


Probing the Higgs b-quark coupling (y_b): bbH Search Feasibility Study

Francis Lance T. Jumawan[a,b]

[a]*University of the Philippines Diliman,* [b]*CERN*

Supervisors: Dr. Lorenzo Santi[b] and Dr. Valentina Cairo[b]

Abstract

This CERN Summer Student Program project focuses on probing the Higgs boson's bottom Yukawa coupling (y_b) through the associated production of a Higgs boson with two b-quarks (bbH). The bbH production cross section is approximately 0.5 pb at $\sqrt{s}$ = 13 TeV at the LHC and have some destructive interference contributions from other Higgs production processes such as gluon-gluon Fusion (ggH) and Higgs Strahlung (VH). While the decay channel is $H \to b\bar{b}$ have the largest branching ratio, its precision is limited by the resolution to $O(10\%)$. The decay channel for this search is $H \to \gamma\gamma$ to avoid the resolution limitation since photon measurement in the ATLAS detector has a resolution of about $O(1\%)$. In this feasibility study, Run 2 and Run 3 samples from the $HH \to b\bar{b}\gamma\gamma$ analysis are used. The analysis utilizes kinematic selections, newly engineered discriminating variables, and boosted decision trees to enhance signal-background separation. Statistical sensitivity is assessed through binned significance and maximum-likelihood fits. The findings indicate limited sensitivity with the current samples, methods, and channel, highlighting the need for increased signal statistics, improved forward b-tagging, and advanced background separation.

2nd October 2025

Contents

1 Introduction

The Higgs boson, which was discovered in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC), is a major breakthrough to the validity of the Standard Model (SM) of particle physics [1][2]. This discovery provided insight into the mechanism of the electroweak symmetry breaking and fermion mass generation through Yukawa couplings. One of these couplings is the Higgs coupling to bottom quarks (y_b) which governs the strength between the Higgs field and the b-quark. In the SM, y_b is a free parameter whose value must be determined experimentally. Precise measurement of this coupling are essential for testing the SM's predictions and probing signs of new physics [3].

The current approach to constraining y_b is through the dominant Higgs decay channel $H \to b\bar{b}$, which accounts to about $53 \pm 8\%$ of the decays of Higgs [4]. However, this channel suffers from experimental challenges such as the resolution of the reconstructed Higgs mass of about $O(10\%)$ due to the high Quantum Chromodynamic Background (QCD) multijet processes [5]. This limitation motivates the search for alternative production and decay channels that can probe y_b with higher precision.

One such process is the associated production of a Higgs boson with two b-quarks, $pp \to b\bar{b}H$ (bbH). The bbH production cross section is approximately 0.5 pb at $\sqrt{s} = 13$ TeV and it directly depends on y_b at tree level. In the SM, bbH production interferes destructively with contributions from gluon-gluon fusion (ggF) and vector boson–associated production (VH) as seen on Figure 2.1, further reducing the expected event yield.

Previous experimental searches have been conducted in alternative Higgs decay channels. For example, the CMS Collaboration studied the bbH process with the Higgs decaying to $\tau^+\tau^-$ and W^+W^- in the leptonic channel [6]. No significant excess above the background-only hypothesis was observed, and the 95% confidence level (CL) upper limit on the bbH production cross section was set at 3.7 times the SM prediction, with a best-fit value of $y_b = 1.58$.

In this project, we will focus on the the bbH production through the $H \to \gamma\gamma$ decay channel, the diphoton decay channel offers excellent mass resolution and a well-defined resonance peak, enabling better discrimination against QCD-induced diphoton production background.

From theoretical predictions, the b-jets in bbH events are expected to be relatively soft in transverse momentum (p_T) and often produced in the forward region ($|\eta| > 2.5$. The planned upgrades of the ATLAS detector for the High-Luminosity LHC (HL-LHC) will extend tracking coverage to $|\eta| < 4$ potentially increasing the yield to bbH process. But in the absence of HL-LHC Monte Carlo samples, this project uses Run 2 and Run 3 samples from the ATLAS $HH \to b\bar{b}\gamma\gamma$ analysis search [7]. The analysis explores kinematic selections, newly engineered discriminating features, and boosted decision tree (BDT) classifiers to separate signal from background. Sensitivity is assessed through binned significance calculations and maximum-likelihood fits, with the aim of evaluating the challenges and potential improvements needed for future bbH searches at the HL-LHC.

2 Analysis Methodology

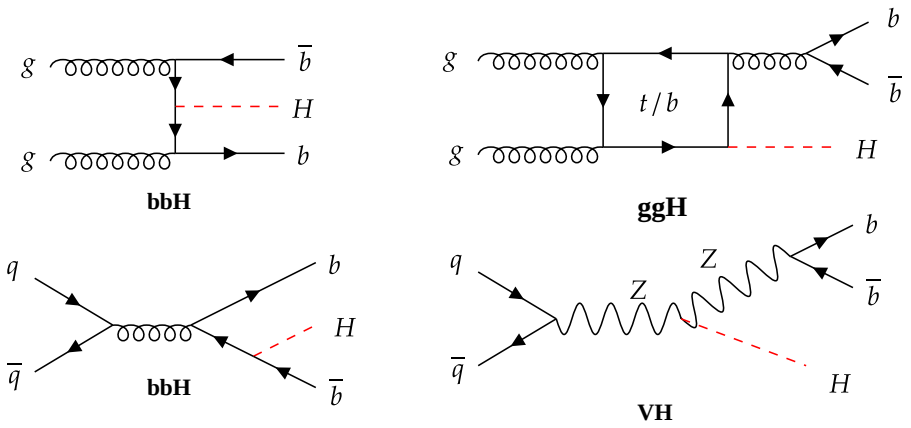

Figure 2.1: bbH Feynman Diagram with the destructive inteference with VH and ggF

2.1 Dataset and Simulation

The analysis was designed to evaluate the feasibility of measuring the bbH process in the $H \to \gamma\gamma$ decay channel using Run 2 and Run 3 samples. The backgrounds considered for this feasibility study are shown on Table 2.1.

DSID	Generator	PDF (ME)	PDF+Tune (PS)	Prod. Mode
345315	POWHEG + PYTHIA8	PDFLHC	A14NNPDF23	bbH (Signal)
343981	NNLOPS + PYTHIA8	PDFLHC	AZNLOCTEQ6	ggH
346214	POWHEG + PYTHIA8	PDFLHC	AZNLOCTEQ6	VBF
345318	POWHEG MiNLO [21] + PYTHIA8	PDFLHC	AZNLOCTEQ6	W^+H
345317	POWHEG MiNLO [21] + PYTHIA8	PDFLHC	AZNLOCTEQ6	W^-H
345319	POWHEG MiNLO [21] + PYTHIA8	PDFLHC	AZNLOCTEQ6	$qq \to ZH$
345061	POWHEG MiNLO [21] + PYTHIA8	PDFLHC	AZNLOCTEQ6	$gg \to ZH$
346525	POWHEG + PYTHIA8	PDFLHC	A14NNPDF23	ttH
346188	MGMCatNLO + PYTHIA8	NNPDF	A14NNPDF23	tHjb
346486	MGMCatNLO + PYTHIA8	NNPDF	A14NNPDF23	tWH
364352	SHERPA 2.2.4 (ME@NLO+PS, AF2)	–	–	$\gamma\gamma$ + 0,1,2,3 jets

Table 2.1: Summary of MC sample generation for background and signal

These backgrounds can be classified as resonant and non-resonant backgrounds. Resonant backgrounds are the background processes that has a Higgs on the production while non-resonant backgrounds do not have Hiigs like $\gamma\gamma$ + jets.

