Search for Higgs Pair Production in $bb\gamma\gamma$ Final States

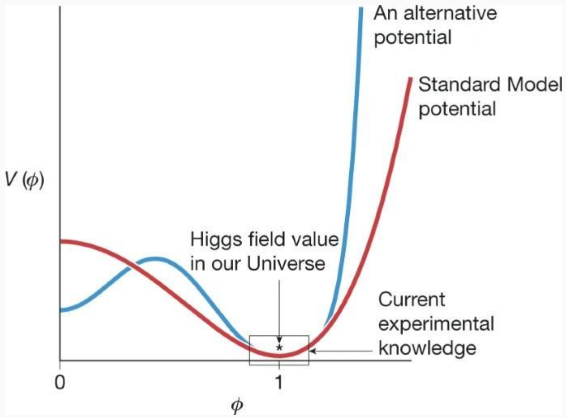
Yibo Zhong (Purdue University)
on behalf of HH$bb\gamma\gamma$ analysis team

IP2I Lyon, PKU, Northwestern, Cornell, FNAL, Caltech,Purdue, USTC, RWTH, NCU, BU, UCSB, UCSD, IHEP CAS, CUA

Motivation

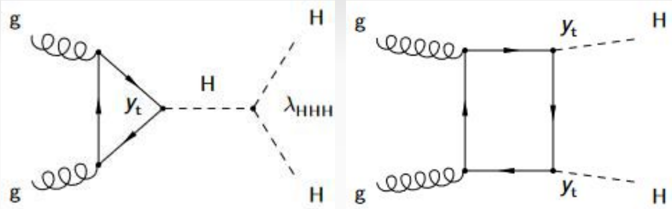
- Higgs trilinear self-coupling λ provides information for the shape of the potential. At LHC, the constant λ is directly accessible via Higgs pair (HH) production.
- $H \to bb$ decays provides a high branching ratio, while $H \to \gamma\gamma$ ensures a high purity.
- BSM can also give HH signatures and we will investigate those signatures, also using boosted categories

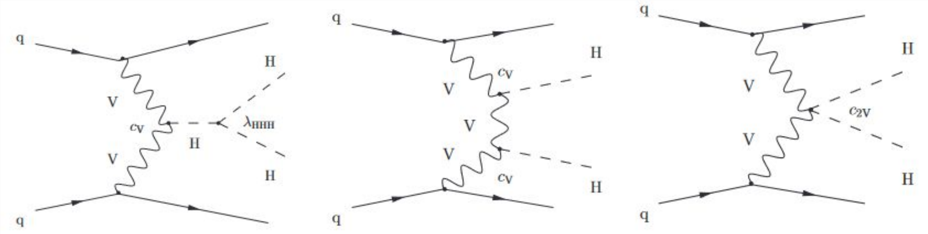
$$V(h)=\lambda_3 v\, h^3 + \frac{1}{4}\lambda_4 h^4$$



Search Strategy

- There are two main production modes, ggF(34.13 fb) and VBF(1.89 fb):

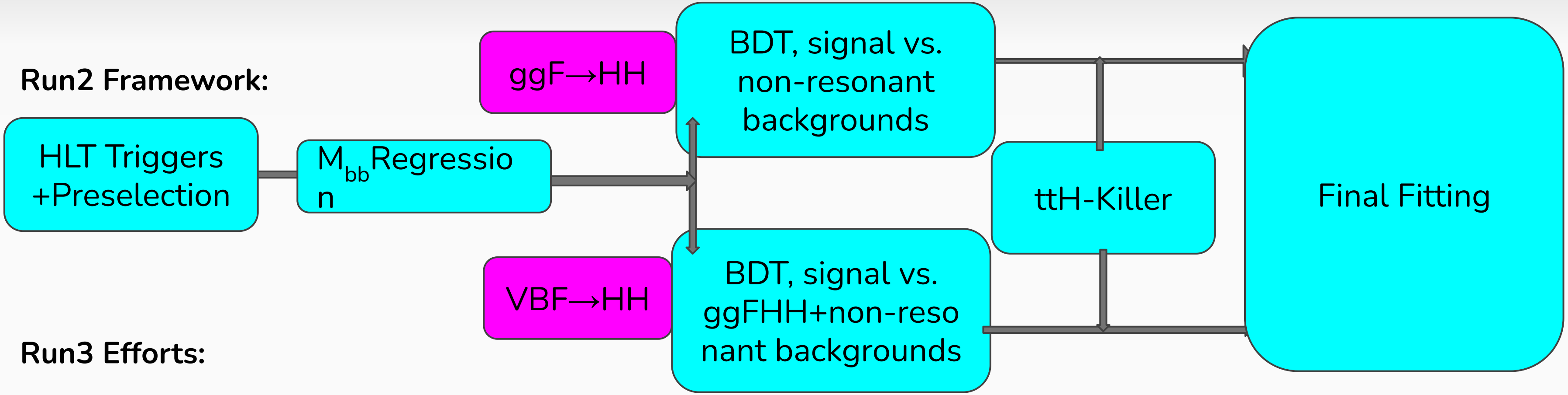




- Nonresonant Background: γγ+jets, γ+fake and fake+fake with one or two jets misidentified as photons.
- Resonant Background: ggH, VBFH, VH, ttH and bbH (MC not ready yet).
- Search for H→γγ candidate and H→bb candidate around Higgs mass ~125 GeV.

Analysis Topics

Run2 Framework:



Run3 Efforts:

HiggsDNA Framework

- Boosted Category
- PaticleNet Ak4 and Ak8 Taggers for b-jets identifications
- Diphoton MVA HLT

- M_{bb} Regression
- Jets pairing
- BDT→New ML based MVA
- Data-driven fake photon bkg estimation

New ttH-Killer

Final Fitting workflow

Single-H DNN
Multi-Class NN for (ggF, VBF, ttH, NonResonant)

BSM efforts (SMEFT, X→HY)

Pre-Selections (Higgs-DNA)

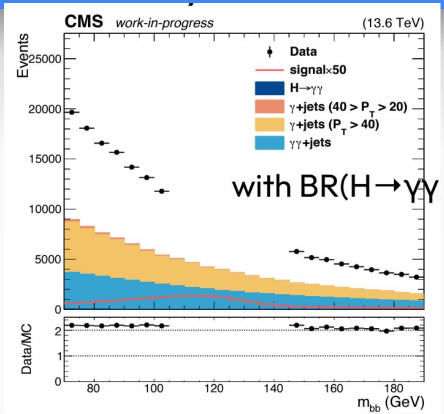
Trigger :
HLT_Diphoton30_22_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass90
HLT_Diphoton30_22_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass95
MET Filters Applied.

