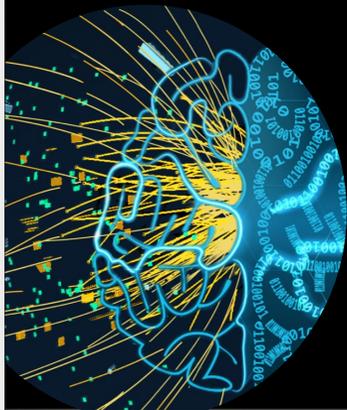


Looking for rare di-Higgs events at the LHC with Machine (Deep) Learning techniques

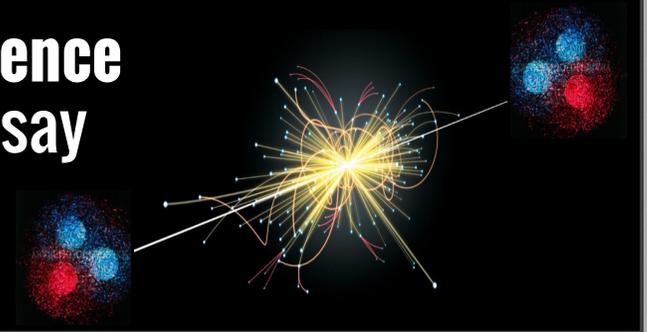
Brunella D'Anzi (Università and INFN, Bari (IT))

contact: brunella.danzi@ba.infn.it



Learning to Discover, AI and physics conference
Institut Pascal, Université Paris-Saclay, Orsay

April 28, 2022



Di-Higgs (HH) physics

Preface: discovery of the Higgs boson and measurements of its main properties

★ Test of the Brout-Englert-Higgs (BEH) mechanism

$$\mathcal{L}_{BEH} = \frac{1}{2} \partial^\mu H \partial_\mu H + [m_W^2 W^{\mu+} W_\mu^- + \frac{1}{2} m_Z^2 Z^\mu Z_\mu] \left(1 + \frac{H}{v}\right)^2 - \frac{1}{2} m_h^2 h^2 - \lambda_{HHH} v h^3 - \frac{1}{4} \lambda_{HHHH} h^4 + \frac{\lambda}{4} v^4$$

where $v \simeq 246 \text{ GeV}$, $m_H = \sqrt{2\lambda v^2}$, $\lambda = \lambda_{HHHH} = \lambda_{HHH}$

★ HH processes are used for measuring Higgs boson self-couplings C_{2V} , C_V , K_λ

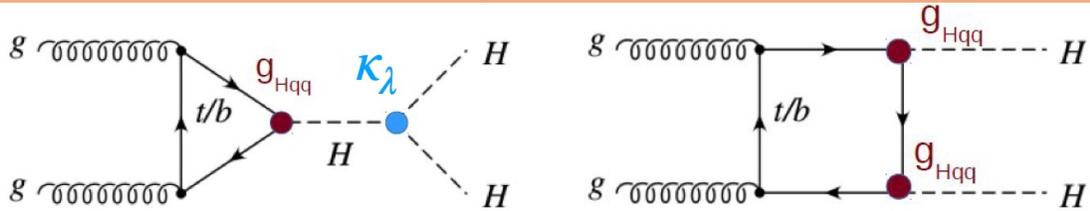
★ HH searches for testing Beyond Standard Model (BSM) physics

The image contains several physics diagrams and equations:

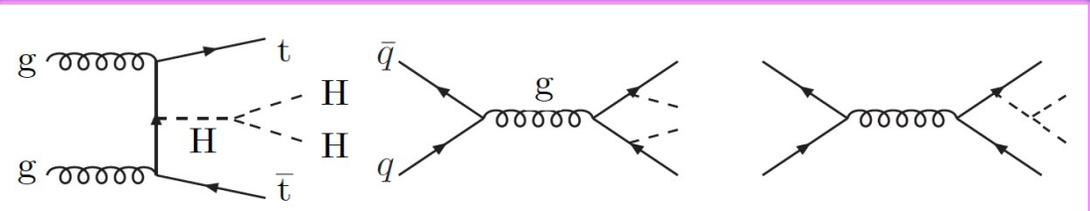
- Top Right:** Two circular diagrams showing gluon fusion (gg) and vector boson fusion (WW) processes.
- Middle Left:** A large oval diagram showing various production channels for a Higgs boson (H) from a gluon (g) or photon (gamma) via a loop of particles like top quarks (t), bottom quarks (b), or W/Z bosons.
- Middle Right:** Handwritten Lagrangian equations:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi} \not{D} \psi + h.c. + \chi_i y_{ij} \chi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$
- Bottom Left:** A circular diagram showing a Higgs boson (H) interacting with two vector bosons (V) via a loop, labeled with c_{2V} .
- Bottom Middle:** A circular diagram showing a Higgs boson (H) interacting with a vector boson (V) via a loop, labeled with c_V .
- Bottom Right:** A circular diagram showing a Higgs boson (H) interacting with two Higgs bosons (H) via a loop, labeled with K_λ .

