Optimizing the sensitivity to the Di-Higgs search in $bbW^{+}W^{-}$ final state using machine learning methods

Pratik Kafle

Department of Physics, Reed College

Abstract

A study of optimization of signal sensitivity in the search for the Di-Higgs event in the ATLAS experiment is presented where one of the Higgs bosons decays via the $H \rightarrow b\bar{b}$ channel and the other via $H \rightarrow W^{+}W^{-}$ with $b\bar{b}lvqq$ as the final state. Large irreducible background contamination from $t\bar{t}$ decay significantly decreases the signal sensitivity in the Di-Higgs search. We employ ROOT's built-in Multivariate analysis tool (TMVA) to train new physics-motivated variables along with currently trained variables. An increment in the rejection of $t\bar{t}$ background while keeping the signal efficiency high is observed.

Introduction: Di-Higgs

What is Di-Higgs?

➢ **Simply a pair of Higgs boson formed in high energy collision!**

Why do we care about Di-Higgs?

➢ One of the consequence of the Brout–Englert–Higgs (BEH) mechanism in the Standard Model (SM) is that it requires the self-coupling of the Higgs boson. And we have yet to verify this experimentally!

❑ **Studying Di-Higgs production at the ATLAS experiment can help us probe the Higgs self-coupling.**

❑ **Measuring the Higgs self-coupling will in turn help in reconstructing the Higgs Potential.**

Higgs Potential

$$V \approx V(v) + \frac{1}{2}m_H^2 h^2 + \underbrace{\frac{m_H^2}{2v}}_{\lambda_{HHH}} h^3 + \frac{m_H^2}{8v^2}\, h^4$$

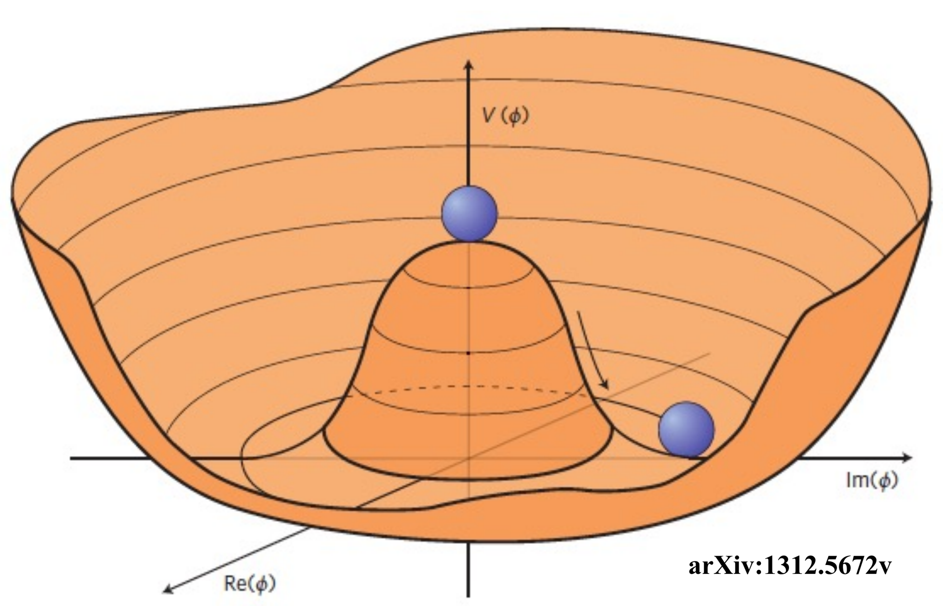


Figure 1. Shape of the SM Higgs Potential

Di-Higgs production at the LHC

Production Mode	Cross Section
$ggf - hh$	~33 fb
$VVF - hh$	~ 2 fb
$hhZ - hh$	~0.4 fb

More than 90% of SM Di-Higgs is produced via ggF, which is only accounted for in this analysis. There are two ways in which Di-Higgs is formed from ggF, as shown in figure 2. Their amplitudes interfere destructively, resulting to a production cross-section of ≈ 31 fb at $\sqrt{s}$ =13 TeV [1] .

