The use of Machine Learning techniques to analyse the $gg \to h \to Z\gamma$ process within the SMEFT framework at the Large Hadron Collider (LHC)

Kutlwano Makgetha[1,2], Njokweni Mbuyiswa[1,2], Srimoy Bhattacharya[1,2], Abduluazem Fadol[1], Mukesh Kumar[1] and Bruce Mellado[1,2]

[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: 2134657@students.wits.ac.za

Abstract. Building on the ATLAS and CMS evidence of the Higgs boson (h) decaying into a Z-boson and a photon (with a 3.4σ significance), the current Standard Model (SM) predictions for the $h \to Z\gamma$ signal rate exceed the measured value by $(2.4 \pm 0.9)\sigma$, indicating possible new physics effects or systematic uncertainties that warrant further investigation. This analysis investigates this rare process using machine learning (ML) techniques, where we use classifiers such as the Extreme Gradient Boost (XGBoost) and the kernel density estimation to analyse the production modes of $h \to Z\gamma$, including gluon-gluon fusion (ggF), within the framework of the Standard Model Effective Field Theory. This machine learning approach aims to constrain the six-dimensional Wilson coefficients and shed light on potential deviations from the SM prediction.

1 Introduction

Recent measurements by ATLAS and CMS have reported evidence on the decay of the $h \to Z\gamma$ process, which cannot be fully explained within the Standard Model (SM). We look for potential answers beyond the SM (BSM), and that is where Standard Model Effective Field Theory (SMEFT) comes into play. The SMEFT [1] provides a systematic approach for BSM physics by introducing higher-dimensional operators. These operators are suppressed by powers of a high-energy scale Λ, allowing for the parameterisation of new physics effects. We build upon the findings presented in [2], where the initial step is to identify the specific operators within the SMEFT framework that play a role in our decay process. These operators are essentially mathematical expressions that represent potential deviations from the SM predictions [3, 4], and once we have identified the relevant operators, we will calculate their corresponding Wilson coefficients. These coefficients are numerical values that determine the strength of the operator's contribution to the decay process. The SMEFT@NLO package is a computer program designed to automate this calculation as it uses models created with `FeynRules` [5], which is another software tool that helps define the interactions between particles in the SM and any potential extensions.

2 Data Analysis

2.1 Object Selections

To optimise the signal events over background noise, we applied minimal cuts, including object and channel-specific selections, based on transverse momentum (p_T), invariant masses, pseudorapidity ($|\eta$), missing transverse energy (E_T^{miss}), and jet reconstruction.

Table 1: Summary of object selections criteria.

Selection	**Criteria**
Photon requirement	(γ) At least one photon candidate
Lepton requirement (ℓ)	At least two same-flavour, opposite-charge leptons ($\ell = e, \mu$)
Muon(μ) selections	$\lvert\eta^{\mu^\pm}\rvert < 2.7$, $p_T^{\mu^\pm} > 25$ GeV
Electron (e) selections	$\lvert\eta^{e^\pm}\rvert < 2.47$ excluding $1.37 < \lvert\eta^{e^\pm}\rvert < 1.52$, $p_T^{e^\pm} > 25$ GeV
Photon selections	$\lvert\eta^{\gamma}\rvert < 2.47$ excluding $1.37 < \lvert\eta^{\gamma}\rvert < 1.52$, $p_T^{\gamma} > 10$ GeV
Jet reconstruction	anti-k_t algorithm, $R = 0.4$
Jet-photon separation	Jets within $\Delta R < 0.4$ of any photon are rejected
Dilepton ($\ell\ell$) mass window	$81 < m^{\ell^+\ell^-} < 101$ GeV
$Z\gamma$ system	$p_T(Z\gamma) > 50$ GeV
Photon-lepton separation	$\Delta R(\gamma, \ell) > 0.4$
Missing transverse energy	$E_T^{\text{miss}} < 40$ GeV
Channel-specific cuts	
Leptonic ($\ell^+\ell^-\gamma$) channel	$100 < m^{Z\ell^+\ell^-\gamma} < 140$ GeV, $m^{\ell\ell} + m^{Z\ell\ell\gamma} > 182$ GeV
Jet ($jj\gamma$) channel	$50 < m^{Zjj\gamma} < 200$ GeV
Neutrino ($\nu\bar{\nu}\gamma$) channel	$30 < E_T^{\text{miss}} < 100$ GeV

The next subsection discusses the ML techniques that contribute to optimising the significance estimation derived from the applied cuts.