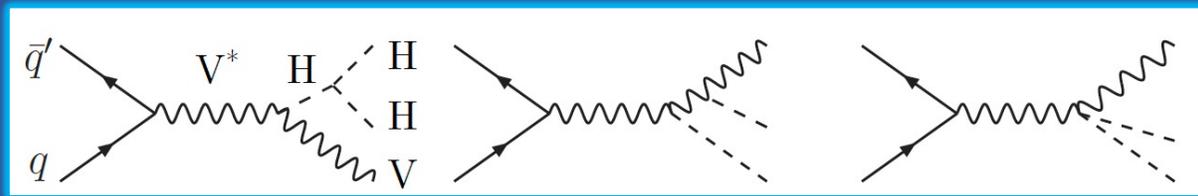
Di-Higgs production mechanisms



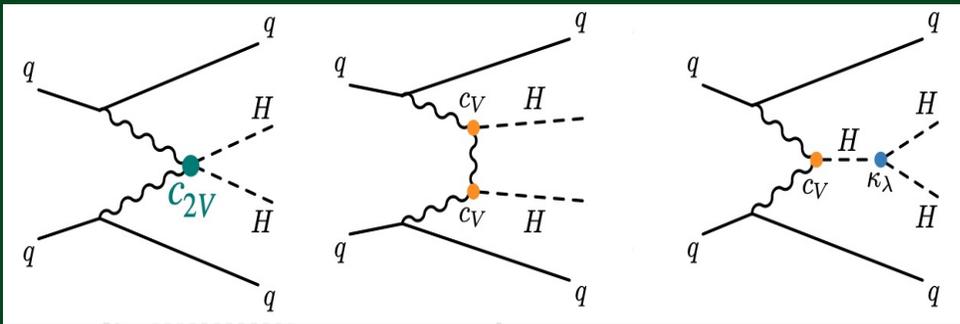
Gluon-gluon fusion (ggF)



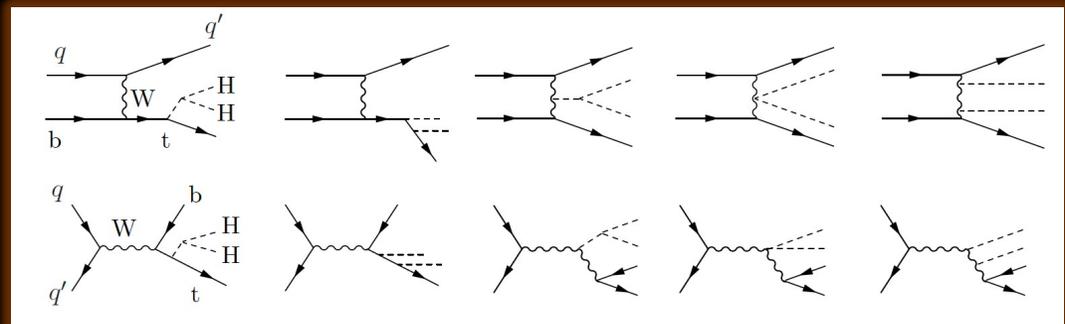
Top quark pair associated production/ Double Higgs bremsstrahlung off top quarks ($t\bar{t}HH$)



Double Higgs-strahlung/Vector boson associated production (VHH , $V = Z, W^+, W^-$)



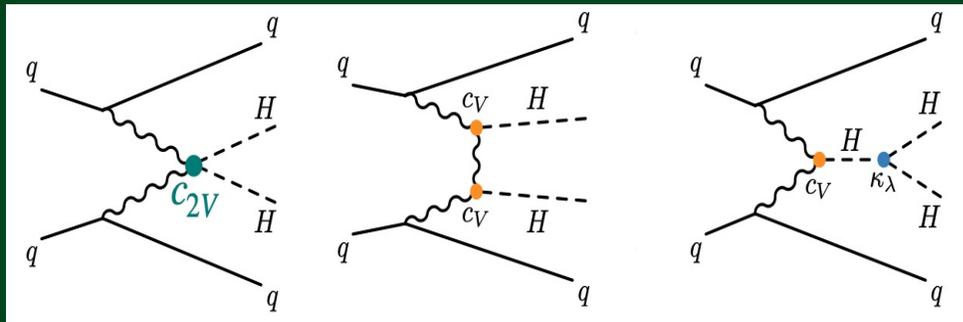
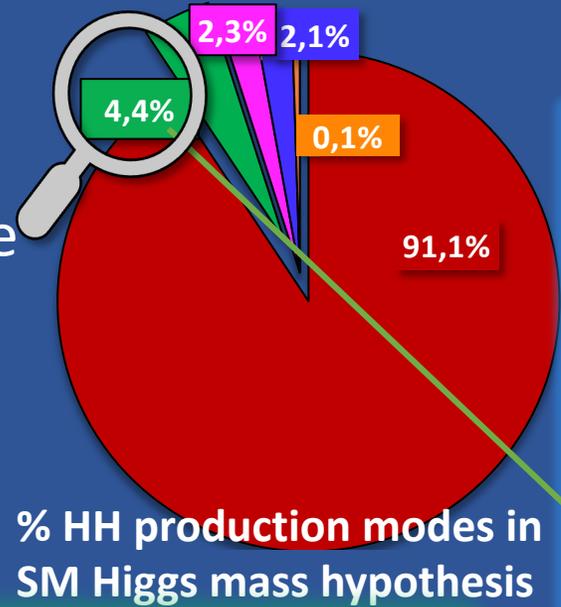
Vector boson fusion (VBF)



