Baku State University: Journal of Physics & Space Sciences, **2025**, v2 (4), p. 32-41

journal homepage: http://bsuj.bsu.edu.az/en

Study of the Higgs boson via the $H \to b\bar{b}$ decay channel using machine learning techniques

Faig N. Ahmadov*

Institute of Physics Ministry of Science and Education, Baku, Azerbaijan;
Received 08-Sep-2025; Accepted 28-Oct-2025
DOİ: https://doi.org/10.30546/209501.101.2025.2.04.0110

Abstract:

This study evaluates the performance of four machine learning algorithms - Boosted Decision Trees, Artificial Neural Networks, Deep Neural Networks, and Transformers - using Monte Carlo simulated data for the $H \to b\bar{b}$ Higgs boson decay channel. The dataset comprises approximately 2 million signal events, along with background events from top quark processes, *V+jets*, and diboson productions. The primary objective is to assess each algorithm's effectiveness in discriminating the Higgs signal from background processes. Performance metrics such as ROC curve AUC values and significance are employed to provide a comprehensive evaluation of signal/background separation and analysis sensitivity. Results demonstrate that all four machine learning approaches excel in separating signal from background, with ANN exhibiting the highest discrimination power. Hyper-parameter optimization for the Transformer model has not yet been performed; however, given its strong performance with high-dimensional low-level variables, further tuning and the incorporation of additional variables are expected to enhance its performance. The findings highlight the potential of advanced machine learning techniques to improve the accuracy and efficiency of signal/background separation in complex datasets typical of Higgs boson analyses. Overall, the study emphasizes the importance of selecting optimal ML algorithms to maximize experimental sensitivity in high-energy physics research.

Keywords: Higgs; ML; BDT; Artificial neural network; DNN
PACS Numbers: 14.80.Bn; 07.05.Mh

* *Corresponding author* – Tel.: (+994) 70 813 08 83
e-mail: fahmadov@cern.ch; ORCID ID: 0000-0003-3644-540X

1. Introduction

The discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations [1, 2] marked a monumental achievement in particle physics, confirming the mechanism of electroweak symmetry breaking and completing the Standard Model (SM) particle spectrum. Despite this milestone, many properties of the Higgs boson remain to be precisely measured, particularly its couplings to fermions and bosons. Among various decay channels, the Higgs boson decaying into a pair of bottom quarks ($H \to b\bar{b}$) is of paramount importance due to its dominant branching ratio of approximately 58% [3], making it a crucial channel for testing the SM and exploring potential new physics. However, the $H \to b\bar{b}$ channel is notoriously challenging to analyze because of the significant background arising from quantum chromodynamics (QCD) processes that produce bottom quark pairs ($b\bar{b}$) with similar signatures, often overwhelming the signal.

To effectively isolate Higgs boson events in this channel, it is advantageous to focus on production modes that provide additional handles for background suppression. One such mode is the associated production of the Higgs boson with a Z boson (ZH), where the Z boson subsequently decays into a pair of charged leptons (electrons or muons). This process offers a cleaner experimental signature, as the leptonic decay of the Z boson provides a distinctive and well-measured final state that can be used to suppress background events significantly. The $ZH \to llb\bar{b}$ (l is e or μ) channel thus provides an optimal environment for studying $H \to b\bar{b}$ decays, combining a manageable background with a clear event topology.

Despite these advantages, the analysis remains complex due to the subtle differences between signal and background events. Traditional cut-based methods often lack the sensitivity to fully exploit all available kinematic information. Recent advances in machine learning (ML) algorithms have demonstrated remarkable success in high-energy physics by significantly improving the discrimination power between signal and background. Techniques such as boosted decision trees and deep neural networks can leverage large datasets and multidimensional feature spaces to enhance signal extraction efficiency.