2.2 Signal and background discrimination using Machine Learning

We use the XGBoost classifier, which is the implementation of the Boosted Decision Trees (BDT) model that uses second-order optimisation. Unlike traditional BDTs that only use first-order gradients, XGBoost takes advantage of both gradients and Hessians for better convergence. For our work, it is trained for 3500 boosting rounds (`n_estimators`) to improve its learning capability. The learning rate is set at 0.05, striking a balance between the speed of convergence and the stability of the model whereby a higher learning rate could lead to fluctuations, while a lower rate might slow down the training process. This method includes Lasso and Ridge regularisation set to 0.0001 to prevent overfitting. We have also added a `gamma` parameter to control the minimum reduction in loss required for making additional splits in the trees. Our objective function, `binary:logistic`, is aimed at optimising binary classification using the logistic loss function (`logloss`). The trees have a maximum depth of 5 (`max_depth`), which helps balance model complexity and performance. We enable early stopping after 50 rounds (`early_stopping_rounds`), ensuring that training stops when no improvement is observed. This strategy promotes stable learning while reducing the risk of overfitting. The next step was to select all the generated signal and background MC samples and ensured that we used the same number of signal and background events such that the signal-to-background ratio was 1. The ggF production is categorised into three distinct categories: $jj\gamma$, $\ell^+\ell^-\gamma$, and $\nu\bar{\nu}\gamma$. For the ggF production categories, we exclusively used the XGBoost classifier. We later trained our classifier on the same MC-simulated samples for performance comparison. Then the performance of both classifiers was assessed using the Kernel Density Estimation (KDE) [6, 7] to estimate the probability density functions of predicted values.

3 Preliminary Results

After generating the MC samples, we used a random forest classifier to evaluate feature importance and used the features with the highest importance to optimise our model. Below, we present the selected features for each channel, their respective channels, their corresponding optimised Receiver Operating Curve (ROC) and associated Area Under the Curve (AUC) scores. We describe the methods used to calculate Z-scores, which are used to quantify the statistical significance of the observed signals using the Machine-Learned Likelihood (MLL) approach [8, 9] and KDE. We also compare different binning values for the classifier's outputs. Z-scores, meanwhile, are used

to quantify the statistical significance of the observed signals. We used a python script to evaluate the classifier's sensitivity by calculating the binned significance for $N_{\text{bins}} = 10, 25, 50, 100$. The background and signal histograms produced were normalised to account for the expected event counts, and the binned significance Z, which is calculated as:

$$Z(i) = \sqrt{2 \sum_{i} \left[(S_i + B_i) \cdot \ln \left(1 + \frac{S_i}{B_i} \right) - S_i \right]}$$

where S_i and B_i represent the normalised bin values for signal and background events in the i-th bin, respectively. Finally, the KDE method offers a continuous analysis, avoiding the limitations of discrete binning.

Table 2: The ggF production Z-scores for jet, neutrino, and lepton channels.

Channel	**Jet**	**Neutrino**	**Lepton**
Z (10 bins)	5.22	6.27	24.08
Z (25 bins)	5.31	6.48	24.00
Z (50 bins)	5.33	6.59	23.92
Z (100 bins)	5.44	6.69	23.82
Z (KDE)	10.66	11.34	27.78