Photon Selections:
pT_1 >35GeV, pT_2>25GeV
(isScEtaEB | isScEtaEE) & $\|\eta\|$ < 2.5
$R9$ > 0.8 or ChI < 20 or ChI/ET < 0.3
H/E < 0.08 & MVAID> −0.9
ElecVeto
isEB_high_r9 | isEB_low_r9 | isEE_high_r9 | isEE_low_r9
pfRelIso03_all_quadratic * pT <10
pT_1 > M(γγ)/3 & pT_2 > M(γγ)/4
EndCap Leak Exclusion for 22_EE

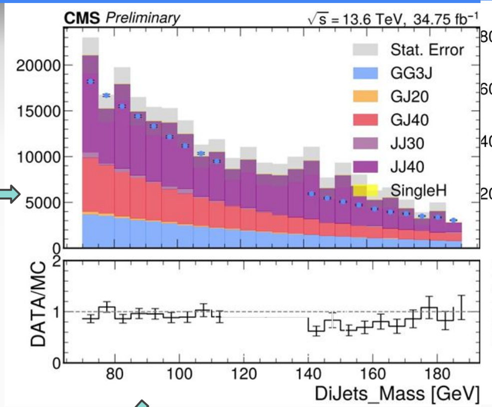
B-jet Selections:
Pt>20 GeV & $\|\eta\|$ < 2.5
ΔR(j,γ) > 0.4
70 GeV <DiJets_Mass<190 GeV
2 Jets with highest PNet Score

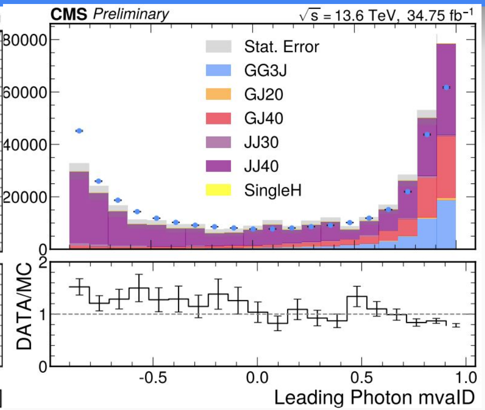
- **Following the recommendation by H→γγ Group**
- **Also studying further ways to improve veto lepton ID**

DATA/MC after Preselection



Add QCD Samples (fake-fake)



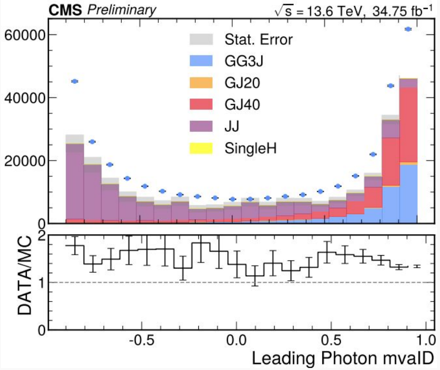


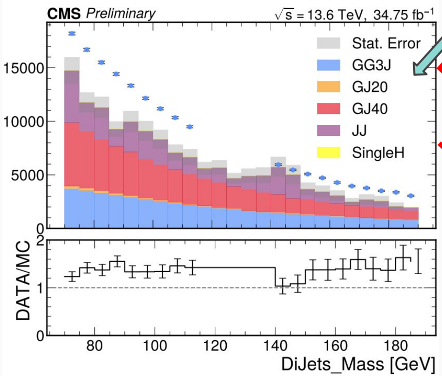
DATA/MC doesn't match well if use only γ+jet, γγ+jet.

Need better agreement of variables for MVA training

Remove real photons from QCD samples

However, QCD (fake-fake) samples has double counting of real photons.

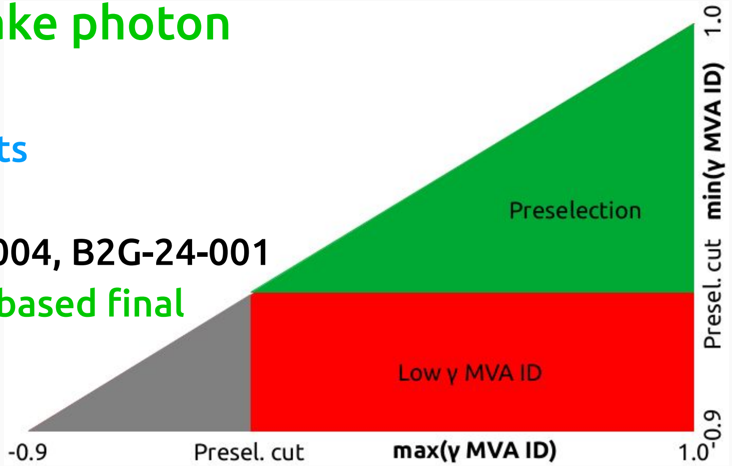




- Data/MC agreement is much better now
- Remainder difference will be fixed by data-driven bkg estimation

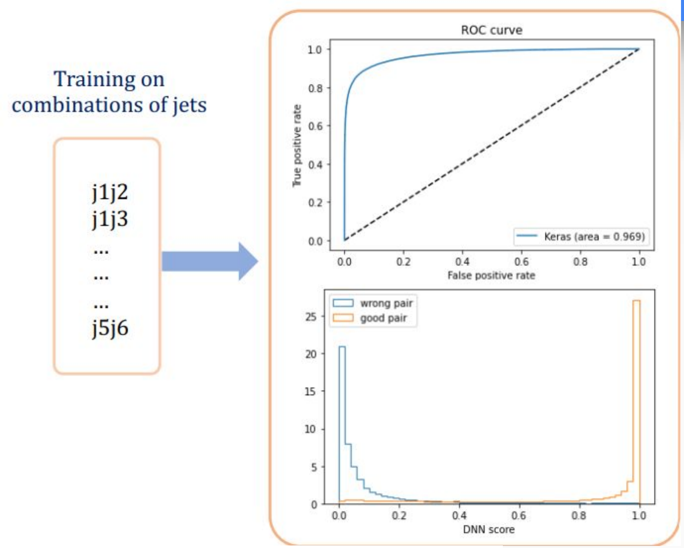
Photon Fakes

- Events with fake photons are dominant at the pre-selection level (QCD, γ+jets):
  - Are poorly modeled in simulation.
  - Have low statistics in the phase space relevant to the γγbb final state.
- The Run-2 non-resonant bbgg analysis was using a 2D mass peak fit to get the fake background normalization
- Investigating a data-driven method to predict fake photon background - more details here
  - Use data events from sideband (red) as a proxy for events with fake photons in preselection region (green).
  - The method developed and used in HIG-19-013, HIG-23-004, B2G-24-001
  - Can improve bkg description and statistics for the MVA-based final selection

