APPLICATION OF MACHINE LEARNING FOR THE ANALYSIS OF HIGGS BOSON PRODUCTION IN ASSOCIATION WITH SINGLE TOP-QUARK

Dubna team: I.Boyko, N.Huseynov, A.Daribayeva, A.Tropina , Oksana Koval,

I.Yeletskikh, A.Didenko

The XXVI International Scientific Conference of Young Scientists and Specialists (AYSS-2022)

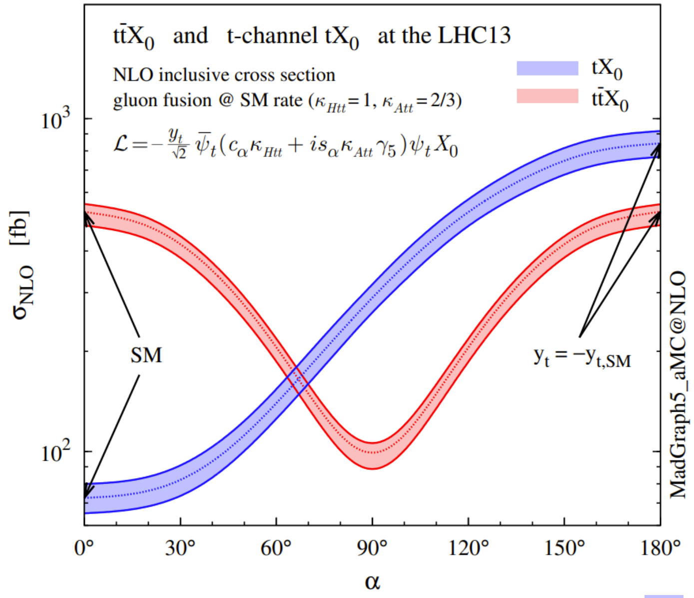
Introduction

1. Goal and tasks

The goal of this work is experimentally research the processes of the production of the Higgs boson in association with a single top quark ($pp \rightarrow tH$), in particular, research the possibilities of kinematic separation of the signal process from the background ones using deep machine learning.

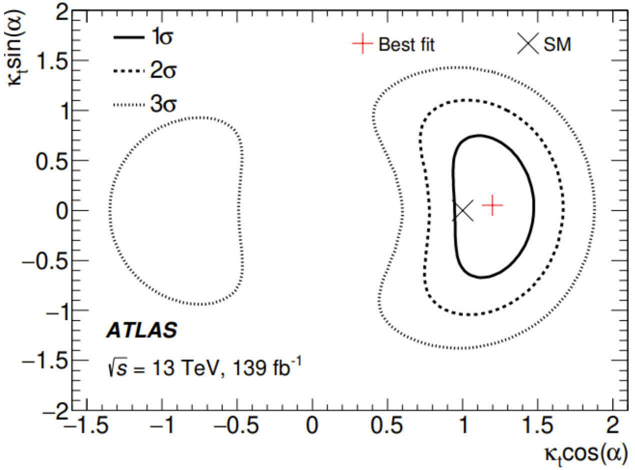
2. Theory of tH production processes
   - Kinematics of the signal process
3. Application of a neural network for process classification
   - Neural network as one of the machine learning models
   - Neural network structure
   - Learning algorithms
4. Results
   - Network response to data and MC events
   - Significance for the tH process

Production of the Higgs boson in association with a single top quark. Top-Higgs coupling



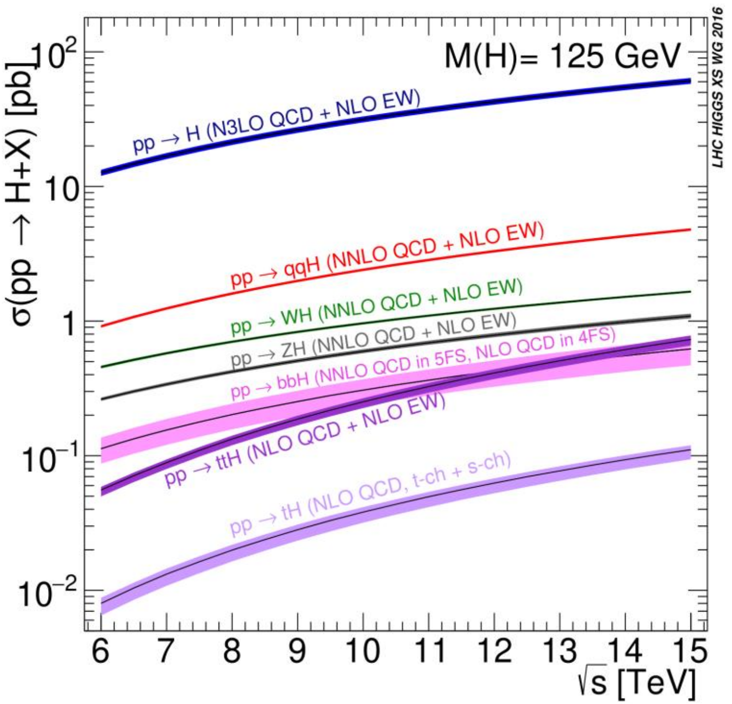
The cross section for the production of tH and ttH in pp collisions at 13 TeV as a function of the CP mixing angle α.
Where the value **α = 0** corresponds to the Standard Model.

[1] Higgs production in association with a single top quark at the LHC [arXiv:1504.00611]



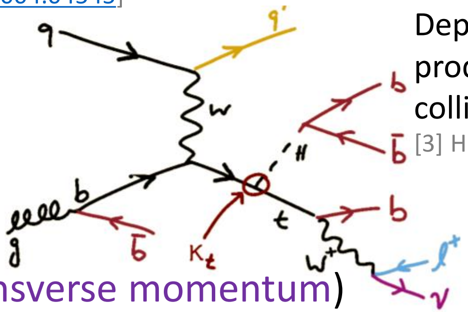
Combined ATLAS results for the measurement of the coupling constant.
The SM predicts alpha = 1.

[2] CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}$H and tH Processes Using H→γγ with the ATLAS Detector[arXiv:2004.04545]



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

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

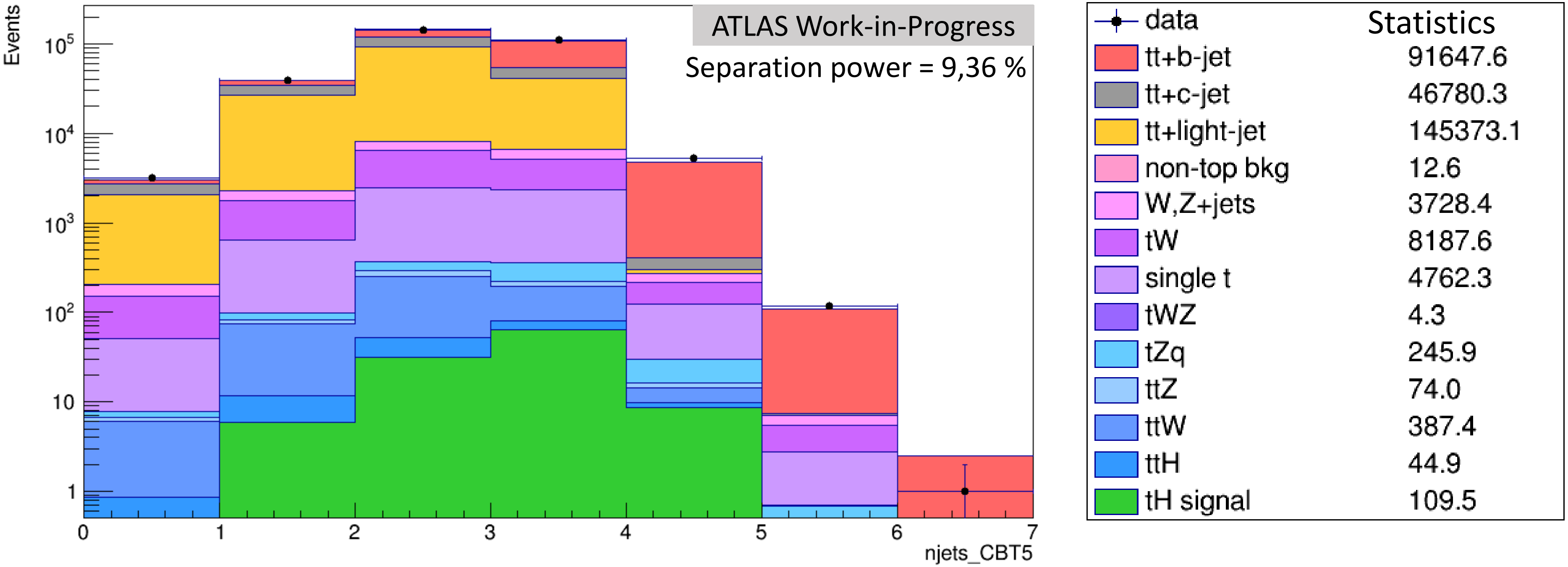


Signature of tH with decay $H \rightarrow b\bar{b}$:
(≥3 b-jets) + (1 light jet) + (1 tight lepton) + (missing transverse momentum)