In this paper, we explore the application of state-of-the-art machine learning algorithms to the study of the $H \to b\bar{b}$ decay channel in the ZH associated production mode, focusing on events where the Z boson decays into charged leptons. By employing ML-based classifiers trained on a comprehensive set of kinematic and topological variables, we aim to improve the sensitivity of Higgs boson measurements in this channel. Specifically, we use four different ML algorithms for this study. Each offers unique advantages in capturing complex correlations within the data, thereby enhancing our ability to distinguish signal events from background processes. This approach not only improves our capacity to observe and characterize the $H \to b\bar{b}$ decay but also provides a versatile framework for future analyses se-

eking to probe the Higgs sector with unprecedented precision.

The article is organized as follows: In the II section, we describe the event selection criteria used to define the dataset and isolate potential signal events. The III section details the machine learning algorithms employed in the analysis, including their theoretical foundations and implementation specifics. In the IV section, we discuss the optimization procedures applied to enhance the performance of these models. Finally, the concluding section summarizes the main findings, discusses their implications, and outlines potential directions for future research.

2. Event selection

For this study, we utilize Monte Carlo (MC) simulated events generated with advanced generators such as Powheg-Box v2 [4] and Sherpa v2.11 [5]. These MC events provide a detailed and accurate modeling of both signal and background processes, serving as a crucial foundation for developing and testing our analysis techniques. The event selection criteria are carefully designed to isolate a clean sample of candidate events consistent with the associated production of the Higgs boson in the $ZH \rightarrow llb\bar{b}$ decay channel, while effectively suppressing the main background processes.

The dominant backgrounds in this analysis are vector boson + jets (*V+jets*), top quark pair production ($t\bar{t}$), and vector boson pair production (*VV*), where *V* is *W* or *Z*. These processes can mimic the signal signature through similar final states, such as dilepton pairs from Z boson decays and jets from quarks or gluons, making their suppression critical for the analysis.

The first selection criterion requires events to contain exactly two leptons - either electrons or muons - of the same flavor and opposite charge. This signature is characteristic of the Z boson decay into leptons and is essential for reconstructing the Z candidate accurately. By focusing on same-flavor pairs, the analysis reduces contamination from background processes with different lepton combinations or multiple leptons.

Next, the selected leptons must satisfy specific kinematic. Each lepton is required to have a transverse momentum (p_T) greater than 25 GeV, and to be located within the pseudorapidity range $|\eta| < 2.5$ (here η is defined as $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle). The η cut corresponds to the geometrical acceptance of the LHC detectors, such as ATLAS [6] and CMS [7].

The invariant mass of the lepton pair, denoted as m_{ll}, must fall within the mass window of 81 to 101 GeV. This range is centered around the Z boson mass (~91 GeV) and ensures that the selected lepton pairs originate predominantly from $Z \rightarrow ll$ decays, thereby reducing background from non-resonant dilepton production and other processes.

In addition to the leptonic criteria, the analysis requires the identification of two

b-jets originating from the Higgs boson decay. To optimize the signal purity, the two jets must satisfy specific transverse momentum cuts: the leading b-jet must have p_T>45 GeV, and the second b-jet must have p_T > 20 GeV. Both jets are required to be within the |η| < 2.5 acceptance, matching the pseudorapidity range of the leptons and reflecting the geometrical limitations of the LHC detectors.

Employing these cuts ensures that the selected events closely resemble the expected signal topology and kinematic properties. Furthermore, the application of geometric acceptance criteria such |η| < 2.5 aligns the simulated event selection with the capabilities of actual LHC detectors, facilitating future application of the results to real experimental data.

Overall, these selection cuts serve to minimize background contamination - particularly from *V+jets*, $t\bar{t}$, and *VV* processes - while maintaining a high efficiency for signal events. The resulting dataset, provides a robust foundation for subsequent machine learning analysis aimed at enhancing the sensitivity to the Higgs boson signal.

3. Machine Learning Algorithms Used

This analysis utilizes four advanced machine learning models: Boosted Decision Trees (BDT) [8], Artificial Neural Networks (ANN) [9], Deep Neural Networks (DNN) [10], and Transformer architectures [11]. Each model is chosen for its unique approach to pattern recognition and ability to model complex relationships within high-energy physics data.