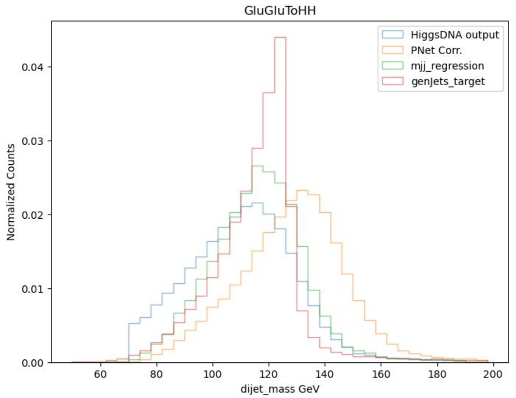


Jets for ggF Category

- Important to select correct b-jet pair to reconstruct M_{bb}
- Default HiggsDNA selection uses highest b-tag scores
- Deploy DNN to improve the jet pairing for ggF→HH:
  - Training on all combinations of 6 jets and uses additional variables (in addition to b-tag info)
  - DNN improves pair choice efficiency by 3%

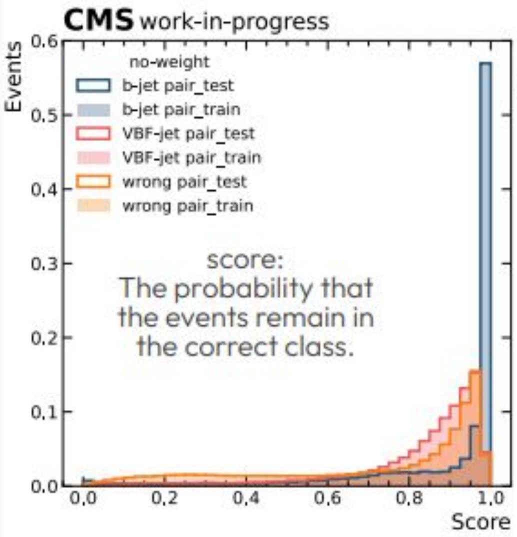


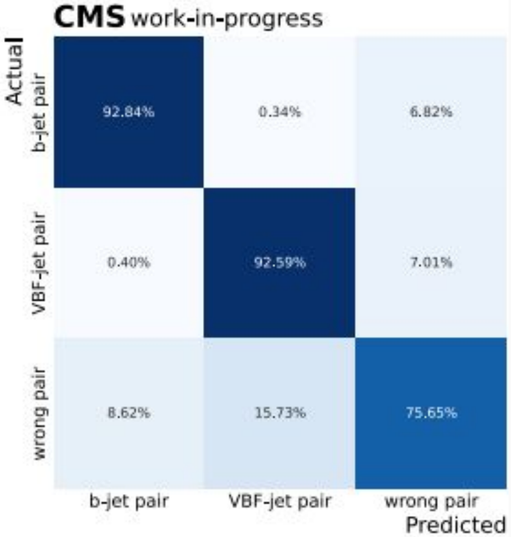
- Mbb Regression: potentially improve Mbb resolution after pre-selection using a DNN
  - Training on GluGluHH signal samples
  - Uses jet kinematics, MET, energy-mass variables



Jet-Pairing for VBF Category

- Need to select correct b-jet pair, and 2 vbf jets in final states.
- A preliminary effort to deploy XGB for [VBF→HH](#) :
  - Separate jet pairs in VBFHH sample into 3 classes: b-jet pair, VBF-jet pair, wrong pair.
  - Preliminary result: 92% selection efficiency for b-jet and VBF-jet pairs





ttH Killer

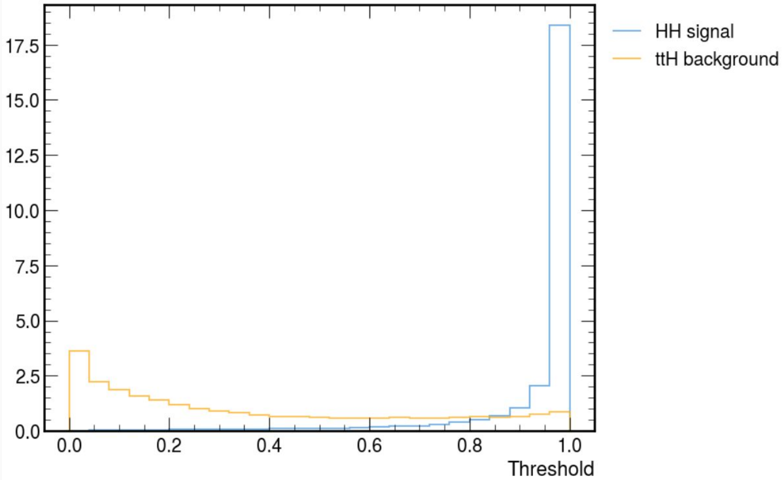
- ttH is a dominant background after signal region selections:
  - At 13.6 TeV and m ~ 125 GeV : σ_{ttH}~579 fb and σ_{HH}~34 fb.
  - b-jet kinematics are similar between tt→WbWb and H→bb
  - ttH is an order of magnitude larger than the signal even after signal region selections
  - A dedicated ttH killer is necessary

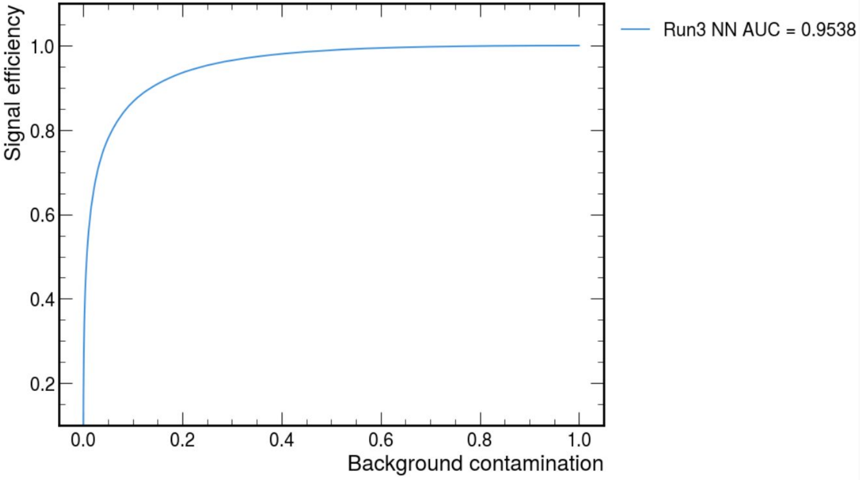
	ggF→HH →bbgg	GG3J	GJ	ttH	GluGlu H	VH	VBFH
Bare Yield	3.1	3083944	40364044	44.8	4115	189.3	321.4
Pre-Selection Yield	1.1	56743.5	69543.2	19.8	216.8	32.3	15.7
Preselection +H Mass +B_tag Yields	$\boxed{0.59}$	2531	2800	$\boxed{7.39}$	14	2.88	1.28