Kinematics of signal and background processes

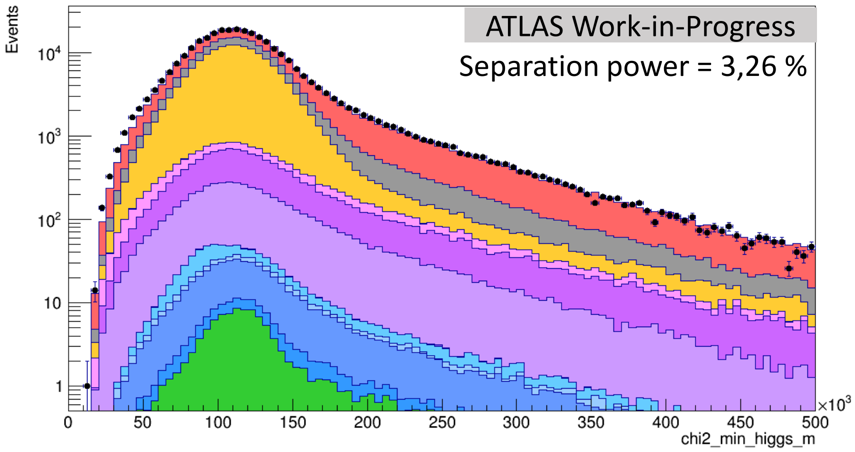
The signal and background processes differ in a large number of kinematic variables, but for each of them the differences are small. The background separation problem can be solved using machine learning.
it is very important for machine learning to have high-quality modeling data.
For events that have passed preselection, the quality of modeling the distributions of the main kinematic variables is checked.

