Separating stop 4 body decay from SM background with NN

Artur Cordeiro and Timothée Cabos

Supervisors: Dr.Pedrame Bargassa and Diogo De Bastos

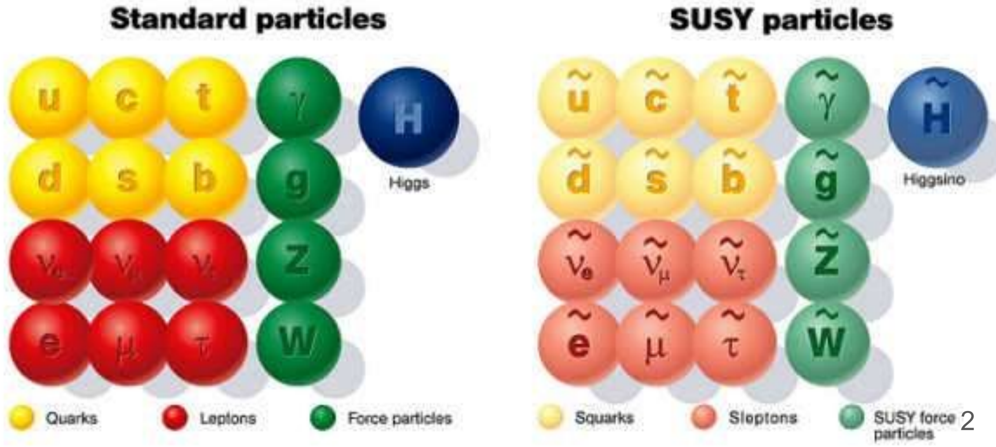
Introduction

SM is the tool we have to describe the world

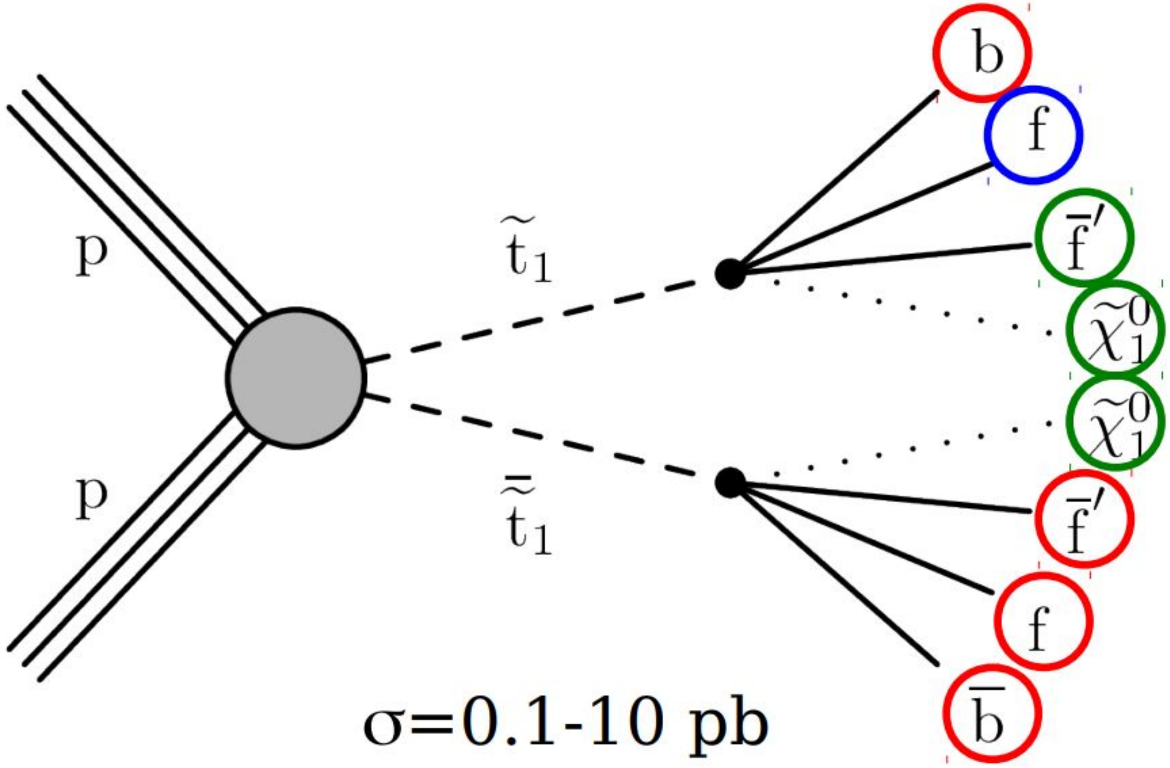
But some phenomena remain unexplained (divergence in higgs' auto-interaction, unification of electro-Weak/Strong/Magnetic forces, dark matter..)

SUperSYmmetry may be the answer!

Top => sTop

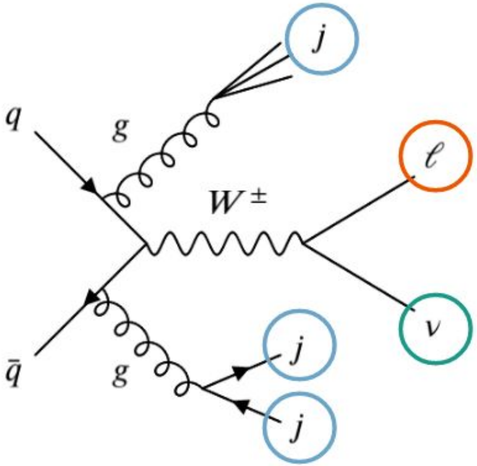


Stop 4 body decay:



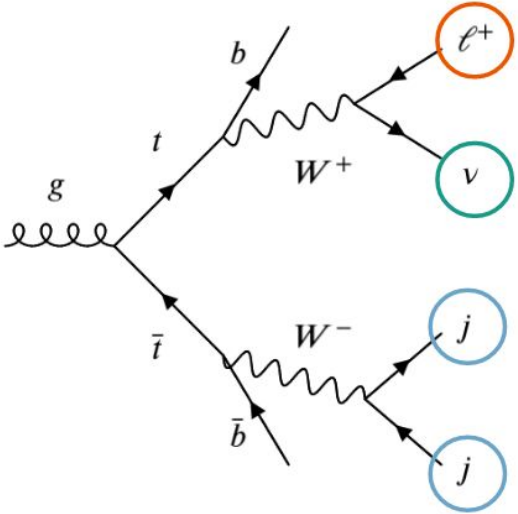
Main backgrounds:

W+Jets



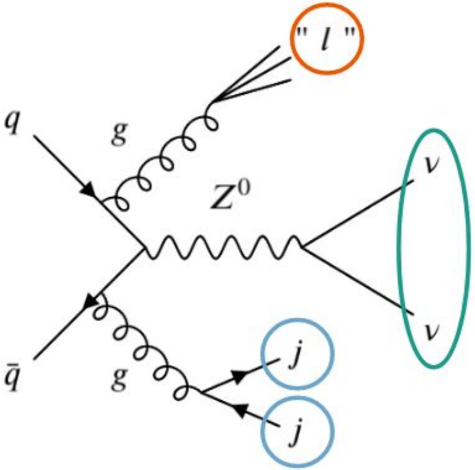
σ = 1395 pb

TTbar+Jets



σ = 832 pb

ZtoNuNu+Jets



σ = 346 pb

Signature: jets + 1ℓ + MET

Other: Drell-Yan+Jets; Single Top; Multiboson; TTbarX; QCD

Neural Network - vs BDT

- Objective is to improve vs BDT:

    → Same input variables

    → Signal : Δm = 30GeV

    → Same preselection:

    ¤ Pt(lep) < 30 GeV ¤ Pt(1jet) > 110 GeV

    ¤ MET > 280 GeV ¤ Ht > 200 GeV

    → Similar training method (data labelling)

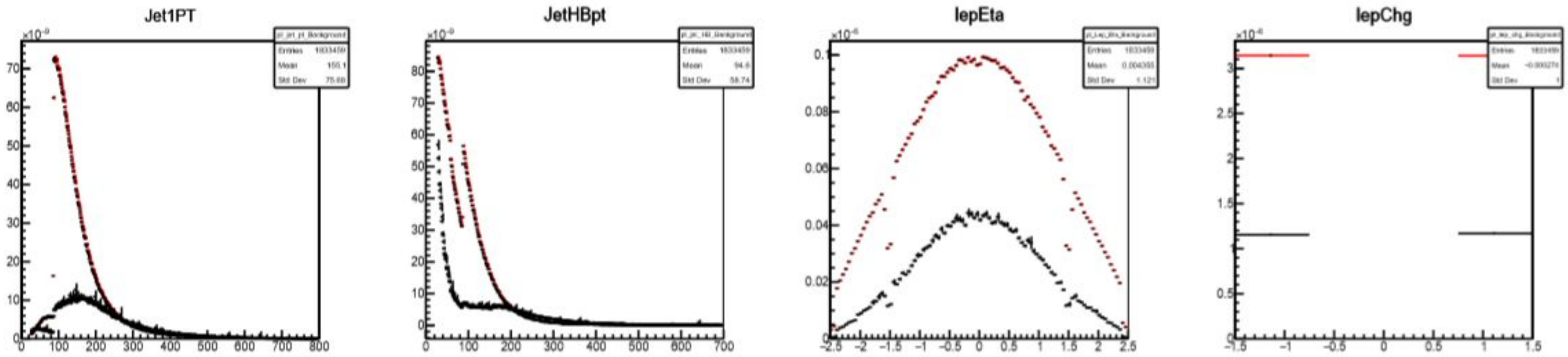
- Our goal is to find the best set of internal NN parameters to achieve the best performance (Figure Of Merit as performance evaluation).

Input variables

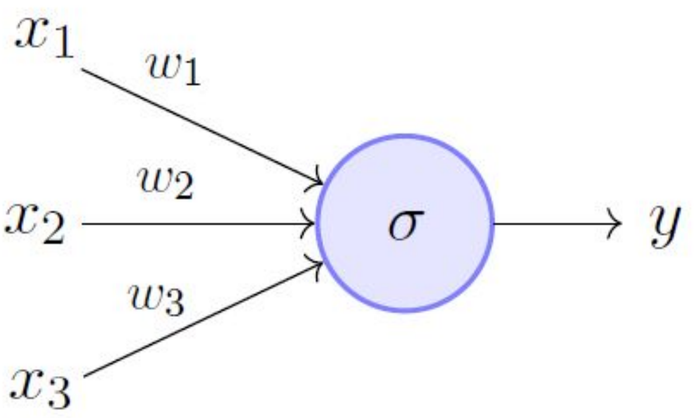
- 12 input variables, deemed most discriminant by 2018's publication
- Most of them are kinematic variables

