Machine learning approaches to the Higgs boson self coupling

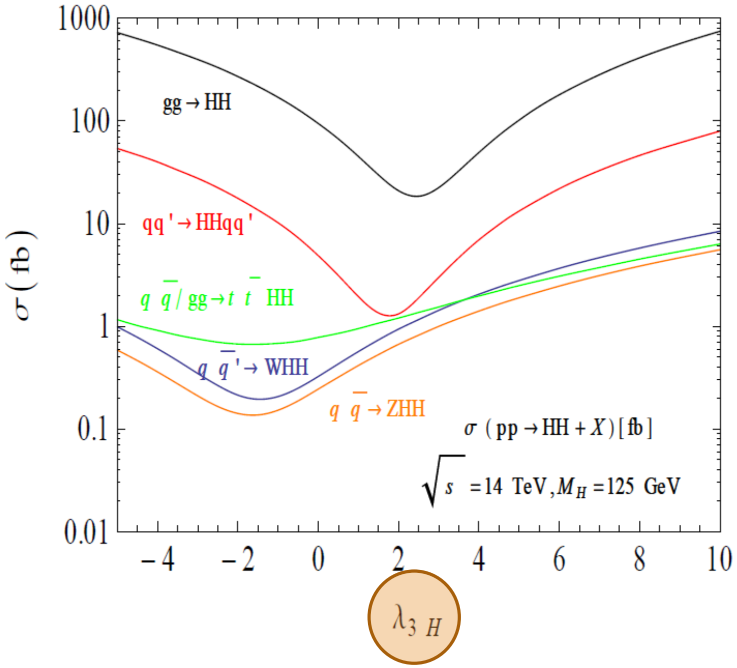
6 JUNE ICHEP 2018 @ COEX SEOUL
JUBIN PARK (CHONNAM NATIONAL UNIVERSITY)
COLLABORATED WITH JUNG CHANG, KINGMAN CHEUNG, JAE SIK LEE, CHIH-TING LU

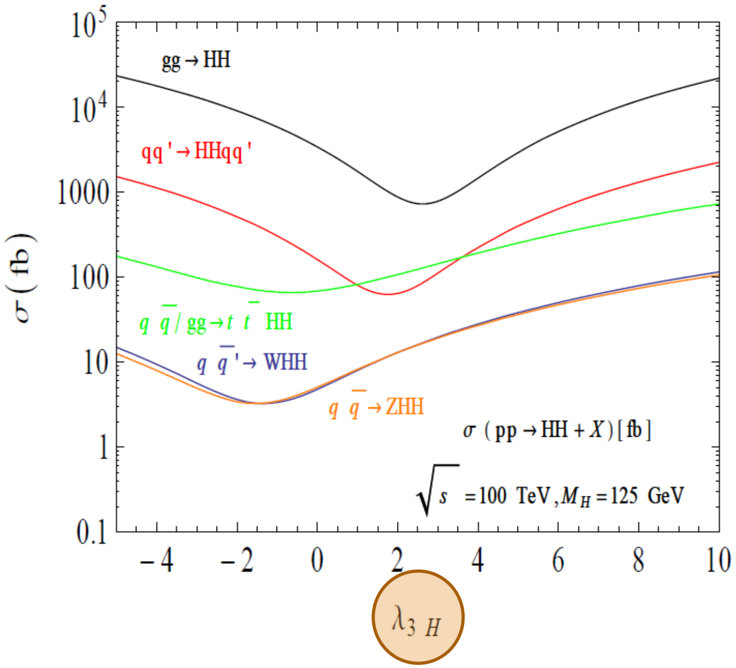
Refs : Higgs-boson-pair production $H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$ from gluon fusion at the HL-LHC and HL-100 TeV hadron collider (Arxiv :1804.07130)

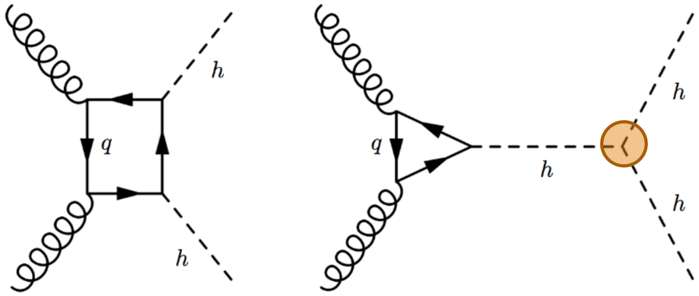
Higgs-boson-pair production $H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$ from gluon fusion with multivariate techenique (Work in Progress)

Contents

Why Higgs pair production so difficult ?







Distrcuctive interference !

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 ?

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

NNLO cross sections including top quark mass effects to NLO
Phys. Rev. Lett. 117, 012001 [S.Borowka, et al.]

