The use of Machine Learning techniques to analyze the gg→ h → Zγ process within the SMEFT framework at the Large Hadron Collider (LHC)

Presenter :Kutlwano Makgetha

In collaboration with: Srimoy Bhattacharya, Abdualazem Fadol, Mukesh Kumar, Njokweni Mbuyiswa, Bruce Mellado

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

69th Annual Conference of the South African Institute of Physics

7-11 July 2025

SOUTH AFRICAN INSTITUTE OF PHYSICS

ATLAS EXPERIMENT

INSTITUTE FOR COLLIDER PARTICLE PHYSICS

UNIVERSITY OF THE WITWATERSRAND

NRF National Research Foundation

iThemba LABS Laboratory for Accelerator Based Sciences

Outline

- Introduction
- Model and Framework
- MC studies with implementation of ML
- Conclusion

INTRODUCTION

- h $\rightarrow Z\gamma$ is a rare process in the SM and it can only happen at NLO.

- Recent study [arXiv:1801.01136] shows that even with NLO QCD corrections and the signal-background interference, this excess in h $\rightarrow Z\gamma$ cannot be explained within the SM which currently overshoots the measured signal rate by 2.2σ.

- With this motivation we perform our analysis of h $\rightarrow Z\gamma$ decay channel with higgs production through gluon gluon-fusion and constraint the corresponding six-dimension Wilson coefficients in SMEFT through the cross section measurement and different sensitive kinematic observables.

Model and Framework

- In this study the Higgs boson decays in the context of the Standard Model Effective Field Theory (SMEFT), which is a framework for describing new physics beyond the Standard Model of particle physics.

- The SMEFT is based on the assumption that new physics effects can be parameterized by higher-dimensional operators, which are suppressed by powers of a high energy scale Λ.

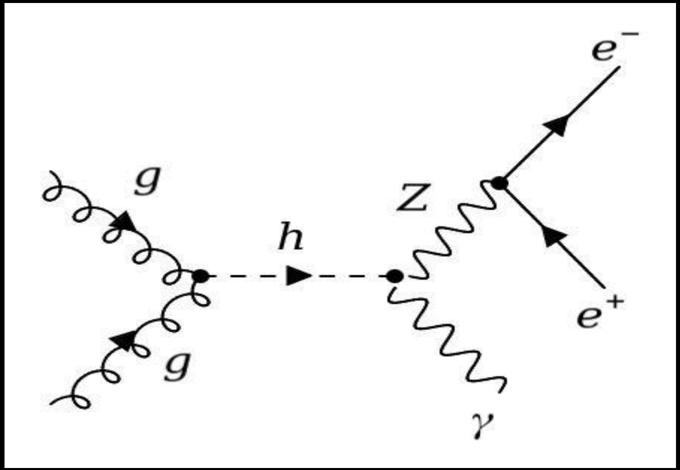
- The paper [arXiv:1801.01136] focuses on two specific decay modes of the Higgs boson: the decay to a pair of Z bosons (ZZ) and the decay to a Z boson and a photon ($Z\gamma$).

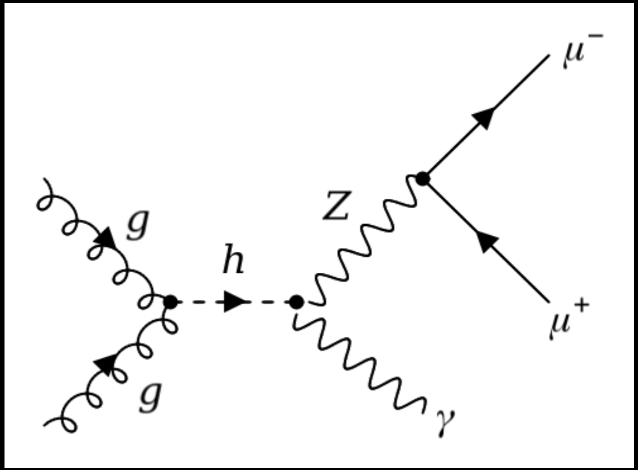
- For more details **N Mbuyiswa** has given a talk on this.

Object Selections

- $|\eta_\mu| < 2.7$
- $p_T^\mu > 10\ GeV$
- $|\eta_e| < 2.47$ excluding $(1.37 < |\eta_e| < 1.52)$
- $p_T^e > 10\ GeV$
- $|\eta_\gamma| < 2.47$ excluding $(1.37 < |\eta_\gamma| < 1.52)$
- $p_T < 10\ GeV$
- Anti-k_t with $\Delta R = 0.4, |\eta_j| < 4.4$
- $p_T^j > 25\ GeV$

