Identifying Dimuon Higgs Decay in Association with Top Quark Pair Using Machine Learning

Paula Bosca[1, 2, 3, *] and supervised by Dr. Pierre Savard[2, 3]

[1] *McMaster University*

[2] *The University of Toronto*

[3] *The European Organization for Nuclear Research (CERN)*

(Dated: August 2021)

Determining all measurable properties of the Higgs boson is one the ways in which we can further test the Standard Model and evaluate additional theories. Specifically, precise measurements of the Higgs coupling to second-generation fermions is the next crucial step and can be investigated via the dimuon Higgs decay channel. Here we present two methods created to examine the dimuon Higgs decay when produced in association with a top-antitop quark pair and differentiate this signal from background top-antitop pair production. The first method involves the use of an algorithm which minimizes a chi-squared-like variable. This showed some discrimination ability but was deemed insufficient for practical use. The second method employs a machine-learning algorithm and displayed correct classification of signal and background events 84 and 85 percent of the time, respectively. Future retraining the neural network with more events and a greater variety of background decays may result in accuracy rates capable of detecting signal events on ATLAS Run II and III data.

I. INTRODUCTION

In 2012, a novel particle with an approximate mass of 125 GeV was discovered [1]. This particle, discovered by both ATLAS and CMS experiments at CERN, has been shown consistent with Standard Model (SM) predictions of the Higgs boson [2, 3]. This Higgs boson describes how Standard Model particles obtain their mass via spontaneous symmetry breaking [4, 5]. In particular, the coupling of the Higgs field to all SM particles (both fermions and bosons) results in their respective masses [6]. Accurate measurements of these couplings allow for rigorous new tests of the SM itself, in addition to probing additional theories whereby multiple Higgs bosons exist [7]. Specific couplings can be investigated through measurements of Higgs production and decay modes.

Since its discovery, the Higgs boson coupling to third-generation fermions have been observed and established consistent with SM predictions [8–10]. However, couplings of this new scalar boson to every particle are necessary for a robust understanding. As such, this study explores a method to enable investigations into the Higgs coupling to muons, a second-generation fermion.

* boscap@mcmaster.ca

The muon is the ideal second-generation fermion on account of the many difficulties associated with detecting strange/charm quarks or the muon neutrino.

Here, we outline our investigation into a specific Higgs production and decay: $t\bar{t}H$, $H \to \mu\mu$. We present a method of identifying events whereby a Higgs boson decays to $\mu^+\mu^-$. The dimuon pair offers a distinct signature in the ATLAS detector: two oppositely charged, isolated muons emerging in opposite transverse directions. The muon calorimeters and spectrometers in ATLAS provide excellent readouts for such signals. However, while this decay presents a clear signature, it is extremely rare. Of the possible hypothesized Higgs decays, the branching ratio of Higgs to $\mu\mu$ is one the smallest, of order 10^{-4} [11]. It thus poses a significant challenge and requires sophisticated detection methods.