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

Xsec(gg -> hh) ~ 40 fb $\Longleftrightarrow$ Xsec(gg -> h) ~ 50 pb @ 14 TeV

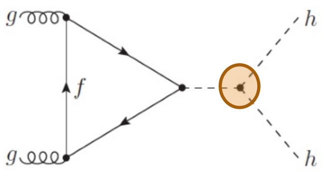
- For the reference, with Xsec ~ 33 fb at 13 TeV,

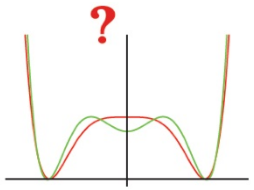
2017 LHC @ 13 TeV with 40 fb^-1 $\Rightarrow$ 1320 Events
14 TeV with 40 fb^-1 $\Rightarrow$ 1600 Events

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 <u>the doublet nature</u> of the Higgs by means of the hhVV coupling

Machine Learning approaches to the Higgs boson self coupling

① BDT(Boosted Decision Tree) : bbYY

1. Phys.Rev. D96 (2017) no.3, 035022 (Alves, Alexandre et al.) arXiv:1704.07395 [hep-ph]

BDT + kinematic cuts → 5 σ (4.6 σ) significance with 10 %(20%) systematics and 3 ab^-1

② (Supervising) Deep Neural Networks (DNN) : bbWW + bbττ

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 ~ 40

→ AUC of ROC = 0.991

Eff(sig)

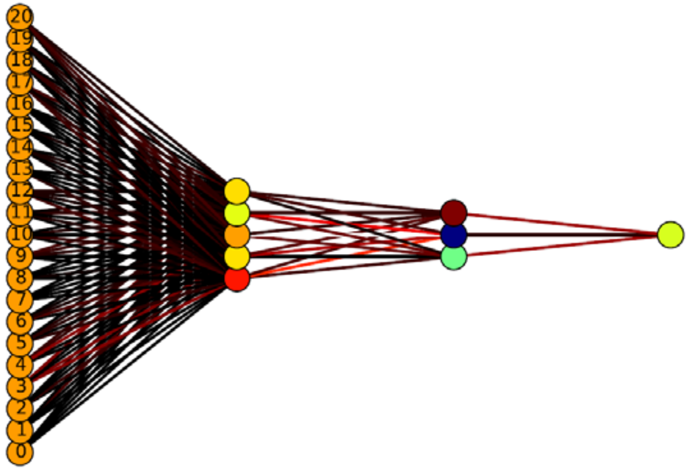
@(Background purity=0.01) = 0.84

Machine Learning approaches to the Higgs boson self coupling

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

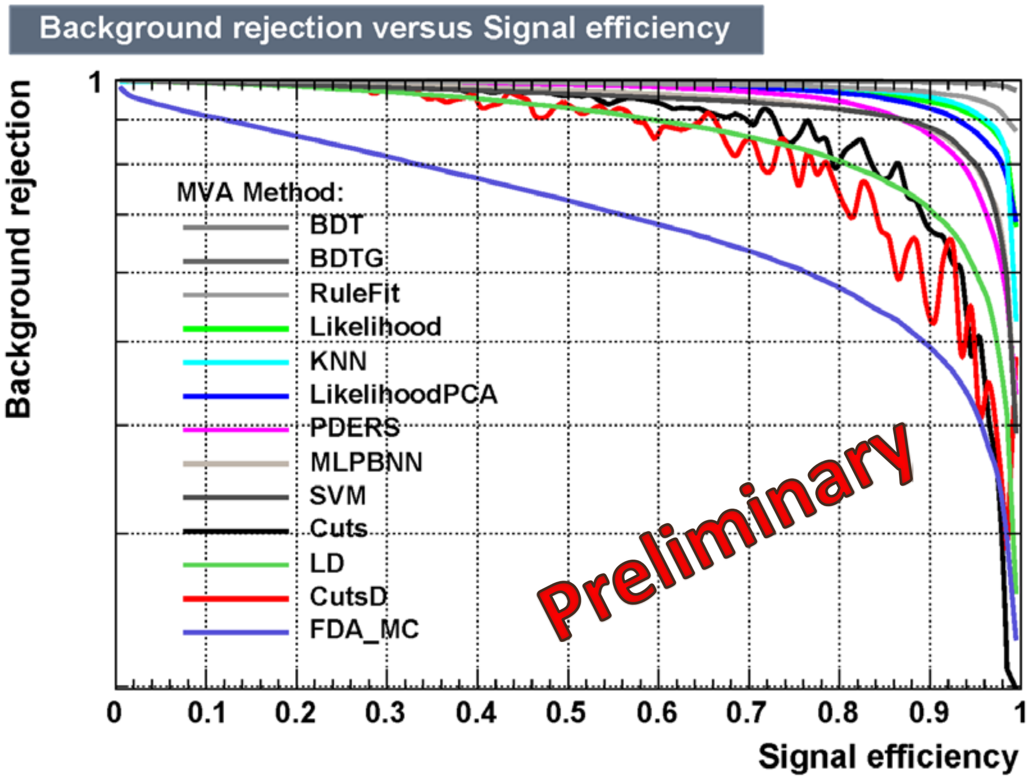
1. Eur. Phys. J. C (2016) 76:386 (Katharina Behr, Bortoletto et al.) arXiv:1512.08928 [hep-ph]