ggF leptonic channel feynman diagrams

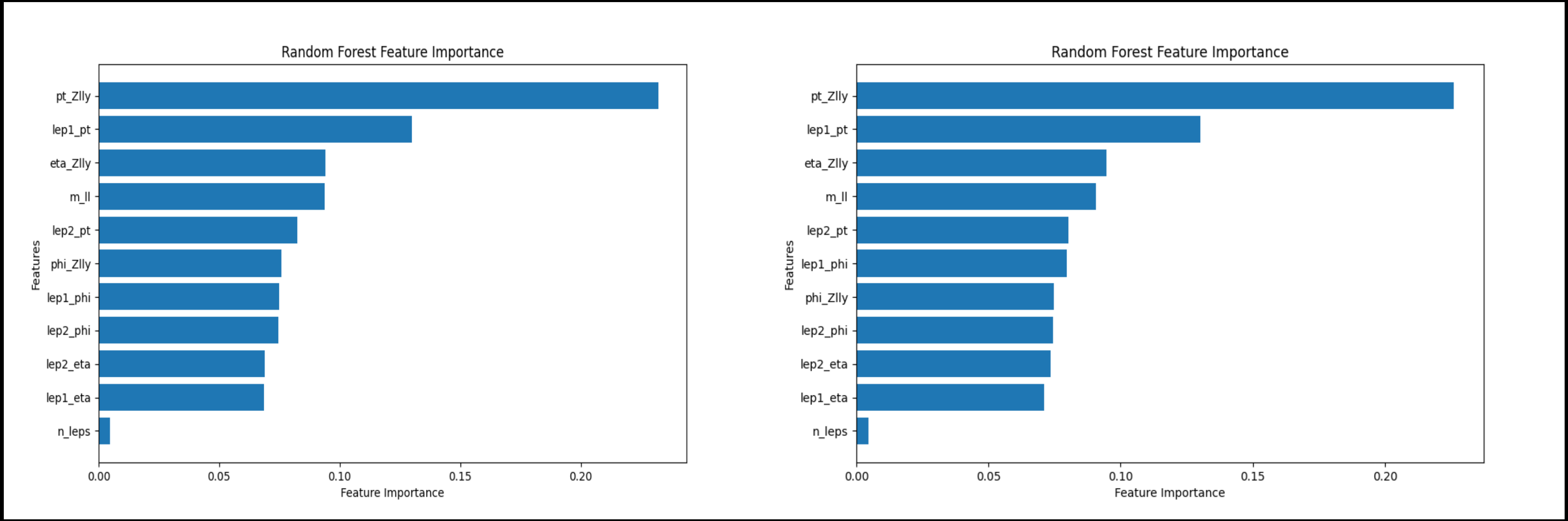




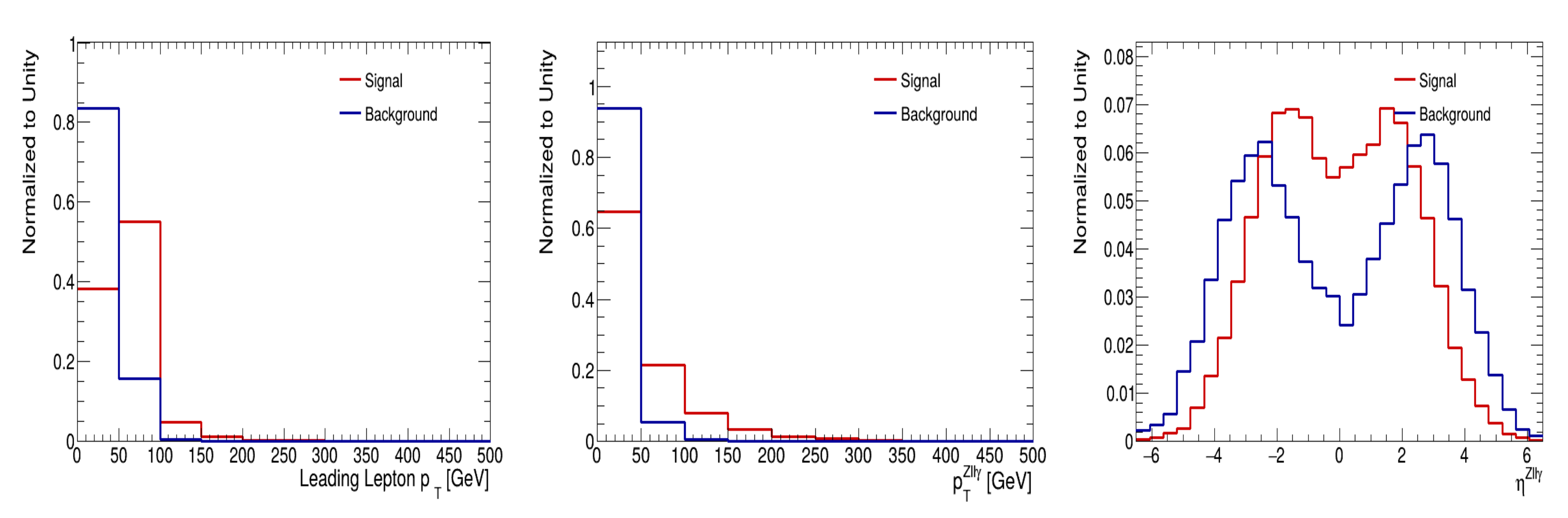
- Cross-section 17.63 pb for background
- Cross-section 0.1525 pb for signal

ggF leptonic channel feature importance

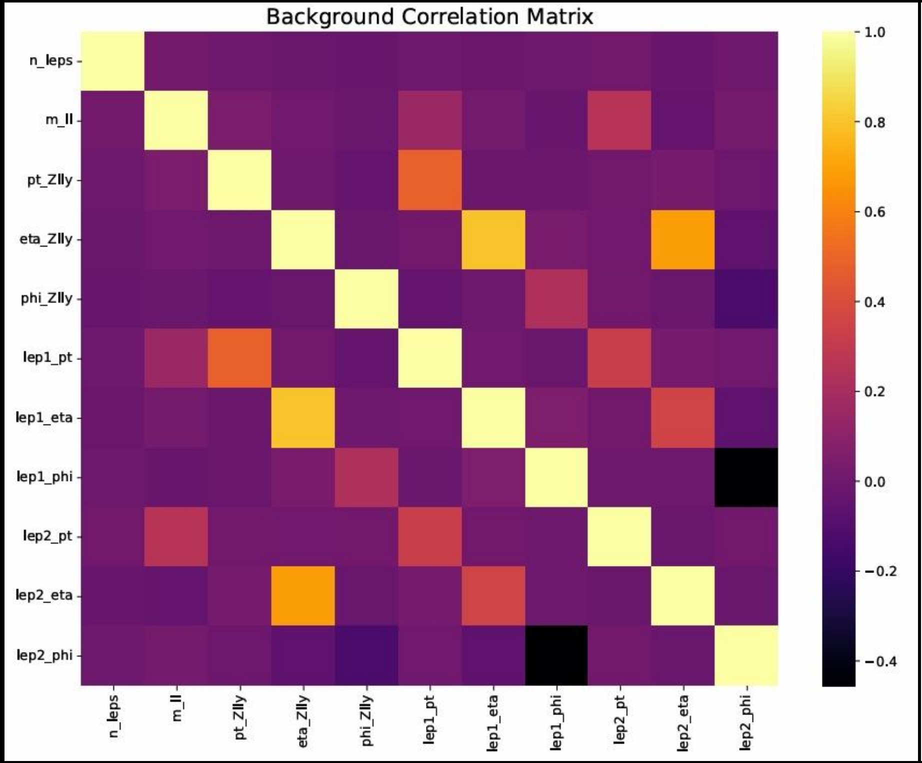
- Left (+1) and right (-1)