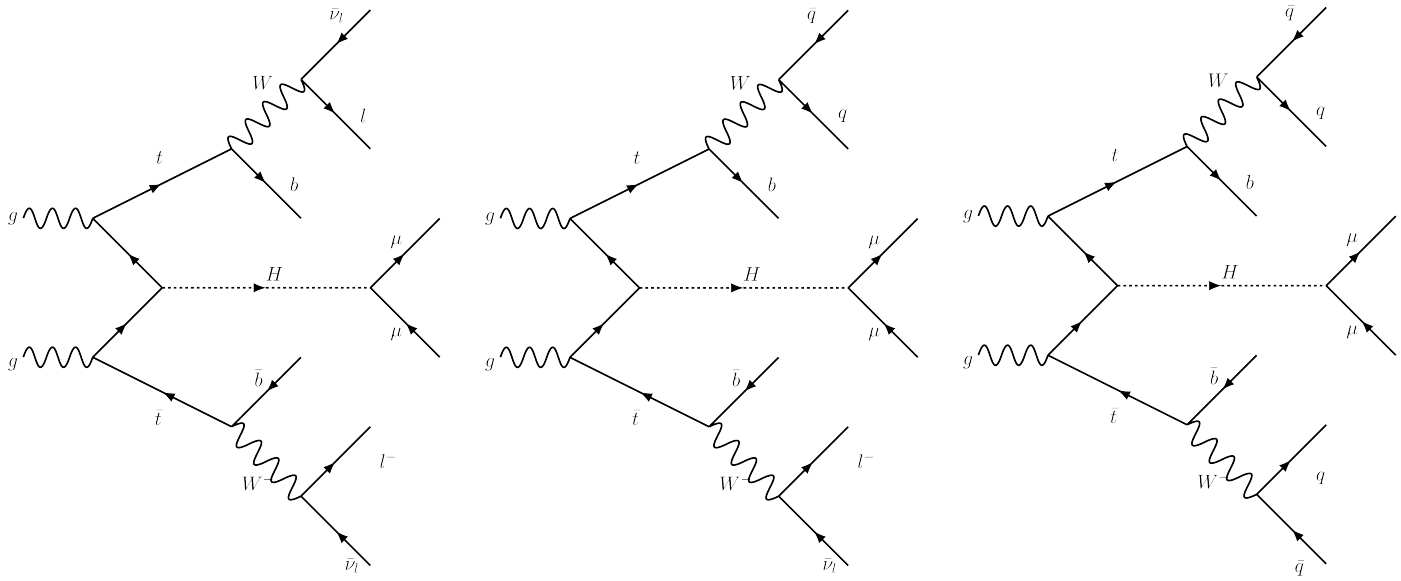
A. $t\bar{t}H$ Production and Decay

All Higgs production modes should be considered to ensure the detection of all possible $H \to \mu\mu$ events. Here, we focus on the production of a Higgs in association with a top-quark pair, namely, $t\bar{t}H$. While this production mechanism is also rare, it once again provides a distinct signature.

$t\bar{t}H$ involves a top-antitop quark pair produced at the same vertex with a Higgs boson, as shown in Figure 1. Due to their incredibly short lifetime, top quarks cannot be detected directly by ATLAS. Instead, decay products provide the distinct signature. Top quarks decay via $t \to bW$ over 99% of the time. This decay creates reliable jets that can be classified as having originated from b quarks, dubbed bjets. A wealth of literature is dedicated to the algorithms involved in classifying bjets [12–14]. It will not be reviewed here, only to say that the longer lifetime of b quarks creates a secondary vertex upon hadronization, thus allowing for identification. Thus, in conjunction with a dimuon signal, the $t\bar{t}H$ decay includes at least two b-tagged jets.

There are two main decay modes for the W boson, resulting in 3 possible channels for the full $t\bar{t}H$ mode due to the combinatorics of the top quark pair. These three possibilities are shown in Figure 1. $t\bar{t}H$ decay modes are labelled by the products of the W boson, which can decay either leptonically, $W \to l\nu$, or hadronically, $W \to qq$. If both tops decay leptonically or hadronically, the $t\bar{t}H$ decay is said to be all-leptonic or all-hadronic, respectively. Furthermore, if the top quark pair decay via different channels, the whole $t\bar{t}H$ decay is said to be semi-leptonic. For any of these three decays, the sheer number of decay products constitutes a unique signal. The decay products are expected to exhibit high energies due to their originating from massive top quarks, adding an extra uniqueness to this decay.

Here, we investigate methods to identify any of these three $t\bar{t}H$ decay modes (where $H \to$



(a) Fully leptonic decay *(b) Semi-leptonic decay* *(c) Fully hadronic decay*

Figure 1: Feynman diagrams describing the signal decays included in the present analysis. All decays include a Higgs produced in association with a top quark pair, followed by a dimuon decay of the Higgs boson, $t\bar{t}H,\ H \to \mu\mu$.

