Machine learning approach for the search of resonances with topological features at the Large Hadron Collider

S E Dahbi[1], J Choma[1], B Mellado[1,2], G Mokgatitswane[1], X Ruan[1] and B Lieberman[1]

[1]School of Physics and Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa.
[2]iThemba LABS, National Research Foundation, PO Box 722, Somerset West 7129, South Africa.

E-mail: salah_eddine.dahbi@cern.ch

Abstract. We propose a new approach to search for new resonances beyond the Standard Model (SM) of particle physics in topological configurations using Machine Learning techniques. This involves a novel classification procedure based on a combination of weak-supervision and full-supervision in conjunction with Deep Neural Network algorithms. The performance of this strategy is evaluated on the production of SM Higgs boson decaying to a pair of photons inclusively and exclusive regions of phase space, for specific production modes at the Large Hadron Collider (LHC), namely through the gluon-gluon fusion, the fusion of weak vector bosons, in associated production with a weak vector boson, or in association with a pair of top quarks. After verifying the ability of the methodology to extract different Higgs signal mechanisms, a search for new phenomena in high-mass diphoton final states is setup for the LHC.

1. Introduction

After the discovery of a Higgs boson (h) [1, 2, 3] at the Large Hadron Collider (LHC) by the ATLAS [4] and CMS [5] experiments, a number of anomalies have been identified based on discrepancies in the production of leptons at these experiments [6]. These anomalies join a number of phenomena that represent a significant experimental evidence, such as Dark Matter, the origin of neutrino mass, the matter-anti-matter asymmetry, in addition to a number of theoretical problems [7].

Machine Learning (ML) can play a significant role in solving these problems, by searching for resonances in corners of the phase-space that are either unexplored or poorly covered by the cut-based analysis strategy. Therefore, in this paper we extend the applications of machine learning in high energy physics to the search for new physics at the LHC, using weak supervision learning with topological requirements. The approach is intended to identify different phase-space of the resonances, since they are expected to be generated with different production mechanisms at the LHC and it is adequate to extract more subtle signals in the data without relying on a specific model of the signal with a particular set of parameters. We have organised this paper as follows: section 2 describes the simulation of both background and signal processes of the production of the SM Higgs boson decaying to a pair of photons at the LHC, as a benchmark. Section 3 briefly introduces weak supervision learning classification and the used deep neural network (DNN) architecture. The performance of the weak supervision classification will be evaluated in section 4. Section 5 proposes an approach for the search for new physics, where weak supervision learning is restricted to the topological configurations of the phase space, such as number of b-tagged jets, number of leptons, vector boson fusion topology, etc... Finally, section 6 summarises the study with suggestions and prospects for the future search of resonances at the LHC.

2. SM Higgs datasets and event selection

The performance of the ML classifier, described in section 3, is tested on simulated Higgs to di-photon ($pp \rightarrow h \rightarrow \gamma\gamma$) events as a benchmark. We focus on the proton-proton collisions at the LHC with a centre-of-mass energy of 13 TeV. Background and signal Samples were produced using `MadGraph5_aMC@NLO 2.6.7` with Next-to-Leading Order accuracy in QCD [8]. The parton showering and hadronization were simulated with `PYTHIA 8.2` [9] using ATLAS A14 event tuning and NNPDF2.3 LO parton distribution function set [10]. Events were processed with `Delphes 3` [11], which provides an approximate fast simulation of the current ATLAS experiment. Events of interest were simulated by imposing a set of generator-level cuts, where the transverse momentum of the photons is required to be greater than 25 GeV and the di-photon invariant mass to be between 105 and 160 GeV. Hadronic jets were reconstructed using the anti-kt algorithm [12] with the radius parameter, $R = 0.4$, as implemented in the `FastJet` 3.2.2 [13] package. Jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$ are considered. In addition, jets originating from bottom quarks are identified as b-jets with b-tagging algorithms [14]. Reconstructed jets overlapping with photons, electrons or muons in a cone of size $R = 0.4$ are removed. Electrons and muons are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Finally, an inclusive event selection of at least two photons, that have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.37$, is applied. The reconstructed di-photon invariant mass spectra, $m_{\gamma\gamma}$, for signal and background are shown in figure 1.