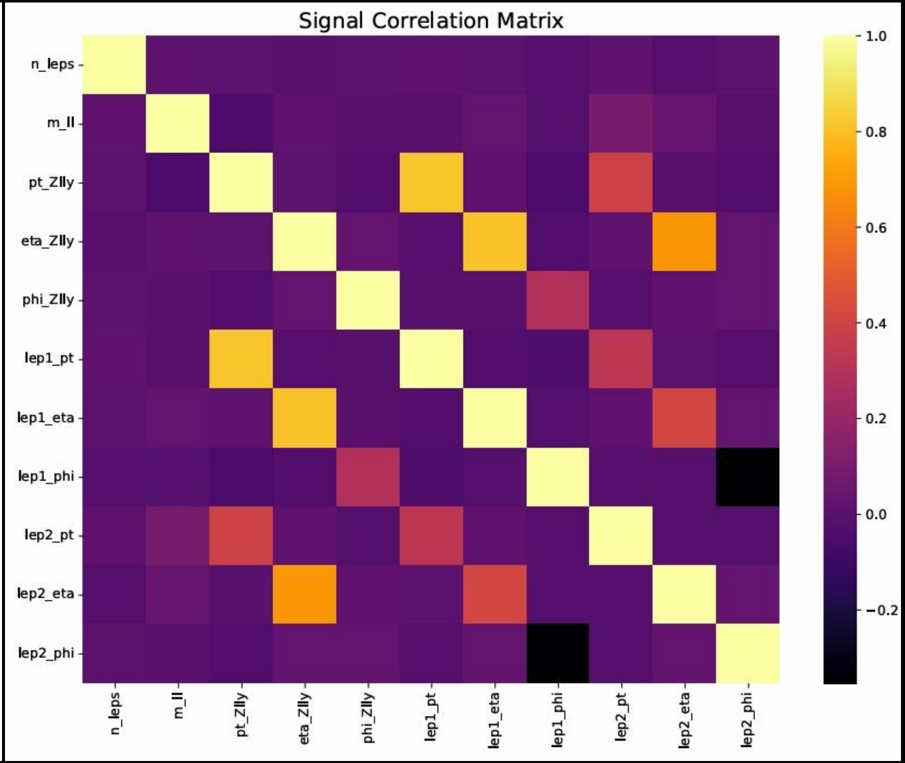


ggF leptonic channel important kinematics



ggF leptonic channel correlation matrix





ML Approach

PARAMETER	VALUE
Classifier	Extreme Gradient Boost
split	70:30:10
N estimators	3500
Learning rate	0.05
reg lambda	1
reg alpha	0.1
gamma	0.1
objective	binary:logistic
eval metric	logloss
early stopping rounds	50
max depth	5

KDE Calculation

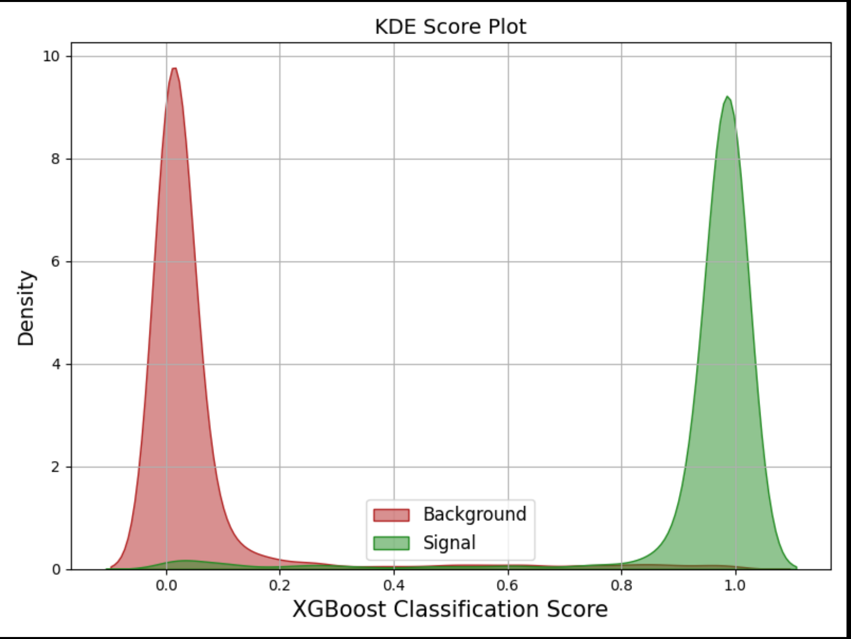
The Z binned significance is calculated as :

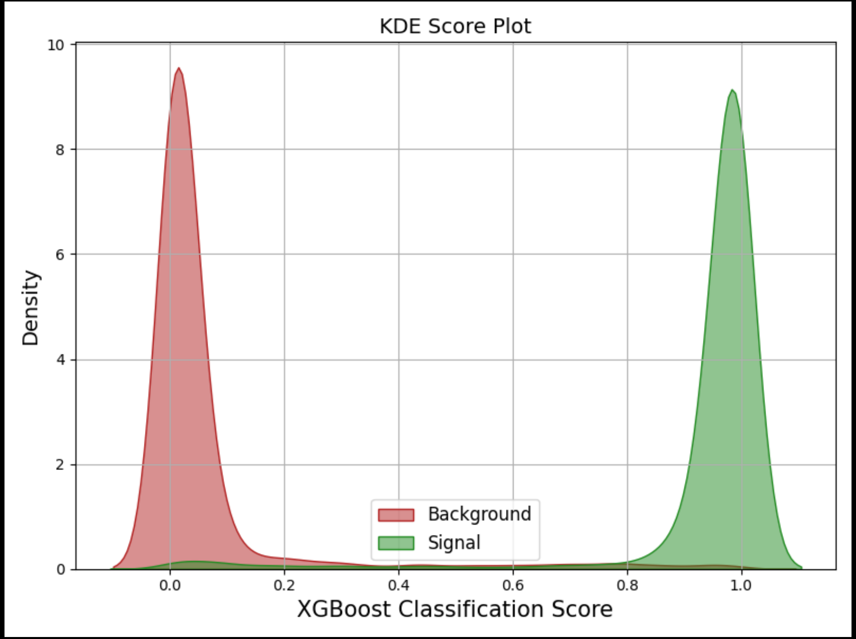
$$Z = \sqrt{2\sum_i \left[(S_i + B_i)\cdot \ln\left(1 + \frac{S_i}{B_i}\right) - S_i\right]}$$