$\mu\mu$), with specific emphasis on the fully-hadronic decay as that is the most probable. The background source considered here is the common production of top-quark pairs, $t\bar{t}$. The decay modes of $t\bar{t}$ are the same as those for $t\bar{t}H$, without the dimuon signal from the Higgs. Thus, $t\bar{t}$ decays can also be categorized as fully leptonic, fully hadronic, semi-leptonic. While the dimuon signal reconstructing to the Higgs mass is the starkest difference between $t\bar{t}H$ and $t\bar{t}$, we expect further distinctions in the boosted nature of $t\bar{t}H$ products.

II. DATASETS AND DECAYS

Full-simulation Monte Carlo (MC) events were used as data for signal and background in all analyses presented herein. MC data is such that it resembles ATLAS detector output, thus allowing for the same analysis to be run on simulated and real datasets without any significant changes. Signal data produced contained all three possible $t\bar{t}H,\ H \to \mu\mu$ decays listed above but were not labelled as such. MC data used for background only contained fully leptonic $t\bar{t}$ decays. One of the reasons for this choice was the similarity in signature as the final leptons produced from the W decay could be muons and thus mimic signal data more closely.

Requirements regarding event selection were identical for both signal and background. The cutoffs outlined in Table I were implemented to select for the expected highly energetic products. Events that did not satisfy all criteria were ex-

cluded from the analysis. The final dataset included approximately 197 000 signal and 404 000 background events.

III. METHOD ONE: CHI-LIKE VARIABLE

The first method to differentiate signal from background reconstructed W and top particles using a chi-like variable. This method was motivated by Husemann 2017[15] and is reproduced here in Equation 1.

$$\chi^2 = \frac{(W_{SM} - W_{recon})^2}{\sigma_W^2} \tag{1}$$

The equation above seeks to minimize a chi-like variable which is calculated by taking the difference of the reconstructed invariant mass of a W, W_{reco} from its SM prediction, W_{SM}. This value is normalized by the predicted SM natural width of the W mass, σ_W. Here, this formula is extended to include two W bosons and two top quarks, as described by Equation 2, to calculate a chi-like variable for the entire $t\bar{t}H$ decay.

Table I: Minimum requirements for event inclusion. Min. count represents the minimum number of objects which must pass the threshold cutoff, where pT is transverse momentum, WP is working point, and the same pT cutoff applies for both btagged and non-btagged jets.

	Muons	Non-btagged Jets	btagged Jets
Min. count	2	4	2
Cutoff	30 pT	25 pT	85 WP

$$\chi^2 = \frac{(W_{SM,1} - W_{recon,1})^2}{\sigma_W^2} + \frac{(W_{SM,2} - W_{recon,1})^2}{\sigma_W^2} + \frac{(t_{SM,1} - t_{recon,1})^2}{\sigma_t^2} + \frac{(t_{SM,2} - t_{recon,1})^2}{\sigma_t^2} \tag{2}$$

As mentioned previously and in Table I, all events contained two oppositely charged muons, at least two b-tagged jets, and at least four non-btagged jets, each satisfying their respective thresholds listed. For every event, a W boson was reconstructed from the kinematics of a pair of non-btagged jets. This W boson was then paired with a bjet to reconstruct a top quark. This was performed a second time with a different combination of jets for a total reconstruction of two unique W bosons and two unique tops. The combination (out of all possible for any given event) of four non-btagged jets and two bjets which minimized Equation 2 was selected.

