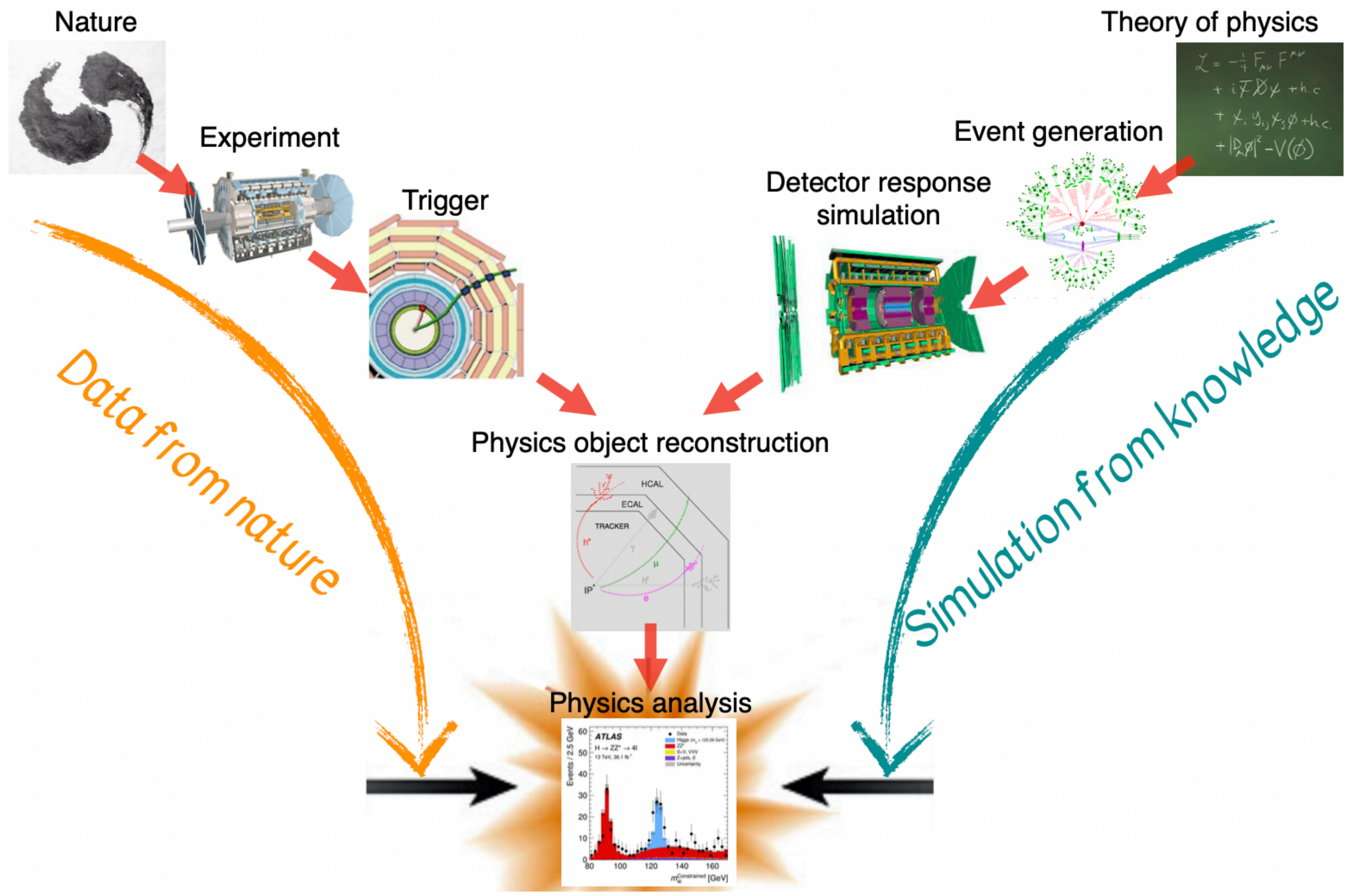
AI-based event classification with CMS

Chen Zhou (Peking University)
On behalf of CMS Collaboration

Boost Conference
2024.07.29

Introduction

- One of the major objectives of the experimental programs at the LHC is the discovery of new physics
- Machine Learning: “application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed”
  - It has become one of the most powerful techniques for High Energy Physics (HEP) data analysis
  - **It greatly enhances our ability of identifying signal from background: important for discovery of new physics**



Back to Higgs discovery era

H→γγ analysis (2012)

Select events with **two photons**

➔Train **Diphoton MVA** using signal and background MC

- Input variables: kinematics and (BDT-based) photon ID MVA of each photon, (BDT-based) vertex probability, etc.

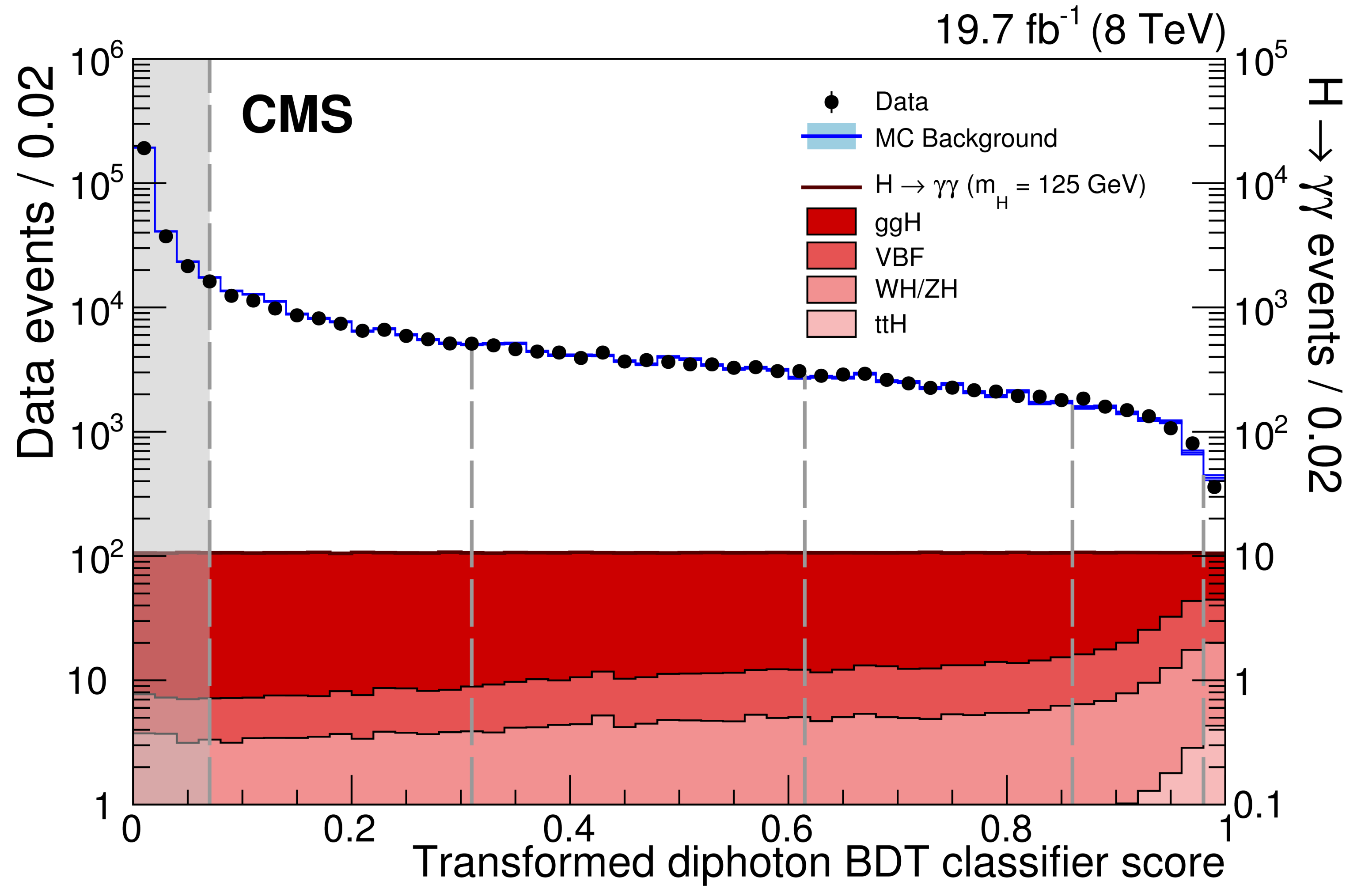
➔ Separate events to **categories** based on BDT score (which is to the first order independent of diphoton mass)

➔Fit **diphoton mass** over all categories

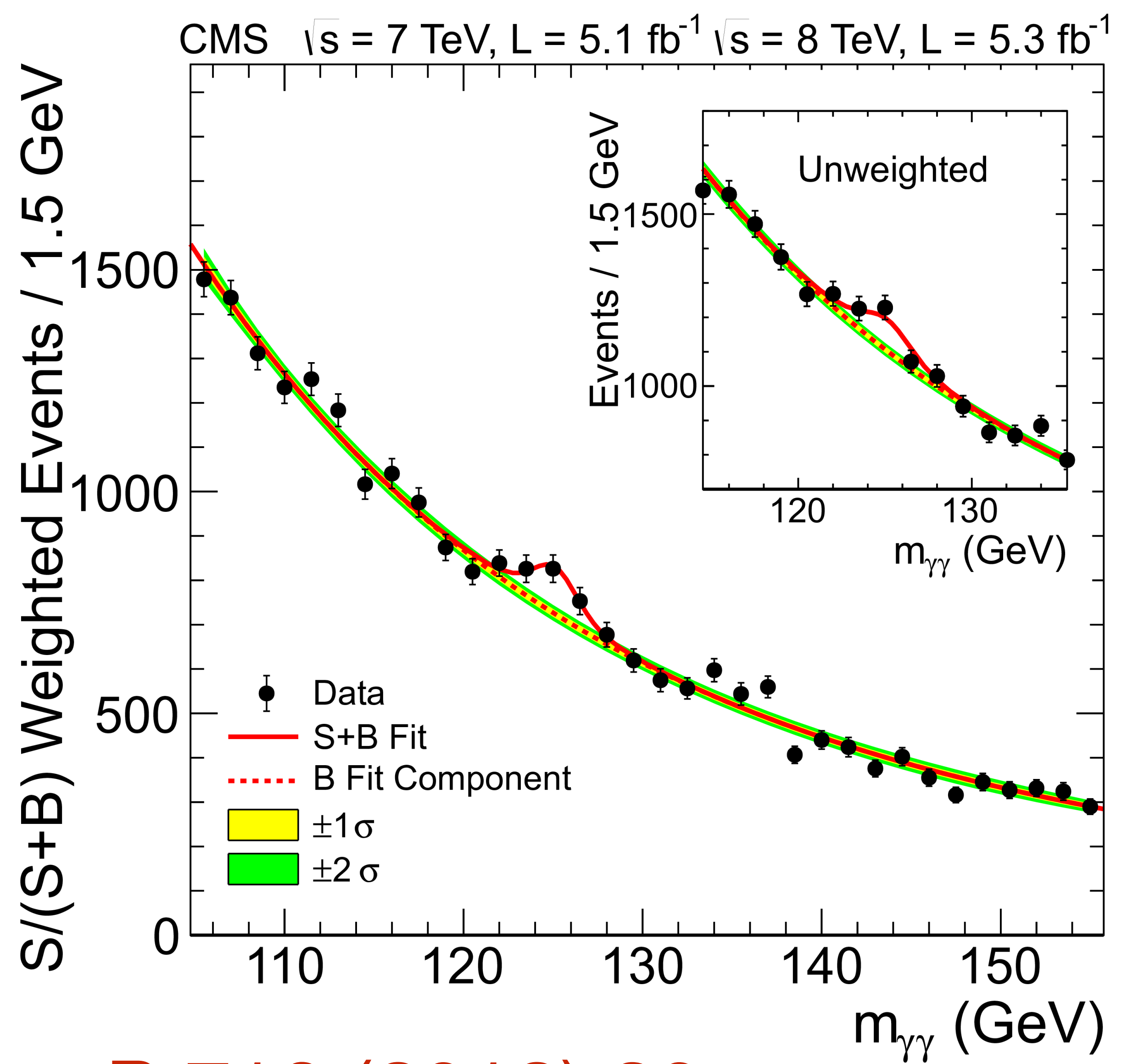
- Signature: a narrow resonance above a smooth background (QCD γγ production, etc.)