- Bandwidth parameter ($\epsilon = 0.1$)
- Smaller ϵ provides finer resolution but risks overfitting
- Larger ϵ provides smoother approximations but may underfit the events
- For more information see paper [arXiv:2211.04806v2]

ggF leptonic channel KDE plot

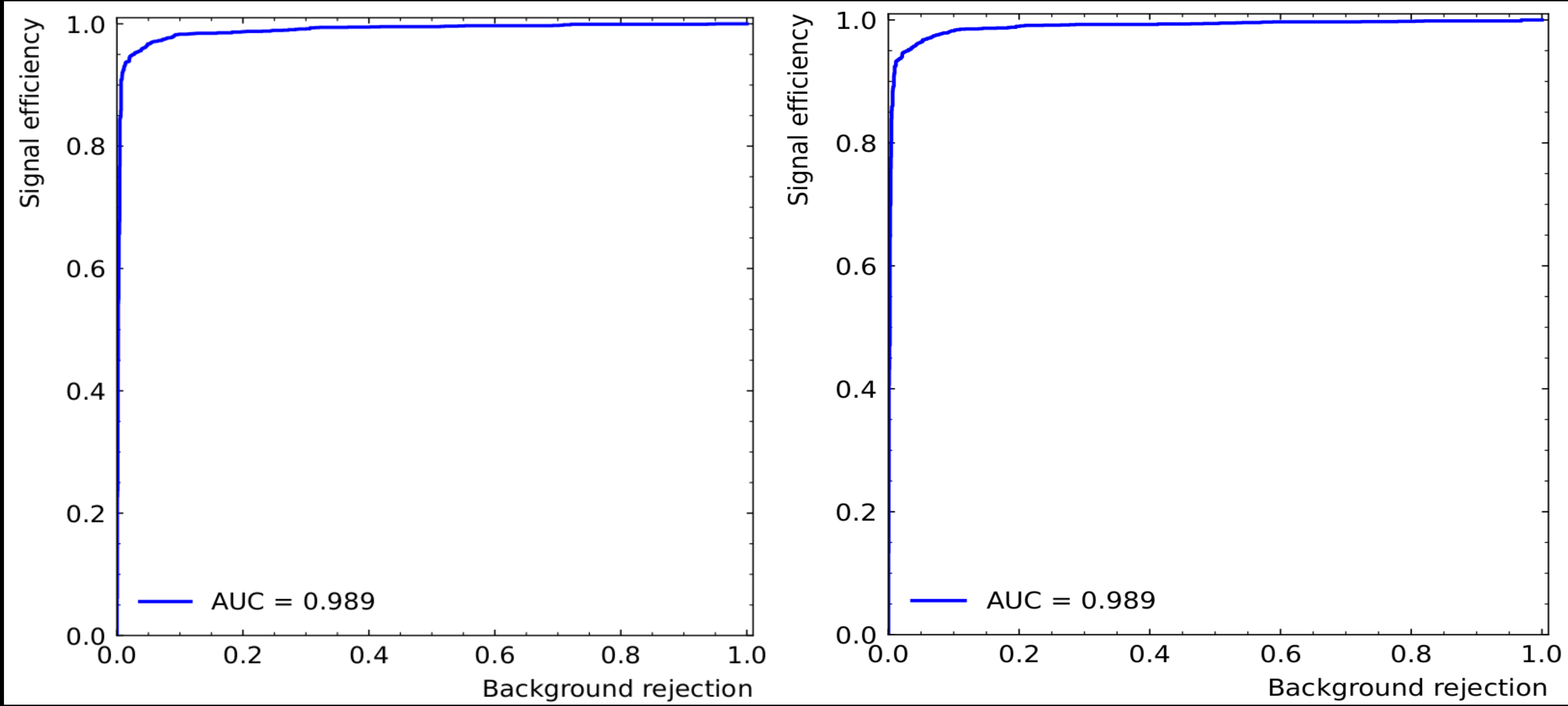
- Left (+1) and right (-1)



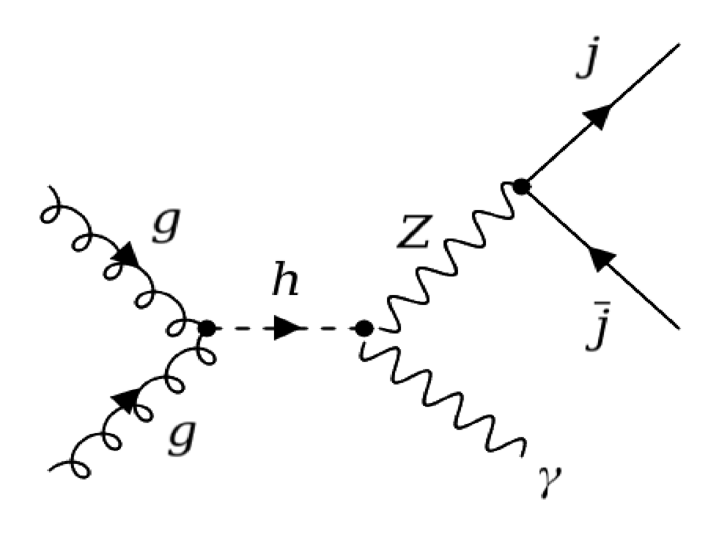


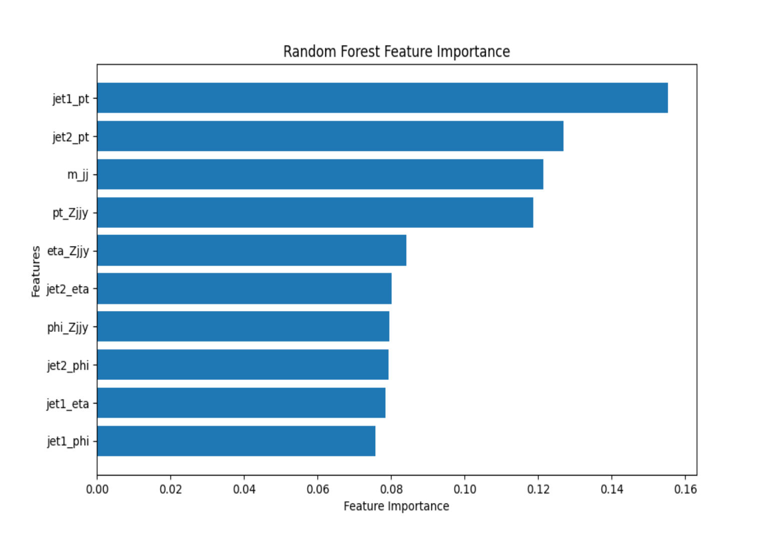
ROC Curves for leptonic channel

- Left (+1) and right (-1)