Equation 2 initially utilized standard model predictions for W and top masses alongside their respective natural widths. However, this was changed to better reflect the data and detector resolution. Standard model values for the w and top mass variables were preserved, while the widths were approximated from the data. The invariant mass distributions of top and W bosons were plotted, and a Gaussian fit was applied to the distributions' core. The σ of the fit was then used as the new W and top widths, respectively. An example is shown in Figure 2, and the final

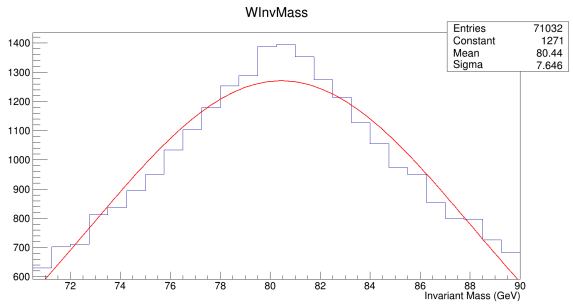


Figure 2: Reconstructed invariant W boson mass with the core distribution overlain with a Gaussian fit. Sigma was used for chi-like variable computation to more accurately reflect data resolution.

utilized numbers were 7.6 for W width and 17.3 for top width.

With these values for W_{SM}, t_{SM}, σ, the algorithm to minimize the chi-like variable was implemented on both signal and background to identify decay products.

This method was then evaluated by examining the invariant mass of the dimuon pair. As expected, the dimuon invariant mass distribution from signal events peaked sharply at about 125 GeV. In contrast, that of the background data peaked at 100 GeV with a much wider observed distribution. Reconstructed W and top mass distributions for signal and background also differed. This is suspected of being influenced by the reconstruction algorithm's assumption of fully hadronic decays while the background data contained only leptonic decays. The differences in all distributions, however, had significant overlaps and thus were not desirable.

The invariant mass distributions (dimuon, reconstructed W and top) were further inspected by examining only those events with a chi-like value less than some arbitrary cutoff. When these events were viewed closely, it was evident that the reconstruction algorithm would falsely reconstruct invariant masses. Attempts were made to optimize the algorithm by varying the value of this cutoff, but the fundamental problem remained. As such, an alternative approach was necessary to better differentiate signal and background events.

IV. METHOD TWO: MACHINE LEARNING

A deep learning feed-forward multi-layer perceptron utilizing back-propagation was constructed to classify signal and background events. The neural net consisted of 61 input variables, three hidden layers, and two output nodes.

Input variables included all kinematic variables for btagged and non-btagged jets: transverse momentum, pT, pseudorapidity, η, azimuthal angle ϕ, and energy, E. Additionally, the working point of the btag algorithm for all btagged jets was included, as well as event missing transverse momentum (energy and azimuthal angle), MET. The eight most energetic non-btagged jets and top four btagged jets which satisfy requirements outlined in Table I were included. If a given event contained less than the maximum number of included jets but satisfied the minimum, the excess variables were set to

zero. This was done to ensure a uniform input dimensionality for all events. Additional variables included the transverse momentum and pseudorapidity of the most energetic $\mu^{\pm}$. Muon variables ϕ and E were purposely excluded to prevent the algorithm from reconstructing the dimuon invariant mass and potentially classifying signal from background by identifying the Higgs boson. Furthermore, having the net blind to this invariant mass allows examination of the dimuon invariant mass distribution after classification as a measure of network performance.

The chi-like variable from the previous section was also included as a potential aid to the neural net. Similarly, the $\delta\phi$ and δR of the two reconstructed top quarks to increase network efficiency and accuracy, where $\delta\phi$ is the difference in azimuthal angle of the two objects, and δR defined by Equation 3.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3}$$

Key neural net properties are included in Table II. A train-test-split was performed whereby 70% of the input data was used for training the network, while the other 30% was used to test (evaluate) the net's performance on novel data. The training sample was further divided wherein 80% of the training data was used for training during epochs and the remainder used for cross-validation.

Due to the abundance of background data, class weights were applied to signal vs background data to prevent biasing the network to background data. This was achieved using the balanced class weight method where class weights are given by

$$\frac{\text{total events}}{(\text{number of classes})(\text{number of events in class } i)}$$

Standard features were additionally implemented to avoid overtraining. These include monitoring loss with early stopping and including dropout in the first two hidden layers. Here we present the optimal configuration which minimized validation loss after hyperparameter tuning. The network presented here was also compared with network configurations with one to eight hidden layers, which had been similarly hypertuned.

