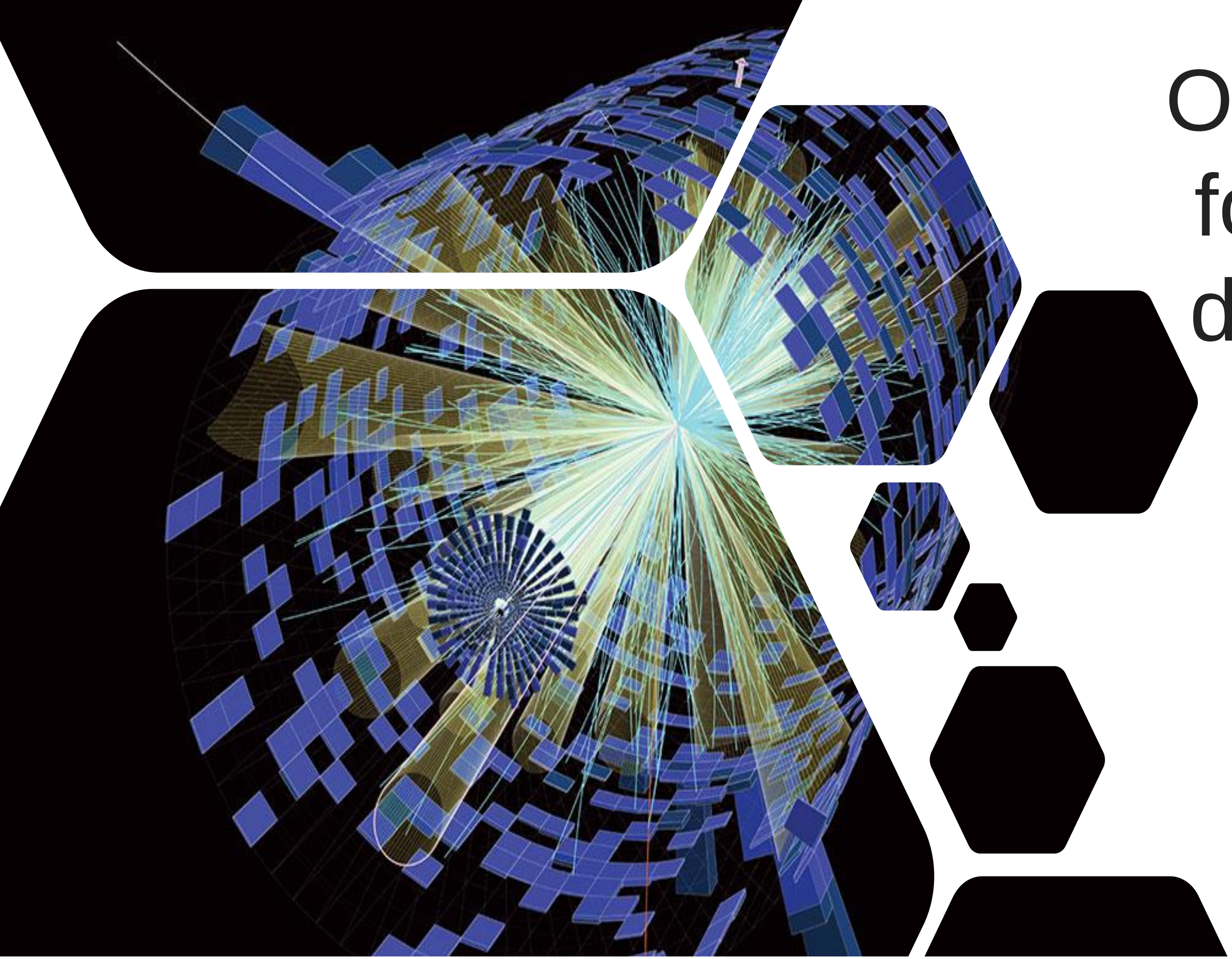
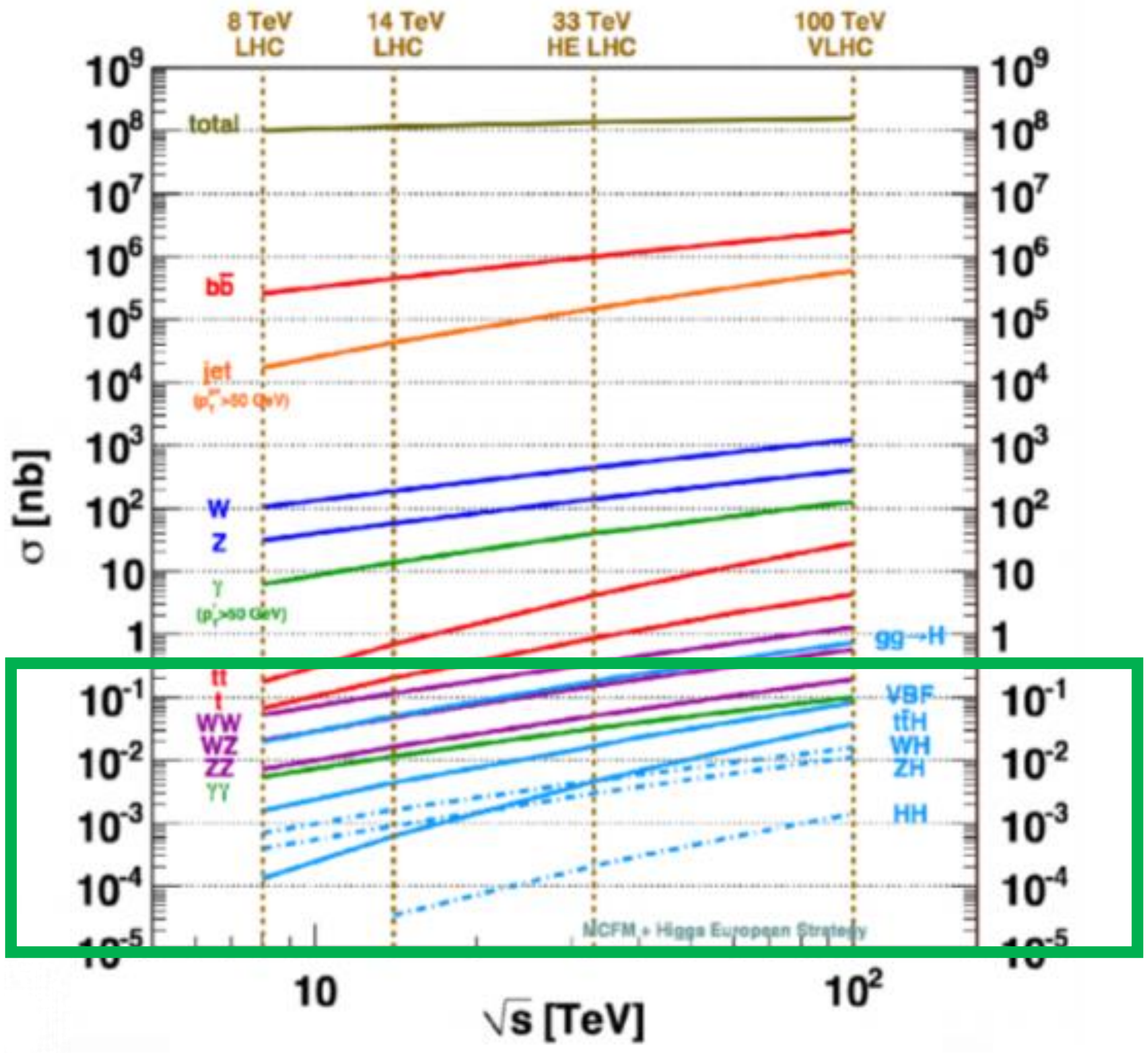


Optimization in the search for rare Higgs processes during the HL-LHC using AI algorithms