H Mass window: 100 < M(jj),M(γγ) < 150, BTag- score: PNet_btag$_{1,2}$>looseWP

ttH Killer

- First ttH killer retrained with Run 3 samples, with the up to date b-tagging etc.
- Model Performance:



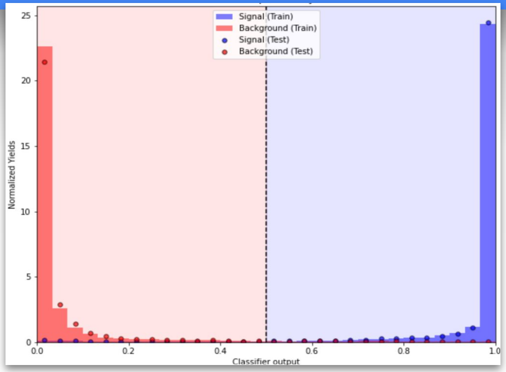


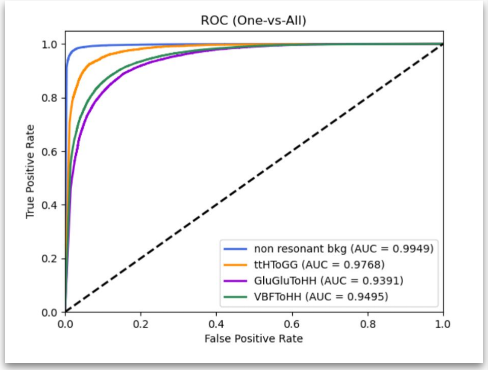
- Evaluated and established similar performance as Run 2
- Further improvements are being investigated

Classification NNs

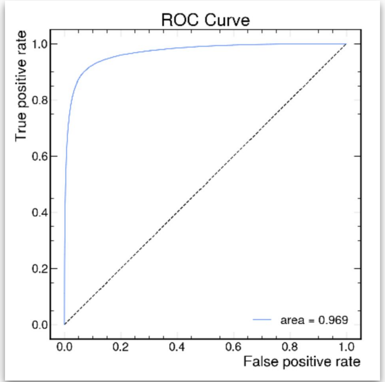
We are developing different classifiers for signal separation from non-resonant and resonant backgrounds

- DNN on ggFHH vs non-resonant backgrounds:
  - To be later developed also for X→YH signals
- Efforts on Multi-Class Classification/binary DNN for single-H bkg:
  - Samples:
    Non-resonant background (γ+jet, γγ+jet)
    single-H bkg: ttH /ggH/VBFH
    Sig: ggHH, VBFHH
  - Started with all the available higher level variables.
  - Will also include BSM signals later





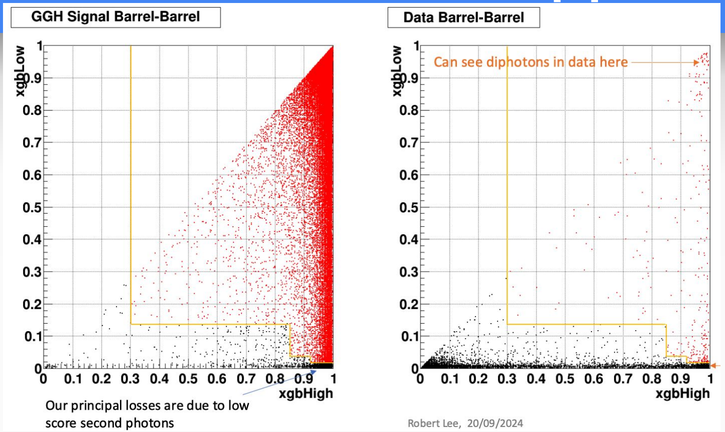
Multi-class DNN

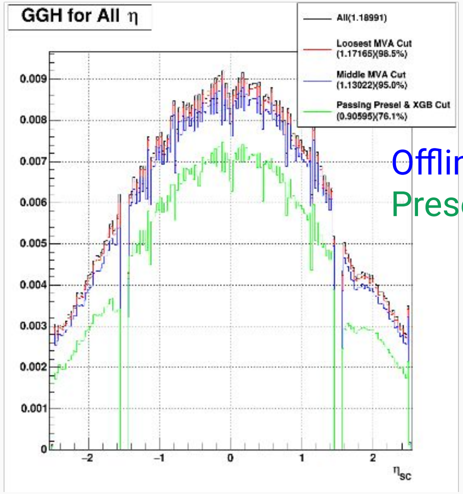


Binary DNN

MVA based HLT and offline selections for H→γγ

- An offline MVA (XGboost) was trained to increase H→γγ selection efficiency *by removing the preselections and relaxing the p_T/M cut on the photons* while keeping the same rate of background
  - Showed improvements of ~30% on reconstructed signal: see talk in Hgg meeting on [Nov 6, 2023]
- An MVA-based HLT was deployed which is very similar to offline MVA-based selection
  - see improved signal efficiency
  - decreases the rate of selected data by a factor of 0.6.
- The MVA based HLT and offline MVA both improve kinematic coverage of H→γγ signal
  - The improvements to HH→bbγγ signal efficiency from this HLT and offline MVA are currently being studied