A softmax activation function was used for the two output nodes of the network. Softmax was chosen for its interpretation as a probability that a given event is from each class (where the probability of each class must sum to unity). Since this network was employed as a binary classifier, only the output of the signal-like class can be considered, where closer to 1 is more signal-like and 0 is background-like. All further model evaluation was performed using these outputs derived from the previously mentioned testing dataset.

A. Neural Network Evaluation

The algorithm's classification of all test data can be seen in Figure 3. Background and signal data are clustered around 0 and 1, respectively.

Table II: Neural network properties. Additional parameters include Learning rate: 1e-3, Optimizer: Adam, Loss: Sparse Categorical Cross-Entropy.

	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output
Neurons	300	300	300	2
Dropout	0.2	0.2	0	N/A
Activation Functions	relu	relu	relu	softmax

The clear differentiation of the network's output is an early suggestion of good performance.

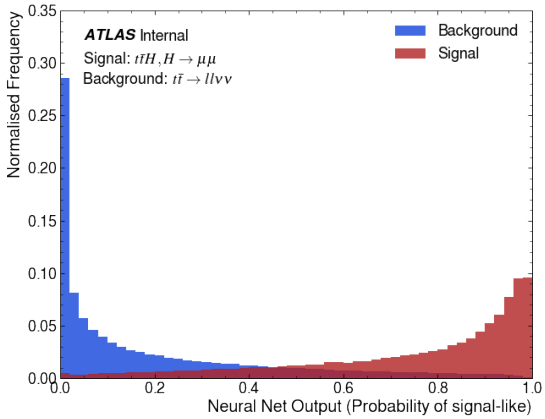


Figure 3: Network output distribution that a given event is signal-like, with signal in red and background as blue. Normalized frequency was performed per class, and is the fraction of events per class at each probability (bin).

Various metrics were utilized to evaluate the performance of the classifier. In Table III we report a confusion matrix where events were considered to be classified as signal-like if they had an output probability greater than 0.5. All other events were considered to be classified as background-like. The table is normalized by row, and we see fair classification of true positives and negatives at 0.85 and 0.84 respectively.

	Predicted	
True	Signal	Background
Signal	0.85 (TP)	0.16 (FN)
Background	0.16 (FP)	0.84 (TN)

Table III: Confusion matrix displaying network classification (predicted) compared to true values, normalized by row where TP, TN, FP, FN are true positive, true negative, false positive, false negative respectively.

We can define recall and precision from values in the confusion matrix, where TP, TN, FP, FN are true positives, true negatives, false positives, and false negatives, respectively.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall and precision can be further combined to define the f-score of the network, a standard metric used to evaluate classification algorithms. Equation 4 presents the definition and network's f-score.

$$\text{f-score} = 2\,\frac{\text{Precision * Recall}}{\text{Precision + Recall}} = 0.779 \quad (4)$$

V. CHARACTERIZING SIGNAL AND BACKGROUND DISTRIBUTIONS

In section IV A events were classified by the net as signal-like or background-like if the probability of signal-like classification was above or below 0.5. Here, we investigate adjusting this cutoff value and the resulting changes in the dimuon invariant mass distribution. Figure 4a is an example distribution plot for all events that were classified by the network as having a probability greater than 0.75 of being signal-like. We see the distribution dominated by correctly-classified signal events, which sharply peak at 125 GeV as expected. 4% of all background events and 62% of all signal events were classified above this cutoff, resulting in a 16:1 signal to background ratio.

Figure 4b displays all events excluded from Figure 4a(which classified probabilities less than 0.75). Figure 4b clearly shows the background events decaying with increasing mass, with some incorrectly-classified signal events clustering and creating a peak at 125 GeV. We can further compute the signal to background ratio by the change in respective events classified. Of all input events, signal was outnumbered to background by approximately 1:2. Implementing a 0.75 cutoff resulted in a ratio of 6:1. These results are summarized in Table IVfor a range of cutoffs from 0 to 0.9.