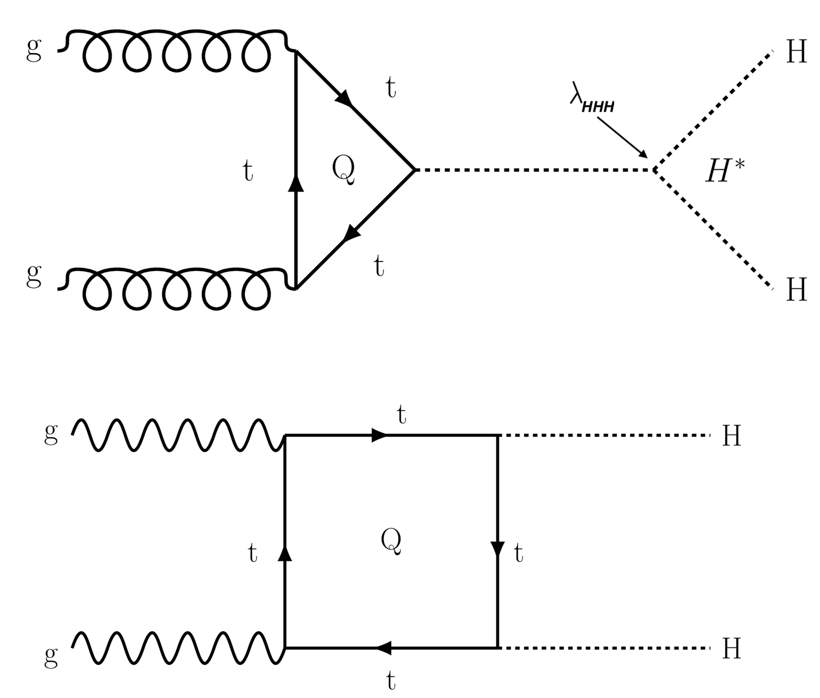


Figure 2. Feynman diagrams showing the production of SM Di-Higgs through gluon-gluon fusion. The figure in the top has the Higgs trilinear self-coupling, whereas the figure in bottom does not.

Di-Higgs Signal and background ($b\bar{b}\ W^{+}W^{-}$)

Di-Higgs Signal:

$pp \rightarrow HH \rightarrow b\bar{b}\ W^{+}W^{-} \rightarrow b\bar{b}lvqq$

Background: $(t\bar{t}\ ,\ t\bar{t}\,W,\ t\bar{t}\,Z,\ W$+jets, Z + jets...)

$t\bar{t} \rightarrow b\bar{b}\ W^{+}W^{-} \rightarrow b\bar{b}lvqq$

➢ Since $t\bar{t}$ accounts for > 80% of background, only its contribution is accounted for the machine learning modelling.