Number of b-jets

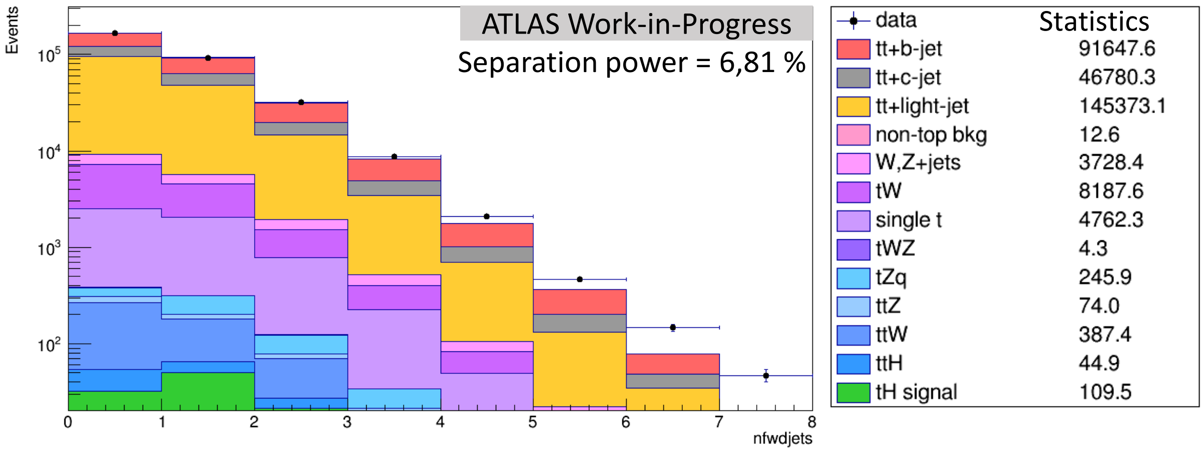


Kinematics of signal and background processes

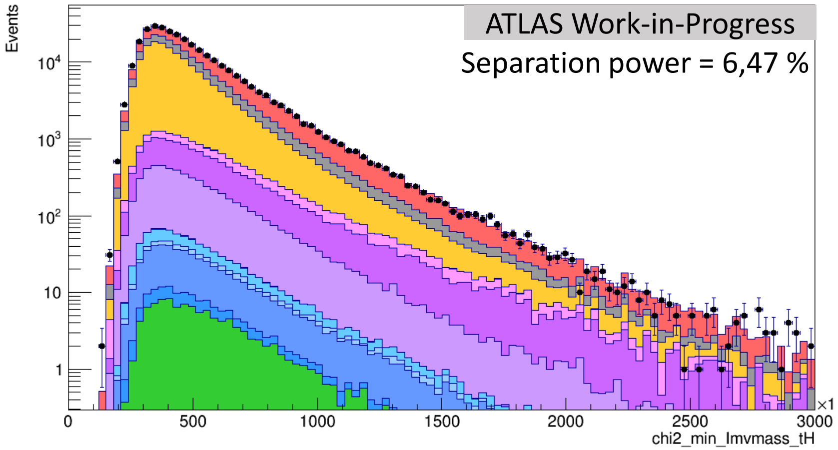
Reconstructed mass of the Higgs boson



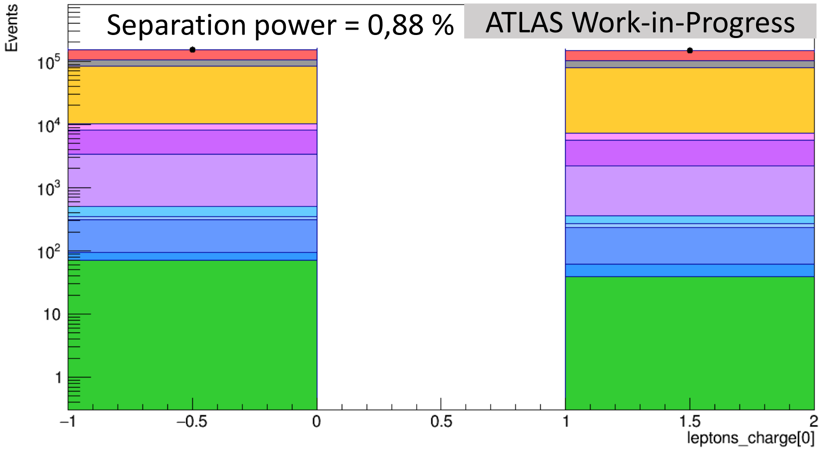
Number of forward jets



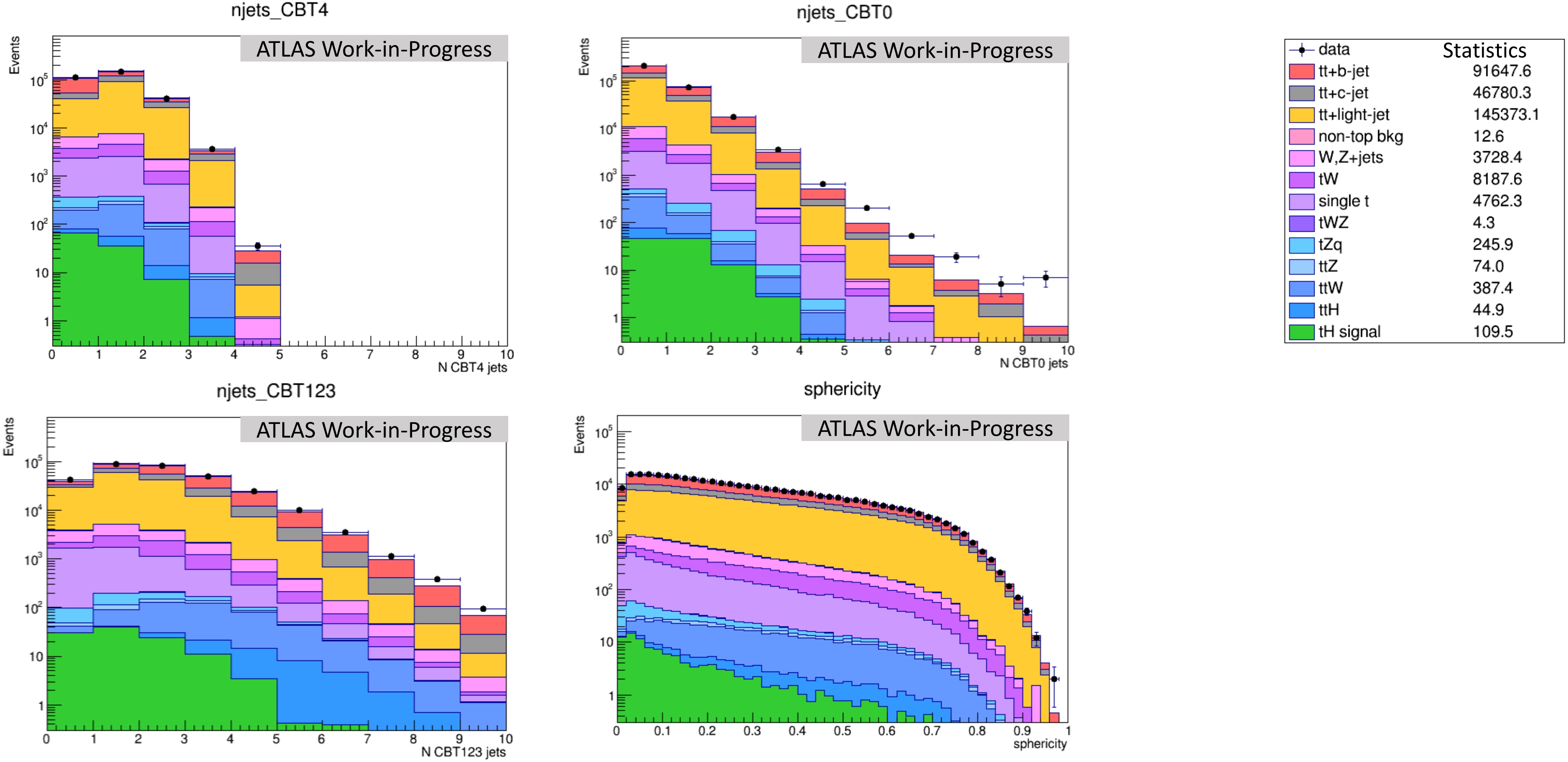
Invariant tH mass



The charge of the leading lepton



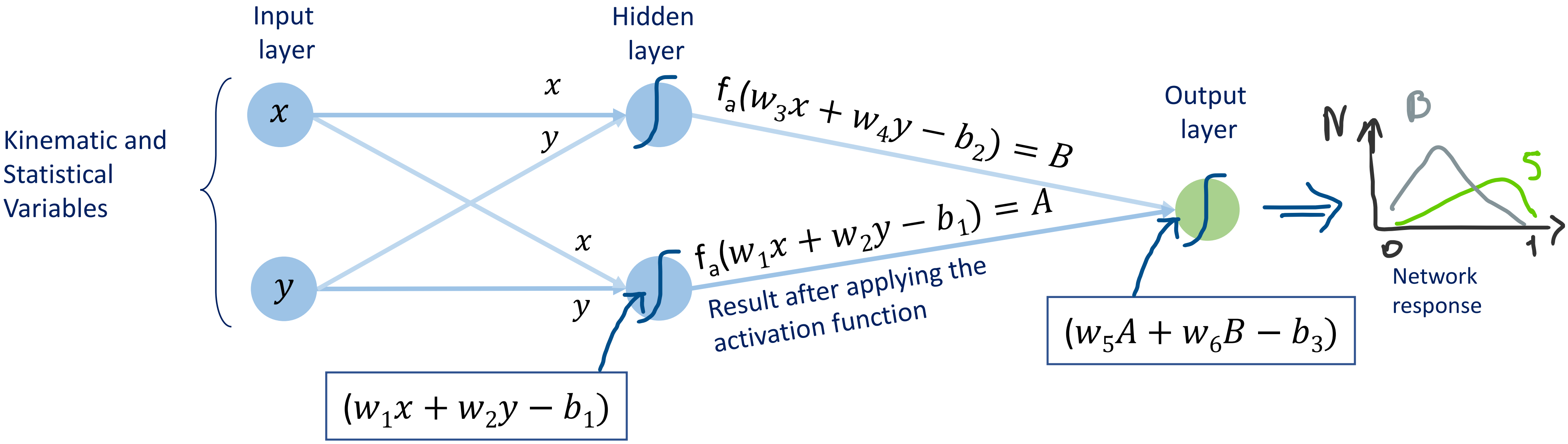
Kinematics of signal and background processes



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:



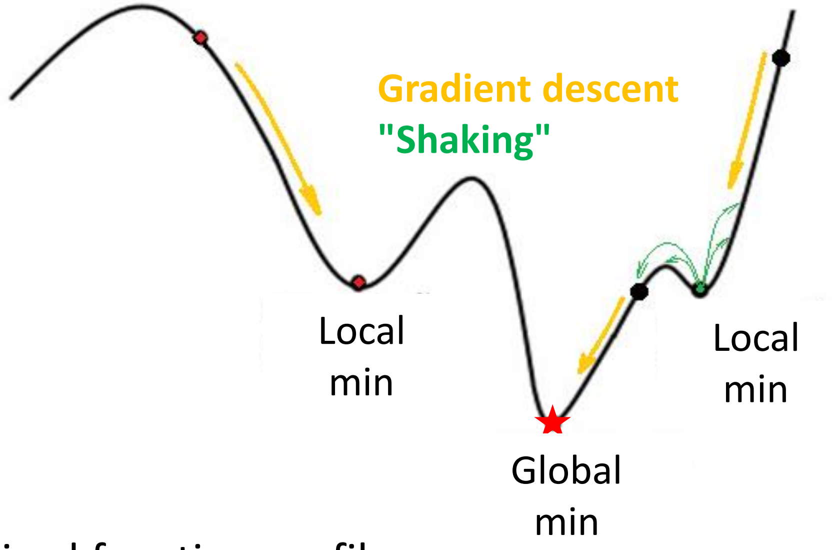
Keras – open sourse library, written in Python for artificial NN.
[Keras for root tmva]

