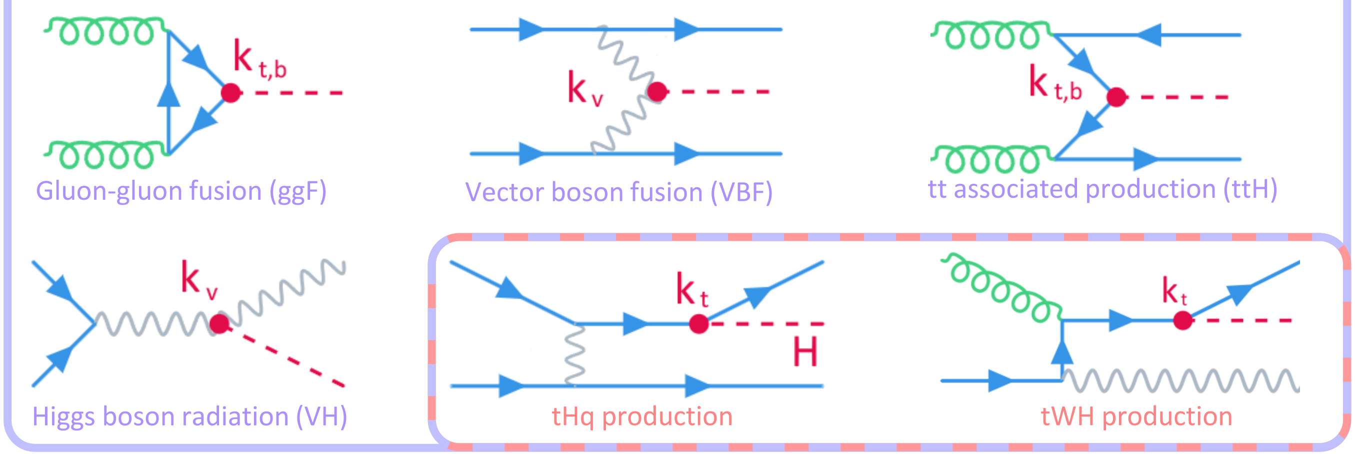
Application of a deep neural networks for Top-Higgs coupling analysis with the ATLAS detector in the H->bb final state

A.Didenko, I.Yeletskikh

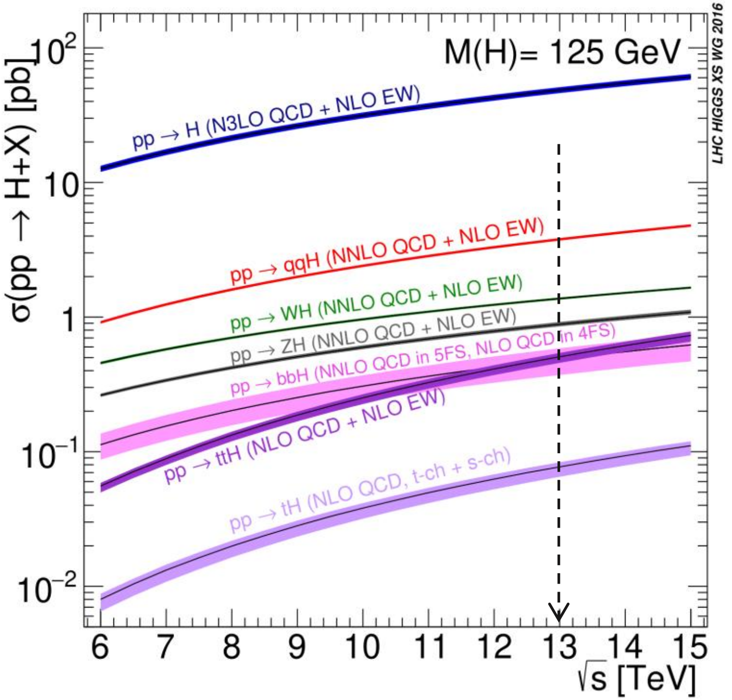
The 29th International Scientific Conference of Young Scientists and Specialists (AYSS-2025)

Production of the Higgs boson in association with a single top quark. Higgs couplings

Feynman diagrams for the **Higgs production** models:



Feynman diagrams for the **Higgs decay** models:



Dependence of the cross section for the production of H_{SM} on the energy of pp collisions.

[1] Higgs Physics, C. Grojean [arXiv:1708.00794]

$\sigma(tHq)^{SM} \approx 74.3$ fb
$\sigma(tWH)^{SM} \approx 15.2$ fb

Signature of tH with decay $H \rightarrow b\bar{b}$:

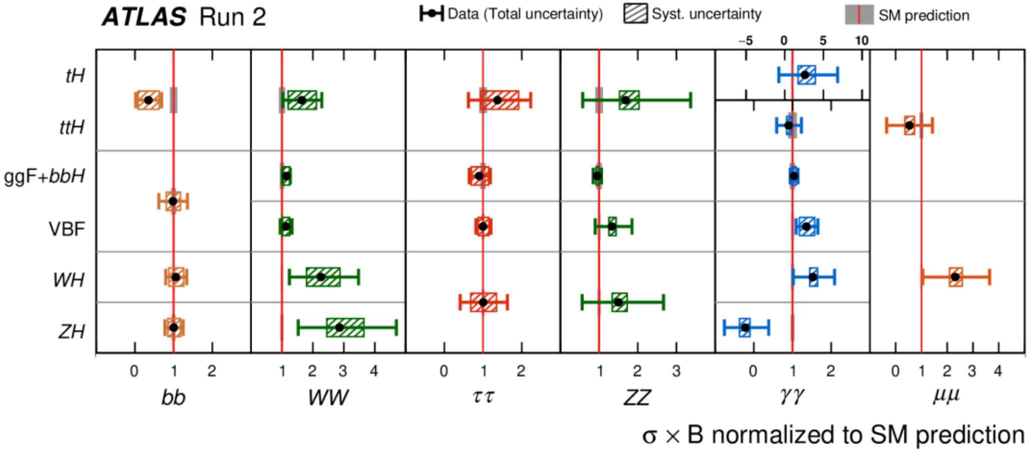
(≥3 b-jets) + (1 light jet) + (1 tight lepton) + (missing transverse momentum)

Higgs boson interaction couplings and hierarchy of particle masses

Effective coupling modifier parameter k and global signal strength μ_i^f

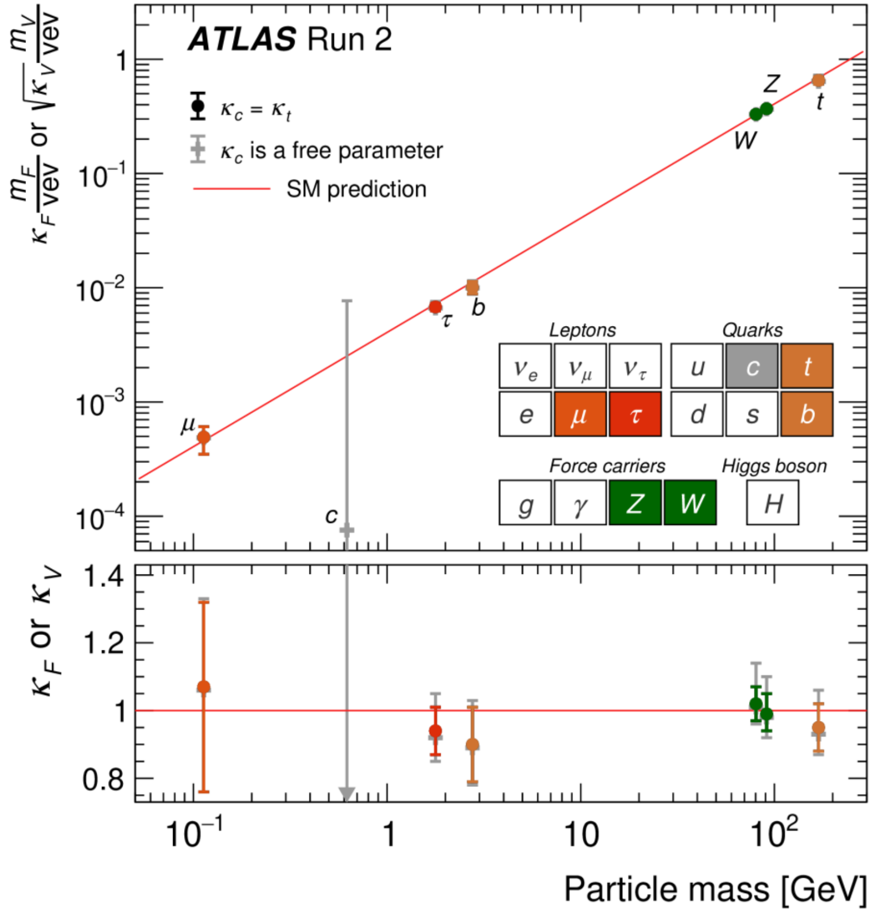
$$\mu_i \equiv \frac{\sigma_i}{(\sigma_i)_{\rm SM}}, \quad \mu^f \equiv \frac{{\rm BR}^f}{({\rm BR}^f)_{\rm SM}}, \quad \mu_i^f \equiv \mu_i \mu^f, \quad \kappa \equiv \frac{Y}{Y^{\rm SM}},$$

$$\mu_{t\bar{t}H} \approx \kappa_t^2 \cdot \frac{{\rm BR}(H \to X)}{{\rm BR}(H \to X)_{SM}}.$$



Ratio of observed rate to predicted SM event rate for different combinations of Higgs boson productionand decay processes.

