The XXVII International Scientific Conference of Young Scientists and Specialists (AYSS-2023)

A new evolutionary algorithm for optimizing the search of a rare Higgs boson production channel

A.Didenko

Dubna team tHbb: I.Boyko, I.Yeletskikh, N.Huseynov, A.Tropina, O.Koval

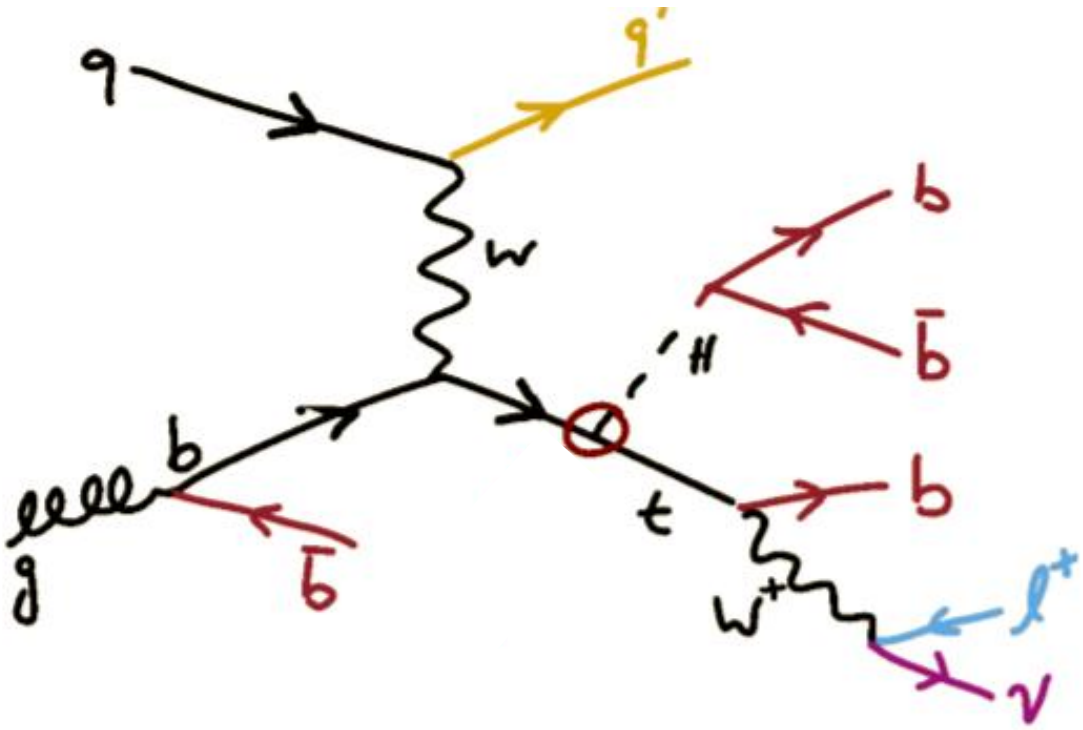
31.10.2023

Outline

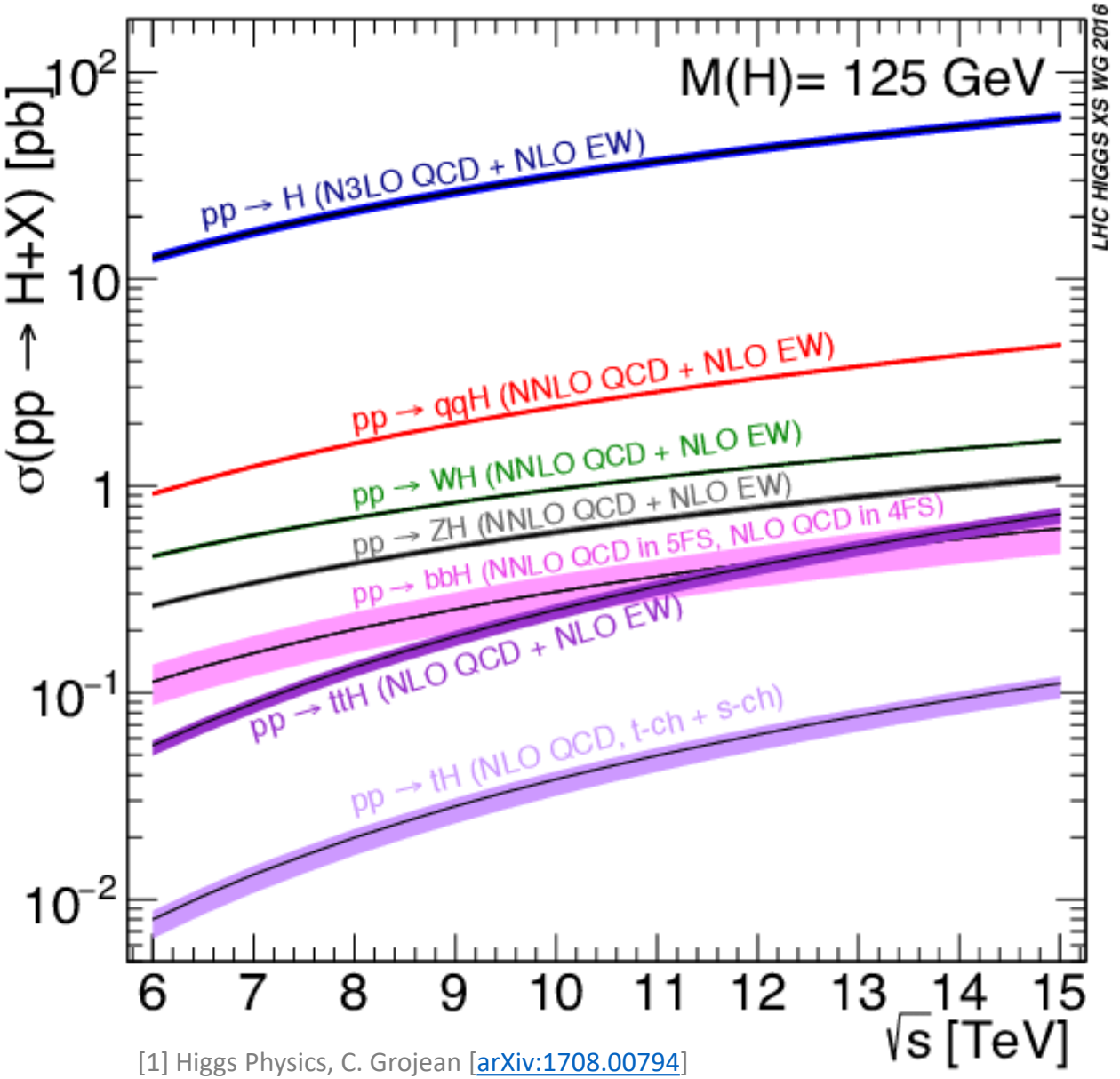
1. Motivation
   - Kinematics of signal and backgrounds
2. Artificial neural networks
   - Basic Concepts
   - Implicit network parameters (hyperparameters)
3. Hyperparameters optimization using evolutionary algorithm
4. Results
   - Optimized Network structure
   - Network response to signal and backgrounds
   - Significance for the tH process after applying the network
5. Conclusion and plans

Kinematics of signal and backgrounds

- The number of signal events is extremely small compared to the number of background events



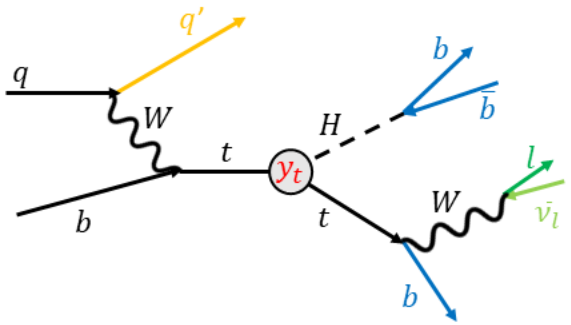
Signature of tH with decay $H \rightarrow b\bar{b}$:
(≥3 b-jets) + (1 light jet) + (1 tight lepton) + (missing transverse momentum)



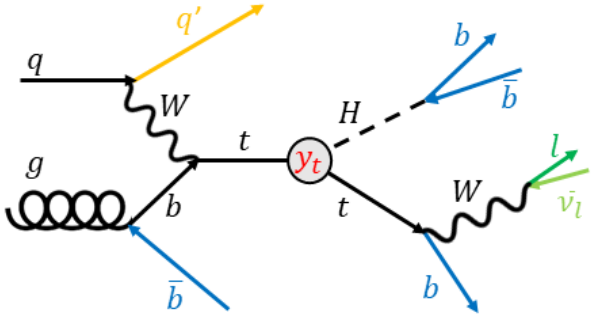
[1] Higgs Physics, C. Grojean [arXiv:1708.00794]