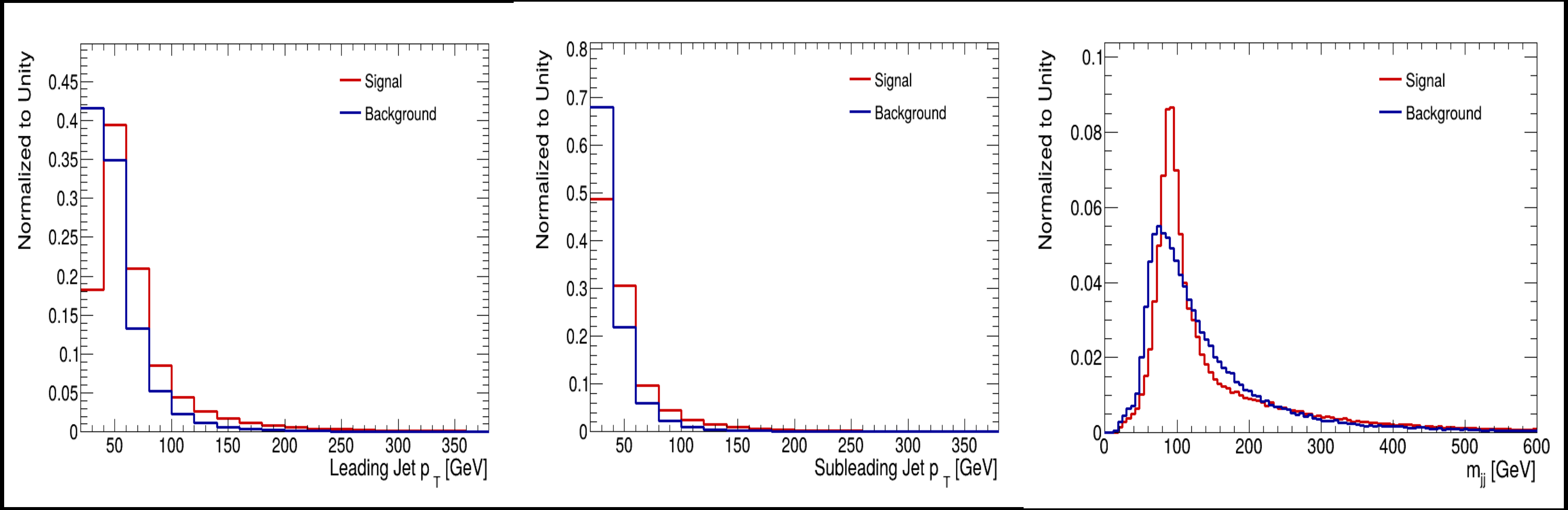


ggF hadronic channel feynman diagram and feature importance

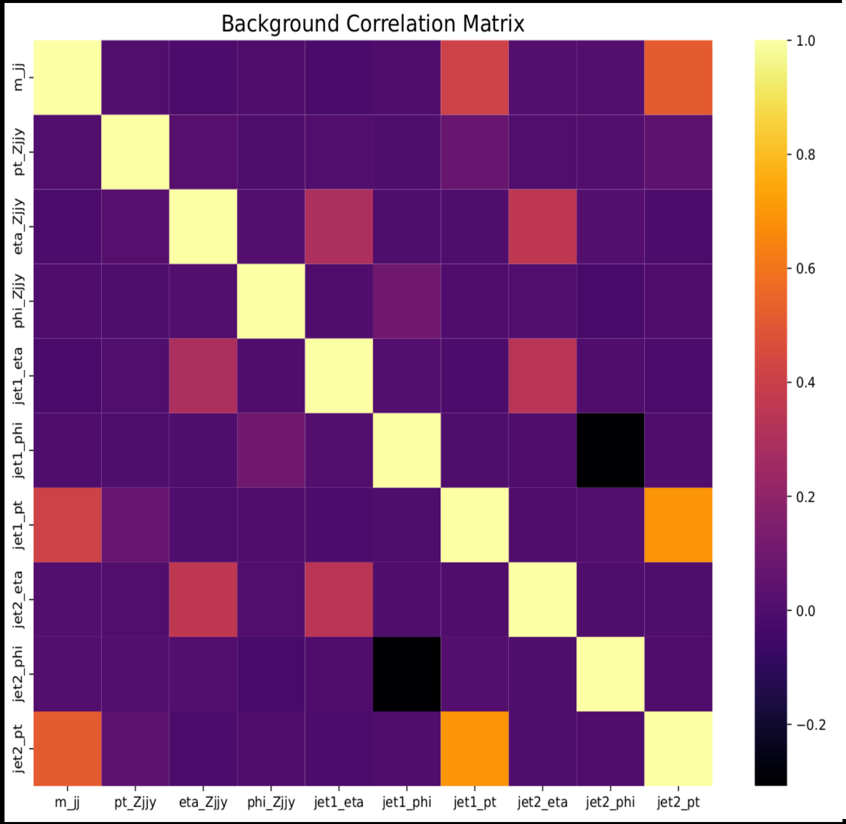


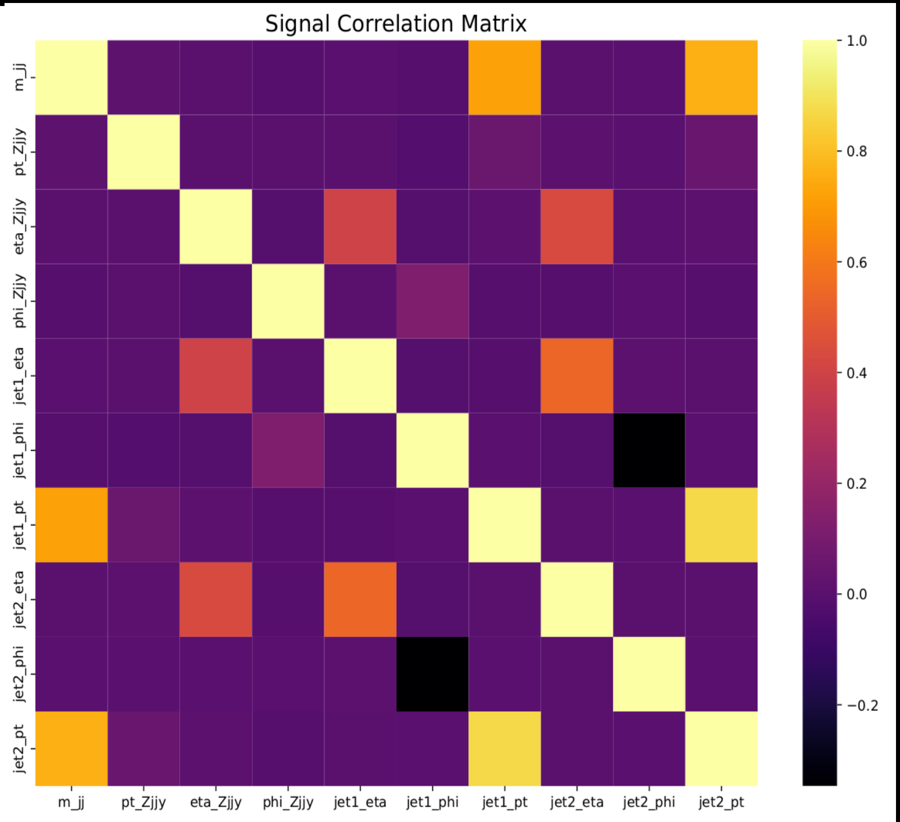