There are two independent signal regions (SR) that this analysis is interested in and that is the events with b-jets number equals 1 (SR1) and b-jets number greater than or equal to 2 (SR2). For the SR1, the samples have initial preselection where the event must have atleast two jets, a diphoton mass requirement of 105 to 160 GeV, and a zero leptons while the SR2 has an extra preselection of having atleast 2 b-jets with the 85% working point. In addition, we explored two different b-tagging efficiencies 77% and 85% working points. Overall, I have 4 SR regions per b-tagging working points, namely SR1_Run2, SR2_Run2, SR1_Run3, and SR2_Run3.

2.2 Feature Engineering

In addition to the features listed in Table 1, several new features are derived to enhance the discriminatory power of the machine learning model. These engineered features capture kinematic relationships, angular distributions, and tagging efficiencies that are particularly relevant for distinguishing the signal from backgrounds in particle physics analyses. Each feature is motivated by physical insights, such as the topology of signal events versus backgrounds. The derived features are described below.

Angular separations between the two leading b-tagged jets and the reconstructed Higgs boson (from the diphoton system). These include the differences in pseudorapidity ($\Delta\eta$), azimuthal angle ($\Delta\phi$), and the angular distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. For events in Signal Region 1 (SR1), where only one b-tagged jet may be explicitly identified, the second b-jet is selected as the one with the next-highest b-tagging score (pcbt). These features exploit the collinear or back-to-back configurations typical in signal events due to the Higgs decay kinematics.

Next is the b-tagging ratio, which quantifies the relative confidence in the leading versus subleading jets being b-jets:

$$b\text{-tag ratio} = \frac{\text{Jet } 1_{pcbt} + \text{Jet } 2_{pcbt}}{\text{Jet } 3_{pcbt} + \text{Jet } 4_{pcbt}} \tag{2.1}$$

where pcbt denotes the b-tagging probability score. In signal events, the two leading jets are more likely to originate from b-quarks, resulting in higher pcbt values for Jets 1 and 2 compared to Jets 3 and 4, leading to a larger ratio than in backgrounds dominated by light-flavor jets.

Next feature is Centrality, a measure that weights jets based on their pseudorapidity to emphasize central activity defined as,

$$\text{Centrality} = \frac{\sum_{i=1}^{4} (\text{Jet } i)_{pT}\, e^{-|\eta|}}{\sum_{i=1}^{4} (\text{Jet } i)_{pT}} \tag{2.2}$$

Signal events often feature forward jets with higher $|\eta|$ and lower p_T due to the associated b-quark production, skewing the centrality distribution lower compared to backgrounds with more central, high-p_T jets.

Next is an improved p_T balance, assessing the overall transverse momentum conservation in the event:

$$p_T \text{ balance} = \frac{\left|\sum_{i=1}^{4} \vec{p_T}^{Jet\ i} + \sum_{i=1}^{2} \vec{p_T}^{\gamma_i}\right|}{\sum_{i=1}^{4} \left|\vec{p_T}^{Jet\ i}\right| + \sum_{i=1}^{2} \left|\vec{p_T}^{\gamma_i}\right|} \tag{2.3}$$

This feature highlights imbalances from missing energy or soft radiation in backgrounds, while signal events tend to exhibit better balance due to the clean Higgs decay topology.

Lastly, p_T symmetry between the bb and $\gamma\gamma$ systems, capturing the relative transverse momenta of the reconstructed Higgs subsystems: - pT Symmetry of bb and $\gamma\gamma$ system defined as

$$p_T^{\gamma\gamma-bb} = \frac{p_T^{\gamma\gamma} - p_T^{bb}}{p_T^{\gamma\gamma} + p_T^{bb}} \tag{2.4}$$

In signal processes, the bb and $\gamma\gamma$ systems are often balanced in p_T.

2.3 Machine Learning

This analysis used a Boosted Decision Tree (BDT) with XGBoost as our classification algorithm to separate the signal and the background. The model is trained as a binary classifier (signal vs. background) using the `binary:logistic` objective, optimized with log-loss and area under the ROC curve (AUC) as evaluation metrics to ensure reliable probabilistic outputs and ranking performance.

Hyperparameter values are defined as shown in Table 2.2. Training proceeds for up to 3000 boosting rounds, with early stopping after 50 unchanged validation rounds to avoid overfitting. Data is split 33/33/33 (train/validation/testing) ensuring no leakage of information.

Parameter	**Value/Range**
objective	binary:logistic
eval_metric	[logloss, auc]
eta (η)	[0.01, 0.02, 0.04]
max_depth	[4, 6, 8]
subsample	0.7
colsample_bytree	0.7
min_child_weight	10
gamma	2.0
lambda	1.0
alpha	1.0
num_boost_round	3000
early_stopping_rounds	50

Table 2.2: Hyperparameters used in the XGBoost classifier

2.4 Statistical Inference

The signifiance of the analysis was calulated by the BDT score range into bins, computing Z defined as

$$Z = \sqrt{2\left((n_s + n_b)\log\left(1 + \frac{n_s}{n_b}\right) - n_s\right)} \tag{2.5}$$

in each bin, and summing in quadrature to get the expected significance.

The fit is performed by minimizing the binned negative log likelihood (N_{LL}), defined as

$$N_{LL} = -\log(\mathcal{L}) \tag{2.6}$$

where $\mathcal{L}$ is the likelihood function, which gives us the probability of seeing n events if the true signal strenght is μ, defined as

$$\mathcal{L}(data|\mu) = \prod_i \mathcal{P}(n_i|\mu \cdot s_i + b_i) \tag{2.7}$$

where, $\mathcal{P}$ means a Poisson distribution defined as

$$\mathcal{P}(n|\mu) = \frac{(s_i\mu + b_i)^n \exp(-(s_i\mu + b_i))}{n!} \tag{2.8}$$

This term means that this observed events n_i follows a Poisson distribution, whose mean is the expected signal (s_i) + background (b_i). In this summer project, I will not include any systematic uncertainties. Minimizing the N_{LL} is the same as maximizing the $\mathcal{L}$, using the N_{LL} is just easier and mathematically better to handle than the likelihood function.

In getting the range of μ that can still explain the the data found with 95% confidence level (CL) is by getting the likelihood ratio test. It is defined as

$$q(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu)}{\mathcal{L}(\mu_{best})}\right) \tag{2.9}$$

The 95% region is defined as all the μ values within $q(\mu) \leq 3.84$, which is the chi-square distribution with 1 degree of freedom.

2.5 Data and code availability

The analysis pipeline used in this study is publicly available on GitLab[1]. The repository provides a complete workflow for preparing datasets and training a Boosted Decision Tree (BDT) classifier for the $pp \to b\bar{b}H$ analysis in the $H \to \gamma\gamma$ channel. It supports both Run 2 and Run 3 datasets, multiple dataset versions, and flexible background selections. Trained models and evaluation results are stored for reproducibility. The repository requires Python 3, ROOT, and standard scientific packages (pandas, numpy, xgboost, uproot, scikit-learn) and provides example workflows for training and evaluation.