Alfredo Castaneda
University of Sonora

XXXIV Annual meeting DPyC-SMF

Motivation



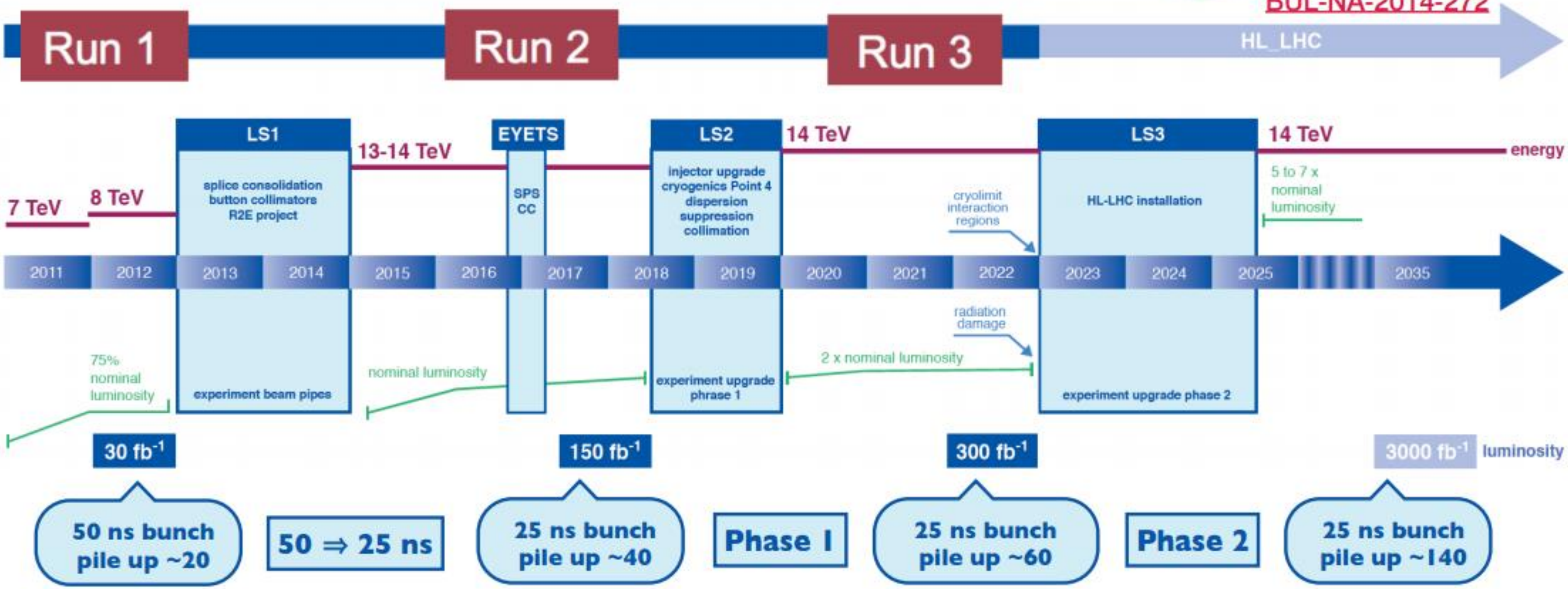
- The LHC and future accelerators (FCC) will allow scientists to measure “rare” processes both in SM and BSM
- With the data collected so far processes as double Higgs production and some Higgs decays are not accessible yet
- Two things are needed:
  - More Luminosity (~10 times more)
  - Better methods to separate signal and background

High Luminosity LHC

LHC / HL-LHC Plan

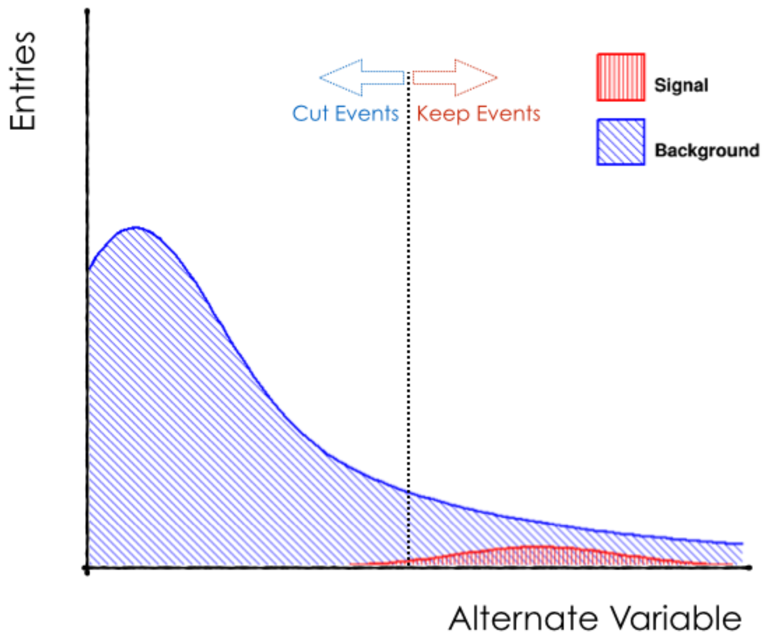
High Luminosity LHC

BUL-NA-2014-272



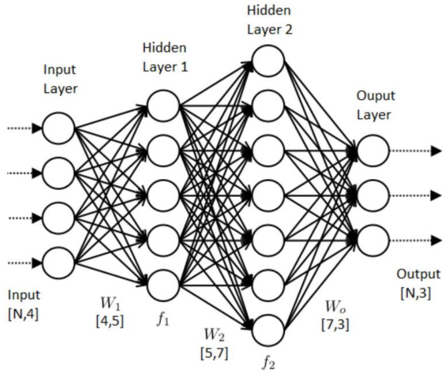
Traditional vs alternative analysis methods

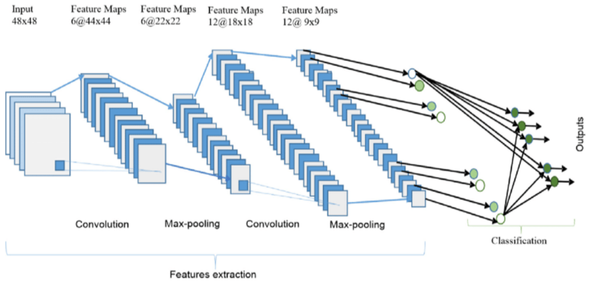
Cut-based analysis



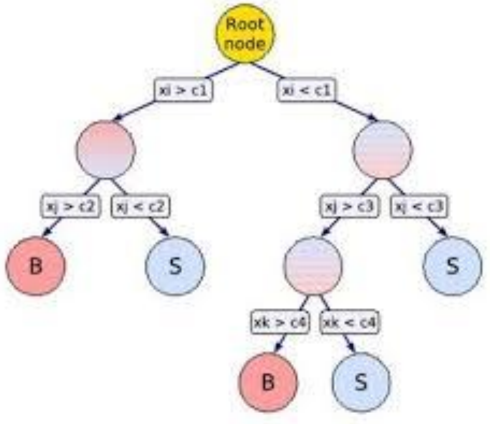
- Based on basic physics assumptions, easy interpretation of systematics
Not optimized separation of signal and background

Artificial Neural networks



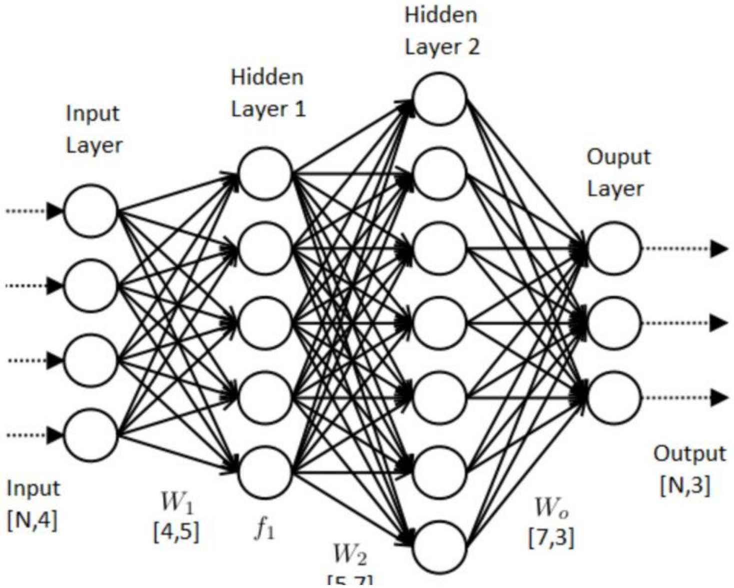


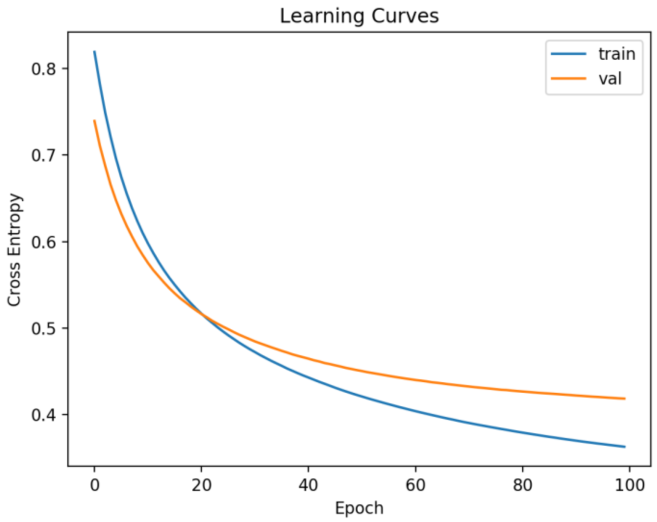
Boosted Decision Trees



- Optimized method to separate signal and background
- Not relying only on first physics principles
Difficult to assign a systematic uncertainty