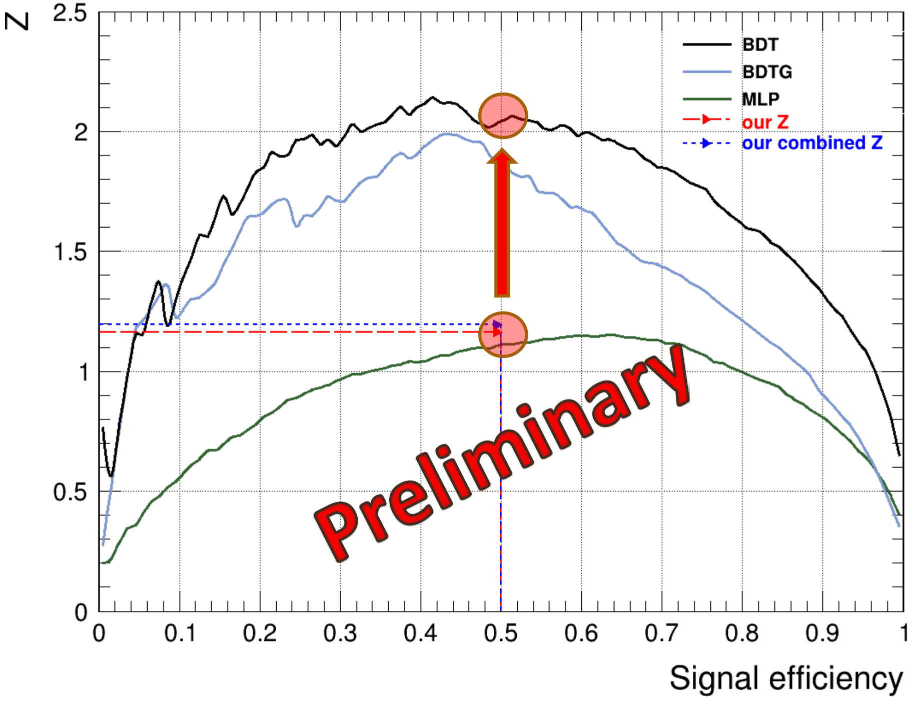
DNN + kinematic cuts → $\frac{S}{\sqrt{B}} \sim$ 3 σ significance with 3 ab^-1



Machine Learning approaches to the Higgs boson self coupling

④





Boosted Decision Trees with gradient boosting (BDTG)

Predictive learning via rule ensembles (RuleFit) **Function discriminant analysis (FDA)**

Summary Table

Channel	Achievable Significance (σ)	Methods	Papers	Remarks
bbbb	~ 3	Kinematic Cuts+ DNN	Eur. Phys. J. C (2016) 76:386	HL-LHC (3 ab^-1)
	~ (3.1 ~ 5.7)	DNN	Arxiv: 1609.002541	100 TeV FCC (10 ab^-1)
bbWW		DNN	Dr. Won Sang Cho’s talk	HL-LHC (3 ab^-1)
bbττ				
WWWW				
bbYY	~ 5 (4.6)	Kinematic Cuts + BDT	Phys.Rev. D96 (2017) no.3, 035022	HL-LHC (3 ab^-1),
	~ 2.1	Kinematic Cuts + BDT	Preriminary	With full BGs.
bbZZ(eemm)				

Summary Table

Channel	Achievable Significance (σ)	Methods	Papers	Remarks
bbbb	~ 3	Kinematic Cuts+ DNN	Eur. Phys. J. C (2016) 76:386	HL-LHC (3 ab^-1)
	~ (3.1 ~ 5.7)	DNN	Arxiv: 1609.002541	100 TeV FCC (10 ab^-1)
bbWW		DNN	Dr. Won Sang Cho's talk	HL-LHC (3 ab^-1)
bbττ				
WWWW				
bbYY	~ 5 (4.6)	Kinematic Cuts + BDT	Phys.Rev. D96 (2017) no.3, 035022	HL-LHC (3 ab^-1),
	~ 2.1	Kinematic Cuts + BDT	Preriminary	With full BGs.
bbZZ(eemm)				

Summary Table

Channel	Achievable Significance (σ)	Methods	Papers	Remarks
bbbb	~ 3	Kinematic Cuts+ DNN	Eur. Phys. J. C (2016) 76:386	HL-LHC (3 ab^-1)
	~ (3.1 ~ 5.7)	DNN	Arxiv: 1609.002541	100 TeV FCC (10 ab^-1)
bbWW		DNN	Dr. Won Sang Cho's talk	HL-LHC (3 ab^-1)
bb$\tau\tau$				
WWWW				
bbYY	~ 5 (4.6)	Kinematic Cuts + BDT	Phys.Rev. D96 (2017) no.3, 035022	HL-LHC (3 ab^-1),
	~ 2.1	Kinematic Cuts + BDT	Preriminary	With full BGs.
bbZZ(eemm)				