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

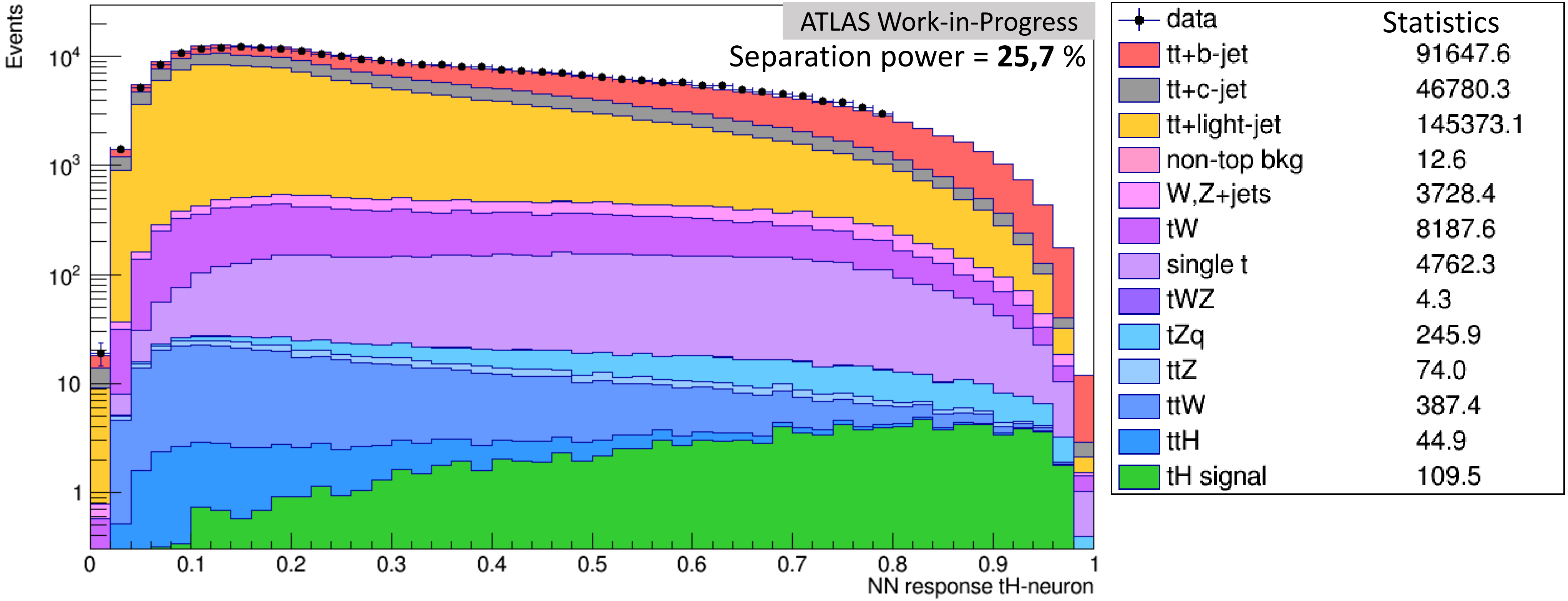
Learning algorithms of the network	
Absolutely random search	Gradient descent (derivative of the function to be minimized)

Results. Network response to data events and MC: signaling neuron

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**.**

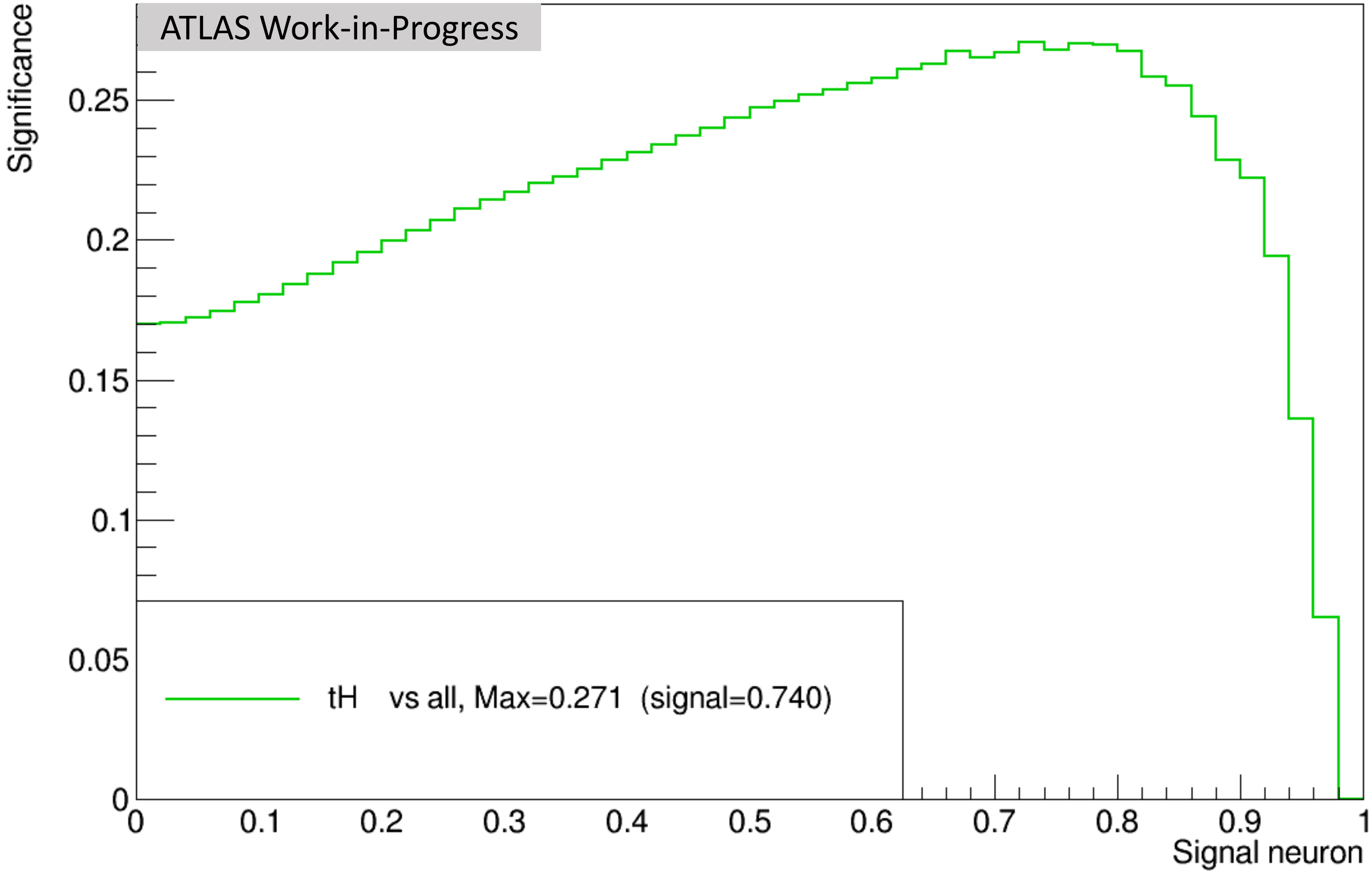
The data is well described MC. Some processes were not included in training (for example, single t).

Experimental data in the region > 0.8 are not shown.



Results. Significance for the tH process

The maximum significance of the signal is achieved for selection for the response of a tH-neuron > **0.74**



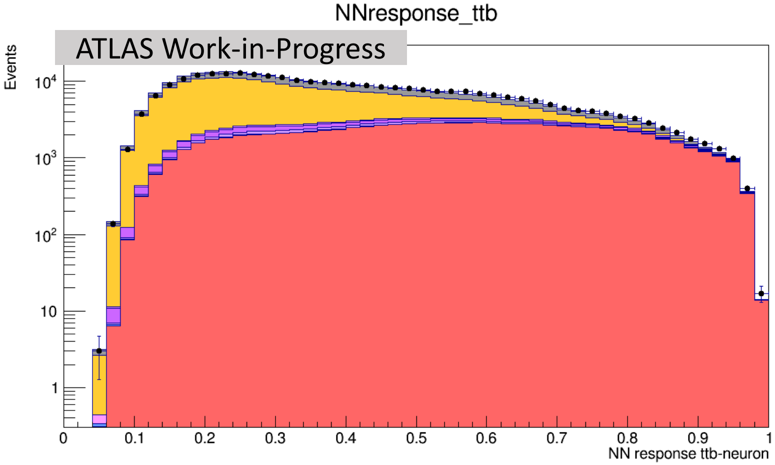
$$Signif_i = \frac{s_i}{\sqrt{s_i + b_i}}$$