[1] https://gitlab.cern.ch/fjumawan/bbhyy-analysis

3 Results

Shown in Figure 3.2 and 3.1 are the defined SR regions for this analysis. The zero b-tag region, while containing signal events, is dominated by large backgrounds and was excluded from further optimization. Shown in the Appendix 4 are the other kinematic distributions of the observables.

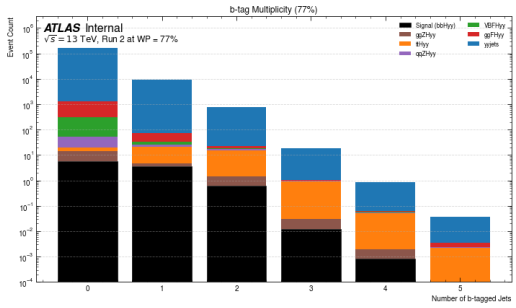

Figure 3.1: n-$b$tag Dsitribution for WP = 77%

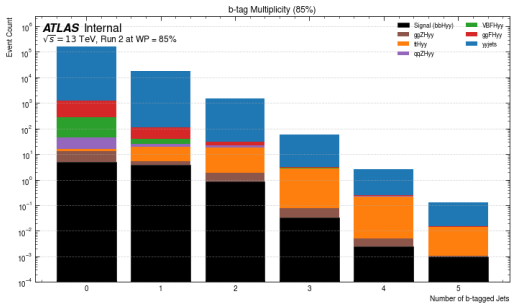

Figure 3.2: n-$b$tag Dsitribution for WP = 77%

The classifier achieved an average AUC of about 0.8 as seen on Figure 3.3 and Figure 3.4, this indicates moderate discrimination power between signal and background. It is significantly better than random guessing (AUC = 0.5), but still far from a good quality classifier (AUC $\approx$ 0.95) that would be achievable with more distinct features or higher-quality statistics. The AUC curve for SR2_Run3 at WP = 77% and SR1_Run3 at WP = 85% indicates that the separation of the signal and background are unstable. This might be because of the overlapping kinematic phase space of bbH and the major resonant backgrounds such as VBF and ggF.

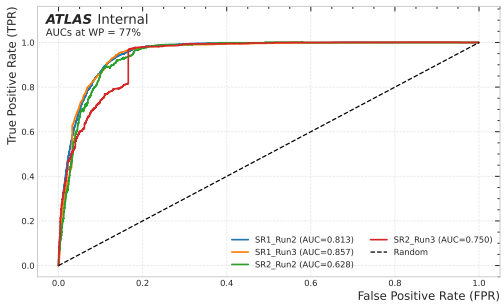

Figure 3.3: AUC curves at WP = 77%

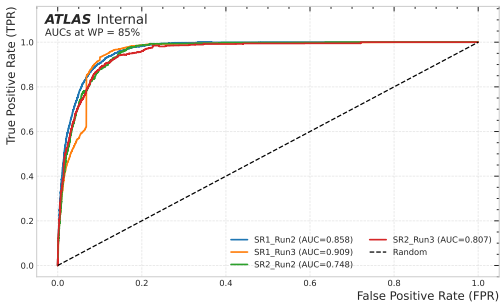

Figure 3.4: AUC curves at WP = 85%

During training, some models stopped at a certain boosting round, as shown in Figures 3.5 and 3.6. While most models converged smoothly, the SR1_Run3 model stopped early for both working point configurations.

Since all models were trained with the same parameters (Table 2.2), it is unlikely that the early stopping was caused by overfitting or an excessively large learning rate. A plausible explanation is that, in this signal region, the available features provide limited information, so the model quickly reaches a point where no additional signal can be learned, resulting in an early plateau in the boosting rounds. This may explain the slightly unstable behavior observed in the AUC curve for Run3.

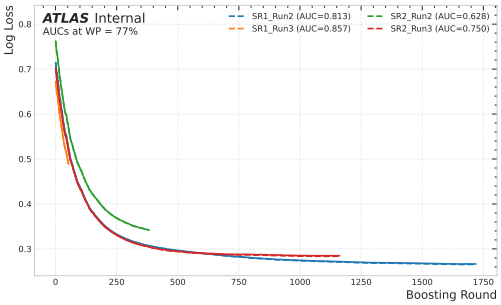

Figure 3.5: AUC curves at WP = 77%

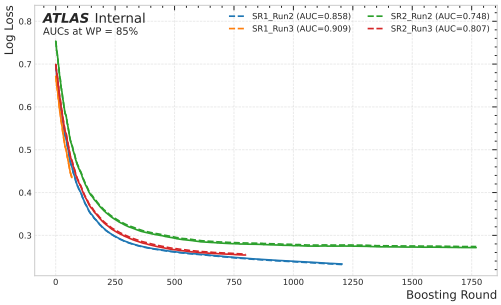

Figure 3.6: AUC curves at WP = 85%

The model metrics of all the BDT trained is summarize in Table 3.1. The performance of our models was evaluated across two working points (WP = 77%, 85%) and two signal regions (SR1, SR2) for different training configurations (Run2, Run3). While some of the AUC curves for Run 3 are unstable, increasing the working point at this run consistently improves AUC and accuracy, indicating better overall discrimination. SR1 consistently achieves higher AUC and F1 than SR2, suggesting that features or signal characteristics in SR1 are more informative since there are more signals in SR1 than SR2. Recall remains very high across all configurations, demonstrating that the model effectively identifies nearly all signal events. However, precision and F1 remains low, indicating significant contamination from background events, a consequence of the strongly imbalanced dataset and the high cross section of resonant backgrounds.

Configuration	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**
		WP = 77%			
SR1_Run2	0.8132	0.5102	0.0063	0.9339	0.0125
SR1_Run3	0.8573	0.6744	0.0258	0.9223	0.0502
SR2_Run2	0.6284	0.2091	0.0037	0.9514	0.0073
SR2_Run3	0.7499	0.3915	0.0132	0.9206	0.0260
		WP = 85%			
SR1_Run2	0.8579	0.6331	0.0060	0.9324	0.0119
SR1_Run3	0.9094	0.7633	0.0215	0.9194	0.0421
SR2_Run2	0.7480	0.3087	0.0040	0.9460	0.0080
SR2_Run3	0.8066	0.4590	0.0131	0.9355	0.0257

Table 3.1: Model Metric Results for 77% and 85% Models

The top 10 observables that the model uses most often in separating the signal and background are shown in Figure 3.7 and Figure 3.8 for SR1 and SR2 respectively. Both SR have the same most discriminating

observable which is the invariant mass of Higgs from the diphoton system ($m_{\gamma\gamma}$). This is expected since the most dominant background, in terms of number, is the non-resonant $\gamma\gamma$ + jets has broad distribution without a peak in the value of the invariant mass of the Higgs. Though not much, it is good that the Centrality as defined in Section 2.2 provide additional separation power beyond the standard kinematic quantities, justifying their inclusion in the features.