ggF hadronic channel important kinematics

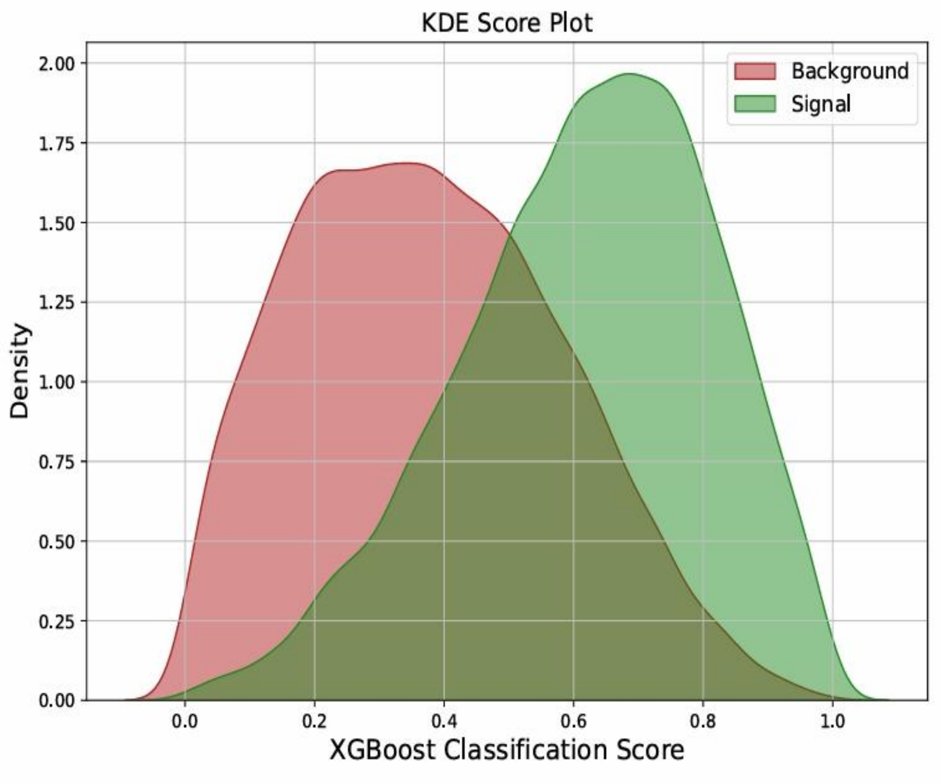


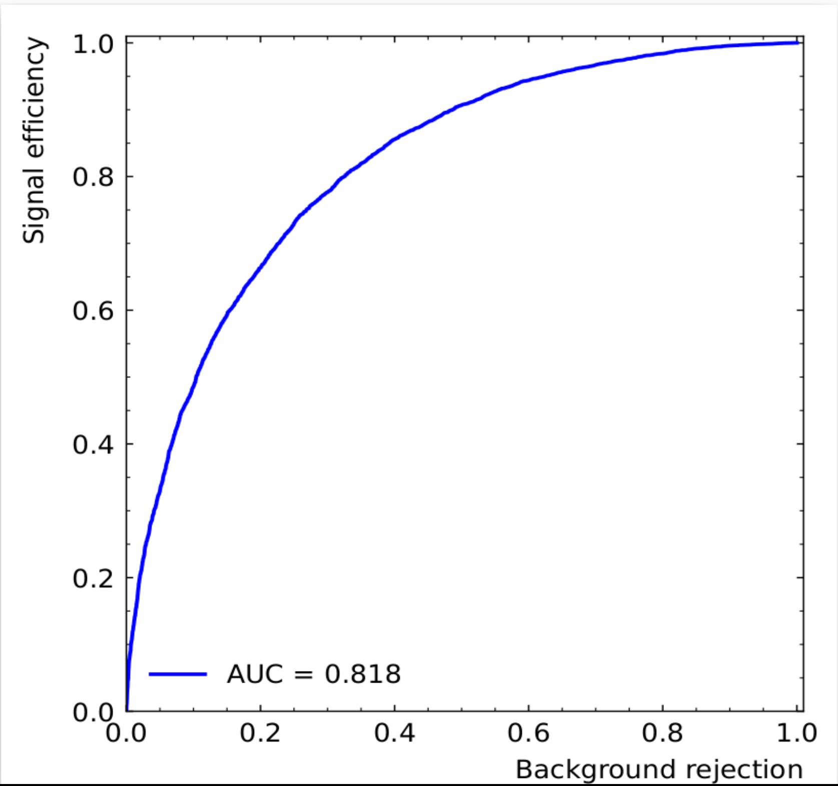
ggF hadronic channel correlation matrix



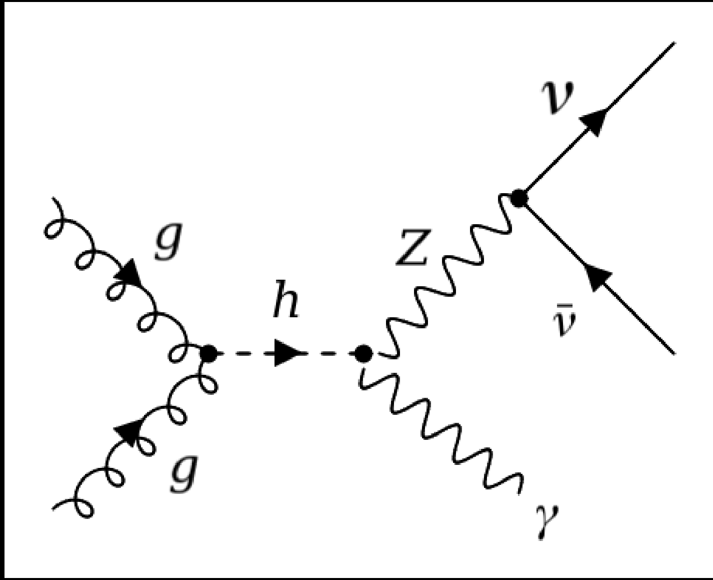