3.1. Boosted Decision Trees

The BDT is an ensemble method that combines multiple decision trees to improve classification accuracy. Each tree is built sequentially, with subsequent trees focusing on the instances misclassified by previous ones. This boosting process effectively reduces bias and variance, resulting in a highly robust classifier. The decision trees themselves partition the feature space through binary splits based on thresholds of input variables, creating a series of decision rules that lead to a final classification. The BDT's structure allows it to naturally handle heterogeneous data types and is relatively straightforward to interpret, as feature importance can be directly derived from the model.

3.2. Artificial Neural Networks

The ANN employed in this study consists of a single- or two-layer feedforward network comprising multiple neurons connected to the input features. Each neuron applies a weighted sum of inputs followed by a non-linear activation function, enabling the network to learn complex non-linear mappings. The network's architecture is designed to be flexible, with the number of neurons and layers tuned to balance model

capacity and computational efficiency. These models are capable of capturing intricate relationships between variables through their non-linear transformations, making them suitable for complex classification tasks inherent to collider data.

3.3. Deep Neural Networks

The DNN extends the ANN architecture by adding multiple hidden layers, allowing the network to learn hierarchical feature representations. Each layer transforms the data into increasingly abstract features, which can enhance the classifier's ability to distinguish subtle differences between signal and background. The depth of the network enables it to model complex, non-linear interactions among input variables that may not be easily captured by shallower models. Regularization techniques such as dropout and weight decay are employed to prevent overfitting, facilitating stable training and robust performance across varied datasets.

3.4. Transformer Models

Transformers are a relatively recent addition to the machine learning toolkit, originally developed for natural language processing. Their core component is the self-attention mechanism, which allows the model to weigh the importance of different parts of the input sequence dynamically. In this analysis, the Transformer architecture processes event data structured as sequences of features, enabling the model to learn long-range dependencies and complex correlations across variables. The positional encoding component helps the model understand the order and structure within the input data, while multi-head attention allows it to focus on different aspects simultaneously. This architecture provides a highly flexible framework capable of capturing subtle and distributed patterns within the dataset.

Each of these models offers distinct advantages: the BDT provides interpretability and robustness; the ANN and DNN offer flexibility and depth for modeling non-linear relationships; and the Transformer brings the power of sequence modeling and attention mechanisms to the analysis. Their diverse approaches enrich the overall analysis strategy, enabling a comprehensive exploration of the data's underlying structures.

4. Optimization of Machine Learning Algorithms

4.1. Input variables

All four machine learning models - BDT, ANN, DNN, and Transformer - use the same set of nine input variables. These variables are carefully selected to capture the most relevant information for the classification task and are listed in Table 1. The variables include kinematic properties, event shape variables, and other discrimi-

nating features derived from the physics analysis. Their uniform usage across models ensures a fair comparison of the models' performance and allows for consistent optimization procedures.

Table 1. Variables used as input to machine learning models.

#	Variable Name	Description
1	m_{bb}	Invariant mass of the two b-jets
2	$\Delta R(b_1, b_2)$	Angular distance between the two b-jets
3	p_T^{b1}	Transverse momentum of the first b-jet
4	p_T^{b2}	Transverse momentum of the second b-jet
5	p_T^Z	Transverse momentum of the Z boson
6	$\lvert\Delta\phi(V, bb)\rvert$	The azimuthal distance between the Z boson and the Higgs boson
7	$\lvert\Delta y(V, bb)\rvert$	The rapidity distance between the Z boson and the Higgs boson
8	m_{ll}	Invariant mass of the two charged leptons
9	$cos\theta(l^-, Z)$	The Z boson polarization sensitive angle

4.2. Optimization of ML Algorithms

To achieve the best possible performance, a comprehensive hyper-parameter tuning process was undertaken. For each model, extensive testing across a wide range of hyper-parameter values was performed. The hyper-parameter combinations listed below represent the most optimal settings identified through this process, yielding the highest classification accuracy and robustness on validation datasets.