In the SR1, since your signal bbH has a b-quark in association, the b-tagging score of the leading jet helps distinguish signal from light-flavor jets in backgrounds like ggH or VBF. While in SR2, the $m_{b\bar{b}}$ distribution relects the kinematics of the associated b-quark from bbH which would have different distribution if the b-quarks are not correlated like the backgrounds ggH and $\gamma\gamma$+ jets.

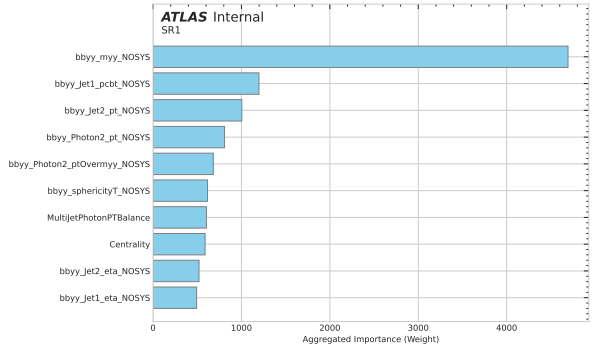

Figure 3.7: Importance Ranking at SR1

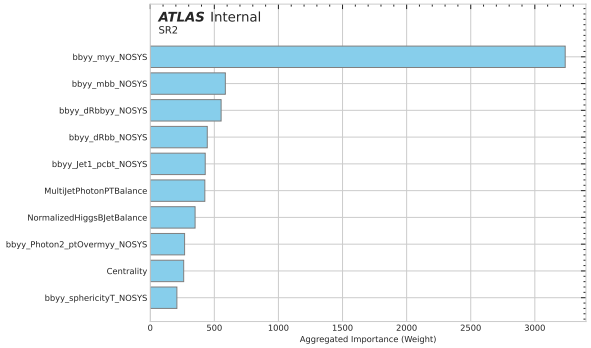

Figure 3.8: Importance Ranking at SR2

The BDT score distribution for all the models in each configuration is shown in Figure 3.9 and Figure 3.10. The BDT combines multiple kinematic and event-level features into a single discriminant variable, the BDT score. High values indicate events more signal-like (1), while low values indicate background-like events (0). All the signal score distribution tends to cluster at high BDT scores, because the BDT has learned the characteristic kinematics of the bbH signal while tends to be more spread out or skewed toward lower BDT scores. However, SR1_Run3 for both WP did not reached a good separation.

The distributions on SR2 usually shows better separation, as events with two genuine b-jets strongly favor signal. In general, SR1 provides more statistics but less information on the quality of the signal while the SR2 provides good kinematics for the signal but has less statistics due to the geometry of the detector.

The BDT scores in the plot are normalized, which highlights a distinct peak for the signal. However, when the event weights and normalization are removed, a clear peak from the background appears in the signal-like region. This peak originates from the resonant backgrounds.

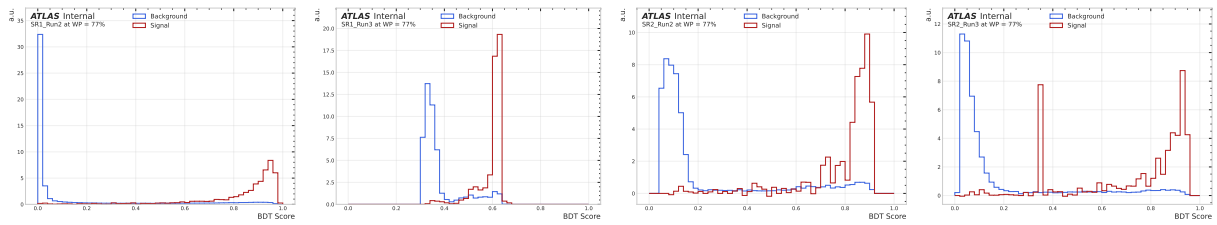

Figure 3.9: BDT Score distributions for WP = 77%

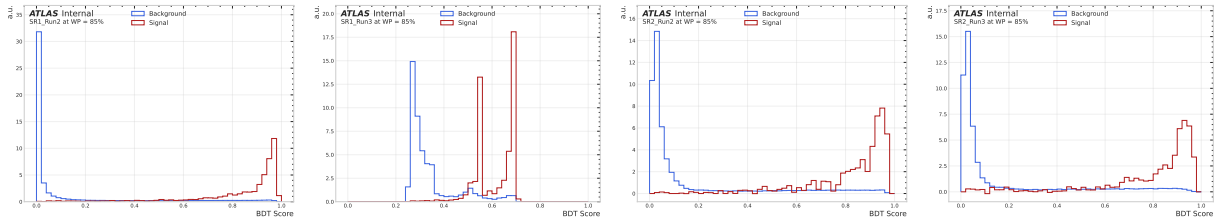

Figure 3.10: BDT Score distributions for WP = 85%

The significance calculated for all the 8 categories are consistently well below 1σ, with SR1 providing slightly higher values in Run 2, and SR2 improving in Run 3 for WP = 77% as seen on Table 3.2.

Configuration	**Z Value**	**Signal Strength (μ)**	**95% CL Interval**
WP = 77 ($Z_T = 0.1292$)			
SR1_Run2	0.0831	1.000	[-10.69, 15.86]
SR1_Run3	0.0517	1.000	[-17.93, 25.52]
SR2_Run2	0.0404	1.000	[-22.76, 32.76]
SR2_Run3	0.0740	1.000	[-10.69, 20.69]
WP = 85 ($Z_T = 0.1300$)			
SR1_Run2	0.0796	1.000	[-10.69, 18.28]
SR1_Run3	0.0586	1.000	[-15.52, 23.10]
SR2_Run2	0.0556	1.000	[-15.52, 25.52]
SR2_Run3	0.0635	1.000	[-13.10, 23.10]

Table 3.2: Statistics Results for WP = 77 and WP = 85

The analysis shows no significant excess of bbH signal in either Run 2 or Run 3, with all Z values below 0.13. Signal strength is fixed to the SM expectation, and the 95% confidence intervals are broad, reflecting limited sensitivity in SR1 and SR2. Differences between b-tagging working points and runs are minor, indicating that the current dataset is not yet sufficient to measure or constrain the bbH production rate in these regions.

4 Conclusion

This CERN Summer Student Program Project investigated the feasibility of probing the Higgs–bottom quark Yukawa coupling (y_b) through the $pp \to b\bar{b}H$ (bbH) process in the $H \to \gamma\gamma$ decay channel, using Run 2 and Run 3 ATLAS Monte Carlo samples from the $HH \to b\bar{b}\gamma\gamma$ analysis as a proxy in the absence of HL-LHC data.

The bbH production cross section is approximately 0.5 pb at $\sqrt{s}$ = 13 TeV and suffers from destructive interference with gluon-gluon fusion (ggF) and vector boson–associated production (VH), further reducing the expected yields. Combined with the low branching ratio of $H \to \gamma\gamma$, this results in very limited statistics, making the observation of a significant signal extremely challenging with the current datasets.

Kinematic studies indicated that bbH events tend to feature forward, relatively soft b-jets, while resonant backgrounds (especially ggF and VBF) share many kinematic similarities with the signal. Engineered features, including jet centrality, b-tagging scores, and event shape variables, were used to train boosted decision tree (BDT) classifiers. These classifiers achieved moderate signal–background separation, with an AUC of approximately 0.75 and stable training–validation performance.