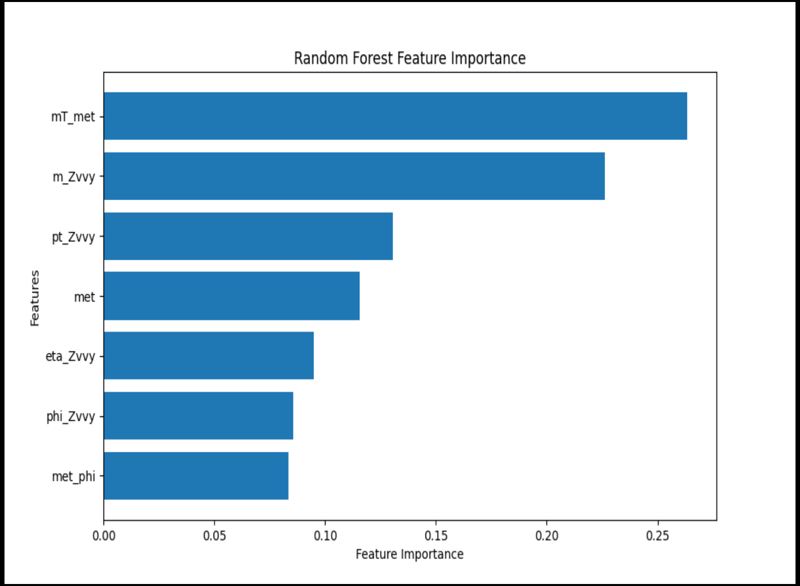
ggF (hadronic channel KDE plot and ROC curve)





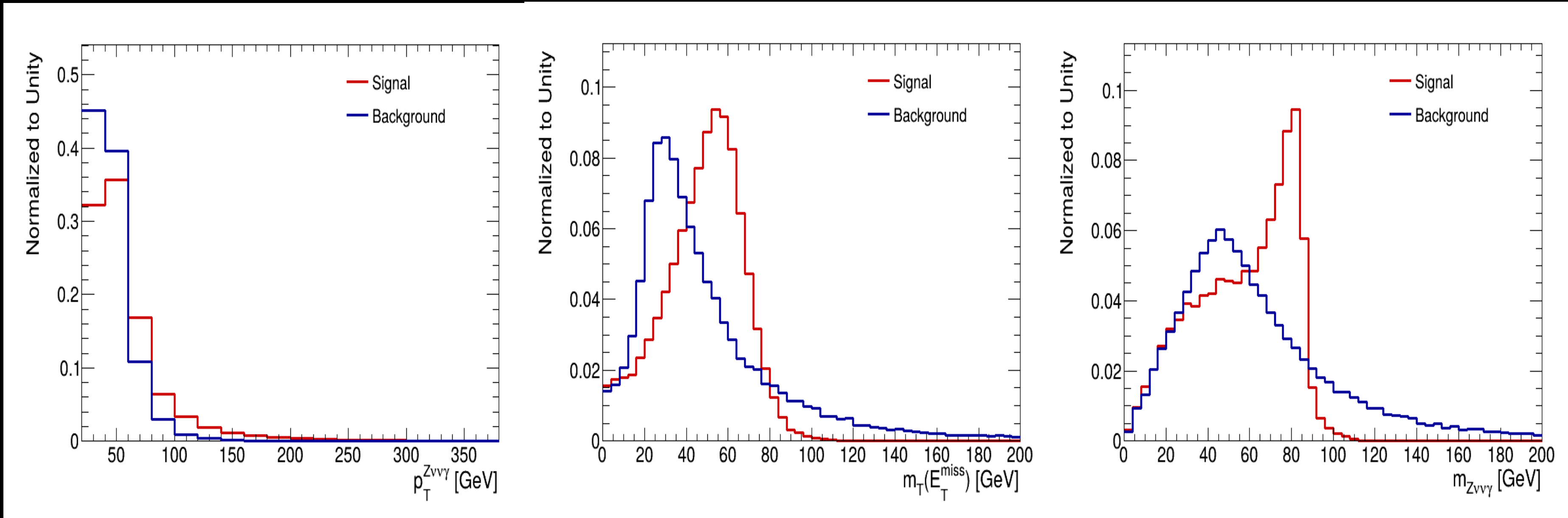
ggF neutrino channel feynman diagram and feature importance