➔Measure signal strength, etc.

Better than cut-based analysis by 15%



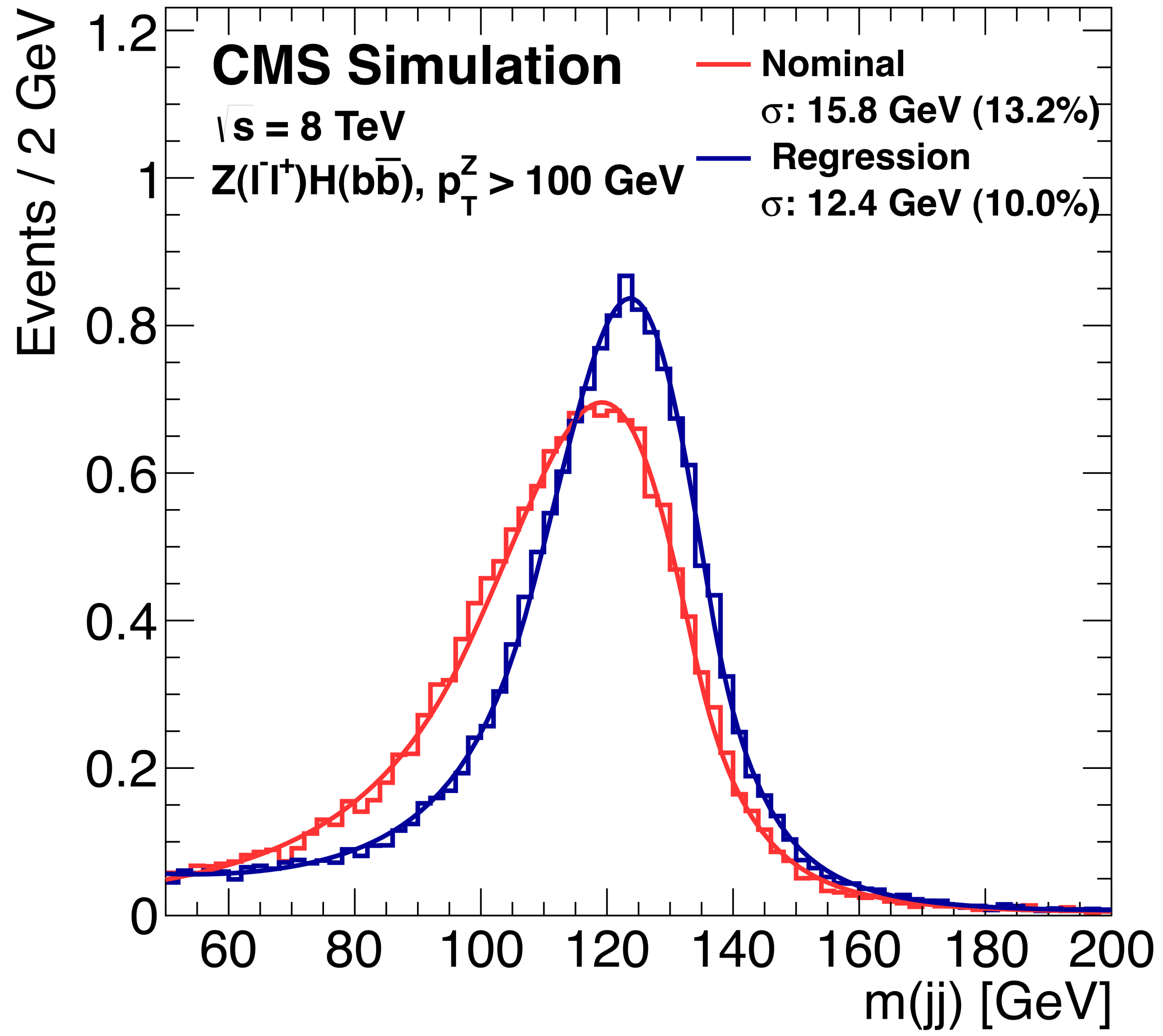
Eur. Phys. J. C 74 (2014) 3076



Phys. Lett. B 716 (2012) 30

H→b$\bar{b}$ analysis (2012)

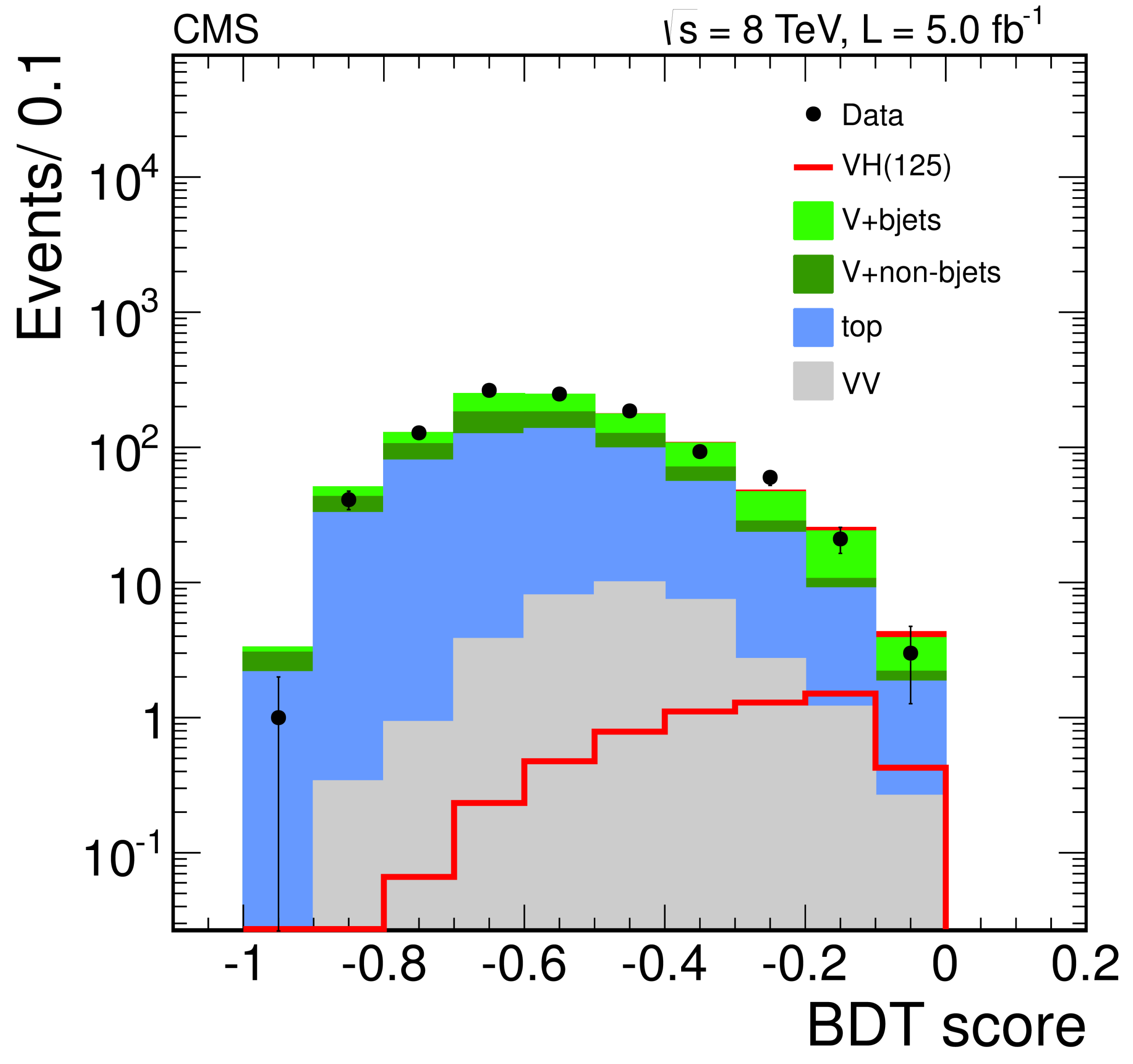
- **Large branch ratio (~58%)**
- Huge background, tackled by requiring associated particles and machine learning



Phys. Rev. D 89 (2014) 012003

VH→Vb$\bar{b}$

- Reconstruct Higgs as two small-radius b-tagged jets
- Tag leptonically decaying W/Z boson
- Main bkg: V+heavy flavor, t$\bar{t}$
- Train BDTs using kinematics of V and H candidates (e.g. $m_{b\bar{b}}$ reconstructed by regression)
- Fit the shape of the BDT output distribution

