Machine learning approaches to the Higgs boson self coupling

6 JUNE ICHEP 2018 @ COEX SEOUL
JUBIN PARK (CHONNAM NATIONAL UNIVERSITY)
COLLABORATED WITH JUNG CHANG, KINGMAN CHEUNG, JAE SIK LEE, CHIH-TING LU

Refs: Higgs-boson-pair production $H(\rightarrow bb^\ast)H(\rightarrow \gamma\gamma)$ from gluon fusion at the HL-LHC and HL-100 TeV hadron collider (Arxiv:1804.07130)
Higgs-boson-pair production $H(\rightarrow bb^\ast)H(\rightarrow \gamma\gamma)$ from gluon fusion with multivariate technique (Work in Progress)
Contents

Why Higgs pair production so difficult?
Why Higgs pair production so interesting?
Machine Learning approaches to the Higgs boson self coupling
Summary Table
Conclusion
Why Higgs pair production so difficult?

In the SM, $hh$ rates are small: In the leading gluon fusion production mode, the cross section at 14 TeV is only 40 fb, further suppressed by each decay branching fractions.
Why Higgs pair production so difficult?

- $\text{Xsec}(gg \rightarrow hh) = 39.64^{+4.4}_{-6.0} \text{ (scale)} \pm 2.1 \text{ (PDF)} \pm 2.2 \text{ } (\alpha_s) \text{ fb @ [14 TeV, } m_h = 125 \text{ GeV]}$

NNLO cross sections including top quark mass effects to NLO

- $O(10^{-3})$ smaller than the single Higgs production (SM)

\[ \text{Xsec}(gg \rightarrow hh) \sim 40 \text{ fb} \quad \text{Xsec}(gg \rightarrow h) \sim 50 \text{ pb} \quad \text{@ 14 TeV} \]

- For the reference, with Xsec $\sim 33$ fb at 13 TeV,

\begin{align*}
\text{2017 LHC @ 13 TeV with 40 fb}^{\text{-1}} & \rightarrow 1320 \text{ Events} \\
\text{14 TeV with 40 fb}^{\text{-1}} & \rightarrow 1600 \text{ Events}
\end{align*}
Why Higgs pair production so interesting?

- Allows accessing crucial components of the Higgs sector !!!

  - can probe the Higgs self-coupling
  - can help to reconstruct the electroweak symmetry breaking potential
  - may reveal the doublet nature of the Higgs by means of the hhVV coupling
Machine Learning approaches to the Higgs boson self coupling

① BDT(Boosted Decision Tree) : $bb\gamma\gamma$
   BDT + kinematic cuts $\rightarrow$ 5 $\sigma$ (4.6 $\sigma$) significance with 10 % (20%) systematics and 3 ab$^{-1}$

② (Supervising) Deep Neural Networks (DNN) : $bbWW + bb\tau\tau$
   1. “Supervising Deep Neural Networks with topological augmentation in search for di-Higgs production at the LHC (Won Sang Cho, next speaker)
      5 classes by the number of leptonic taus
      Optimass & its compatibility distance with dim. Of vars $\sim$ 40
      \[ \text{AUC of ROC} = 0.991 \]
      \[ \text{Eff(sig)} \]
      @ (Background purity=0.01) = 0.84
Machine Learning approaches to the Higgs boson self coupling

③ DNN (ANN: a multi-layer feed-forward artificial neural network) : bbbb


DNN + kinematic cuts $\frac{S}{\sqrt{B}} \sim 3 \sigma$ significance with 3 ab$^{-1}$
Machine Learning approaches to the Higgs boson self coupling

Boosted Decision Trees with gradient boosting (BDTG)