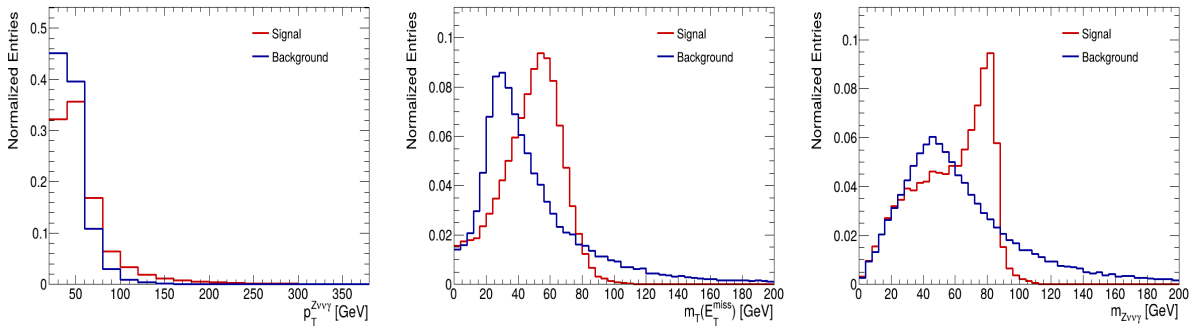

Figure 1: The three most important kinematics, i.e, the transverse momentum (left), missing transverse energy (center), and the invariant mass of the entire system (right) normalised to unity of the ggF production ($\nu\bar{\nu}\gamma$ category).

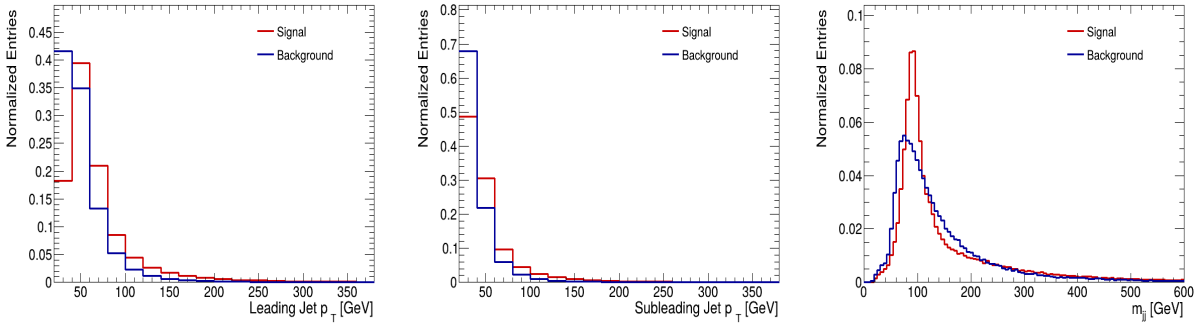

Figure 2: The three most important kinematics, i.e, the leading jet transverse momentum (left), subleading jet transverse momentum (center), and dijet mass system (right), normalised to unity of the ggF production ($jj\gamma$ category).

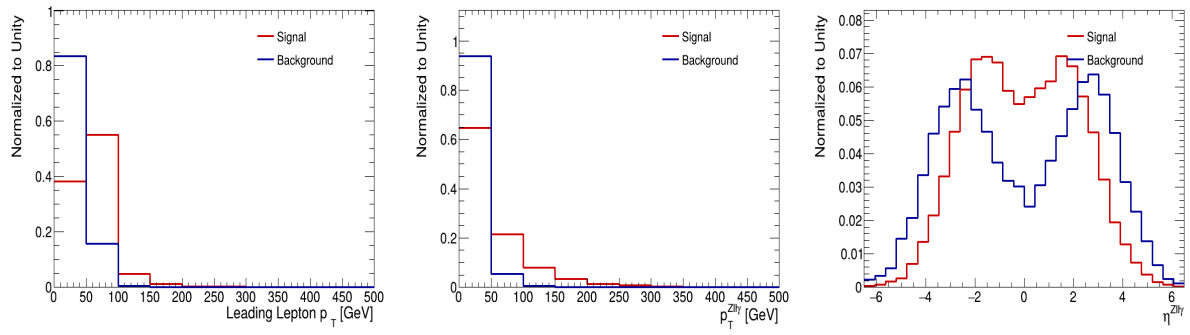

Figure 3: The three most important kinematics, i.e, the leading lepton transverse momentum (left), transverse momentum of the system (center), and pseudorapidity of the system (right) normalised to unity of the ggF production ($\ell^+\ell^-\gamma$ category).