Kinematics of signal and backgrounds

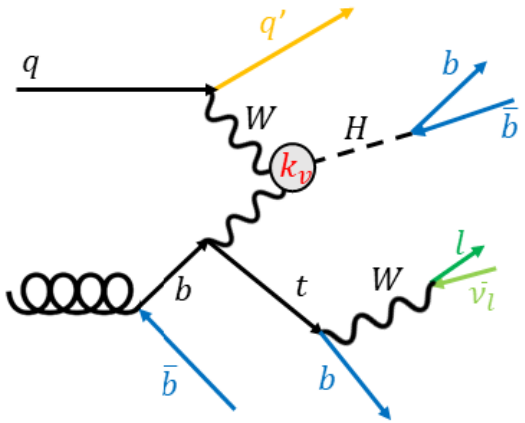
- The kinematics of signal and background processes are very close



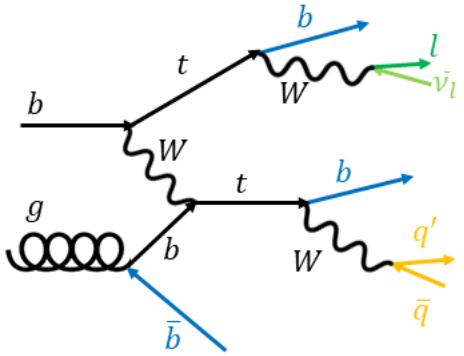
(a) t-channel of the signal tH process



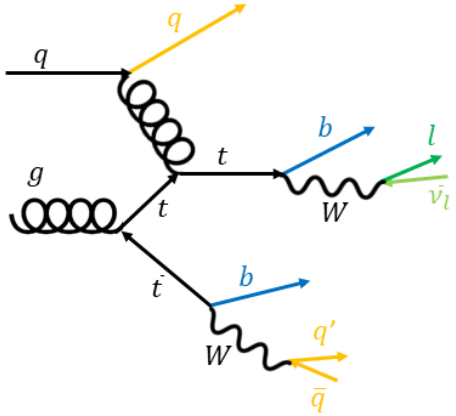
(b) s-channel of the signal tH process



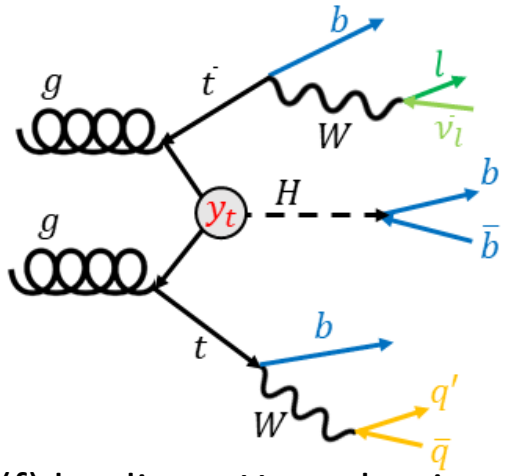
(c) Higgs boson production channel with top quark, where the Higgs boson interacts with the W boson



(d) leading ttb production channel



(e) leading ttc and ttL production channels

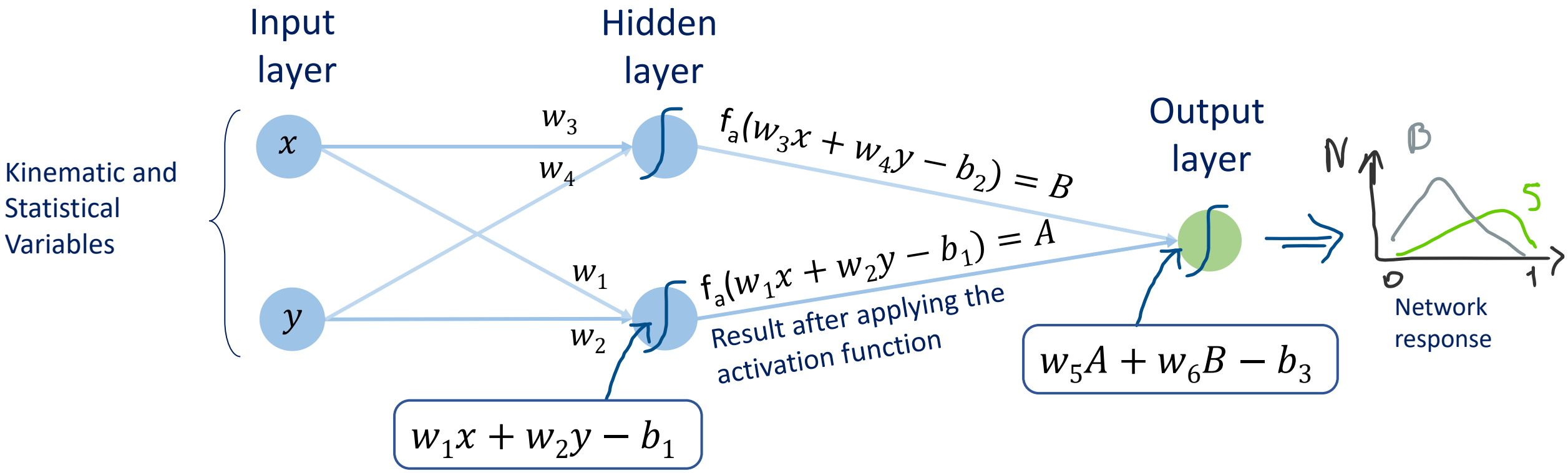


(f) leading ttH production channel

Neural network

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

Parameters: weights ($w_1 - w_6$) and shifts ($b_1 - b_3$).



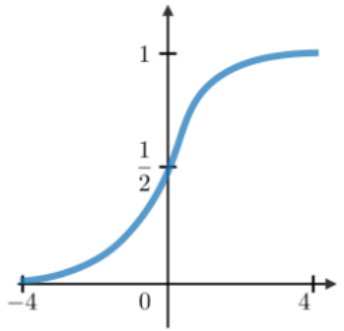
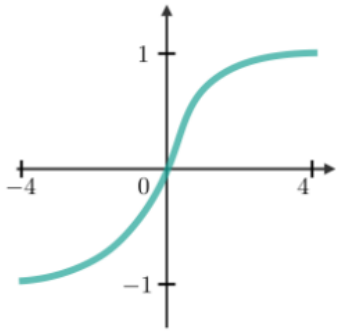
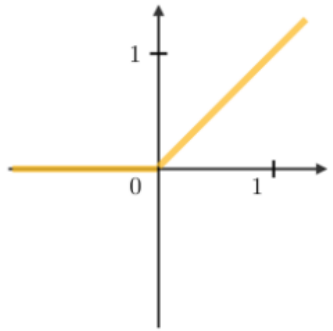
Keras – open source library, written in Python for artificial NN.

Keras for root tmva

Neural network. Implicit network parameters (hyperparameters)

- Number of layers
- Number of neurons in a layer
- Activation functions:

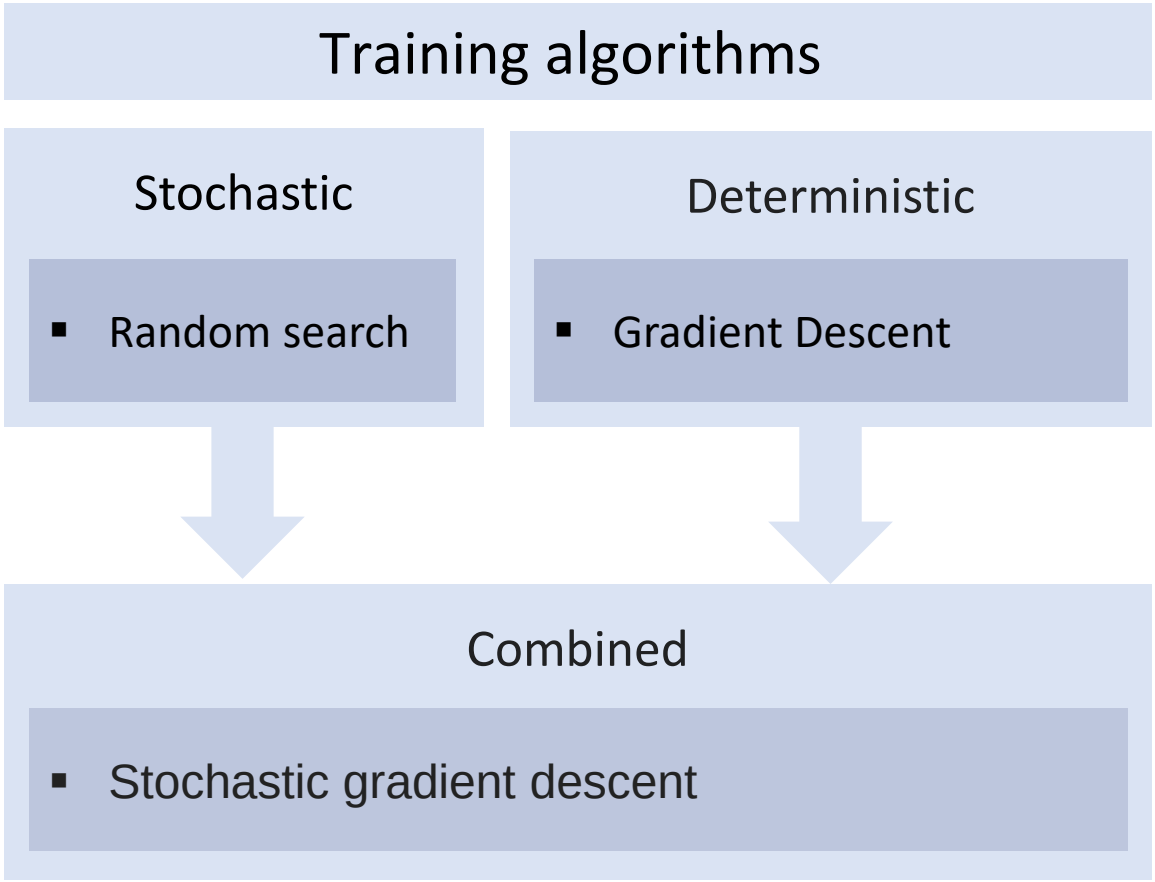
} The network structure determines the number of explicit parameters