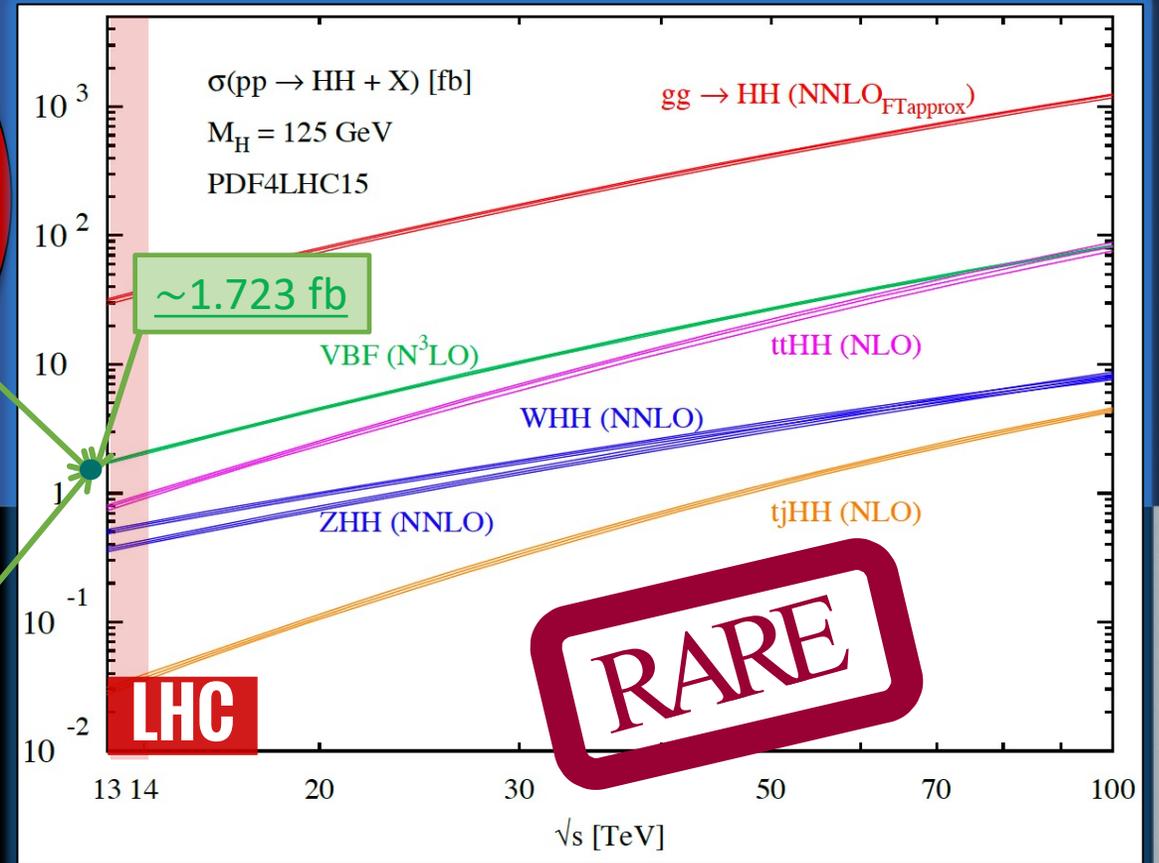
Single top quark associated production ($tjHH$)

Vector Boson Fusion (VBF) production mechanism

★ It is a unique mean for probing the $VVHH$ ($V=Z^*, W^\pm$) C_{2V} Higgs self-coupling!

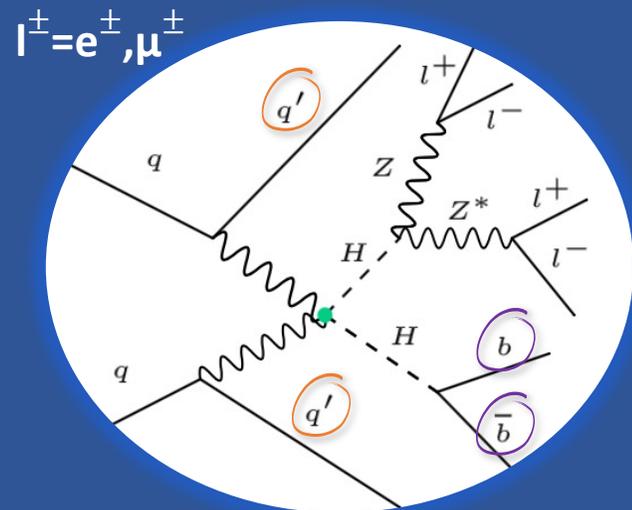


Vector boson fusion (VBF)



CERN Technical report, 2019

Signal sample: HH to four charged lepton + 2 b-jets final state



Sample generated using MadGraph5_aMC@NLO + PYTHIA8 + TuneCP5

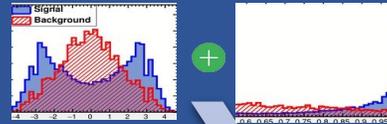
- ★ $H \rightarrow b\bar{b}$ has the highest Higgs boson BR decay mode 5.75×10^{-1}
- ★ $H \rightarrow ZZ^* \rightarrow l^+l^-l'^+l'^-$, $l = e, \mu, \tau$ ($\text{BR} = 2.79 \times 10^{-4}$) has an excellent four-leptons invariant mass resolution, one of the best Signal/Bkg ratio decay channel