Artificial neural networks in a nutshell

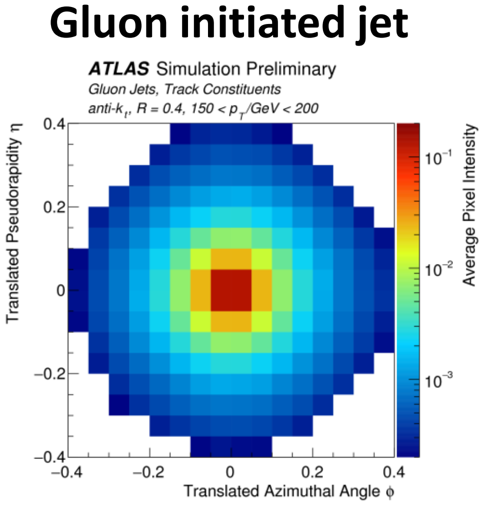


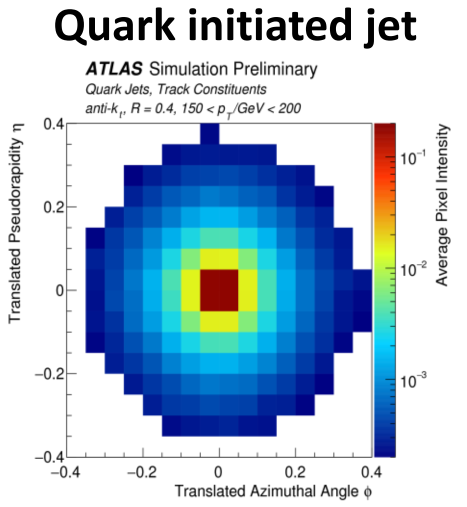


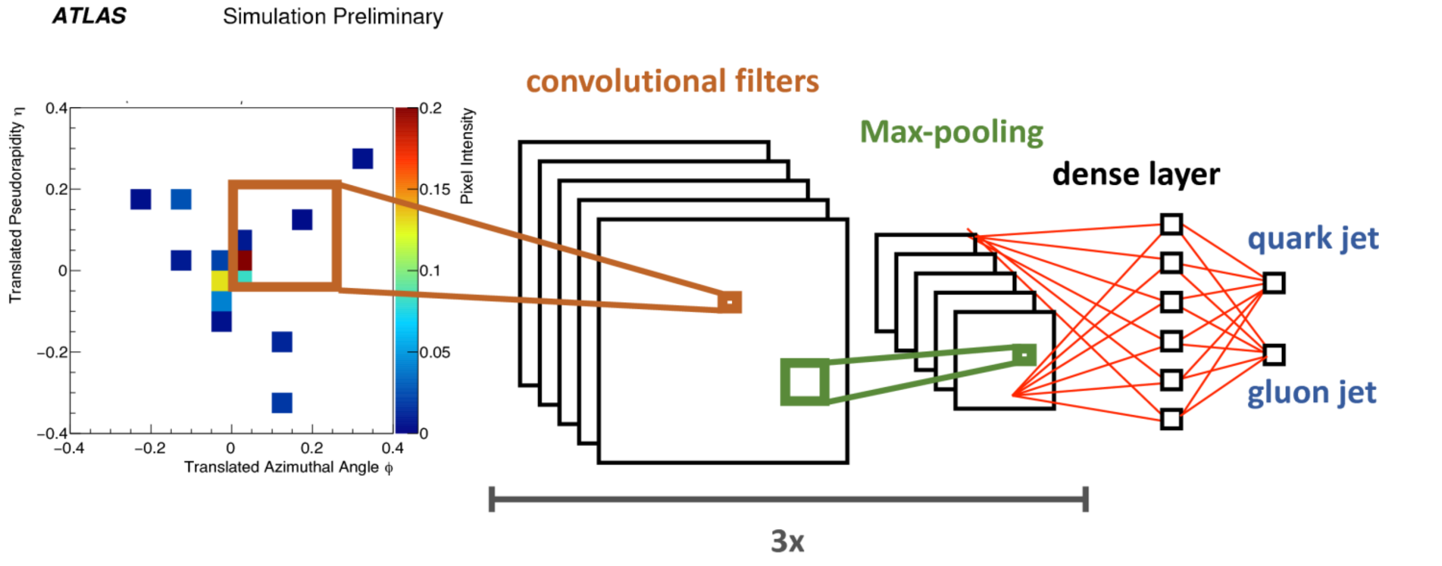
- ANN are composed of layers or artificial neurons interconnected by a weight factor
- The first layer is the input data, which usually is a collection of variables
- A learning step is completed once output is compared to the "true" answer (i.e. signal/background) if the results is wrong the weights are updated (using an error minimization process) and the sequence is repeated
- In every iteration the error decreases, once a threshold is achieving the model is declared as trained

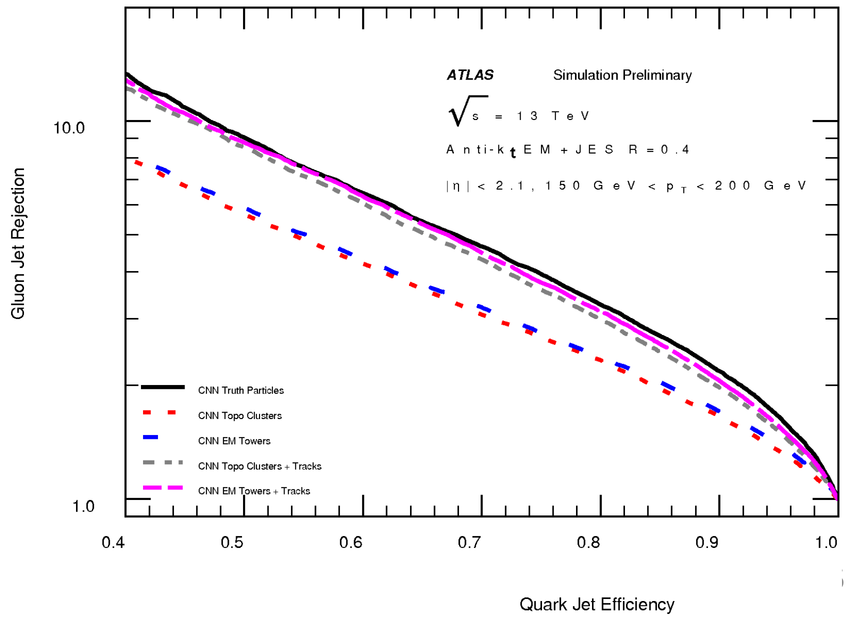
Convolutional neural networks in a nutshell

- CNN are a type of network designed to process data as images
- In HEP is extensively used to identify jets and classify according to origin and flavor