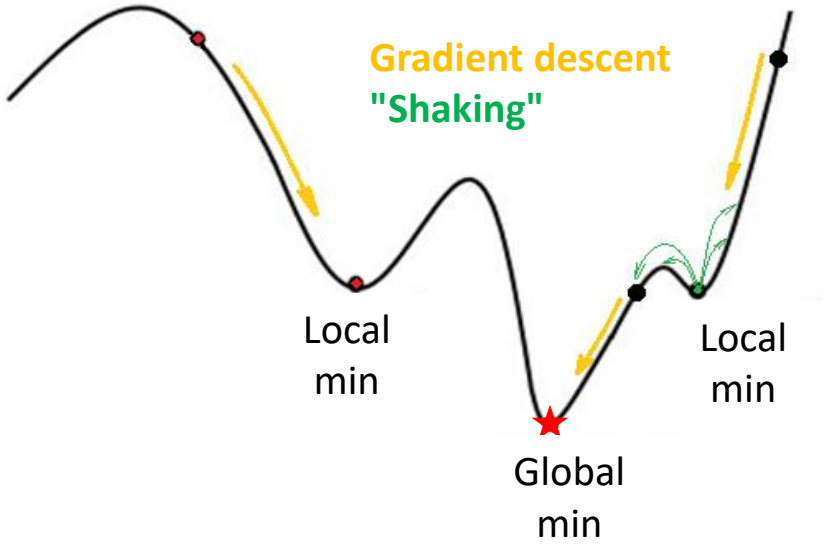
Sigmoid	Tanh	ReLU
$g(z) = \frac{1}{1+e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$
		

- Input variables
- Number of training iterations
- The size of the batch of parameters trained at a time
- Training algorithms: Adam, SGD, RMSprop, …

Neural network. Training algorithms

Training algorithms - algorithms for searching parameter values. The algorithm minimizes the measure of difference between the “true” value of the target variable and the value predicted by the neural network.





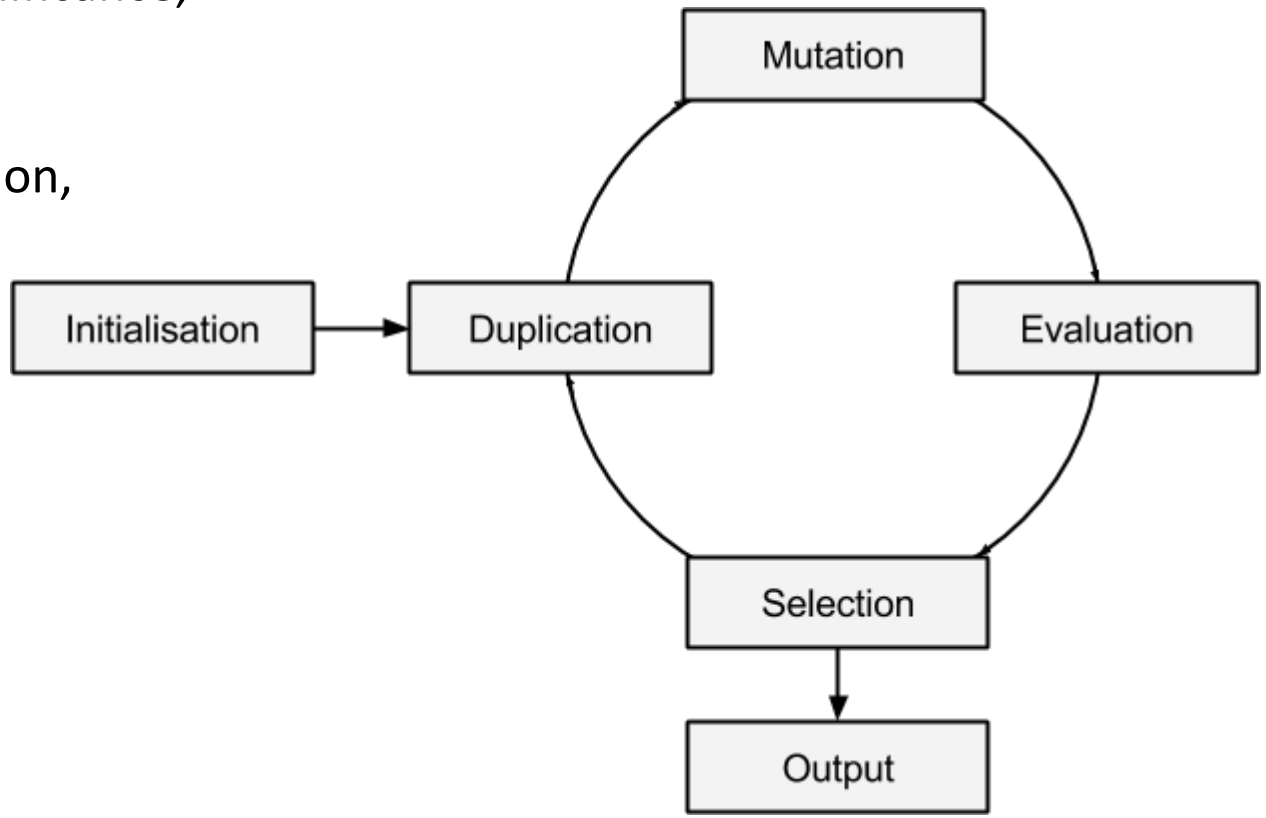
Really used algorithms try to combine the advantages of deterministic and stochastic methods. In most problems, a sufficiently **deep local minimum** is a satisfactory solution.

Neural network. Optimization methods. Evolutionary algorithms

Evolutionary algorithms are a subset of optimization algorithms that is inspired by natural selection:

- Single set of hyperparameters – individual,
- 30 individuals – population,
- Goodness of individual: ROC integral, signal significance,
- 10-50 generations,
- Each generation – selection, duplication, mutation,

  goodness evaluation



Neural network. Input variables

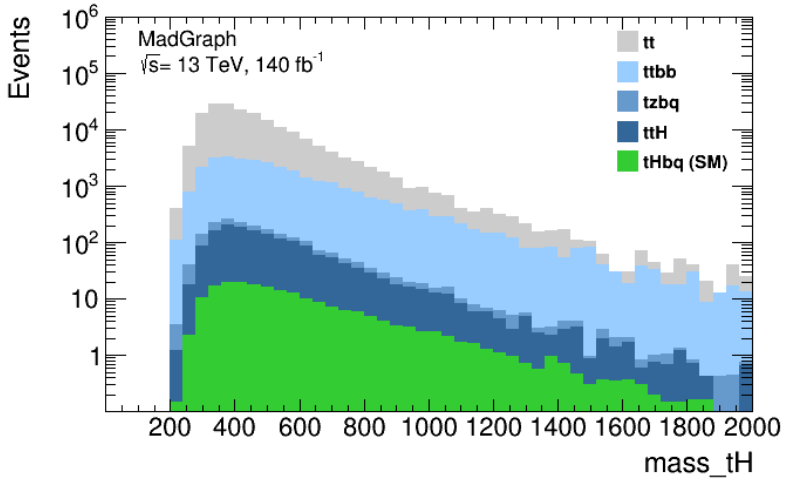
$$SP = \frac{1}{2}\left(\sum_{i=0}^{nbins} \frac{(s_i - b_i)^2}{s_i + b_i}\right) \times 100$$