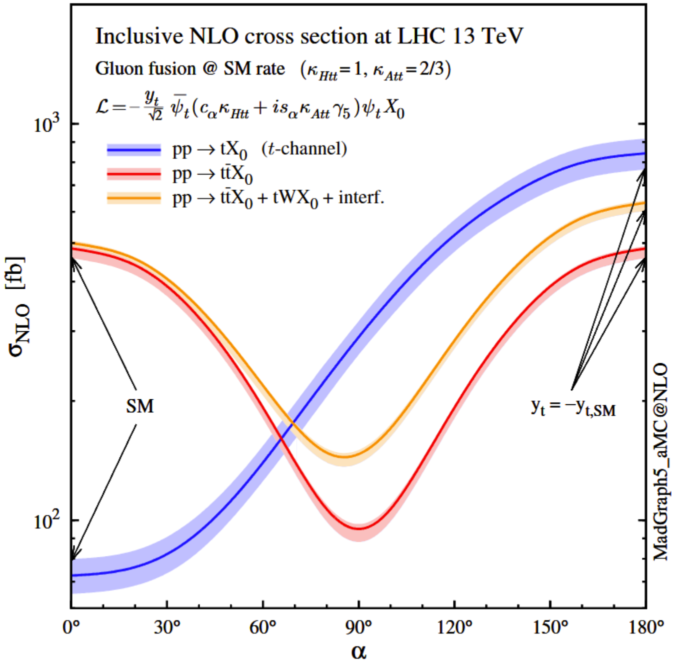
[2] A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery, ATLAS collaboration [arXiv:2207.00092v2]



[2] A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery, ATLAS collaboration [arXiv:2207.00092v2]

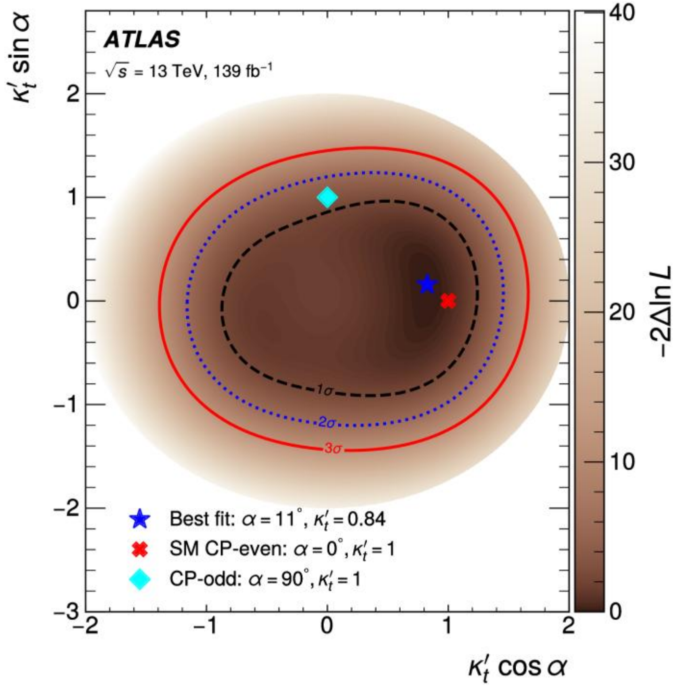
Effective Top-Higgs Yukawa coupling (κ_t) and Higgs CP

Effective Lagrangian describing top Yukawa coupling can be paramaterized as: $\mathcal{L} = -\frac{m_t}{v}\left\{\bar{\psi}_t \kappa_t \left[\cos(\alpha) + i\sin(\alpha)\gamma_5\right]\psi_t\right\} H$



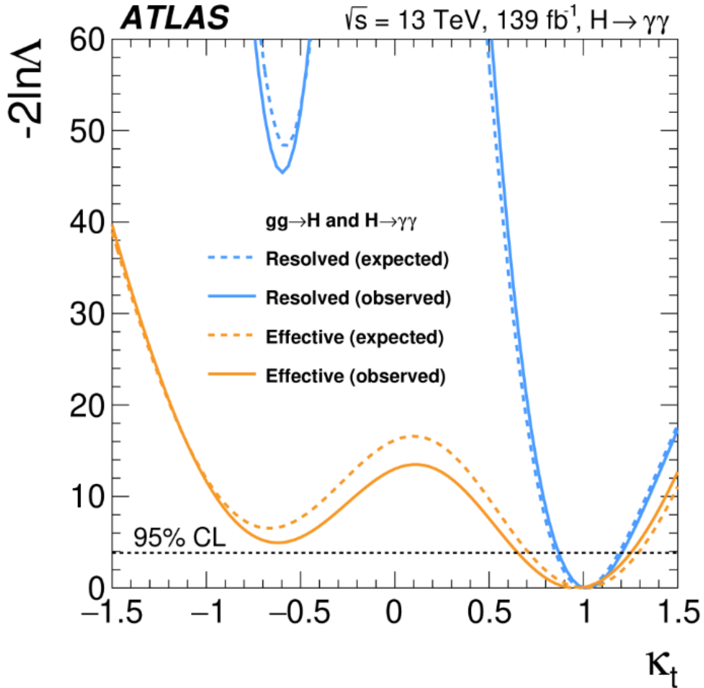
NLO cross sections at the 13-TeV LHC as a function of the CP-mixing angle α, where k(Htt) and κ(Att) are set to reproduce the SM gluon-fusion cross-section for every value of α.

[3] tWH associated production at the LHC, F.Demartin, B.Maier, F.Maltoni, K.Mawatari, M.Zaro [arXiv:1708.00794]



CP-mixing angle in the ttH(H → bb) channel using the fullRun-2 dataset.

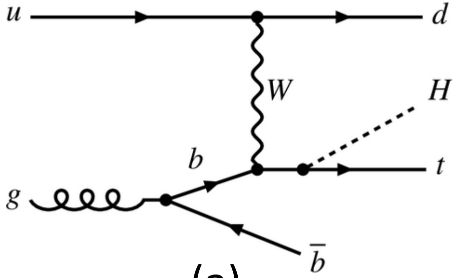
[4] Probing the CP nature of the top-Higgs Yukawa coupling in ttH and tH events with H→bb decays using the ATLAS detector at the LHC, ATLAS Collaboration [arXiv:1708.00794]



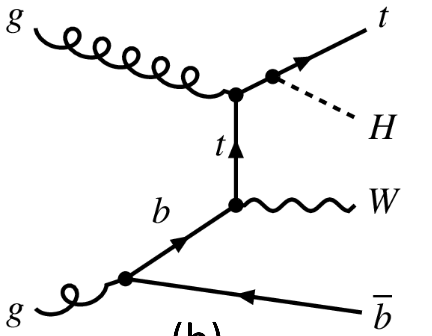
In the latter case, sensitivity to the sign of κt is provided by the tH process, and to a lesser degree by the $gg \rightarrow ZH$ process.

[5] Measurement of the properties of Higgs boson production at √13 TeV in the H→γγ channel using 139 fb-1 of pp collision data with the ATLAS experiment, ATLAS Collaboration [arxiv:2207.00348]

Kinematics of signal (**tHbq**) and background processes (Z+jets, W+jets, tZq, tWZ, rare t-quark and others)



(a)



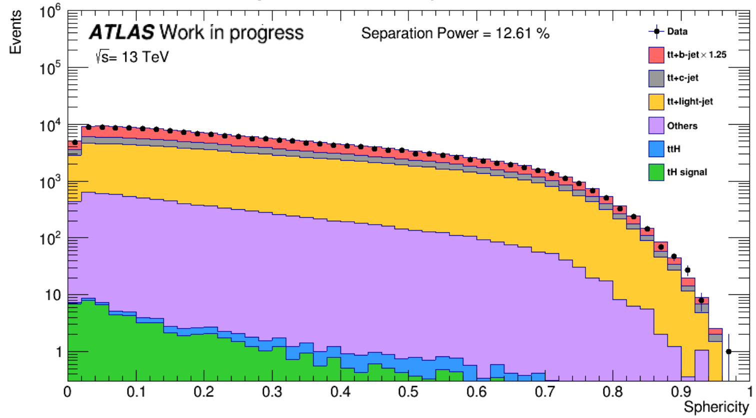
(b)

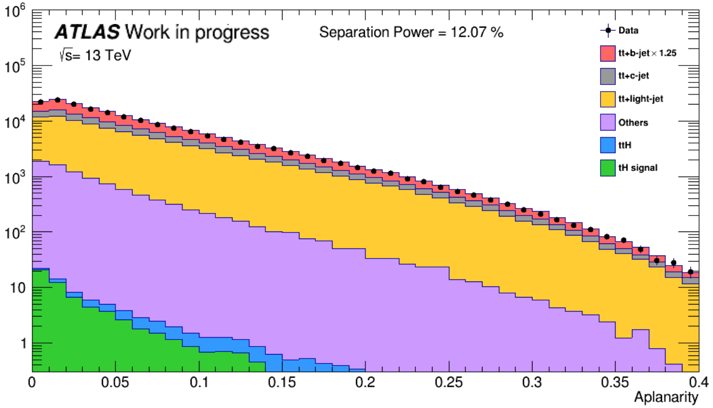
Dominant Feynman diagram for **tHq (a)** and tWH (b).

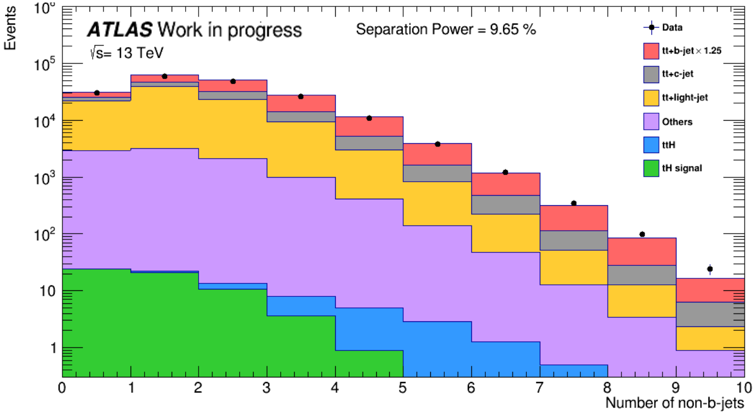
[6] Search for the production of a Higgs boson inassociation with a single top quark in pp collisionsat √*s* = 13 TeV with the ATLAS detector, ATLAS Collaboration [arXiv:2508.14695]

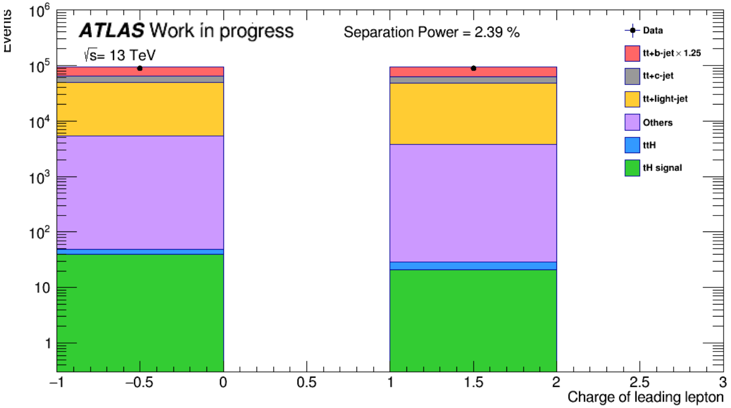
The signal and background processes differ in a large number of kinematic variables, but for each of them the differences are small. The background separation problem can be solved using machine learning.