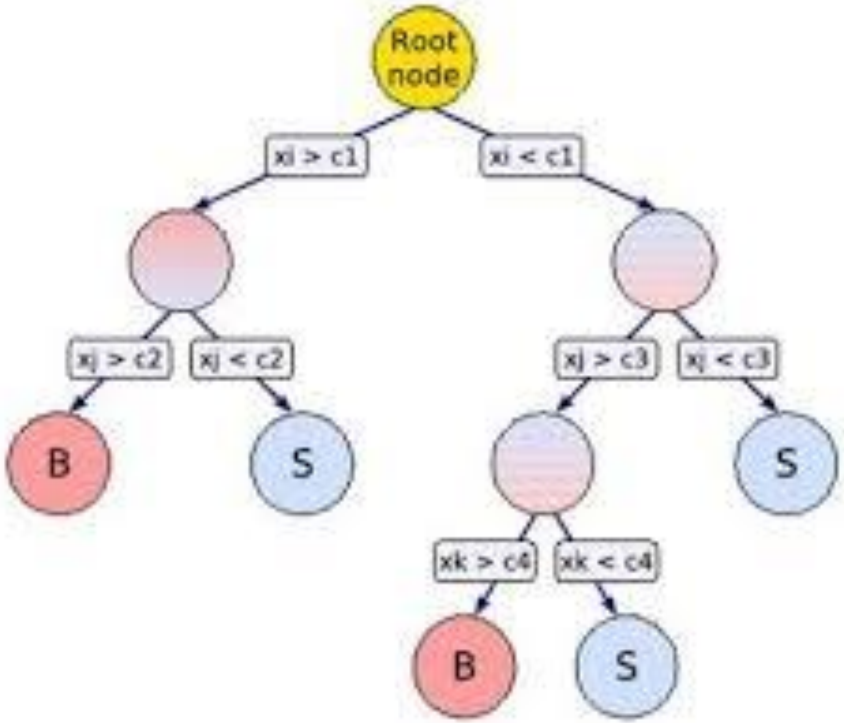




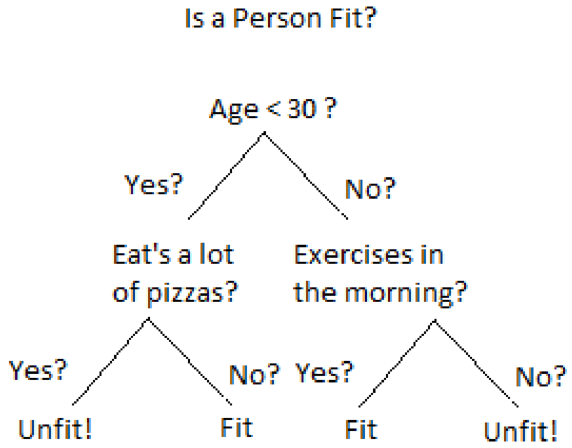


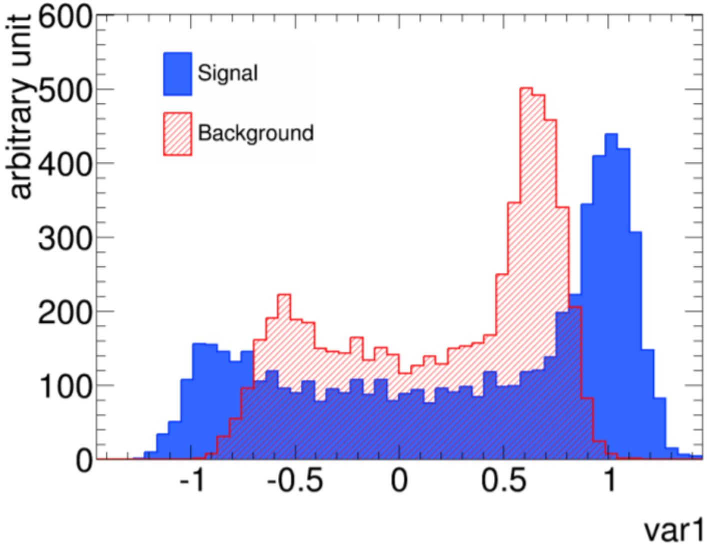
ATL-PHYS-PUB-2017-017

Boosted Decision Trees



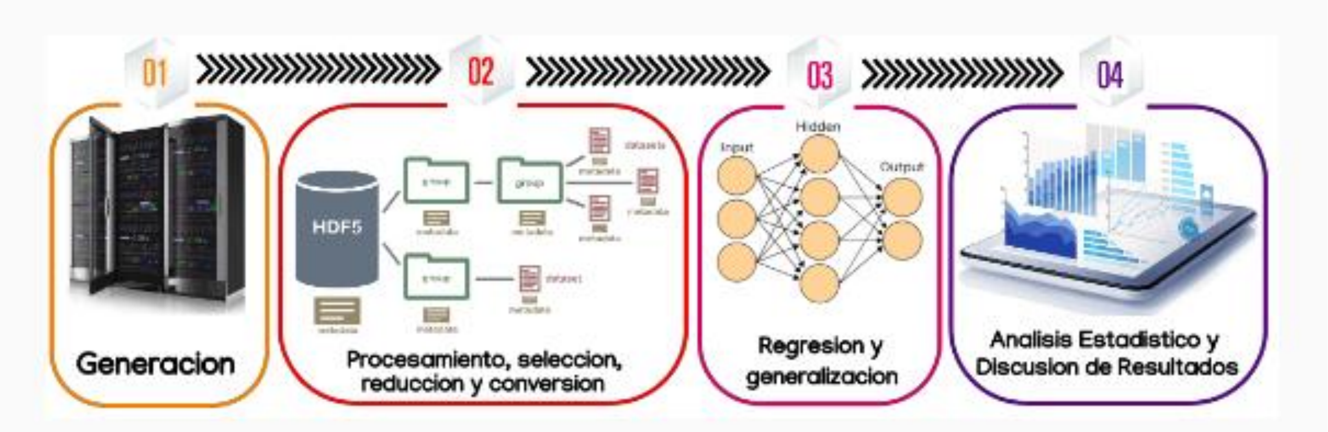
- Decision Trees are non-parametric supervised learning
- Decision Trees learn from the data
- Each terminal node representing the entropy to find a way to split the set until all data belongs to the same class
- The output of a BDT is a classifier between -1 and 1





Computational resources

- In order to use these alternative methods a combination of common CERN tools and GPU graphics card processing is needed
- Training data is obtained commonly from MC simulated samples (signal and background)
- Basic event selection to obtain input for neural network
- Training of NN with common tools (Tensorflow, Keras) and clusters of GPU cards for accelerating the process
- Interpretation of results



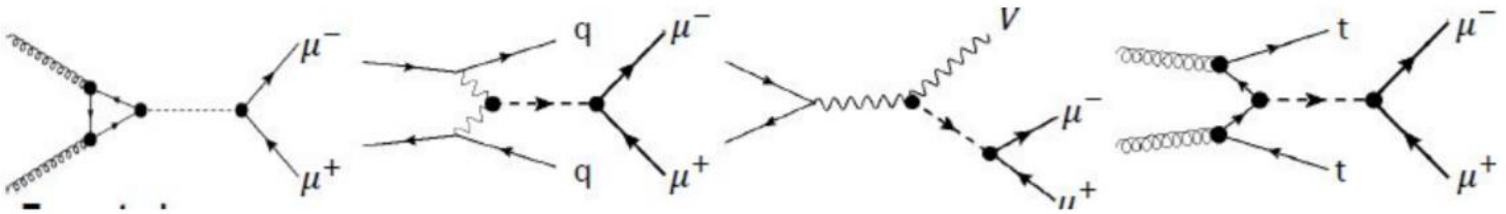
Rare processes

- As example we will mention three processes that share at least two of the following characteristics:
  - ❑ Low Cross section and Branching ratios
  - ❑ Large background contribution
  - ❑ Large object multiplicity

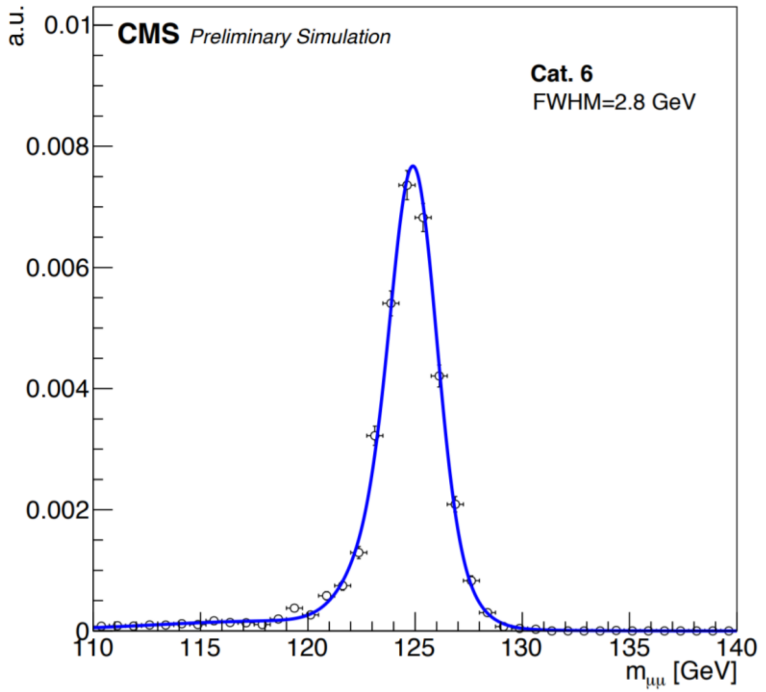
H to μμ

- Higgs decay to a pair of opposite muons would provide the first evidence of Higgs couplings to fermions (outside third generation)
- This search is difficult because the low branching ratio (~0.022) and large background contribution

	gluon-fusion	VBF	VH	$t\bar{t}H$
Cross sections (pb)	44.08	3.76	2.25	0.51



Simulated Higgs to dimuon

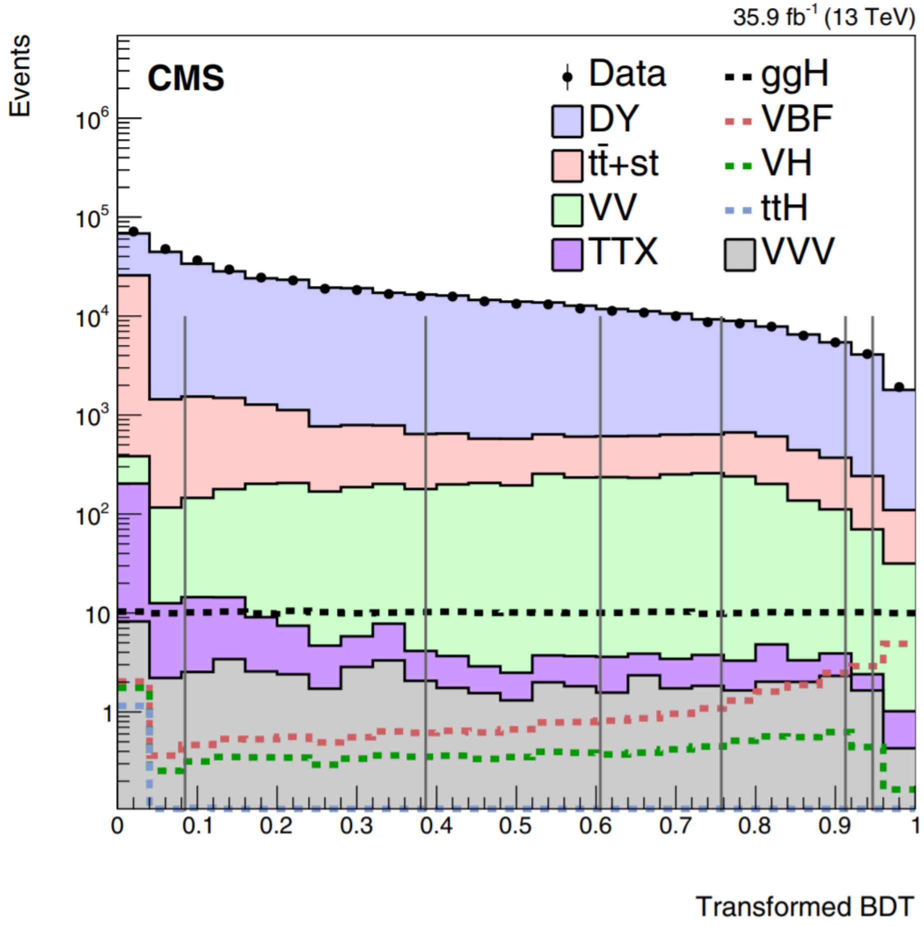


EPJ Web of Conferences 214, 06002 (2019)

H to μμ

- BDT classifier is built using following kinematical variables:

- The p_T and η of the dimuon system
- The $|\delta\eta|$ and $|\delta\phi|$ between the muons
- The η values of the two highest–p_T jets
- The masses of the two highest–mass dijet pairs
- The $|\delta\eta|$ between the jets in the two highest–mass pairs
- The number of jets with $|\eta| < 2.4$ and $|\eta| > 2.4$
- The number of jets identified as coming from b-hadrons [8]
- The missing transverse energy E_T.