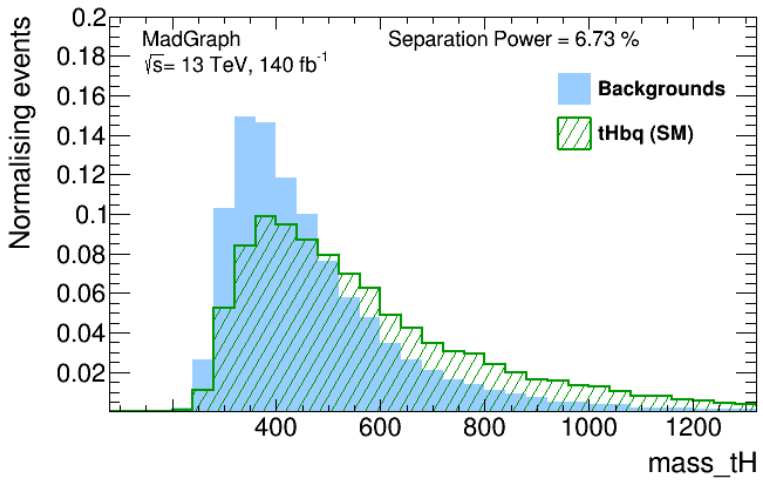
	Name	Separation (SM) [%]	for SM	for BSM	
1	lead_lep_charge	3.32			Charge of the leading lepton
2	N_b	10.71			Number of jets generated by b-quarks
3	n_nonb	2.79	excluded		Number of jets generated by quarks other than the b-quark
4	HT_alljets	1.34			Algebraic Sum of all transverse momenta
5	delta_eta_tH	8.41			Difference between the pseudorapidities of the top quark and the Higgs boson
6	sphresity_alljets	8.85			A measure of the uniformity of jet distribution in space
7	sphresity_lnu4maxje	8.83			A measure of the uniformity of jet distribution in space
8	sphresity_allobjects	8.62			A measure of the uniformity of jet distribution in space
9	aplanarity_allobjects	8.23	excluded		A measure of the deviation of jets from one common plane
10	higgs_m	5.94			Recovered mass of the Higgs boson
11	mass_tH	6.73			Invariant mass of the t-quark and Higgs boson
12	aplanarity_alljets	8.54	excluded	excluded	A measure of the deviation of jets from one common plane
13	aplanarity_lnu4maxjet	7.94	excluded	excluded	A measure of the deviation of jets from one common plane
14	delta_eta_FWD_t	7.35			Pseudo-rapidity difference between the t-quark and the front jet
15	min_chi	3.84			Quality of Higgs and Top Mass determinations
16	mass_H_CenJet	5.05			Invariant mass of the Higgs boson and the central light jet
17	mass_H_FWD	5.37			Invariant mass of the Higgs boson and the front jet
18	FWD_pt	5.37			Transverse momentum of a quark scattered forward
19	fwmlnujet1	5.51	excluded		First Fox-Wolfram moment composed of jets, lepton and neutrino
20	FWD_eta	4.95	excluded	excluded	Pseudofastness of a forward scattered quark
21	fwm1	4.67			First Fox-Wolfram moment composed of jets only
22	top_m	3.52		excluded	Recovered t-quark mass
23	fwm2	2.50	excluded		Second Fox-Wolfram moment composed only of jets
24	DeltaR_qqW	2.89		excluded	Angle between jets of hadronic decay of w boson
25	RapGap_maxptb	1.93			Difference between the pseudo-velocities of the front jet and the b-jet with the highest pt
26	RapGap_closest	1.79			The difference between the pseudo-velocities of the front jet and the b-jet closest to it
27	Central_non_b_maxpt_pt	1.68			Highest transverse impulse among light jets
28	FWD_m	1.63			Invariant mass of the front jet and t quark
29	lead_lep_eta	2.04	excluded	excluded	Pseudofastness of the leading lepton
30	jet_b2_e	1.48		excluded	Energy of the b-jet second in transverse momentum
31	W_T_m	1.63			Transverse Mass of all jets
32	InvMass_3Jets	12.24		excluded	Invariant mass of three jets

Neural network. Input variables. Transformation

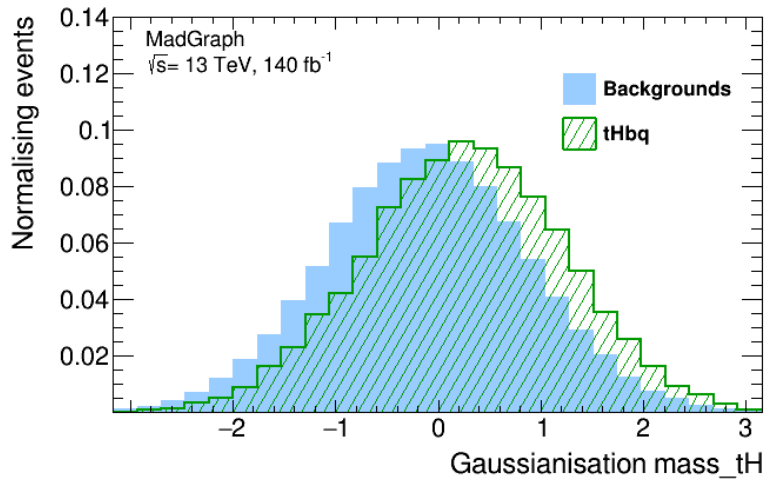
$$SP = \frac{1}{2}\left(\sum_{i=0}^{nbins} \frac{(s_i - b_i)^2}{s_i + b_i}\right) \times 100$$

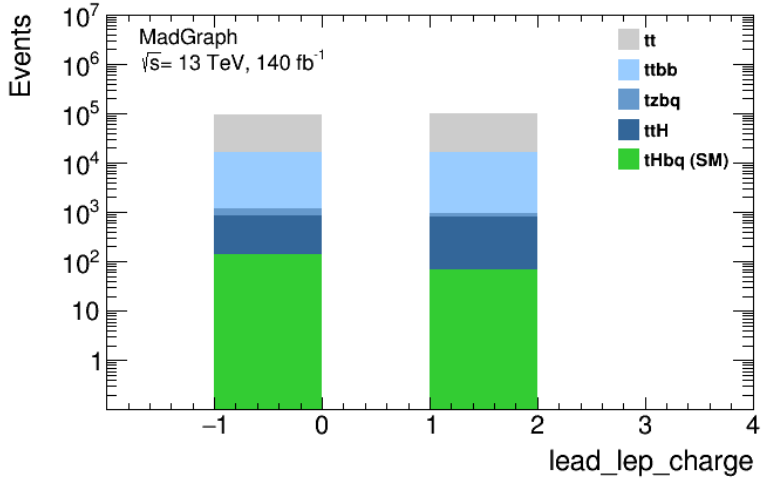
Kinematic variables must be brought to a comparable domains (before training)

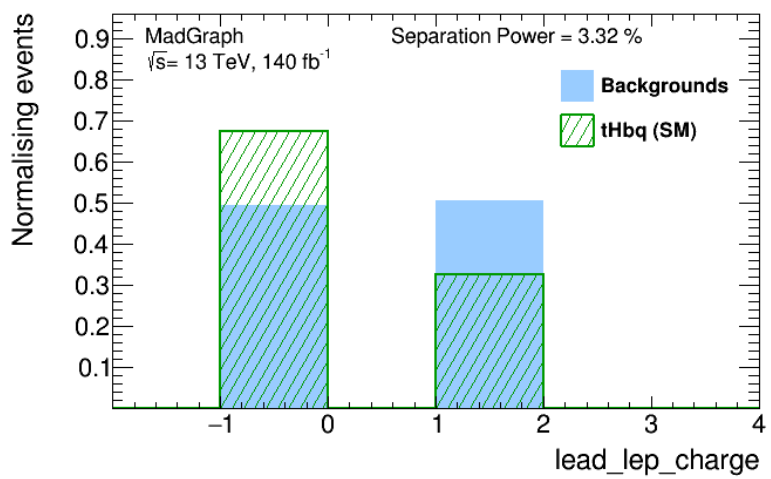




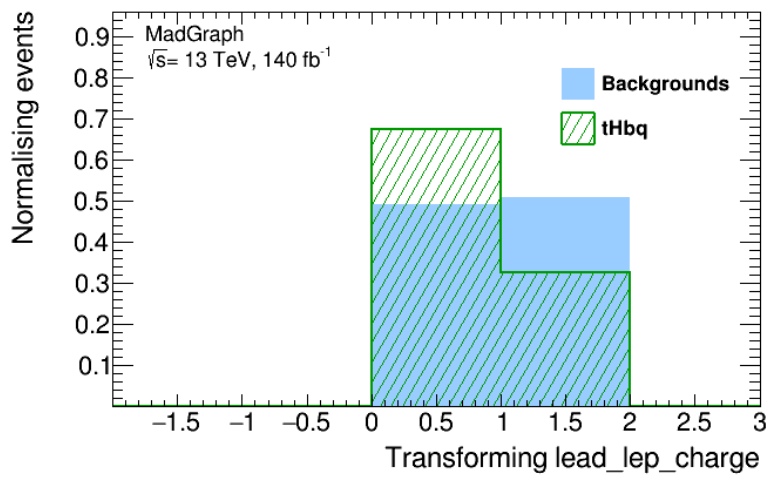
Gaussization for input variables of float type