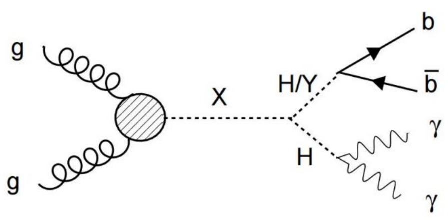




Offline MVA

Preselection

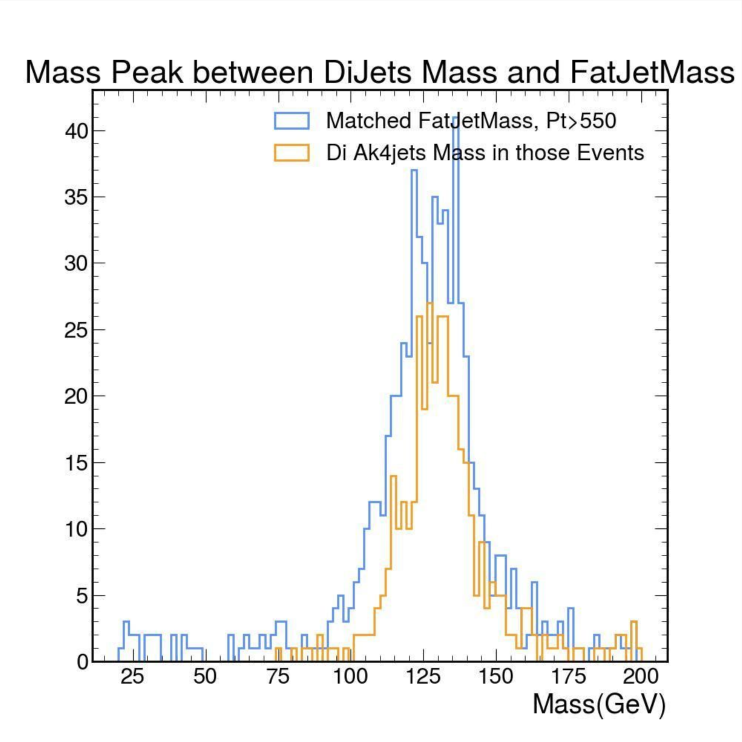
BSM Efforts



- $X \to HH$:
  - Focus on the interference between resonant and nonresonant HH production.
    - Ongoing: Model construction and validation to tackle the interference (more information here)
    - Private samples generated
  - Plan for first results:
    - Use the signal extraction strategy and MVA developed for the SM signal
    - After full analysis workflow is set and modeling of interferences validated, we will extend the MVA training to consider the resonance mass
- $X \to HY$:
  - Add BSM X→YH signal to main classifier for different (X, Y) mass ranges.
- EFT interpretations :
  - Run 2 analysis only had HEFT interpretations; will add both HEFT and SMEFT now
  - SMEFT gridpacks being currently tested - see talk in HH meeting on 9 Sep 2024

Add Boosted Category

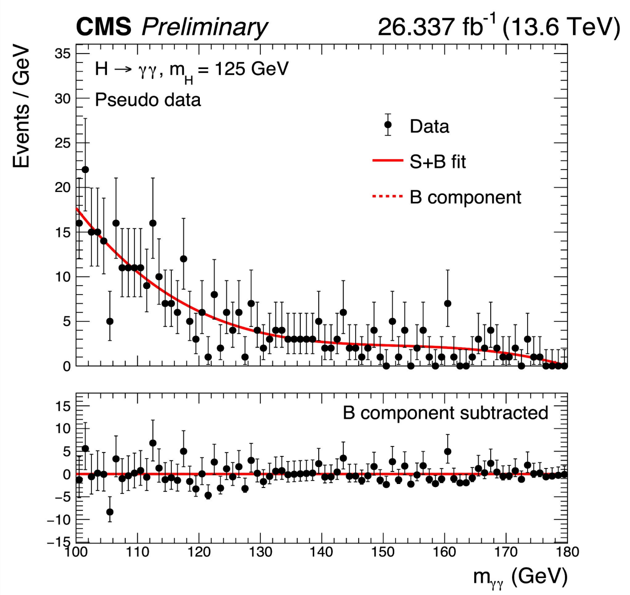
- Starting to evaluate sensitivity from boosted channel using SM HH signal
- Studying interplay between boosted and resolved categories
- Boosted category is particularly interesting for high mass BSM



Fitting

- Integration of flashggFinalFits in progress
- FinalFit work flow :
  - Convert parquet file to RooWorkSpace
  - Signal modeling (Including fTest, photon systematics computation ,signalFits ...)
  - Background modeling (Including fTest, bkg ploting...)
  - Datacard creation and Limit computation
- Plan to add more finalFit tasks, like including Single H background, into the workflow.
- Investigating different fitting strategies, such as the possibility of 2D fitting with $M_{\gamma\gamma}$ and M_{bb}.

PostFit S+B with Pseudo data



Summary

- Making good progress with Run 3 HHbbγγ analysis
  - Establishing baseline analysis with Run 3 data: Data/MC comparison, ttH killer, new jet-pairing methods for both gg and VBS categories.
  - Actively pursuing potential improvements: better offline and HLT preselection with γγ MVA, multi-class classification, study interference, adding boosted categories etc.
  - Establishing analysis framework with HiggsDNA and flashggFinalFits for statistical interpretation.

