XXXI Annual International Seminar Nonlinear Phenomena in Complex Systems

Automatic optimization of neural network parameters for data analysis of the Large Hadron Collider

V.Kiseeva

JINR team: I.Boyko, A.Didenko, O.Dolovova, N.Huseynov, I.Yeletskikh

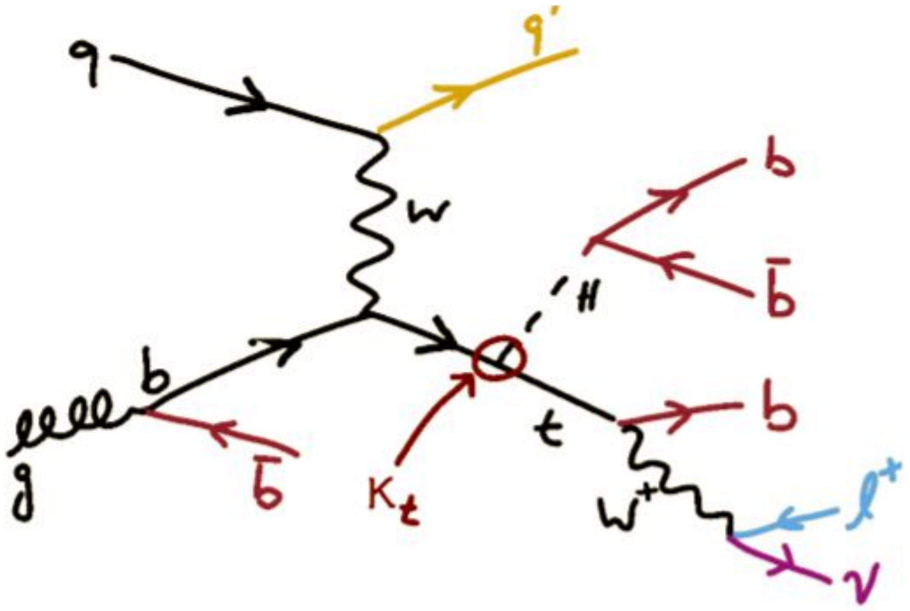
26.06.2024

Outline

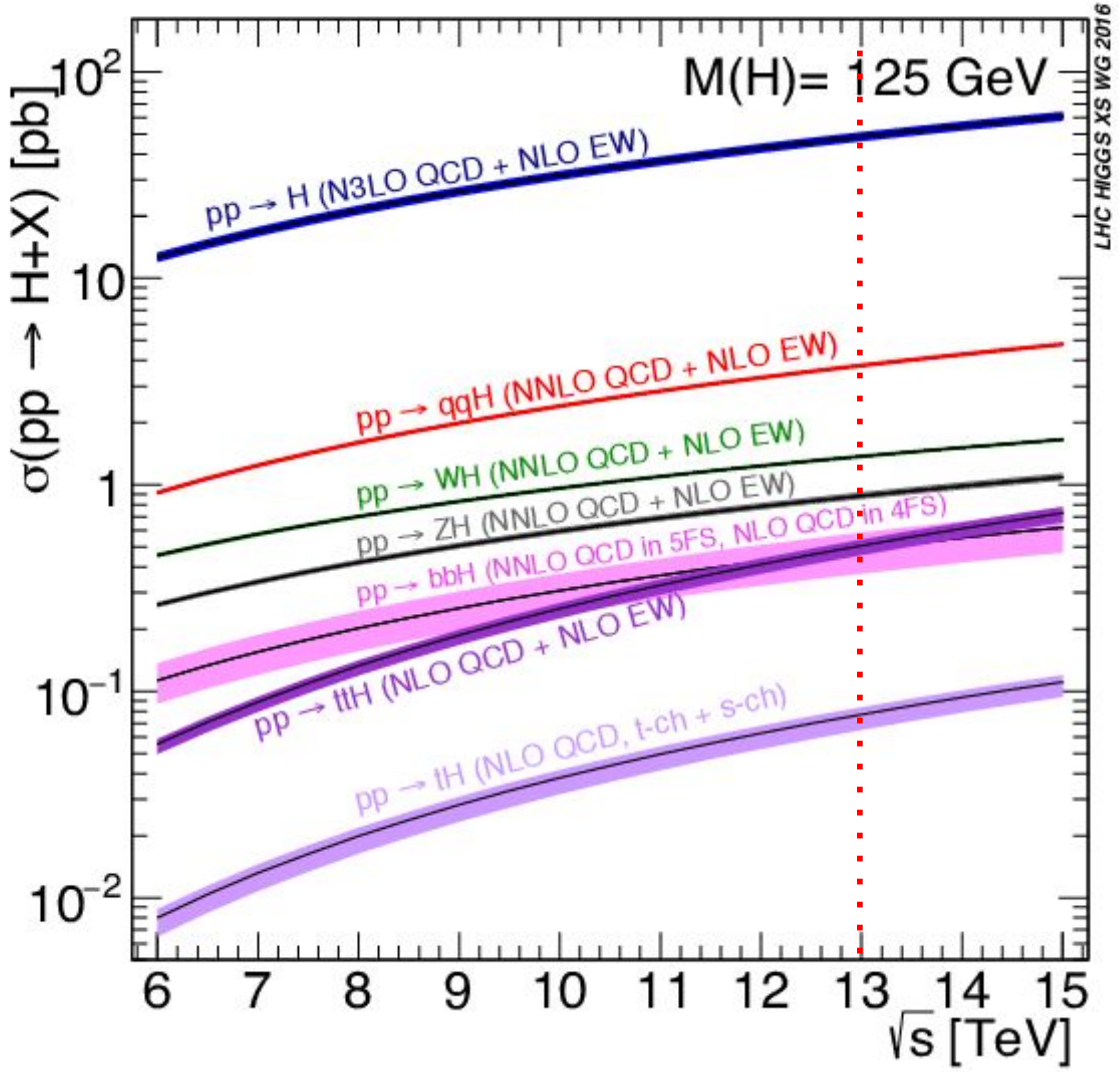
Motivation: tH signal

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

for LHC Run2 expected: ~100 tH; ~200k total background

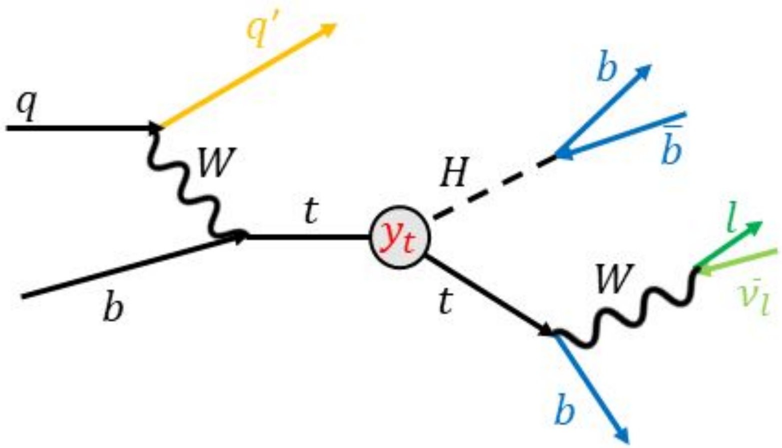


Signature of tH with decay $H \rightarrow b\bar{b}$:
(≥3 b-jets) + (1 light jet) + (1 tight lepton) + (missing transverse momentum)

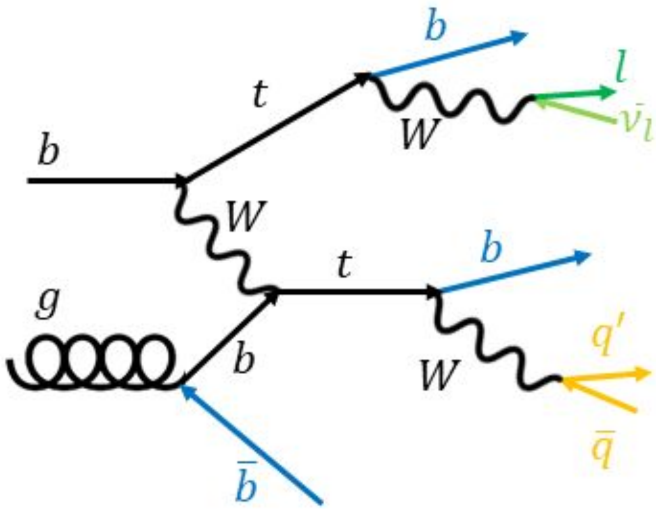


[1] Higgs Physics, C. Grojean [arXiv:1708.00794]