4.2.1. Optimal hyper-parameters for BDT:

- ***n_estimators = 200*** – Number of trees in the ensemble, balancing performance and computational cost.
- ***learning_rate = 0.5*** – Controls the contribution of each tree; higher values lead to faster learning but may risk overfitting.
- ***max_depth = 3*** – Maximum depth of each tree; limits complexity to prevent overfitting.
- ***min_samples_split = 2*** – Minimum number of samples required to split an internal node.
- ***min_samples_leaf = 5*** – Minimum number of samples in newly created leaves.
- ***subsample = 0.8*** – Proportion of samples used to fit each base learner, introducing randomness for robustness.
- ***random_state = 42*** – Seed for reproducibility of results.

4.2.2. Optimal hyper-parameters for ANN:

- ***hidden_units = [256, 128]*** – Number of neurons in each hidden layer.
- ***learning_rate = 0.0001*** – Step size for updating weights during training.
- ***batch_size = 256*** – Number of samples processed before updating model weights.
- ***epochs = 50*** - Number of complete passes through the training dataset.
- ***activation = ReLU*** – Activation function for hidden layers, introducing non-linearity.
- ***optimizer = Adam*** – Optimization algorithm used for training.
- ***patience = 35*** – Number of epochs without improvement before early stopping is triggered.

4.2.3. Optimal hyper-parameters for DNN:

The DNN shares most hyper-parameters with the ANN but differs in its architecture: ***hidden_units = [256, 128, 64]*** – Adding a third layer to increase model depth and capacity for hierarchical feature learning.

The remaining hyper-parameters are identical to those used in the ANN.

4.2.3. Hyper-parameters for Transformer (not optimized, chosen according to ANN or DNN):

- ***embed_dim = 256*** – Dimensionality of the embedding space, allowing rich feature representations.
- ***num_heads = 4*** – Number of attention heads, enabling the model to focus on different parts of the input simultaneously.
- ***ff_dim = 256*** – Dimension of the feed-forward network within transformer blocks.
- ***num_transformer_blocks = 2*** – Number of stacked transformer layers, capturing complex dependencies.
- ***mlp_units = [256, 128]*** – Size of the multilayer perception in the final stages of the transformer.
- ***dropout_rate = 0.05*** – Dropout rate to prevent overfitting during training.
- ***learning_rate = 0.001*** – Learning rate for the optimizer.
- ***batch_size = 256*** – Number of samples per training batch, balancing memory use and training stability.
- ***epochs = 50***

This thorough hyper-parameter tuning process, involving testing a broad spectrum of values, ensures that each model operates at its most effective configuration, facilitating a fair and optimized comparison of their classification capabilities.

Using the hyper-parameters outlined for each machine learning model, the mo-

dels were trained on the training dataset and evaluated on the validation dataset (size - 10% of the training dataset). The performance was primarily assessed through the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), which measures the model's ability to distinguish between signal and background events across various threshold settings. Additionally, the signal significance - an important metric in physics analyses - was calculated to determine the statistical strength of the signal detection for each model. The significance here is defined as the Asimov significance with the formula:

$$Z = \sqrt{2[(S+B)\ln(1+S/B) - S]}. \quad (2)$$

The results indicate that the BDT achieved an AUC of 0.926, while the ANN and DNN both reached slightly higher AUC values of 0.930, demonstrating excellent discrimination power. The Transformer model, however, yielded a lower AUC of 0.835, indicating comparatively reduced classification performance. In terms of signal significance, the models produced the following maximum values: 2.66 for BDT, 2.77 for ANN, 2.75 for DNN, and 1.00 for the Transformer. These significance metrics reflect the potential of each model to enhance the detectability of the signal within the data.