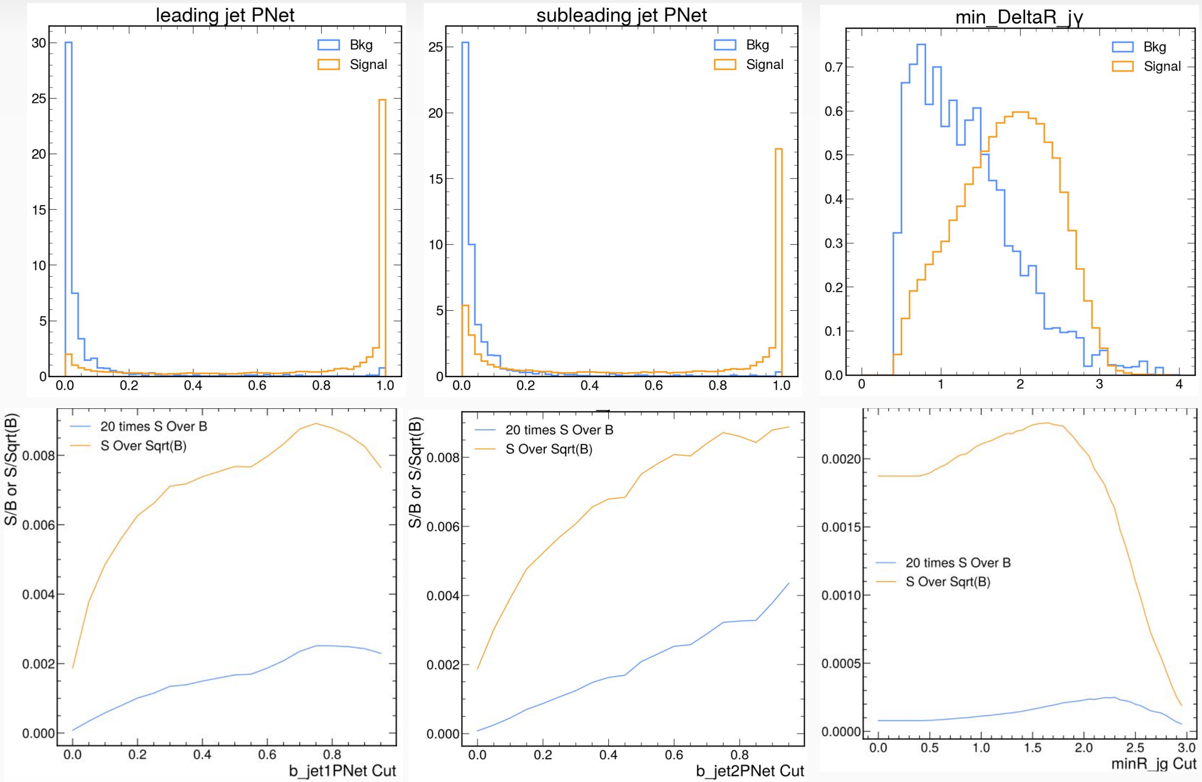
BackUp

Variables Distribution and Significance Scan

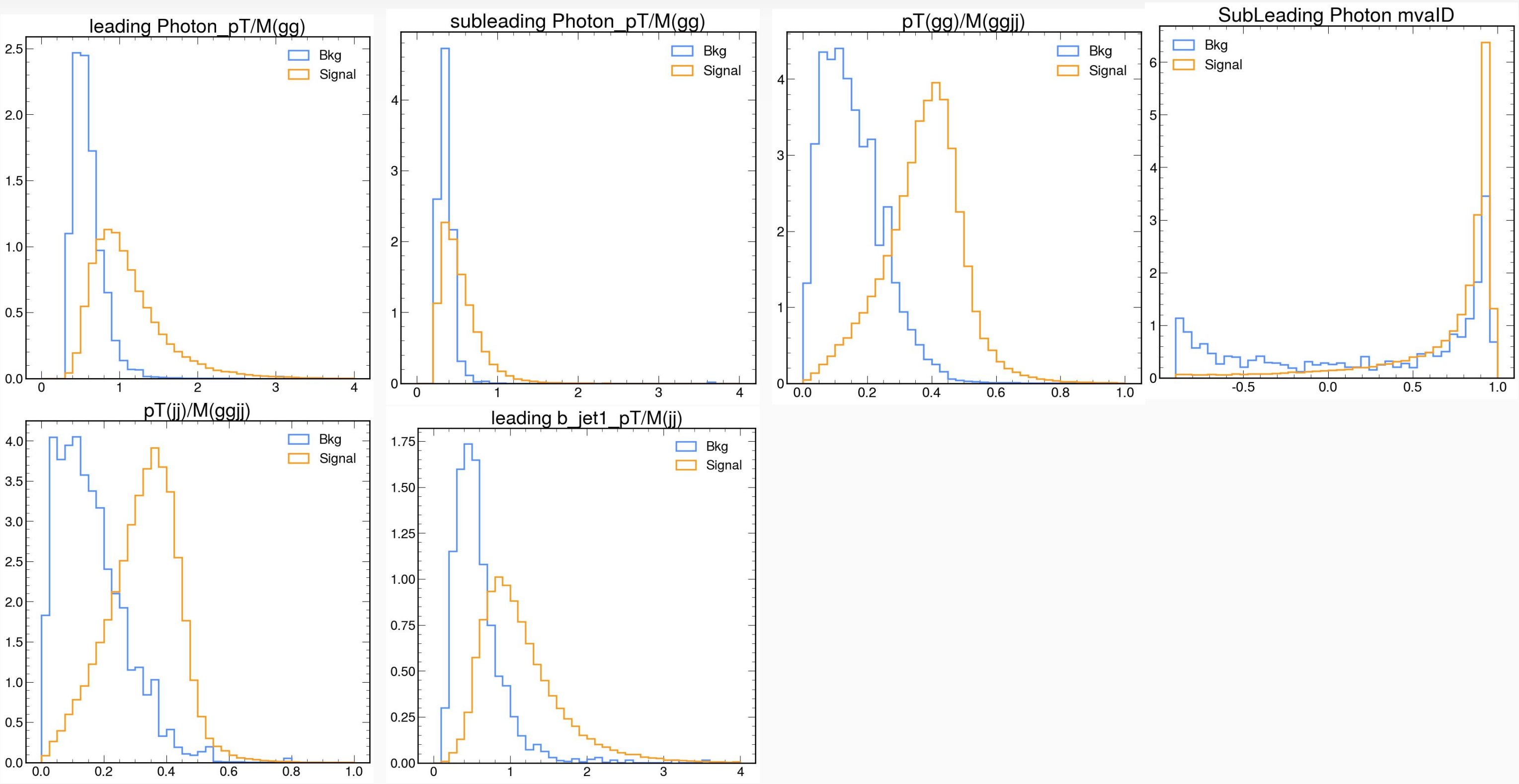
- Weighted sum of all bkg samples are included.
- Fully MC based study

S : Number of Signal greater than the cut

B: Number of Bkg greater than the cut



Discriminating Variables Distribution

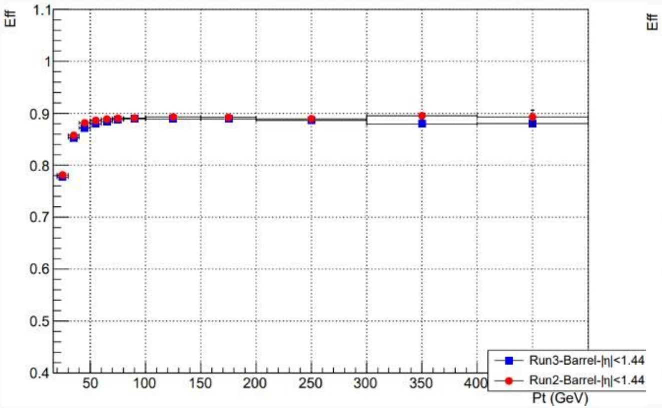


Single Photon Efficiency

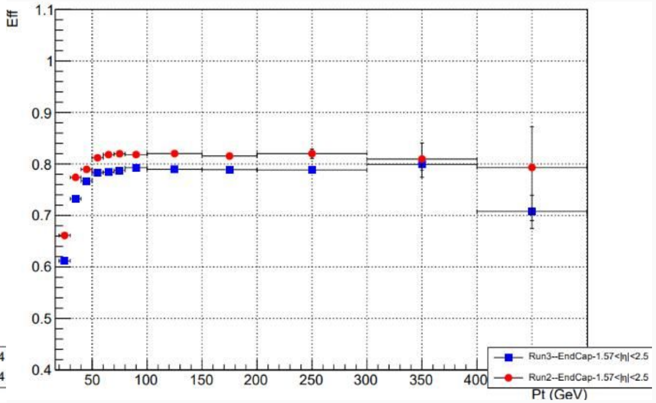
- In the Barrel, Run 3 and Run 2 have similar efficiencies and in genera equal to 90%.
- In the Endcap, Run3 drops due to Z+ Ecal leakage in 2022EE.
- Checked comparable photon efficency with Run3 and Run2 samples.