Additionally, these values are provided normalized to the total amount in their class. For example, 100% of background events and 100% of signal events are classified above a threshold of 0. However, for a threshold of 0.75, only 6% of the total background events and 65% of the signal events pass. These are given both as fractions and ratios in the table, where ratios are provided for easier interpretation.

VI. CONCLUSIONS

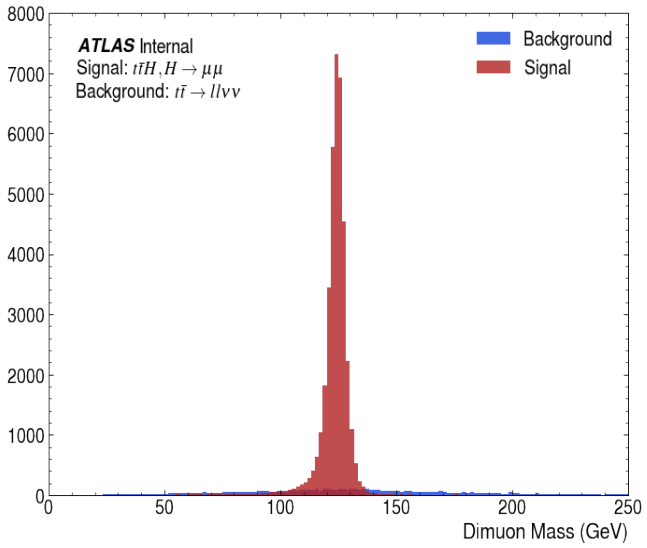
Here, we presented two methods to discriminate signal dimuon Higgs decay produced in association with a top-quark pair, $t\bar{t}H,\ H \to \mu\mu$, from a background of top quark pair leptonic decays, $t\bar{t} \to ll\nu\nu$. The first method involved creating an algorithm utilizing a chi-like variable to select objects to reconstruct essential decay products. This method was shown to be ineffective at discriminating background from signal. We present the creation of a neural network as a second method where critical outputs from the first method were used as network inputs. The network correctly identified signal and background events 82 and 83 percent of the time, respectively. The model is primarily limited by small datasets where an increase in event counts for both signal and background may result in model improvements. With a greater sample size, model complexity can increase without the risk of overtraining. Additional variables such

as high-energy leptons beyond the dimuon pair may also prove helpful. Feature importance of the networks where input variables are ranked on their relative importance to the model should also be performed. Like all metrics and analyses on machine learning algorithms, caution must be taken as feature importance is neural net specific, and any configuration changes may drastically change variable importance. Hyperparameter tuning should also be conducted for any new model configuration. Future model improvements should also consider incorporating a variety of new background sources. Key decays to improve the model's usefulness would include all $t\bar{t}$ decay modes, as well as $t\bar{t}Z$.

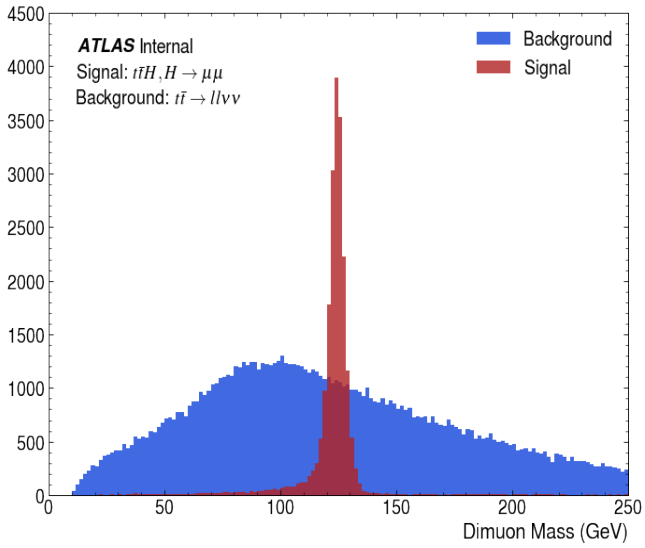
Furthermore, classifying signal into specific decay modes (training the network to classify an event as fully leptonic, semi-leptonic, or fully-hadronic) could provide further insights into branching ratios and decay modes. Finally, this analysis hopes to eventually merge with all others investigating the Higgs dimuon decay with the hopes of gaining more insight into the standard model and potentially beyond.