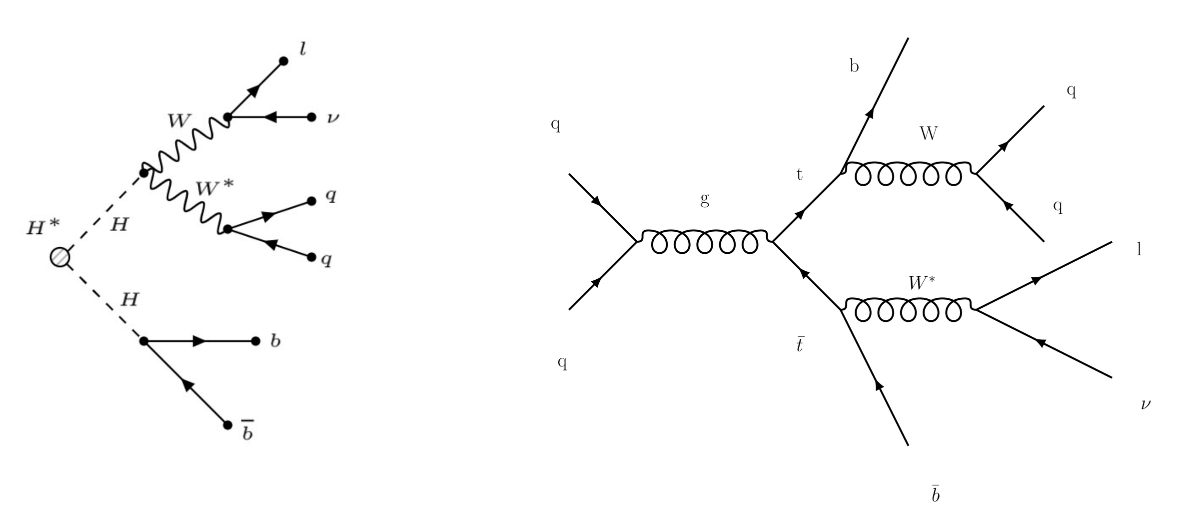


Figure 3: Di-Higgs decaying to $b\bar{b}lvqq$ (left), and $t\bar{t}$ decaying to $b\bar{b}lvqq$ (right), making the background irreducible.

Project description:

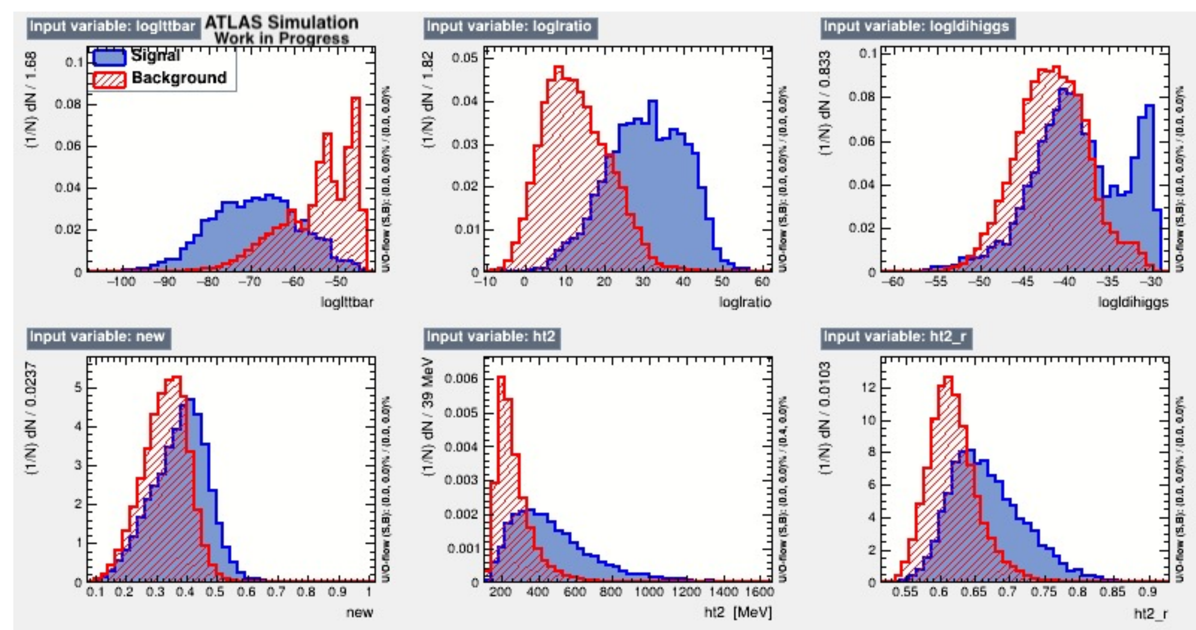
- Increase the discrimination of the background from signal for MC simulated data based on the signal significance, defined as $\frac{S}{\sqrt{S+B}}$

How?

- Introduce new kinematic variables, train machine learning algorithms (BDT and DNN) as implemented in ROOT's TMVA.