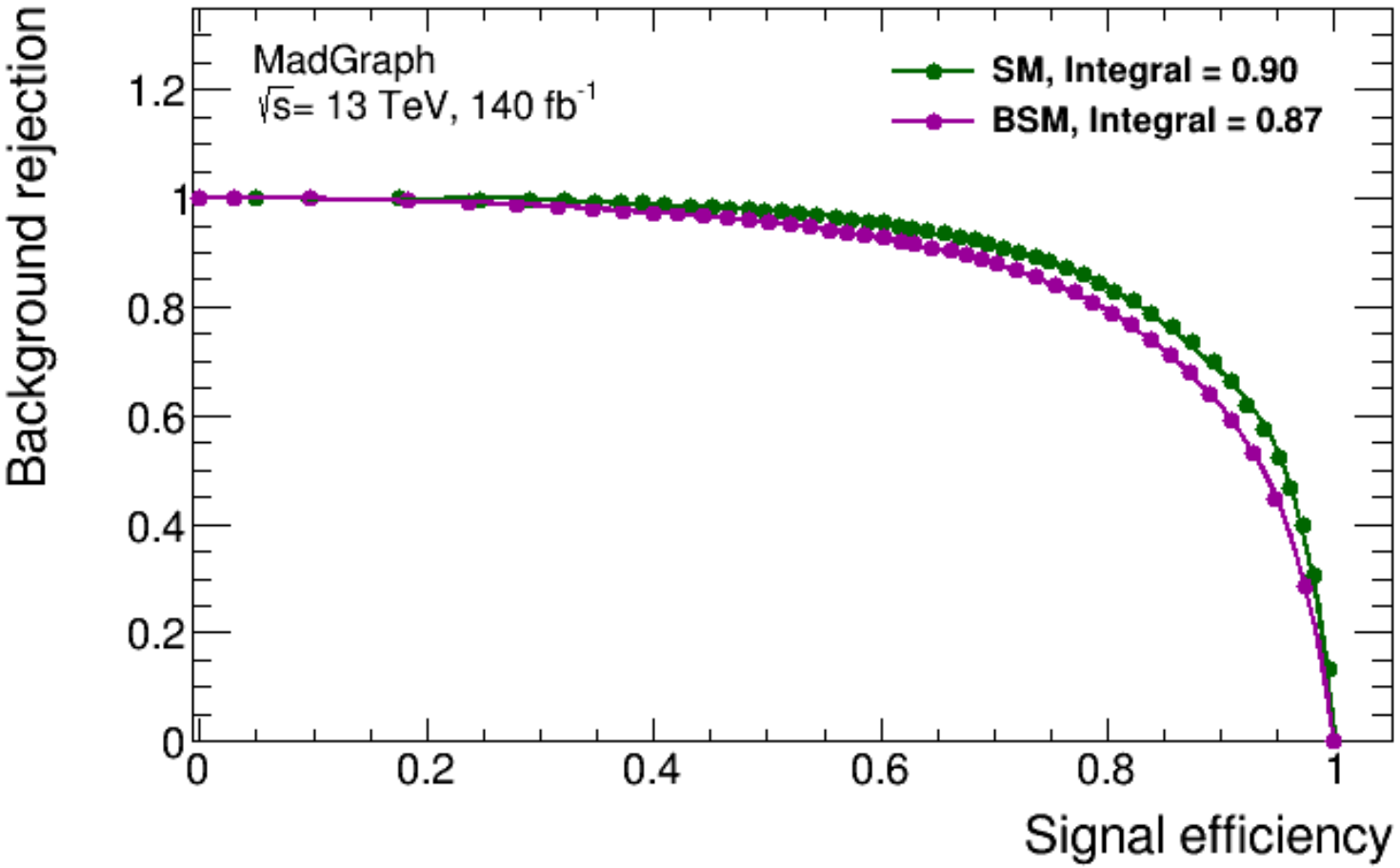


Normalization for input variables of int type



Results. Optimized NN structure and ROC-curve integral

Set of NN hyperparameters optimized by evolutionary algorithms

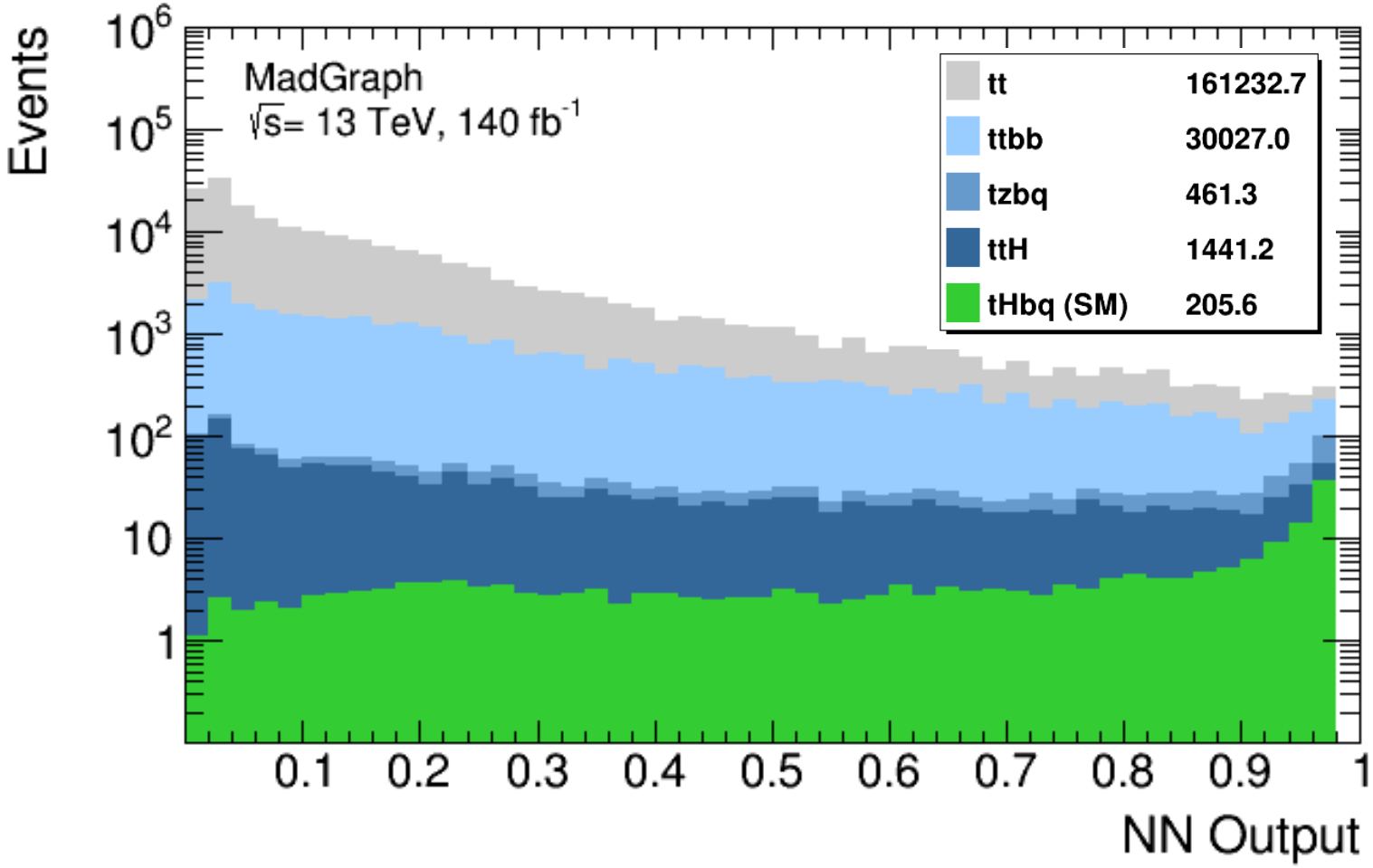


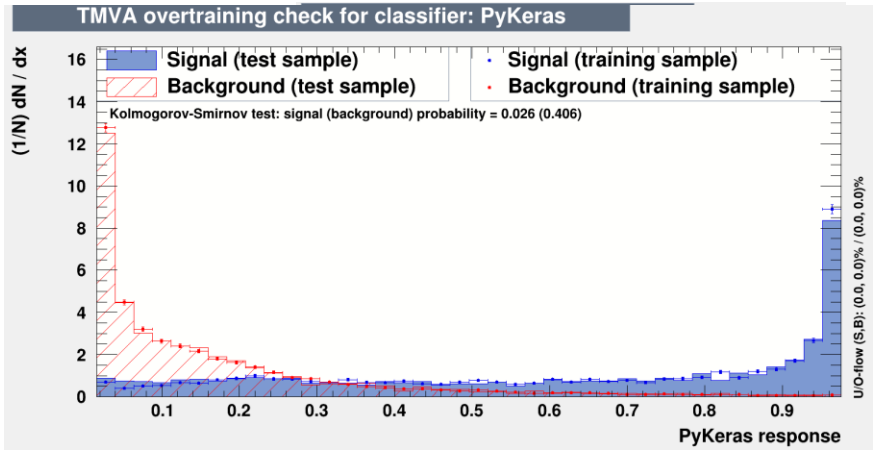
	SM signal	BSM signal
Generation	16	10
Input Layer	24 vars	24 vars
Layer1	576 (relu)	608 (relu)
Layer2	1440 (relu)	832 (softmax)
Layer3	48 (softmax)	32 (exponential)
Layer4	80 (relu)	32 (softmax)
Total variable parameters	918 530	549 666
Batch size	101	159
Optimizer	Adam	RMSprop

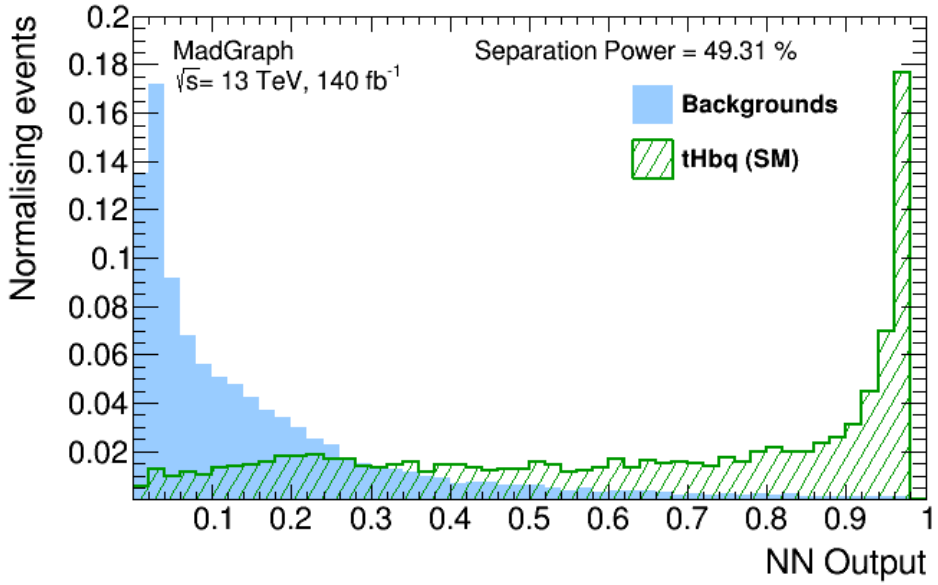
The network efficiency curve is the dependence of the cut signal on the number of background events captured.

