Discrimination between the final state of $t\bar{t}H$ and $t\bar{t}A$

using neural network

H P Pratama¹, S Sumowidagdo², R Primulando³ and T Mart¹

¹Department of Physics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
²Research Center for Physics, Indonesian Institute of Sciences (LIPI), Puspiptek Area, Serpong 15134, Indonesia
³Department of Physics, Faculty of Information Technology and Sciences, Catholic University of Parahyangan, Bandung 40141, Indonesia

Corresponding author’s email: suharyo.sumowidagdo@lipi.go.id

Abstract. The Higgs Boson discovered in 2012 at the Large Hadron Collider (LHC) is still assumed to be the Standard Model Higgs Boson -a scalar particle. But there are still possibilities for a pseudo-scalar Higgs Boson. Currently, LHC has not been able to discriminate whether the particle is a scalar, pseudo-scalar, or mixed-scalar-pseudoscalar particle. The focus of this research is in the formulation of the neural network algorithm to discriminate between event signals from a scalar Higgs Boson and pseudo-scalar Higgs Boson in the Higgs production channel of associate production of Higgs Boson with a pair of top-antitop quark in a proton-proton collision $p\bar{p} \rightarrow t\bar{t}H$ with center of mass energy $\sqrt{s} = 13$ TeV with only Higgs decaying to a pair of bottom-antibottom quark had been considered. The set of discriminating variables used is the 3-momentum of all the final state particles. Discriminating efficiency of 0.412 was obtained which showed relatively small discriminating power.

Keywords: Higgs Boson, LHC, top quark, bottom quark, neural network

1. Introduction

The Higgs Boson discovered in 2012 is still assumed to be a scalar particle as it is predicted by Standard Model. So far, Standard Model has been the best theory we have in explaining all phenomena in the particle physics [1]. However, there are many unexplained phenomena that lead us to improve the theory by a small (or sometime even radical) extension usually called the theory of Beyond the Standard Model (BSM) [2]. One of these extensions is the minimal supersymmetric extension of the standard model (MSSM) [3] which was formulated (one of the reasons) to explain a phenomenon called Hierarchy Problem [4]. In its implementation, the MSSM require expansion to the Higgs Sector. The extended sector is called Two-Higgs-Doublet Model (2HDM) of type 2 which is composed of five Higgs Bosons, light scalar ($h$), heavy scalar ($H$), pseudoscalar ($A$), and two charged ($H^\pm$) [5]. This theoretical background is one of the reason why experimentalists need input from the phenomenologist regarding how the Higgs event data should be treated.

Neural Network (NN) is one of many Machine Learning method. It learns to do a specific task by considering examples without being programmed with any task-specific rule [6]. The idea of using NN...
is because it has been widely used in phenomenological High Energy Physics for a long time and proven to be reliable [7].

The data used in this research is simulated according to the best theory so that it behaves and possesses properties like actual data that come from an actual detector that detects real physical events. Simulated data is used because real data do not come with a label, whether they represent a scalar or pseudoscalar Higgs event. In order to build a NN algorithm, we need to train the machine on how to separate signal (pseudoscalar Higgs coupling or $t\bar{t}A$) and background (scalar Higgs coupling or $t\bar{t}H$). That’s why it requires data with label on them to learn. Then, we will extract from the data the 3-momentum of all the final state particles in the associate production of Higgs Boson with a pair of top-antitop quark $pp \to t\bar{t}H/ t\bar{t}A$ with center of mass energy of $\sqrt{s} = 13$ TeV which will then be used as input for the NN. The Feynman diagram of both reactions is shown in figure 1. In both reactions we used the semileptonic final state where one top quark decays leptonically ($t \to b e^+\nu_e$ or $t \to b \mu^+\nu_\mu$) and another top quark decays hadronically ($t \to b q\bar{q}$). The only Higgs decay channel considered in this research is the decay to a pair of bottom-antibottom quark $H/A\to b\bar{b}$ which is the dominant decay channel of Higgs.

The novelty of this research is in the final state used, the semileptonic final state with one neutrino and one electron/muon compared to previous related research [8]. For the record, observation of associate production of Higgs with a pair of top quark has only been confirmed in 2018 at ATLAS [9] and CMS [10].

2. Event generation, detector simulation and event selection
Event data used in this research is generated from Monte Carlo simulation. The simulation was modeled using MadGraph5_aMC@NLO [11] for matrix element calculation with next-to-leading-order (NLO) accuracy in QCD. The mass of scalar and pseudoscalar Higgs was set at 125 GeV and only the decay to a pair of bottom-antibottom quark ($H/A\to b\bar{b}$) was considered.

After generating the physical event at parton level, particles in the final state will be converted to the upcoming long-range physics process depending on the underlying physical event. Such as decay process after the reaction or the conversion of quarks and gluon into jets. A program named PYTHIA [12] was used for the modeling.

Resulting events will then be passed on to a generic detector framework named DELPHES [13] to simulate a real detector response if it were to detect actual events. DELPHES simulates response of a certain detector consisting of an inner tracker, electromagnetic calorimeter (ECAL), hadronic calorimeter (HCAL), and a muon system.Each with their own sub-detector responses.