- The kinematics of signal and background processes are very similar

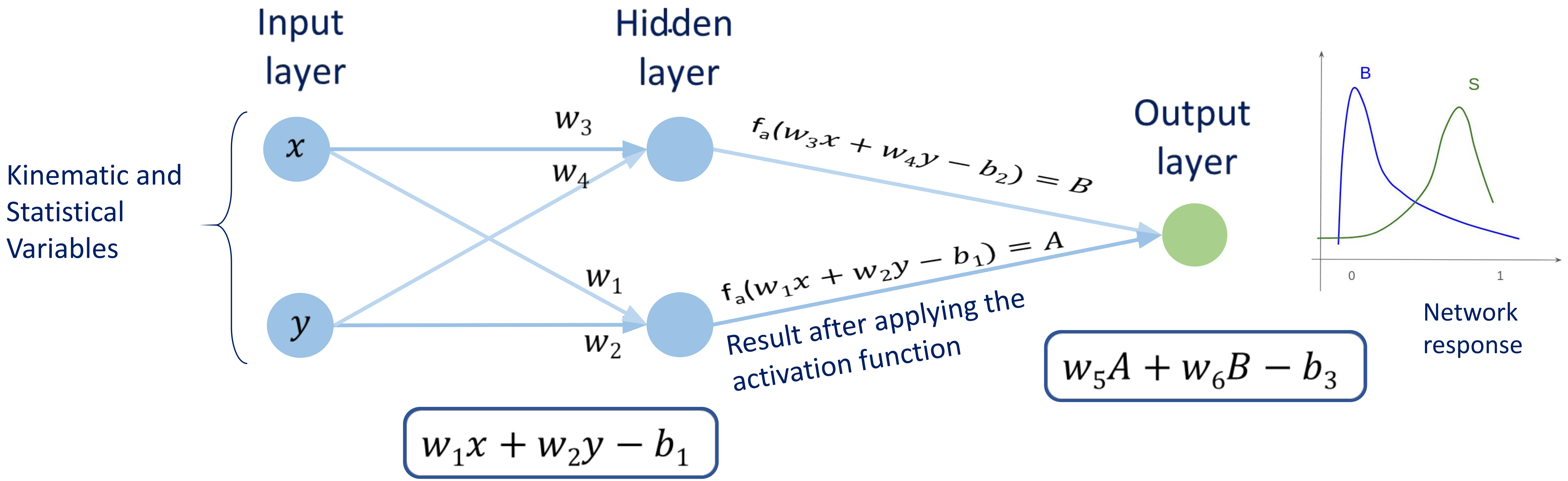


signal: tH process



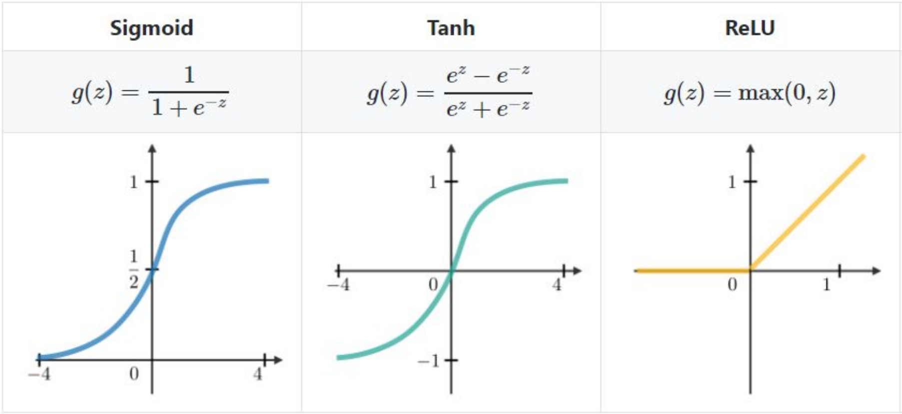
main background: ttb production channel

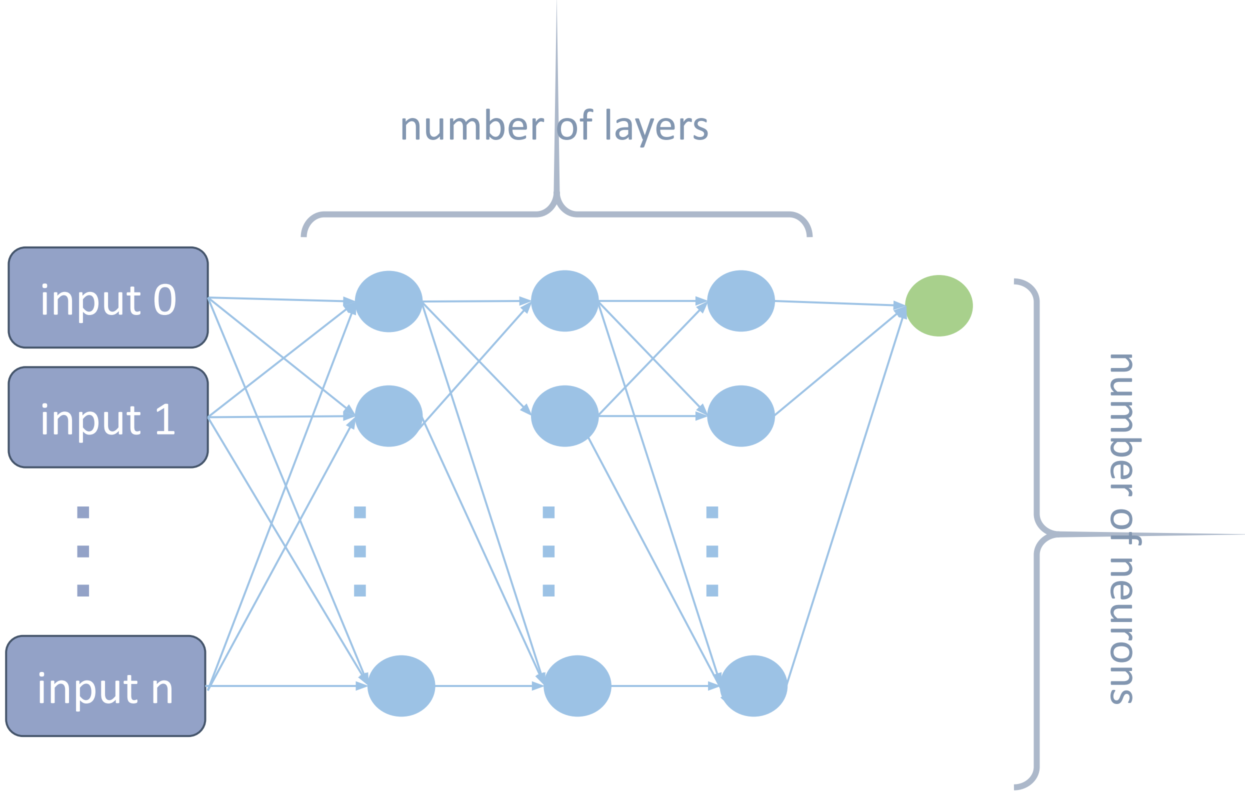
A **neural network** is a function with large number of internal parameters.
Internal parameter: w1…w6 and b1…b6



Neural network. Parameters neural network structure (hyperparameters)

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





- 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.

Training algorithms

Stochastic

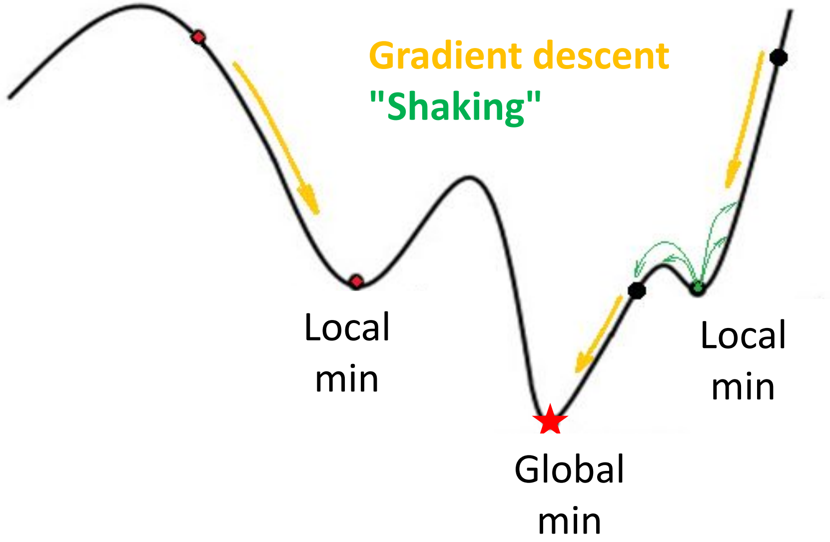
- Random search

Deterministic

- Gradient Descent

Combined

- Stochastic gradient descent

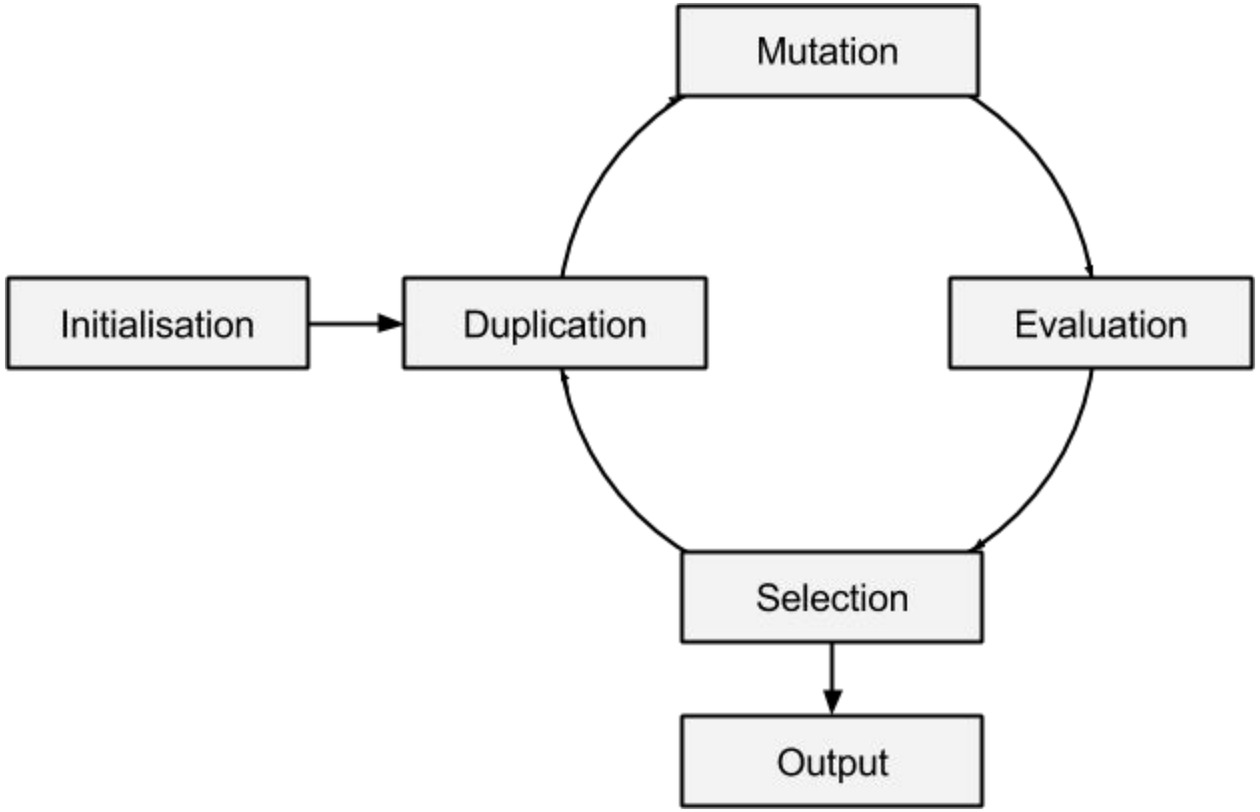


Practically used algorithms try to combine the advantages of deterministic and stochastic methods. In most problems, a sufficiently **deep local minimum** is a satisfactory solution.