Results. NN output and Separation power for SM tH signal

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**.**

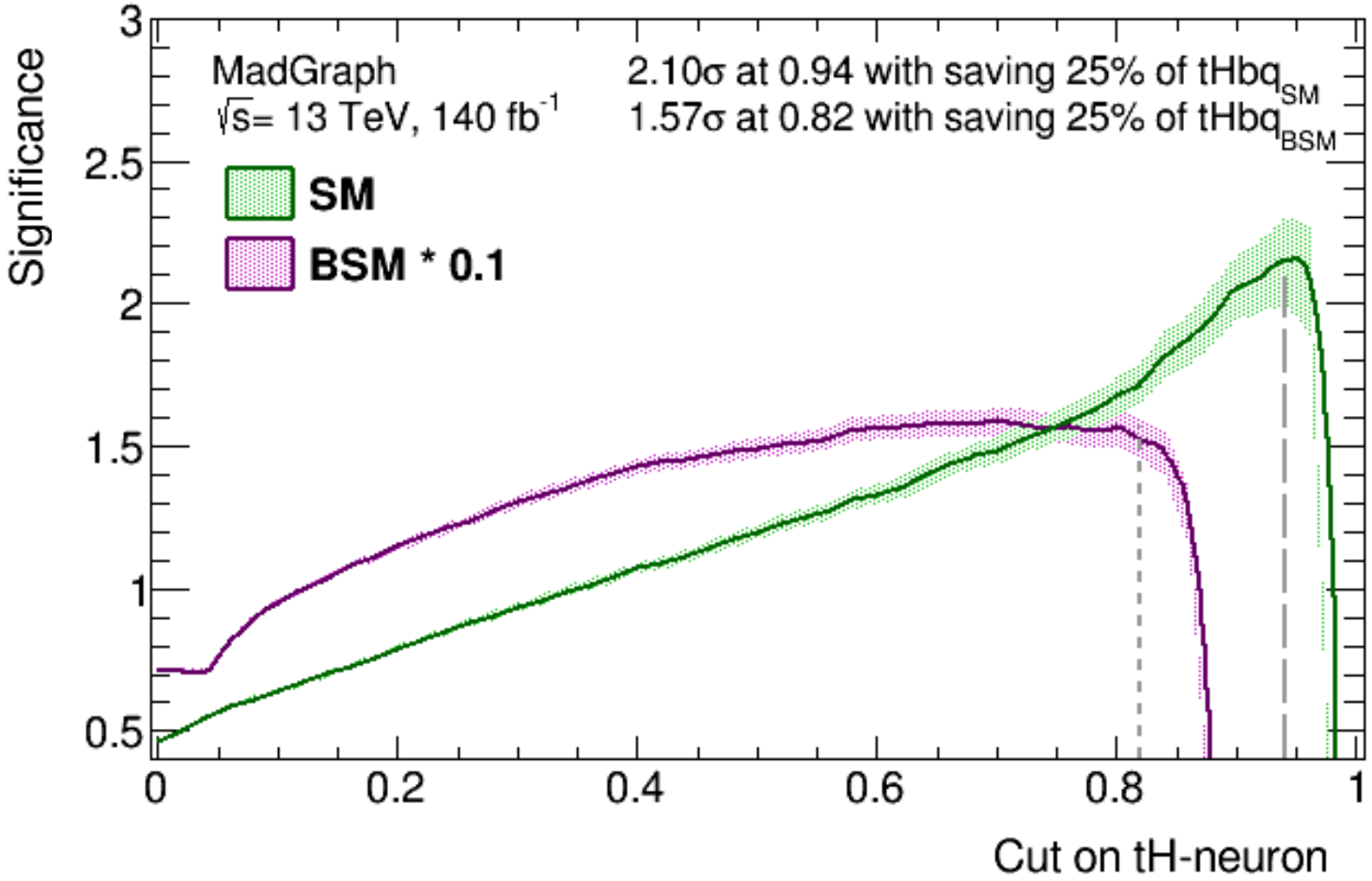






Results. Significance

The maximum significance of the SM signal is achieved with cut on tH-neuron > **0.94**
Applied requirement to save at least 25% of the signal



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

For pp collisions at 13 TeV :
140 fb^{-1}

- Without the NN, the significance of the SM signal is **0.47** σ
- After applying the NN: **2.1** σ

Thus, the NN allows to increase the significance of the SM signal by **4.47** times.