Summary Table

Channel	Achievable Significance (σ)	Methods	Papers	Remarks
bbbb	~ 3	Kinematic Cuts+ DNN	Eur. Phys. J. C (2016) 76:386	HL-LHC (3 ab^-1)
	~ (3.1 ~ 5.7)	DNN	Arxiv: 1609.002541	100 TeV FCC (10 ab^-1)
bbWW		DNN	Dr. Won Sang Cho's talk	HL-LHC (3 ab^-1)
bbττ				
WWWW				
bbYY	~ 5 (4.6)	Kinematic Cuts + BDT	Phys.Rev. D96 (2017) no.3, 035022	HL-LHC (3 ab^-1),
	~ 2.1	Kinematic Cuts + BDT	Preliminary	HL-LHC (3 ab^-1), With full BGs.
bbZZ(eemm)				

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 bbYY 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.

Conclusion II

From Chih-Ting Lu's talk

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 (λ / λ _SM=1)**.

1.194 σ

⇩

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

2.1 σ

It may be enough to establish the Higgs self-coupling at the SM value **(λ / λ _SM=1) !**

Question : Is it possible to establish the general Higgs self-coupling (for instance, λ / λ _SM=2) at the HL-LHC ?

Conclusion II

From Chih-Ting Lu's talk

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

1.194 σ

⇩

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

2.1 σ

It may be enough to establish the Higgs self-coupling at the SM value **(λ / λ _SM=1) !**

Thank you for your attention !

Question : Is it possible to establish the general Higgs self-coupling (for instance, λ / λ _SM=2) at the HL-LHC ?