Neural network. Optimization of hyperparameters. 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

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

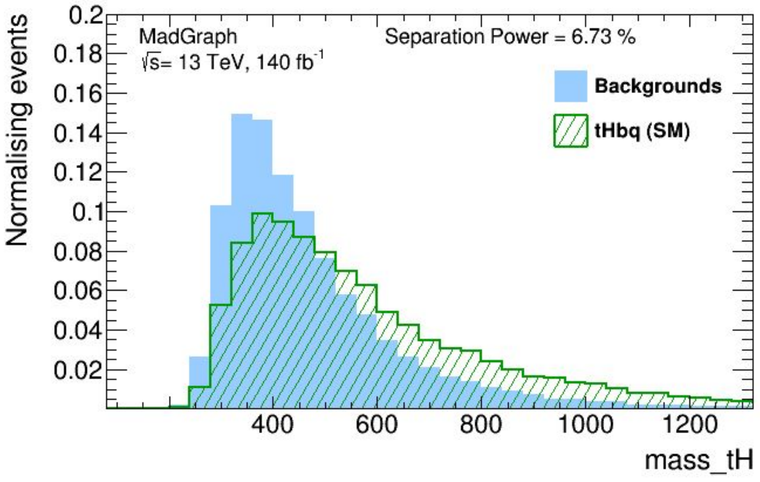
	Name	Separation (SM) [%]	for SM	for BSM	
1	lead_lep_charge	3.32			Charge of the leading lepton
2	N_b	10.71			Number of jets generated by b-quarks
3	n_nonb	2.79	excluded		Number of jets generated by quarks other than the b-quark
4	HT_alljets	1.34			Algebraic Sum of all transverse momenta
5	delta_eta_tH	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

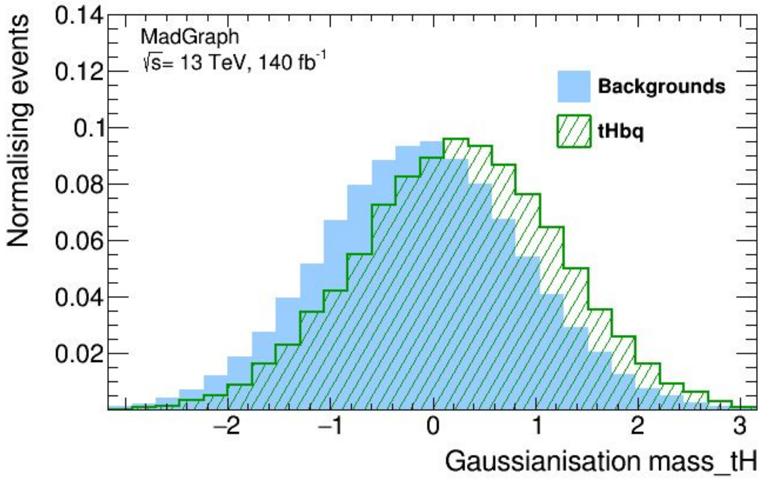
$$SP = \frac{1}{2}\left(\sum_{i=0}^{nbins}\frac{(s_i-b_i)^2}{s_i+b_i}\right)\times 100$$

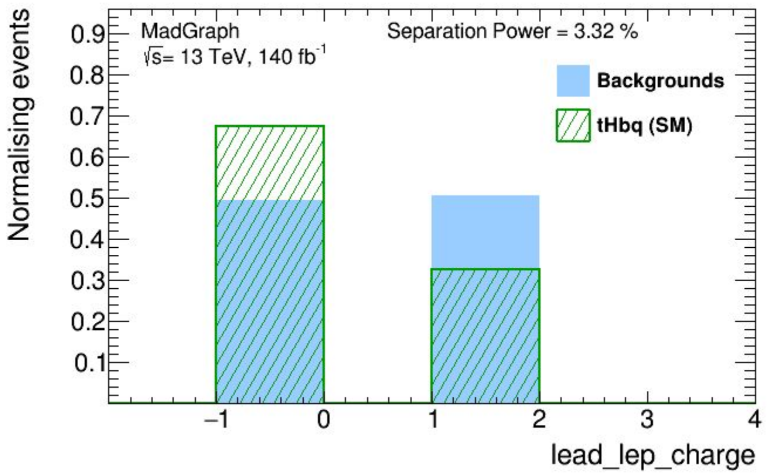
Variables must be brought to a comparable domains (before training)

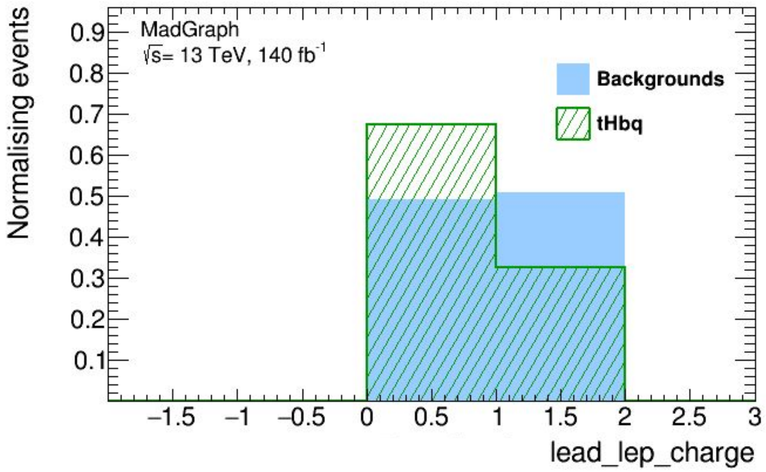
BEFORE

AFTER



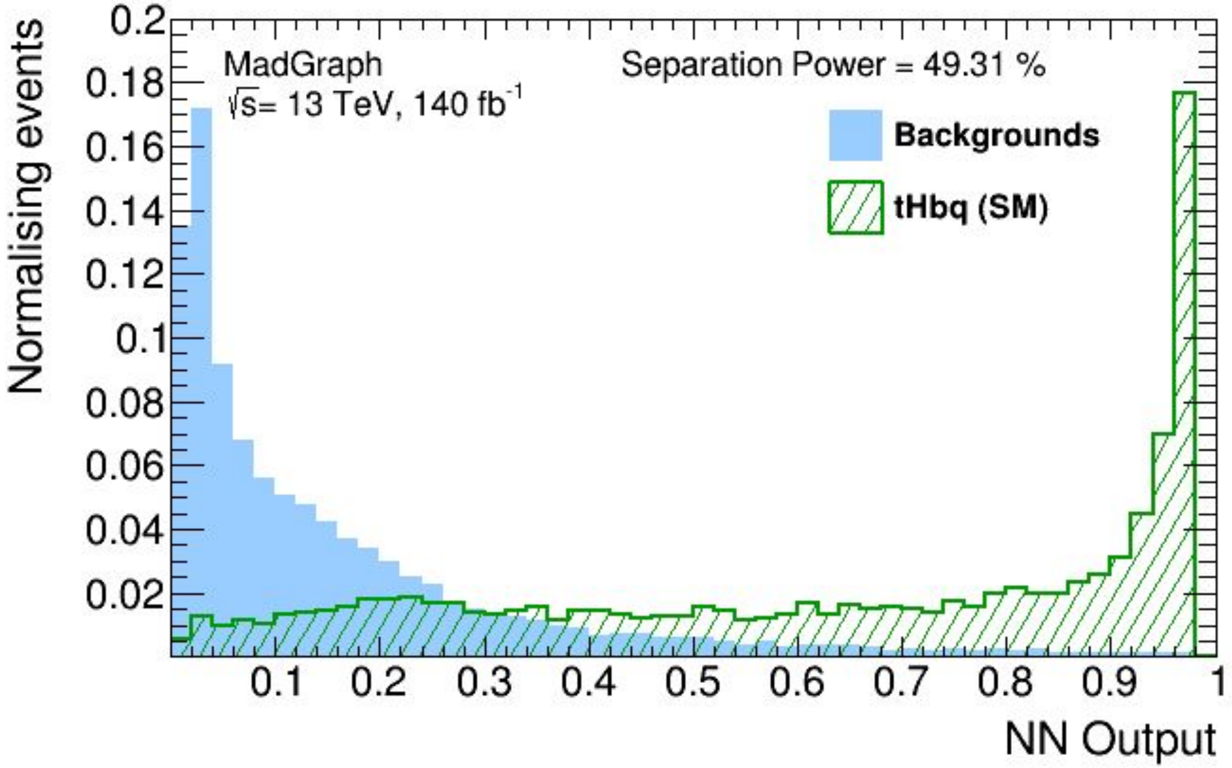






Results for SM tH signal

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**.**

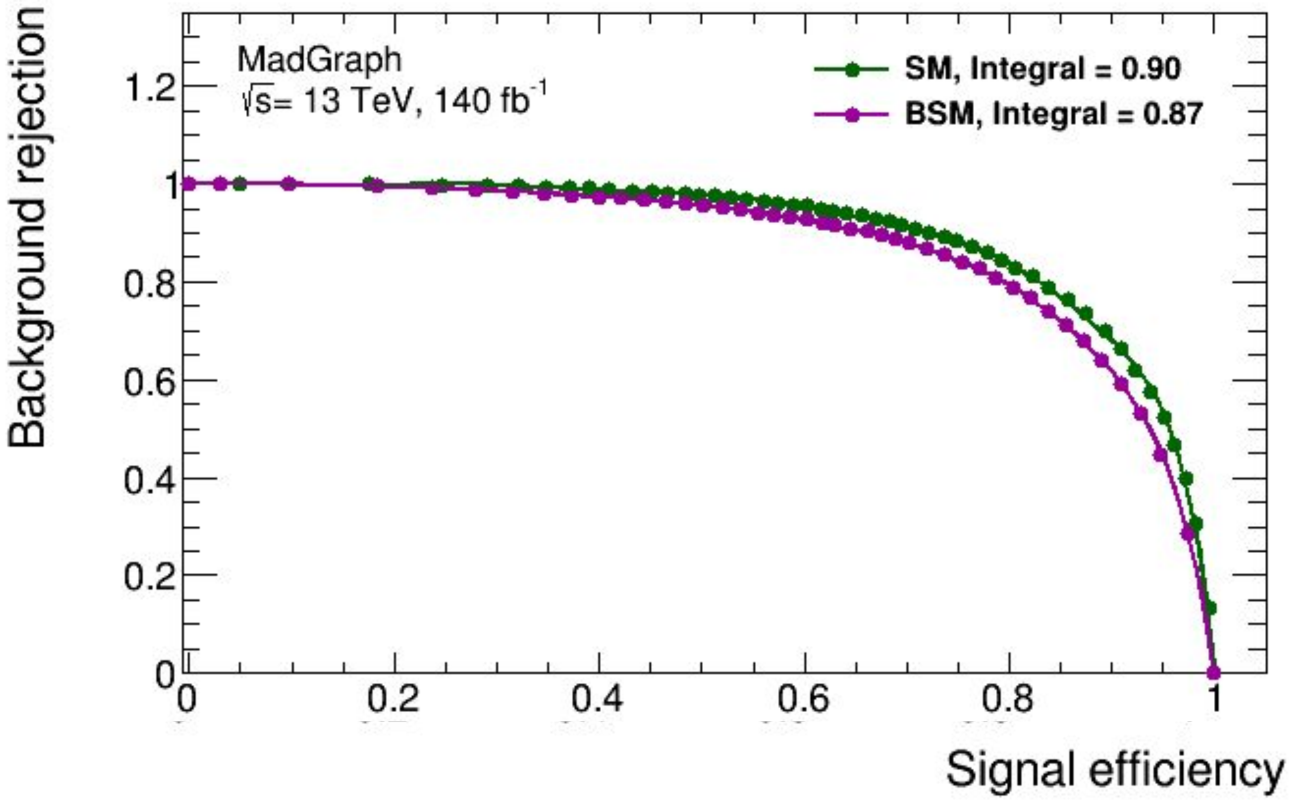


The maximum significance 2.1σ of the SM signal is achieved with cut on tH-neuron > 0.94 Applied requirement to save at least 25% of the signal. For full Run-2 luminosity 140 fb-1

The optimal ROC-curve integral for SM **0.90**

The optimal Separation Power 49.31%

Results for Evolutionary algorithm. Optimized NN structure and ROC-curve integral



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

[Boiko, I.R., Guseinov, N.A., Eletskikh, I.V. et al. Phys. Part. Nuclei Lett. 21, 481–488 (2024).]