![Figure 1. One of many Feynman diagrams for $t\bar{t}H$ process (left) and $t\bar{t}A$ process (right).](image-url)
Out of all the 50000 events generated from the simulation process above, only some of them are used as input. Event will be filtered to enrich the signal significance because even though both signal and background are generated by the same amount in this simulation, in real data they will be highly outnumbered by the actual background which is currently beyond the scope of this research. The result of the selection is presented in table 1.

There are a set of selection criteria for the events used in this research. Events used in this research are required to have a definite semileptonic final state. Also, recorded lepton in the events are required to have transversal momentum of \( p_T > 25 \) GeV for electron and muon. Tau lepton has been excluded from consideration and any event containing tau lepton will be dismissed due to the fact that it could decay hadronically. Next, event will be checked for the presence of additional lepton. If there exist an additional lepton(s), the additional lepton will need to be considered insignificant for the event to continue be used, that it needs to have transversal momentum below a significance value of \( p_T < 10 \) GeV.

Apart from the above leptonic selection, events are also required to contain at least 6 jets and at least 4 of them are b-tagged jets or jets that come from b quark. Other than that, each six of those jets are required to have a minimum transversal momentum of \( p_T > 25 \) GeV. For event with more than 6 jets and more than 4 b-tagged jets, the b-tagged jets will be gathered and ranked based on its momentum and 4 highest jets are designated as b jets and the remaining jets will again be gathered and ranked based on momentum and 2 highest will be designated as the remaining required jets and the rest will be abandoned.

### 3. Neural network construction

The NN was constructed using a toolkit named TMVA [14] from a framework named ROOT [15]. As can be seen in figure 2, the network architecture consist of 1 input layer, 2 hidden layer, and 1 output layer. The first input layer are made up of 24 neurons, each representing the input variable plus one additional neuron as bias node. Next, the first hidden layer are made up of 25 neuron plus 1 additional bias node. Then, the second hidden layer which consist of 8 neuron plus 1 additional bias node. Finally, it is ended with one output layer with one output neuron. The type of NN architecture used is called the Multilayer Perceptron (MLP) where every neuron in one layer can only be connected with the neuron in neighboring layer.

Actually, for a simple NN one hidden layer is enough to approximate any continuous discriminating function. However, the choice of number of layer and also number of neuron in a layer in this research are determined by a simple process of trial and error where we tried a lot of neuron-layer configuration and ended up with the current layer configuration which is the one with the highest signal labelling efficiency.

Out of 3986 events after selection used as input to the NN, 2000 events were used for training and 1986 for testing. In the training process, there are 842 weights that need to be fitted for every connection between neuron for the best discrimination, \( 25 \times 25 \) (from input layer and the first hidden layer connection) + \( 26 \times 8 \) (from the first hidden layer and the second hidden layer connection) + \( 9 \times 1 \) (from the second hidden layer and output layer). Training consists of 600 cycles (epoch) because the training already converge as the cycle turns to 600.

| Initial event number | Final event number | \( \epsilon_{\text{selection}} \) |
|----------------------|--------------------|------------------|
| \( t\bar{t}A \)      | 50000              | 2132             | 4.264 %          |
| \( t\bar{t}H \)      | 50000              | 1854             | 3.708 %          |
On network construction, there are 24 neurons used as discriminating variables for the NN. Variables used are just the 3-momentum of the final state particles and, as can also be seen in figure 3 and figure 4, they do not show any significant separation between signal and background. These variables are the transverse momentum $p_T$, azimuthal and polar component $\phi$ and $\eta$ (pseudorapidity) of the outgoing final state particles except neutrino due to inability for the detector to completely detect it hence 1 degree of freedom is lost.

**Figure 2.** NN architecture with 2 hidden layer

**Figure 3.** Comparison of values between signal ($t\bar{t}A$) and background ($t\bar{t}H$) part 1. Values from signal and background are almost overlapping in all variables.
Figure 3 (continued). Comparison of values between signal (t\bar{t}A) and background (t\bar{t}H) part 1. Values from signal and background are almost overlapping in all variables.

What is left to detect from neutrino is the two remaining components of the momentum represented in the variable of transverse missing energy \(MET\) and its azimuthal angle \(MET_\phi\). Finally, the last variable is the summation of all the jet momentums in an event including the excess jets that was ignored in event selection \(H_T\).

The discrimination strength of those variables can be guessed from how much signal and background histogram doesn’t overlap in figure 3 and figure 4. Even after those variables are ranked by their discriminating strength, the strongest variable does not give us a significantly better separation than the weakest variable.

4. Results and discussion
The result of the NN construction is a probabilistic equation that gives each event a designated value from 0 to 1 which is the probability of the event to be a signal (pseudoscalar). The output function can also be seen as a 23-dimensional surface that divide a 24 dimensional space from the number of input variables.