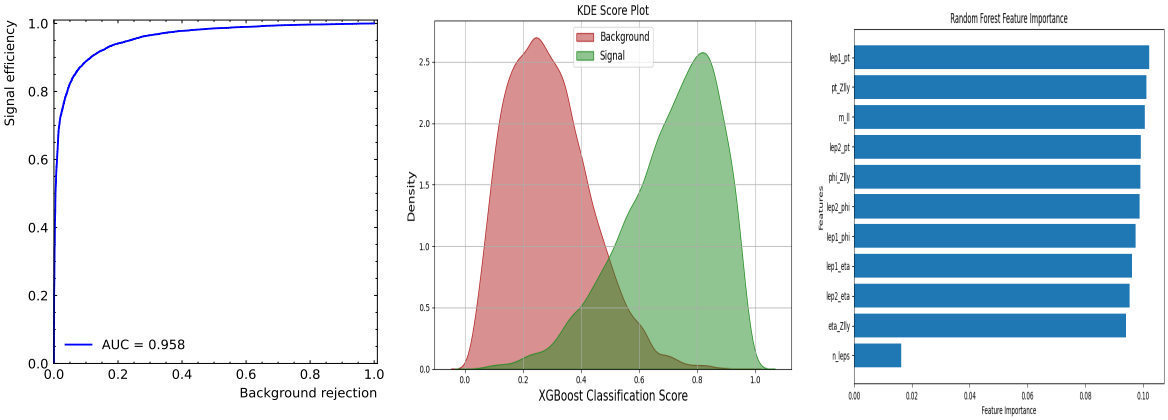

Figure 4: Comparison of the ROC Curve with its associated AUC score (left), XGBoost performance with KDE (center), and feature importance (right) for the ggF production ($\ell^+\ell^-\gamma$ category) without the invariant mass.

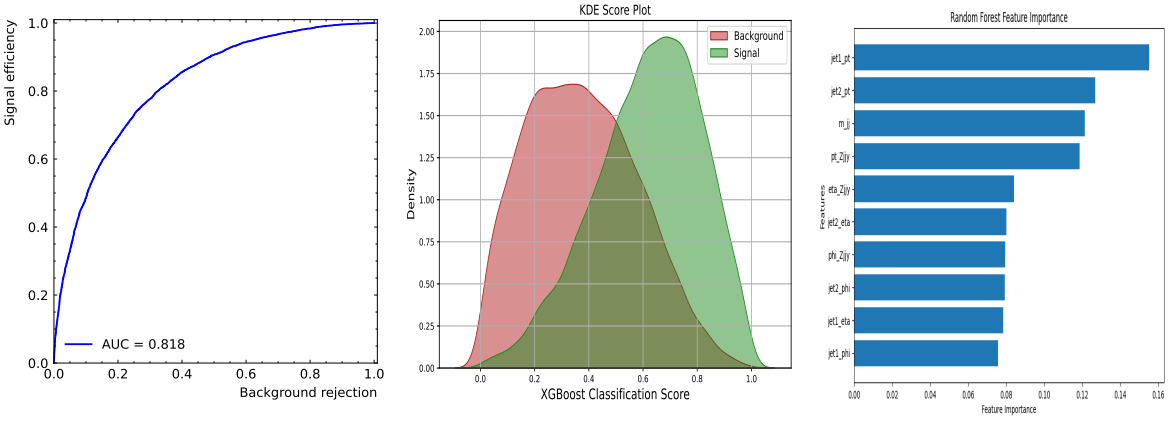

Figure 5: Comparison of the ROC Curve with its associated AUC score (left), XGBoost performance with KDE (center), and feature importance (right) for the ggF production ($jj\gamma$ category) with the invariant mass.