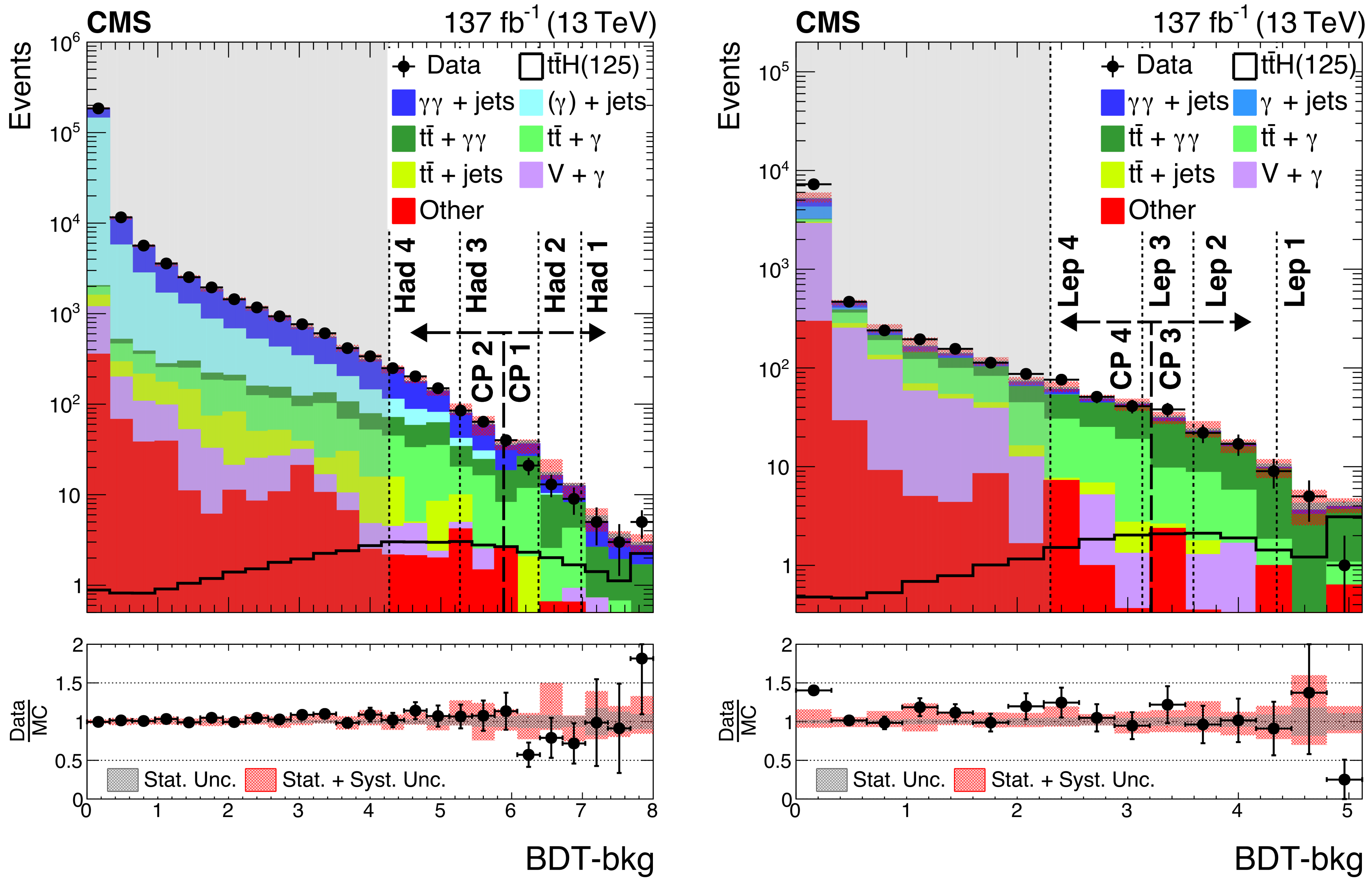


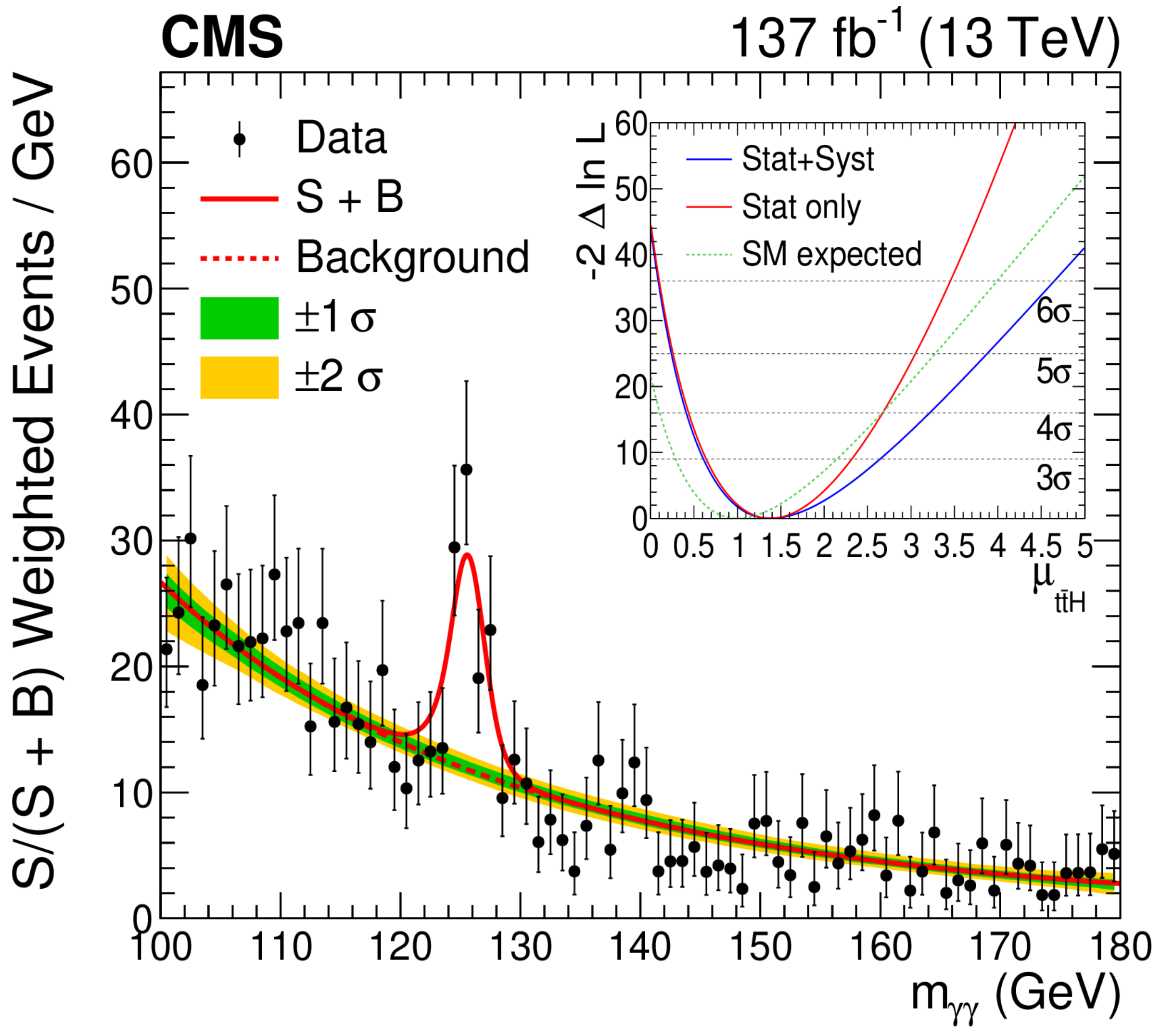
Phys. Lett. B 716 (2012) 30

AI-based event classification in Higgs measurement era

ttH (H→γγ) analysis (2020)

- **The signal strength analysis** trains a dedicated BDT ("BDT-bkg") with XGBoost to distinguish between ttH and background
- Inputs include kinematic variables of photons, jets & leptons, as well as outputs of other ML algorithms:
  - top quark tagger BDT to identify events with top quarks decaying into three jets
  - long short-term memory based DNNs exploiting low-level information including full four-vectors of each jet & lepton and jet flavor scores
- **First published single-channel observation of the ttH process, with a significance of 6.3σ**

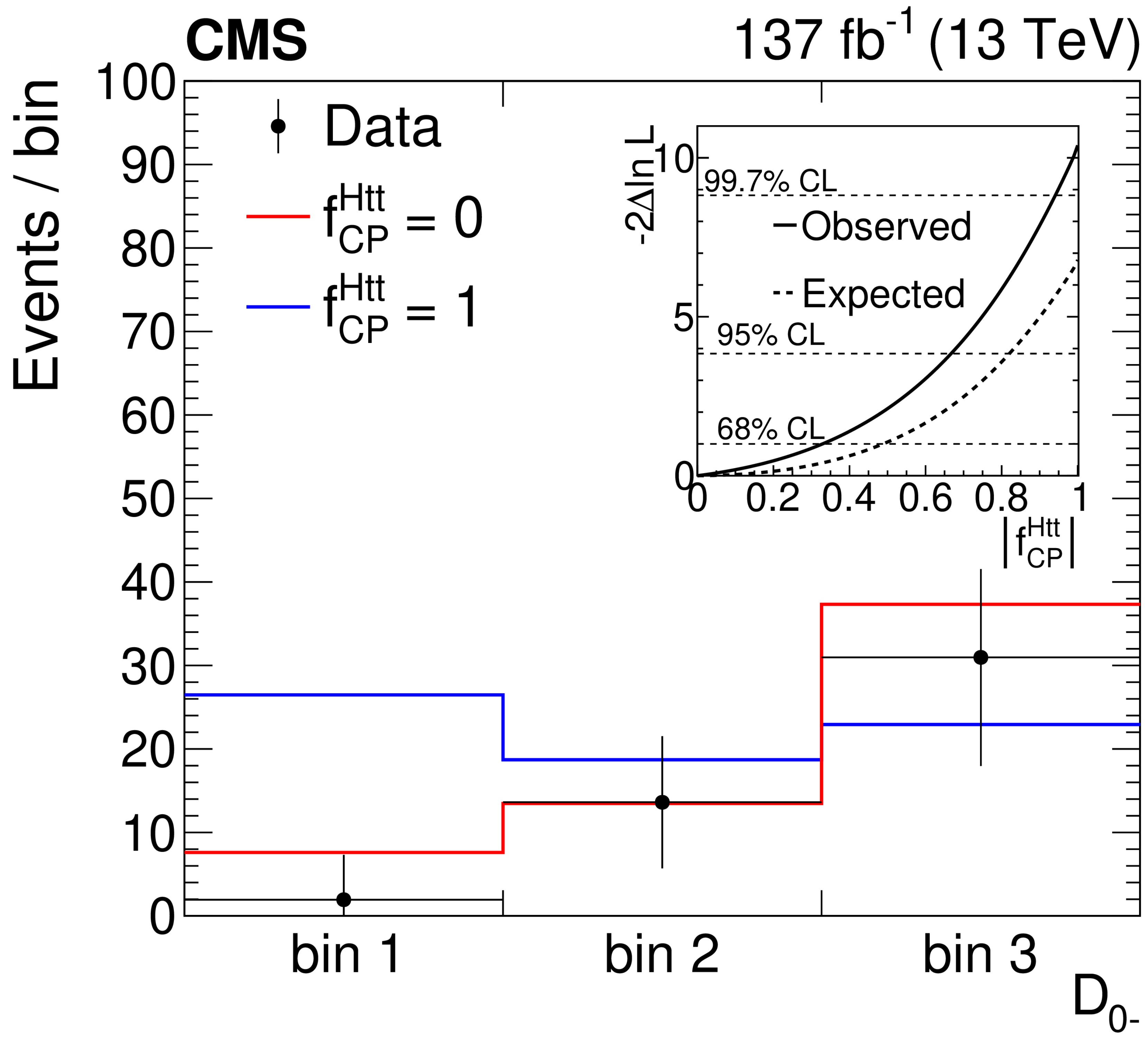




Phys. Rev. Lett. 125, 061801 (2020)

ttH (H→γγ) analysis (2020)

- **The CP analysis** trains a BDT to distinguish CP-even and CP-odd contributions
  - Simulation shows that BDT-CP discriminant has negligible correlation with BDT-bkg discriminant
- Events selected for the signal strength measurements are split into 12 categories, leptonic or hadronic, two BDT-bkg categories, and three BDT-CP bins
- **First measurement of the CP structure of the Htt coupling using the H → γγ channel, disfavor the pure CP-odd model of the Htt coupling at 3.2σ**