Despite this, the statistical analysis across all signal regions (SR1 and SR2), runs (Run 2 and Run 3), and b-tagging working points (WP = 77% and 85%) showed no significant excess, with all Z values below 0.13. The signal strength fits were fixed to the Standard Model expectation (μ = 1.0), and the resulting 95% confidence intervals were extremely wide, demonstrating that the current datasets provide no meaningful constraint on bbH production in the diphoton channel.

These results indicate that, with the current Run 2 and Run 3 statistics and acceptance, the bbH process in $H \to \gamma\gamma$ cannot be measured with useful precision. Future HL-LHC searches could improve sensitivity through several avenues: extending forward b-jet acceptance ($|\eta| < 4$) to increase signal yield, applying advanced multivariate techniques (e.g., multiclass BDTs or graph neural networks) for better separation from resonant backgrounds, optimizing b-tagging efficiency in the forward region, and developing novel discriminating features tailored to the unique kinematics of bbH. With these improvements, the bbH channel in $H \to \gamma\gamma$ could provide a complementary and direct probe of y_b at the HL-LHC, helping to refine our understanding of the Higgs sector and its couplings.

Bibliography

[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, arXiv: 1207.7214 [hep-ex] (cit. on p. 3).

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS detector at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 3).

[3] L. Santi, *The Beauty and Charm Yukawa Couplings of the Higgs Boson with the ATLAS Detector at the LHC. - 4D Tracking, Particle Flow and Jet Flavour Reconstruction Algorithms Development*, 2024 (cit. on p. 3).

[4] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **110** (2024) 030001 (cit. on p. 3).

[5] V. Cairo, *Collider Physics*, Presentation at CHIPP Winter School, 2025, URL: https://indico.cern.ch/event/1463355/sessions/569297/attachments/2998746/5283773/ColliderPhysics_CHIPPWinterSchool_VMMCAIRO_20January2025.pdf (cit. on p. 3).

[6] CMS Collaboration, *Measurement of the Higgs boson production via vector boson fusion and its decay into bottom quarks in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, (2023), arXiv: 2308.01253 [hep-ex] (cit. on p. 3).

[7] A. Collaboration, *Study of Higgs boson pair production in the* $HH \to b\bar{b}\gamma\gamma$ *final state with 308* fb^{-1} *of data collected at* $\sqrt{s} = 13$ *TeV and 13.6 TeV by the ATLAS experiment*, (2025), arXiv: 2507.03495 [hep-ex] (cit. on p. 3).