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

Further optimization of implemented method

Strategies that can possibly to contribute to the robustness and effectiveness of neural network method:

1. Integrating Evolutionary Algorithms in Neural Network Training
   - Can help to efficiently explore and exploit the solution space
2. Implementing Algorithms with Arbitrary Optimization Functions
   - The goal is to allow user to choose any function for optimization
   - Cross-entropy is now optimized during network training.
   - ROC-AUC is an example of optimization metric with clear physical meaning.

Cross-entropy

$$H(P^* | P) = -\sum_i \underbrace{P^*(i)}_{\text{TRUE CLASS DISTRIBUTION}} \log \underbrace{P(i)}_{\text{PREDICTED CLASS DISTRIBUTION}}$$

ROC-AUC

common sense

AUC = ∫TPRd(FPR)

physical sense

AUC = ∫Background_Rejection d(Signal_Efficiency)

Further optimization of implemented method

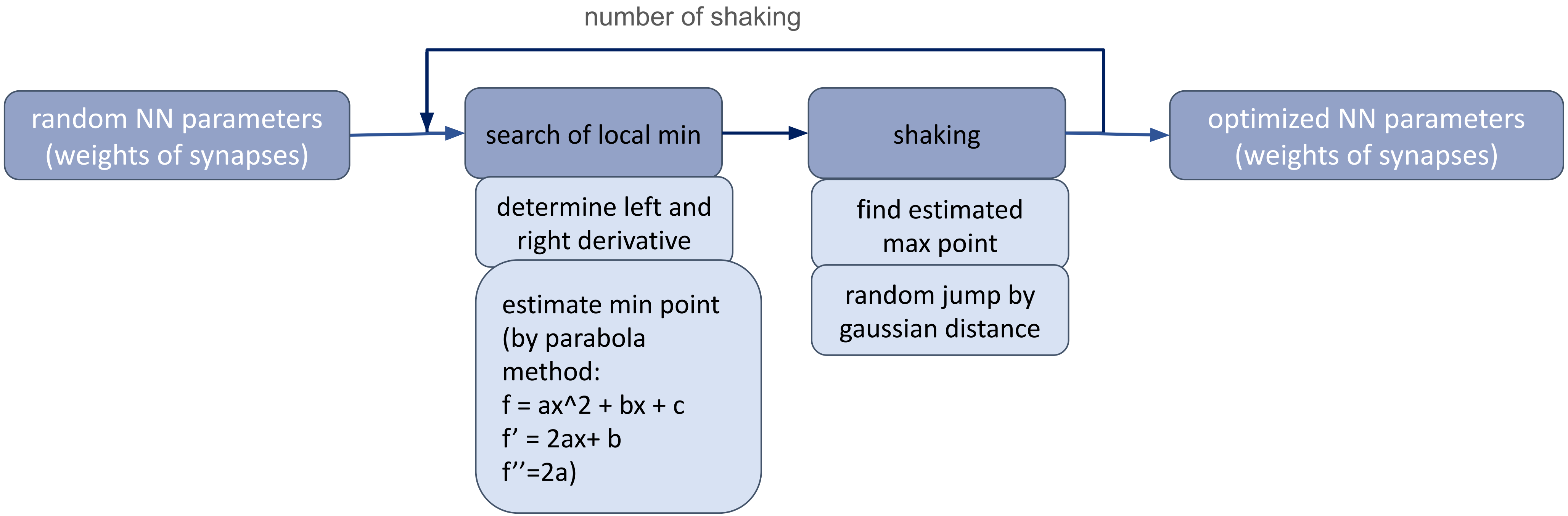
Steps to Improve Neural Network Robustness and Effectiveness

1. Develop a Custom Library (without using of Keras, TensorFlow etc) (done)
   - Create a specialized library to build neural network structures from the ground up.
2. Create a New Training Algorithm (currently in progress)
   - Design an algorithm for training neural networks.
3. Make the Algorithm Optimization-Ready (currently in progress)
   - Ensure the algorithm easily supports optimizing neural network parameters/hyperparameters and allows for any user-defined optimization function.
4. Start with Parameter Optimization
   - Begin by optimizing the neural network parameters (currently in progress).
5. Expand to Hyperparameter Optimization
   - Once performance is good, include the optimization of hyperparameters.

Status of work in progress

Library was implemented through making Network, Layer, Synapse and Neuron classes.

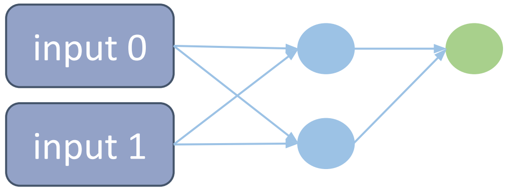
Algorithm was implemented as follows:



Status: test of optimization NN parameters

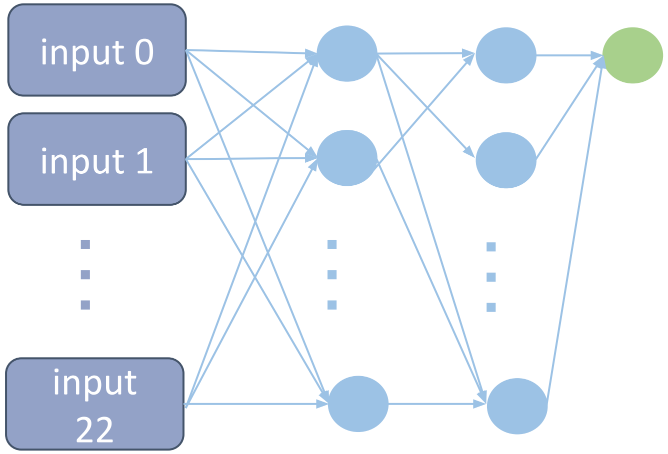
Network test configuration n.1:

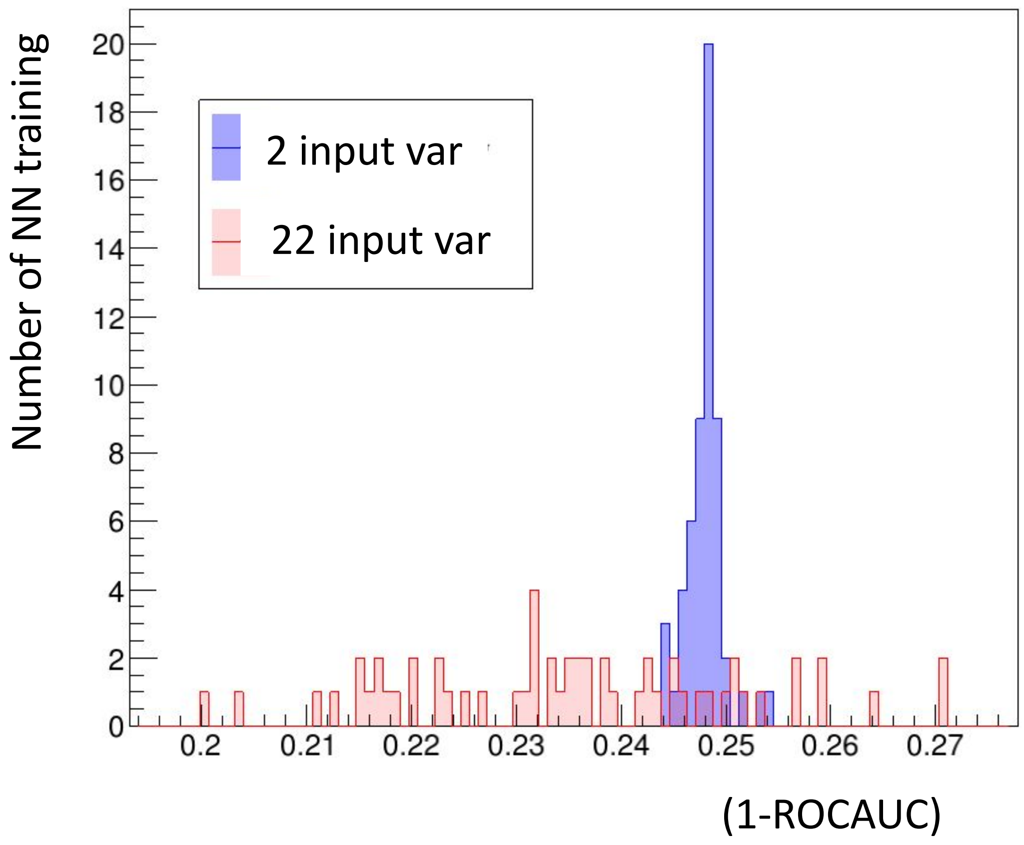
Input variables	2
Hidden Layers	1
Neurons per layer	2
Activation function	Tanh
Optimization Function	1-ROCAUC
Number of “shaking”	2



Network test configuration n.2:

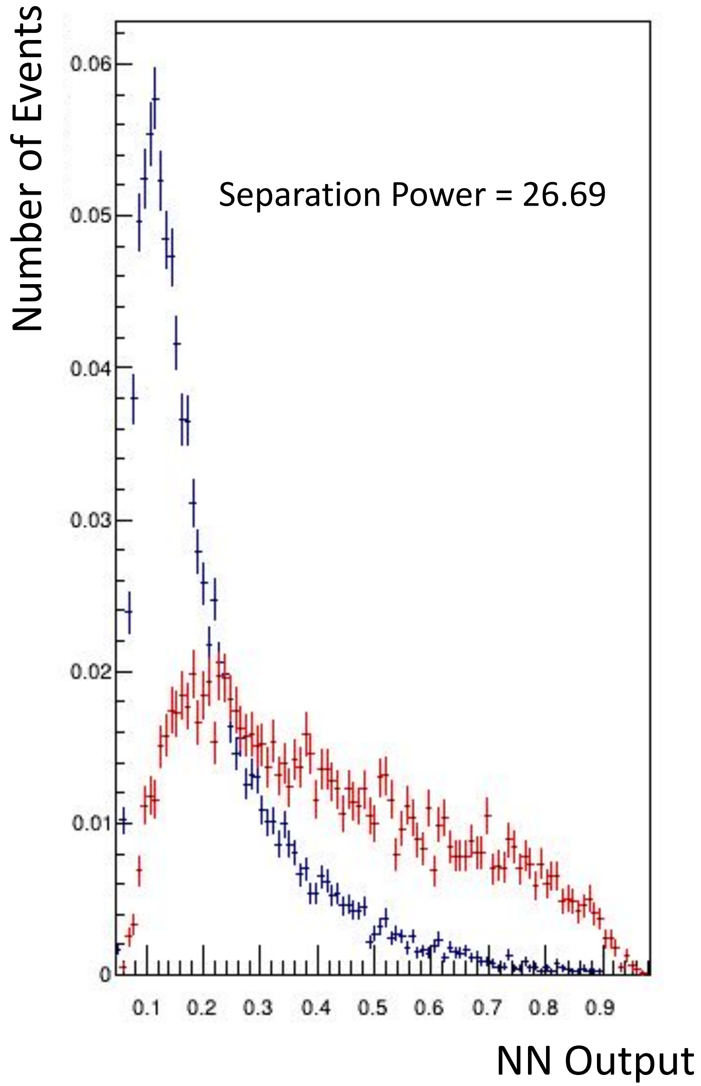
Input variables	22
Hidden Layers	2
Neurons per layer	30:30
Activation function	Tanh
Optimization Function	1-ROCAUC
Number of “shaking”	3



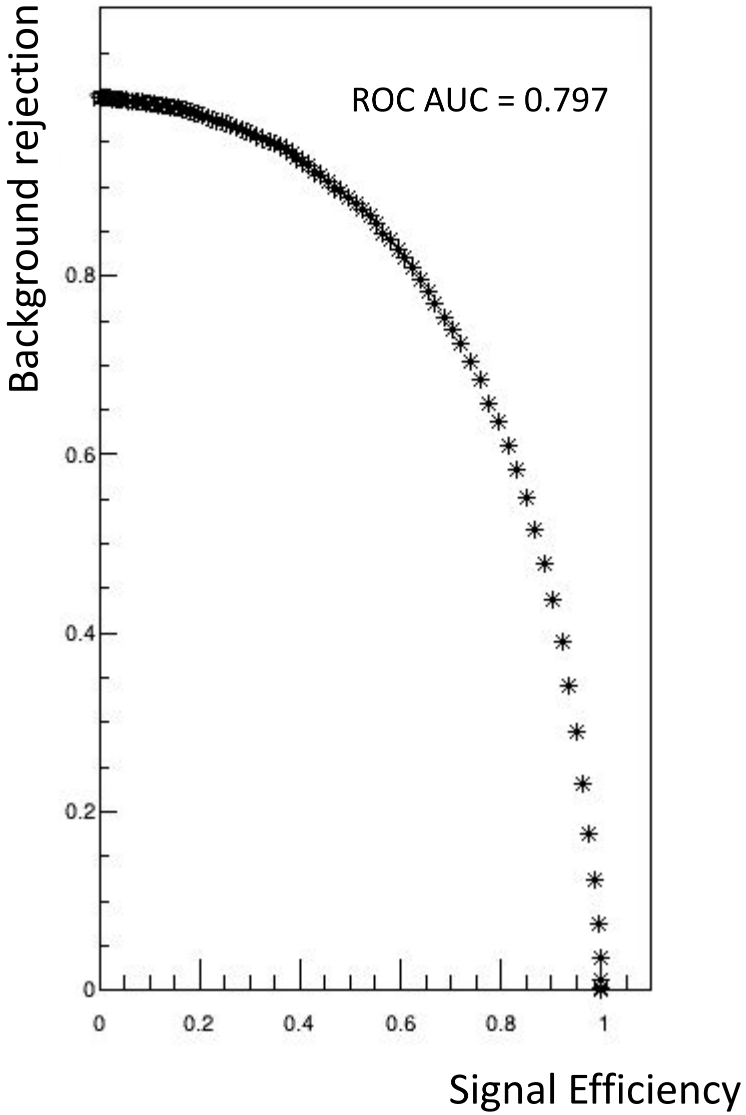


Status: test of optimization NN parameters

Signal vs Background comparison



ROC Curve



Conclusion

Results

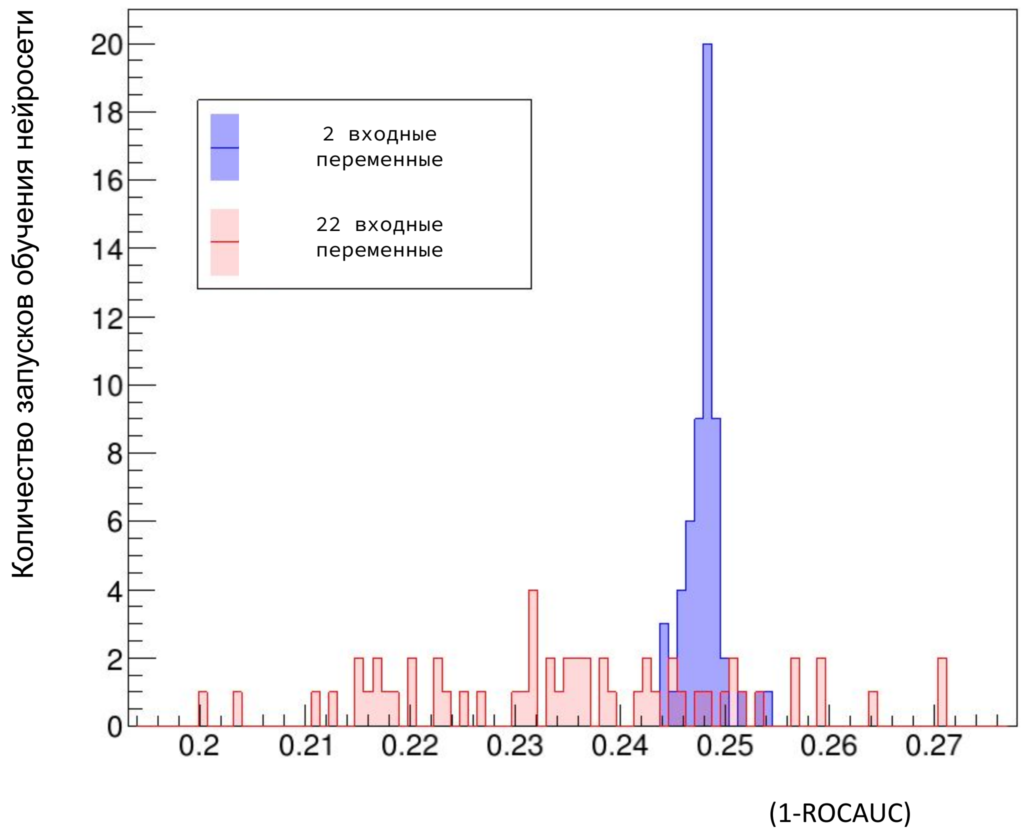
- Evolutionary algorithm working on top of existing NN was created. As a result, a very good improvement of ATLAS tH signal selection is achieved.
- Started to develop a new package which will optimize NN parameters and hyperparameters simultaneously.
- Optimisation of parameters of network with current version of package was tested on 22 input variables with Network architecture as follows: 2 hidden layers, 30 neurons each
- Test shows that even without any optimisation of hyperparameters implemented neural network trained with such algorithm provide ROC-AUC up to ~ **0.8**

Future Plans

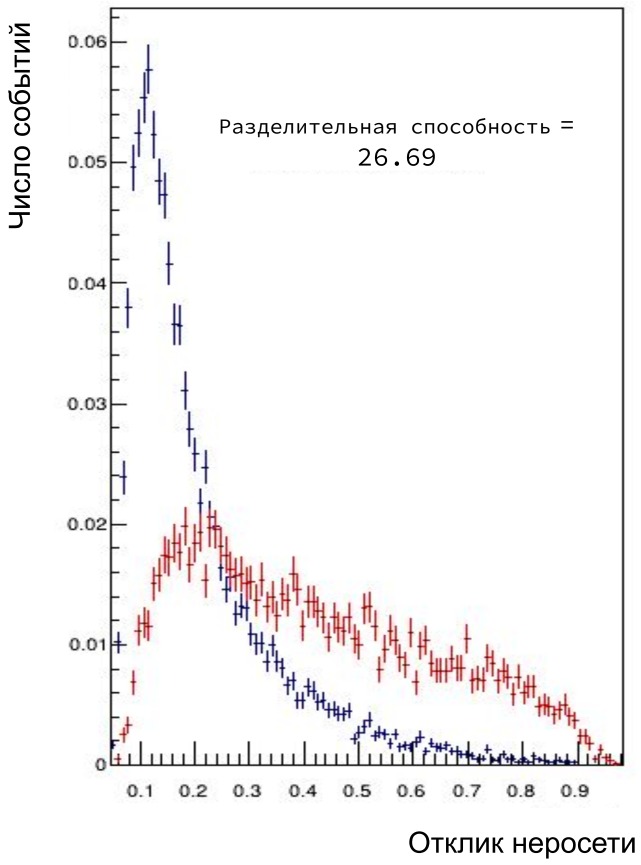
- Implement optimization of hyperparameters
- Integrating evolutionary algorithm into training process