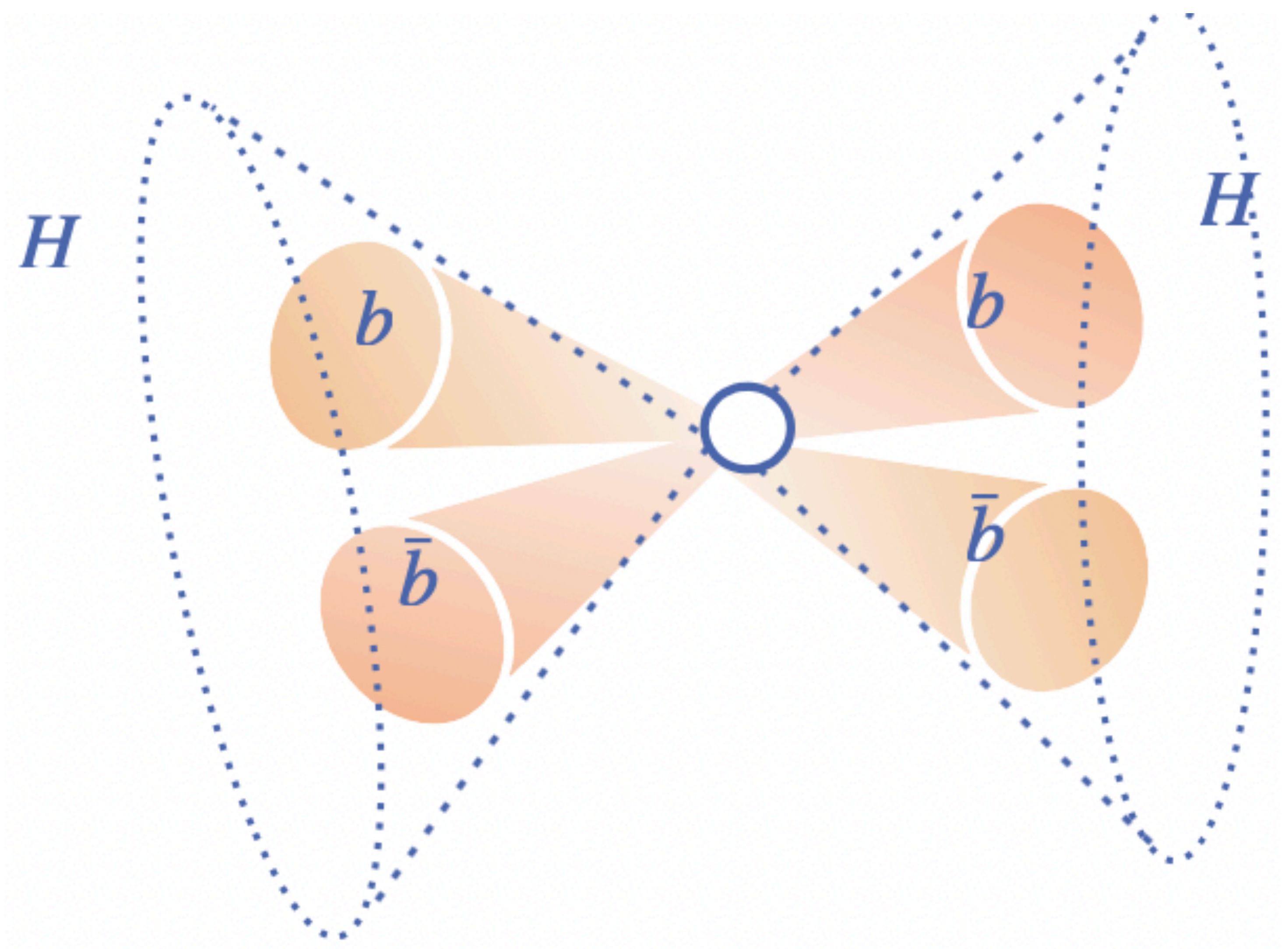


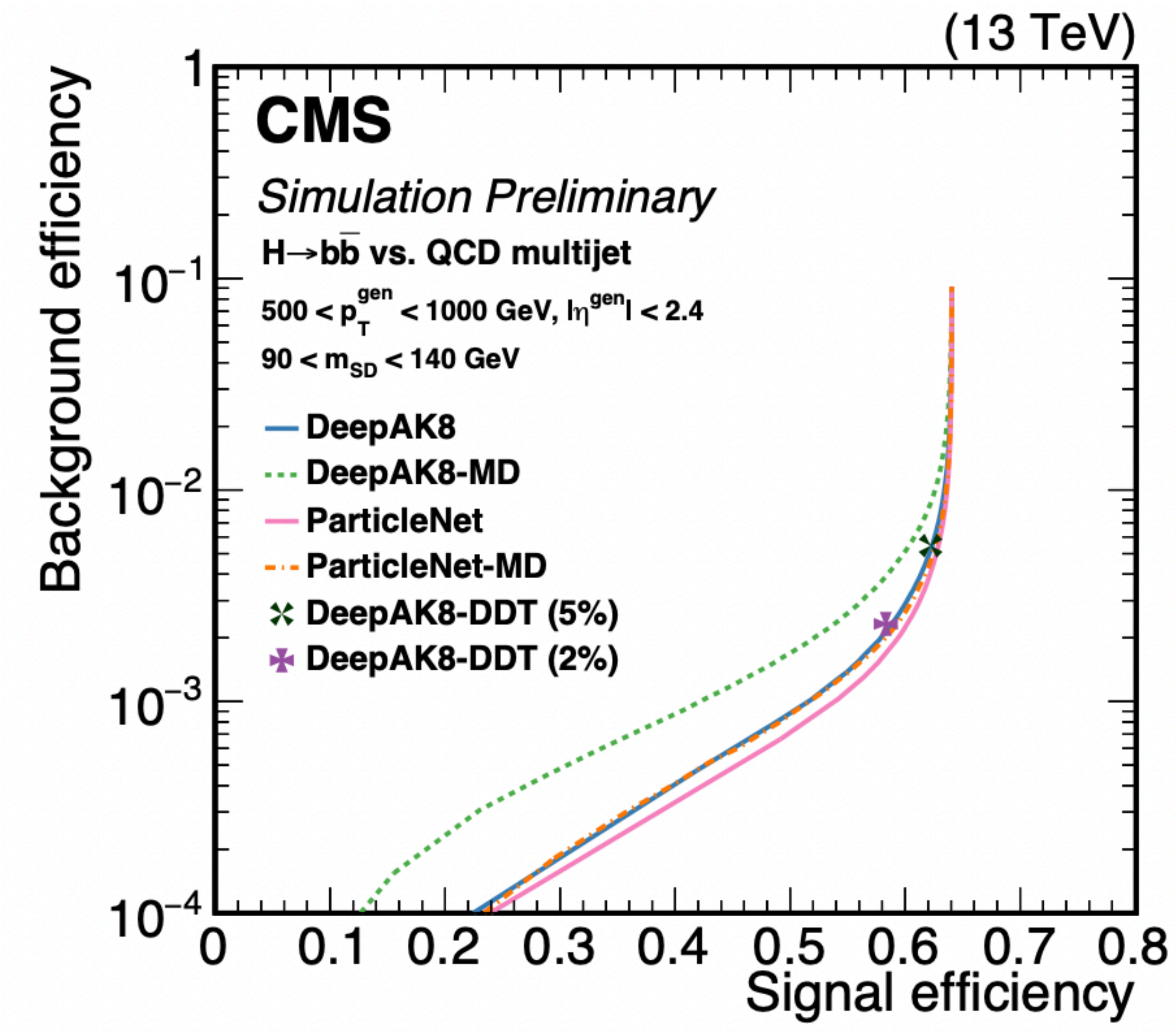
Phys. Rev. Lett. 125, 061801 (2020)

Use of deep learning with low-level inputs

Non-resonant boosted HH→bbbb analysis (2022)

- Focus on phase space region where both Higgs bosons are highly Lorentz boosted
- Reconstruction and identification of b quark pair from Higgs decay is achieved with **ParticleNet, a graph neural network algorithm**
  - Using PF candidates and secondary vertices as inputs, yielding substantial gains over other approaches