Appendices

Kinematic Distribution of Observables

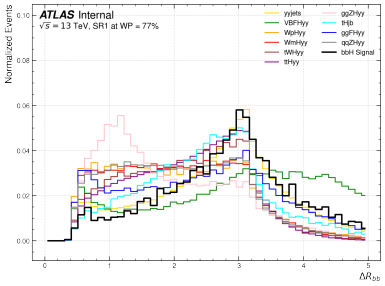

Figure 1: dRbb

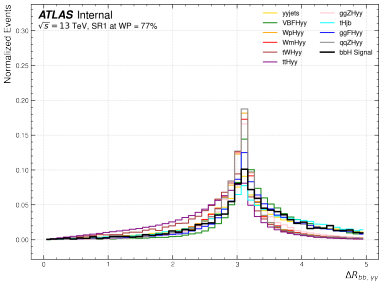

Figure 2: dRbbyy

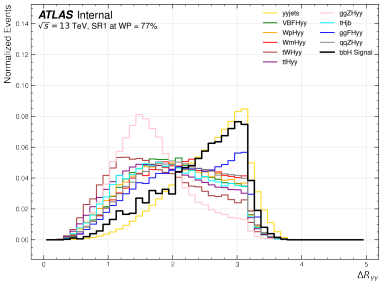

Figure 3: dRyy

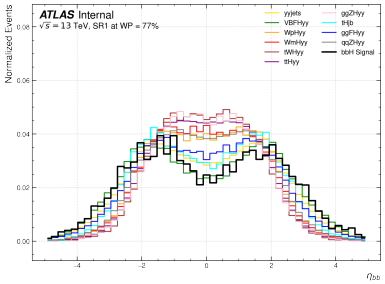

Figure 4: Etabb

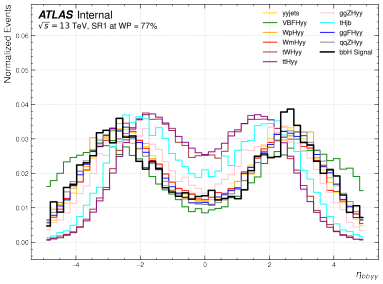

Figure 5: Etabbyy

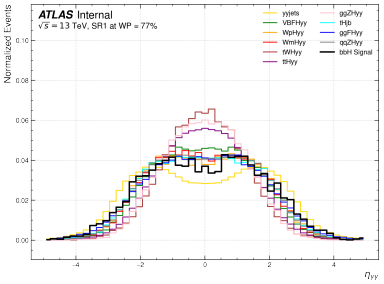

Figure 6: Etayy

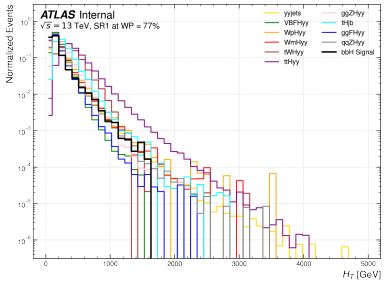

Figure 7: HT

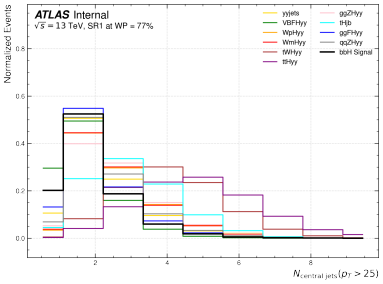

Figure 8: nCentralJets25

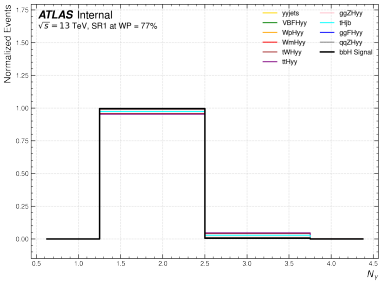

Figure 9: nPhotons

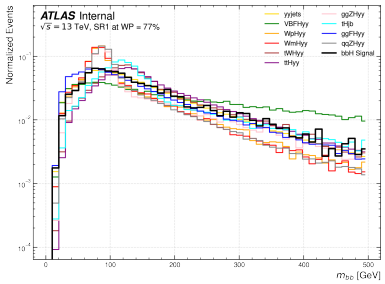

Figure 10: mbb

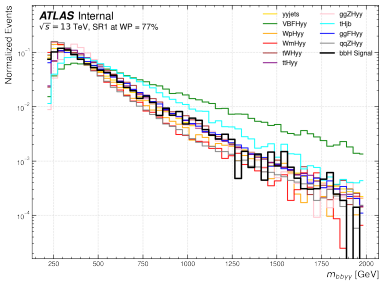

Figure 11: mbbyy

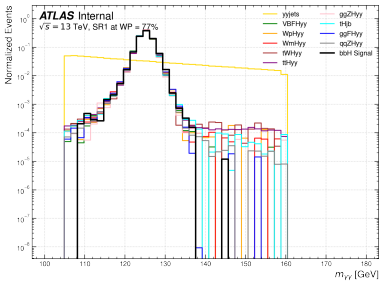

Figure 12: myy

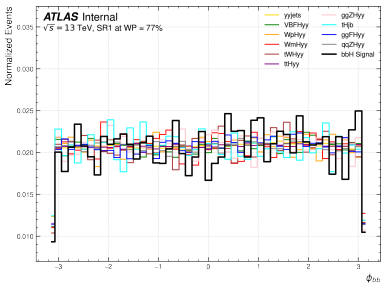

Figure 13: Phibb

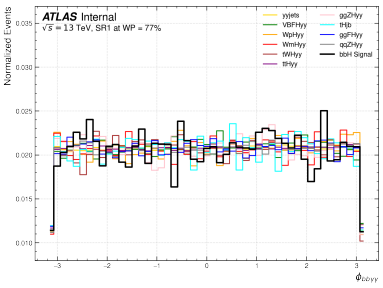

Figure 14: Phibbyy

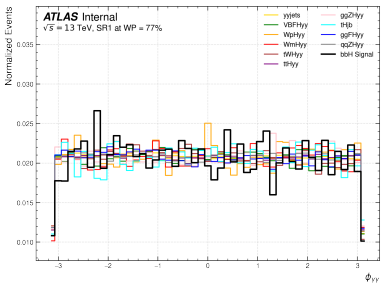

Figure 15: Phiyy

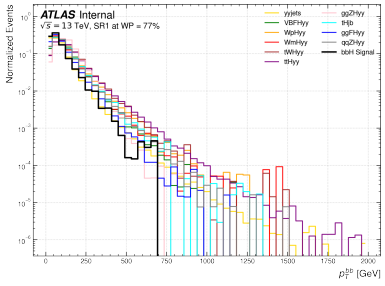

Figure 16: pTbb

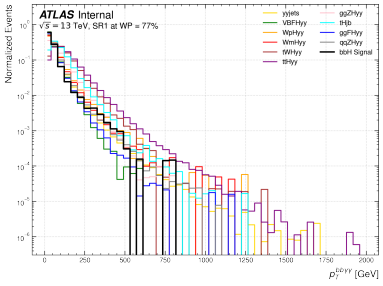

Figure 17: pTbbyy

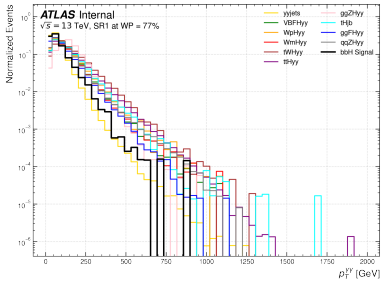

Figure 18: pTyy

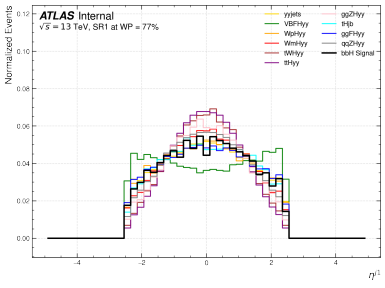

Figure 19: Jet1 eta

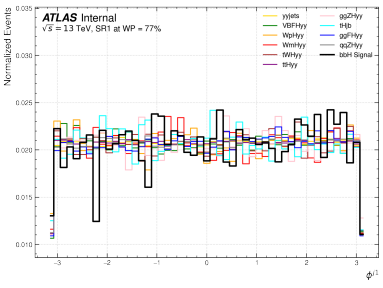

Figure 20: Jet1 phi

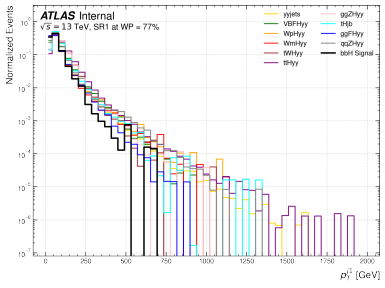

Figure 21: Jet1 pt

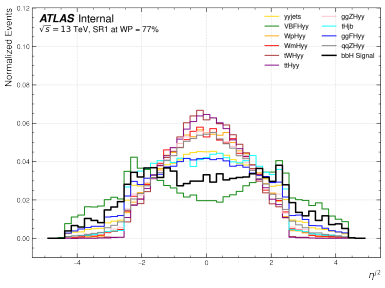

Figure 22: Jet2 eta

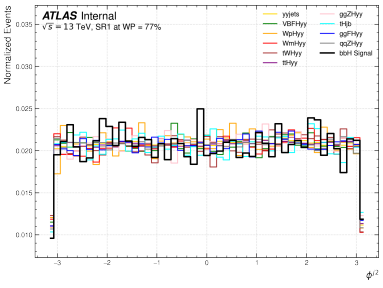

Figure 23: Jet2 phi

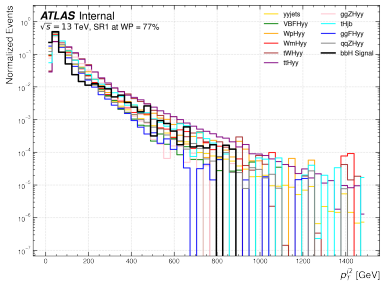

Figure 24: Jet2 pt

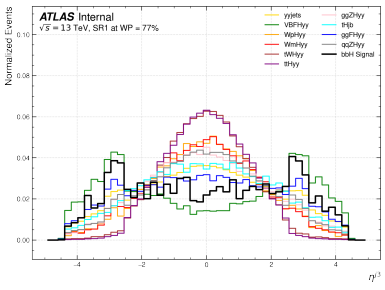

Figure 25: Jet3 eta

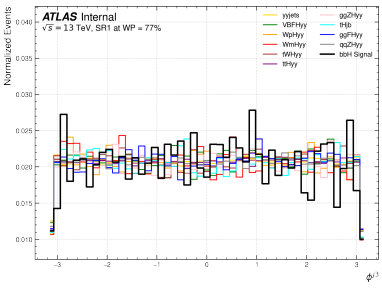

Figure 26: Jet3 phi

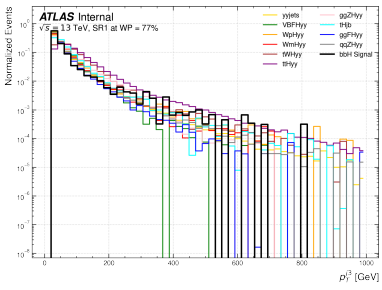

Figure 27: Jet3 pt

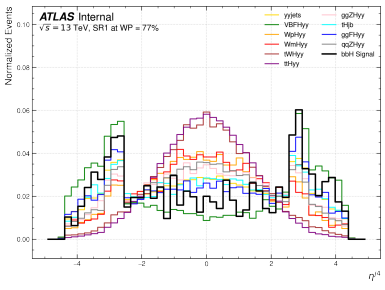

Figure 28: Jet4 eta

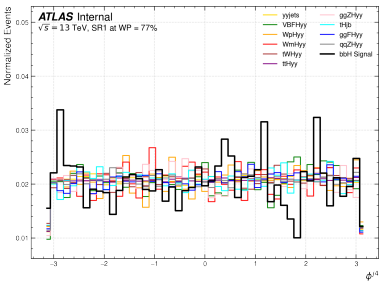

Figure 29: Jet4 phi

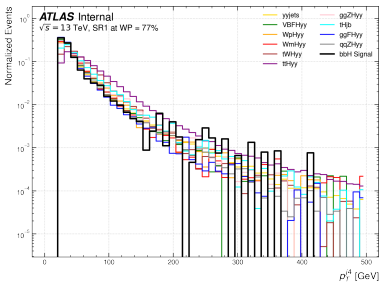

Figure 30: Jet4 pt

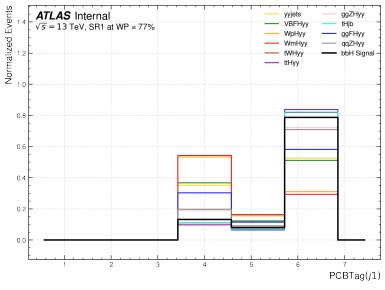

Figure 31: Jet1 pcbt

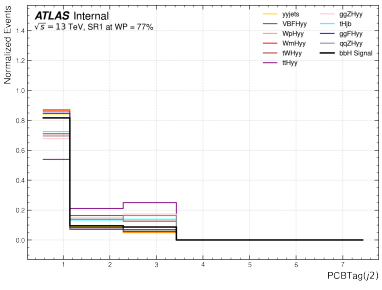

Figure 32: Jet2 pcbt

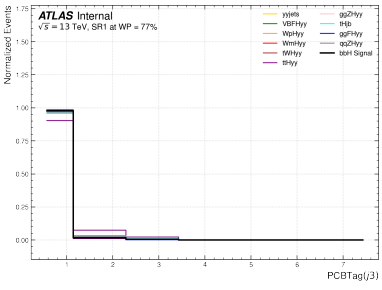

Figure 33: Jet3 pcbt

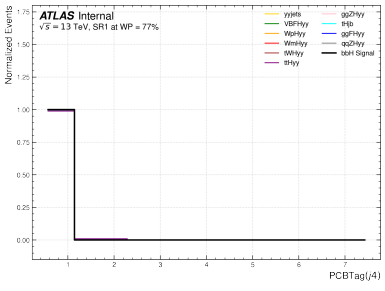

Figure 34: Jet4 pcbt

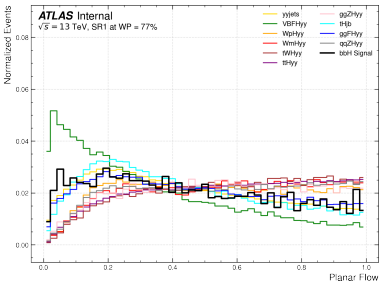

Figure 35: Planar Flow

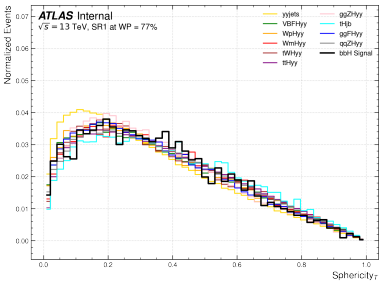

Figure 36: SphericityT