That was Presented for Master Diploma









Kinematics of signal (**tHbq**) and background processes (Z+jets, W+jets, tZq, tWZ, rare t-quark and others)

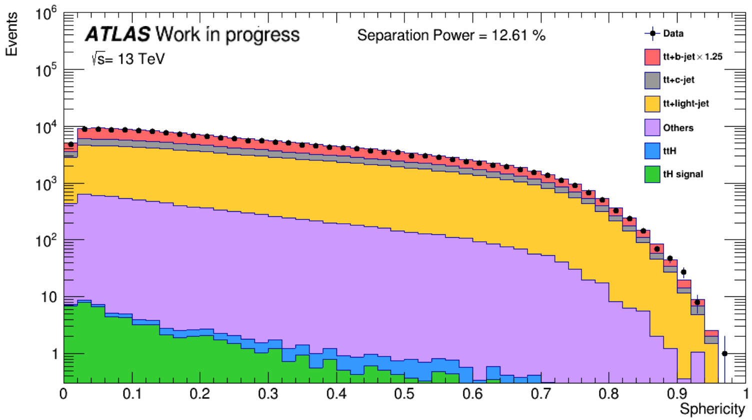
After pre-selection cuts:

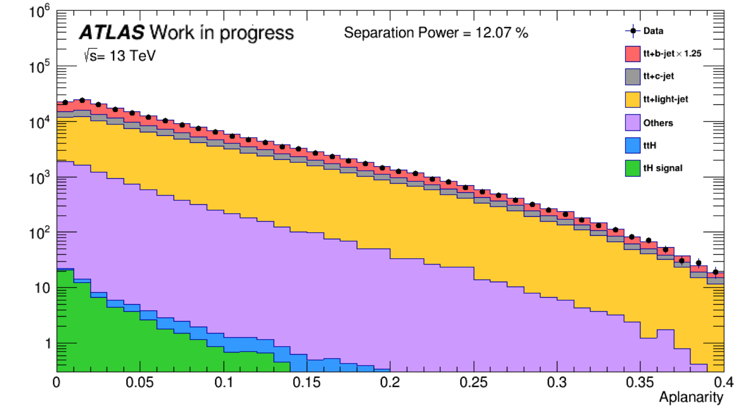
- Exactly 1 tight trigger-matched lepton with p_T above 27 GeV;
- No reconstructed hadronic τ;
- At least 3 b-tagged jets;
- $E^{miss}_T > 25$ GeV;
- Veto for events with ≥ 5 jets and ≥ 4 b-tagged jets;

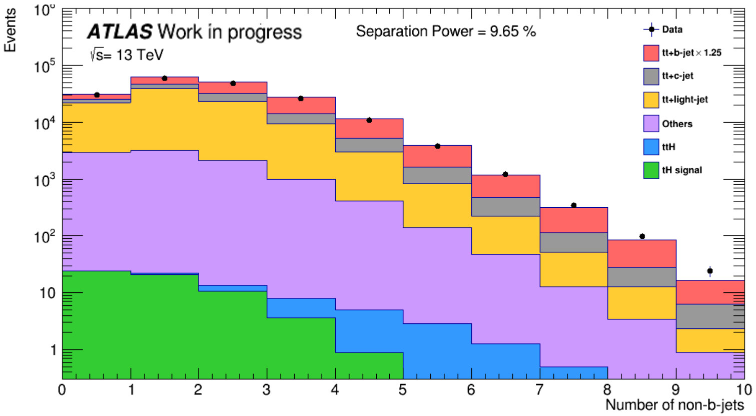
Data	**188296**
tt+b-jet × 1.25	**62229.3**
tt+c-jet	**27901.9**
tt+light-jet	**88533.7**
non-top bkg	**6.3**
W,Z+jets	**3873.7**
tW	**2445.3**
single t	**2838.9**
tWZ	**2.1**
tZq	**135.5**
ttZ	**42.8**
ttW	**217.0**
ttH	**17.0**
tH signal	**52.5**

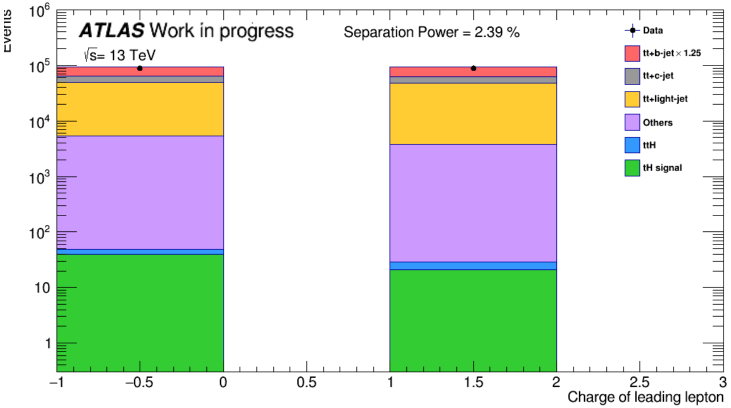
Using ***v34_minintuples_v1*** *Which was Presented for Master Diploma*

The most sensitive to signal-background separation 24 kinematic variables were selected. Variables with correlation more then 84% were excluded. *That was Presented for Master Diploma*







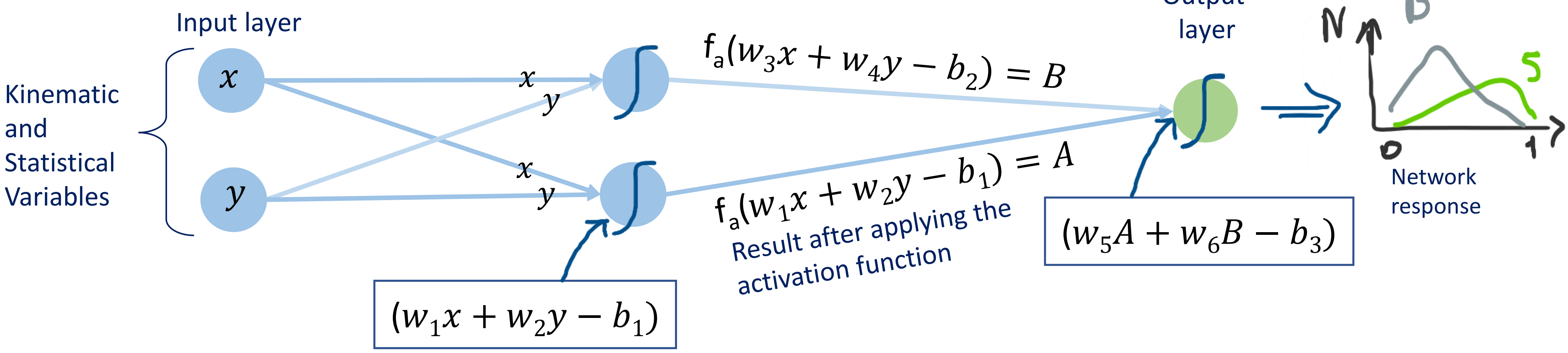


Application of a neural network for the classification of processes. Network structure

A **neural network** is a function with a large number of variable parameters. Optimal values of these parameters provide the best separation of signal and background.

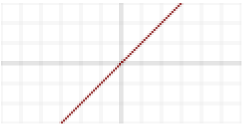
Variable (free) parameters: weights ($w_1 - w_6$) and shifts ($b_1 - b_3$).

A simple neural network:



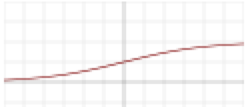
some of activation functions:

Identity



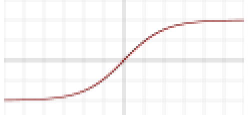
x

sigmoid



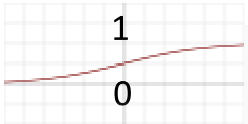
$\sigma(x) \doteq \frac{1}{1+e^{-x}}$

Hyperbolic tangent



$\tanh(x) \doteq \frac{e^x - e^{-x}}{e^x + e^{-x}}$

softmax



$\frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}$

J

Results. DNN for classification signal events

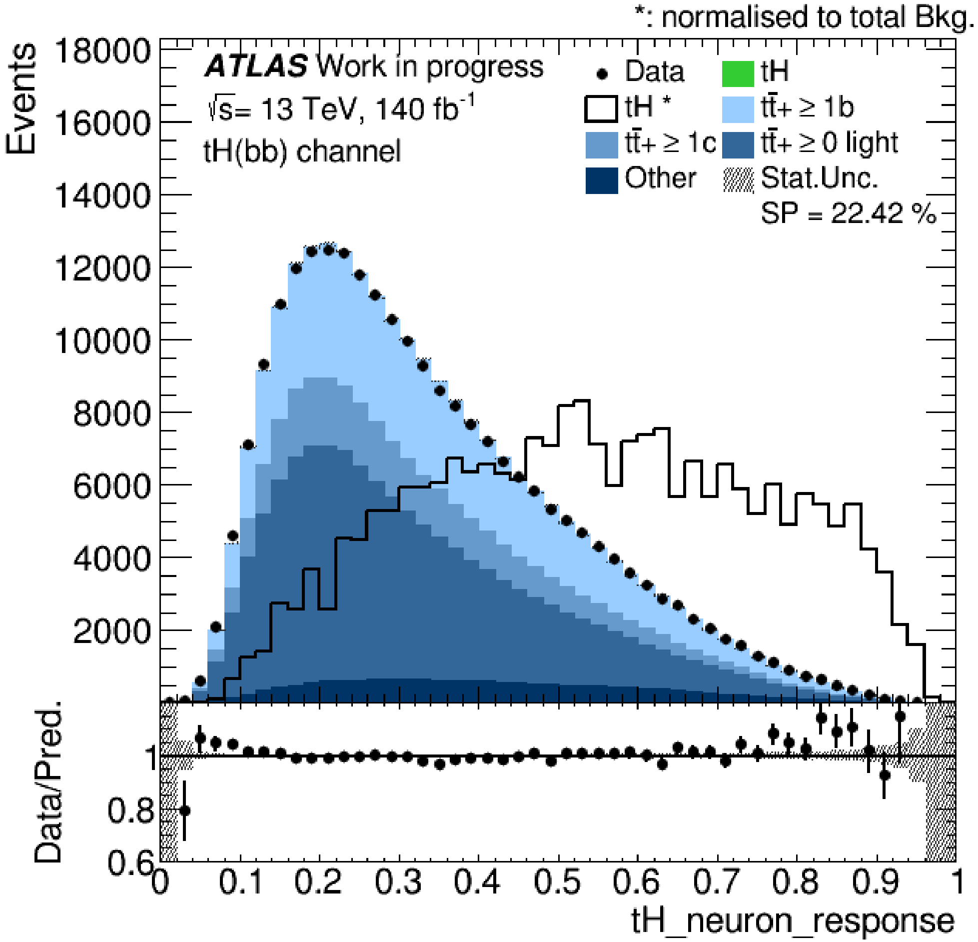
DNN method

The network receives 24 variables as input and outputs 1 variable (network response), which accumulates the differences between the signal and the background contained in all 24 input variables. This variable gives an increase in the significance of the signal by 1.4 times (expected from MC).

Only tHbq, ttb, ttc, ttL и ttH were used for DNN learning.

- Fully-connected feedforward neural network
- Adam optimizer
- Number of Epochs = 20
- Batch Size = 100

MLP Architectures :
MLP **tHbq**: 24 vars -> 150(tanh) -> 150(tanh) -> 2(softmax)
MLP **ttb**: 24 vars -> 240(tanh) -> 240(tanh) -> 2(softmax)
MLP **ttc**: 28 vars -> 150(tanh) -> 150(tanh) -> 2(softmax)
MLP **ttL**: 23 vars -> 230(tanh) -> 23(tanh) -> 2(softmax)

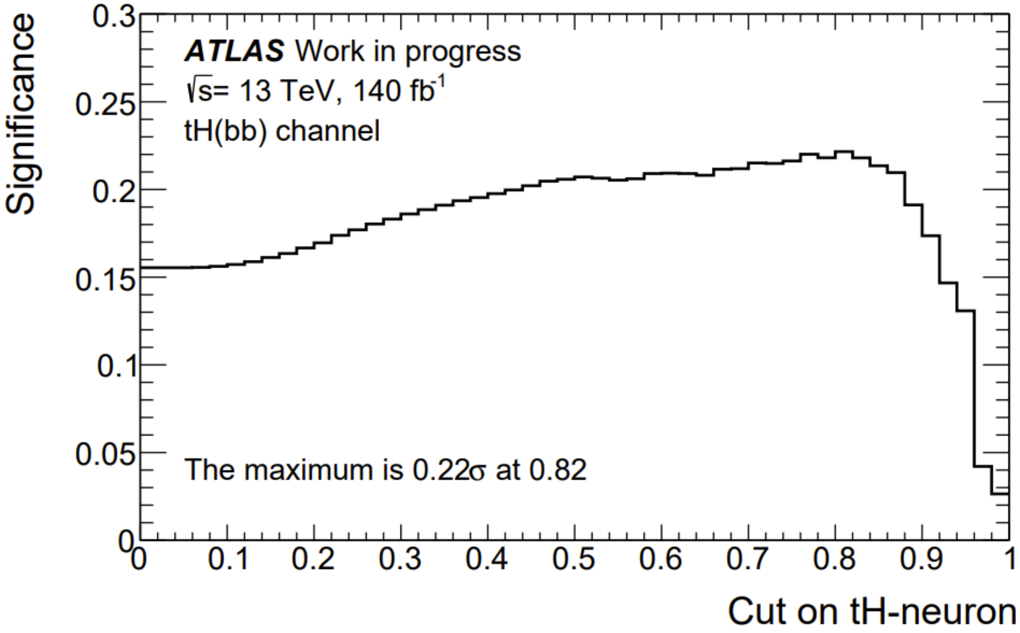


Keras for root tmva

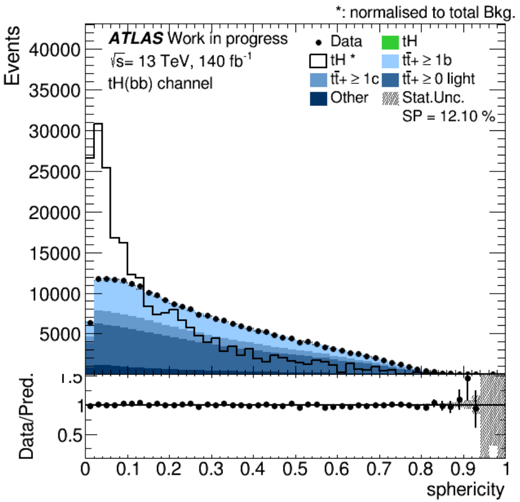
Results. DNN for classification signal events

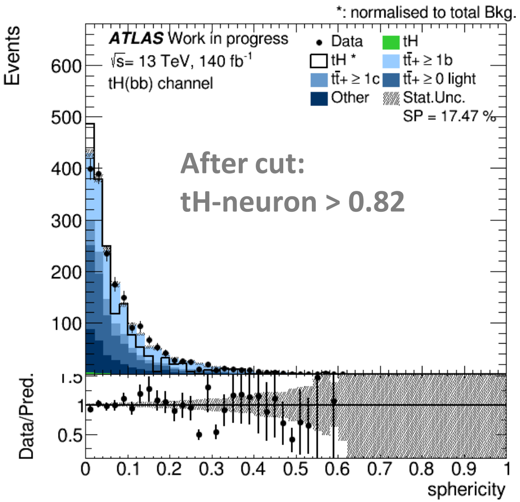
DNN method

The applying of DNN allows to increase the significance of the SM signal by **1.4** times.
The maximum significance of the SM signal is achieved with cut on **tH-neuron > 0.82**.



Process	Pre-selection yield Before cut on tH-neuron	After cut on tH-neuron
tH	74 (0.03 %)	9 (0.53 %)
tt + ≥ 1b *1.24	74447	688
tt + ≥ 1c	33516	194
tt + light	106055	509
others	16231	261
total	230326	1661
data	236832	1867





Results. DNN response vs BDT in Signal and Control Regions

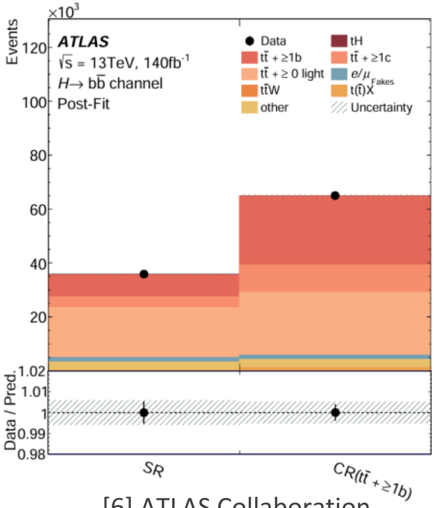
BDT method

Defining the rectangular regions:

SR:
$N^{70}_{not-b} < 2$
$N^{jet}_{forward} \geq 1$

CR(tt + ≥1b):
$N^{70}_{not-b} \geq 2$
$N^{jet}_{forward} = 0$

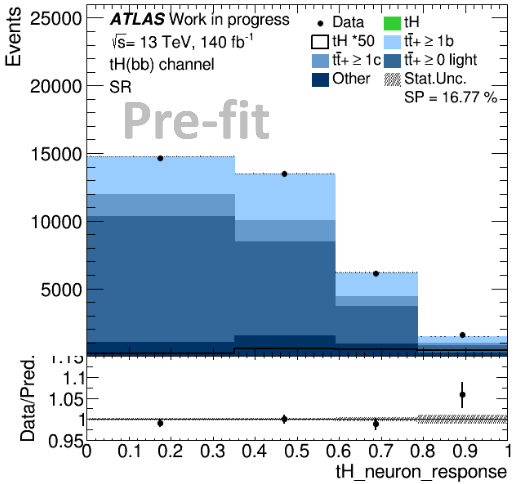
The discriminant variable in the SR is the BDTG response for tHq signal-like events, P(tHq); in the CR(tt+ ≥1b), the discriminant is the BDTG response for tt+ ≥1b-like events, P(tt+ ≥1b).

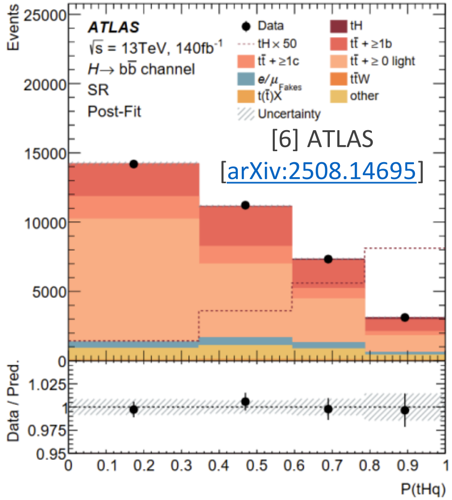


[6] ATLAS Collaboration [arXiv:2508.14695]

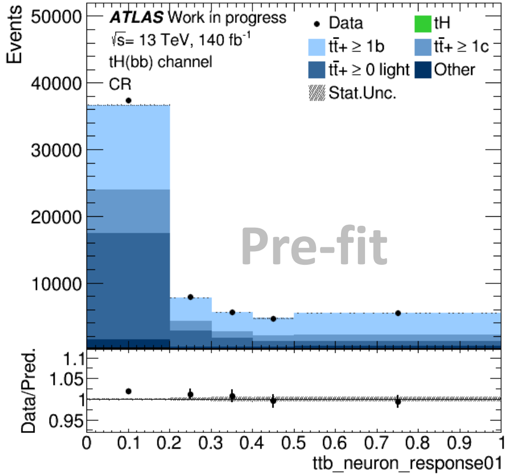
- **ML methods** are used to separate different background components to better constrain them, as well as to define the analysisregions.