Figure 5 summarizes the NN training (dots) and testing (histogram). Even though from the figure there is no overtraining and the training result is not quite bad, but from the testing histogram the separation is not good enough to discriminate signal and background. Signal histogram is still peaking in the near-zero region that is supposed to be background region. The signal labelling efficiency of the constructed NN on testing (on training) is 0.412 (0.630). This means for every 100 event classified as signal by the NN, only about 41 of them is actually signal. The bottom part of figure 5 shows us the change of significance, signal efficiency, and background efficiency in response to the cut applied on the NN output. The graph also tells us that signal efficiency are almost linearly reduced and only slightly higher than background efficiency as we shift to the signal region (near 1).
Figure 4. Comparison of values between signal ($t\bar{t}A$) and background ($t\bar{t}H$) part 2. Values from signal and background are almost overlapping in all variables.

This means by applying cut to the response, we are rejecting as much background as signal (only slightly more background are being rejected). A better result would show us that by applying cut we will reduce a much larger background efficiency while maintaining high signal efficiency.

As mentioned before, the inability for the constructed NN to give a satisfying result was caused by the lack of strong discriminating variables. It turns out to discriminate between pseudoscalar and scalar coupling of Higgs we need more than just the 3-momentum of the final state particles.
Figure 5. Summary of the discrimination result of the NN. (a) The output of the NN in separating signal and background during training and testing, (b) Effects on applying cut to output.

From the Feynman diagram, the definite difference between the two lies on the vertices of $H$ and $A$ where the pseudoscalar vertices contains $i\gamma^5$ factor while scalar one doesn't. Therefore, in order to find a more discriminating variables, we need to do reconstruction of the 3-momentum of each pre-final state
particles all the way up to the top quark. After reconstruction, one example of suspected significant variable is the angular distribution between the top quark and Higgs.

5. Conclusion
The Neural Network Machine Learning algorithm constructed was not good enough to discriminate between pseudoscalar coupling Higgs Boson and scalar coupling Higgs Boson using input variable of 3-momentum of all the final state particles (minus neutrino). The result of efficiency of signal labeling on testing (training) is 0.412 (0.630). Further variables reconstruction and calculation of all the particles 3-momentum all the way up to the top quark pair and/or more simulated data will be needed to improve the result of this research.

Acknowledgments
This work was financially supported by Universitas Indonesia under research grant Q1Q2 for T M with grant contract number NKB-0277/UN2.R3.1/HKP.05.00/2019.

References
[1] Altarelli G 2005 The Standard Model of Particle Physics arXiv:hep-ph/0510281 [hep-ph] available at https://arxiv.org/pdf/hep-ph/0510281.pdf
[2] Rosenfeld R 2017 Physics Beyond the Standard Model arXiv:1708.00800 [hep-ph] available at https://arxiv.org/ftp/arxiv/papers/1708/1708.00800.pdf
[3] Vempati S K 2012 Introduction to MSSM arXiv:1201.0334 [hep-ph] available at https://arxiv.org/pdf/1201.0334.pdf
[4] Bhattacharyya G 2017 Pramana 89 53
[5] Spira M 2017 Higgs Boson Production and Decay at Hadron Colliders arXiv:1612.07651v2 [hep-ph] available at https://arxiv.org/pdf/1612.07651.pdf
[6] Gerven M v and Bohte S 2017 Front. Comput. Neurosci. 11 114
[7] Therhaag J 2013 EPJ Web Conf. 55 02005
[8] Santos S A d et al. 2017 Probing the CP Nature of the Higgs Coupling in t̅tH Events at the LHC arXiv:1704.03565v1 [hep-ph] available at https://core.ac.uk/download/pdf/159405435.pdf
[9] The ATLAS Collaboration 2018 Observation of Higgs Boson Production in Association with a Top Quark Pair at the LHC with the ATLAS Detector arXiv:1806.00425 [hep-ex] available at https://arxiv.org/pdf/1806.00425.pdf
[10] The CMS Collaboration 2018 Observation of t̅tH Production arXiv:1804.02610 [hep-ex] available at https://arxiv.org/pdf/1804.02610.pdf
[11] Alwall J et al. 2014 The Automated Computation of Tree-level and Next-to-leading Order Differential Cross Sections, and Their Matching to Parton Shower Simulations arXiv:1405.0301v2 [hep-ph] available at https://arxiv.org/pdf/1405.0301.pdf
[12] Sjostrand T et al. 2014 An Introduction to PYTHIA 8.2 arXiv:1410.3012v1 [hep-ph] available at https://arxiv.org/pdf/1410.3012v1.pdf
[13] Favereau J d et al. 2014 DELPHES 3 : A Modular Framework for Fast Simulation of a Generic Collider Experiment arXiv:1307.6346v3 [hep-ex] available at https://arxiv.org/pdf/1307.6346.pdf
[14] Hoecker A et. al. 2009 TMVA: Toolkit for Multivariate Data Analysis with ROOT arXiv:physics/0703039v5 [physics.data-an] available at https://arxiv.org/pdf/physics/0703039.pdf
[15] Brun R and Rademakers F 1997 Nucl. Inst. Meth. Phys. Res. A 389 81-6