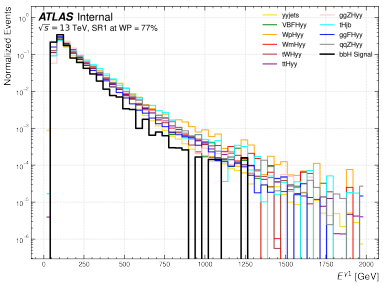

Figure 37: Photon1 E

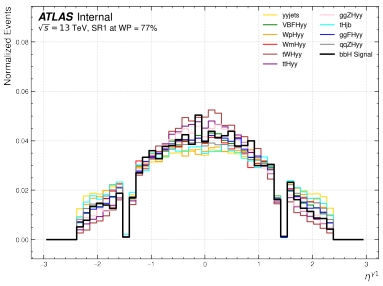

Figure 38: Photon1 eta

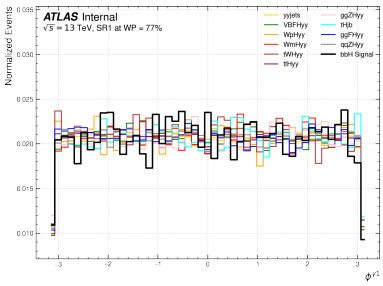

Figure 39: Photon1 phi

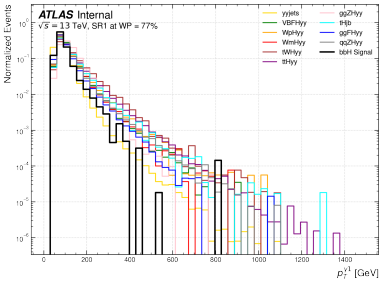

Figure 40: Photon1 pt

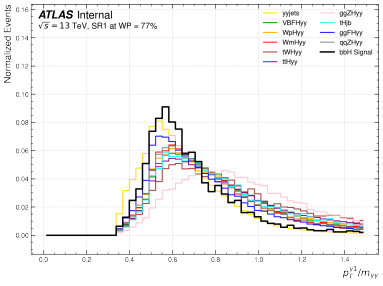

Figure 41: Photon1 ptOvermyy

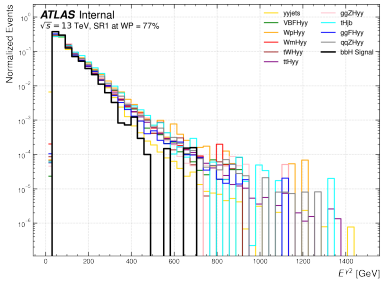

Figure 42: Photon2 E

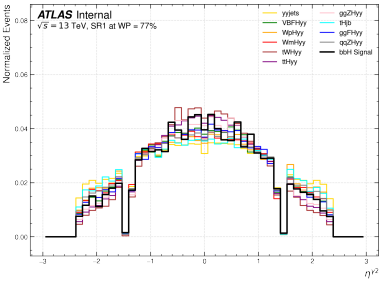

Figure 43: Photon2 eta

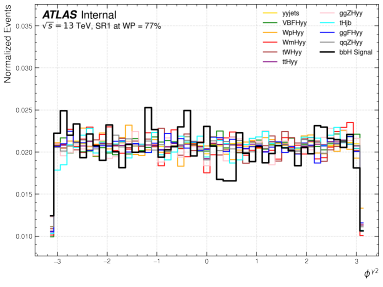

Figure 44: Photon2 phi

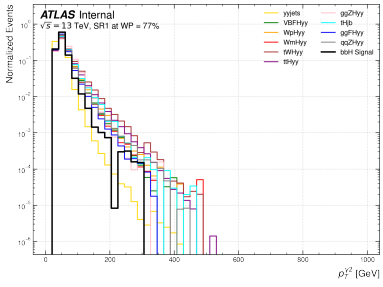

Figure 45: Photon2 pt

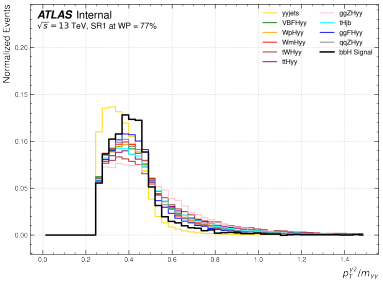

Figure 46: Photon2 ptOvermyy

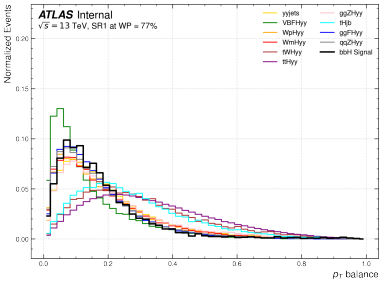

Figure 47: pTBalance

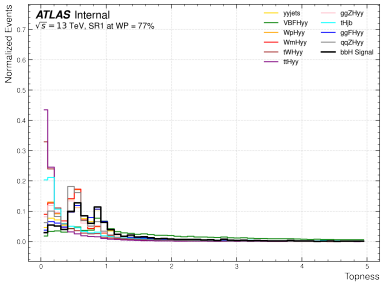

Figure 48: Topness

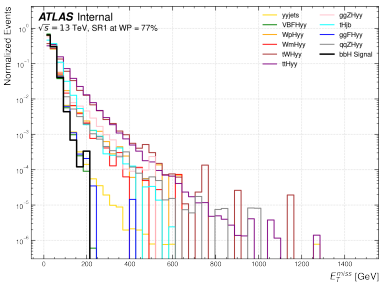

Figure 49: MET

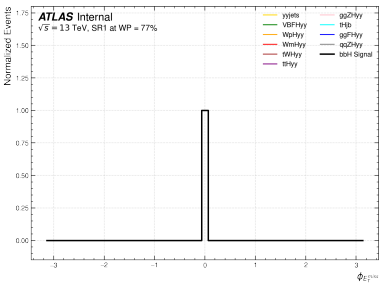

Figure 50: MET phi

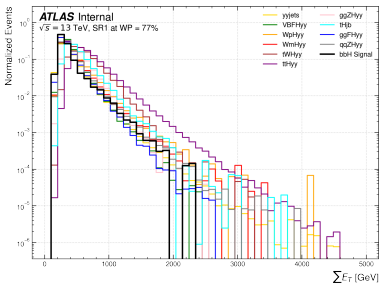

Figure 51: MET sumet

Features

Feature (Variable Type)	Description
bbyy_Photon1_E_NOSYS (F)	Energy of the leading photon.
bbyy_Photon1_eta_NOSYS (F)	Pseudorapidity (η) of the leading photon.
bbyy_Photon1_phi_NOSYS (F)	Azimuthal angle (ϕ) of the leading photon.
bbyy_Photon1_pt_NOSYS (F)	Transverse momentum (pT) of the leading photon.
bbyy_Photon1_ptOvermyy_NOSYS (F)	pT of the leading photon divided by the di-photon invariant mass.
bbyy_Photon2_E_NOSYS (F)	Energy of the subleading photon.
bbyy_Photon2_eta_NOSYS (F)	Pseudorapidity of the subleading photon.
bbyy_Photon2_phi_NOSYS (F)	Azimuthal angle of the subleading photon.
bbyy_Photon2_pt_NOSYS (F)	Transverse momentum of the subleading photon.
bbyy_Photon2_ptOvermyy_NOSYS (F)	pT of the subleading photon divided by the di-photon invariant mass.
bbyy_nCentralJets25_NOSYS	Number of central jets
bbyy_Jet1_eta_NOSYS (F)	Pseudorapidity of the leading jet.
bbyy_Jet1_phi_NOSYS (F)	Azimuthal angle of the leading jet.
bbyy_Jet1_pt_NOSYS (F)	Transverse momentum of the leading jet.
bbyy_Jet1_pcbt_NOSYS (I)	b-tagging score for the leading jet (probability of being a b-jet).
bbyy_Jet2_eta_NOSYS (F)	Pseudorapidity of the subleading jet.
bbyy_Jet2_phi_NOSYS (F)	Azimuthal angle of the subleading jet.
bbyy_Jet2_pt_NOSYS (F)	Transverse momentum of the subleading jet.