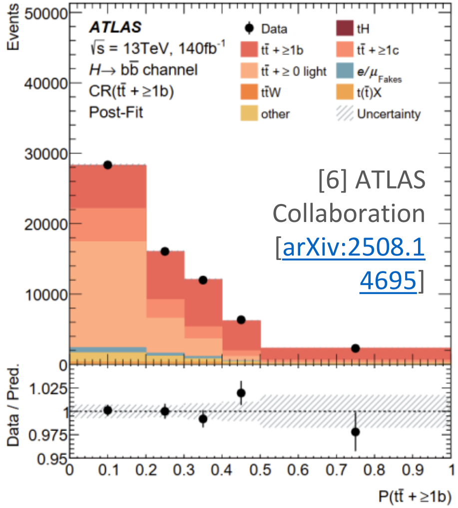
The normalisation factor of the tt+ ≥ 1b background found in the combined fit to data is $k = 1.24^{+0.17}_{-0.15}$.





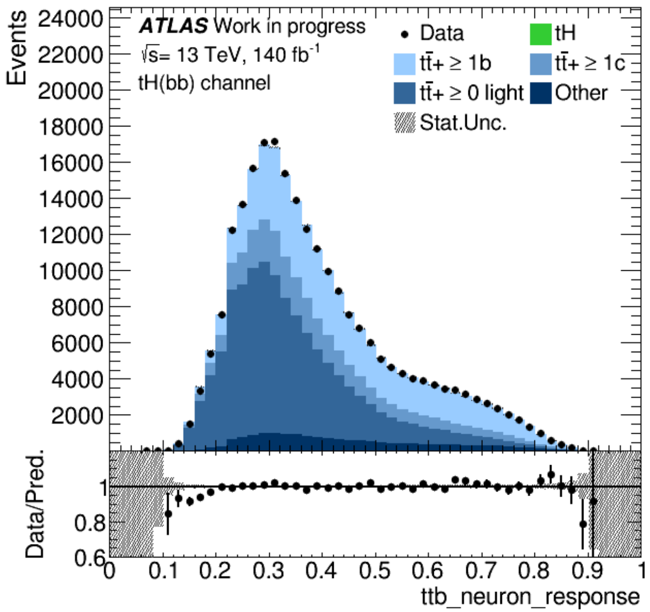
[6] ATLAS [arXiv:2508.14695]

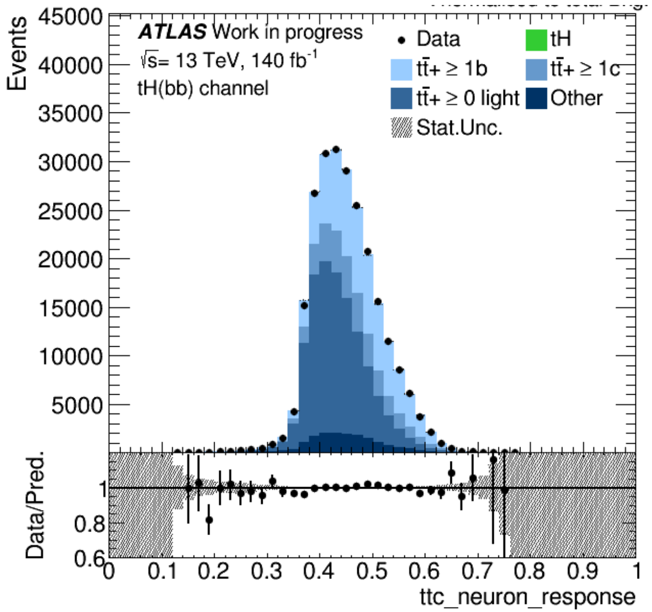


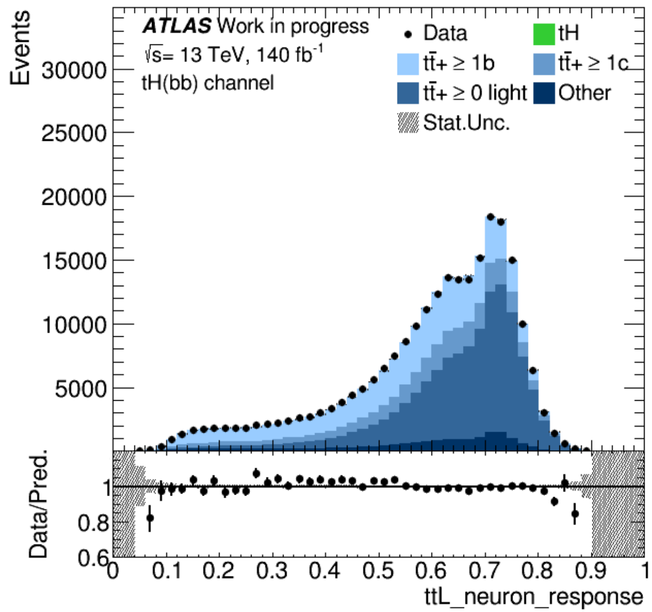


[6] ATLAS Collaboration [arXiv:2508.14695]

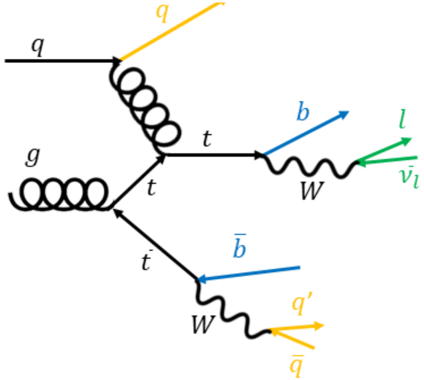
Results. DNN for classification of background events and control regions definitions







ML methods are used to separate different background components to better constrain them, as well as to define the analysis regions. The result of the fit in the control regions determined by the **multiclassifier**: 1.25 for ttb, 1.1 for ttc, 0.98 ttL.
Bad separation for ttc is due to its kinematic properties, not to network learning.



$t \rightarrow b + \ell + \nu$

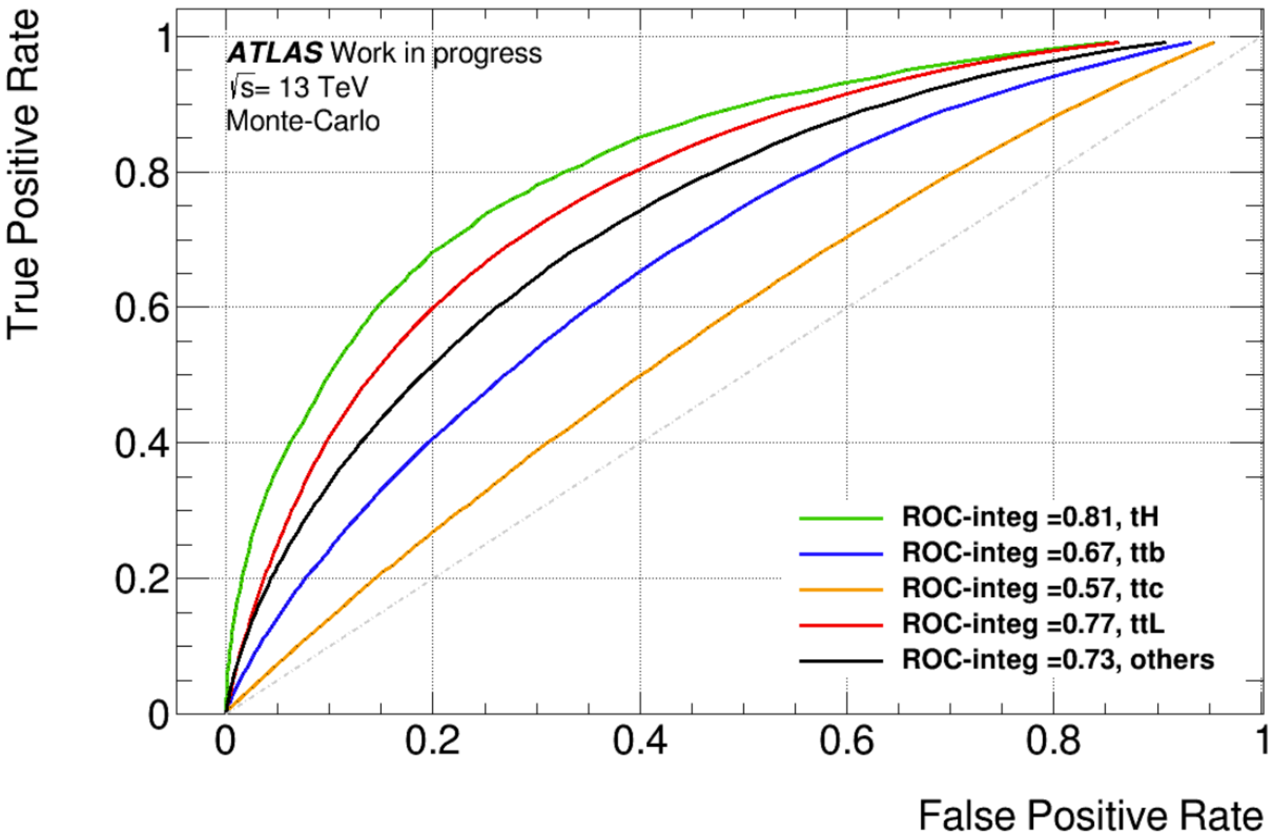
$t <$ $b + d + u$ (50%)
$b +$ **c** $+ s$ (50%)

Efficiency plot. Network vs BDT

The efficiency curve is the dependence of the cut signal on the number of background events captured. The largest integral value shows the better the classification performance.