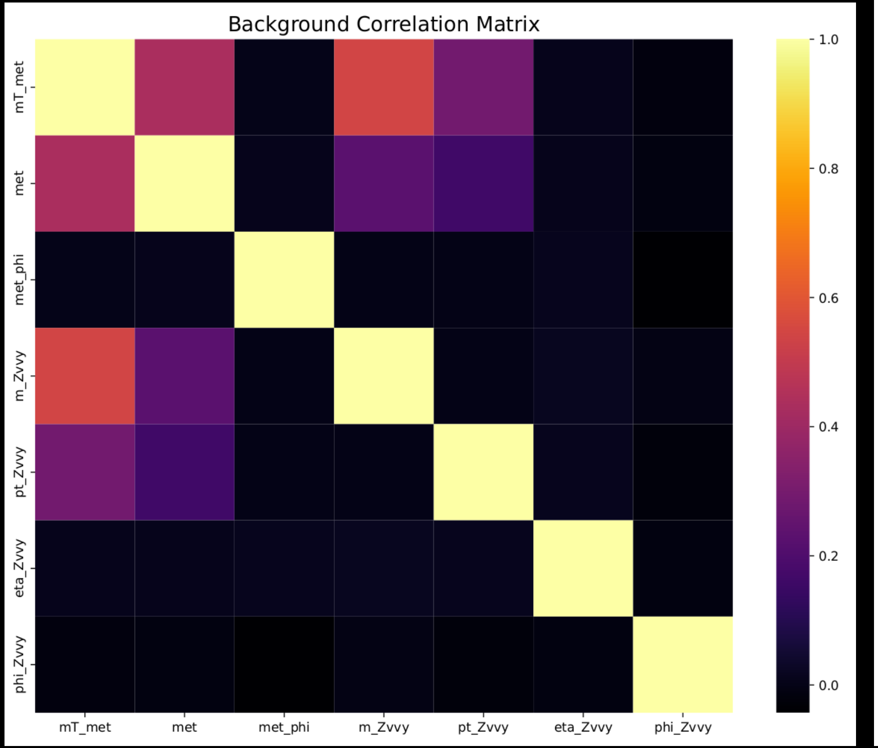


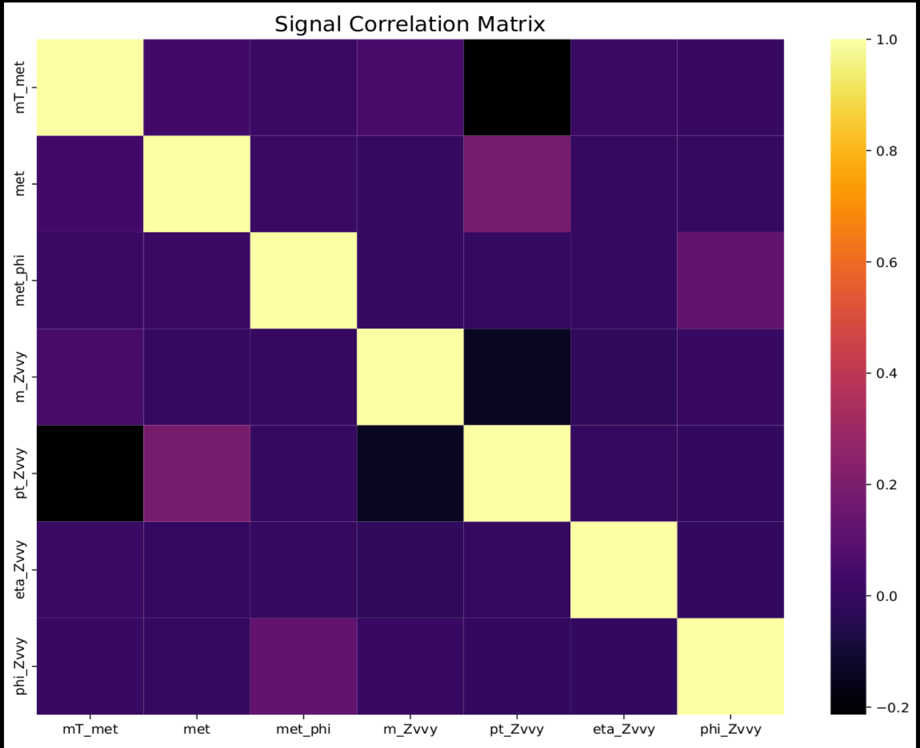
- Cross-section 13.94 pb for background
- Cross-section 0.015 pb for signal

ggF neutrino channel important kinematics

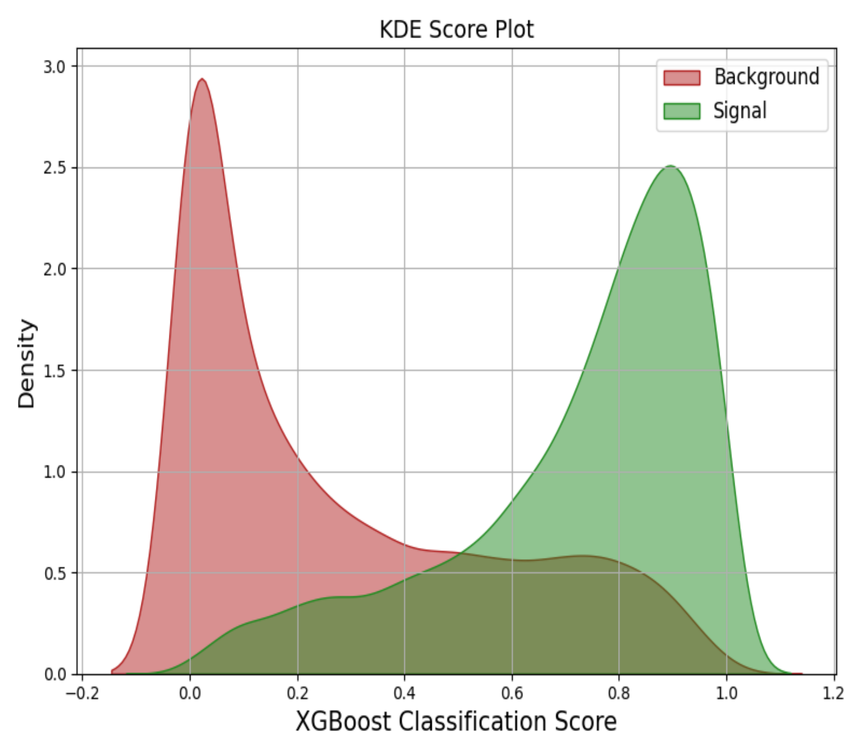


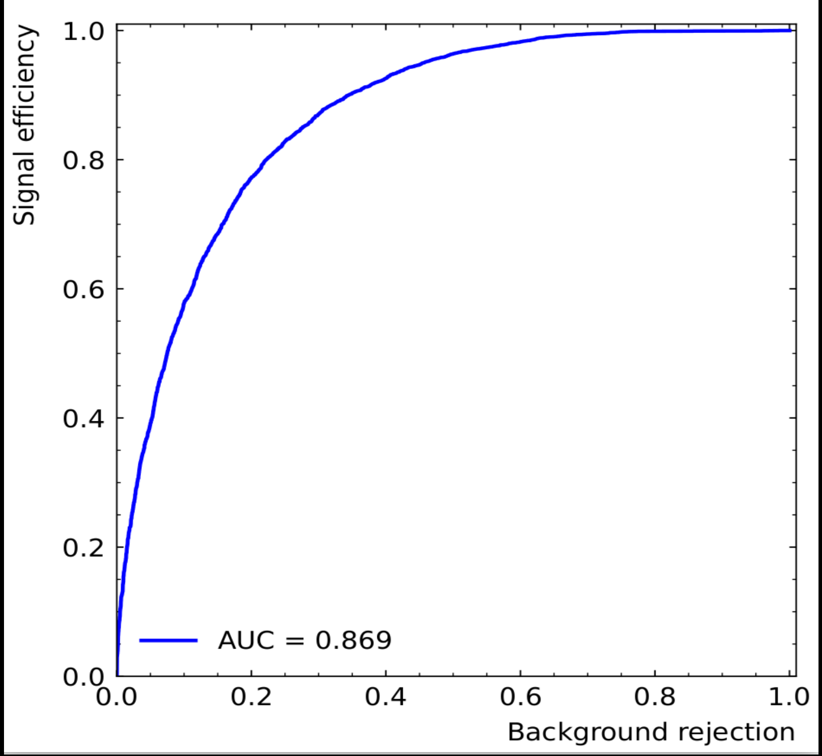
ggF Neutrino channel correlation matrix





Neutrino channel KDE Plot and ROC Curve





KDE Scores

Z Scores	ggF leptonic (+1)	ggF leptonic (-1)	ggF jet	ggF Neutrino
10 Bins	17.48	17.52	4.89	6.27
25 bins	18.28	18.65	4.96	6.48
50 bins	18.68	18.68	4.99	6.59
100 bins	18.83	19.01	5.04	6.67
KDE	24.78	25.04	8.72	11.34

Conclusions

- KDE improves with more ensembles since it avoids binning our classifiers' outputs while the binning method benefits from increased bin counts.

- We are working on introducing three more production channels (VBF,VH and ttH).

- We are about to generate new samples using the best fit values of the Wilson Coefficients.

Thank You

email:

kutlwano.chongo.makgetha@cern.ch