EPJ Web of Conferences 214, 06002 (2019)

H to μμ

BDT response quantile [%]		Muon $\|\eta\|$ range	ggH [%]	VBF [%]	WH [%]	ZH [%]	$t\bar{t}H$ [%]	Signal	Bkg./GeV @125 GeV	FWHM [GeV]	$S/\sqrt{B}$ @ FWHM
0	8	0.0-2.4	4.9	1.3	3.3	6.3	32	21.2	3.13×10^3	4.2	0.12
8	39	1.9-2.4	5.6	1.7	3.9	3.5	1.3	22.3	1.34×10^3	7.2	0.16
8	39	0.9-1.9	10	2.8	6.5	6.4	5.2	41.1	2.24×10^3	4.1	0.29
8	39	0.0-0.9	3.2	0.8	1.9	2.1	3.5	12.7	7.83×10^2	2.9	0.18
39	61	1.9-2.4	2.9	1.7	2.7	2.7	0.3	11.8	4.37×10^2	7.0	0.14
39	61	0.9-1.9	7.2	3.3	6.1	5.2	1.3	29.2	9.70×10^2	4.0	0.31
39	61	0.0-0.9	3.6	1.1	2.6	2.2	0.9	14.5	4.81×10^2	2.8	0.26
61	76	1.9-2.4	1.2	1.5	1.8	1.7	0.2	5.2	1.48×10^2	7.6	0.11
61	76	0.9-1.9	4.8	3.6	4.5	4.4	0.7	20.3	5.12×10^2	4.2	0.29
61	76	0.0-0.9	3.2	1.6	2.3	2.1	0.6	13.1	3.22×10^2	3.0	0.28
76	91	1.9-2.4	1.2	3.1	2.2	2.1	0.2	5.8	1.04×10^2	7.1	0.14
76	91	0.9-1.9	4.4	8.7	6.2	6.0	1.1	20.3	3.60×10^2	4.2	0.35
76	91	0.0-0.9	3.1	4.0	3.8	3.6	0.9	13.7	2.36×10^2	3.2	0.34
91	95	0.0-2.4	1.7	6.4	2.5	2.6	0.5	8.6	96.0	4.0	0.28
95	100	0.0-2.4	2.0	19	1.5	1.4	0.7	13.7	83.4	4.1	0.48
0	100	0.0-2.4	59	61	51	52	49	253	1.30×10^4	3.9	

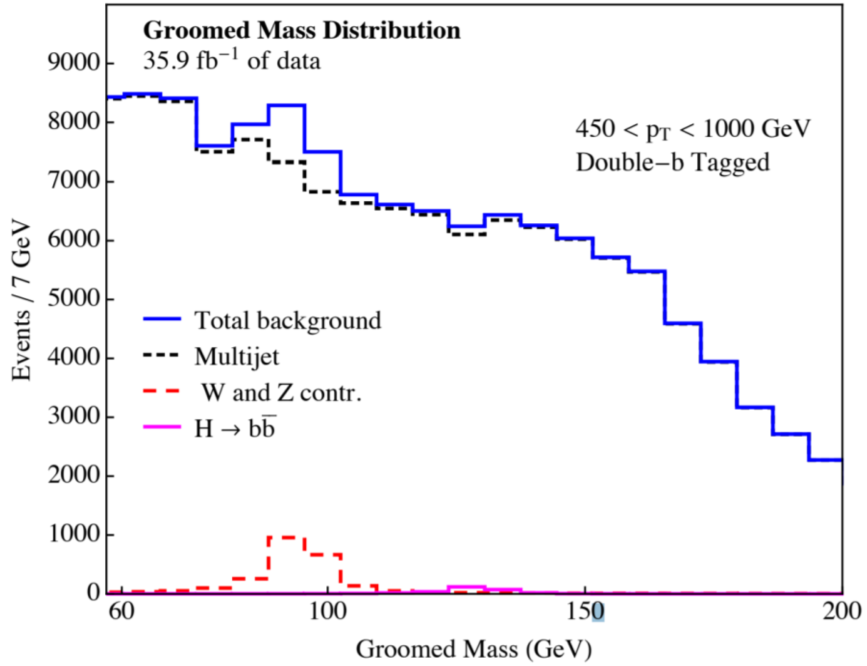
- In order to further increase the sensitivity, deep neural networks with several layers and multi-node architectures are being explored to improve the already good performance of the BDT

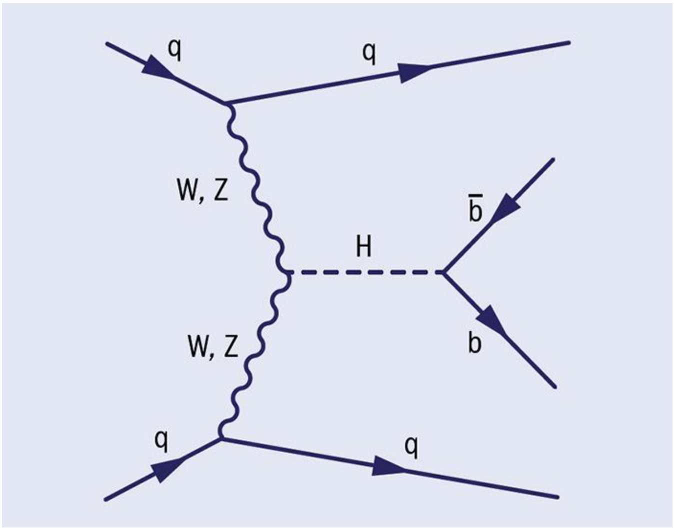
EPJ Web of Conferences 214, 06002 (2019)

H to bb

- Final state with largest branching ratio B(H->bb) ~ 58%
- Large and nearly irreducible background processes (QCD)
- High-pT Higgs production challenging to probe with cleaner final states (H->gammagamma, H->4l) due to lower branching ratios
- H->bb provide access to highest pT Higgs bosons at the LHC

10.1007/JHEP10(2018)101

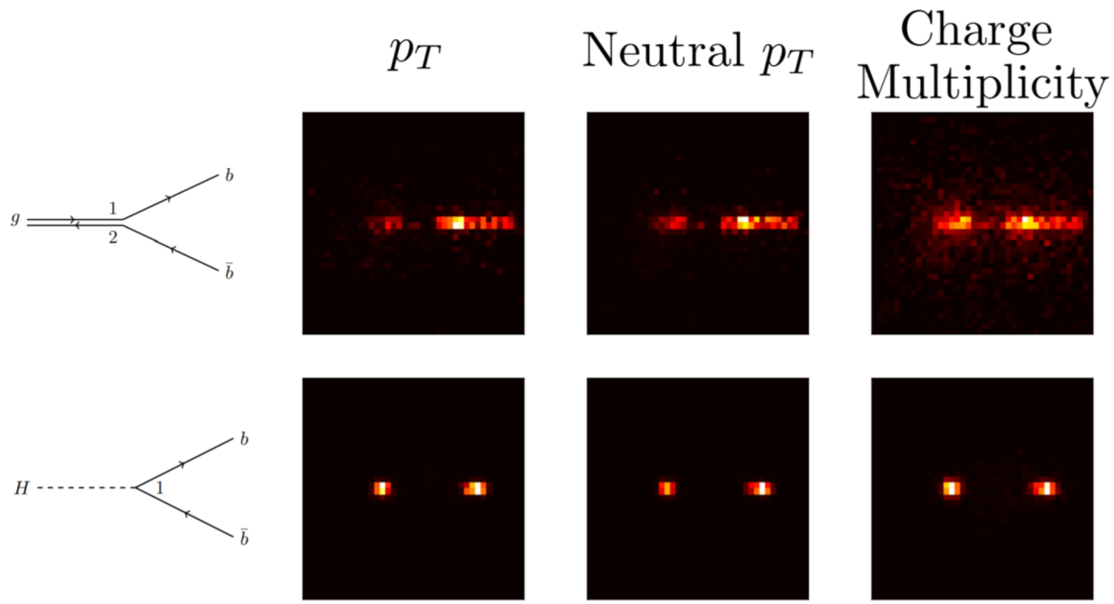


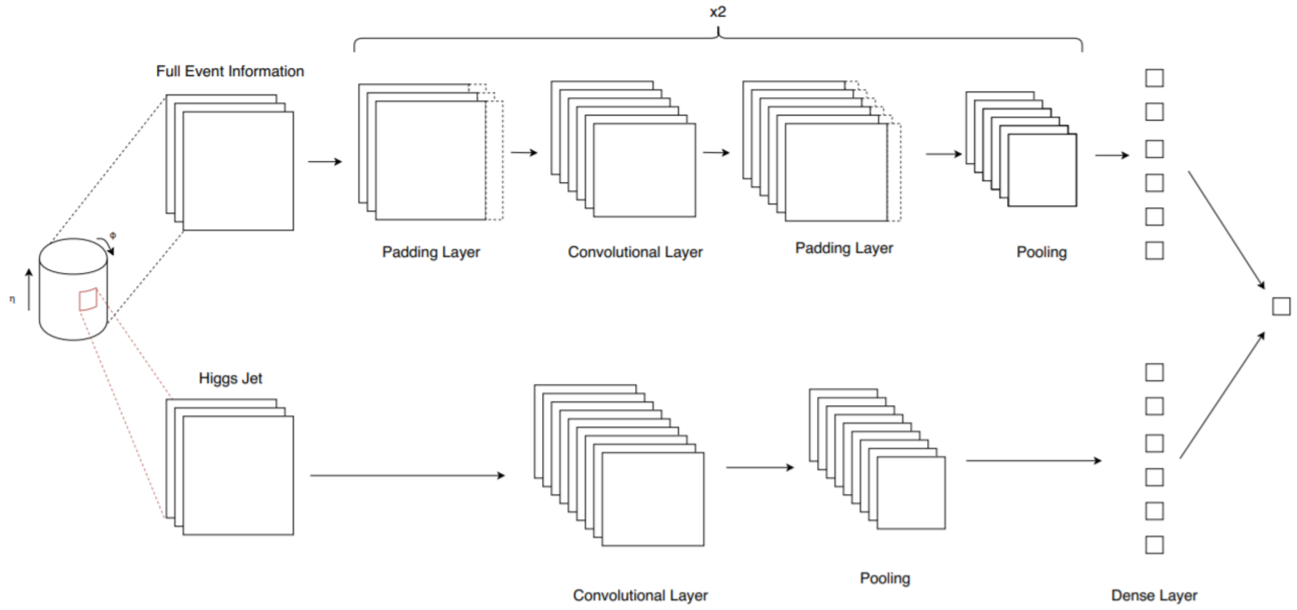


H to bb

- Two stream CNN used in this study
- The first stream uses the full event information
- The second stream uses the jet substructure information