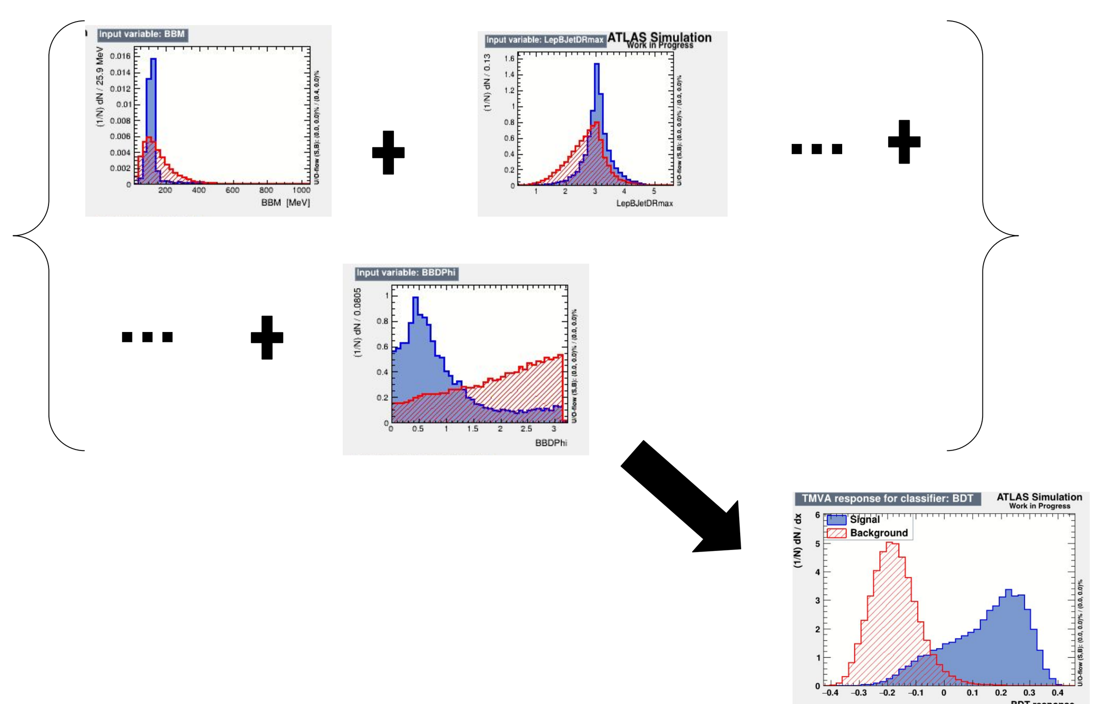
Data and methods

- 43 Kinematic Variables obtained from Monte Carlo Simulation of pp collision.



Machine Learning methods

- Deep Neural Network (DNN) algorithm,
- Boosted Decision Tree (BDT) algorithm
  - Many decision trees are trained together and integrated to form a "forest".
  - Several weak classifiers are amalgamated to construct a strong classifier.