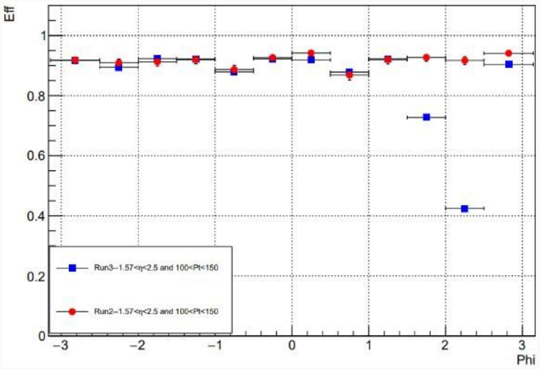
e.g. Efficiency as a function of ϕ, in pT in (100, 150) GeV, η in (1.57,2.5)

Barrel



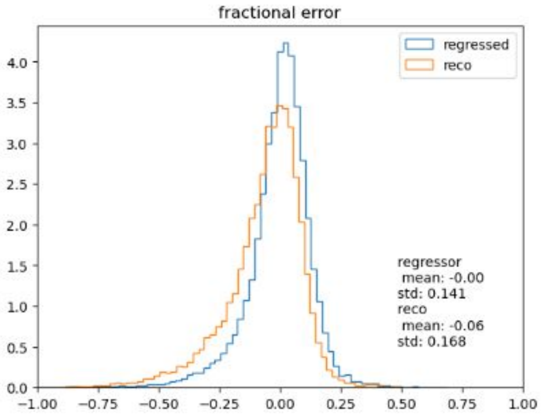
Endcap

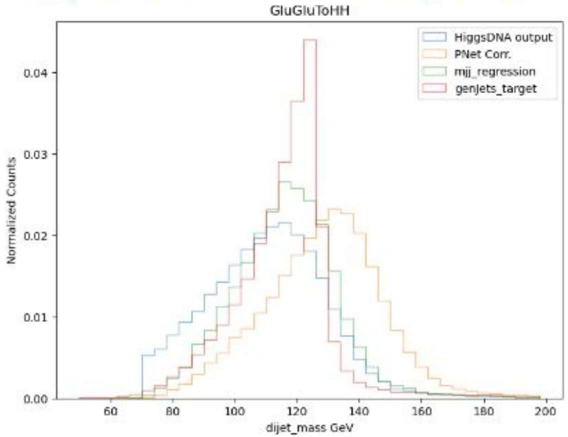




Mbb Regression

- Regress dijet after preselection to boost mbb resolution
  - Training on GluGluHH signal samples
- DNN with 3 hidden layers, 200 neurons
  - Hyper-parameters selected via a grid search
- Input variables:
  - jet kinematics, MET
  - ParticleNet pT correction
  - Energy mass variables divided by reco dijet mass, to remove effect of H mass scale
- Targets a correction term, to be applied to the reco dijet mass.



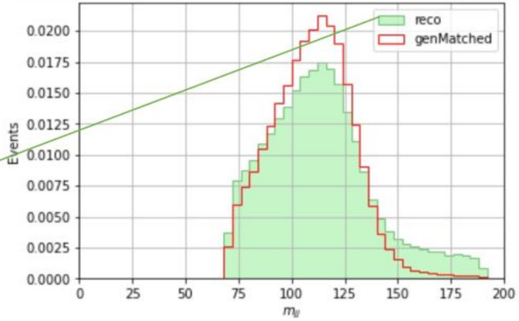


Pair Matching

Selection efficiency

Selection :

- $P_T^{jets} > 20$ GeV
- $100 < m_{\gamma\gamma} < 180\ GeV$
- $70 < m_{jj} < 190\ GeV$
- b-jet candidates taken as the **highest sum of b-tagging scores (PNet)**



Ratios wrt to all the reconstructed events

	lead	sublead	pair
Ratio of gen-matched candidates (%)	93	73	68

77 % of reconstructed events contain the « True » pair

	lead	sublead	pair
Ratio of gen-matched candidates (%)	97	92	88

Ratios wrt to the reconstructed events having two b-jets matched in the jet collection

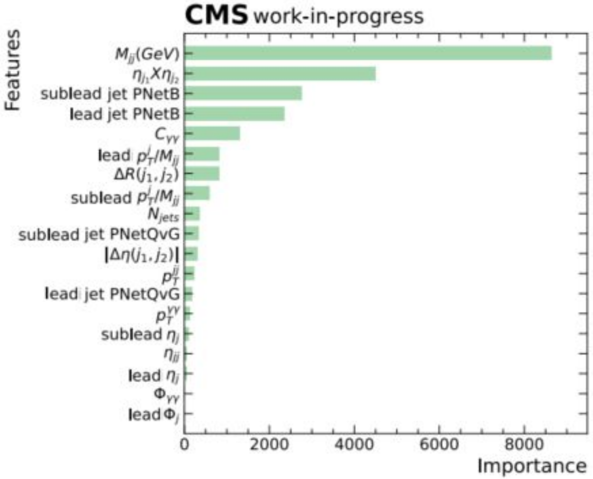
Jet pairing for VBFHH category

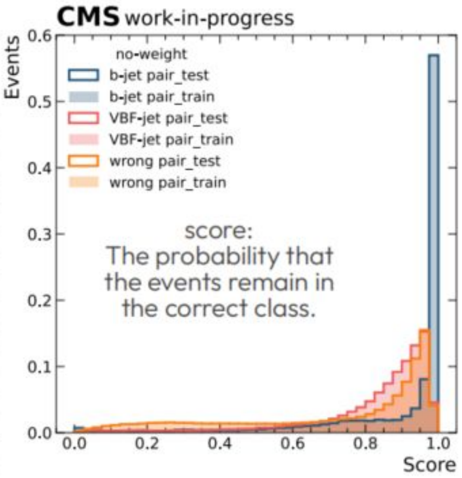
- Distinguish the correct pairing for b-jets and VBF jets by XGBoost/DNN.
  - Separate VBFHH sample into 3 classes.
  - Input features: kinematics of jet pair, b-tag score, QvG, Centrality $C_{\gamma\gamma}$.

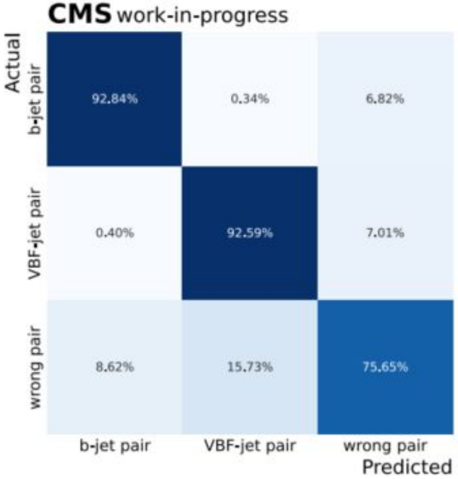
② VBFHH → b-jet pair | VBF-jet pair | wrong pair

→ important access to C_{2V} → match to genJet → match to LHE outgoing parton

- Pair combinations: (jet 1, jet 2-6), will add all combinations soon.
- 92% b-jet and VBF-jet pair can be predicted correctly.