Predictive learning via rule ensembles (RuleFit)  Function discriminant analysis (FDA)
<table>
<thead>
<tr>
<th>Channel</th>
<th>Achievable Significance ($\sigma$)</th>
<th>Methods</th>
<th>Papers</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ (3.1 $\sim$ 5.7)</td>
<td>DNN</td>
<td>Arxiv: 1609.002541</td>
<td>100 TeV FCC (10 ab$^{-1}$)</td>
</tr>
<tr>
<td>bbWW</td>
<td></td>
<td>DNN</td>
<td>Dr. Won Sang Cho’s talk</td>
<td>HL-LHC (3 ab$^{-1}$)</td>
</tr>
<tr>
<td>bb\tau\tau</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWWWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bbYY</td>
<td>~ 5 (4.6)</td>
<td>Kinematic Cuts + BDT</td>
<td>Phys.Rev. D96 (2017) no.3, 035022</td>
<td>HL-LHC (3 ab$^{-1}$),</td>
</tr>
<tr>
<td></td>
<td>~ 2.1</td>
<td>Kinematic Cuts + BDT</td>
<td>Preriminary</td>
<td>With full BGs.</td>
</tr>
<tr>
<td>bbZZ(eemm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Summary Table

<table>
<thead>
<tr>
<th>Channel</th>
<th>Achievable Significance ($\sigma$)</th>
<th>Methods</th>
<th>Papers</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ (3.1 ~ 5.7)</td>
<td>DNN</td>
<td>Arxiv: 1609.002541</td>
<td>100 TeV FCC (10 ab^-1)</td>
</tr>
<tr>
<td>bbWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bb\tau\tau</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bbYY</td>
<td>~ 5 (4.6)</td>
<td>Kinematic Cuts + BDT</td>
<td>Phys.Rev. D96 (2017) no.3, 035022</td>
<td>HL-LHC (3 ab^-1),</td>
</tr>
<tr>
<td></td>
<td>~ 2.1</td>
<td>Kinematic Cuts + BDT</td>
<td>Preriminary</td>
<td>With full BGs.</td>
</tr>
<tr>
<td>bbZZ(eemm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Summary Table

<table>
<thead>
<tr>
<th>Channel</th>
<th>Achievable Significance (σ)</th>
<th>Methods</th>
<th>Papers</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ (3.1 ~ 5.7)</td>
<td>DNN</td>
<td>Arxiv: 1609.002541</td>
<td>100 TeV FCC (10 ab^-1)</td>
</tr>
<tr>
<td>bbWW</td>
<td></td>
<td>DNN</td>
<td>Dr. Won Sang Cho’s talk</td>
<td>HL-LHC (3 ab^-1)</td>
</tr>
<tr>
<td>bbττ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bbYY</td>
<td>~ 5 (4.6)</td>
<td>Kinematic Cuts + BDT</td>
<td>Phys.Rev. D96 (2017) no.3, 035022</td>
<td>HL-LHC (3 ab^-1),</td>
</tr>
<tr>
<td></td>
<td>~ 2.1</td>
<td>Kinematic Cuts + BDT</td>
<td>Preliminary</td>
<td>With full BGs.</td>
</tr>
<tr>
<td>bbZZ(eemm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Summary Table

<table>
<thead>
<tr>
<th>Channel</th>
<th>Achievable Significance ($\sigma$)</th>
<th>Methods</th>
<th>Papers</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~ (3.1 ~ 5.7)</td>
<td>DNN</td>
<td>Arxiv: 1609.002541</td>
<td>100 TeV FCC (10 ab^-1)</td>
</tr>
<tr>
<td>bbWW</td>
<td></td>
<td>DNN</td>
<td>Dr. Won Sang Cho’s talk</td>
<td>HL-LHC (3 ab^-1)</td>
</tr>
<tr>
<td>bbtt</td>
<td></td>
<td>DNN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WWWW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bbYY</td>
<td>~ 5 (4.6)</td>
<td>Kinematic Cuts + BDT</td>
<td>Phys.Rev. D96 (2017) no.3, 035022</td>
<td>HL-LHC (3 ab^-1),</td>
</tr>
<tr>
<td></td>
<td>~ 2.1</td>
<td>Kinematic Cuts + BDT</td>
<td>Preliminary</td>
<td>With full BGs.</td>
</tr>
<tr>
<td>bbZZ(eemm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion I

1. Higgs pair production can allow us to reconstruct the EWSB potential and to understand the nature of the EWSB mechanism!