Thank you for your attention

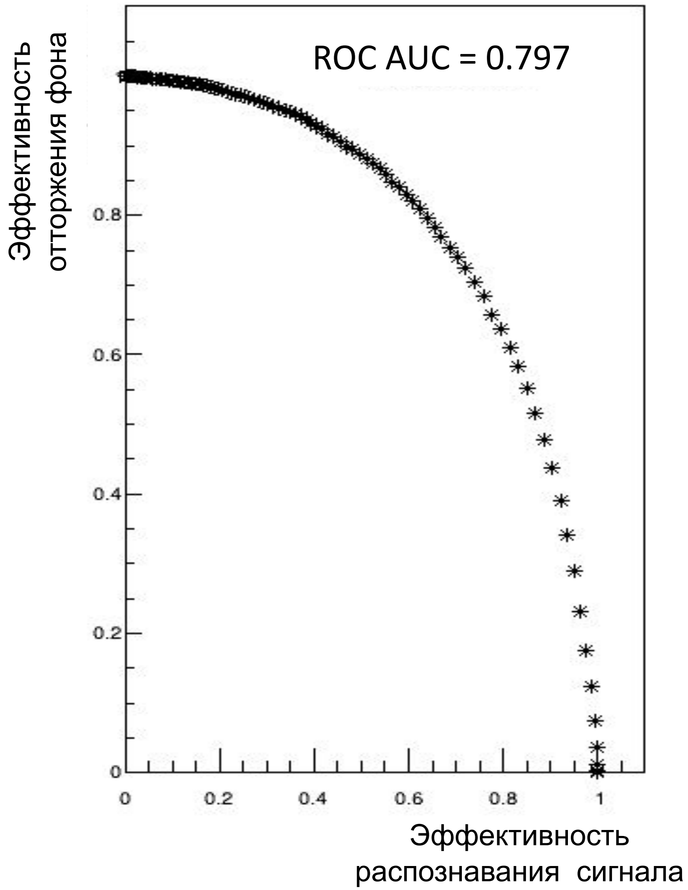
Contact: kiseevavi@jinr.ru


Количество запусков обучения нейросети
20
18
16
14
12
10
8
6
4
2
0
2 входные переменные
22 входные переменные
0.2
0.21
0.22
0.23
0.24
0.25
0.26
0.27
(1-ROCAUC)

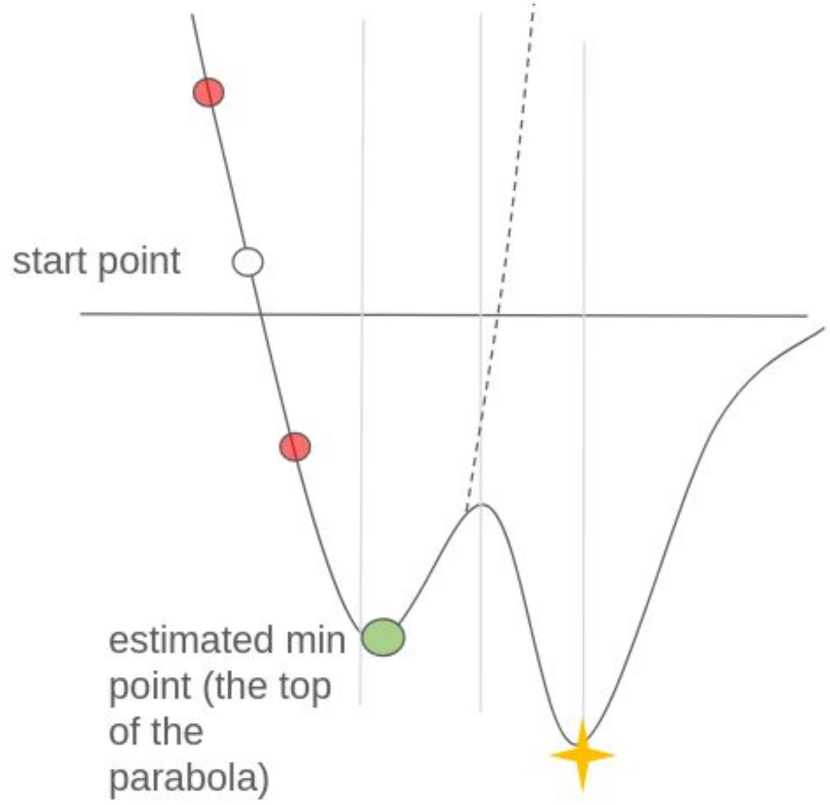
Распределение сигнала и фона



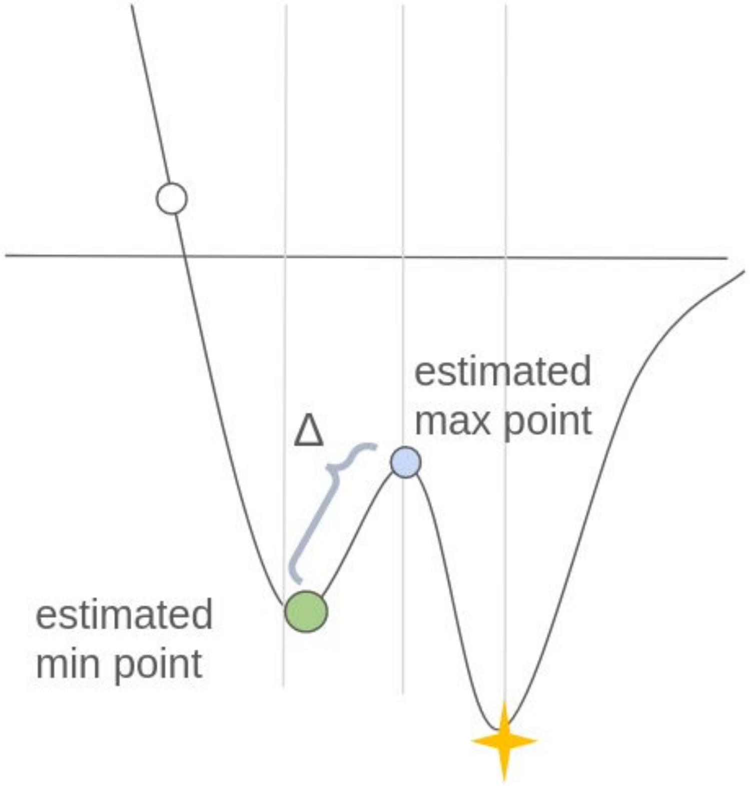
ROC кривая



Status of work // change



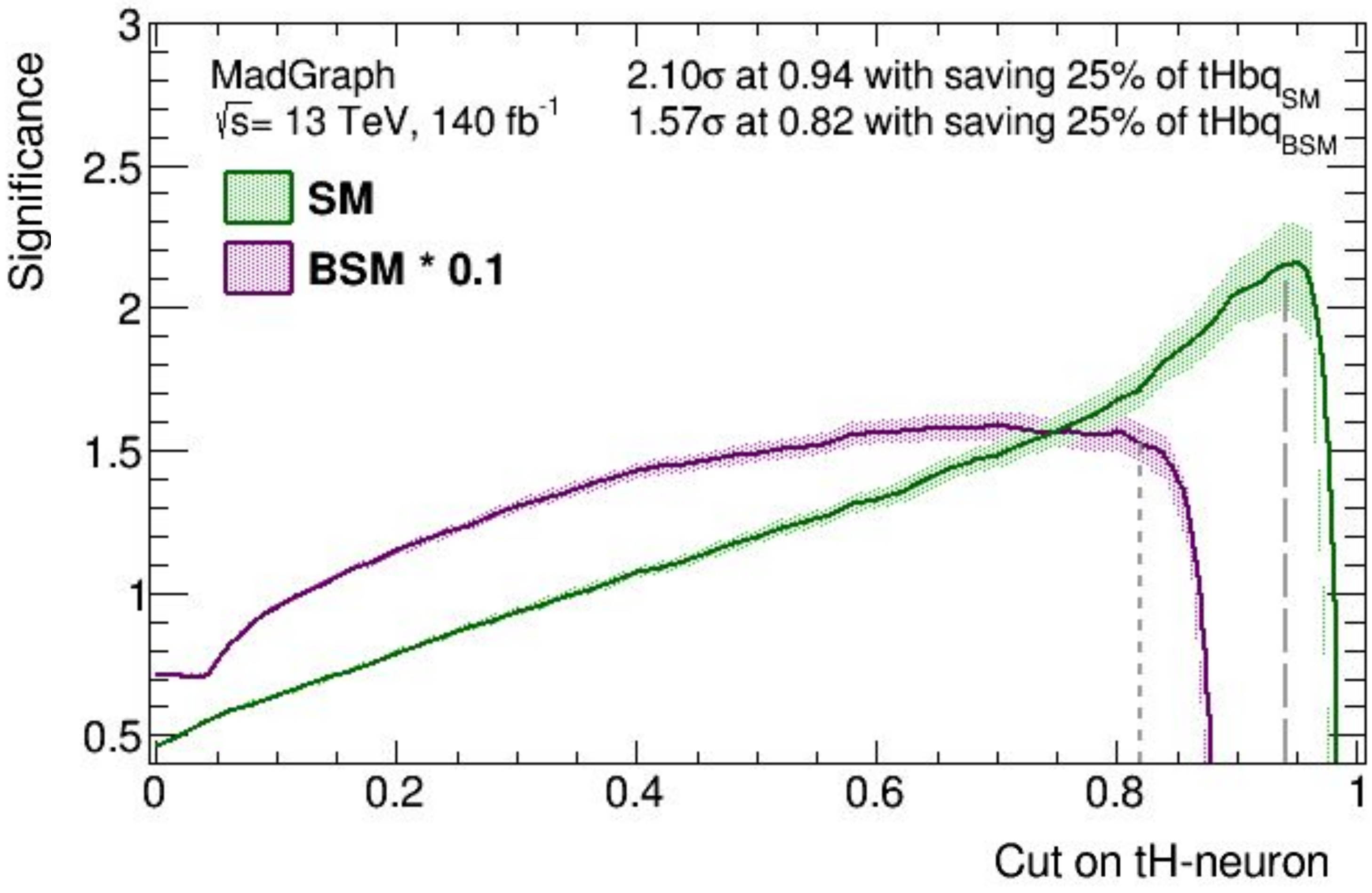
f = ax^2 + bx + c
f' = 2ax+ b
f''=2a



Δ*gauss_point = step

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}

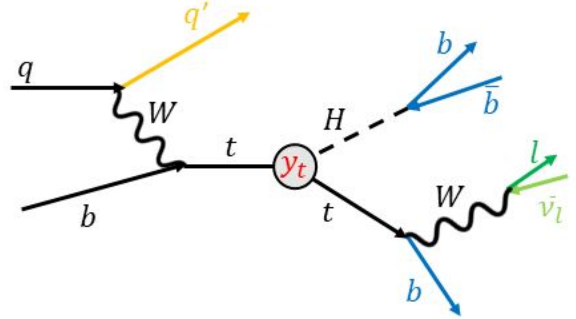
- Without the NN, the significance of the SM signal is **0.47** σ
- After applying the NN: **2.1** σ