bbyy_Jet2_pcbt_NOSYS (I)	b-tagging score for the subleading jet.
bbyy_Jet3_eta_NOSYS (F)	Pseudorapidity of the third jet.
bbyy_Jet3_phi_NOSYS (F)	Azimuthal angle of the third jet.
bbyy_Jet3_pt_NOSYS (F)	Transverse momentum of the third jet.
bbyy_Jet3_pcbt_NOSYS (I)	b-tagging score for the third jet.
bbyy_Jet3_truthLabel_NOSYS (I)	Truth-level label for the third jet.
bbyy_Jet4_eta_NOSYS (F)	Pseudorapidity of the fourth jet.
bbyy_Jet4_phi_NOSYS (F)	Azimuthal angle of the fourth jet.
bbyy_Jet4_pt_NOSYS (F)	Transverse momentum of the fourth jet.
bbyy_Jet4_pcbt_NOSYS (I)	b-tagging score for the fourth jet.
bbyy_myy_NOSYS (F)	Invariant mass of the di-photon system.
bbyy_pTyy_NOSYS (F)	Transverse momentum of the di-photon system.
bbyy_Etayy_NOSYS (F)	Pseudorapidity of the di-photon system.
bbyy_Phiyy_NOSYS (F)	Azimuthal angle of the di-photon system.
bbyy_dRyy_NOSYS (F)	Angular distance (ΔR) between the two photons.
bbyy_mbb_NOSYS (F)	Invariant mass of the di-b-jet system.
bbyy_pTbb_NOSYS (F)	Transverse momentum of the di-b-jet system.
bbyy_Etabb_NOSYS (F)	Pseudorapidity of the di-b-jet system.
bbyy_Phibb_NOSYS (F)	Azimuthal angle of the di-b-jet system.
bbyy_dRbb_NOSYS (F)	Angular distance (ΔR) between the two b-jets.
bbyy_mbbyy_NOSYS (F)	Invariant mass of the bbyy (di-b-jet, di-photon) system.
bbyy_pTbbyy_NOSYS (F)	Transverse momentum of the bbyy system.
bbyy_Etabbyy_NOSYS (F)	Pseudorapidity of the bbyy system.
bbyy_Phibbyy_NOSYS (F)	Azimuthal angle of the bbyy system.
bbyy_dRbbyy_NOSYS (F)	Angular distance (ΔR) between the di-b-jet and di-photon systems.
bbyy_HT_NOSYS (F)	Scalar sum of transverse momenta of all jets and photons (HT).
bbyy_pTBalance_NOSYS (F)	Transverse momentum balance of the bbyy system.
bbyy_planarFlow_NOSYS (F)	Planar flow variable, describing event shape in the transverse plane.
bbyy_sphericityT_NOSYS (F)	Transverse sphericity, a measure of event isotropy.
bbyy_topness_NOSYS (F)	Variable to distinguish top quark events from signal (e.g., HH).
met_met_NOSYS (F)	Magnitude of missing transverse energy.
met_phi_NOSYS (F)	Azimuthal angle of the missing transverse energy.
met_sumet_NOSYS (F)	Scalar sum of transverse energy contributing to MET.

Table 1: List of branches used for analysis