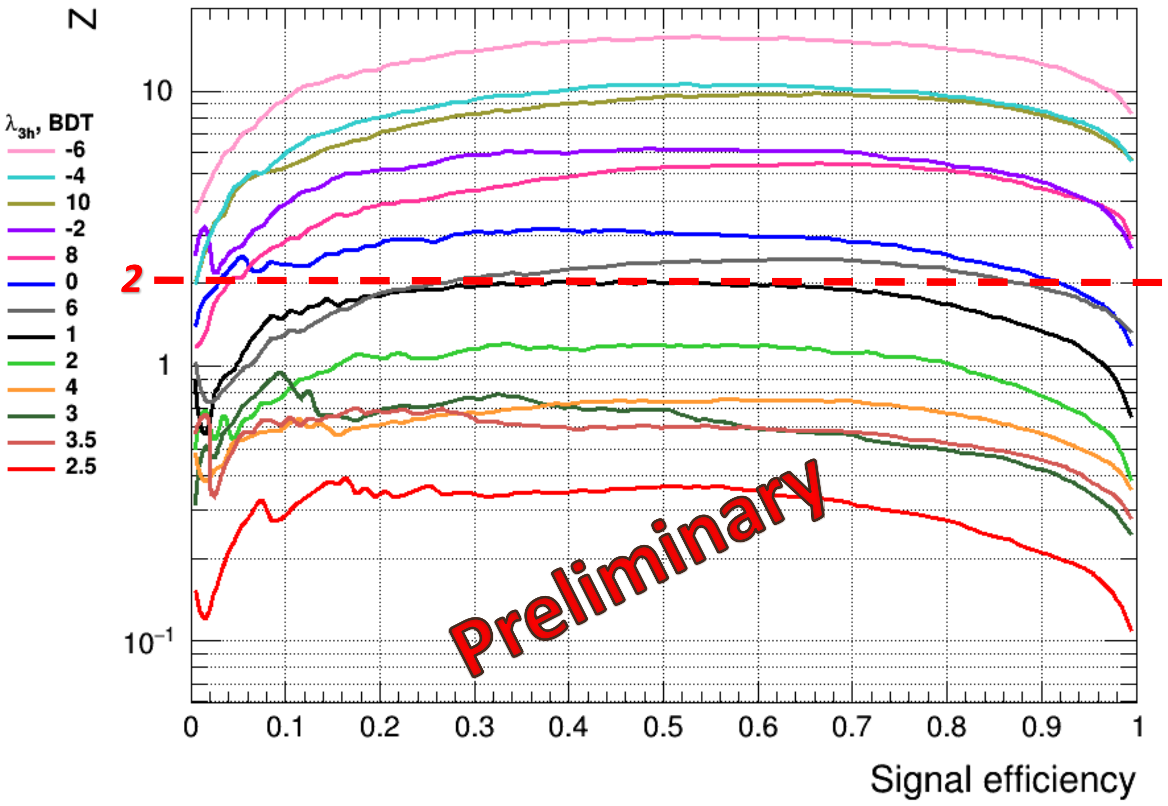
BACKUP SLIDES

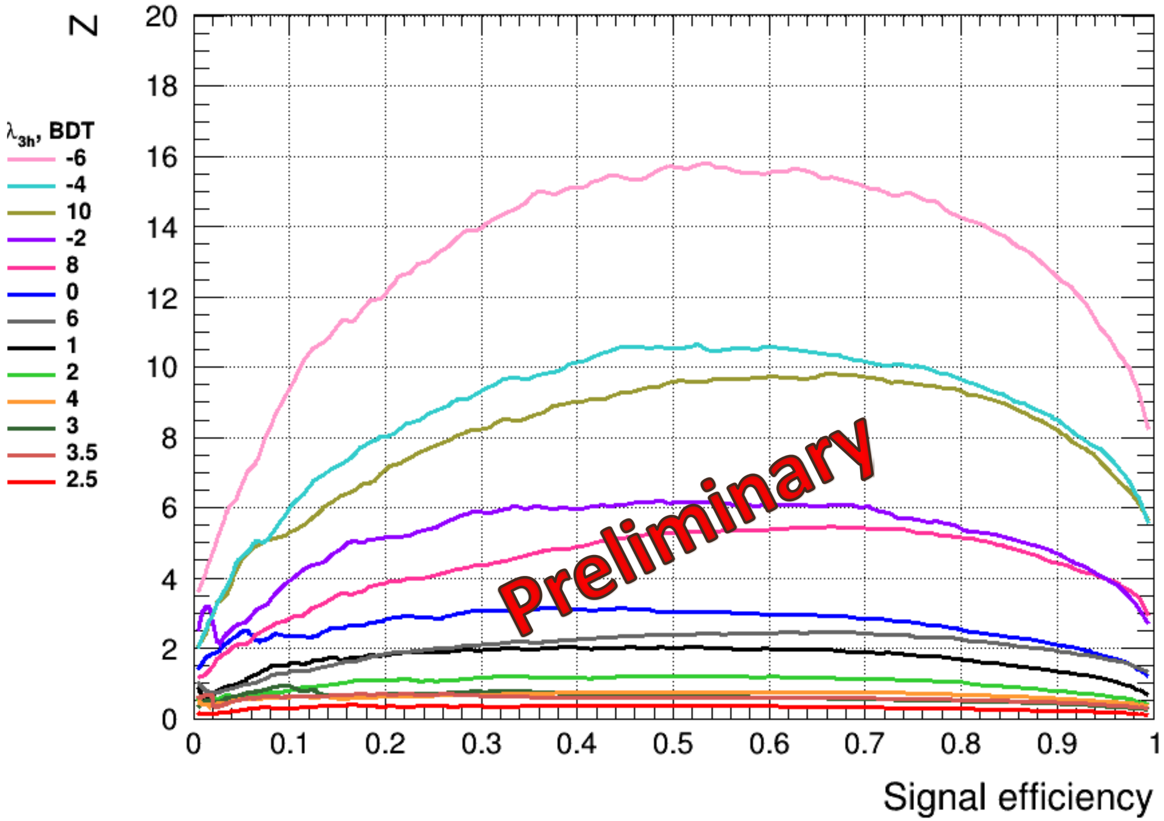
Our event selection cuts and TMVA variables

Sequence	Event Selection Criteria at the HL-LHC
1	Di-photon trigger condition, ≥ 2 isolated photons with $P_T > 25$ GeV, $\lvert\eta\rvert < 2.5$
2	≥ 2 isolated photons with $P_T > 30$ GeV, $\lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.37$, $\Delta R_{j\gamma} > 0.4$
3	≥ 2 jets identified as b-jets with leading(subleading) $P_T > 40(30)$ GeV, $\lvert\eta\rvert < 2.4$
4	Events are required to contain ≤ 5 jets with $P_T > 30$ GeV within $\lvert\eta\rvert < 2.5$
5	No isolated leptons with $P_T > 25$ GeV, $\lvert\eta\rvert < 2.5$
6	$0.4 < \Delta R_{b\bar{b}} < 2.0$, $0.4 < \Delta R_{\gamma\gamma} < 2.0$
7	$122 < M_{\gamma\gamma}/\mathrm{GeV} < 128$ and $100 < M_{b\bar{b}}/\mathrm{GeV} < 150$
8	$P_T^{\gamma\gamma} > 80$ GeV, $P_T^{b\bar{b}} > 80$ GeV