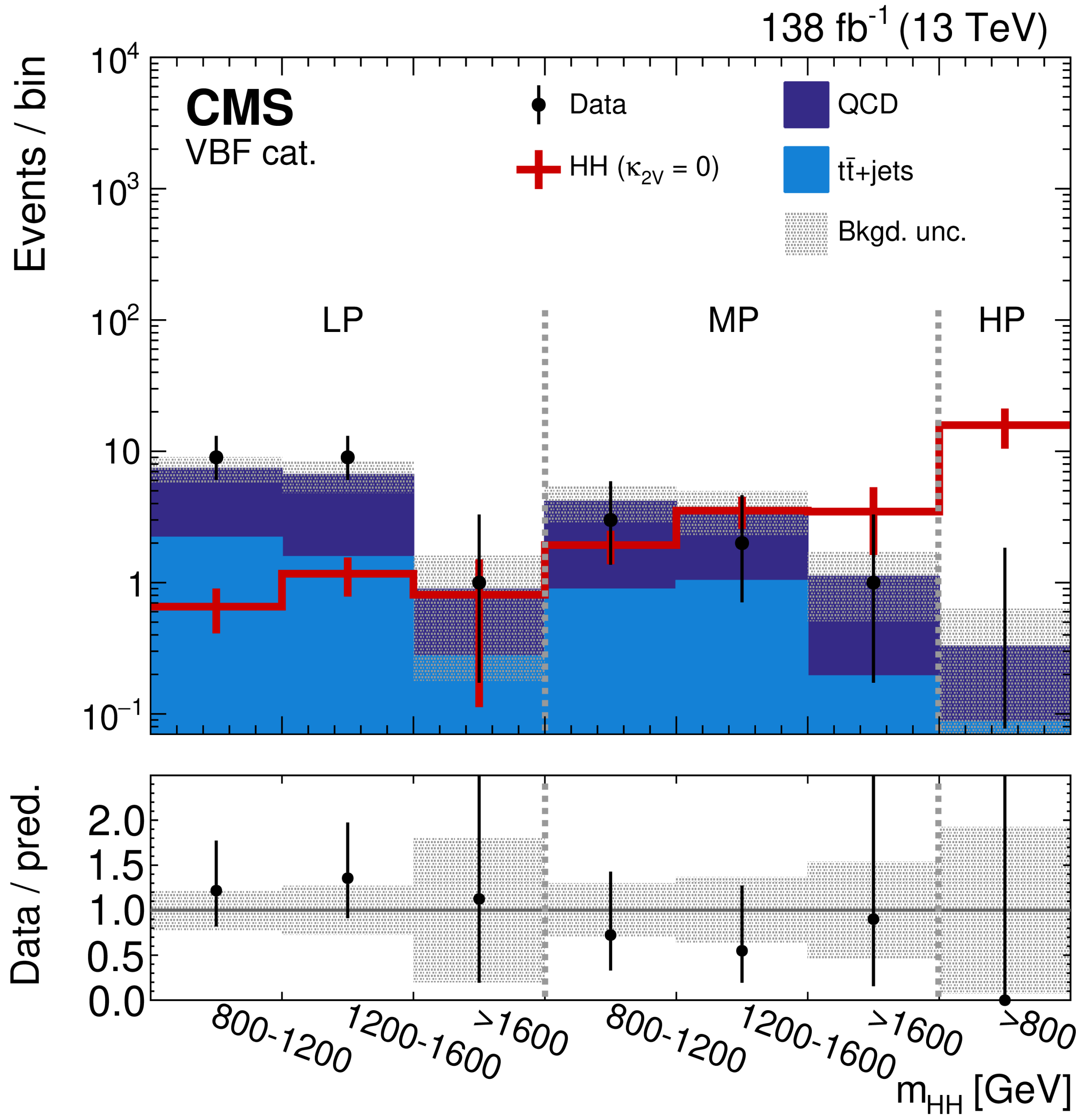


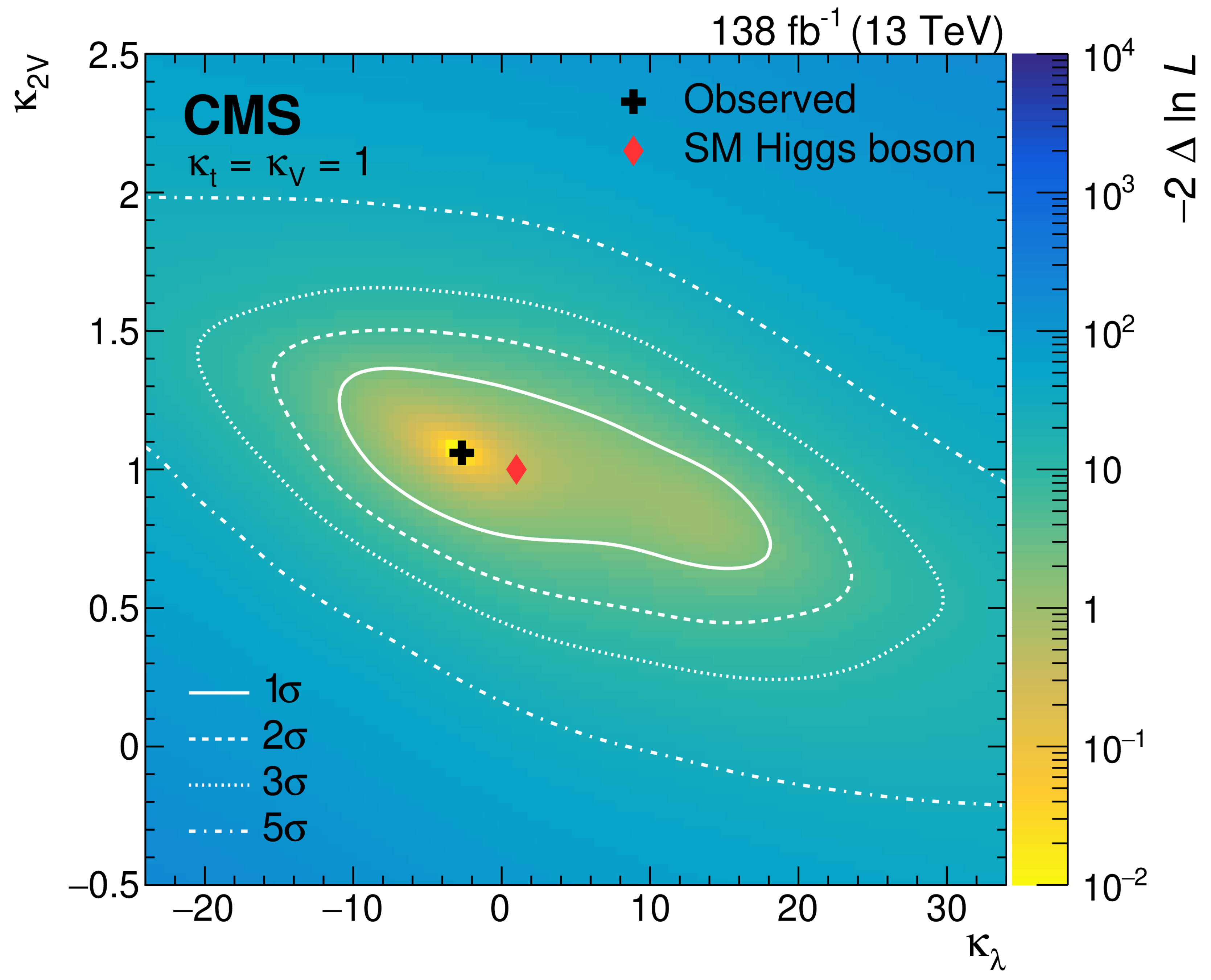


Phys. Rev. Lett. 131 (2023) 041803

Non-resonant boosted HH→bbbb analysis (2022)

- HH candidate mass is taken as final discriminant
- Constrains the H self-coupling strength and the quartic VVHH coupling strength κ_{2V}
  - **Excluding κ_{2V}=0 for the first time**, with a significance of 6.3σ





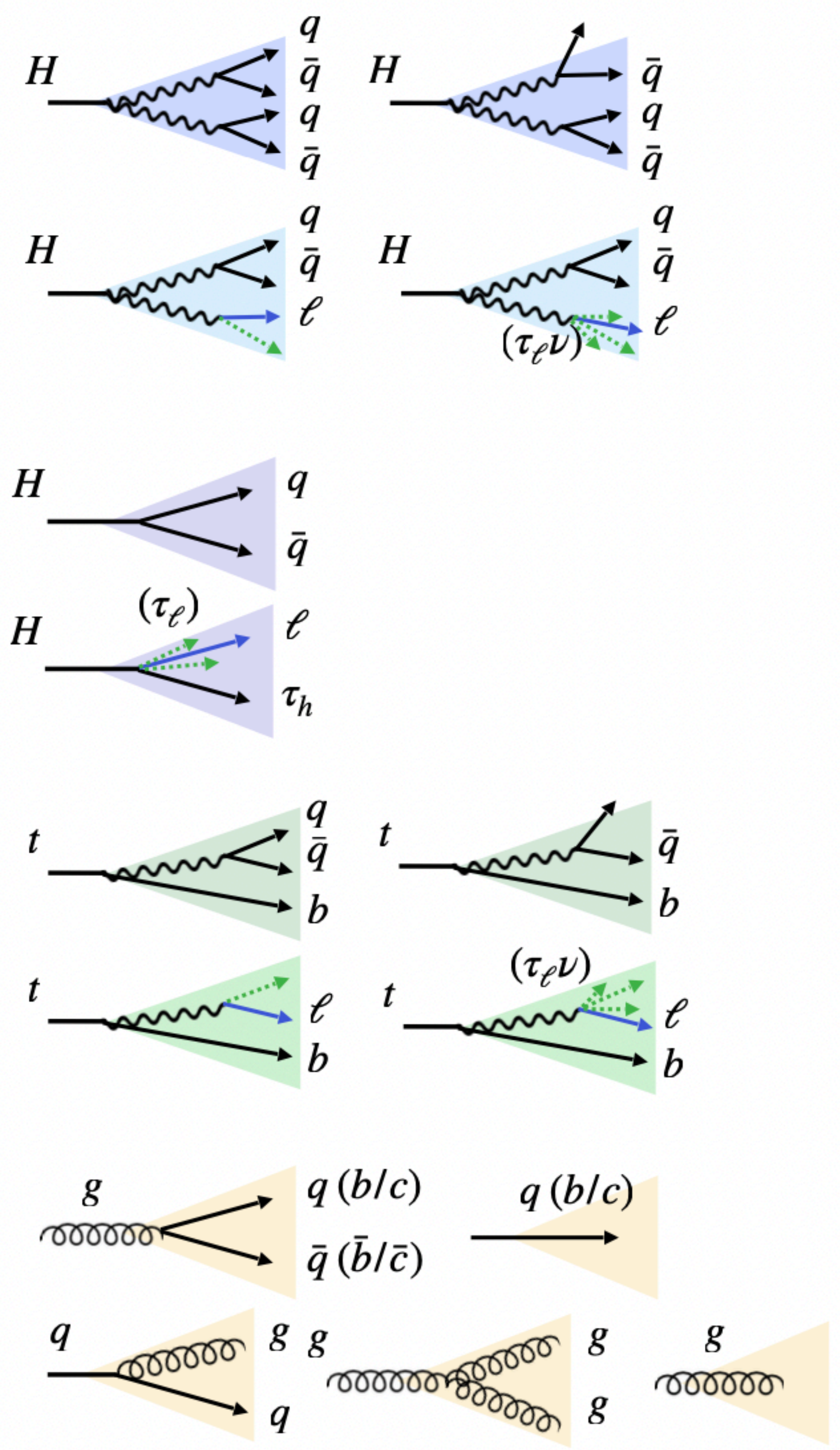
[Phys. Rev. Lett. 131 (2023) 041803]

Non-resonant boosted HH→bbVV analysis (2024)

- Extend to a large array of final states, including H→VV, all-hadronic, and semi-leptonic modes

- **Global Particle Transformer algorithm (GloParT)** uses learned “attention” to give more weight to certain particles in order to infer the origin of jets

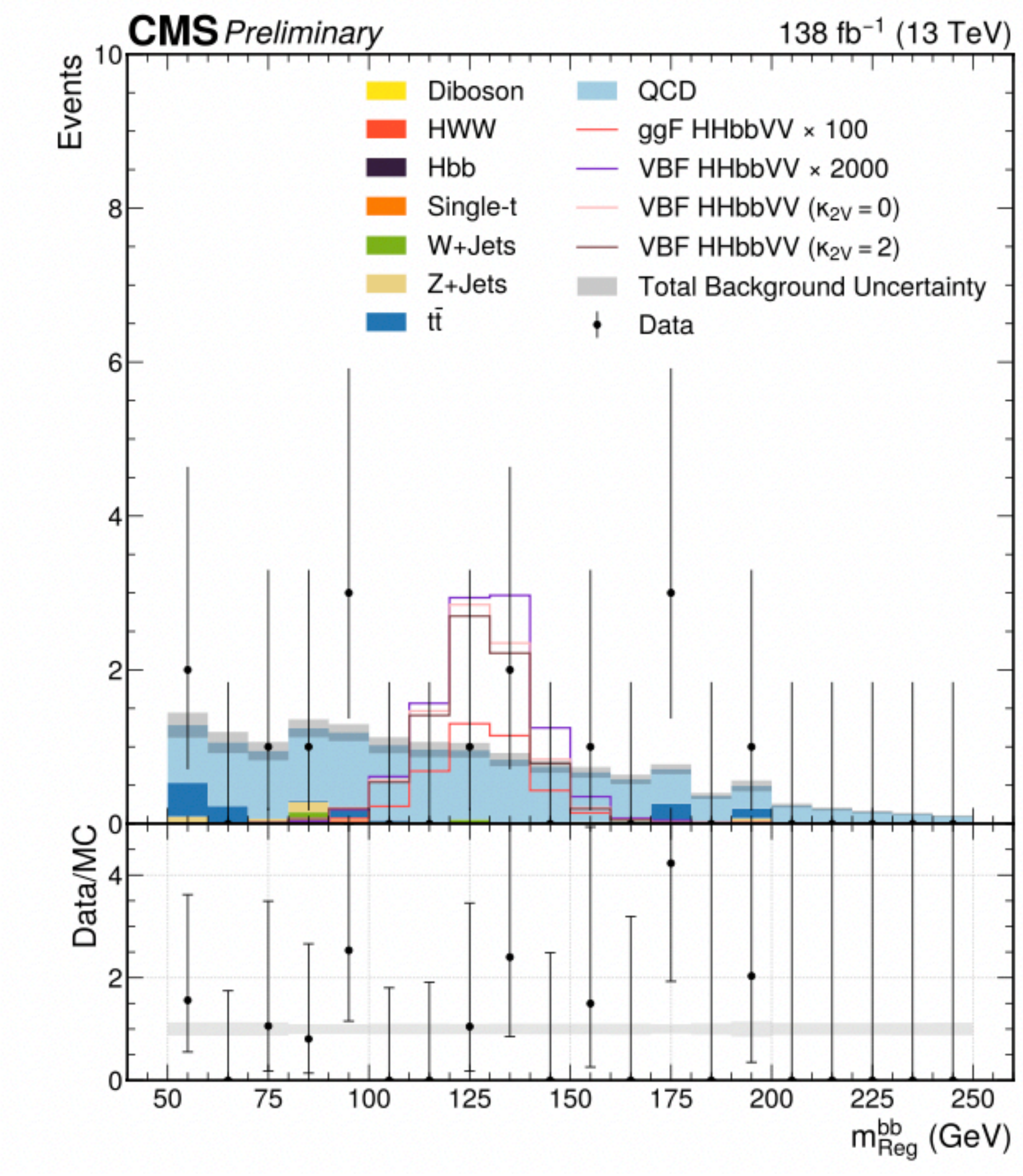
Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	$e\nu qq$	0c/1c	2
	$\mu\nu qq$		2
	$\tau_e \nu qq$		2
	$\tau_\mu \nu qq$		2
	$\tau_h \nu qq$		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e \tau_h$		1
	$\tau_\mu \tau_h$		1
	$\tau_h \tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	$be\nu$	1b	1
	$b\mu\nu$		1
	$b\tau_e\nu$		1
	$b\tau_\mu\nu$		1
	$b\tau_h\nu$		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1