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

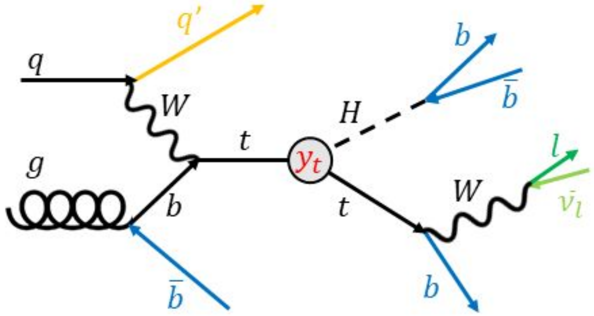
- Without the NN, the significance of the BSM signal is **7** σ
- After applying the NN: **15** σ

Kinematics of signal and backgrounds

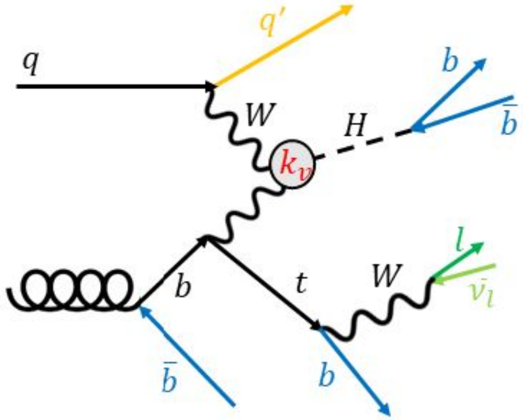
- The kinematics of signal and background processes are very close



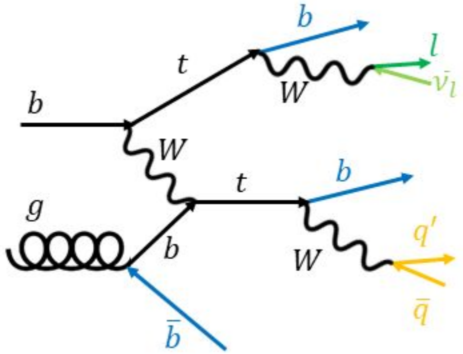
(a) t-channel of the signal tH process



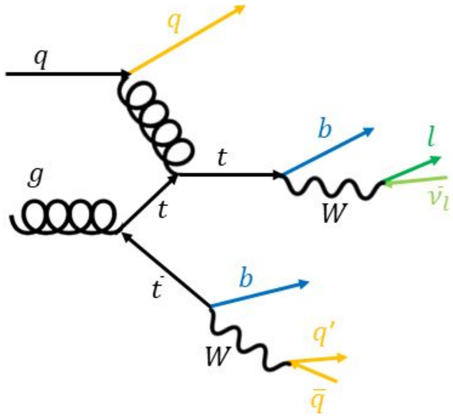
(b) s-channel of the signal tH process



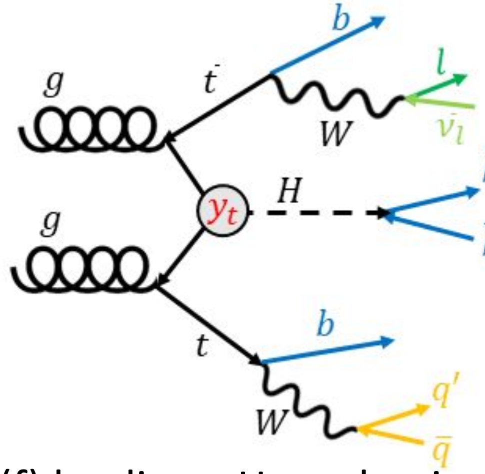
(c) Higgs boson production channel with top quark, where the Higgs boson interacts with the W boson



(d) leading ttb production channel



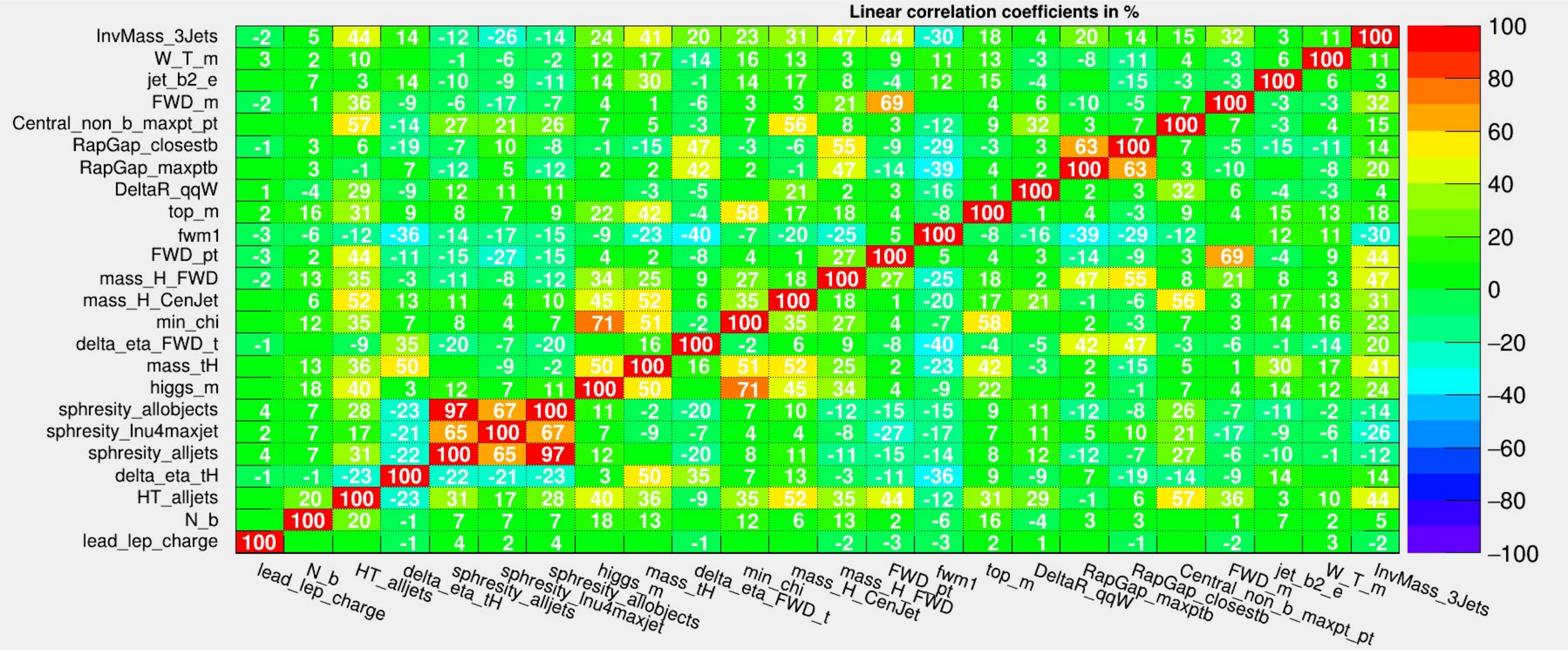
(e) leading ttc and ttL production channels