ACKNOWLEDGMENTS

The author would like to give immense thanks to Dr. Savard for his supervision, kindness, patience answering many questions, in addition to his invaluable teaching and feedback. Additionally, many thanks are owed to the entire ATLAS collaboration at the University of Toronto for their feedback. Finally, the author would like to thank the Institute of Particle Physics in Canada and the CERN Summer Student Program, as well as NSERC Canada for making this project possible.



(a) Events classified with probability greater than 0.75



(b) Events classified with probability less than 0.75

Figure 4: All events classified by the neural network as having a probability (a) greater than or (b) less than 0.75 of being signal-like. Event true classification is shown in red (signal) and blue (background).

	Threshold					
	0	0.25	0.5	0.75	0.8	0.9
S:B Events	1:2	1.5:1	2.7:1	5.6:1	7:1	13:1
Norm. S:B Events	1:1	0.94:0.31	0.84:0.15	0.65:0.06	0.59:0.04	0.41:0.02
Norm. S:B Events ratio	1:1	3:1	5.6:1	10.8:1	14.8:1	20.5:1

Table IV: Signal to Background ratios for various probability cutoffs. S:B Events indicates relative amount of events where initial (threshold cutoff of 0) there are twice as many background events. Norm. S:B Events shows a normalization of events per threshold by class, where all signal events are classified above a threshold of 0, but only 41% of signal evens are classified above 0.9. Norm. S:B Events ratio converts Norm. S:B Events normalized further to the fraction of background events included, for easier interpretation.

[1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, *et al.* (ATLAS), Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, Physics Letters B **716**, 1 (2012).

[2] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, *et al.* (ATLAS), Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc, Physics Letters B **716**, 1–29 (2012).

[3] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, and W. o. Adam, Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc, Physics Letters B **716**, 30–61 (2012).

[4] S. Weinberg, A model of leptons, Phys. Rev. Lett. **19**, 1264 (1967).

[5] T. W. B. Kibble, Symmetry breaking in non-abelian gauge theories, Phys. Rev. **155**, 1554 (1967).

[6] P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, Phys. Rev. Lett. **13**, 508 (1964).

[7] Search for the decay of a heavy higgs boson H into two lighter higgs bosons h and h_s in the $h(\tau\tau)h_s(bb)$ final state at 13 TeV (2021).

[8] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, *et al.* (ATLAS), Observation of $h\to bb^{-}$ decays and vh production with the atlas detector, Physics Letters B **786**, 59 (2018).

[9] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, *et al.* (ATLAS), Observation of higgs boson production in association with a top quark pair at the lhc with the atlas detector, Physics Letters B **784**, 173 (2018).

[10] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, *et al.* (ATLAS Collaboration), Cross-section measurements of the higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector, Phys. Rev. D **99**, 072001 (2019).

[11] K. Olive, Review of particle physics, Chinese Physics C **40**, 100001 (2016).

[12] M. Paganini, Machine learning algorithms for b-jet tagging at the atlas experiment (2017), arXiv:1711.08811 [hep-ex].

[13] G. Aad, B. Abbott, D. C. Abbott, A. A. Abud, K. Abeling, *et al.* (ATLAS), Atlas b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} =$ 13 tev, The European Physical Journal C **79**, 10.1140/epjc/s10052-019-7450-8 (2019).

[14] G. Aad *et al.* (ATLAS), Configuration and performance of the ATLAS b-jet triggers in Run 2 (2021), arXiv:2106.03584 [hep-ex].

[15] U. Husemann, Top-quark physics: Status and prospects, Progress in Particle and Nuclear Physics **95**, 48–97 (2017).