2. The $b b \Upsilon \Upsilon$ channel can offer the appropriate yields and clean(?) signal.

3. Various multivariate classification methods based on machine learning techniques are used to consider the enhancement of significance in measuring the Higgs self coupling.

4. We found that the **BDT-related methods (+ cut-based analysis) can give the best results** compared with other methods.

5. Presently, we are checking the consistencies of our methods.
We find that even for the most promising channel $HH \rightarrow b\bar{b}\gamma\gamma$ at the HL-LHC with a luminosity of, the significance is still not high enough to establish the Higgs self-coupling at the SM value ($\lambda / \lambda_{SM}=1$).

With the multivariate classification methods, for example, BDT based on machine learning techniques.

It may be enough to establish the Higgs self-coupling at the SM value ($\lambda / \lambda_{SM}=1$)!

Question: Is it possible to establish the general Higgs self-coupling (for instance, $\lambda / \lambda_{SM}=2$) at the HL-LHC?
Conclusion II

We find that even for the most promising channel \( HH \rightarrow b\bar{b}\gamma \) at the HL-LHC with a luminosity of, the significance is still not high enough to establish the Higgs self-coupling at the SM value \( \frac{\lambda}{\lambda_{SM}} = 1 \).

With the multivariate classification methods, for example, BDT based on machine learning techniques.

It may be enough to establish the Higgs self-coupling at the SM value \( \left( \frac{\lambda}{\lambda_{SM}} = 1 \right) \)!

Question: Is it possible to establish the general Higgs self-coupling (for instance, \( \frac{\lambda}{\lambda_{SM}} = 2 \)) at the HL-LHC?
BACKUP SLIDES
Our event selection cuts and TMVA variables

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Event Selection Criteria at the HL-LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Di-photon trigger condition, ( \geq 2 ) isolated photons with ( P_T &gt; 25 \text{ GeV} ), (</td>
</tr>
<tr>
<td>2</td>
<td>( \geq 2 ) isolated photons with ( P_T &gt; 30 \text{ GeV} ), (</td>
</tr>
<tr>
<td>3</td>
<td>( \geq 2 ) jets identified as b-jets with leading(subleading) ( P_T &gt; 40(30) \text{ GeV} ), (</td>
</tr>
<tr>
<td>4</td>
<td>Events are required to contain ( \leq 5 ) jets with ( P_T &gt; 30 \text{ GeV} ) within (</td>
</tr>
<tr>
<td>5</td>
<td>No isolated leptons with ( P_T &gt; 25 \text{ GeV} ), (</td>
</tr>
<tr>
<td>6</td>
<td>( 0.4 &lt; \Delta R_{bb} &lt; 2.0, 0.4 &lt; \Delta R_{\gamma\gamma} &lt; 2.0 )</td>
</tr>
<tr>
<td>7</td>
<td>( 122 &lt; M_{\gamma\gamma}/\text{GeV} &lt; 128 ) and ( 100 &lt; M_{bb}/\text{GeV} &lt; 150 )</td>
</tr>
<tr>
<td>8</td>
<td>( P_T^{\gamma} &gt; 80 \text{ GeV}, P_T^{bb} &gt; 80 \text{ GeV} )</td>
</tr>
</tbody>
</table>
$\lambda$ dependency with BDT

Preliminary
$\lambda$ dependency with MLP

Preliminary
Machine Learning (ML)

Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

- **Supervised Learning**  Data With label
- **Unsupervised Learning**  Data Without label
- **Reinforcement Learning**
Higgs pair productions

Gluon Fusion

Vector Boson Fusion

Top associated productions

Higgs strahlung
Why Higgs pair production so difficult?
## Search channel for Higgs pair production

<table>
<thead>
<tr>
<th>Channel</th>
<th>BR(%)</th>
<th>Events with 3 ab⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbbb</td>
<td>~ 33</td>
<td>40080</td>
</tr>
<tr>
<td>bbWW</td>
<td>~ 25</td>
<td>30000</td>
</tr>
<tr>
<td>bbττ</td>
<td>~ 7.3</td>
<td>9000</td>
</tr>
<tr>
<td>WWWW</td>
<td>~ 4.3</td>
<td>5200</td>
</tr>
<tr>
<td>bbYY</td>
<td>~ 0.27</td>
<td>5200</td>
</tr>
<tr>
<td>bbZZ(eemm)</td>
<td>~ 0.015</td>
<td>19</td>
</tr>
</tbody>
</table>
TMVA methods

Rectangular cut optimization (binary splits, Sec. 8.1).