DNN MLP	BDT
0.81	0.85

Using ***v34_minintuples_v1***
Which was Presented for Master Diploma



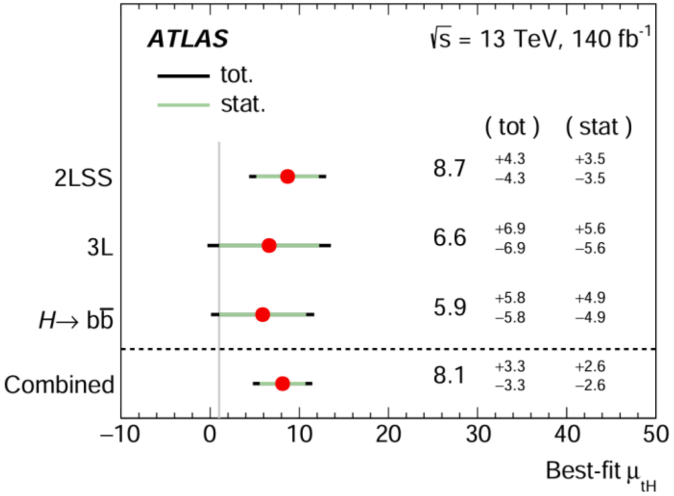
Higgs boson in association with a single top quark. Top-Higgs coupling. Run2 ATLAS and CMS

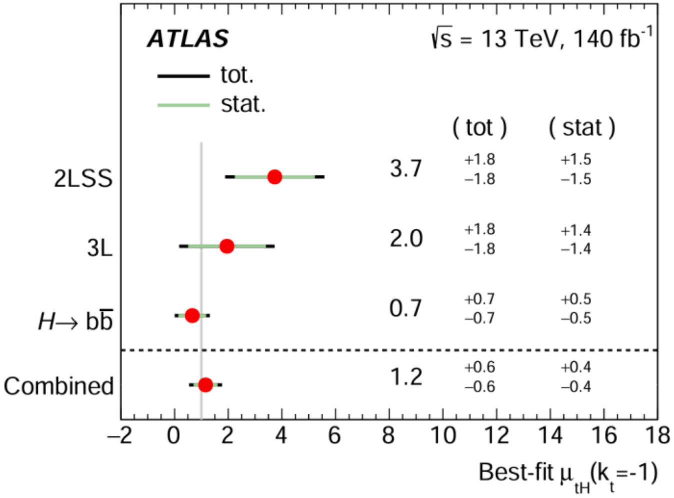
[6] Search for the production of a Higgs boson inassociation with a single top quark in pp collisionsat $\sqrt{s}$ = 13 TeV with the ATLAS detector, ATLAS Collaboration [arXiv:2508.14695]

ATLAS: BDT(tHq) is trained on the SM signal simulation, andits highly ranked input features are highly sensitive to the sign of the top-Yukawa coupling. The excess in data favours the ITC hypothesis, but both hypotheses are compatible with the data. The best-fit value of the tHq and tWH signal strength (left for SM & right for ITC):

μ_{tH} = 8.1 ± 2.6 (stat.) ± 2.0 (syst.) corresponding to σ(tH) = 720 ± 270 fb.

μ_{tH} (κt = −1) = 1.2 ± 0.4 (stat.) ± 0.5 (syst.)



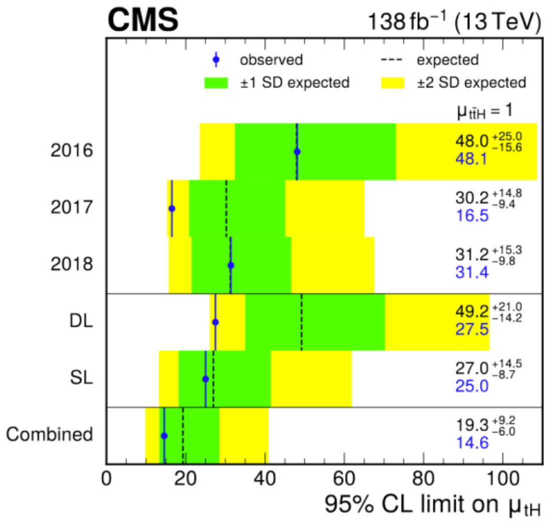


The $H \rightarrow bb$ channel is most sensitive to the H →bb decay, while the 2ℓSS and 3ℓ channels mostly probe H → WW∗. In the 2ℓSS and 3ℓ channels, subleading contributions arise from H → $\tau\tau$ and from H → ZZ∗ decays.

[7] Measurement of the ttH and tH production rates in the H → bb decay channel using proton-proton collision data at √s = 13 TeV, CMS Collaboration [arXiv:2407.10896]

CMS: ANN output values are used to compute the final discriminant observables.

An observed (expected) upper limit on the tH production cross section relative to the SM expectation of 14.6 (19.3) at 95% confidence level (CL) is derive.



Conclusions and Results of Deep Neural Networks.

Deep Neural Networks, were developed and tested **as cross-checks** of the main analysis method based on Boosted Decision Trees.

1. Found:
   - optimal set of variables
   - optimal network structure and settings
   - corrections for background processes
   - area of maximum signal-to-background ratio: **DNN output for tH-neuron>0.82**
2. DNN produces a separation comparable to the basic analysis method BDT.
3. Application of a NN increases the significance of the signal by **1.4**.

- Further optimization of network hyperparameters can increase the efficiency of signal-background separation.

BackUp

References

[1] Higgs Physics, C. Grojean [arXiv:1708.00794]

[2] A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery, ATLAS collaboration [arXiv:2207.00092v2]

[3] tWH associated production at the LHC, F.Demartin, B.Maier, F.Maltoni, K.Mawatari, M.Zaro [arXiv:1708.00794]

[4] Probing the CP nature of the top-Higgs Yukawa coupling in ttH and tH events with H→bb decays using the ATLAS detector at the LHC, ATLAS Collaboration [arXiv:1708.00794]

[5] Measurement of the properties of Higgs boson production at √13 TeV in the H→γγ channel using 139 fb-1 of pp collision data with the ATLAS experiment, ATLAS Collaboration [arxiv:2207.00348]

[6] Search for the production of a Higgs boson inassociation with a single top quark in pp collisionsat $\sqrt{s}$ = 13 TeV with the ATLAS detector, ATLAS Collaboration [arXiv:2508.14695]

[7] Measurement of the ttH and tH production rates in the H → bb decay channel using proton-proton collision data at √s = 13 TeV, CMS Collaboration [arXiv:2407.10896]

[8] Measurement of the associated production of atop-antitop-quark pair and a Higgs boson decayinginto a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV usingthe ATLAS detector at the LHC , ATLAS Collaboration [arXiv:2407.10904]

[9] Development of systematic uncertainty-aware neural network trainings for binned-likelihood analyses at the LHC, CMS Collaboration [arXiv:2502.13047]

Higgs
Di Higgs

Geneva, Switzerland, September 8th, 2025

To whom it may concern,

This letter confirms the participation of Alisa Didenko (JINR, Dubna) within the **ATLAS Higgs and Di-Higgs Physics** working group. Ms Didenko has made significant contributions to the recently published search for the production of a Higgs boson in association with a single top quark. Her contributions are detailed below.

- ***Search for the production of a Higgs boson in association with a single top quark in*** pp ***collisions at*** $\sqrt{s}$ ***= 13 TeV with the ATLAS detector***, [arXiv:2508.14695]
    - development of deep neural network (NN) for classification of tH events with respect to main backgrounds,
    - development of algorithms for optimization of NN hyperparameters,
    - optimization of the list of input variables,
    - comparison of NN and boosted decision tree performances,
    - validation of NN results across signal and control regions.

Ms Didenko has actively participated in all related internal group meetings and contributed to the discussions for this analysis.

Please do not hesitate to contact us in case of further questions.

Sincerely,

Arely Cortes Gonzalez | Paolo Francavilla

ATLAS Higgs and Di-Higgs conveners
atlas-phys-higp-conveners@cern.ch

Lists of kinematic variables.

njets_CBT5	number of b-jets (in 5th CBT bin)
nnonbjets	number of non-b-jets
sphericity	uniform distribution of jets in space
aplanarity	deviation of jets from one common plane
nonbjets_eta[1]	eta angle for the second soft jet
rapgap_top_fwdjet	difference between the top-quark pseudorapidity and forward-jet
fwdjets_pt[0]	transverse momentum of a forward scattered quark
chi2_min_DeltaEta_tH	difference between the pseudorapidity of top-quark and Higgs
tagnonb_eta	pseudorapidity of the tagging-jet
tagnonb_topb_m	invariant mass of the tagging-jet and the b-jet from the top-quark
nfwdjets	number of jets in the front region of the detector
chi2_min_tophad_m_ttAll	reconstruction of the top quark mass