Results

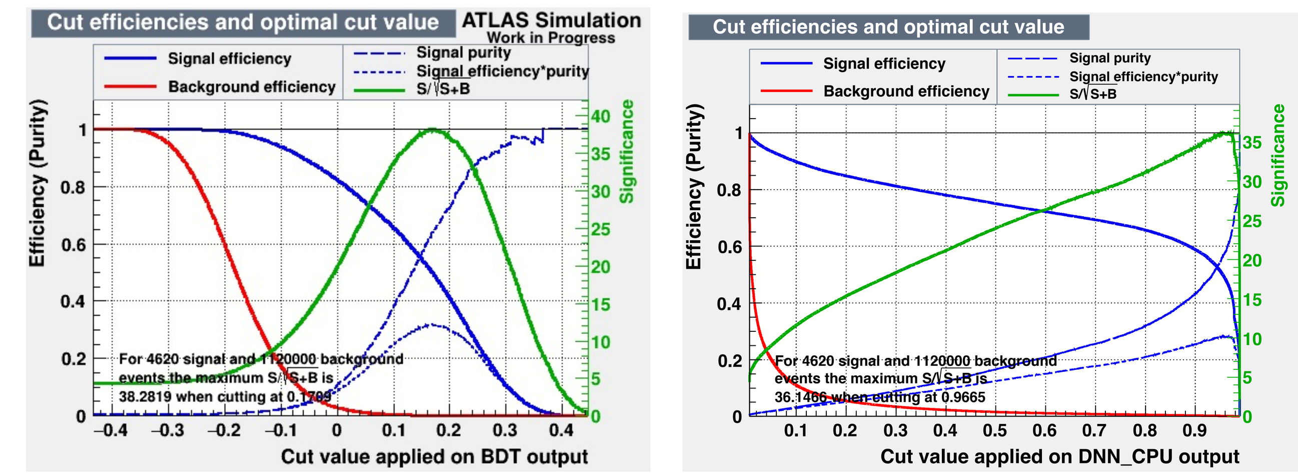


Figure 4: Signal Significance peaked at 38.2819, and 36.1466 for 43 BDT trained kinematic variables (left) and DNN trained variables (right) for 4620 signal and 1120000 background events.

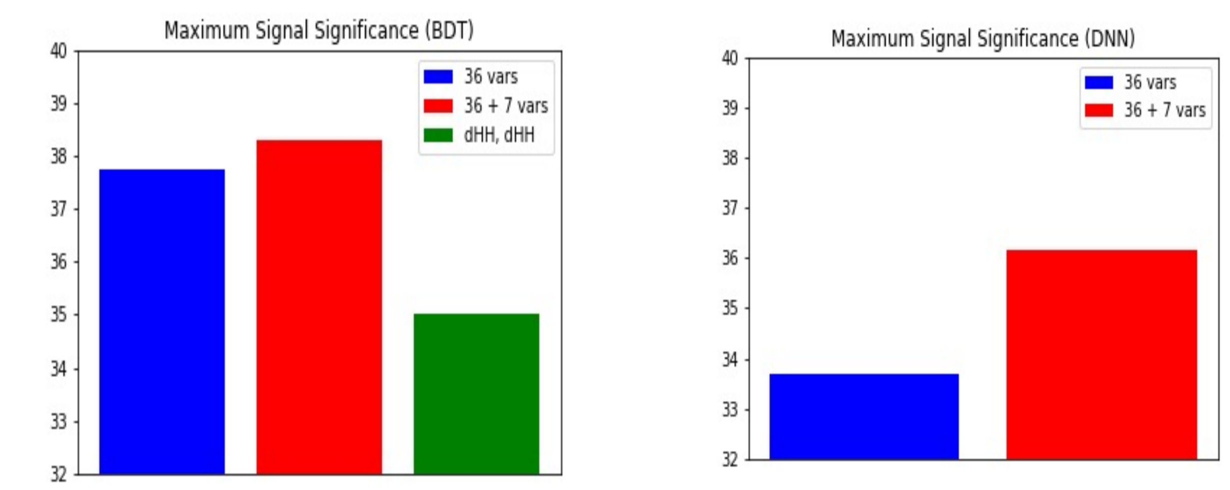


Figure 5: Comparison of performance of BDT (left) and DNN (right) performance on previously trained variables and (previously trained variables + new Kinematic variables) .

Conclusions

❑ **Adopting kinematic variables from similar channels and training machine learning models increased the signal significance,**

❑ **BDT algorithm performed better than DNN algorithm for a same number of input kinematic variables.**

References

1. LHC Higgs Cross Section HH Sub-group
2. P. Kafle, Optimization of sensitivity for the Di-Higgs search in $b\bar{b}lvqq$ final state in the LHC data recorded by the ATLAS detector in pp collisions at $\sqrt{s}$ =13 TeV, (2021)
3. ATLAS Collaboration, "Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS Detector," Physics Letters B800, 135103 (2020)

Acknowledgments

Special thanks to Dr. Suyog Shrestha (OSU/CERN) for the project idea and for numerous hours of instruction and for suggestions on improvement. Thanks also to CLBR for the Summer Research Fellowship Award. Thanks finally to Prof. Harris Kagan (OSU).