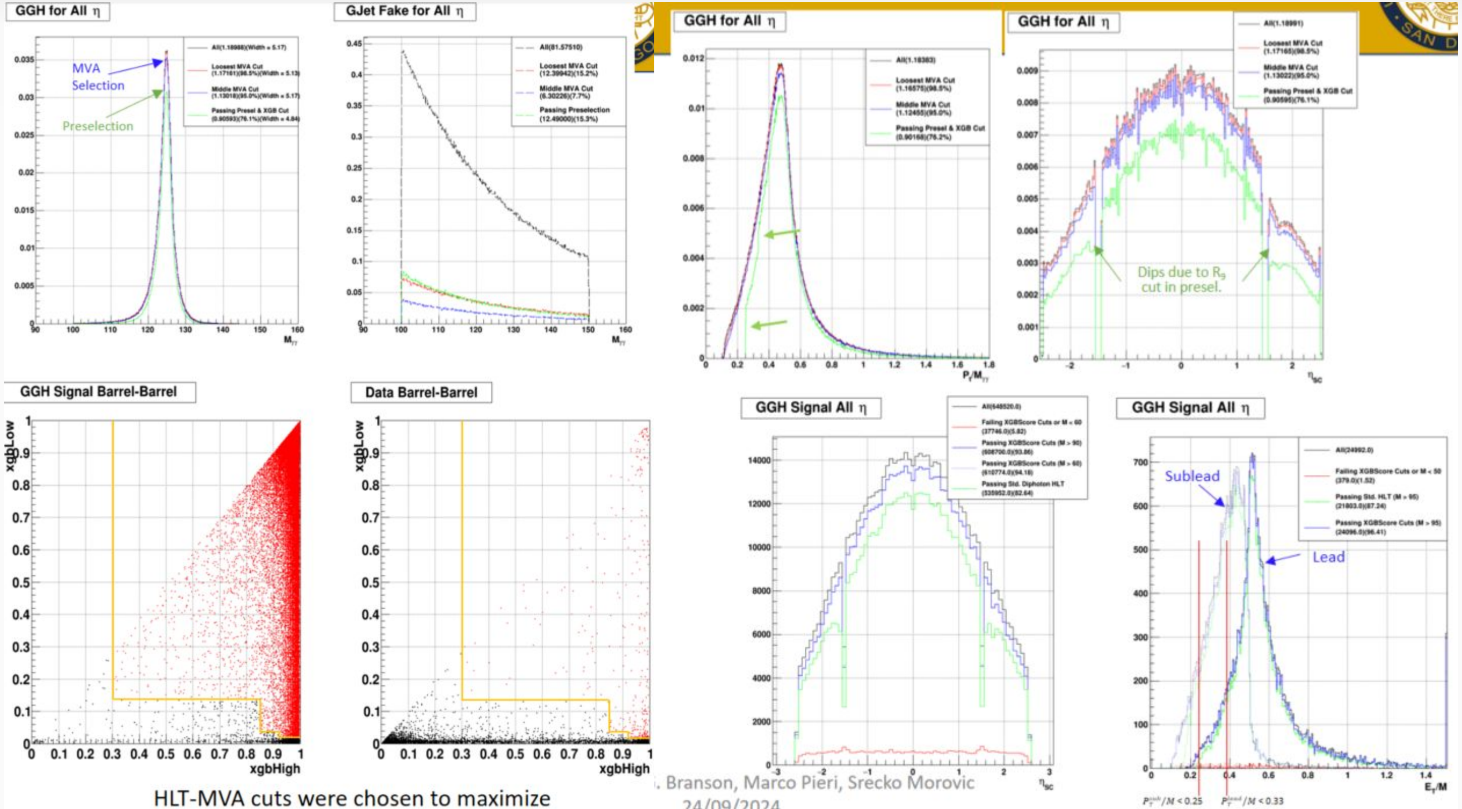


Multi-Classification

Input features

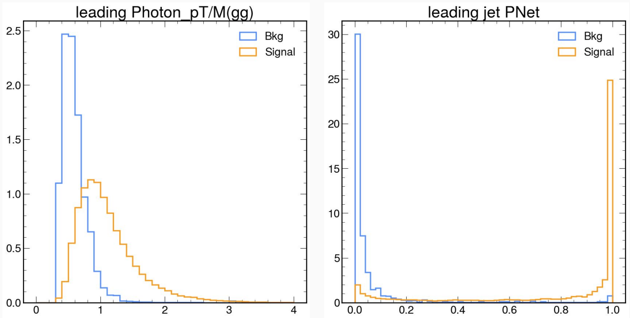
- $p_T^{\gamma\gamma}/m^{\gamma\gamma bb}, p_T^{bb}/m^{\gamma\gamma bb}$
- Photons and b-tagged jets (leading and sub-leading):
  - $p_T^{\gamma}/m^{\gamma\gamma}, p_T^{b}/m^{bb}, \eta, \phi$, MVA ID or b-tagging score
- $cos(\theta^*_{CS}), cos(\theta^*_{bb}), cos(\theta^*_{\gamma\gamma}), \Delta R(\gamma_{1,2}, b_{1,2})$
- p_T, η, ϕ, b-tagging score or MVA ID for up to 6 jets, 4 leptons (order of p_T)
- $p_T^{VBF}/m_{jj}^{VBF}, \eta, \phi$, b-tagging score, quark-vs-gloun score, $min\Delta R(b, VBF), min\Delta R(\gamma, VBF)$ and C_H for VBF-tagged jets
- MET p_T and ϕ, $\Delta\phi(MET, b_1)$ $\Delta\phi(MET, b_2)$
- χ^2_{top} for events with at least 2 and 4 additional jets besides the b-tagged jets
- Number of jets and leptons
- $\tilde{M}_X = m_{\gamma\gamma bb} - (m_{bb} - m_H) - (m_{\gamma\gamma} - m_H)$ as a condition

DiPhoton MVA



ggF: Cut Based Study

- An internal cut-based selection allows to better understand the data, and provide a reference point for future more complex MVA studies.
- In a few sensitive CATs of Run2, the total signal yield is ~1. Use Run2 as a criteria, and normalize to year 2022, should get ~ 0.27
- Do a manual cuts of some variables and a grid-search over the phase space of other variables under condition (max S/√B | S>=0.27):
- S/√B 0.175, Sig 0.270, Bkg 2.374
- Upper Limit, scale to Run2 Lumi (137 fb-1) : 15.13



e.g. Distributions of some variablles