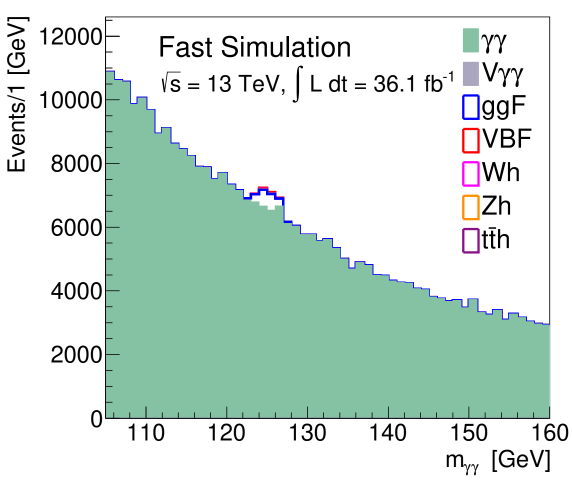

Figure 1. Di-photon invariant mass spectrum at 13 TeV of proton-proton collision, corresponds to an integrated luminosity of $36.1\,\mathrm{fb}^{-1}$. The different production mode of the Standard Model Higgs boson samples, are normalised to the cross sections times the di-photon decay branching ratio.

3. Weak supervision classification and DNN architecture

The main purpose of supervised learning is to train a model on fully labelled data where each training example $\vec{x_i}$ comes with a label $y_i \in \{0, 1\}$, in a case of binary classification task. The model is trained to minimise the loss function which can be in a form of binary cross-entropy:

$$\ell(y, \hat{y}) = -y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}), \tag{1}$$

where $\hat{y}$ is the model output and y is the target output (1 for signal and 0 for background). In the case of weak supervision [15], the learning phase during this paradigm takes place on partially/weakly labelled data set [16]. These weakly labelled data is then used by the classifier for the training. Contrariwise of the the full supervision (see equation 1), weak supervision uses the following equation:

$$f_{weak} = argmin_{f:\mathbb{R}^n \to [0,1]} \sum_K \ell \left(\frac{1}{|K|} \sum_{i \in K} \hat{y}_i, y_K \right), \tag{2}$$

where K denotes training data batches and y_K is the signal ratio in each batch. The performance of weak supervision was compared to that of full supervision in [18] and weak supervision achieved reasonable results, despite knowing only a fraction of the labels.