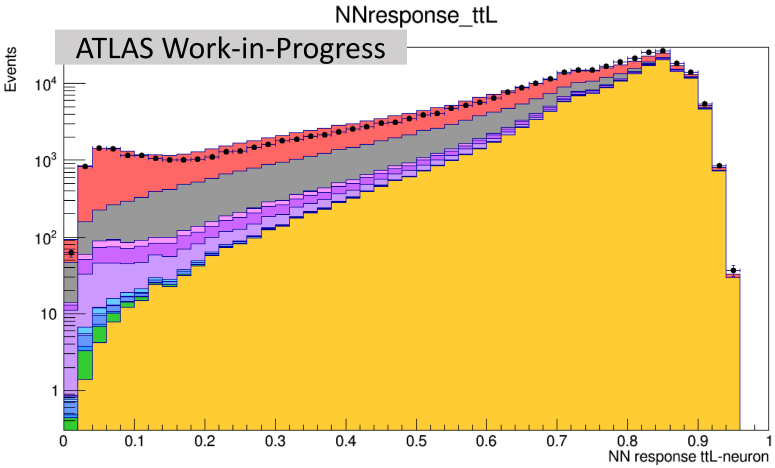
For pp collisions at 13 TeV, the integral luminosity is 139 fb^{-1} :

- Without the NN, the significance of the signal is **0.17** σ
- After applying the NN: **0.27** σ

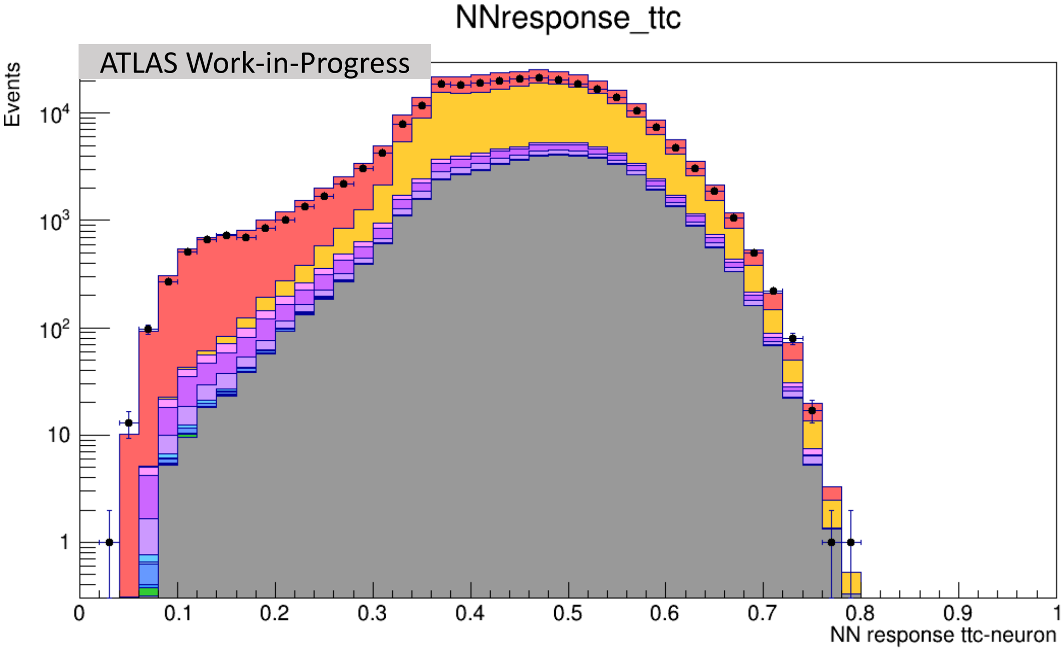
Thus, regardless of the amount of experimental data, the NN allows to increase the significance of the signal by **1.6**.
Regarding data sensitivity to potential deviation from SM.

Results. Network response to data events and MC: neurons ttb, ttc, ttL

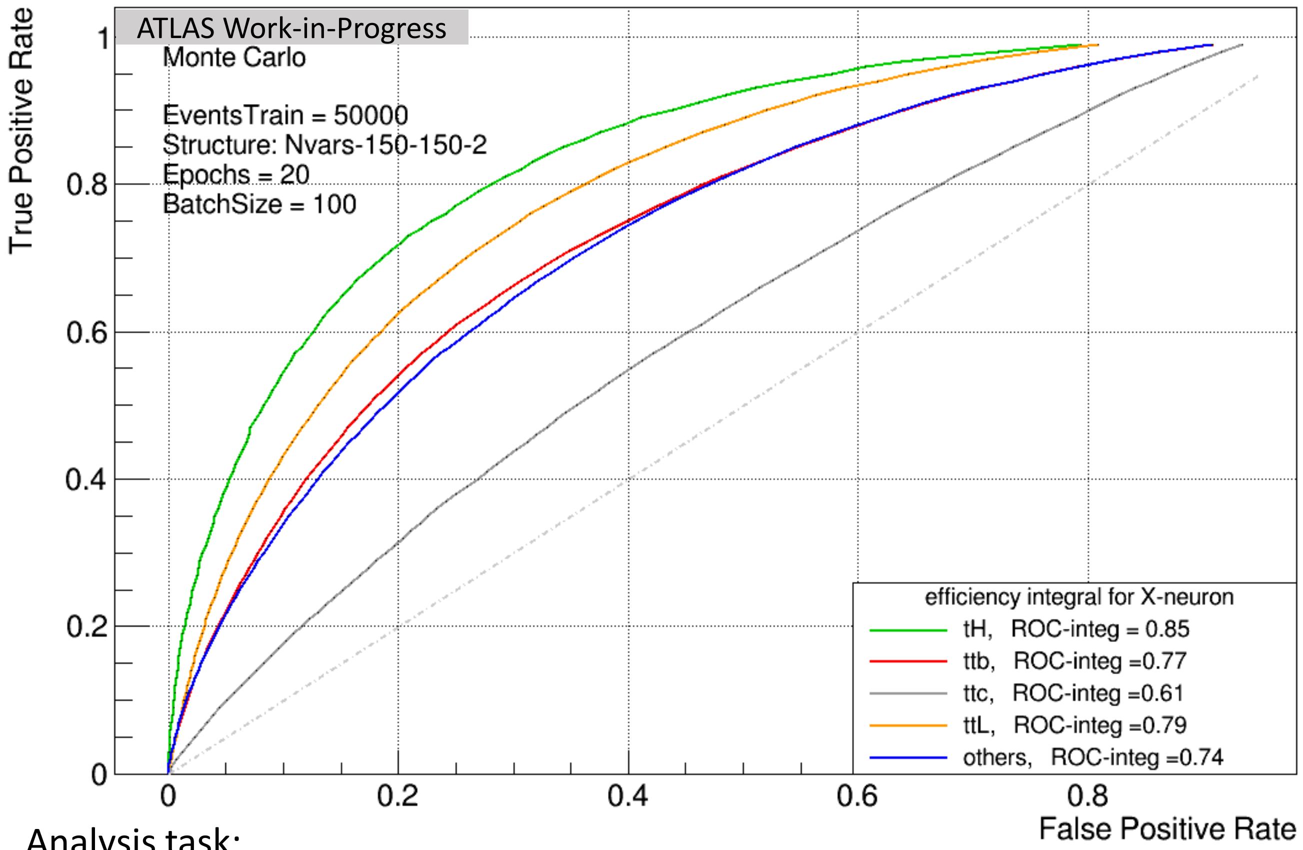




Through the use of a network, corrections to background process cross sections can be extracted from the data by fitting.



Efficiency plot for the multiclassifier. Network structure 24->150/150->2



ROC-integ (TMVA pyKeras)	0.85
BDT 2021	0.86

Analysis task:

- separate the signal from the background,
- separate background processes from each other..

* The ttc, ttb, ttL cross section in the simulated data is not accurate enough..

Results

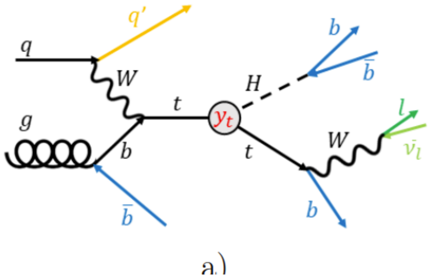
1. found:
    - optimal network structure
    - optimal network settings
    - optimal set of variables
    - area of maximum signal-to-background ratio
2. Application of a NN increases the significance of the signal by **1.6**.