- Without the NN, the significance of the BSM signal is **7** σ
- After applying the NN: **15** σ

Conclusion

Results

- A new evolutionary algorithm has been created that allows optimization in an irregular space of hyperparameters;
- Application of evolutionary algorithm made it possible to significantly improve the significance of the tH signal;
- Standard Model signal significance increased from 0.47 to 2.1 σ
- Signal significance of ITC BSM increased from 7 to 15 σ
- According to RUN2, the ITC signal is expected to be observed or eliminated

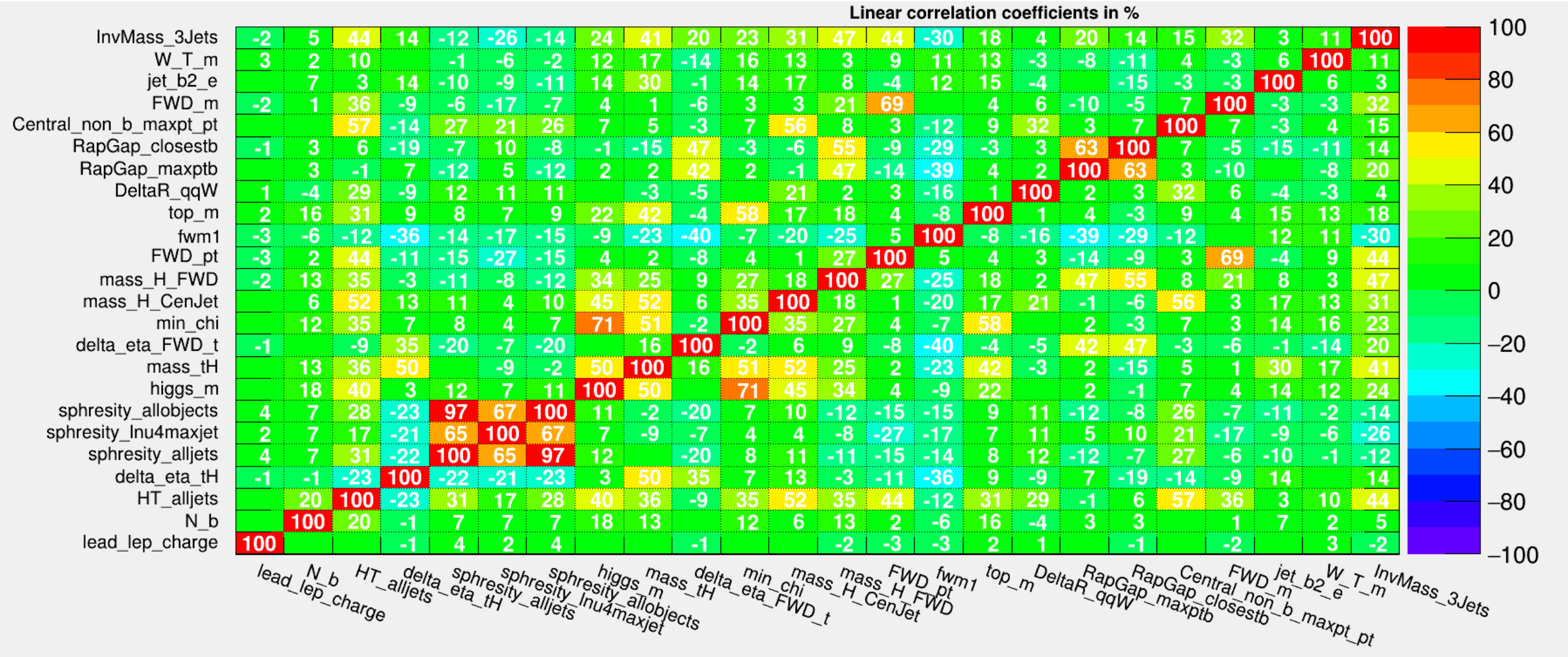
Future plans

- Implement diagonalization of input variables, which will eliminate their correlation
- “Global optimization” of the entire procedure, allowing simultaneous optimization of both parameters and hyperparameters of the neural network

Thank you for your attention

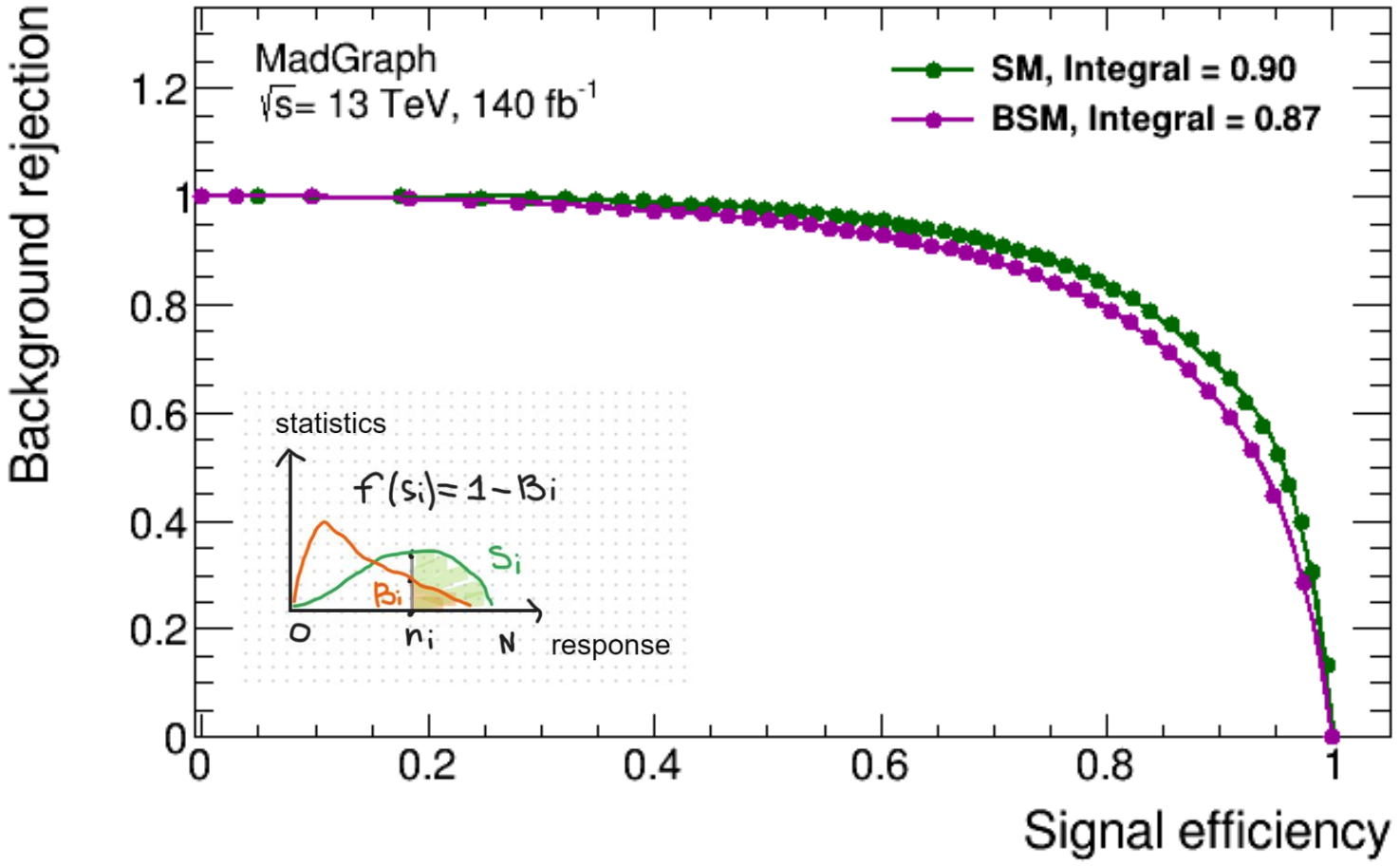
Back-up. Correlation matrix

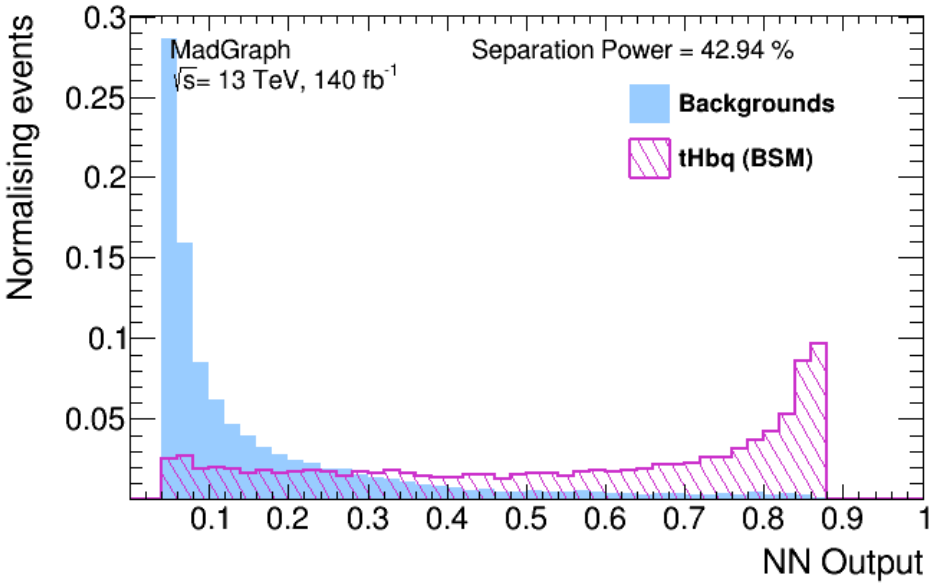
24 Variables Most Sensitive to $tHbq_{SM}$