Red curves = background

Black curves = signal



Neural Network - Neuron



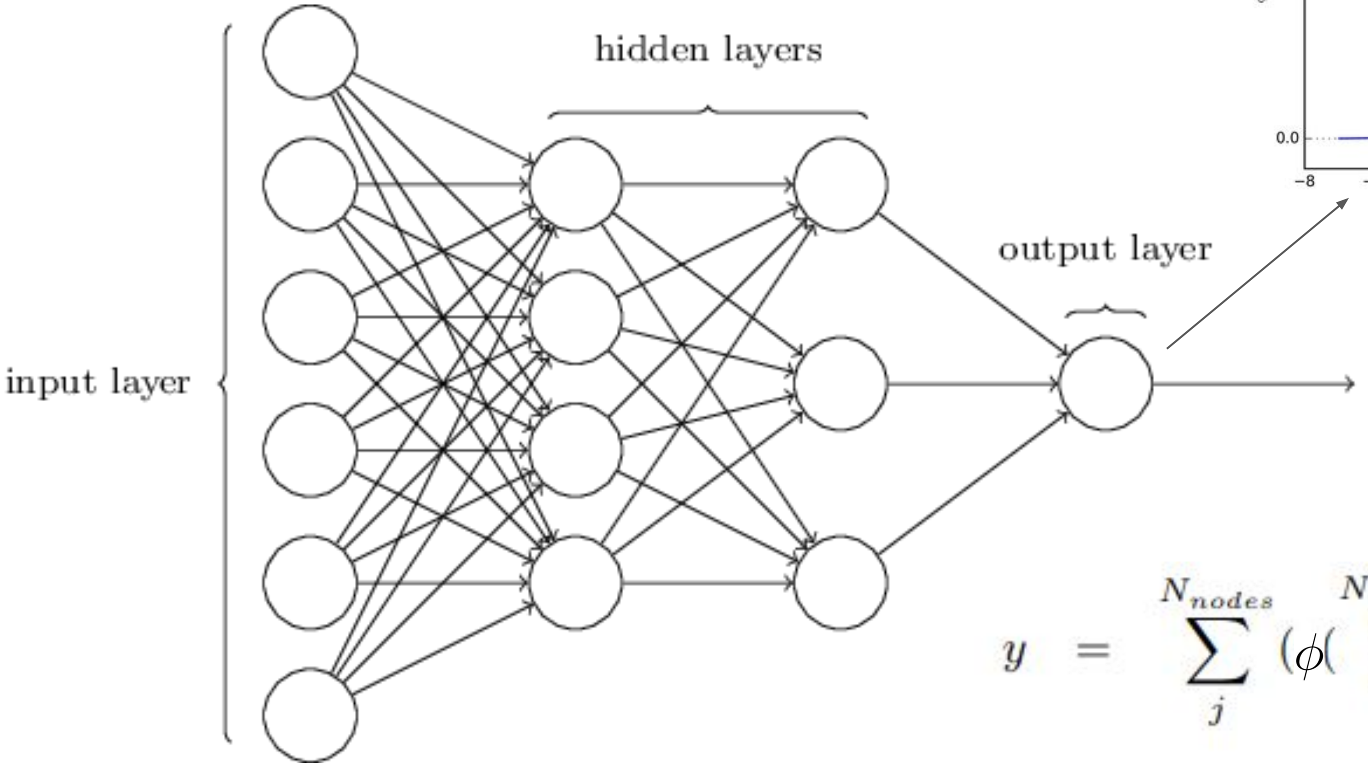
Artificial Neuron

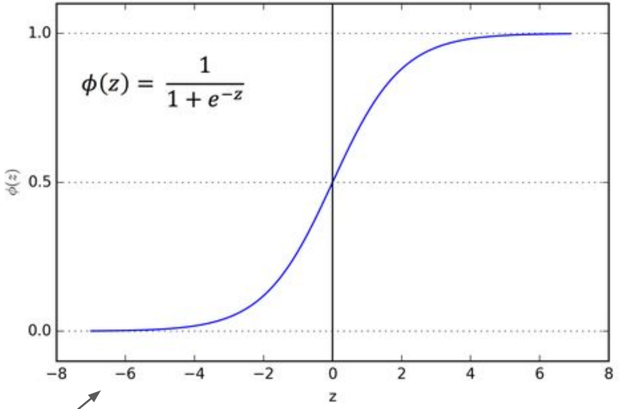
$$y = \sigma\left(\sum_{i}^{n} w_i . x_i\right)$$

Each neuron:

- Is a function that takes all the previous neuron's output as inputs
- Applies a factor (weight) to each input
- Applies the sum to its activation function
- Resulting into one output

Neural Network - Architecture





$$y = \sum_{j}^{N_{nodes}} (\phi(\sum_{i}^{N_{inputs}} w_{ij} \cdot x_i) \cdot w_j) + O$$

Neural Network - Training

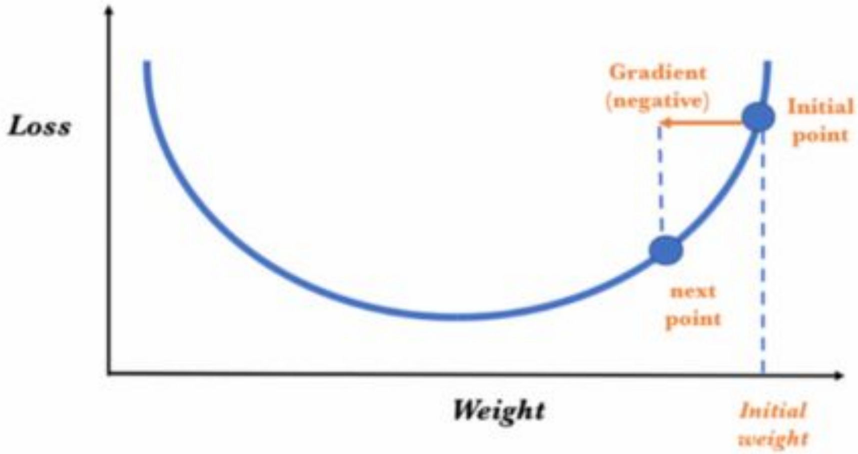
Training = minimizing the "Loss function": $$L \simeq (y - \hat{y})^2$$

(Where y is the NN's output and $\hat{y}$ the label.)

$\longrightarrow$ Back propagation: adapting the weights so that the Loss is minimized.

$$w_i{}' = w_i - \lambda \frac{\partial L}{\partial w_i}$$

λ = Learning rate or step size



$\longrightarrow$ When $\frac{\partial L}{\partial w_i} = 0$, weights do not change and the NN stops learning.

Method

1) Try an architecture

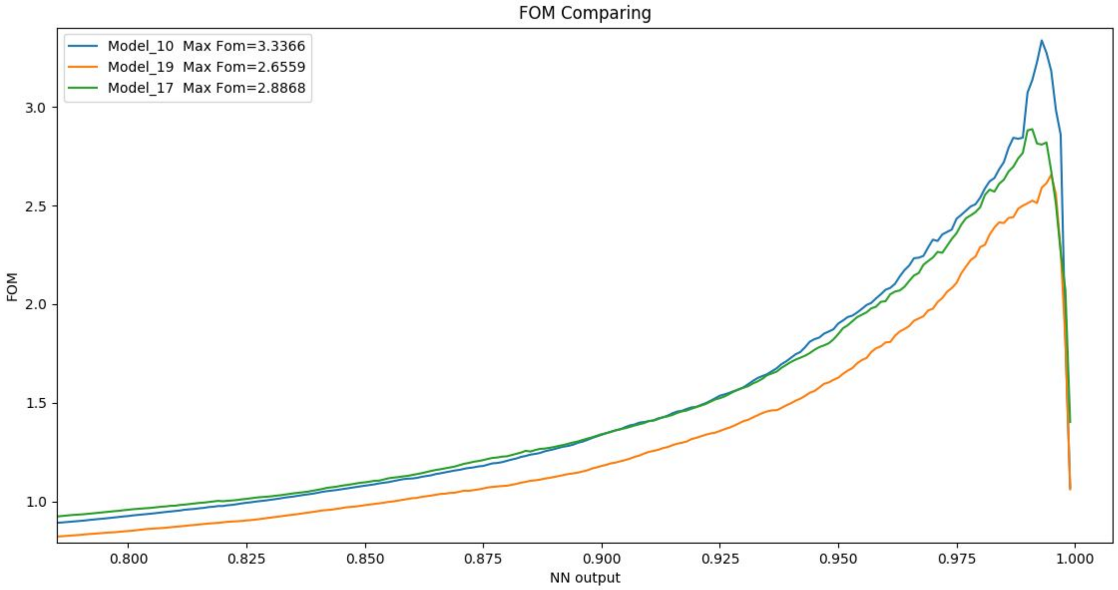
- Change learning rate
- Change weight initialiser
- Repeat