Data analysis strategy

Monte Carlo event generation



Signal region definition



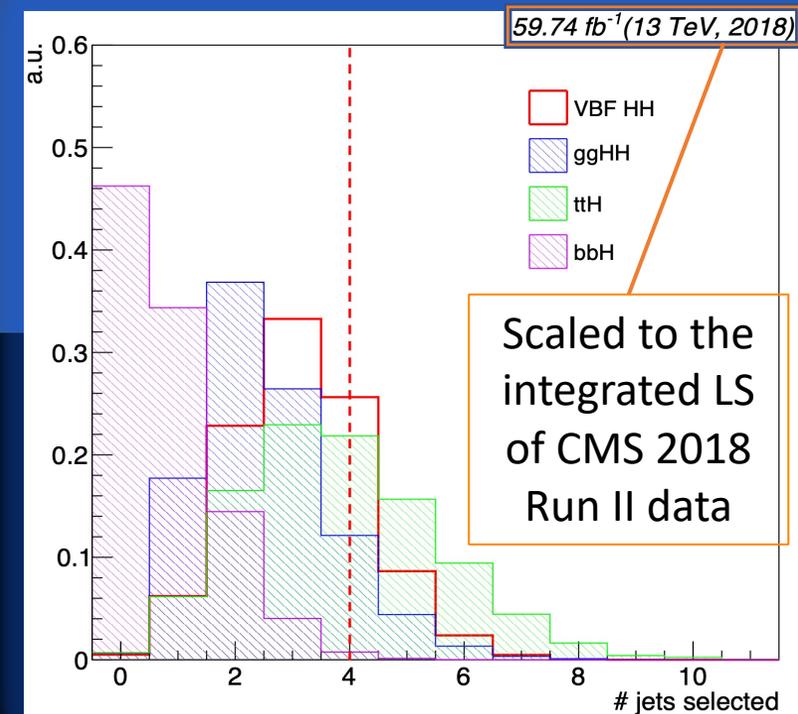
Optimize the separation using AI tools



Ready for statistical analysis!

- 7 SM and BSM VBF HH signals - vary C_V, C_{2V}, K_λ ;
- SM irreducible backgrounds;
- SM HH ggF production.

- One good Primary Vertex;
- Z candidates with $12 < m_{ll(\gamma)} < 120 \text{ GeV}/c^2$;
- Channels $4e, 4\mu$ and $2e2\mu$ are selected separately;
- Higgs candidate from ZZs;
- $4 \leq \# \text{ good jets} \leq 6$, with $\Delta R_{j1} > 0.3, |\eta| < 4.7, \text{ high } p_T$.



Application of binary classification AI techniques

Hundreds of multiple training jobs to ReCaS Bari computing center (CPU only-usage)

- Pre-processing of the events and merging of the three (4e, 4 μ and 2e2 μ) decay modes
- Discriminator hyper-parameters scan (including parameters of regularization functions such as the dropout)
- Inclusion of event weights ($\sigma \times B.R.$)

~ 30 minutes for getting results using different signal samples!

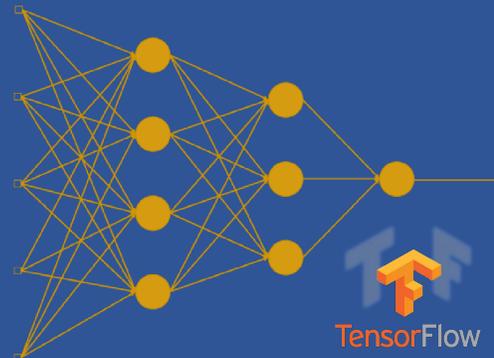
Evaluation of each ML/DL model

Choose the best model configuration in terms of the maximum purity (π) = TP/(TP+FP) x signal efficiency (ϵ_S) = TP/(TP+FN) metrics on the test data set

$\kappa_\lambda, C_V, C_{2V}$	$\sigma \cdot BR$ [fb]	Signal samples name
1,1,1	0.00053581950	VBFHHTo2B2Z_CV_1_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8
1,1,2	0.0044084790	VBFHHTo2B2Z_CV_1_C2V_2_C3_1_TuneCP5_13TeV-madgraph-pythia8
2,1,1	0.00044277300	VBFHHTo2B2Z_CV_1_C2V_1_C3_2_TuneCP5_13TeV-madgraph-pythia8
0,1,1	0.0014277825	VBFHHTo2B2Z_CV_1_C2V_1_C3_0_TuneCP5_13TeV-madgraph-pythia8
1,0.5,1	0.0033560910	VBFHHTo2B2Z_CV_0.5_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8
1,1.5,1	0.020470230	VBFHHTo2B2Z_CV_1.5_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8
1,1,0	0.0014277825	VBFHHTo2B2Z_CV_1_C2V_0_C3_1_TuneCP5_13TeV-madgraph-pythia8