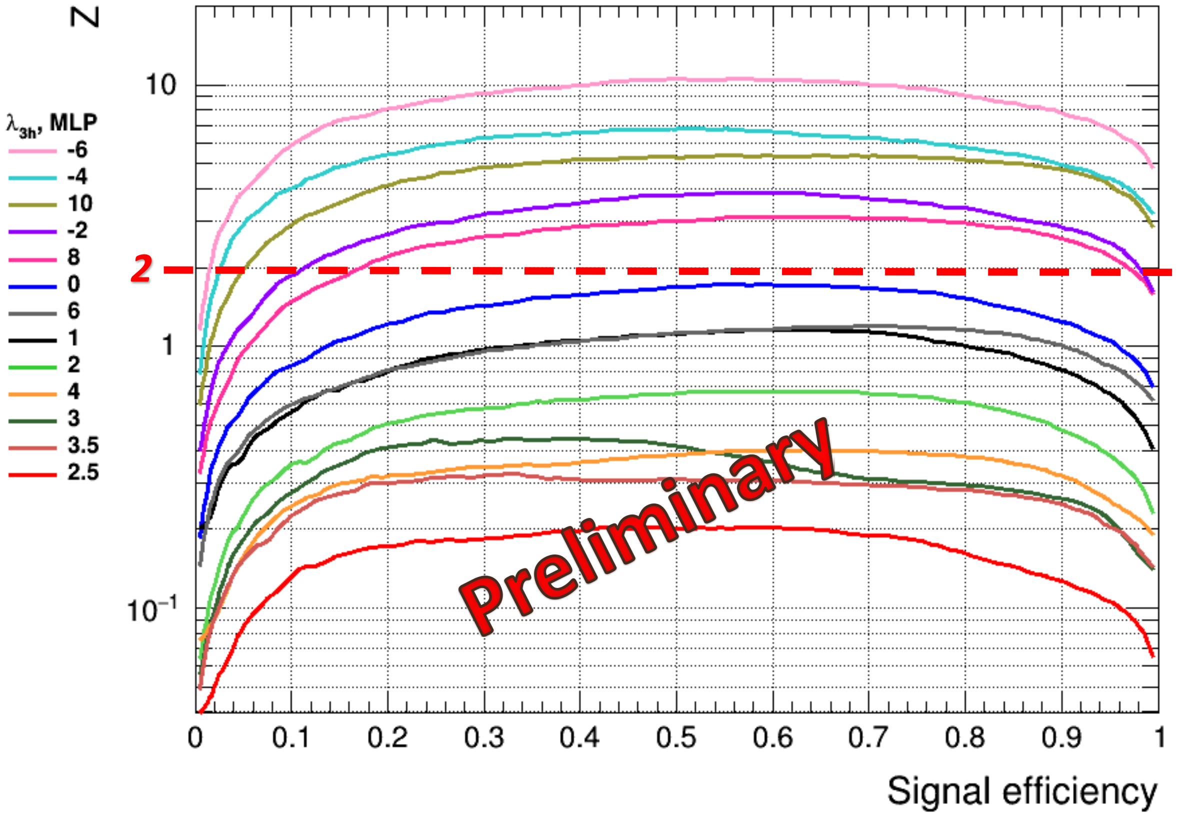
TMVA variables

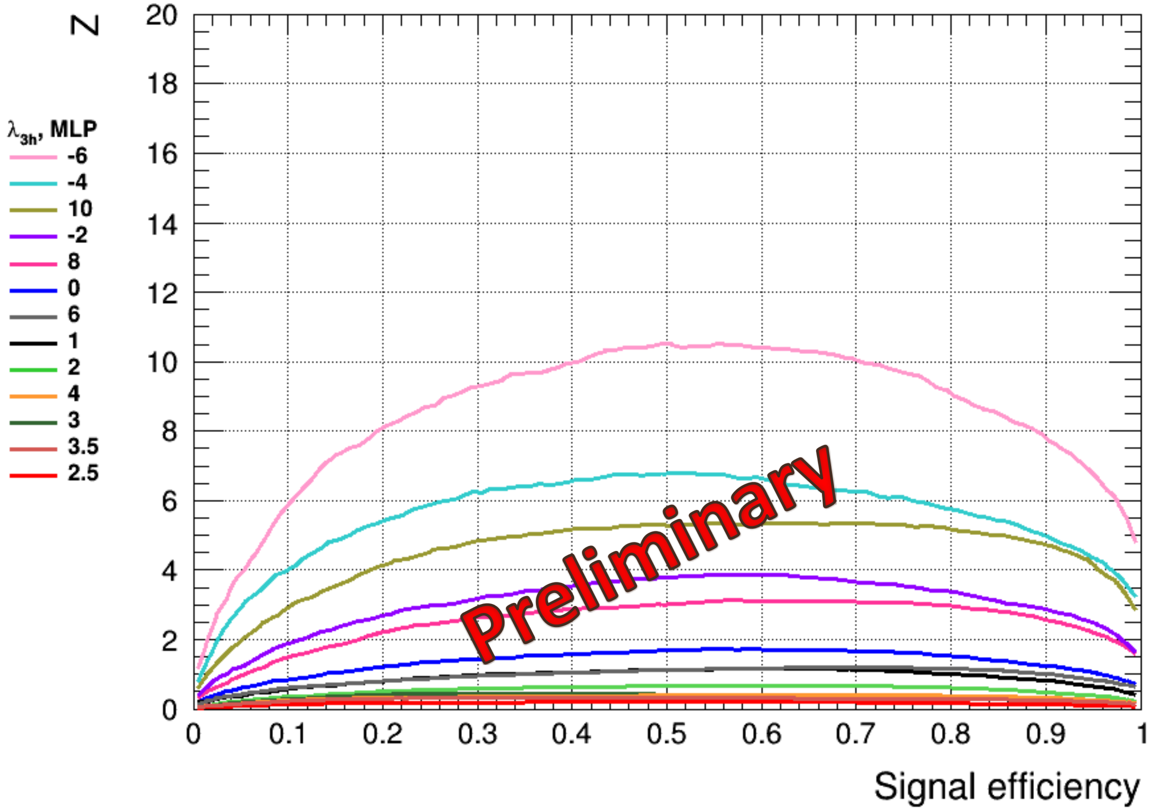
λ dependency with BDT





λ dependency with MLP





Machine Learning (ML)