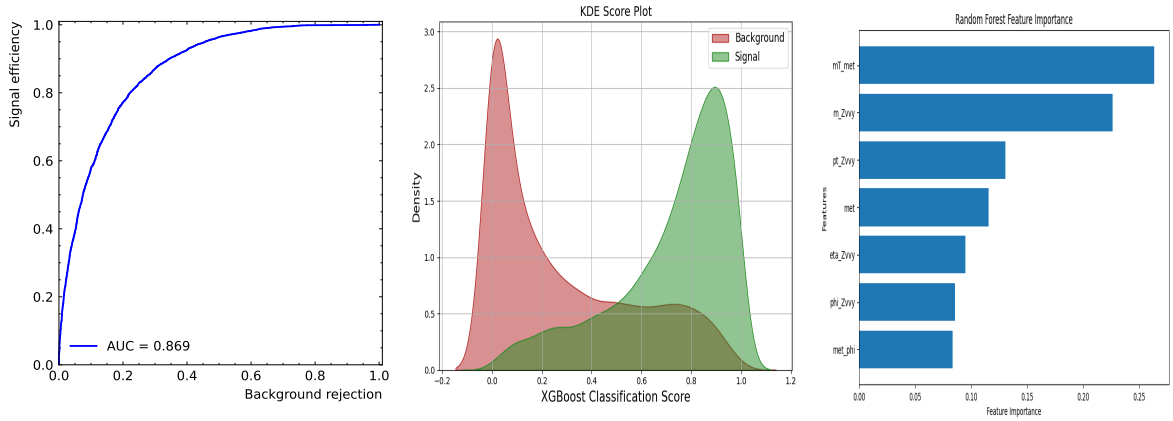

Figure 6: Comparison of the ROC Curve with its associated AUC score (left), XGBoost performance with KDE (center), and feature importance (right) for the ggF production ($\nu\bar{\nu}\gamma$ category) with the invariant mass.

4 Conclusion and Future Plans

We evaluated the XGBoost classifier's performance in the ggF channel using ROC scores, demonstrating its signal-background separation capability. Next month, we will extend this analysis to the VBF, $t\bar{t}H$, and VH production channels, assessing performance both before and after optimization using each channel's key features. We also plan to compare the performance of a BDT classifier against XGBoost. Our results indicate that Kernel Density Estimation (KDE) improves with larger ensemble sizes because it avoids output binning, while the binning method benefits from increased bin counts; KDE consistently outperforms binning.

5 Acknowledgments

I would like to express my gratitude to the SA-CERN Excellence bursary for funding this research.

References

[1] I. Brivio and M. Trott, "The standard model as an effective field theory," *Physics Reports*, vol. 793, pp. 1–98, 2019.

[2] S. Dawson and P. P. Giardino, "Higgs decays to zz and z γ in the standard model effective field theory: An nlo analysis," *Physical Review D*, vol. 97, no. 9, p. 093003, 2018.

[3] J. Elias-Miro, J. Espinosa, E. Masso, and A. Pomarol, "Higgs windows to new physics through d= 6 operators: constraints and one-loop anomalous dimensions," *Journal of High Energy Physics*, vol. 2013, no. 11, pp. 1–42, 2013.

[4] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization group evolution of the standard model dimension six operators iii: gauge coupling dependence and phenomenology," *Journal of High Energy Physics*, vol. 2014, no. 4, pp. 1–47, 2014.

[5] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks, "Feynrules 2.0—a complete toolbox for tree-level phenomenology," *Computer Physics Communications*, vol. 185, no. 8, pp. 2250–2300, 2014.

[6] M. Rosenblat, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Stat*, vol. 27, pp. 832–837, 1956.

[7] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[8] E. Arganda, X. Marcano, V. M. Lozano, A. D. Medina, A. D. Perez, M. Szewc, and A. Szynkman, "A method for approximating optimal statistical significances with machine-learned likelihoods," *The European Physical Journal C*, vol. 82, no. 11, p. 993, 2022.

[9] E. Arganda, M. de los Rios, A. D. Pérez, and R. S. Seoane, "Imposing exclusion limits on new physics with machine-learned likelihoods," *PoS ICHEP2022*, vol. 1226, 2022.