(f) leading ttH production channel

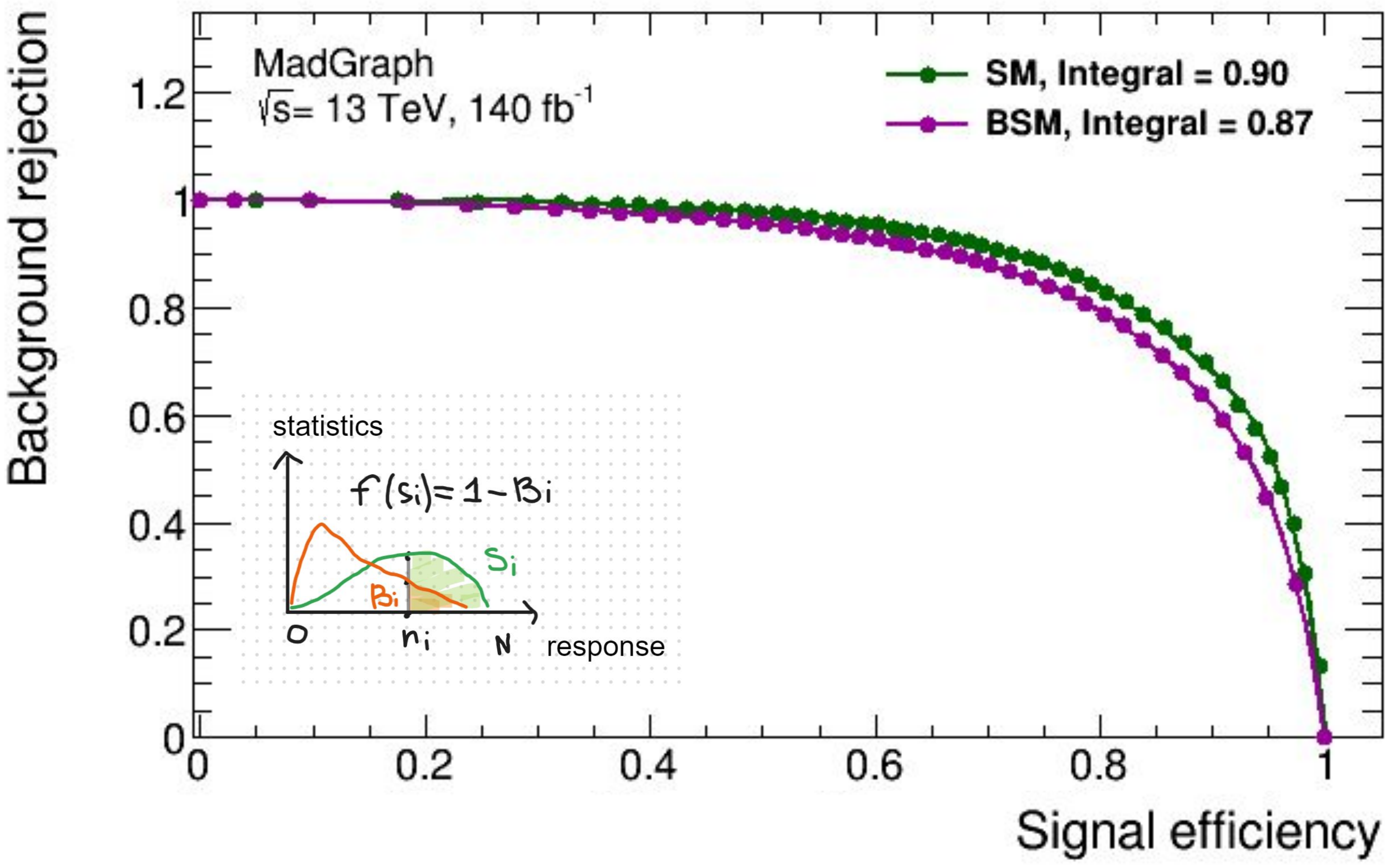
Back-up. Correlation matrix

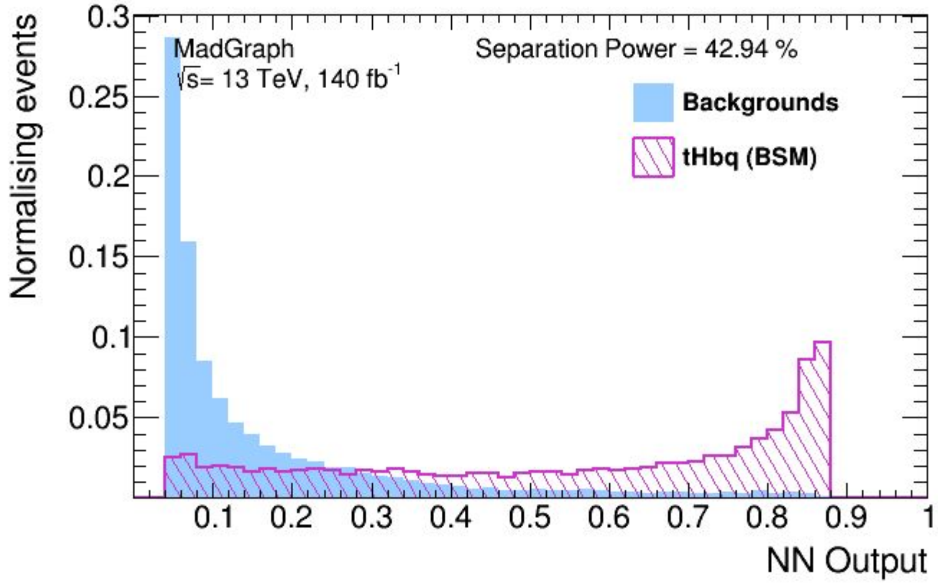
24 Variables Most Sensitive to $tHbq_{SM}$

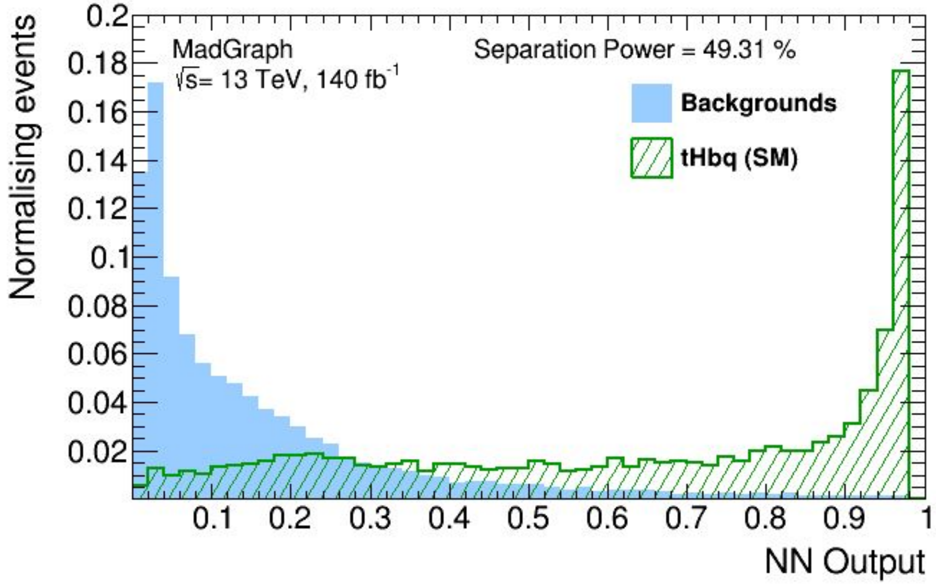


Back-up. ROC-curve integral

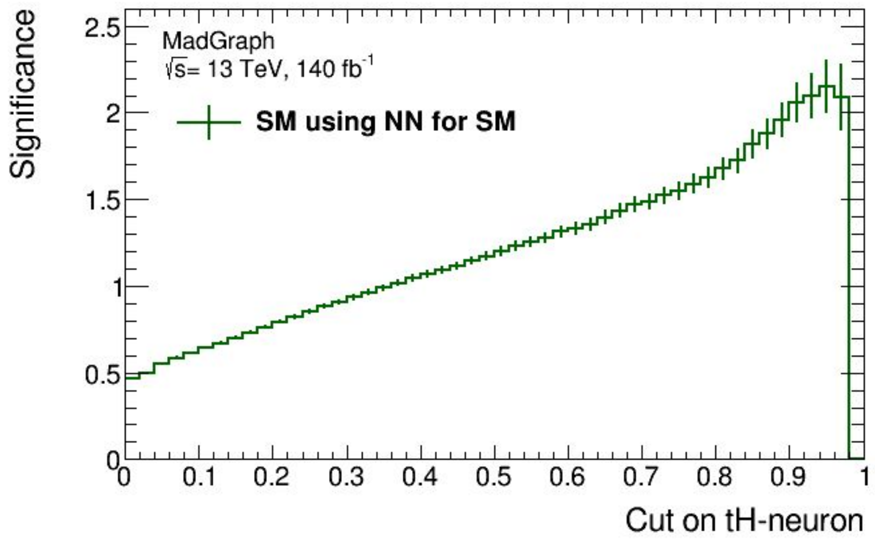
The network efficiency curve is the dependence of the cut signal on the number of background events captured.

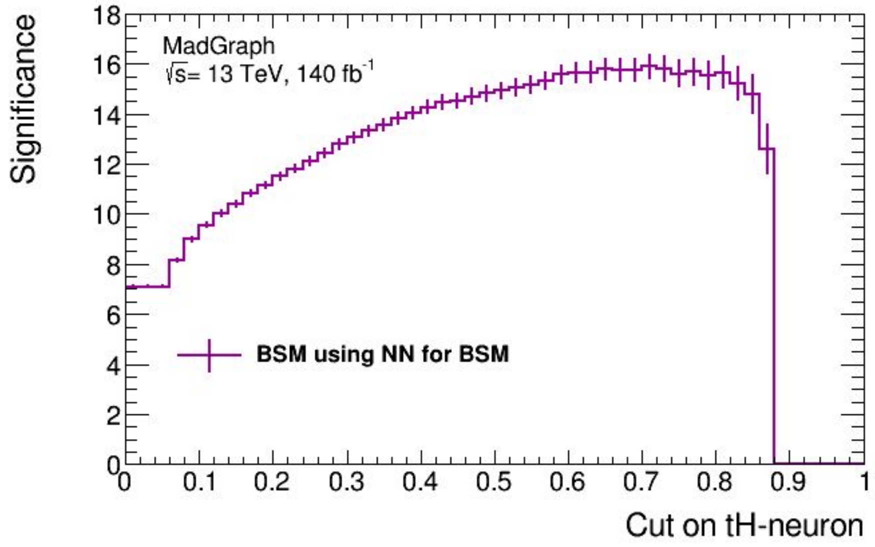


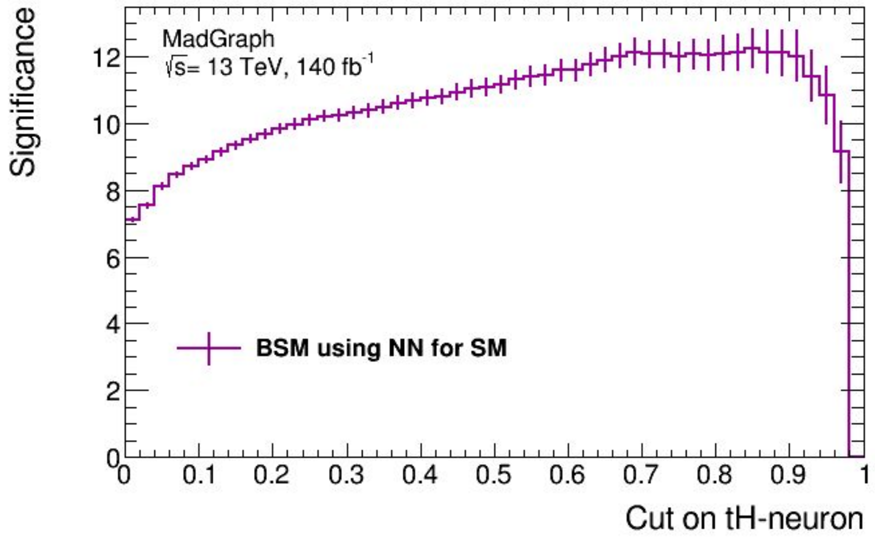


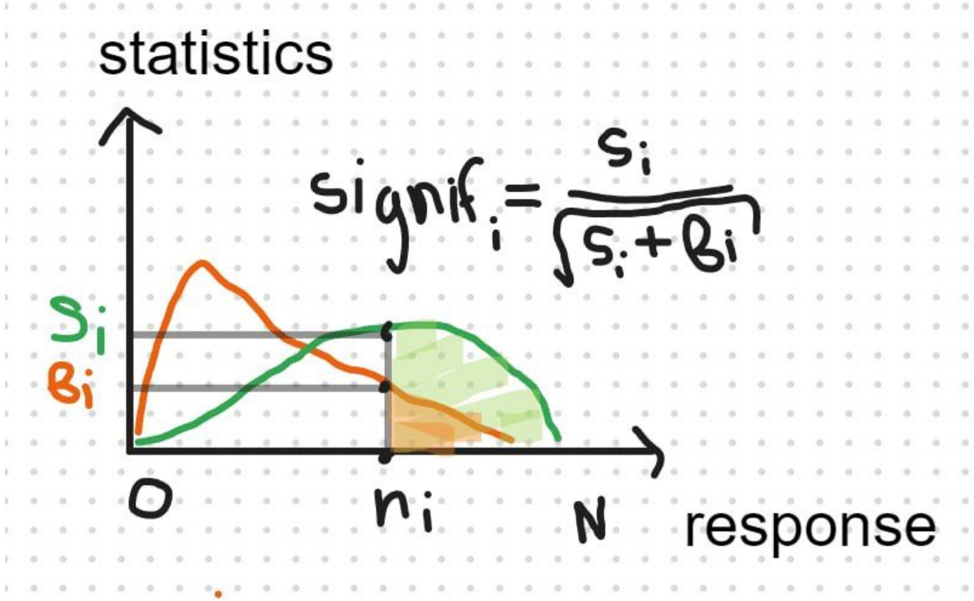


Back-up. Significances









Back-up. Network optimization using evolutionary algorithm