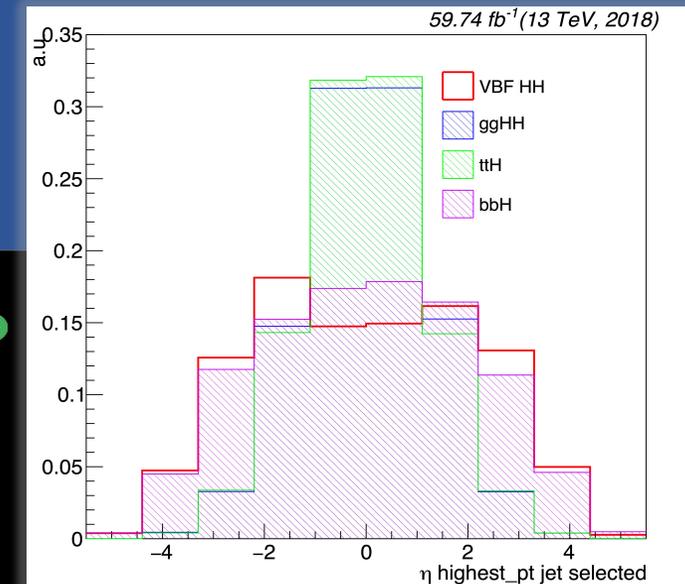
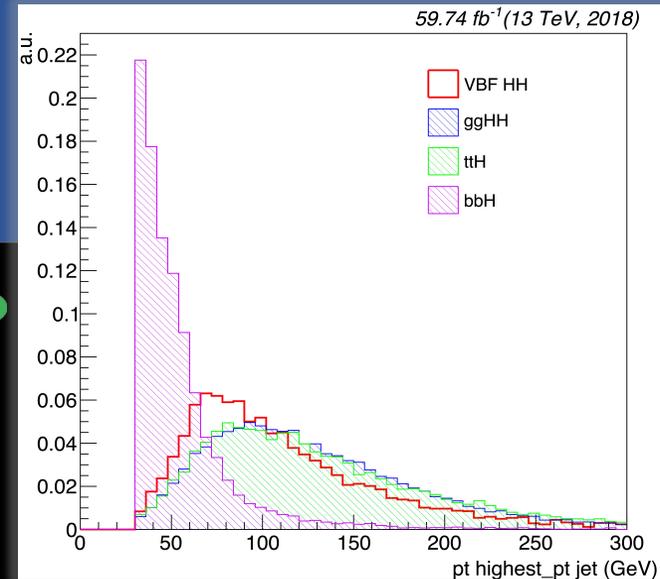
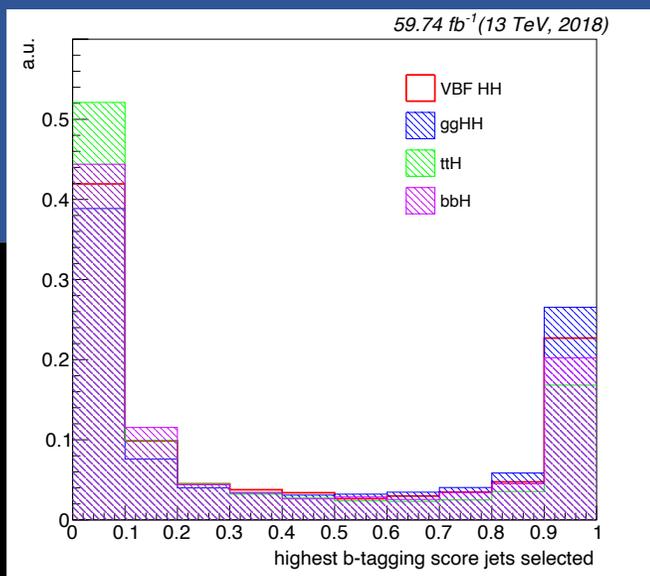
Feed-forward Deep Neural Network (DNN)

- Four charged leptons p_T, η, ϕ
- Two highest p_T jets p_T, η, ϕ , b-tag score
- Four highest p_T jets p_T, η, ϕ
- Size of train (20% validation) set = 207065
- Size of test set = 51786