2) Checks

- Take most promising models
- Check overtraining (training vs validation samples)
- Average (Statistical fluctuations)
  - Run “same” model 5 times

3) Compare models’ FOMs:

- Compare the highest values of the FOM
- Check if FOM is higher globally or not



$$\mathrm{FOM} = \sqrt{2\left((S+B)\ln\left[\frac{(S+B)\,(B+\sigma_B^2)}{B^2+(S+B)\,\sigma_B^2}\right]-\left(\frac{B^2}{\sigma_B^2}\right)\ln\left[1+\frac{\sigma_B^2\,S}{B\,(B+\sigma_B^2)}\right]\right)}$$

S = Signal
B = Background
σ = 0.2 * B

https://arxiv.org/pdf/1805.05784.pdf

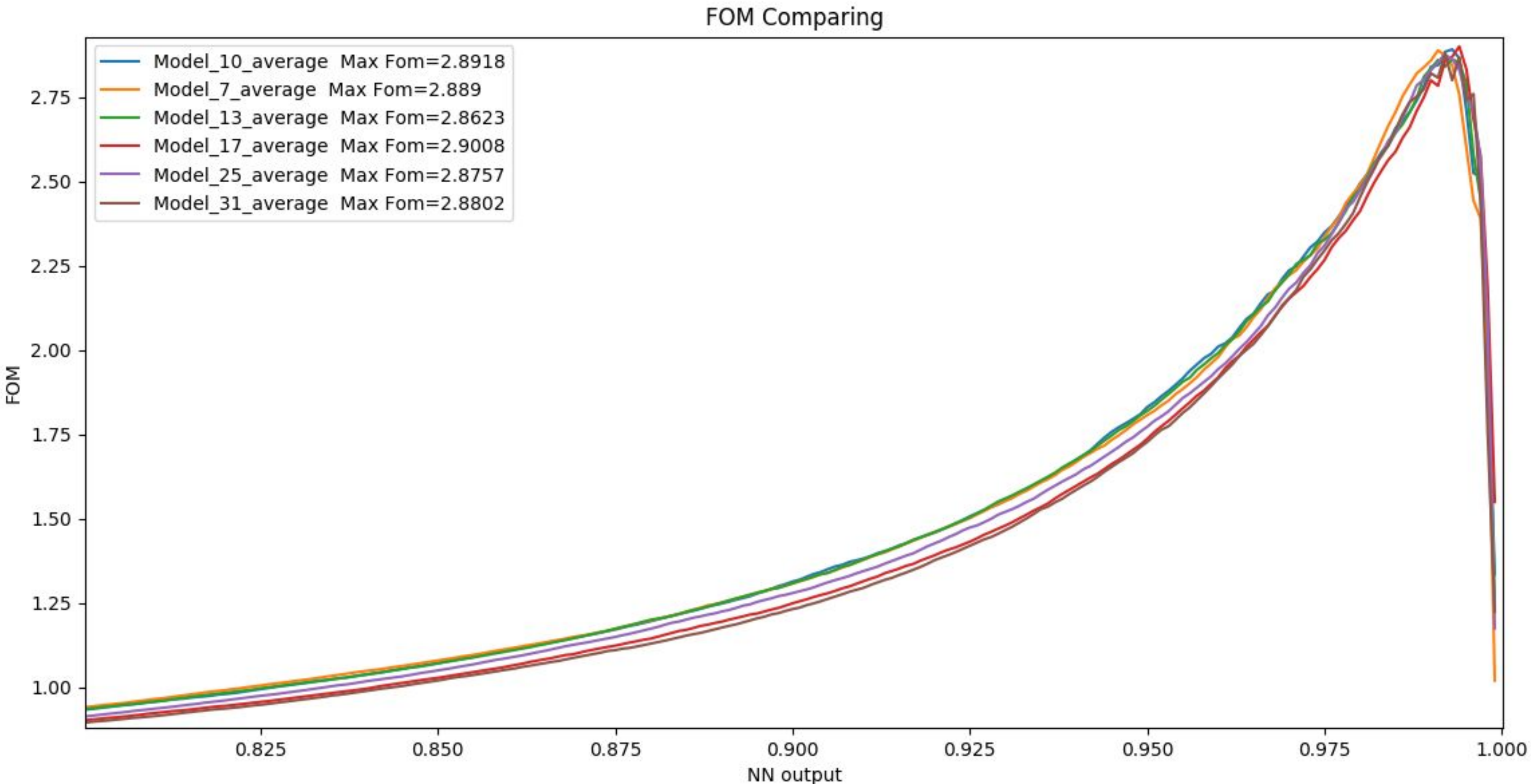
Results - Models

Best options out of 30+ models

Model version	Architecture	Learning rate	Weight initializer
4	12 24 18 12 6 1	0.001	Glorot Uniform
7	12 24 18 12 6 1	0.01	Glorot Uniform
10	12 24 18 12 6 1	0.01	He_normal
13	12 24 18 12 6 1	0.011	He_normal
17	12 24 22 20 10 6 1	0.01	He_normal
25	12 24 22 20 10 6 1	0.008	He_normal
30	12 32 24 16 10 6 1	0.015	He_normal
19 - Diogo's	12 13 13 1	0.003	He_normal