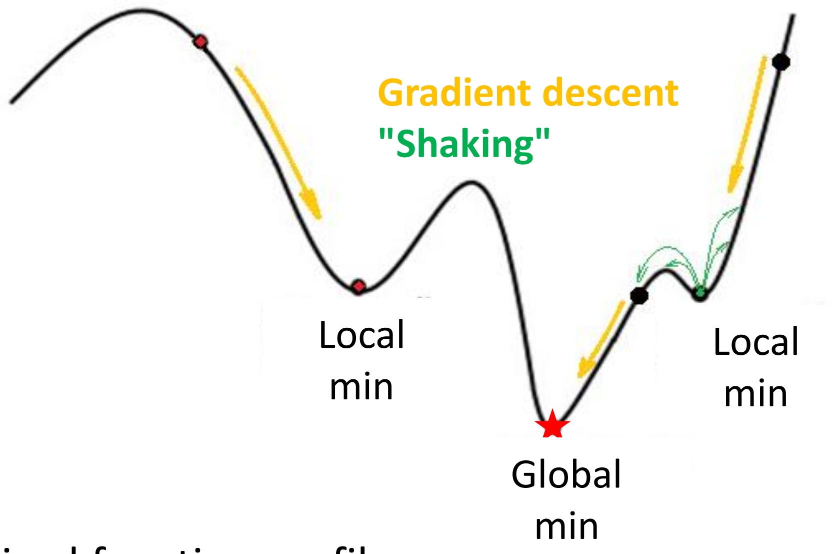
rapgap_maxptjet	difference between the pseudorapidity of forward jet and jet with the largest pt
inv3jets	mass of the three jets with the largest pt
nbjets	number b-jets = njets_CBT5 + njets_CBT4
chi2_min_toplep_pt	transverse momentum of the jet from the semi-hadron decay of the t-quark
nonbjets_pt[0]	transverse momentum of the leading light jet
chi2_min_deltaRq1q2	angular distance between the jets from the hadron decay of W
chi2_min_Whad_m_ttAll	reconstruction mass of W decaying through the hadron channel
leptons_charge[0]	charge of the leading lepton
foxWolfram_2_momentum	geometric correlations between jets
chi2_min_Imvmass_tH	invariant mass of the top quark and the Higgs boson
chi2_min_bbnonbjet_m	Invariant mass of the Higgs boson and light-jet
chi2_min_higgs_m	reconstruction mass of the Higgs boson

Application of a neural network for the classification of processes. Learning algorithms

Learning algorithms - algorithms for finding parameter values.

Really used algorithms try to combine the advantages of deterministic and stochastic methods. In most problems, a sufficiently **deep local minimum** is a satisfactory solution.

The algorithm minimizes the measure of difference between the "true" value of the target variable and the value predicted by the neural network.



Optimized function profile

Settings of the network

Number of layers	Number of training iterations
Number of neurons in a layer	The size of the batch of parameters trained at a time
Input variables	Activation functions and optimization

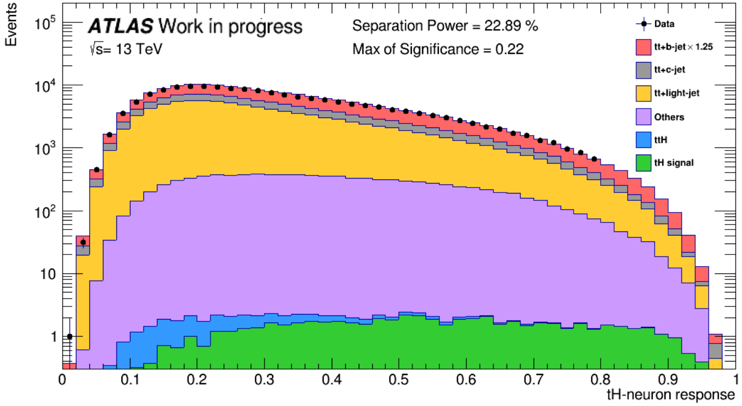
Learning algorithms of the network

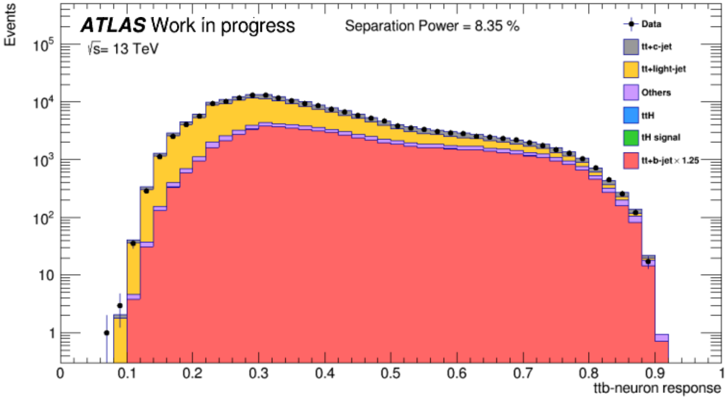
Absolutely random search	Gradient descent (derivative of the function to be minimized)

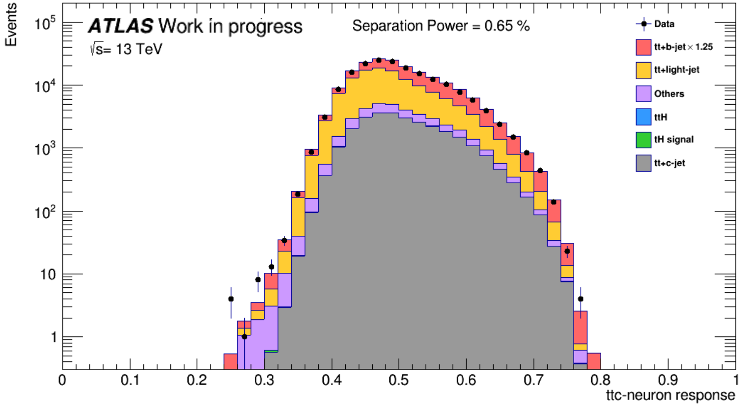
NN outputs on ***v34_minintuples_v1***, *reported on* ***Htop_tHbb*** *meeting (8.02.23, 15.02.23)*