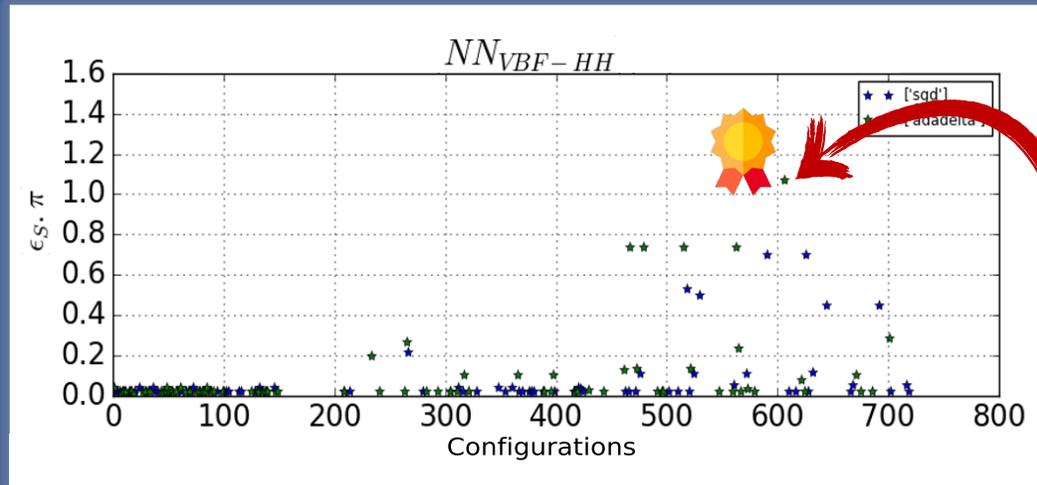
Signal/Background label

ACAT2021 lightning talks

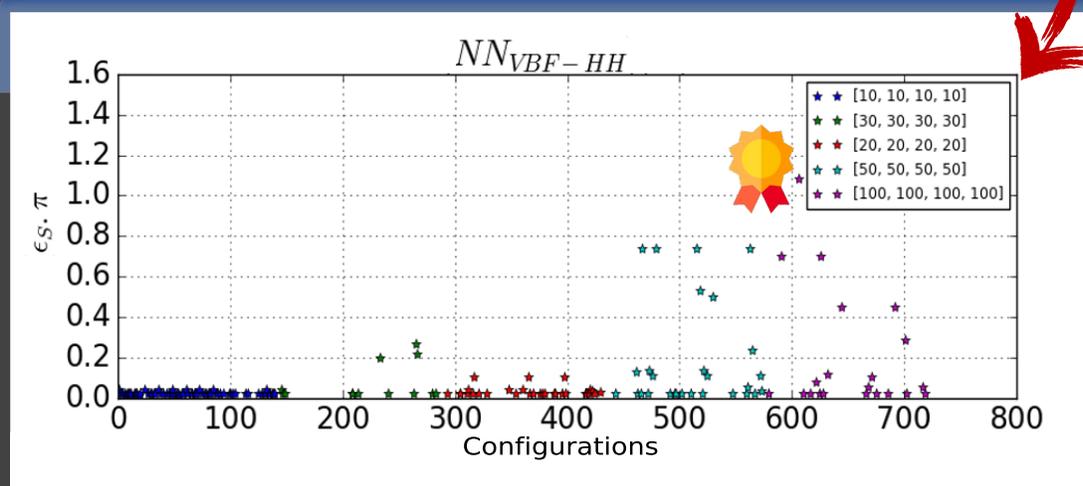
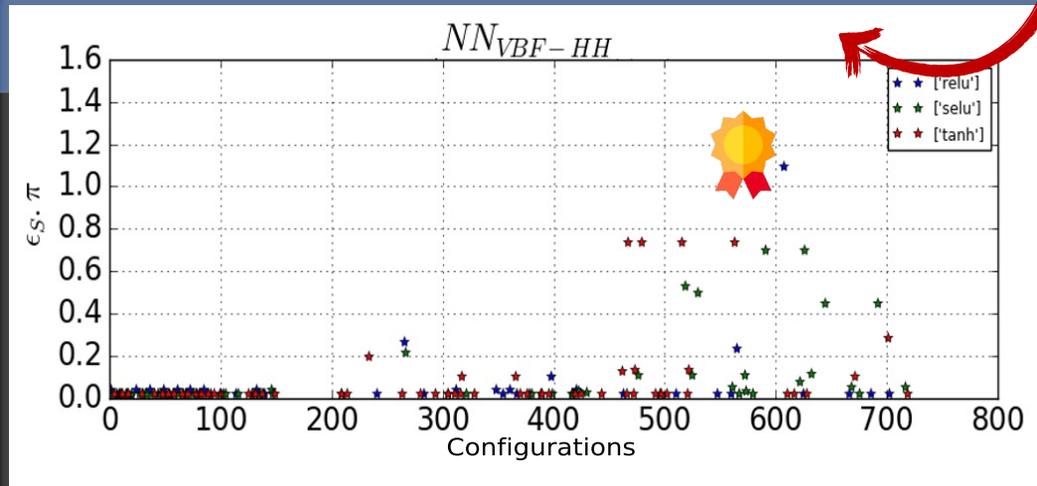


Algorithm with an ancient history as explained in [Bruce Denby's talk](#) and very popular today!

DNN parallel training



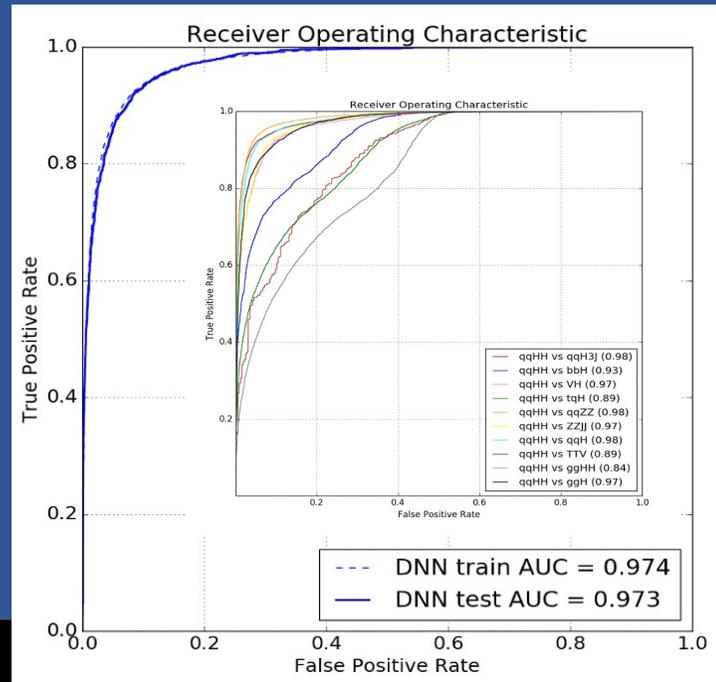
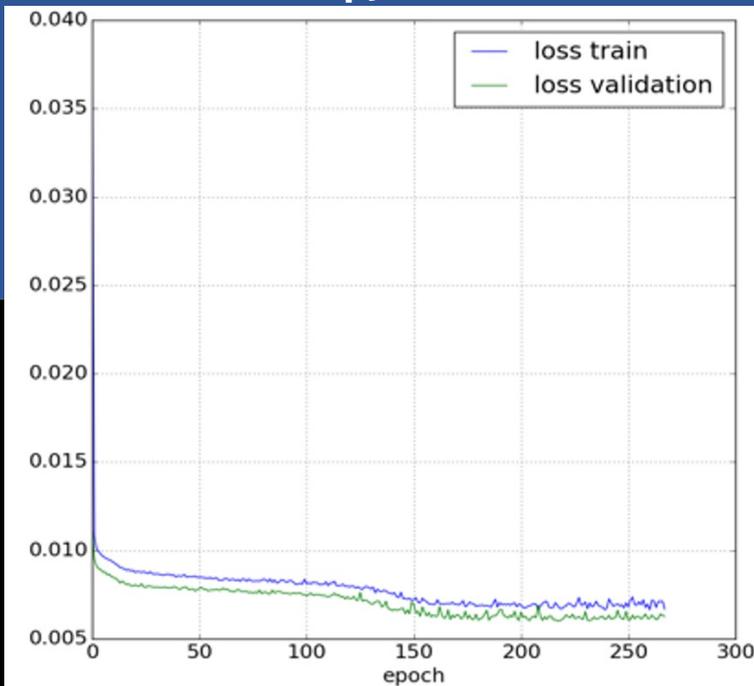
NN hyper-parameters	Tested options
Training of more than 700 DNNs configurations!	
Dropout rate	10%, 20%
Topologies	10:10:10:10, 30:30:30:30, 20:20:20:20, 50:50:50:50, ...
Early stop	50, 100, 600, 3000
Minimizer	SGD, Adadelta
Batch size	5, 32, 64, 128, 786
Neuron	ReLU, SeLU, Tanh