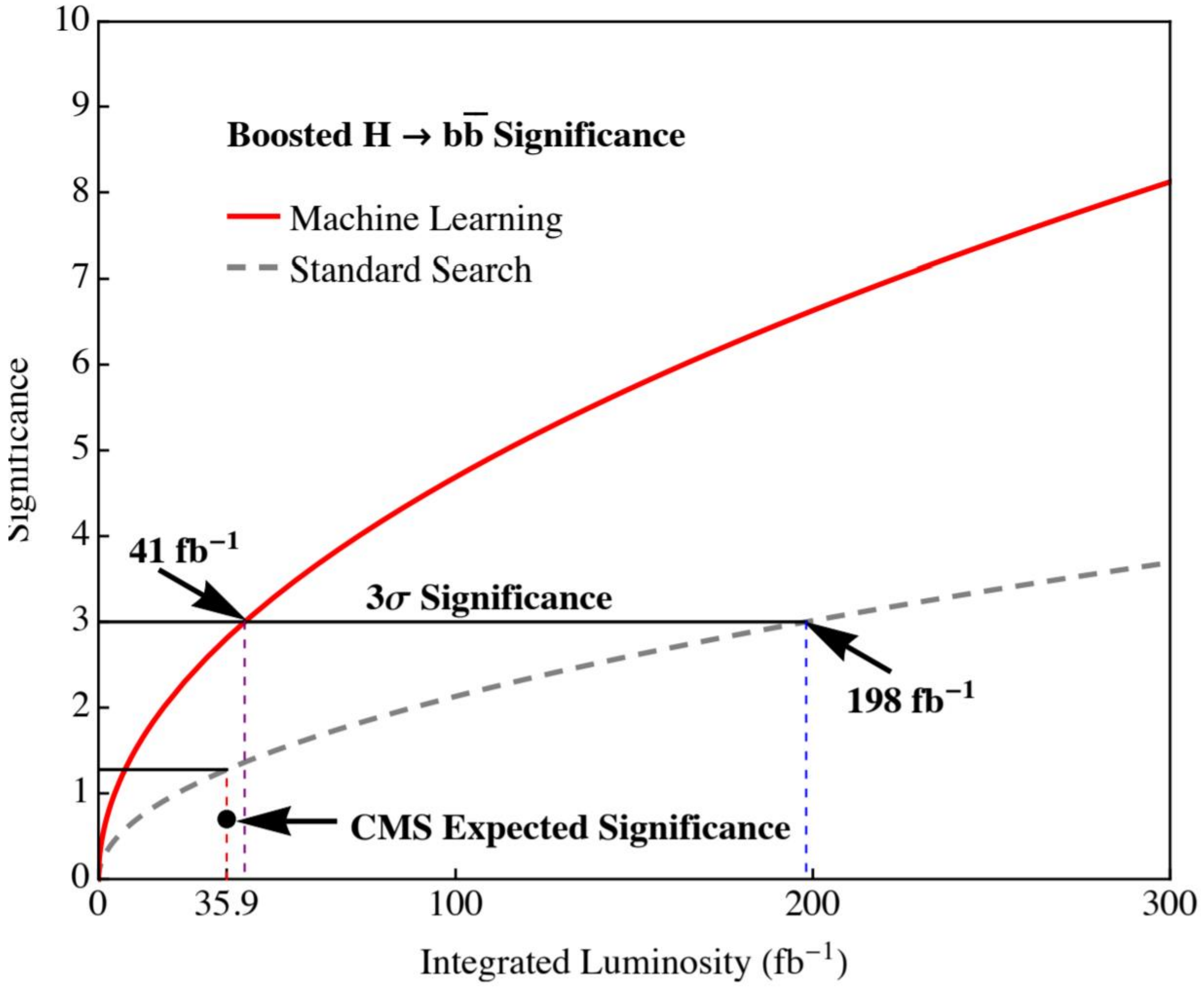
10.1007/JHEP10(2018)101





H to bb

[10.1007/JHEP10(2018)101](https://doi.org/10.1007/JHEP10(2018)101)

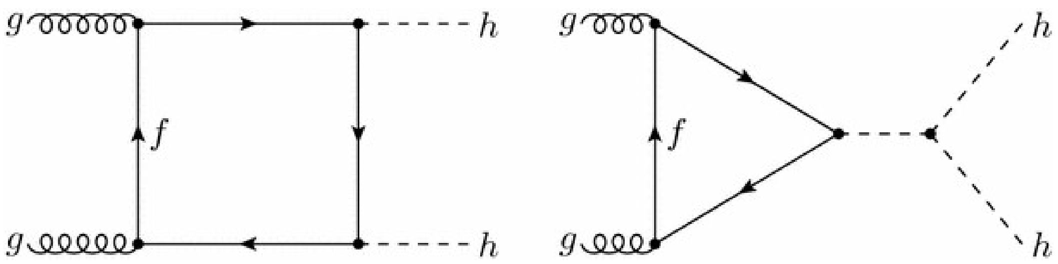


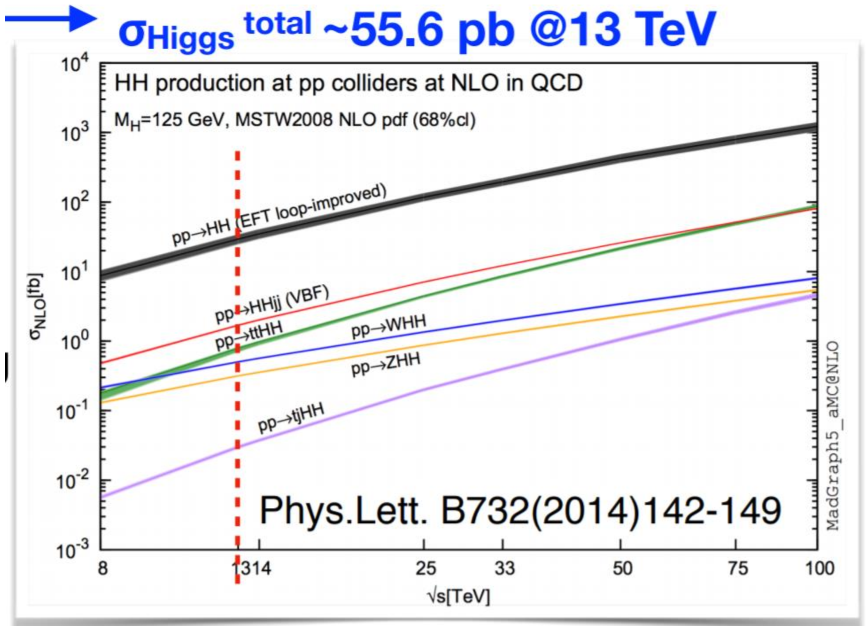
- Without ML “evidence” (3 sigma) may be achieved after LHC (2023) and “observation "may only be possible with HL-LHC

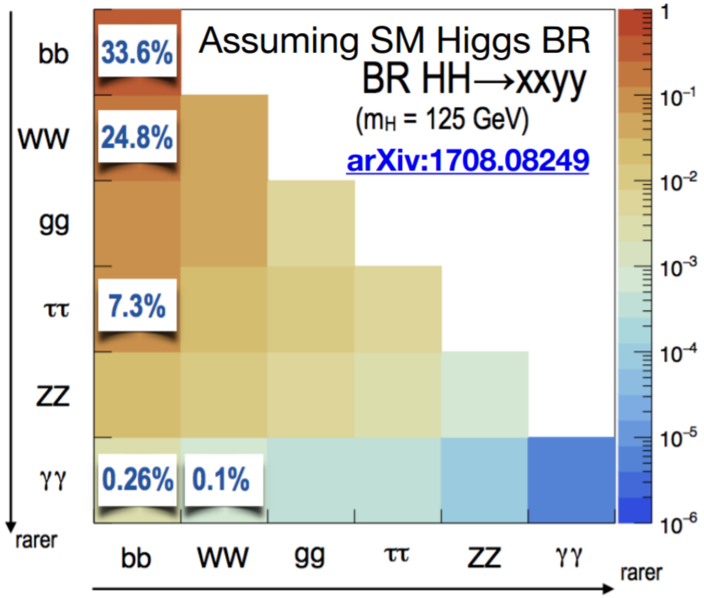
- With neural network evidence may be achievable with full Run 2 (2015-2018) 150pb-1 and observation possible well before the end of the HL-LHC