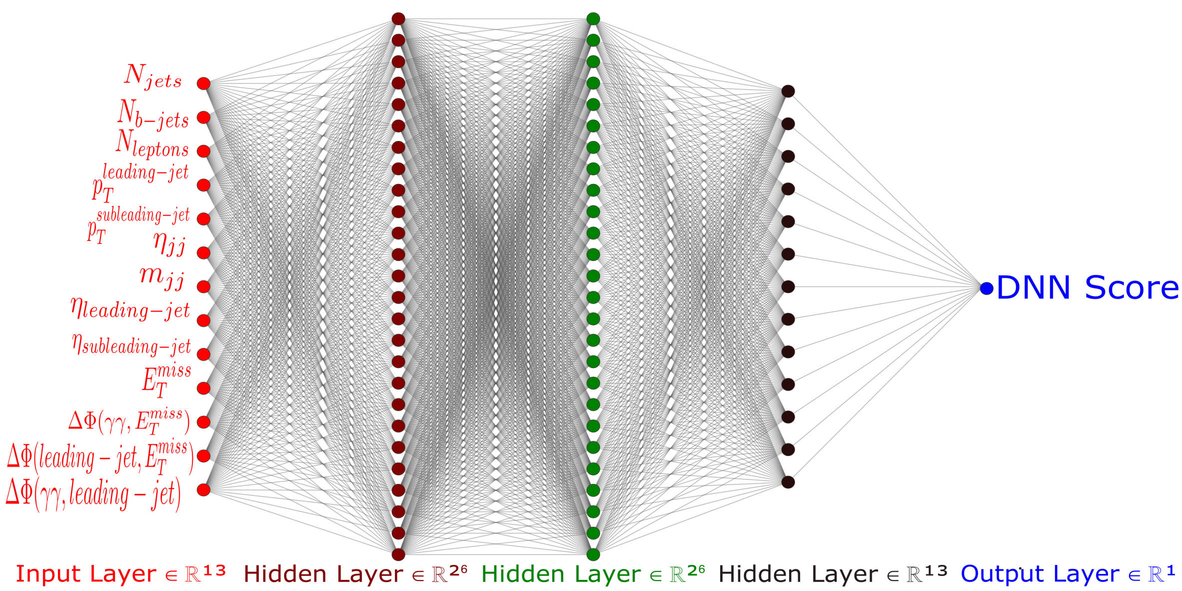

Figure 2. DNN architecture for the classification of events as background or signal in this study, with the `"Layout=TANH|26,TANH|26,TANH|13,LINEAR" configuration.`

Events from the simulated dataset are classified using a DNN algorithm, as implemented in the package ROOT [19] and the TMVA [20] data processing framework. The DNN architecture implemented in this study consists of an input layer, three hidden layers and an outer layer. The input layer consists of 13 neurons, which represent the kinematic features of the dataset. The three hidden layers are made up of 26, 26 and 13 neurons, respectively. The output layer consists of a single neuron. Two activation functions are used here: a hyperbolic tangent for the hidden layers, and a sigmoid function for the output layer. Since we are dealing with a binary classification problem, where the events are classified as signal or background, the cross-entropy (see equation 1) is used as a loss function. The training of the classifier is divided into three stages. The first stage begins with a learning rate of 10^{-2} and a momentum of 0.9. In order to avoid overtraining the classifier, the weights are regularised using the $L2$ option, which multiplies a scaling factor of 10^{-3} to the norms of the weight matrices. For the second training stage, learning rate and momentum are reduced to 10^{-3} and 0.5, respectively. For the final training phase, the learning rate is set to 10^{-2} and the momentum to 0.3. Figure 2 describes the network topology and the input variables used in this study.

4. Weakly supervised learning for the classification of the SM Higgs boson

For the weak supervised learning study on the Higgs production mechanisms, the signals and background processes are mixed in the Higgs invariant mass window of 120 GeV to 130 GeV. The sideband contains only pure background in the region between 105 GeV and 160 GeV, excluding the signal mass window. The purpose of this procedure is to evaluate the ML algorithm's ability to classify a signal from an unknown mixture of signal and background, with respect to pure background in the sideband. According to the receiver operator characteristic (ROC) analysis results shown in figure 3, it is evident that weak supervision is capable of separating different signal candidates. However, the signal and background classification is not satisfactory when compared to full supervised learning (see table 1). Given these points, weak supervision seems to be inefficient due the complexity of the phase-space explored here. For that reason a guided weak supervision method with topological requirements is introduced. This will be discussed explicitly in section 5.