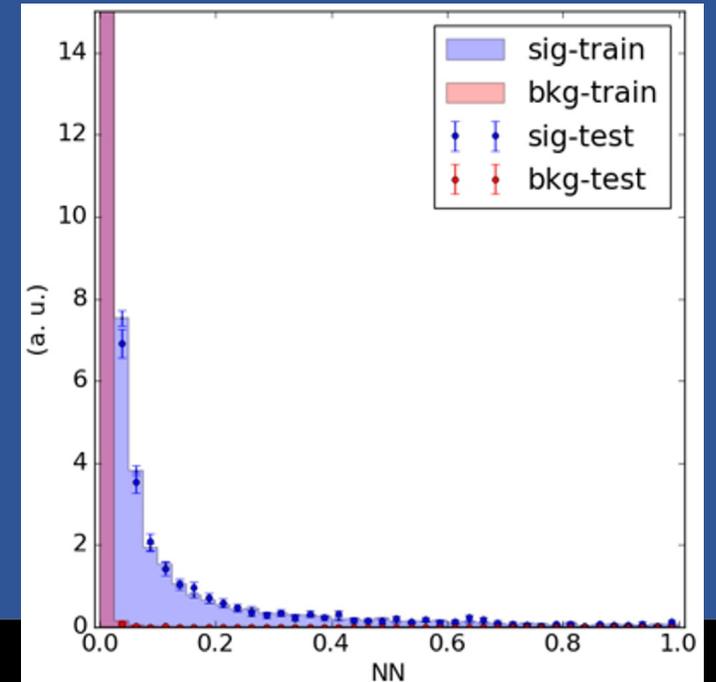
Best DNN model discriminating performance

No overfitting issues due to regularization!

Cross-entropy loss function



Total and partial ROCs



DNN score during the evaluation phase

An home-made optimizer procedure gave already good results!



Random Forest (RF) algorithm



Particle Physicists at LHC experiments really like Decision Trees and Boosted Decision Trees-related binary classification algorithms.

We haven't chosen to use a RF because of the location of this conference!

Nevertheless, we gave a chance to the RF looking at the algorithm definition including the growing of a large number of individual decision trees.

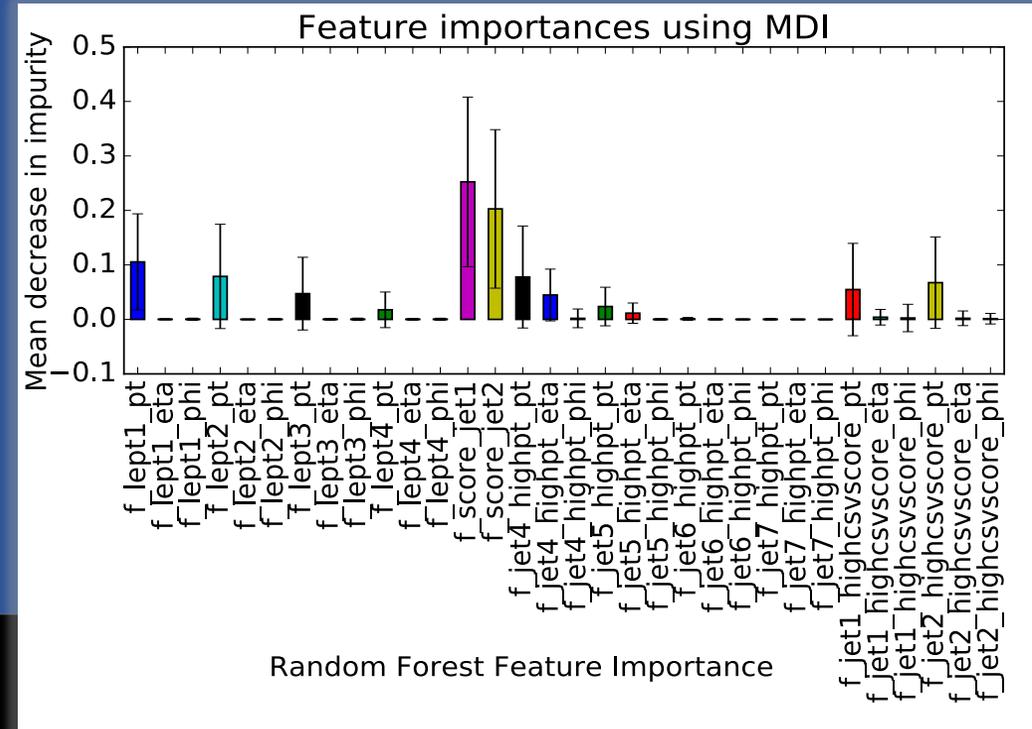
Stay tuned for future comparisons!



