Signal to background discrimination for the production of double Higgs boson events via vector boson fusion mechanism in the decay channel with four charged leptons and two b-jets in the final state at the LHC experiment

B. D'Anzi1, G. Miniello1, N. De Filippis2, W. Elmetenawee2

1University of Bari and National Institute for Nuclear Physics (INFN) Bari, Italy 2Politecnico of Bari and INFN Bari, Italy
Motivation

- At the LHC experiment, the non-resonant double Higgs (HH) production via vector-boson fusion (VBF) represents the unique means to probe the VHH (V=Z,W⁺⁻) Higgs coupling (C_{2V}).

- A rare signal cannot be separated efficiently from huge backgrounds by applying a few-observables cut-based selection. Here, a deep learning algorithm is used to decide whether event is more signal- or background-like. They are easy to implement, but require optimization and validation!
H → b̅b: the highest Branching ratio (BR)
H → ZZ* → l+ l- l'+ l'− (l, l′ = e, μ): one of the best signal to background ratio (S/B)
2. Analysis strategy

- Event selection (generator level)
  - At least one Primary Vertex;
  - Z candidates $12 < m_{l(\ell)} < 120 \text{ GeV}/c^2$;
  - ZZ candidates are built from a pair of Z candidates which do not have common leptons (non-overlapping);
  - SM Higgs candidate from ZZ pairs, channels 4e, 4\(\mu\) and 2e2\(\mu\) selected separately.

- VBF signal region (SR)
  - Full selection of $H \rightarrow 4\ell$;
  - Four charged leptons;
  - Number of jets $\geq 4$;
  - $\Delta R_{\ell\ell} > 0.3$, $|\eta| < 4.7$.

- Backgrounds (bkgd)
  - SM single Higgs processes (irreducible);
  - HH gluon-gluon fusion (ggF) events (irreducible);
  - QCD backgrounds.

A rare signal cannot be detected by standard backgrounds by applying this selection. Here, a decision must be made to decide whether the event is from signal and background or not.
3. Multivariate analysis


b) Deep Neural Network (DNN) hyper-parameters scanning

Plots of $\pi \times \epsilon_3$ - purity ($\pi$) = TP/(TP+FP), sign_eff [$\epsilon_3$] = TP/(TP+FN) vs ith model

[Graphs and plots showing data distributions and analysis results]

Brunella D’Anzi

ACAT 2021

Lightning talks session, 3 Dec 2021
Brunella D'Anzi

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results

4. Results

4. Results

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results

4. Results

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results

4. Results

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results

4. Results

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results

4. Results

Merging the 3 channels (4e, 4µ and 2e2µ) for the DNN training.

Results
Conclusions

- Despite the signal rarity, an area under the ROC curve (AUC) ≈ 98% with the DNN algorithm has been computed.
- A similar binary classification task can be performed for discriminating the VBF HH production under Effective Field Theory (EFT) models vs SM bkgs. The former have enhanced cross-sections, and therefore are simpler to be selected.
- The use of Deep Learning techniques will be proposed to Ph.D. students in the 2nd edition of the ML-INFN hackathon (13-15 December 2021).
Thank you!

brunella.d'anzi@cern.ch

**References**

- Search for double Higgs events produced via a vector boson fusion mechanism in the decay channel bb4l with the CMS experiment at the LHC

- Signal/background discrimination for the VBF Higgs four lepton decay channel with the CMS experiment using Machine Learning classification techniques, ML_INFN project