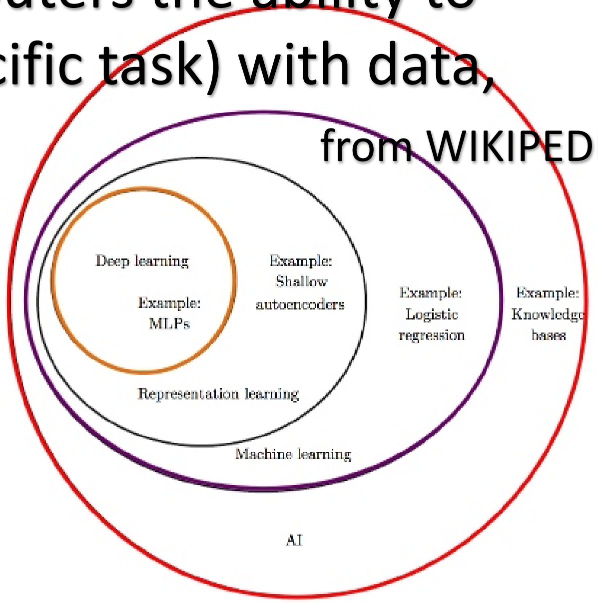
Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

from WIKIPEDIA

Supervised Learning **Data With label**

Unsupervised Learning **Data Without label**

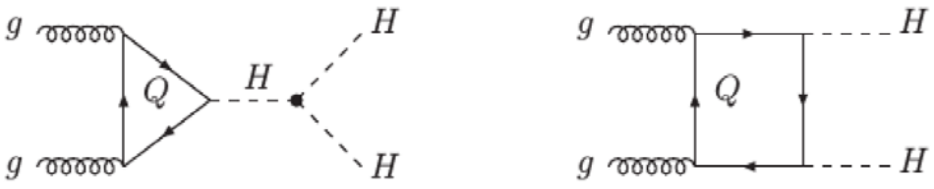
Reinforcement Learning



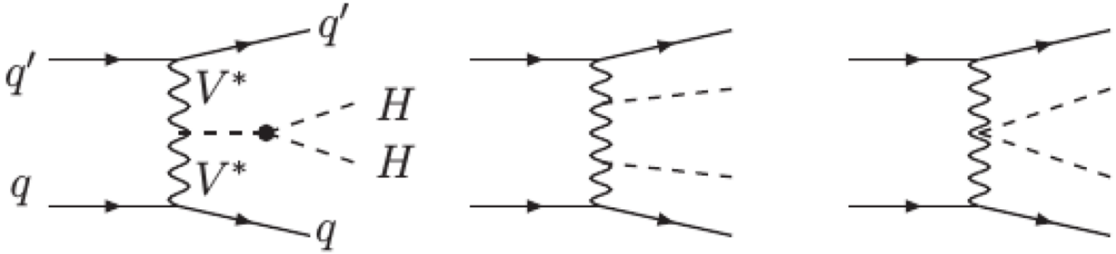
Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, MIT Press, 2016
http://www.deeplearningbook.org/