Projective likelihood estimation (Sec. 8.2).

Multi-dimensional likelihood estimation (PDE range-search { Sec. 8.3, PDE-Foam { Sec. 8.4, and k-NN { Sec. 8.5).

Linear and nonlinear discriminant analysis (H-Matrix { Sec. 8.6, Fisher { Sec. 8.7, LD { Sec. 8.8, FDA { Sec. 8.9).

Artificial neural networks (three different multilayer perceptron implementations { Sec. 8.10).

Support vector machine (Sec. 8.11).

Boosted/bagged decision trees (Sec. 8.12).

Predictive learning via rule ensembles (RuleFit, Sec. 8.13).

A generic boost classier allowing one to boost any of the above classiers (Sec. 9).

A generic category classier allowing one to split the training data into disjoint categories with independent MVAs.
// --- Cut optimisation
Use["Cuts"] = 1;
Use["CutsD"] = 1;
Use["CutsPCA"] = 0;
Use["CutsGA"] = 0;
Use["CutsSA"] = 0;

// --- 1-dimensional likelihood ("naive Bayes estimator")
Use["Likelihood"] = 1;
Use["LikelihoodD"] = 0; // the "D" extension indicates decorrelated input variables (see option strings)
Use["LikelihoodPCA"] = 1; // the "PCA" extension indicates PCA-transformed input variables (see option strings)
Use["LikelihoodKDE"] = 0;
Use["LikelihoodMIX"] = 0;

// --- Multidimensional likelihood and Nearest-Neighbour methods
Use["PDERS"] = 1;
Use["PDERSD"] = 0;
Use["PDERSPCA"] = 0;
Use["PDEFoam"] = 1;
Use["PDEFoamBoost"] = 0; // uses generalised MVA method boosting
Use["KNN"] = 1; // k-nearest neighbour method

// --- Linear Discriminant Analysis
Use["LD"] = 1; // Linear Discriminant identical to Fisher
Use["Fisher"] = 0;
Use["FisherG"] = 0;
Use["BoostedFisher"] = 0; // uses generalised MVA method boosting
Use["HMatrix"] = 0;
// --- Function Discriminant analysis
Use["FDA_GA"] = 1; // minimisation of user-defined function using Genetics Algorithm
Use["FDA_SA"] = 0;
Use["FDA_MC"] = 0;
Use["FDA_MT"] = 0;
Use["FDA_GAMT"] = 0;
Use["FDA_MCMT"] = 0;

// --- Neural Networks (all are feed-forward Multilayer Perceptrons)
Use["MLP"] = 0; // Recommended ANN
Use["MLPBFGS"] = 0; // Recommended ANN with optional training method
Use["MLPBN""] = 1; // Recommended ANN with BFGS training method and bayesian regulator
Use["CFMlpANN"] = 0; // Depreciated ANN from ALEPH
Use["TMlpANN"] = 0; // ROOT's own ANN

// --- Support Vector Machine
Use["SVM"] = 1;

// --- Boosted Decision Trees
Use["BDT"] = 1; // uses Adaptive Boost
Use["BDTG"] = 0; // uses Gradient Boost
Use["BDTB"] = 0; // uses Bagging
Use["BDTD"] = 0; // decorrelation + Adaptive Boost
Use["BDTF"] = 0; // allow usage of fisher discriminant for node splitting

// --- Friedman's RuleFit method, ie, an optimised series of cuts ("rules")
Use["RuleFit"] = 1;