Di-Higgs search

- One of the most challenging measurements at the HL/HL-LHC
- Di-Higgs production measurements probes trilinear self-coupling of the Higgs boson
- Main SM production mechanism is the gluon fusion

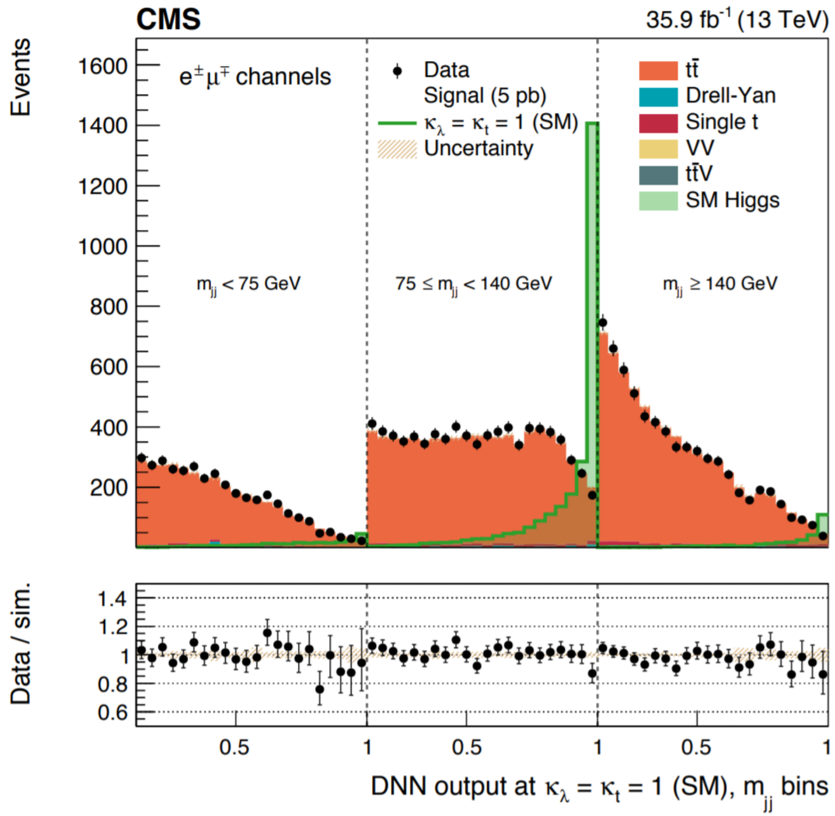


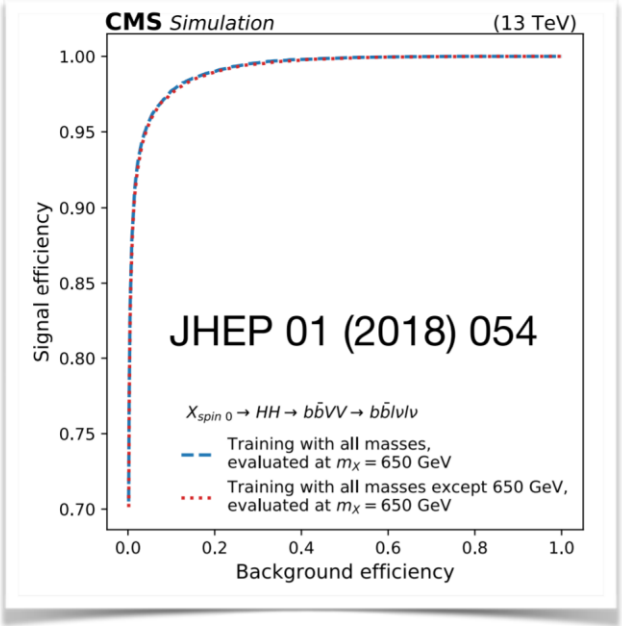




Di-Higgs search: bbVV

- Discriminating variables:
  - Deep neural network using Keras library and parameterized machine learning technique, to optimize sensitivity in one training for full range of signal hypothesis tested (m_X; κ_λ, κ_t)
  - Trained on event kinematics: m_{ll}, ΔR_{ll}, ΔR_{jj}, $\Delta\Phi(ll,jj\Phi)$, pT_{ll}, pT_{jj}, $\min(\Delta R_{jl})$, mT.

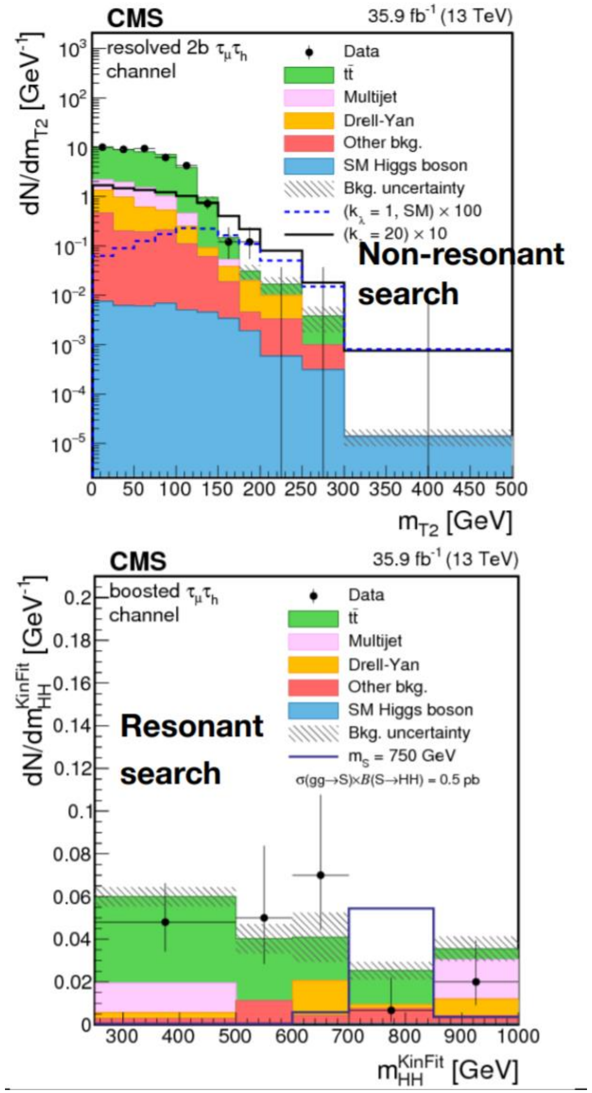




Di-Higgs search: bb τ τ

- Object selection:
  - A pair of tau leptons
  - Two b-tagget jets
- Signal region and discriminant variables
  - Both Higgs candidates' masses consistent with the expected reconstructed mass of the SM Higgs
  - Use of BDT to reject ttbar background