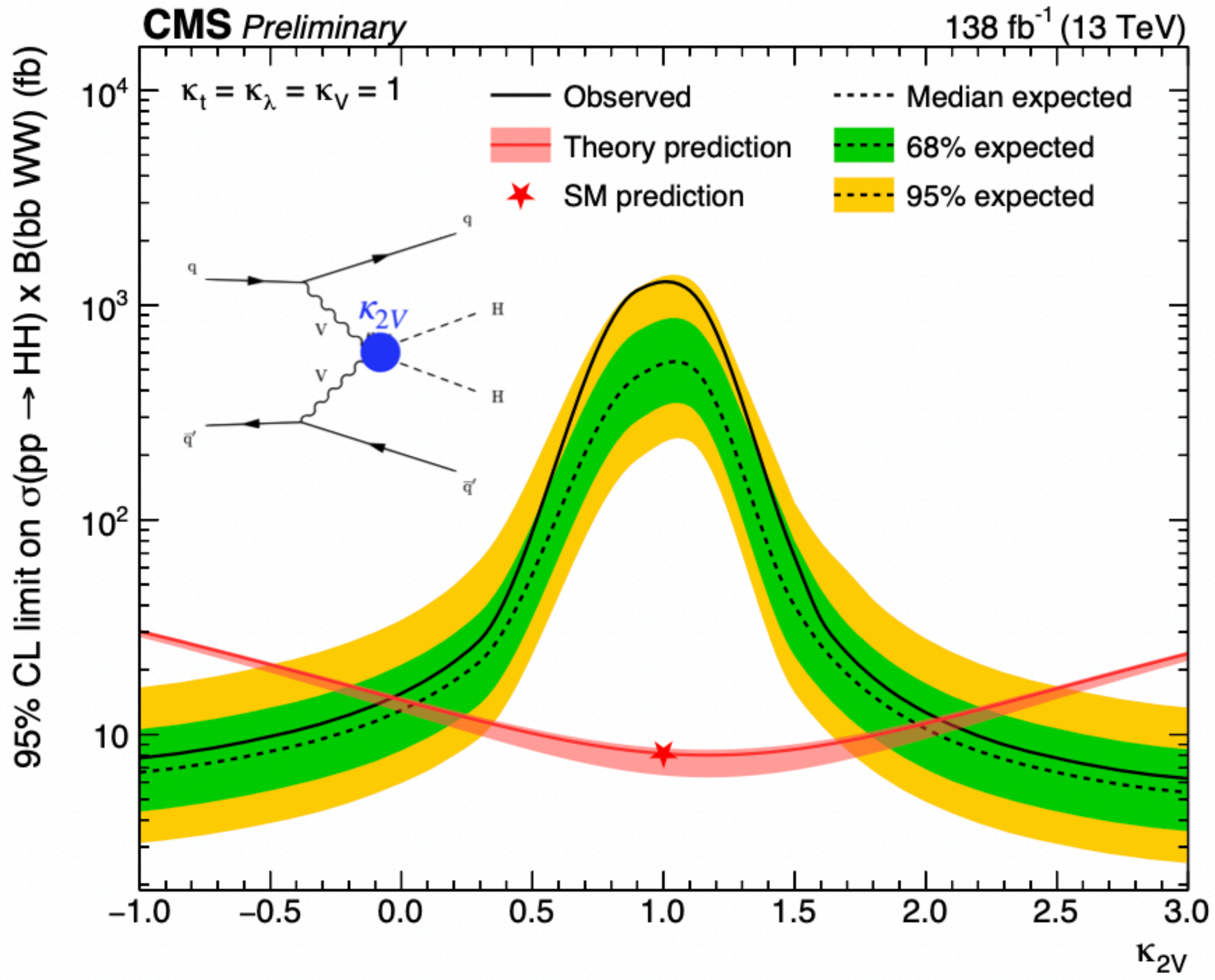


[CMS-PAS-HIG-23-012](CMS-PAS-HIG-23-012)

Non-resonant boosted HH→bbVV analysis (2024)

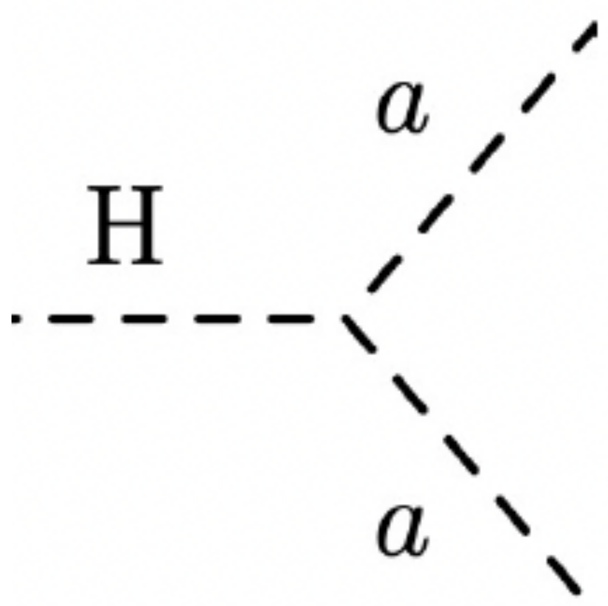
- Enables a new search for boosted HH → bbVV → bb4q
  - established ParticleNet mass-decorrelated tagger for H → bb jets
  - new high-performing GloParT tagger for H → VV jets
- **Provides second-best constraint on HHVV coupling K_{2V}**



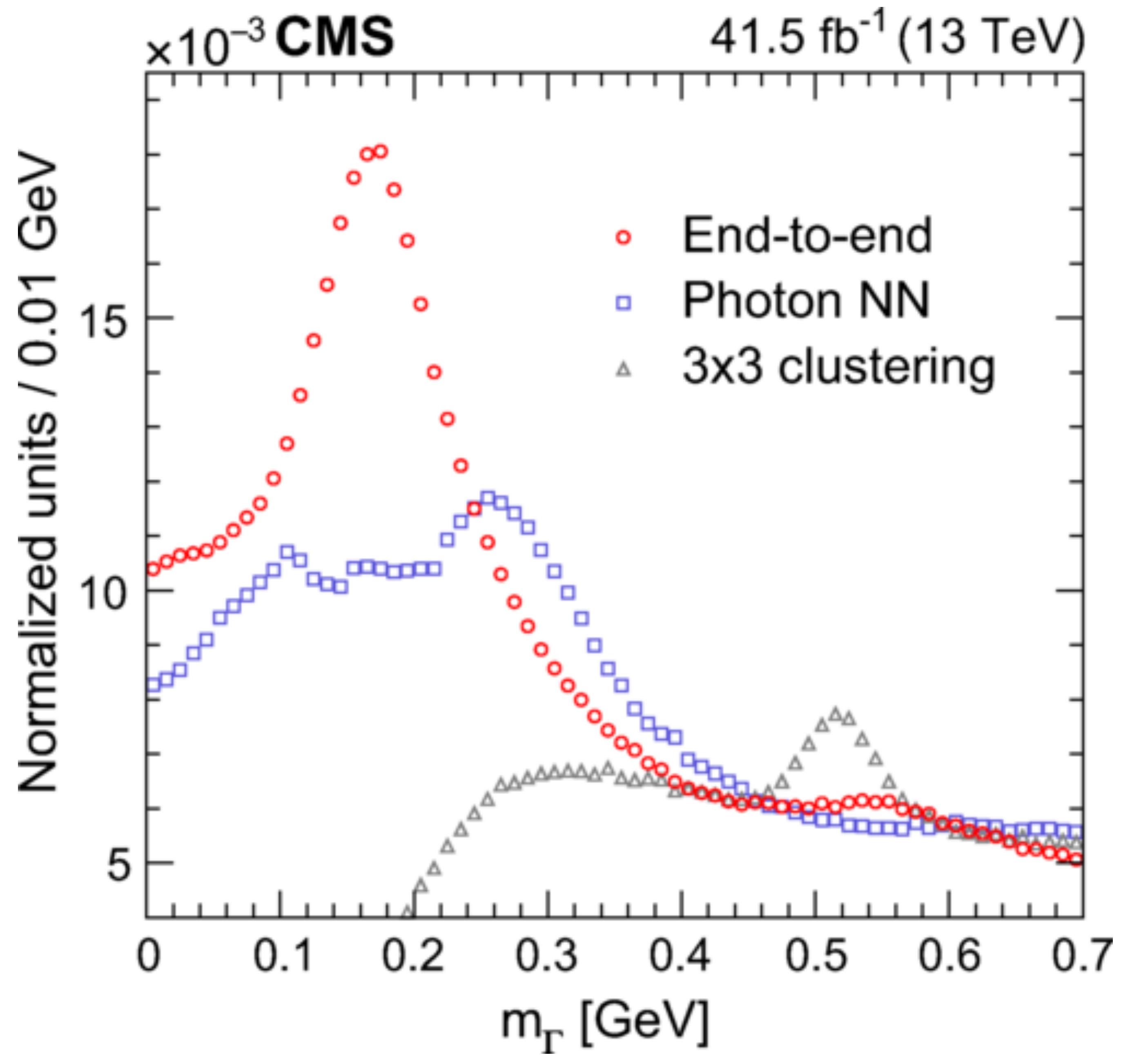


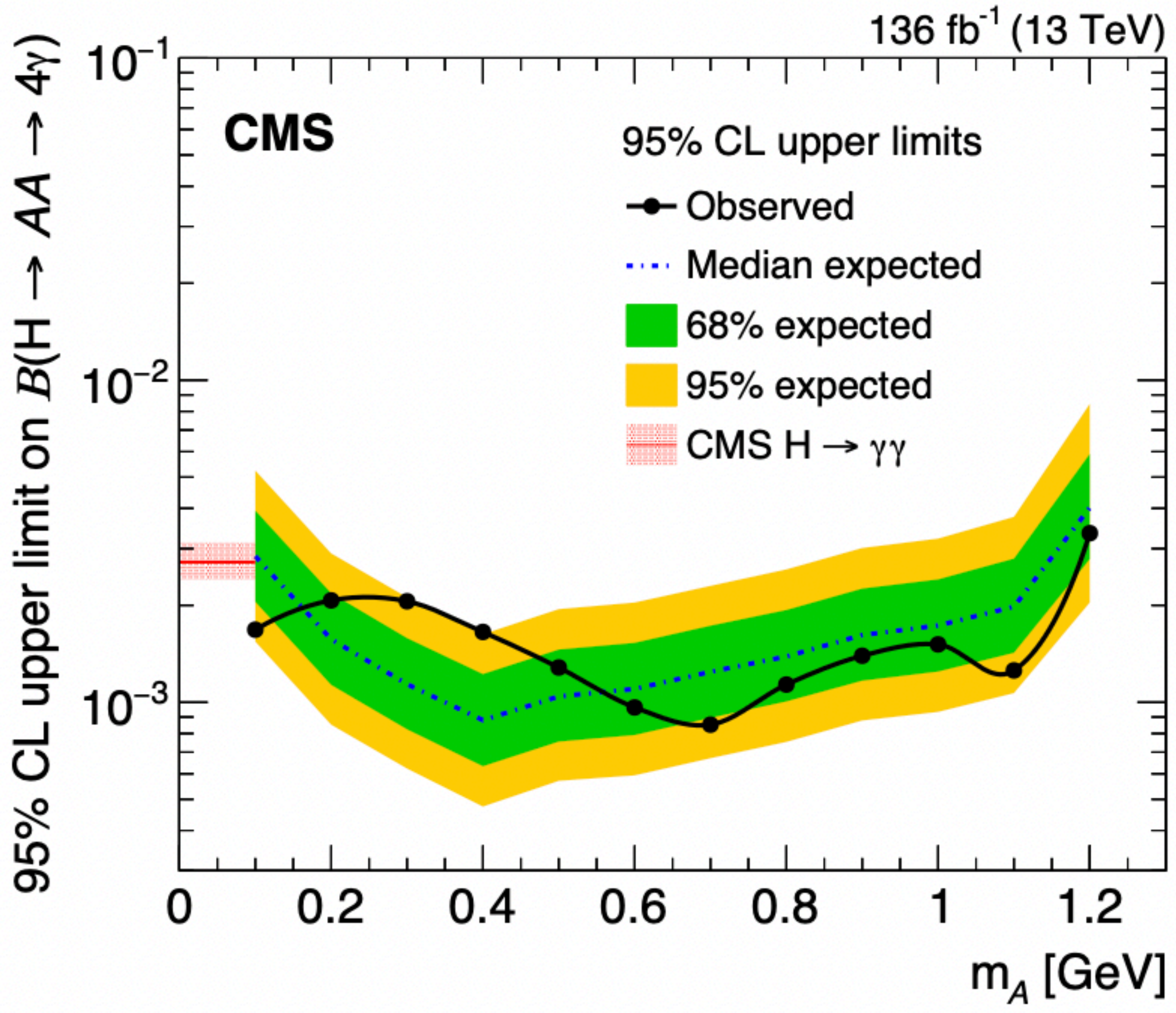
CMS-PAS-HIG-23-012

Search for Higgs exotic decay H→AA→γγγγ



- Very low-mass scalar A decays to two highly merged photons, reconstructed using an *end-to-end* deep learning strategy
  - It uses minimally processed detector data as input and directly outputs particle properties of interest
- Set upper limits on B(H → AA → 4γ) for masses of A in the range 0.1– 1.2 GeV