Plans

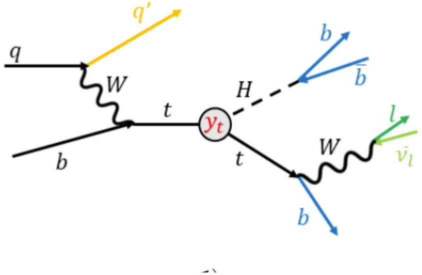
Using a NN to refine the relative contribution of the main background prosses (tt+b-jet, tt+c-jet, tt+light-jet) to each other.

BackUp

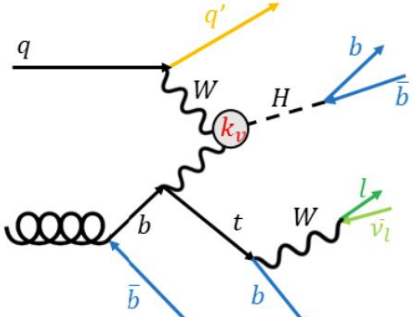
Feynman diagrams of signal tH process and main background.



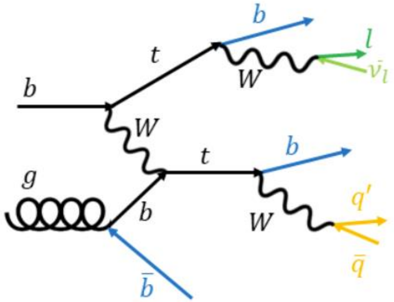
a) t-channel of tH



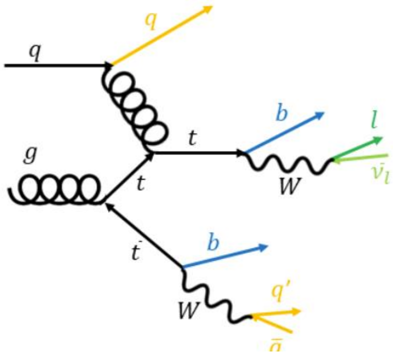
b) s-channel of tH



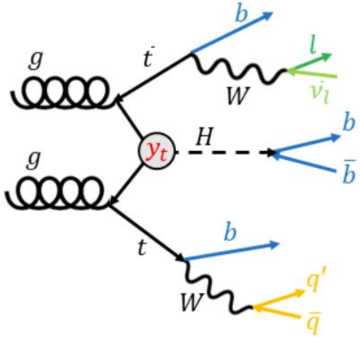
c) tH production, but Higggs interaction with W-boson