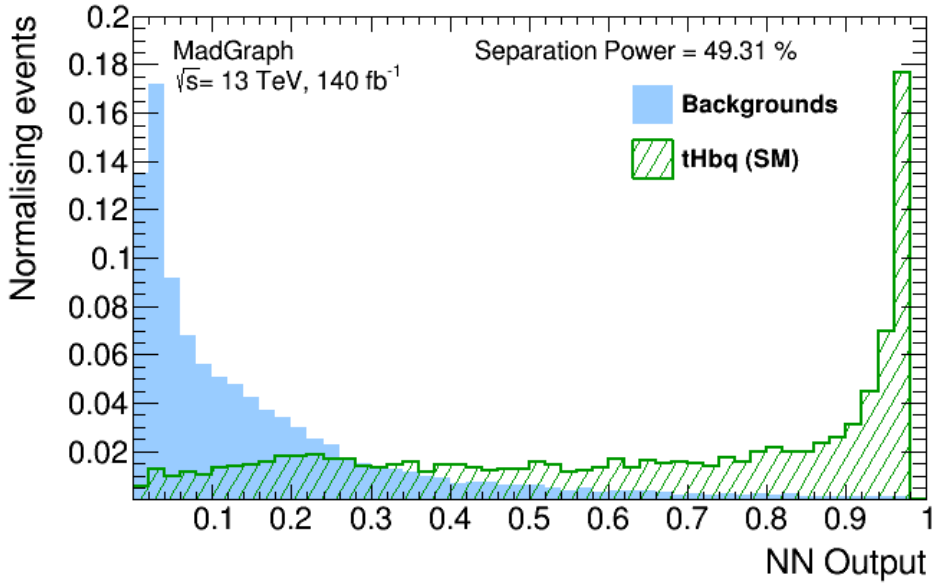


Back-up. ROC-curve integral

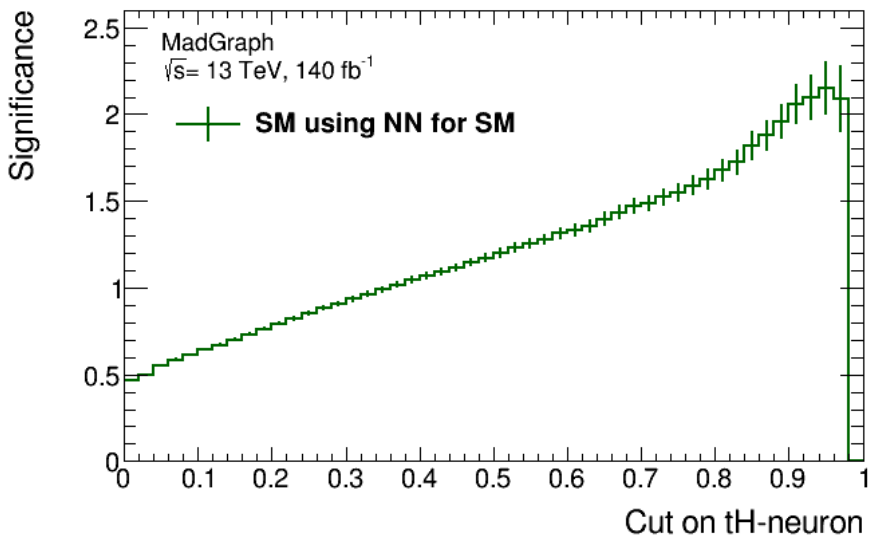
The network efficiency curve is the dependence of the cut signal on the number of background events captured.

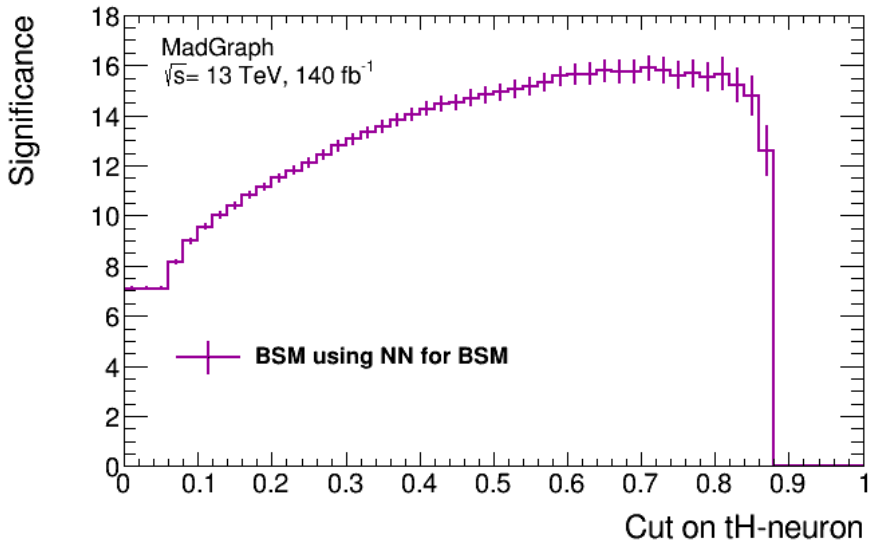


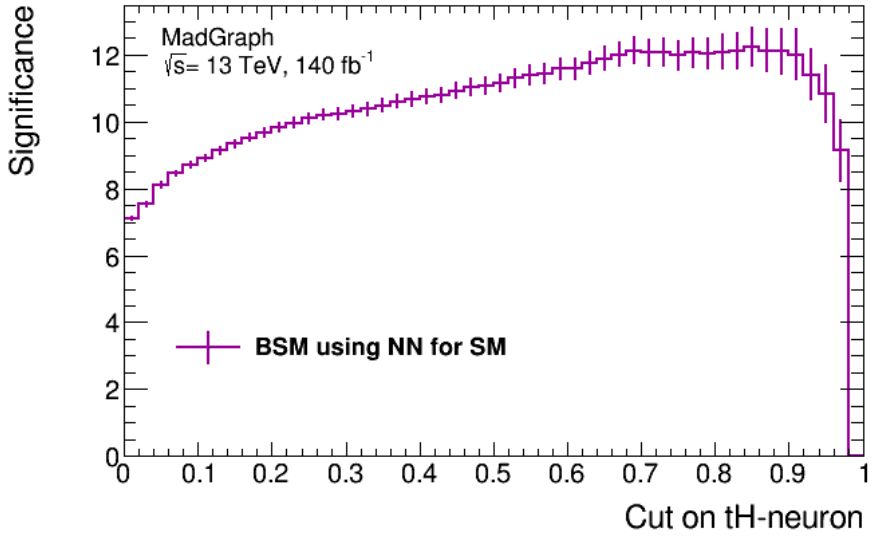


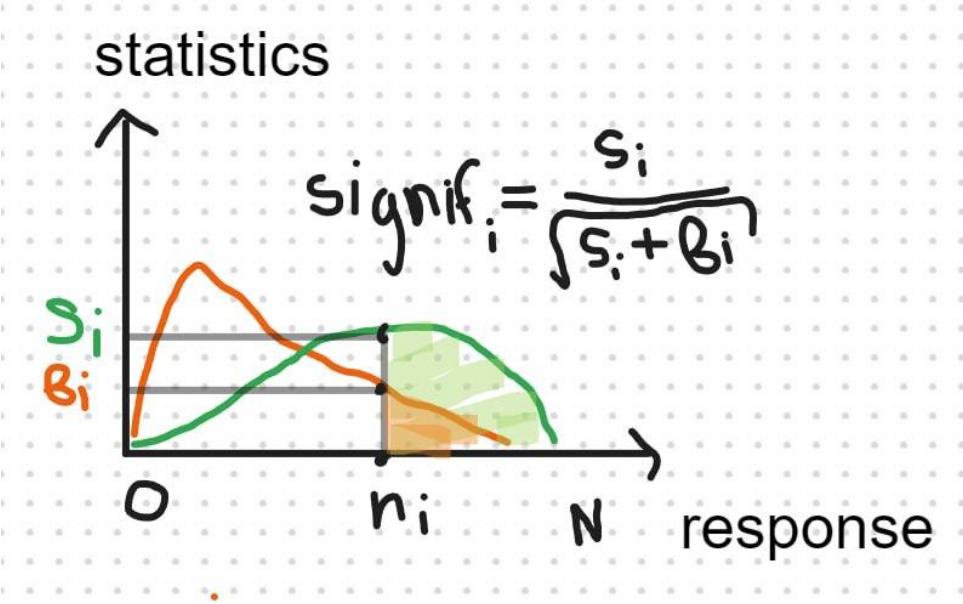


Back-up. Significances









Back-up. Network optimization using evolutionary algorithm