Backup slides
Motivation: the HH production mode

<table>
<thead>
<tr>
<th>Production mode</th>
<th>$\sigma[f_b]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gluon fusion</td>
<td>33.40 $^{+11.17}_{-9.21}$ (scale) $\pm 2.1%$ (PDF) $\pm 2.3%$ (pdf) $\pm 5.0%$ (top)</td>
</tr>
<tr>
<td>VBF</td>
<td>1.62 $^{+1.45}_{-1.12}$ (scale) $\pm 2.3%$ (PDF) $\pm 2.1%$ (pdf) $\pm 8.6%$ (top)</td>
</tr>
<tr>
<td>$t\bar{t}$HH</td>
<td>0.77 $^{+0.22}_{-0.17}$ (scale) $\pm 3.2%$ (PDF) $\pm 1.9%$ (pdf) $\pm 7.4%$ (top)</td>
</tr>
<tr>
<td>WW'HH</td>
<td>0.320 $^{+0.022}_{-0.018}$ (scale) $\pm 2.2%$ (PDF) $\pm 1.9%$ (pdf) $\pm 10.3%$ (top)</td>
</tr>
<tr>
<td>W'H'H</td>
<td>0.153 $^{+0.016}_{-0.013}$ (scale) $\pm 2.8%$ (PDF) $\pm 1.9%$ (pdf) $\pm 11.6%$ (top)</td>
</tr>
<tr>
<td>ZHH</td>
<td>0.362 $^{+0.030}_{-0.028}$ (scale) $\pm 1.9%$ (PDF) $\pm 1.9%$ (pdf) $\pm 5.3%$ (top)</td>
</tr>
<tr>
<td>$t\bar{t}$HH</td>
<td>0.0281 $^{+0.0022}_{-0.0019}$ (scale) $\pm 4.5%$ (PDF) $\pm 1.9%$ (pdf) $\pm 19.2%$ (top)</td>
</tr>
</tbody>
</table>

$\sqrt{s} = 13$ TeV

LHC center-of-mass energy!
Motivation: the HH decay mode

<table>
<thead>
<tr>
<th>Decay mode</th>
<th>$B$ [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H \rightarrow bb$</td>
<td>58.00$^{+0.72}_{-0.73}$</td>
</tr>
<tr>
<td>$H \rightarrow W^+W^-$</td>
<td>21.52$^{+0.31}_{-0.32}$</td>
</tr>
<tr>
<td>$H \rightarrow gg$</td>
<td>8.18$^{+0.42}_{-0.42}$</td>
</tr>
<tr>
<td>$H \rightarrow \tau^+\tau^-$</td>
<td>6.27$^{+0.10}_{-0.10}$</td>
</tr>
<tr>
<td>$H \rightarrow c\bar{c}$</td>
<td>2.88$^{+0.16}_{-0.16}$</td>
</tr>
<tr>
<td>$H \rightarrow ZZ^*$</td>
<td>2.641$^{+0.040}_{-0.040}$</td>
</tr>
<tr>
<td>$H \rightarrow \gamma\gamma$</td>
<td>0.2270$^{+0.0007}_{-0.0007}$</td>
</tr>
<tr>
<td>$H \rightarrow Z\gamma$</td>
<td>0.1541$^{+0.0090}_{-0.0090}$</td>
</tr>
<tr>
<td>$H \rightarrow \mu^+\mu^-$</td>
<td>0.02171$^{+0.00026}_{-0.00026}$</td>
</tr>
</tbody>
</table>
Scanning of the DNN hyper-parameters

<table>
<thead>
<tr>
<th>NN hyper-parameters</th>
<th>Tested options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input variables</td>
<td>leptons/jets ($p_T, \eta, \phi$), jets (Q/G Likelihood, DeepCsV)</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>10%, 20%</td>
</tr>
<tr>
<td>Topologies</td>
<td>10:10:10:10:10, 30:30:30:30, 20:20:20:20, 50:50:50:50, ...</td>
</tr>
<tr>
<td>Early stop</td>
<td>50, 100, 600, 3000</td>
</tr>
<tr>
<td>Minimizer SGD</td>
<td>SGD, Adadelta</td>
</tr>
<tr>
<td>Batch size</td>
<td>5, 32, 64, 128, 786</td>
</tr>
<tr>
<td>Neuron</td>
<td>ReLU, SeLU, Tanh</td>
</tr>
<tr>
<td>Loss Scaling</td>
<td>MC event weight</td>
</tr>
</tbody>
</table>
DNN training strategy

- The channels \(4e, 4\mu\) and \(2e2\mu\) selected separately are merged into just a sample for each MC (the channels will be randomized inside the sample).

- Each merged sample are divided into two independent sets:

  A. **Training set**: contains 80% of all events (from each sample) and is used to train ANNs;

  B. **Testing set**: contains remaining 20% of events and is used to test ANN after training;

- ANNs are built and trained via the open-source Keras (standard ML community tool) python package, not linked to TMVA.
DNN over-training (over-fitting)

Gap between training and validation accuracy indicates the amount of overfitting.

- If validation error curve shows small accuracy compared to training it indicates overfitting ⇒ add regularization (Early Stopping & Dropout) or use more data.

![Early Stopping](image)

![Dropout](image)

- Early Stopping: remove nodes randomly during training to avoid co-adaptation on training data.