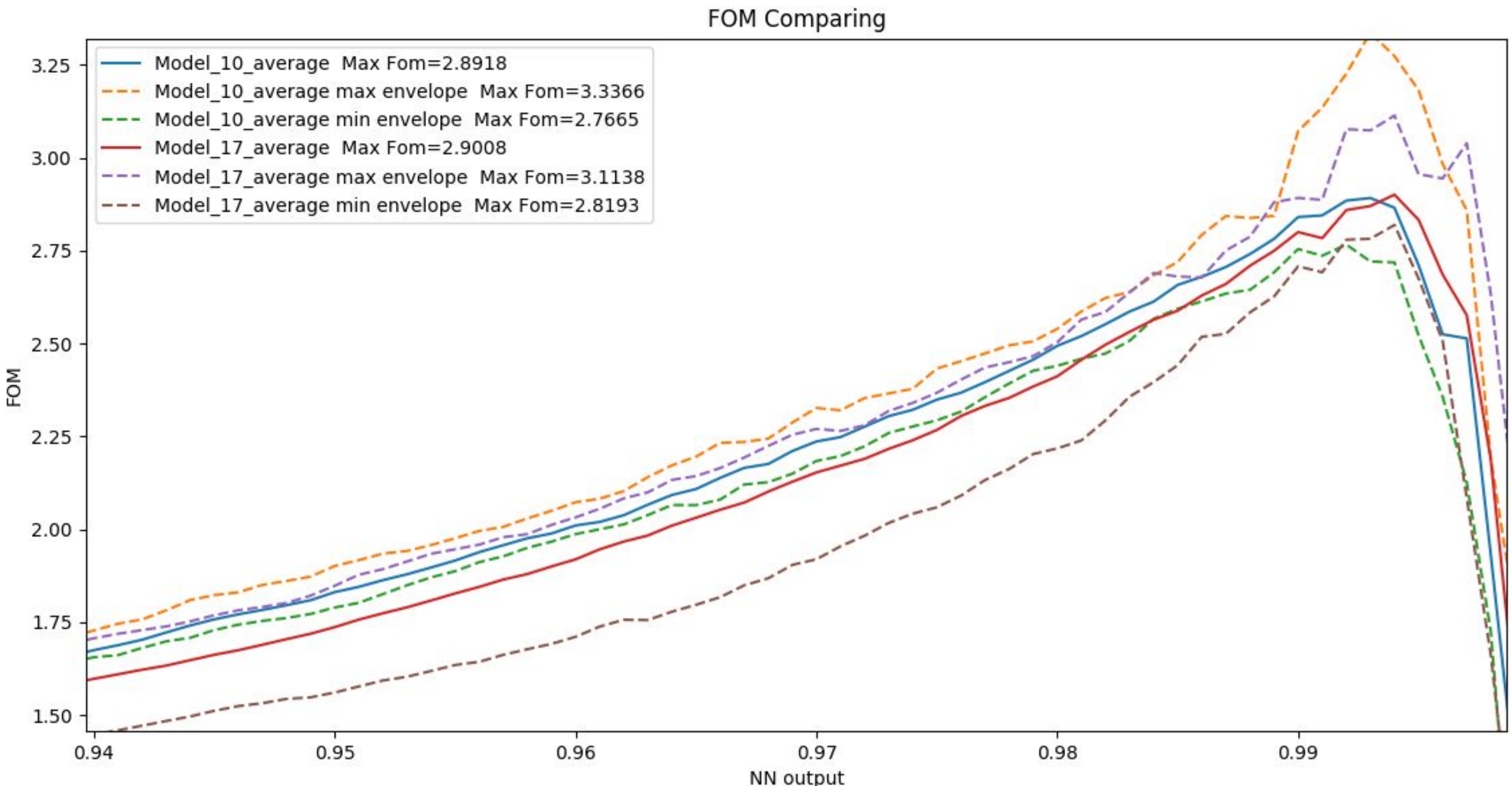
Results - Comparing Averages

- To select the best model: we first compare the FOM average curves



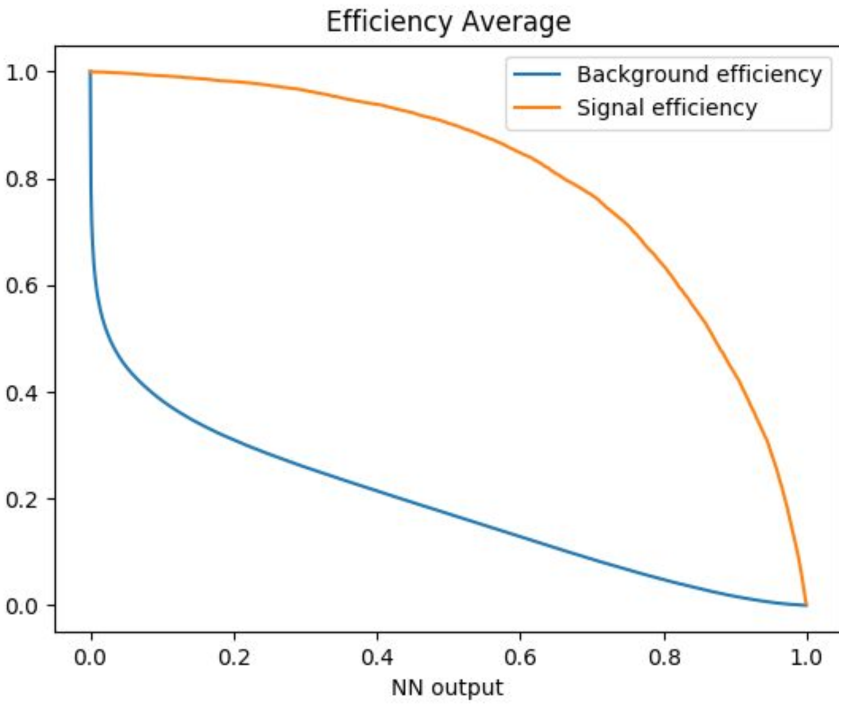
Results - Comparing Averages

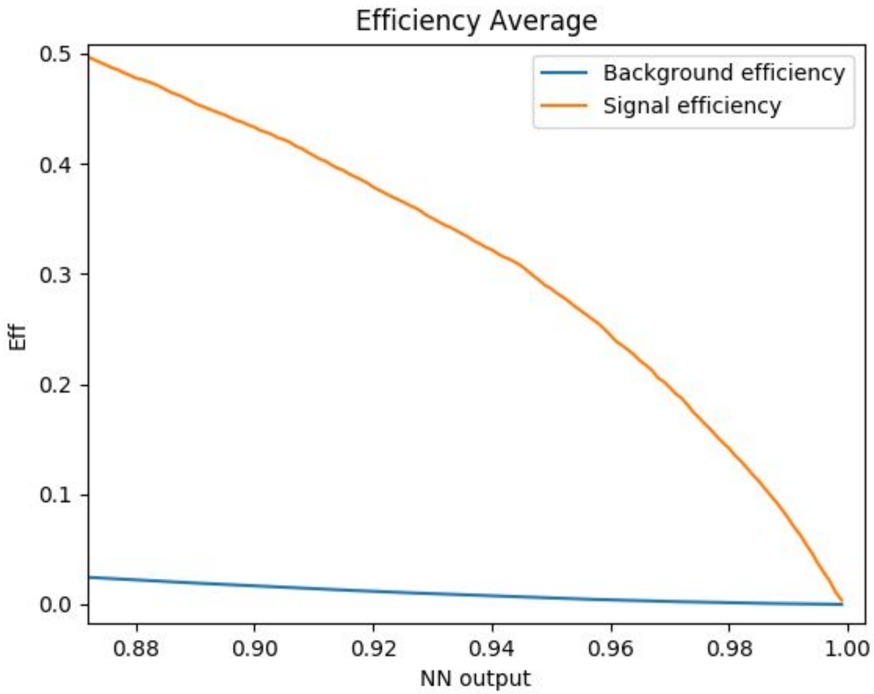
- To select the best model: we then compare how much the curves fluctuate



Results - Comparing Averages. Efficiency curves

To select the best model: we then look the highest FOM values and check fluctuations in the efficiency curves (S and B)

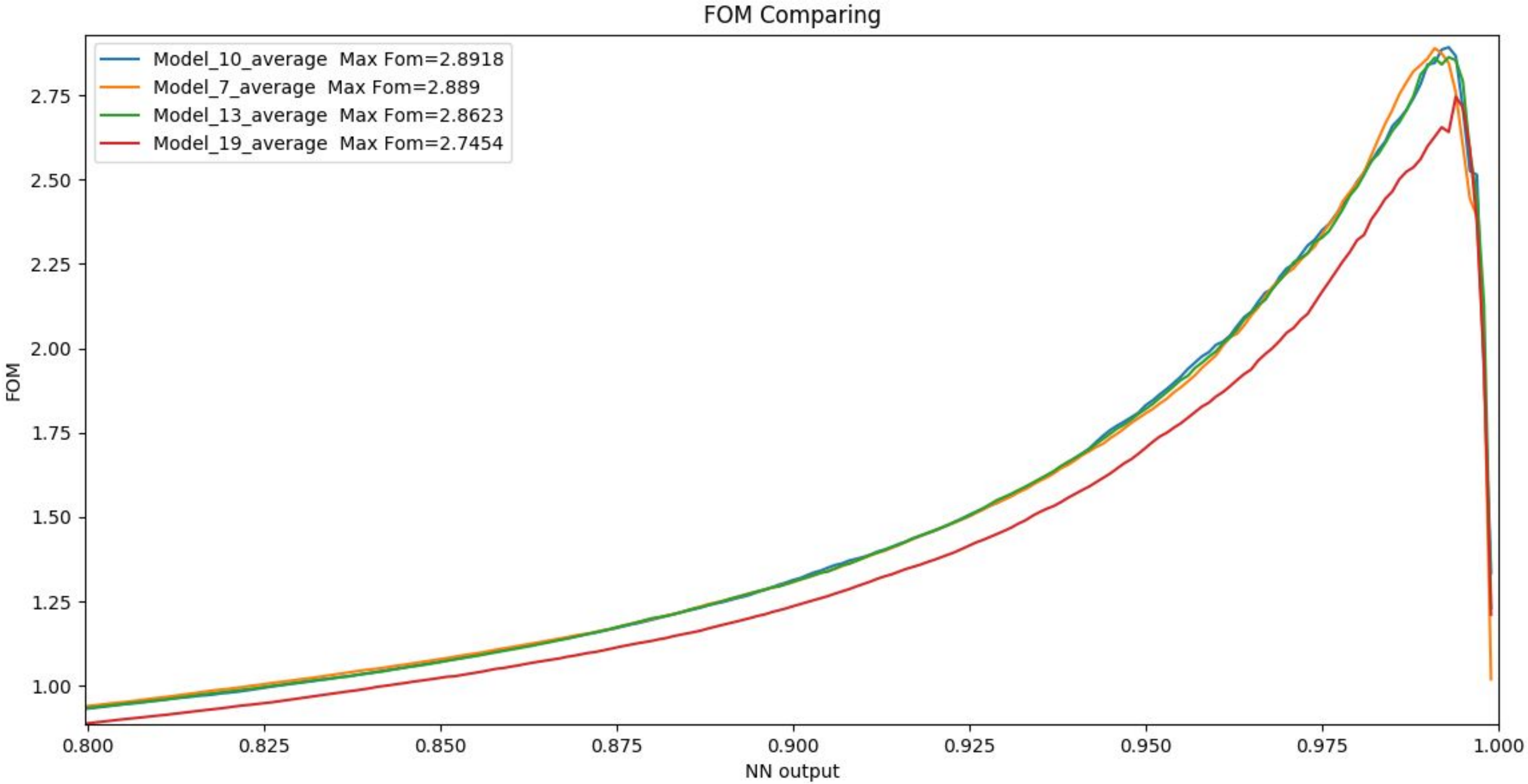




No statistical fluctuation observed for signal and background samples

Results - Final Comparaison

Model 10 systematically higher than others:



Results - Cuts

FOM Cut at the max FOM

Version	Background yield	Signal yield
Model 10	44,36	17,39
Model 19 - Diogo`s	56,38	20,94

FOM Cut at 0,97

Version	Background yield	Signal yield
Model 10	313,33	61,74
Model 19 - Diogo`s	413,28	67,84

Despite the fact that those high FOMs are taking place at high NN outputs (0.99), they are still relatively high compared with those obtained by the BDT in the 2018 paper.