Scikit-learn

Random Forest training

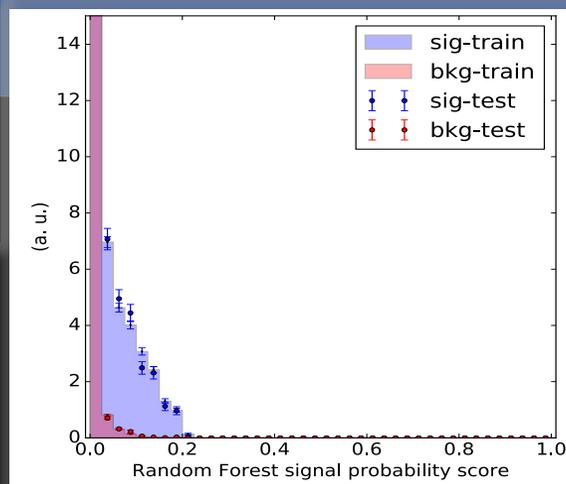
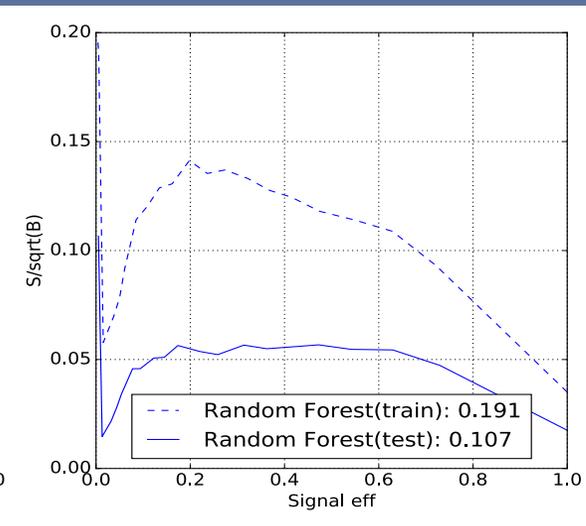
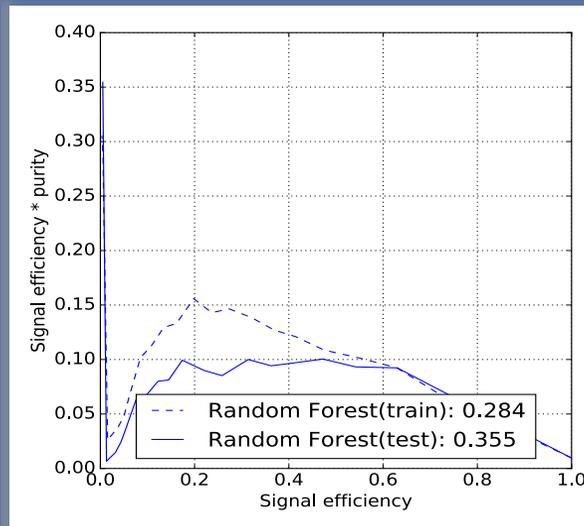
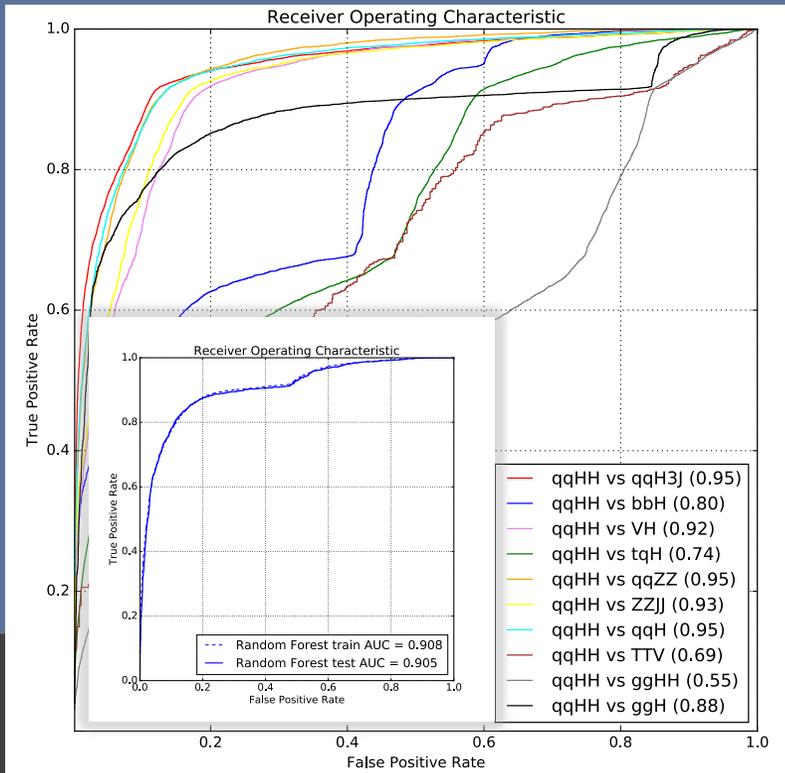
- We computed an estimate of the predictive power of each feature...



...compatible with the known physical distributions!

Best model hyper-parameters: split betw train/test: 0.8, criterion: ['gini'], n_estimators: 500, max_depth: 5, min_samples_leaf: 300, min_samples_split: 200, max_features : 5

Best RF model discriminating performance



Actually we got worst results but we used the algorithm for “training” ML newcomers!

We fixed some hyper-parameters:
 bootstrap: False, preproc: ['none'], split
 betw train/test: 0.8 etc.



Jupyter Notebook VBF Higgs-related
 exercises [Grid search optimizer]

Summary

- We presented a HEP scenario at LHC in which implementing an efficient discriminant for separating a very low-rate signal from a huge background is extremely important.
- We investigated the usage of a DNN and a RF (to be IMPROVED) via open-source platforms to boost the sensitivity to not-yet investigated HH decay mode via VBF production mechanism.
- We showed the advantage of implementing a highly parallelizable AI algorithms training encouraging current LHC analyses to use this approach.

Thank you for the attention!