d) tt-b-jet



e) tt-c-jet
and tt-light-jet



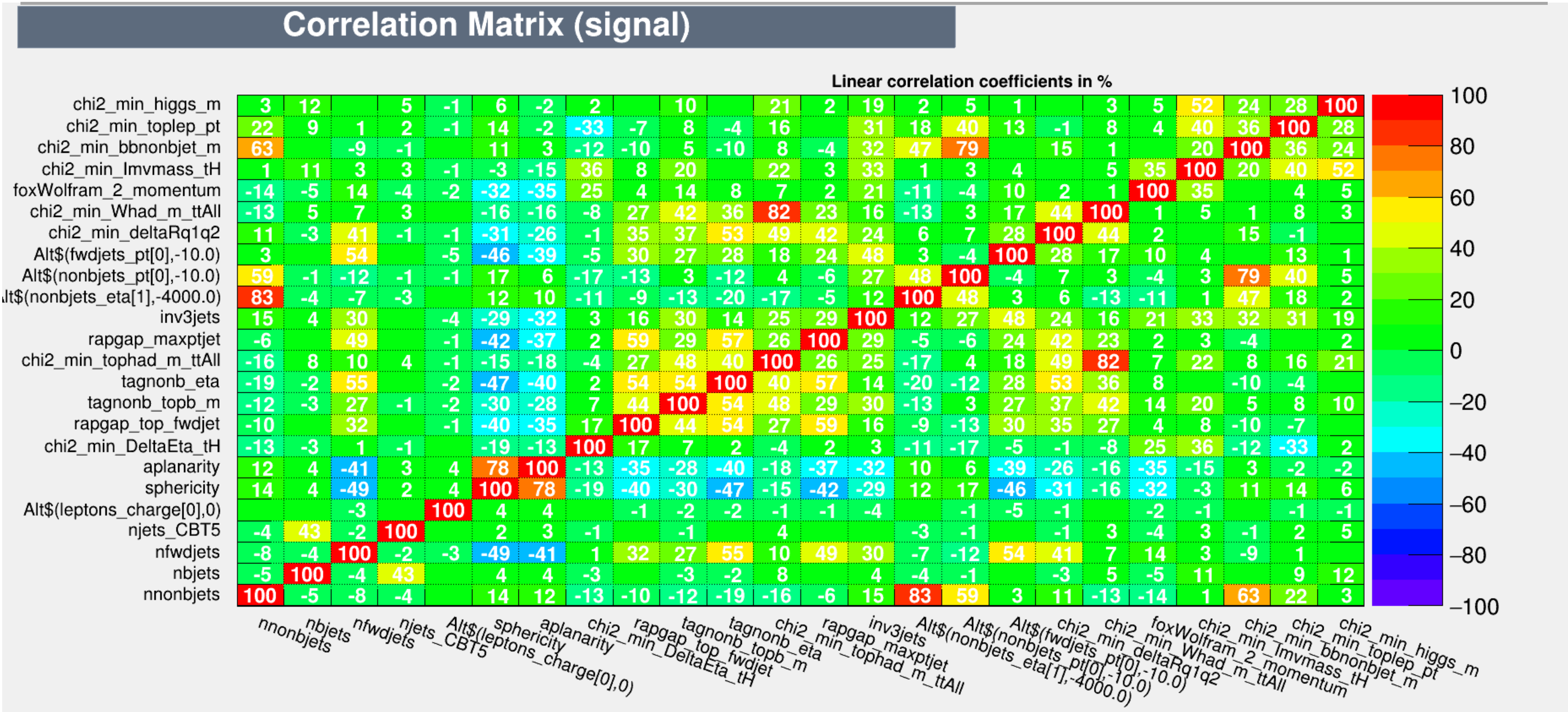
f) ttH

Lists of kinematic variables. TMVA pyKeras NN v34_minintuples_v1

tH	ttb	ttc	ttL	Others (ttH, ttZ, ttW, tZq)
1 : njets_CBT5	1 : njets_CBT5	1 : njets_CBT5	1 : njets	1 : njets
2 : nnonbjets	2 : njets_CBT4	2 : n_nontophad_jets_CBT4_ttAll	2 : Alt$(nonbjets_eta[1],-4000.0)	2 : Alt$(nonbjets_eta[2],-4000.0)
3 : sphericity	3 : nbjets	3 : nbjets	3 : Alt$(nonbjets_eta[2],-4000.0)	3 : Alt$(nonbjets_pt[0],-10.0)
4 : aplanarity	4 : njets	4 : n_tophad_jets_CBT123_ttAll	4 : chi2_min_bbnonbjet_m	4 : Alt$(nonbjets_eta[1],-4000.0)
5 : Alt$(nonbjets_eta[1],-4000.0)	5 : Alt$(nonbjets_eta[1],-4000.0)	5 : top1_m	5 : Alt$(nonbjets_pt[0],-10.0)	5 : chi2_min_bbnonbjet_m
6 : rapgap_top_fwdjet	6 : chi2_min_bbnonbjet_m	6 : n_tophad_jets_CBT5_ttAll	6 : njets_CBT5	6 : chi2_min_tophad_pt_ttAll
7 : Alt$(fwdjets_pt[0],-10.0)	7 : chi2_min_Imvmass_tH	7 : top2_m	7 : inv3jets	7 : inv3jets
8 : chi2_min_DeltaEta_tH	8 : chi2_min_DeltaEta_tH	8 : bbs_top_m	8 : chi2_min_Imvmass_tH	8 : chi2_min_toplep_pt
9 : tagnonb_eta	9 : inv3jets	9 : njets_CBT123	9 : nbjets	9 : chi2_min_tophad_m_ttAll
10 : tagnonb_topb_m	10 : Alt$(nonbjets_eta[2],-4000.0)	10 : chi2_min_deltaRq1q2	10 : bbs_top_m	10 : chi2_min_Whad_m_ttAll
11 : nfwdjets	11 : chi2_min_deltaRq1q2	11 : chi2_min_DeltaEta_tH	11 : chi2_min_toplep_pt	11 : Alt$(fwdjets_pt[0],-10.0)
12 : chi2_min_tophad_m_ttAll	12 : Alt$(nonbjets_pt[0],-10.0)	12 : Alt$(leptons_charge[0],0)	12 : chi2_min_tophad_pt_ttAll	12 : chi2_min_deltaRq1q2
13 : rapgap_maxptjet	13 : chi2_min_tophad_m_ttAll	13 : chi2_min_deltaR_Wq1q2	13 : chi2_min_deltaRq1q2	13 : chi2_min_Imvmass_tH
14 : inv3jets	14 : nfwdjets	14 : n_tophad_jets_CBT0_ttAll	14 : chi2_min_tophad_m_ttAll	14 : Alt$(leptons_charge[0],0)
15 : nbjets	15 : tagnonb_topb_m	15 : chi2_min_Imvmass_tH	15 : Alt$(fwdjets_pt[0],-10.0)	15 : tagnonb_topb_m
16 : chi2_min_toplep_pt	16 : rapgap_maxptjet	16 : chi2_min_toplep_pt	16 : tagnonb_topb_m	16 : nfwdjets
17 : Alt$(nonbjets_pt[0],-10.0)	17 : aplanarity	17 : Alt$(nonbjets_eta[2],-4000.0)	17 : nfwdjets	17 : chi2_min_DeltaEta_tH
18 : chi2_min_deltaRq1q2	18 : Alt$(fwdjets_pt[0],-10.0)	18 : chi2_min_bbnonbjet_m	18 : chi2_min_Whad_m_ttAll	18 : nbjets
19 : chi2_min_Whad_m_ttAll	19 : sphericity	19 : chi2_min_higgs_m	19 : tagnonb_eta	19 : tagnonb_eta
20 : Alt$(leptons_charge[0],0)	20 : rapgap_top_fwdjet	20 : chi2_min_tophad_pt_ttAll	20 : rapgap_maxptjet	20 : aplanarity
21 : foxWolfram_2_momentum	21 : chi2_min_toplep_pt	21 : foxWolfram_2_momentum	21 : aplanarity	21 : rapgap_maxptjet
22 : chi2_min_Imvmass_tH	22 : chi2_min_tophad_pt_ttAll	22 : nfwdjets	22 : chi2_min_DeltaEta_tH	22 : rapgap_top_fwdjet
23 : chi2_min_bbnonbjet_m	23 : tagnonb_eta	23 : inv3jets	23 : rapgap_top_fwdjet	23 : njets_CBT5
24 : chi2_min_higgs_m	24 : Alt$(leptons_charge[0],0)	24 : rapgap_top_fwdjet	24 : sphericity	24 : foxWolfram_2_momentum
	25 : chi2_min_Whad_m_ttAll	25 : chi2_min_tophad_m_ttAll	25 : Alt$(leptons_charge[0],0)	25 : sphericity
	26 : chi2_min_tophad_eta_ttAll	26 : Alt$(fwdjets_eta[0],-4000.0)		
		27 : rapgap_maxptjet		
		28 : W1_m		
		29 : tagnonb_topb_m		
		30 : sphericity_t		

ROC-integ		ROC-integ		ROC-integ		ROC-integ		ROC-integ	
TMVA pyKeras	0.85	TMVA pyKeras	0.77	TMVA pyKeras	0.61	TMVA pyKeras	0.79	TMVA pyKeras	0.74
BDT 2021	0.86	Old BDT 2021	0.75	BDT 2021	0.60	BDT 2021	0.78	BDT 2021	0.64

Task of separating. Selecting a set of variables. Correlation matrix.



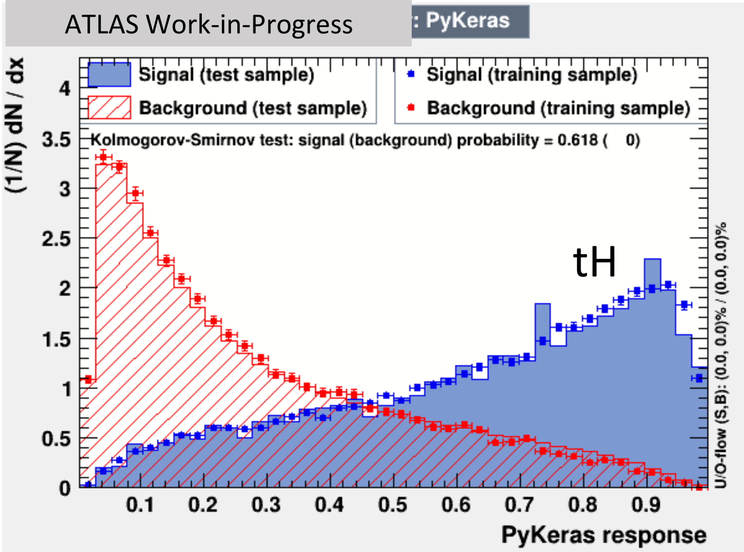
The analysis uses up to 30 variables that are most sensitive to the main processes. Correlation < 88%.

Lists of kinematic variables.

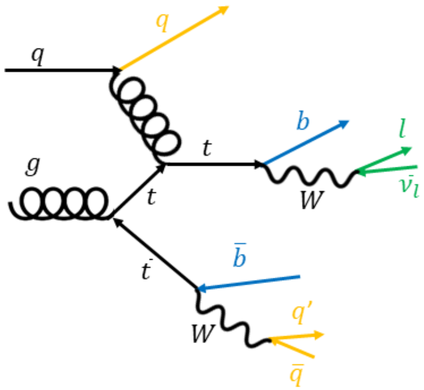
njets_CBT5	number of b-jets (in 5th CBT bin)
nnonbjets	number of non-b-jets
sphericity	uniform distribution of jets in space
aplanarity	deviation of jets from one common plane
nonbjets_eta[1]	eta angle for the second soft jet
rapgap_top_fwdjet	difference between the top-quark pseudorapidity and forward-jet
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

Training and testing.

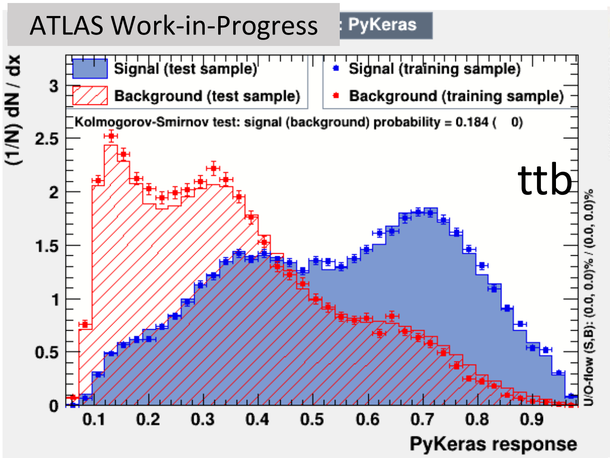


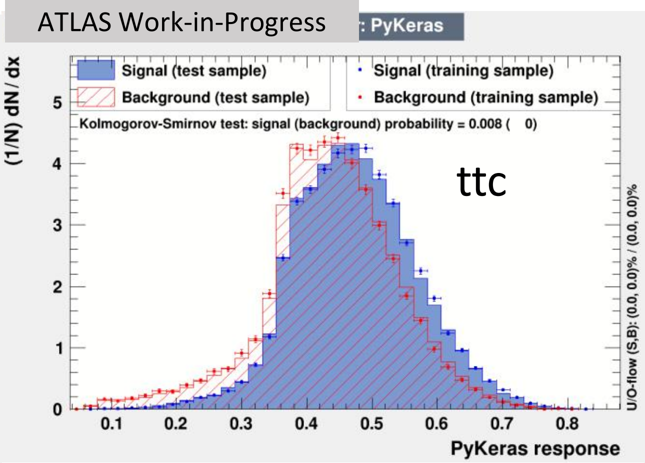
Testing set of events don't independent from training data set .
The same network response to training events and test events is expected.