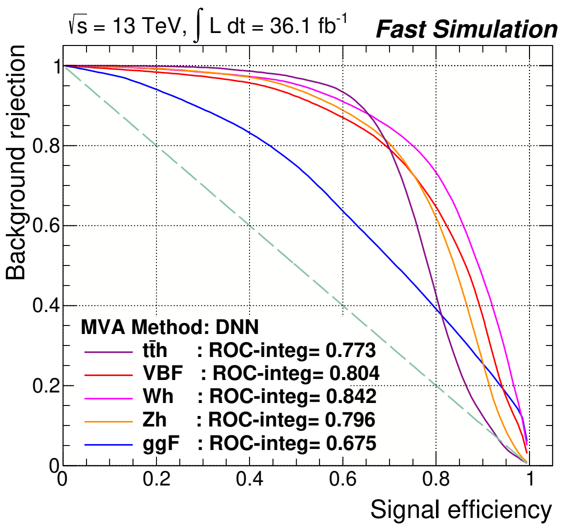

Figure 3. Weakly supervised ROC curves, using DNN classifier for different Higgs production mechanisms.

Table 1. Comparison of the integral of the ROC curve response from different supervised learning methods.

Process	Full Supervision	Weak Supervision
ggF	0.780	0.675
VBF	0.965	0.804
Wh	0.951	0.842
Zh	0.936	0.796
$t\bar{t}$h	0.997	0.773

ISBN: 978-0-620-97693-0

5. Weak supervision learning with topological requirements

The performance of the weak supervised learning seems to be inefficient to separate different signal candidates, due the complexity of the phase-space explored here. Therefore there is a need to implement the method that we refer to as the guided weak supervision method. In guided weak supervision, the training is restricted to the topological signatures derived from selected production mechanisms. The effectiveness of this methodology is evaluated and verified on the VBF, Wh and $t\bar{t}h$ production mechanisms. The results are quantified and compared to the inclusive weak supervision in table 2.

The SM Higgs boson production mode through VBF represents the second largest cross section at the LHC. The VBF mechanism in the SM entails the scattering of two quarks, leading to two hard jets in the forward and backward regions of the ATLAS detector with high di-jet invariant mass, m_{jj}. These topological characteristics can be exploited by the DNN to distinguish VBF from backgrounds and other production mechanisms. In this context, the VBF topology is defined as events with at least two jets. This is followed by a signal significance scan using m_{jj} and the pseudorapidity separation between the two leading jets, $\Delta\eta_{jj}$. We use the signal significance scan to maximise the relative contribution of the VBF production mechanism and minimise the background yield as much as possible.

A similar test is performed on the associated production of Higgs boson, with a W gauge boson, where the neural networks are trained separately with two topological features from the Wh final states. This includes leptonic and hadronic decays of the W boson. The performance of weak supervision classification also experiences improvements with topological requirements and is able to isolate the Wh process.

In addition, the guided weak supervision is tested on the production of the Higgs boson in association with top-antitop quarks. Since the relative contribution of the top-antitop quark to the total Higgs boson production is less than 1%, it becomes challenging to extract this signal process. Most prominently, this is found to be the case for weak supervision. However, the test is performed by requiring two topological characteristics, representing a fully hadronic (at least two central jets) and semi-leptonic (at least one b-jets and one lepton) decays of top quark from $t\bar{t}h$ process.

Table 2. Summary of the ROC integrals for VBF, Wh and $t\bar{t}h$ with different topological requirements, compared to inclusive weak supervision learning.

Process	Weak Supervision	ROC Integral
VBF	Inclusive	0.804
	$N_{jets} \geq 2$, $\Delta\eta_{jj} > 0$, $m_{jj} >$ 1000 GeV	0.951
	$N_{jets} \geq 2$, $\Delta\eta_{jj} > 0$, $m_{jj} >$ 800 GeV	0.881
	$N_{jets} \geq 2$, $\Delta\eta_{jj} > 0$, $m_{jj} >$ 600 GeV	0.912
	$N_{jets} \geq 2$, $\Delta\eta_{jj} > 0$, $m_{jj} >$ 400 GeV	0.873
Wh	Inclusive	0.842
	Hadronic	0.907
	Leptonic	0.902
$t\bar{t}h$	Inclusive	0.773
	$N_{b-jets} \geq 1$, $N_{leptons} \geq 1$	0.880
	$N_{b-jets} \geq 2$	0.950