Conclusions

- We developed new NN architectures to separate Stop 4 body decay signal from SM background
  - Different learning rates
  - Different (N layer, N node)
- In doing so, we were guided by two criteria:
  - Performance: checked by FOM and efficiency curves
  - Validity: checked by over training test
- Best model overall is the model 10:
  - Architecture : 12 nodes input layers (activation ReLu) ; 24 ,18 ,12 and 6 nodes hidden layers (activation ReLu) ; 1 node output layer (activation Sigmoid)
  - Weight initializer : he_normal
  - Learning rate : 0.01
  - Optimizer : Adam
  - Did not use decay rate nor dropouts

Conclusions

- When comparing with the BDT in the publication of 2018, we can see that NN achieves higher FOMs than the BDT (max FOM is increased by almost 50%).

Version	2018 BDT	Diogo's NN	Model 10
MAX FOM (Average)	1.96	2.75	2.89

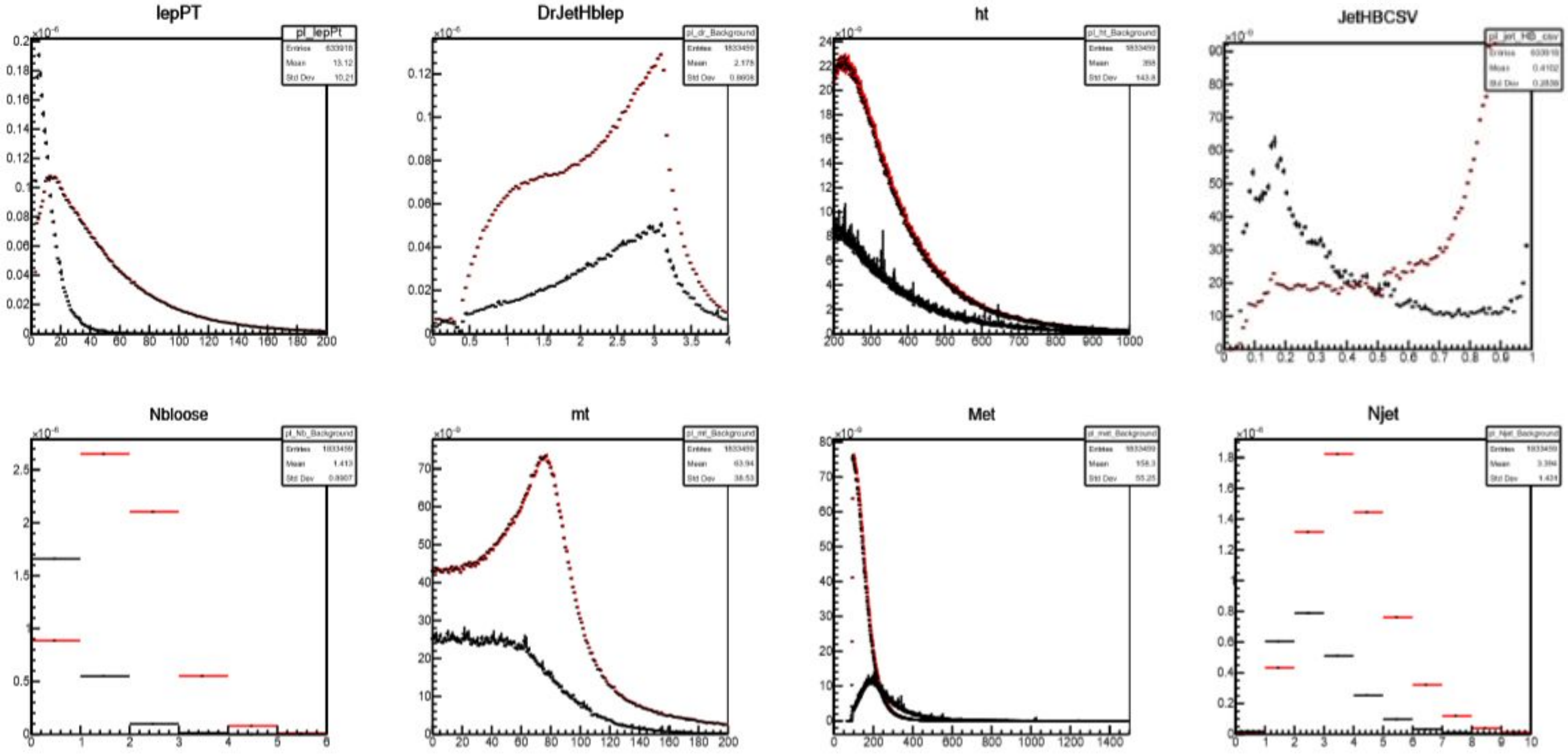
- This result is very promising because the FOM is a hard number to change and this gives a hint to do a more complete study of BDT vs NN

Thanks!

Summary:

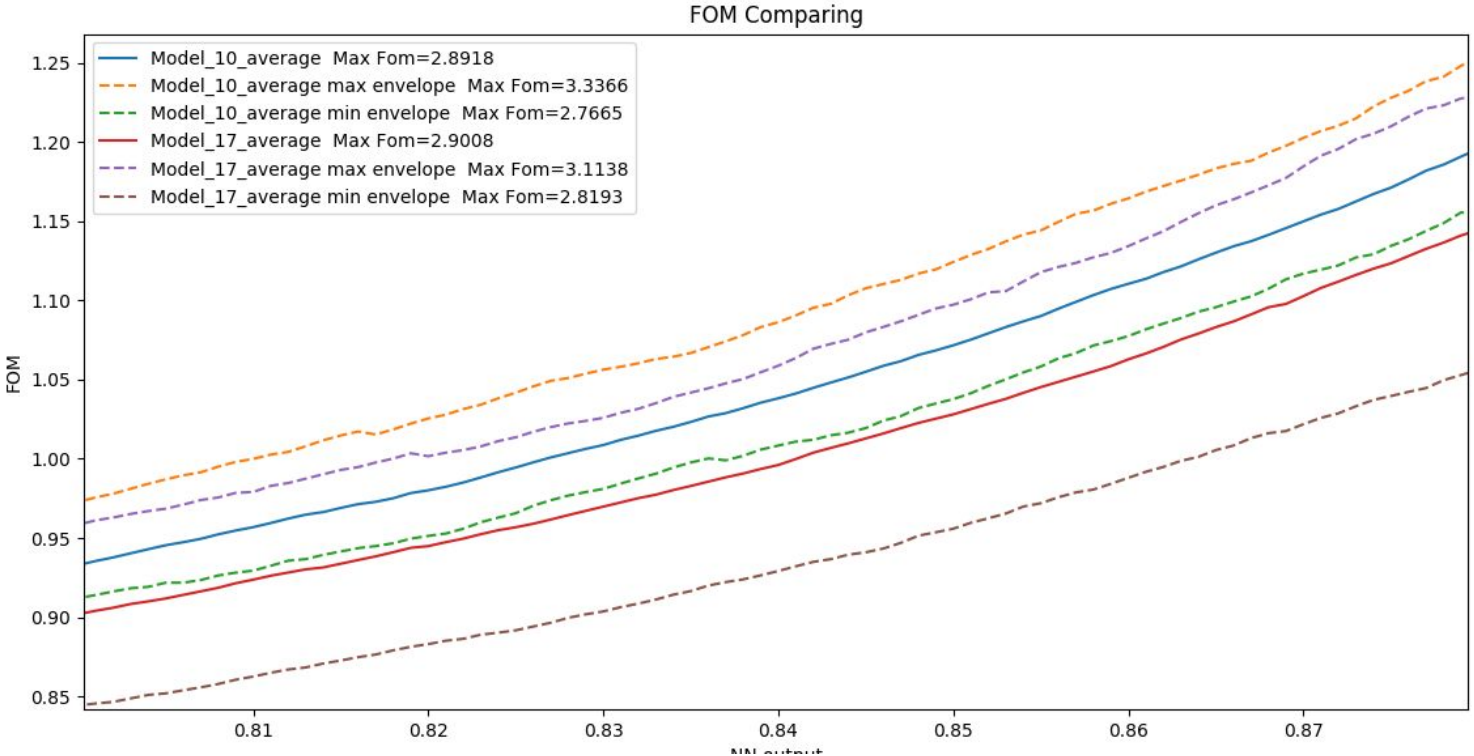
- Introduction
- Stop 4 body decay
- Main backgrounds
- Neural Network
  - NN vs BDT
  - Basic knowledges on how NN works
  - Maximising NN's performance
- Results
- Perspectives

Input variables



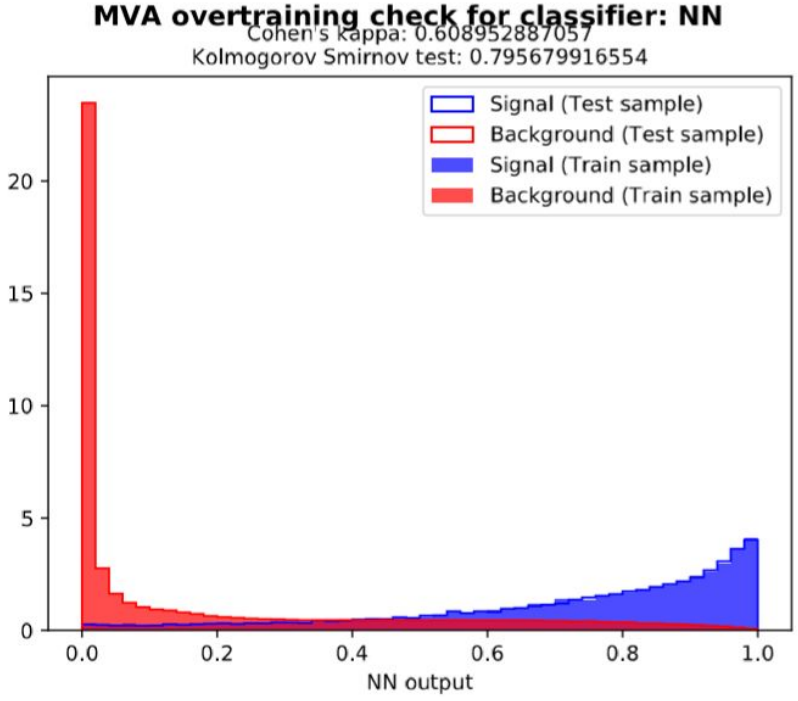
Results - Comparing Averages

- To select the best model: we compare the best models with their envelopes (Max, min of a given model > to another)

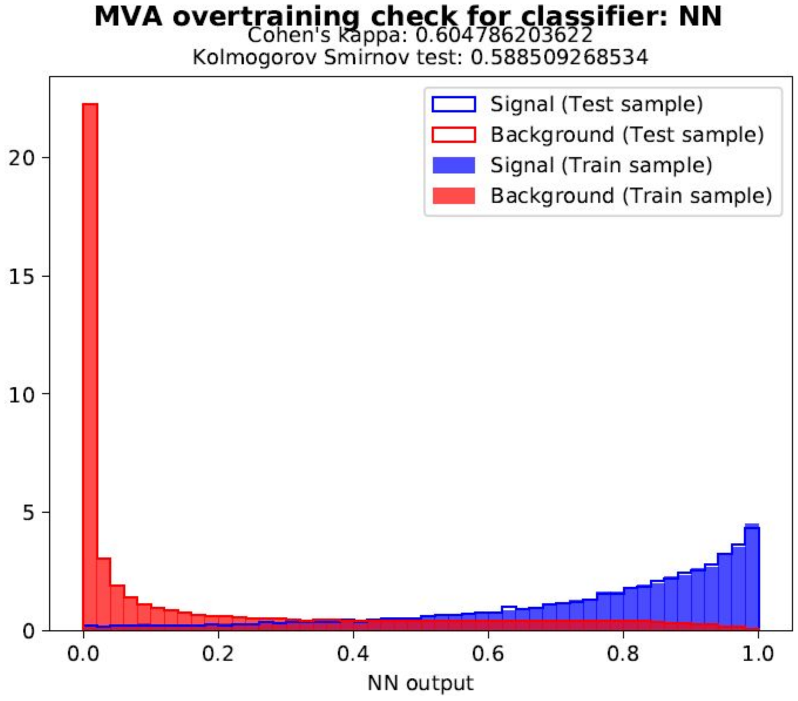


Results - Overtraining

Model 10



Model 19 (Diogo's)



No over training observed for all models