Figures 1 and 2 visually summarize these results. Figure 1 displays the ROC curves for all four models, illustrating their true positive versus false positive rates across different thresholds. Figure 2 presents the maximum significance values obtained from each model, highlighting their relative effectiveness in identifying the signal. Overall, these performance metrics demonstrate that while the BDT, ANN, and DNN exhibit comparable and high classification capabilities, the Transformer's performance is comparatively lower in this context.

5. Conclusions

This study presents an approach to optimizing machine learning algorithms for the classification of signal and background events in the $ZH \to llb\bar{b}$ process. Beginning with a detailed selection of relevant events, we established a consistent set of nine input features to ensure fair and effective training across all models.

In the subsequent section, we explored four advanced machine learning algorithms - BDT, ANN, DNN, and Transformer models - each with carefully designed architectures. Recognizing the importance of hyper-parameter tuning, we conducted an extensive search across broad parameter ranges to identify the most optimal configurations. These hyper-parameters were selected based on their ability to maximize model performance, as validated through rigorous testing.

Using these optimized hyper-parameters, the models were evaluated via ROC curves and signal significance metrics. The results demonstrated that the BDT, ANN, and DNN models achieved high discrimination capabilities, with AUC values of

aproximately 0.926 to 0.930 and maximum significances around 2.66 to 2.77. In contrast, the Transformer model, in its current unoptimized state, showed comparatively lower performance, with an AUC of 0.835 and a maximum significance of 1.0. It is important to note that the Transformer architecture has not yet undergone comprehensive hyper-parameter optimization; therefore, there is significant potential for performance improvements. With further tuning and architectural adjustments, the Transformer could potentially surpass the other models in both discrimination power and signal significance.

Figures 1 and 2 visually depict these performance metrics, emphasizing the strengths and current limitations of each model. Overall, the results highlight the effectiveness of traditional ensemble methods and deep neural networks in this context, while also pointing to the promising prospects of Transformer-based models pending further optimization.

References

[1] ATLAS collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B 716 (2012) 1, https://doi.org/10.1016/j.physletb.2012.08.020

[2] CMS collaboration, Observation of a new boson at a mass of 125GeV with the CMS experiment at the LHC, Phys. Lett. B 716 (2012) 30, https://doi.org/10.1016/j.physletb.2012.08.021

[3] LHC Higgs Cross Section Working Group collaboration, Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, arXiv:1610.07922, https://doi.org/10.23731/CYRM-2017-002

[4] S. Alioli, P. Nason, C. Oleari and E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, JHEP 06 (2010) 043, https://doi.org/10.1007/JHEP06(2010)043

[5] Sherpa collaboration, Event generation with Sherpa 2.2, SciPost Phys. 7 (2019) 034, https://doi.org/10.21468/SciPostPhys.7.3.034

[6] ATLAS Collaboration, JINST, 3 (2008), p. S08003, http://dx.doi.org/10.1088/1748-0221/3/08/S08003

[7] CMS experiment at the CERN LHC, JINST, 3 (2008), p. S08004, http://dx.doi.org/10.1088/1748-0221/3/08/S08004

[8] Byron P. Roe, Hai-Jun Yang, and Ji Zhu, Boosted Decision Trees, a Powerful Event Classifier, Statistical Problems in Particle Physics, Astrophysics and Cosmology, pp. 139-142 (2006), https://doi.org/10.1142/9781860948985_0029

[9] Wang, SC. (2003). Artificial Neural Network. In: Interdisciplinary Computing in Java Programming. The Springer International Series in Engineering and Computer Science, vol 743. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-0377-4_5

[10] Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. Neural Networks, Volume 61, January 2015, Pages 85-117,
https://doi.org/10.1016/j.neunet.2014.09.003

[11] A. Vaswani et al., “Attention is all you need”, in Advances in Neural Information Processing Systems, I. Guyon et al., eds., volume 30. Curran Associates, Inc., 2017, https://doi.org/10.48550/arXiv.1706.03762