6. Discussion and conclusions

Machine learning may play a significant role in the deeper exploration of the phase-space available at the LHC. We introduce signatures and loose topological requirements before the implementation of weak supervised training. The impact on the signal efficiency and background rejection, in this approach, is evaluated for the different production mechanisms, where significant improvements are observed with respect to the implementation of weak supervision inclusively. This is referred to as guided weak supervision. While weak supervision, as setup here, has the advantage of not relying on a model for the signal it is necessary to restrict the phase-space where the sideband and the signal region are confronted with each other. While signature and topological requirements are driven by physics considerations, the search is not biased by the phenomenology of a model with a particular set of parameters.

The performance of the approach has been evaluated with the SM Higgs boson as a benchmark. The signatures and topologies used here in principle can be used in the search for heavier or lighter SM Higgs-like bosons. Provided that the topological requirements are loose enough, these can be used for more generic searches of bosons. For instance, it is well motivated to apply weak supervision in the presence of a b-tagged jet, while requiring a maximum amount of jets in order to remove potential signals in association with top quarks. This can be extended to other types of searches, such as search for resonances in multi-particle cascades.

Acknowledgments

The authors are grateful for support from the South African Department of Science and Innovation through the SA-CERN program and the National Research Foundation for various forms of support. The authors are also indebted to the Research Office of the University of the Witwatersrand for grant support.

References

[1] P. W. Higgs, Phys. Lett. **12**, 132-133.
[2] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321-323.
[3] P. W. Higgs, Phys. Rev. Lett. **13**, 508-509.
[4] G. Aad *et al.* [ATLAS], Phys. Lett. B **716**, 1-29, [arXiv:1207.7214 [hep-ex]].
[5] S. Chatrchyan *et al.* [CMS], Phys. Lett. B **716**, 30-61, [arXiv:1207.7235 [hep-ex]].
[6] S. von Buddenbrock, R. Ruiz and B. Mellado, Phys. Lett. B **811**, 135964, [arXiv:2009.00032 [hep-ph]].
[7] S. Troitsky, Phys. Usp. **55** (2012), 72-95 [arXiv:1112.4515 [hep-ph]].
[8] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, JHEP **07**, 079, [arXiv:1405.0301 [hep-ph]].
[9] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen and P. Z. Skands, Comput. Phys. Commun. **191**, 159-177, [arXiv:1410.3012 [hep-ph]].
[10] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre and J. Rojo, *et al.* Nucl. Phys. B **867**, 244-289, [arXiv:1207.1303 [hep-ph]].
[11] J. de Favereau *et al.* [DELPHES 3], JHEP **02**, 057, [arXiv:1307.6346 [hep-ex]].
[12] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04**, 063, [arXiv:0802.1189 [hep-ph]].
[13] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896, [arXiv:1111.6097 [hep-ph]].
[14] L. Vacavant [ATLAS], PoS **2008LHC**, 064.
[15] L. M. Dery, B. Nachman, F. Rubbo and A. Schwartzman, JHEP **05**, 145, [arXiv:1702.00414 [hep-ph]].
[16] Patrini, Giorgio and Nielsen, Frank and Nock, Richard and Carioni, Marcello, International conference on machine learning, arXiv:1602.02450, pages:708–717.
[17] Junwei Han, Dingwen Zhang, Gong Cheng, Lei Guo, Jinchang Ren, IEEE Transactions on Geoscience and Remote Sensing, **53**, 6, 3325–3337, 2014, IEEE.
[18] T. Cohen, M. Freytsis and B. Ostdiek, JHEP **02**, 034 [arXiv:1706.09451 [hep-ph]].
[19] R. Brun and F. Rademakers, Nucl. Instrum. Meth. A **389**, 81-86.
[20] J. Therhaag [TMVA Core Developer Team], AIP Conf. Proc. **1504**, no.1, 1013-1016 (2012)