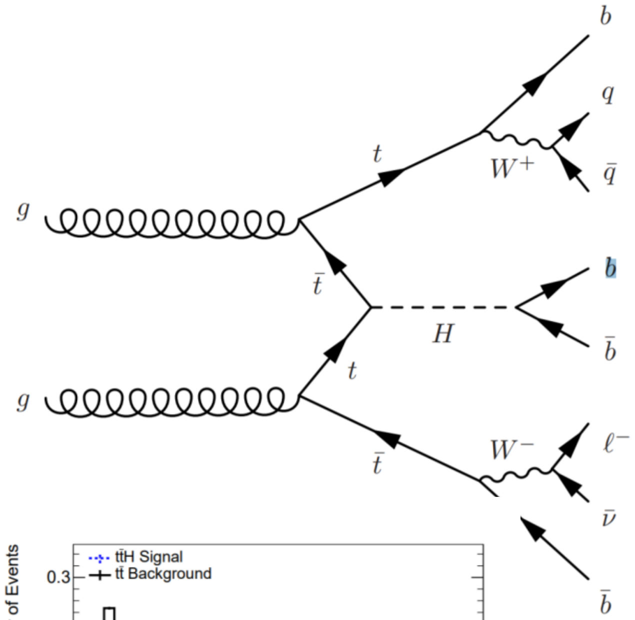
Phys. Lett. B 778 (2018) 101

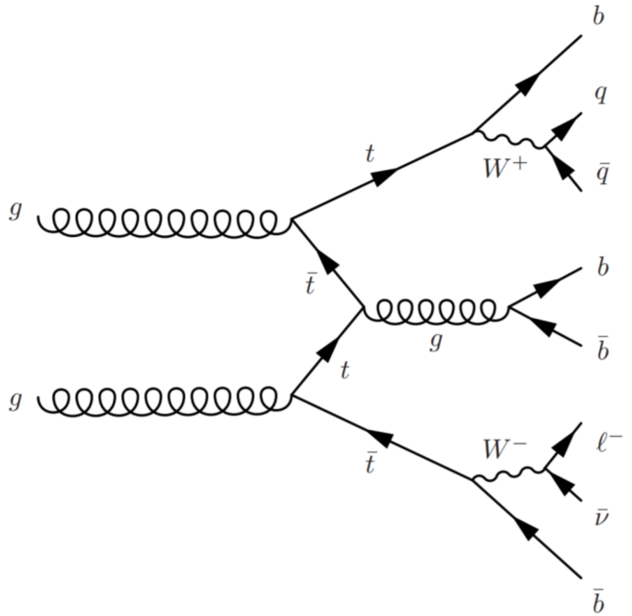


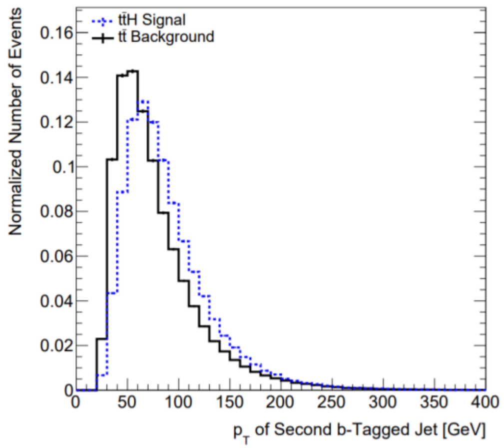
Higgs associated with top quark

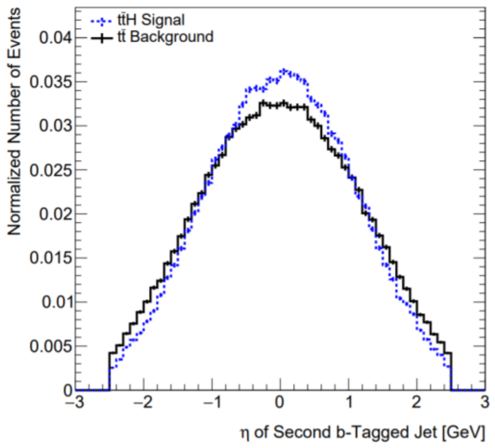
$$pp \to t\bar{t}h \; W^+bW^-\bar{b}b\bar{b} \to \ell\nu q q' b\bar{b}b\bar{b}$$

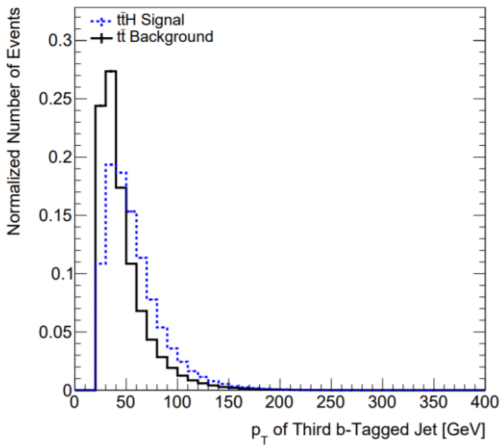
- Higgs production associated with a pair of top quarks (ttH) in a final state with WbWbbb
- 8 objects in the final state
- A collection of variables used to test different neural network arquitectures



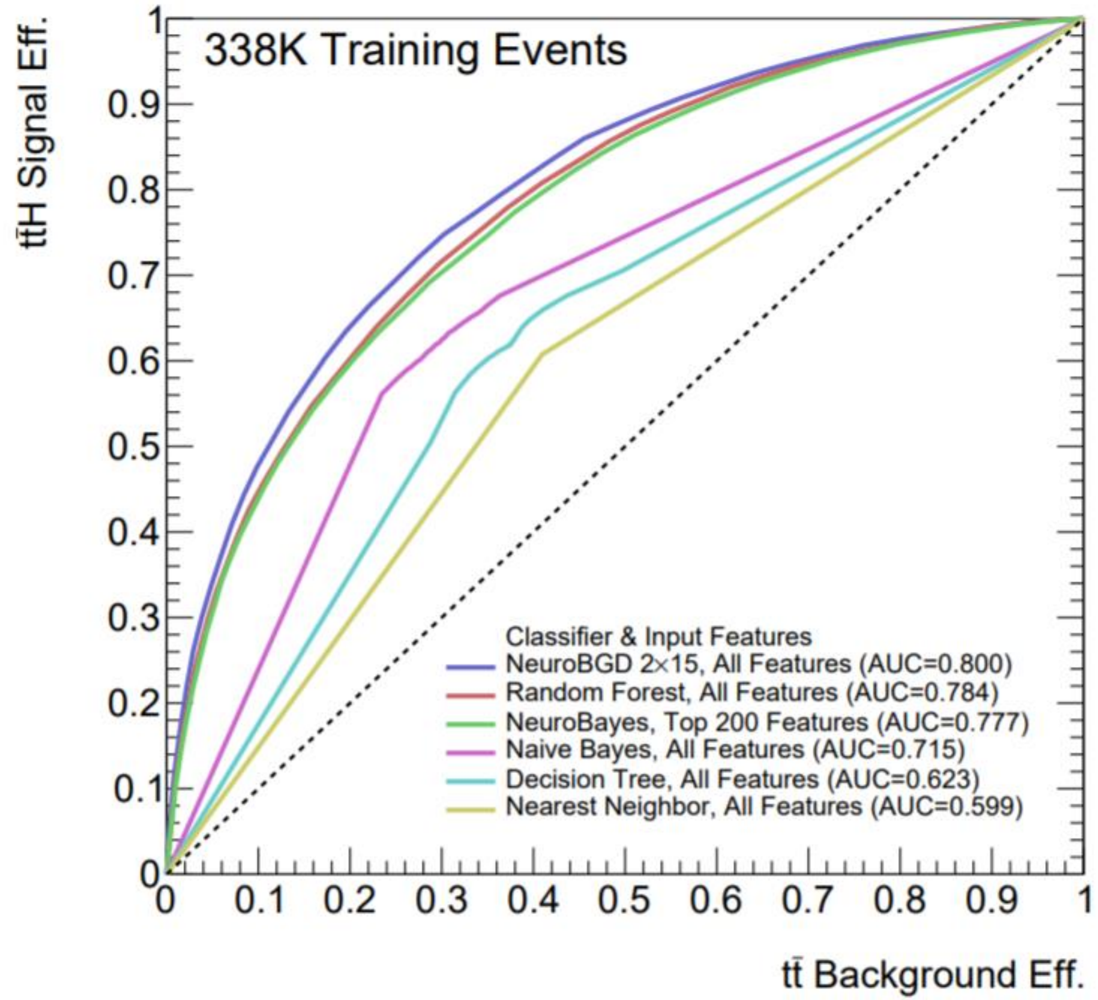




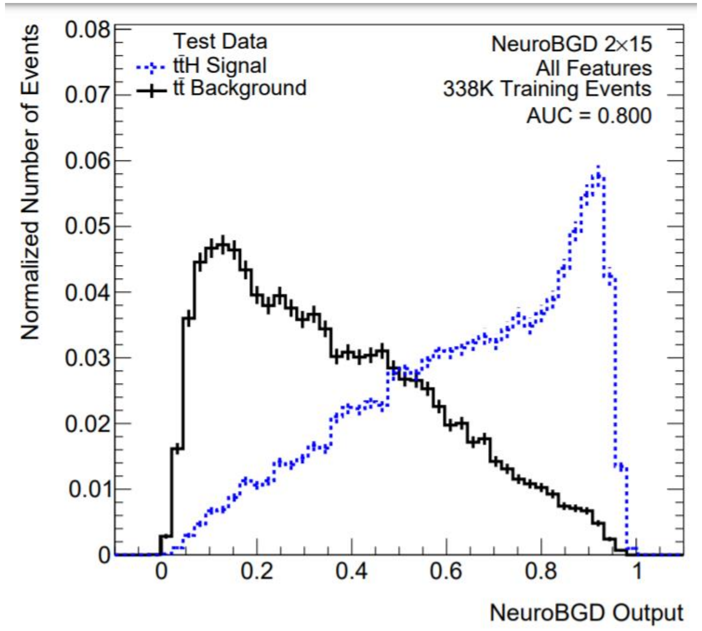


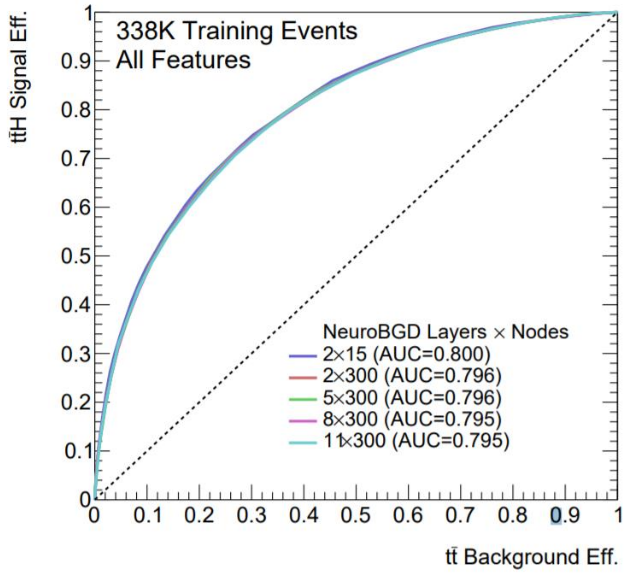


Higgs associated with top quark



- Several ANN comparison
- ROC curves are one of the best tool to compare performance
- Best result with NeuroBGD which is a multilayer feed forward network that uses tanh activation function for hidden layers





Conclusions

- The observation of fundamental processes such as the di-Higgs production and Higgs rare decays depend on the data accumulated in the following years and the development of new analysis techniques
- The integration of ML techniques for several final states increases signal purity and possibility of observation in the following years
- The scientific community needs to spend more computational resources and manpower to collaborate in such developments