Phys. Rev. D 108 (2023) 052002

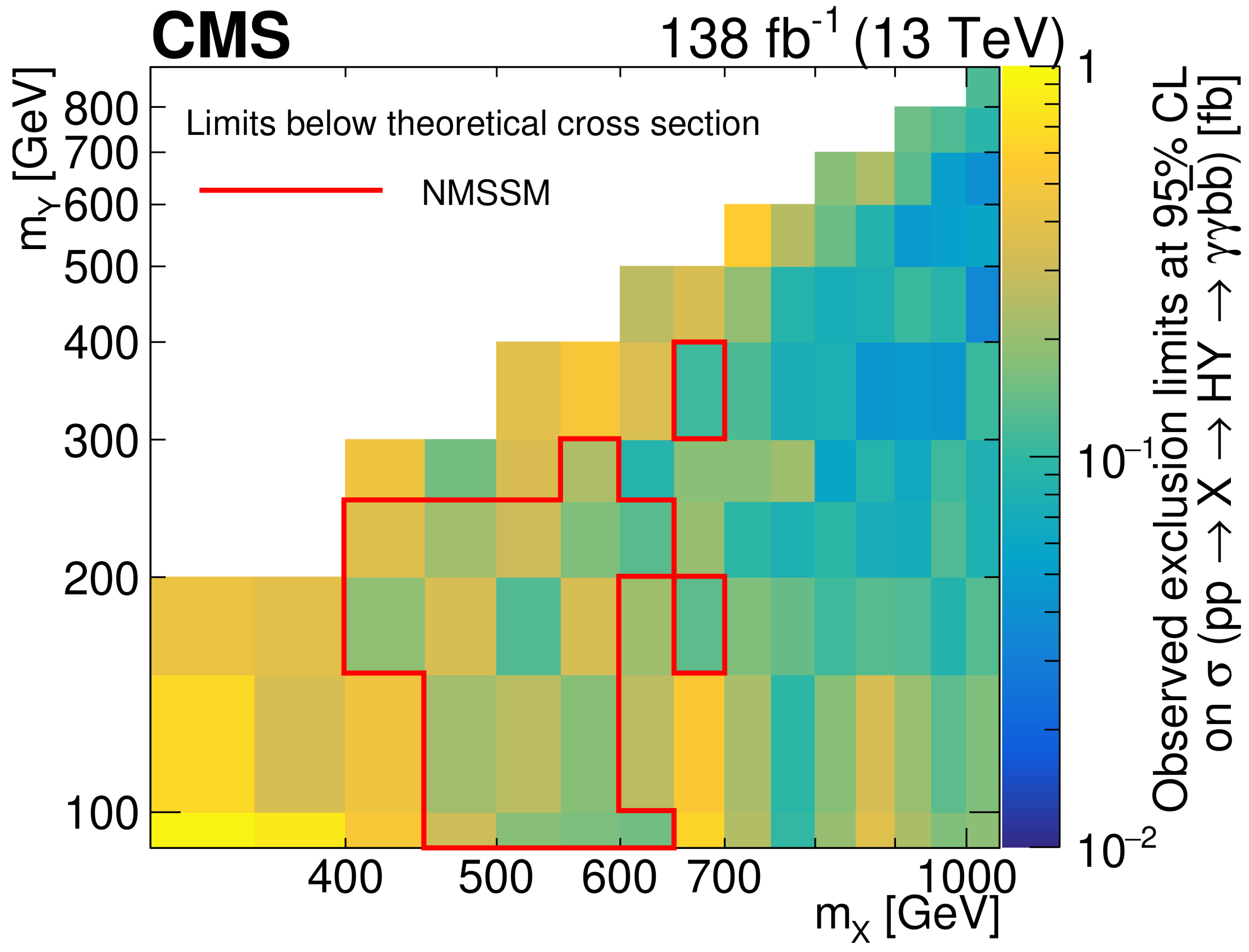
Phys. Rev. Lett. 131 (2023) 101801

AI-based event classification in heavy resonance searches

$X \rightarrow Y(bb)H(\gamma\gamma)$

- Six exclusive kinematic regions are defined based on hypothesised values of m_X and m_Y
- In each kinematic region, a BDT with 3 output classes (2 for backgrounds and 1 for signal) is trained
  - all contained signal samples and the two background samples are used with equal weight
- In each kinematic region, 3 event categories are defined based on output of corresponding BDT
  - for each m_X hypothesis, signal is inferred from a fit in 2D distributions of $m_{\gamma\gamma}$ and m_{jj}

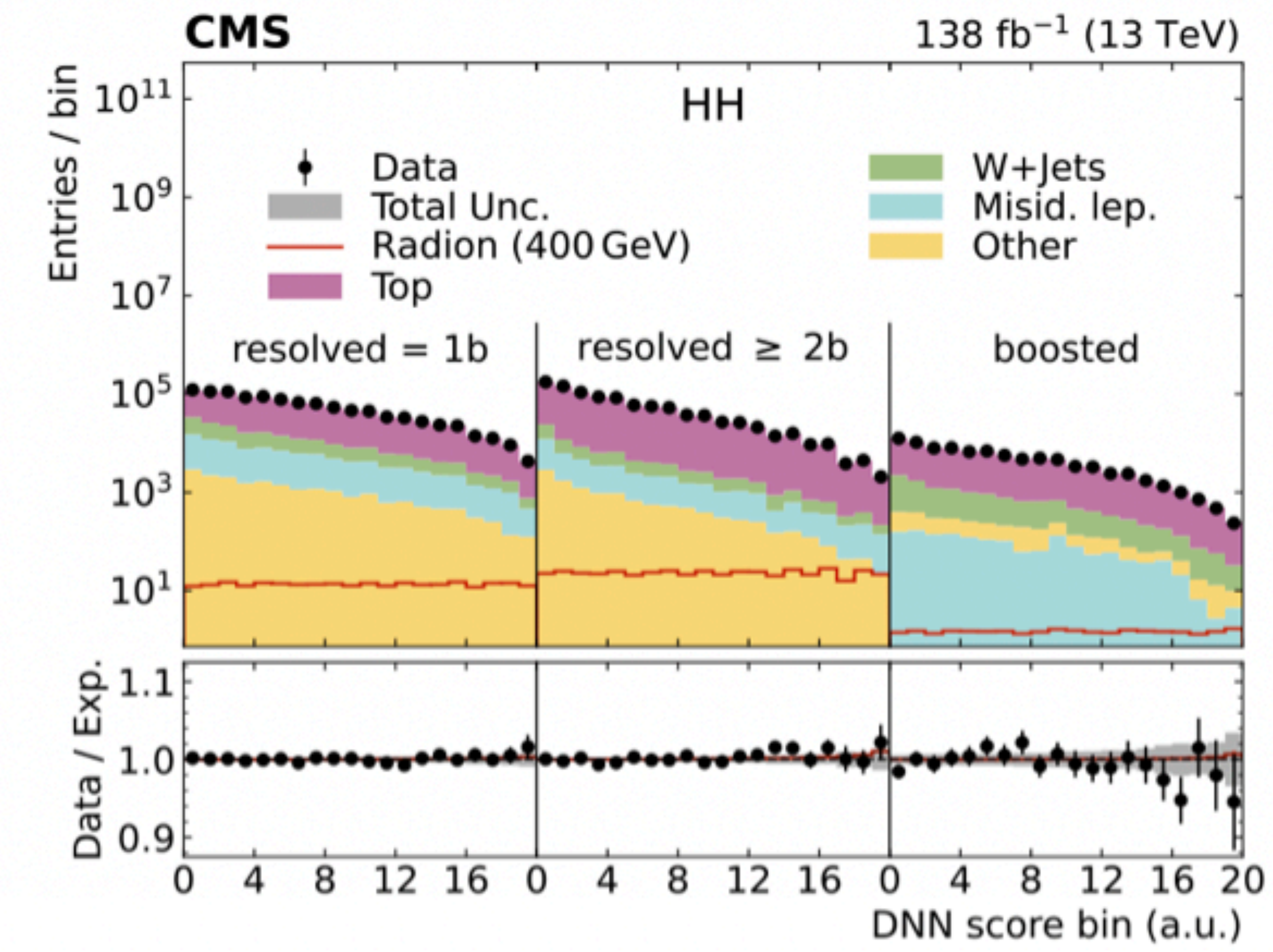
	$m_Y < 300$ GeV	$m_Y = [300–500]$ GeV	$m_Y > 500$ GeV
$m_X < 500$ GeV	CAT 0 = 0.63–1.0 CAT 1 = 0.33–0.63 CAT 2 = 0.17–0.33		
$m_X = [500–700]$ GeV	CAT 0 = 0.55–1.0 CAT 1 = 0.40–0.55 CAT 2 = 0.21–0.40	CAT 0 = 0.60–1.0 CAT 1 = 0.35–0.60 CAT 2 = 0.18–0.35	
$m_X > 700$ GeV	CAT 0 = 0.50–1.0 CAT 1 = 0.30–0.50 CAT 2 = 0.21–0.30	CAT 0 = 0.35–1.0 CAT 1 = 0.24–0.35 CAT 2 = 0.18–0.24	CAT 0 = 0.40–1.0 CAT 1 = 0.29–0.40 CAT 2 = 0.13–0.29