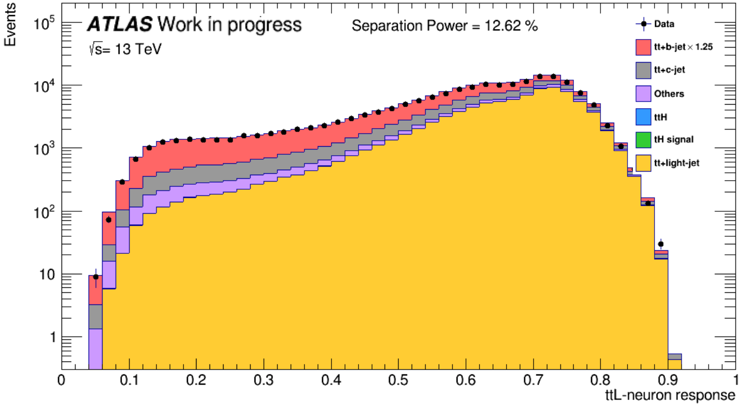
https://indico.cern.ch/event/1252954/

https://indico.cern.ch/event/1255098/

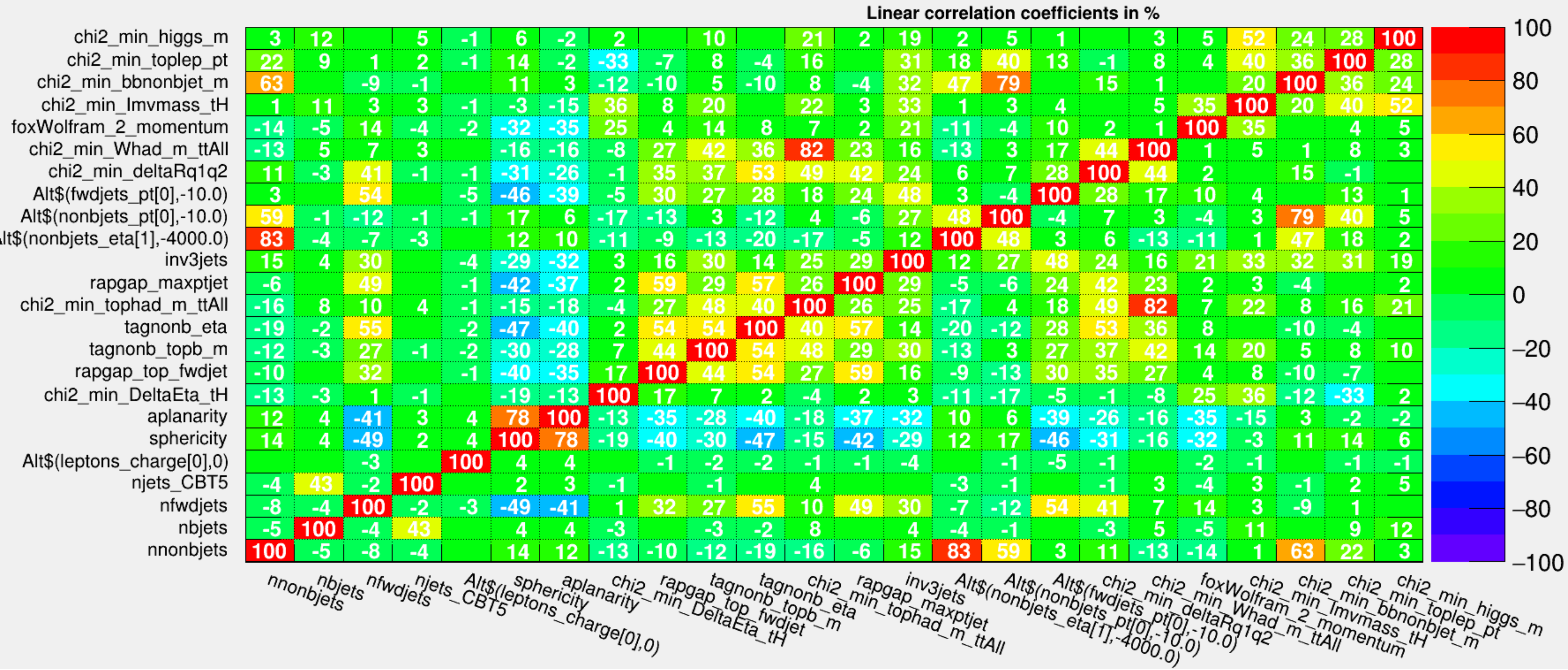




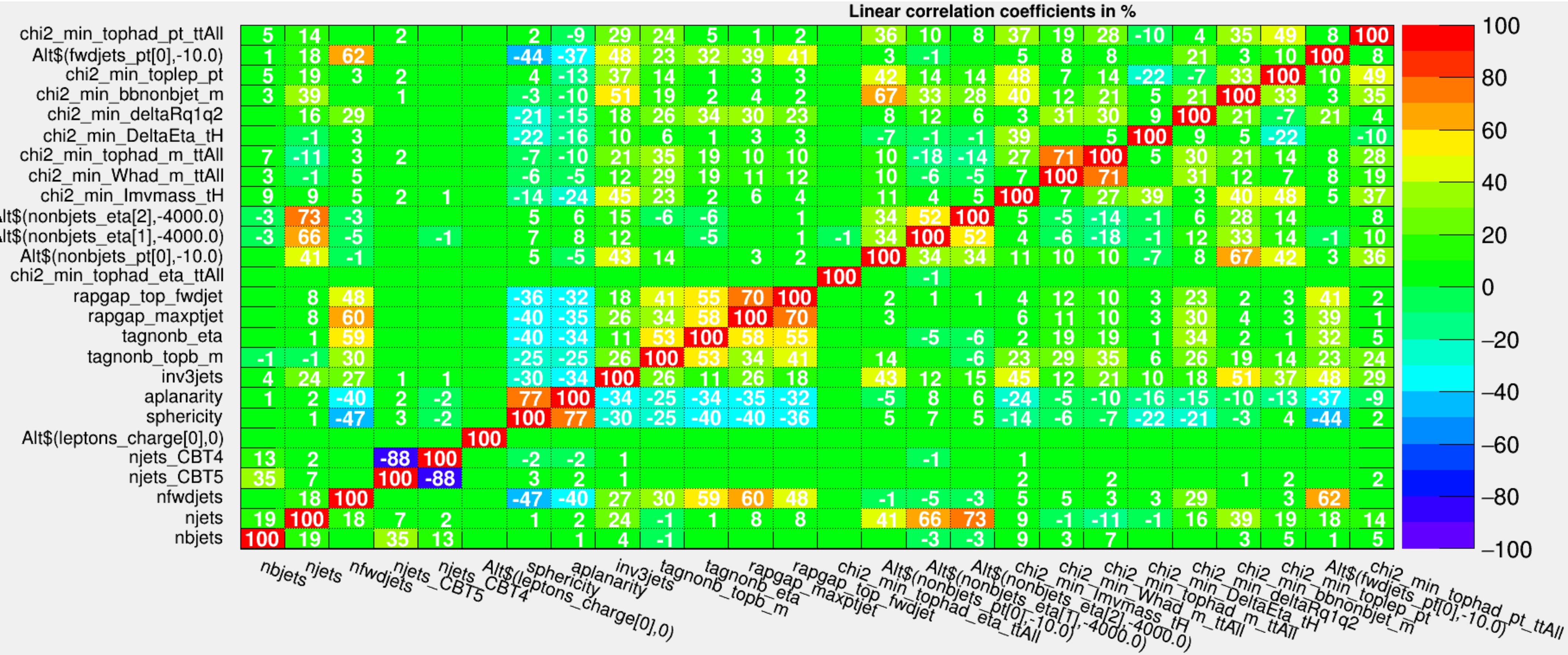




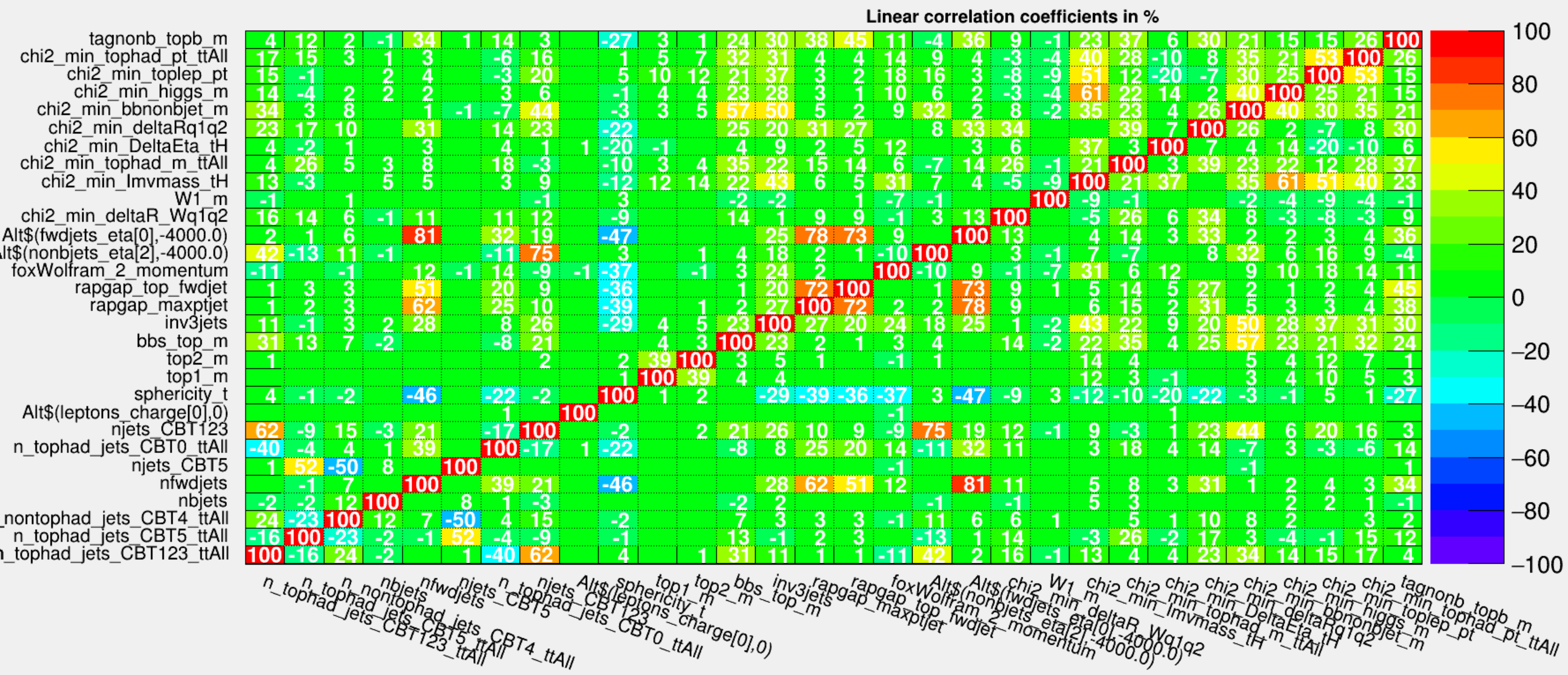
Lists of kinematic variables.



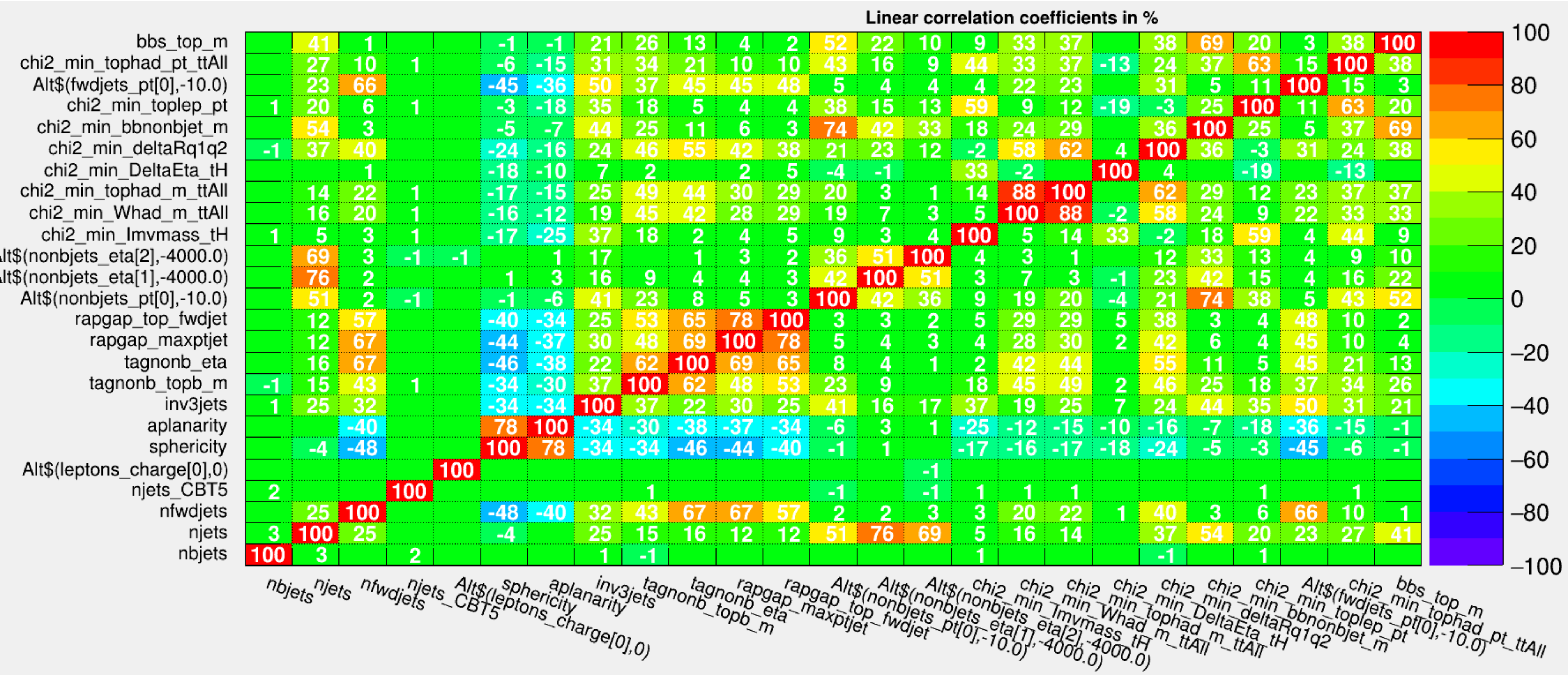
Lists of kinematic variables.



Lists of kinematic variables.



Lists of kinematic variables.



Network responses from 0 to 1

F(x) = (x - 0.06) * 1.1395 - 0.173