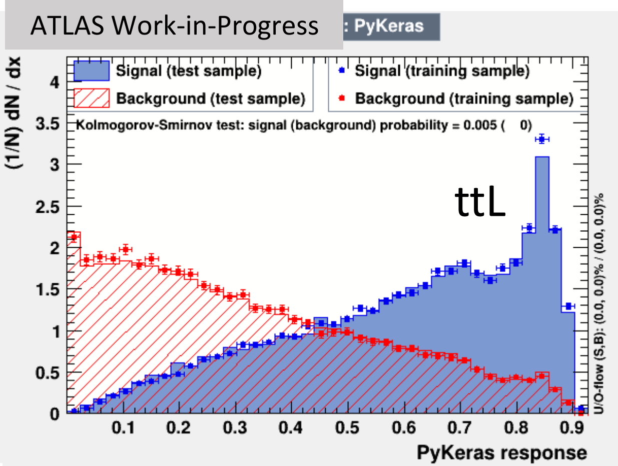


$t \rightarrow b + \ell + \nu$

$t < \begin{matrix} b + d + u\ (50\%) \\ b + c + s\ (50\%) \end{matrix}$





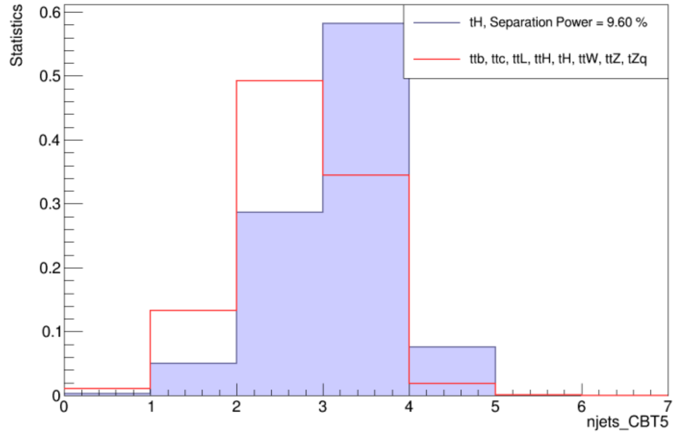


Separation power for tH.

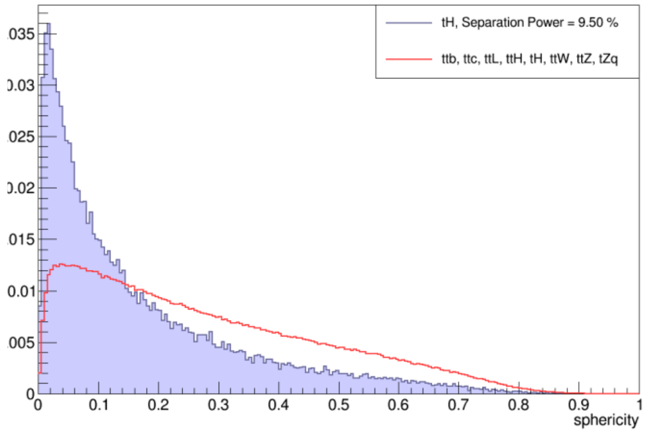
$$S = \frac{1}{2}\left(\sum_{i=0}^{nbins}\frac{(s_i-b_i)^2}{s_i+b_i}\right)\times 100$$

v34_minintuples _v1

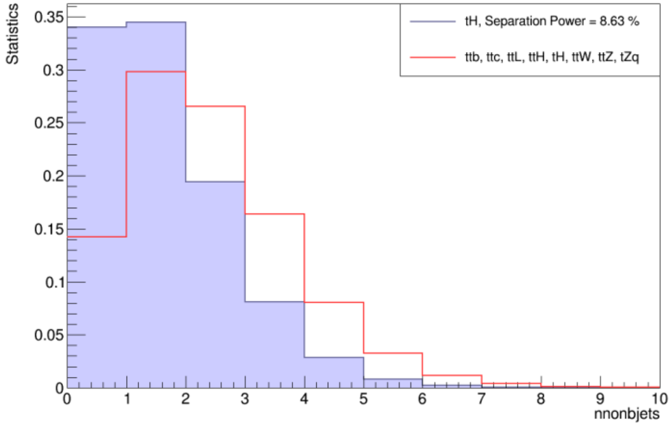
ATLAS Work-in-Progress njets_CBT5



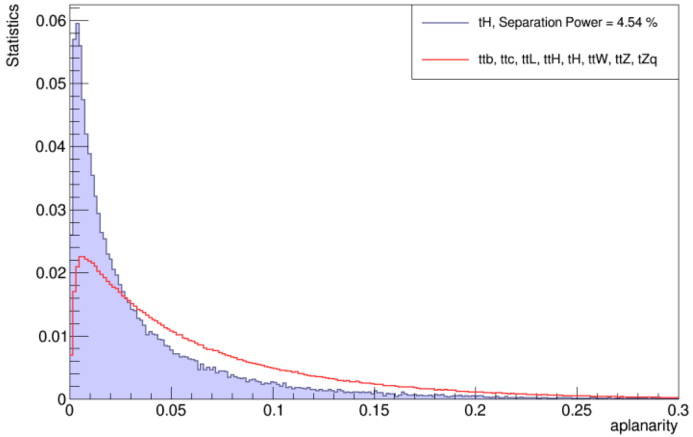
ATLAS Work-in-Progress sphericity



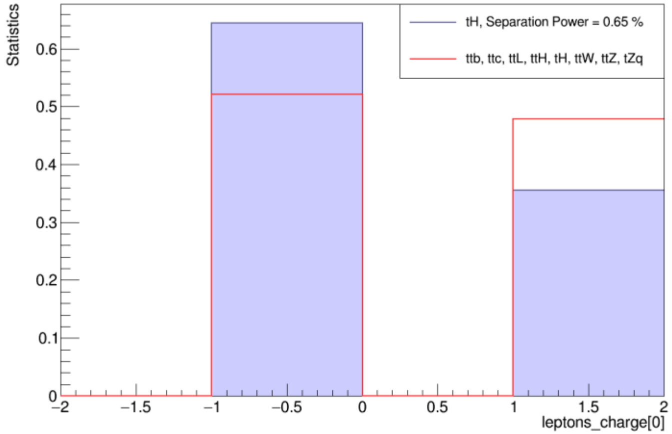
ATLAS Work-in-Progress nnonbjets



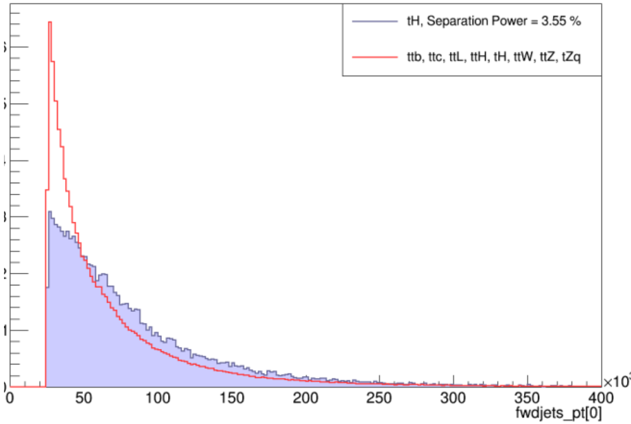
ATLAS Work-in-Progress aplanarity



ATLAS Work-in-Progress leptons_charge



ATLAS Work-in-Progress fwdjets_pt



Transformation.

Before training, the kinematic variables must be dimensionless and brought to a common range of values.
In TMVA ROOT:

- Normalisation,
- Uniformisation,
- Gaussianisation,
- Decorrelation.