Higgs pair productions

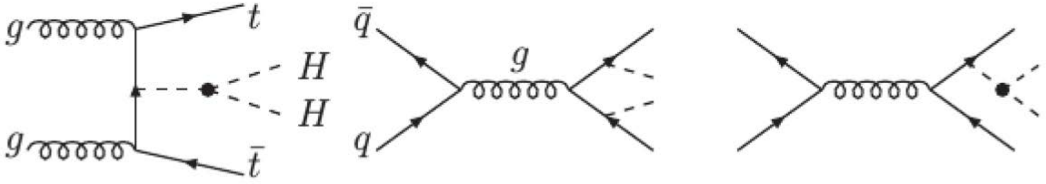
Gluon Fusion



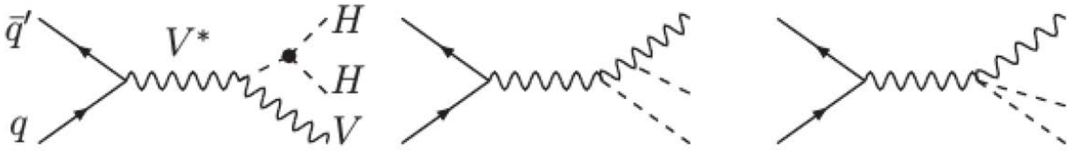
Vector Boson Fusion



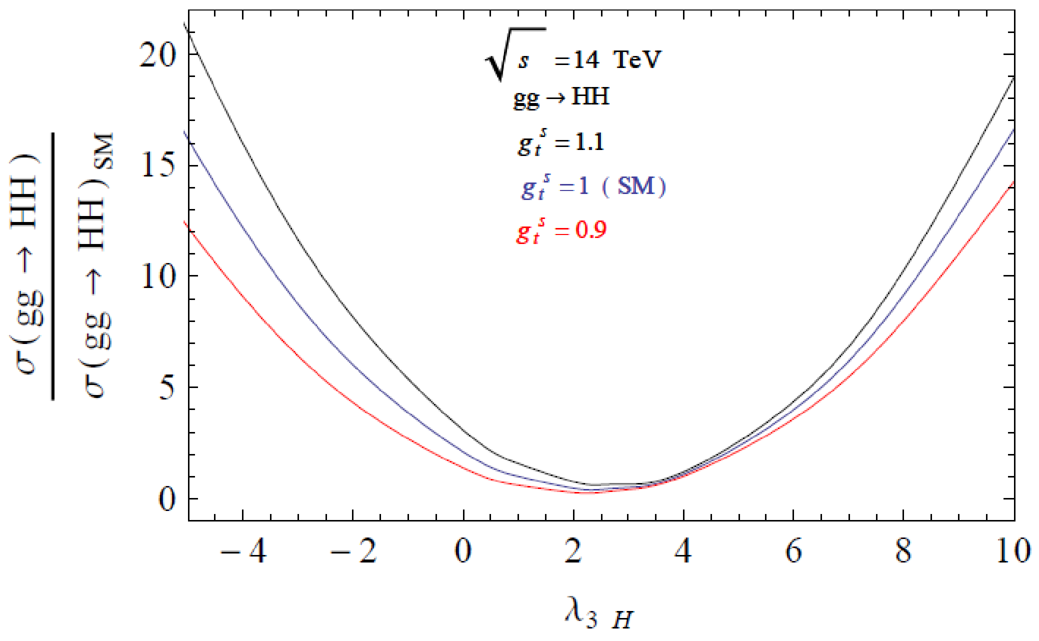
Top associated productions

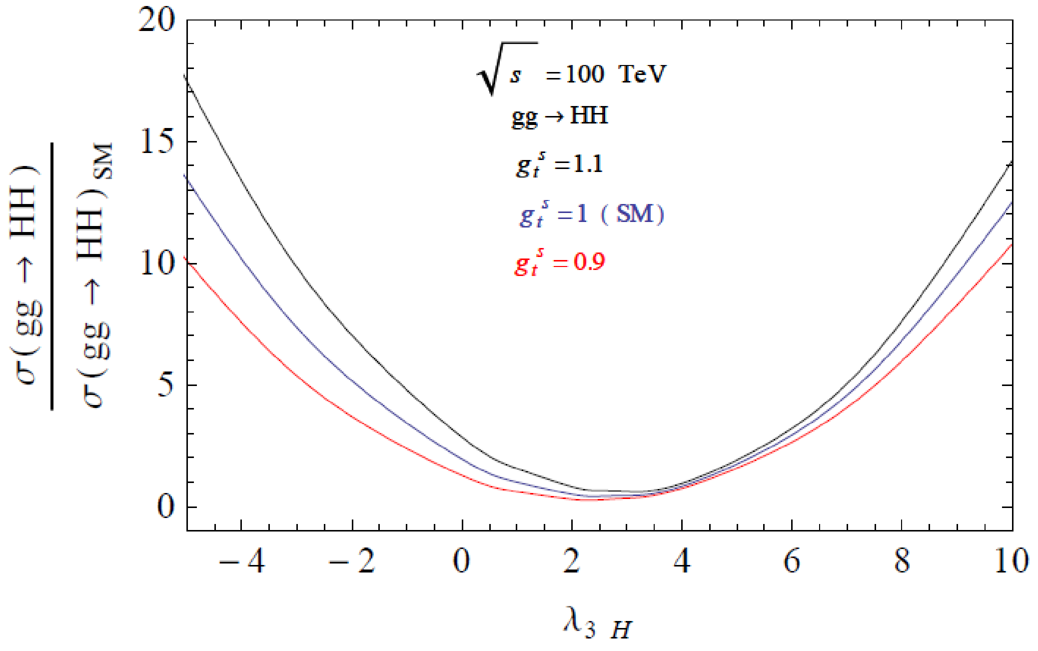


Higgs strahlung



Why Higgs pair production so difficult ?





Search channel for Higgs pair production

Channel	BR(%)	Events with 3 ab^-1
bbbb	~ 33	40080 Huge hadronic BG
bbWW	~ 25	30000 Huge ttbar BG
bb$\tau\tau$	~ 7.3	9000
WWWW	~ 4.3	5200
bbYY	~ 0.27	5200
bbZZ(eemm)	~ 0.015	19

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).

Articial 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;
   //
   // --- Mutidimensional 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["MLPBNN"]        = 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;
```