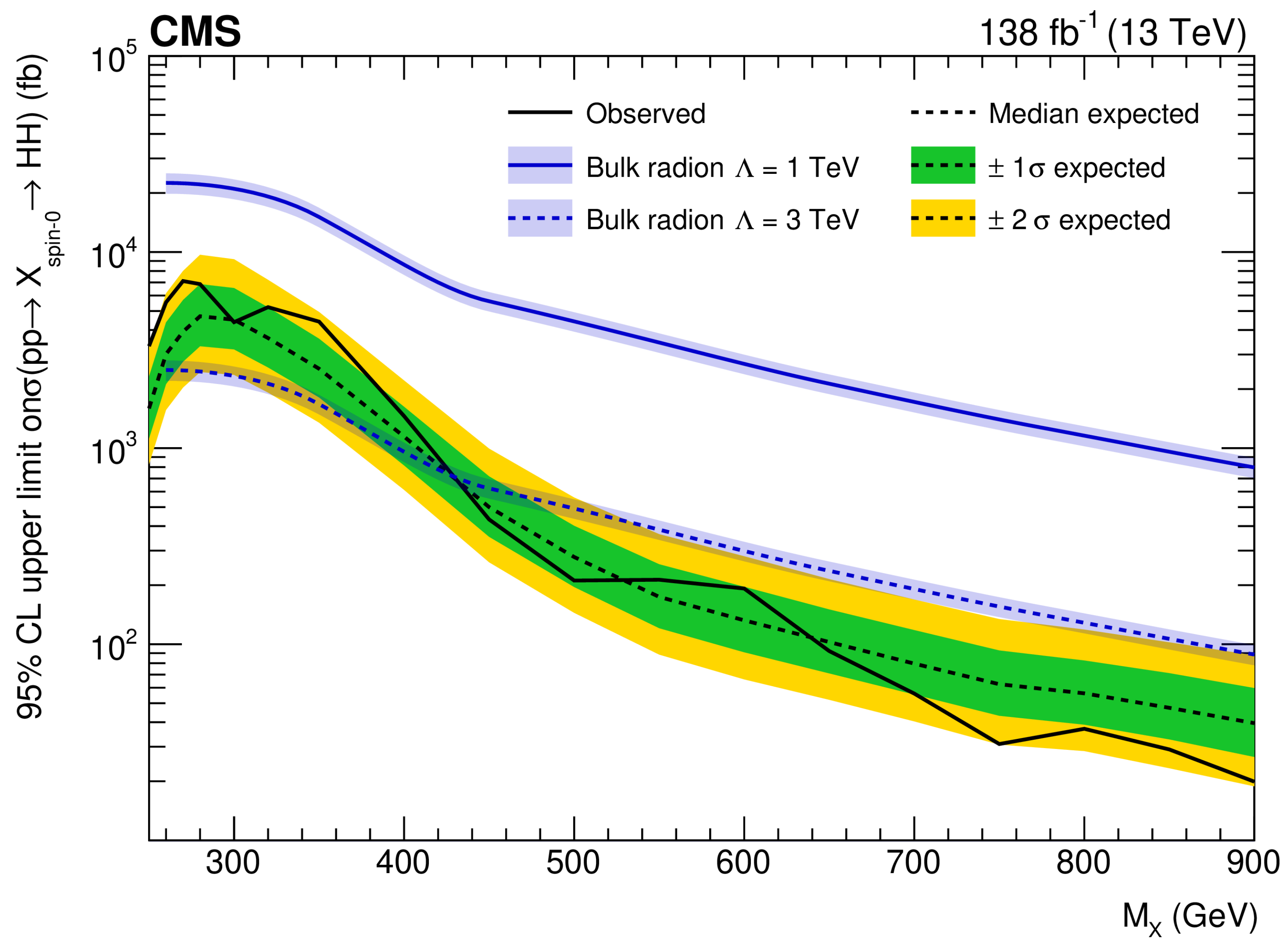


JHEP 05 (2024) 316

X → H(bb)H(WW)

- DNNs feature output nodes for a number of backgrounds and one signal node
  - DNNs are trained on all signal samples; they are parameterized in nominal signal mass
- DNN architecture is complemented by a Lorentz Boost Network acting as input preprocessor
  - takes four-vectors of reconstructed particles as input and creates additional observables
- Depending on the highest scoring node, events are subdivided into signal and background categories
  - signal extraction is performed by a fit to DNN output distributions

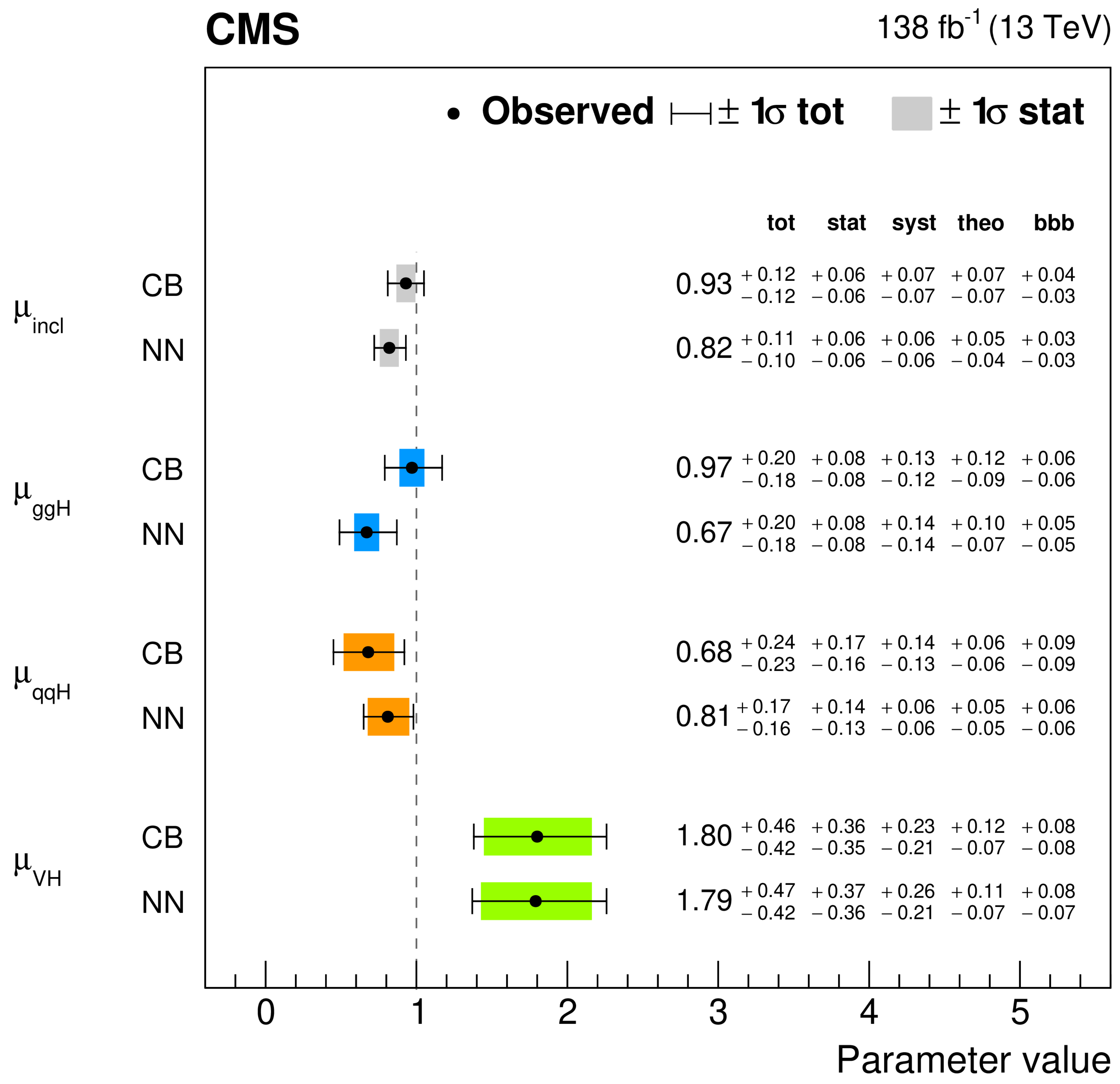




arxiv:2403.09430

Systematic-aware neural network for binned-likelihood-analyses

Systematic-aware neural network

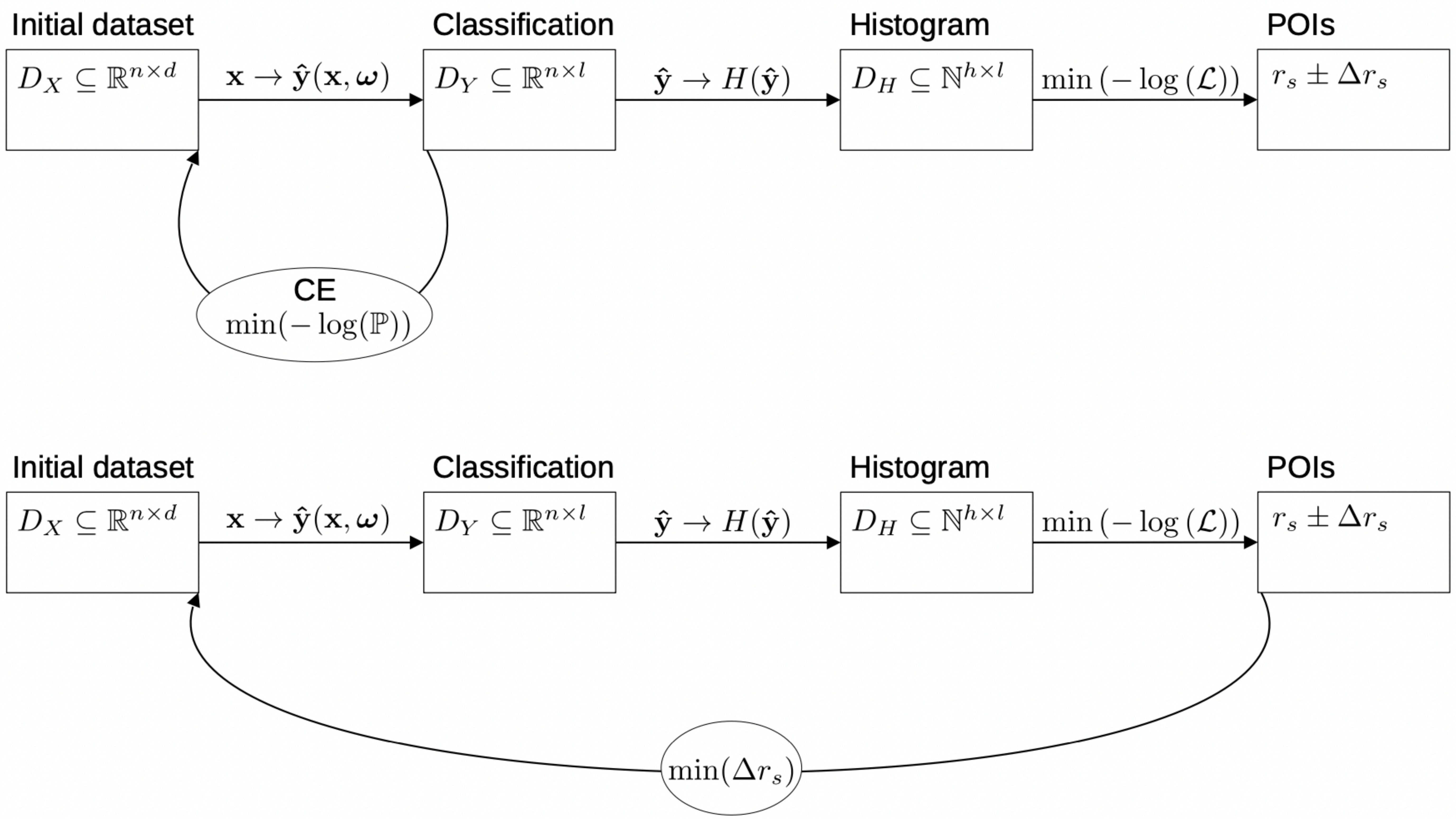


- Prospects for future measurements:
  - Statistical uncertainties will decrease for Run3 and HL-LHC
  - Systematic uncertainties will become more important

Eur. Phys. J. C 83 (2023) 562

Systematic-aware neural network

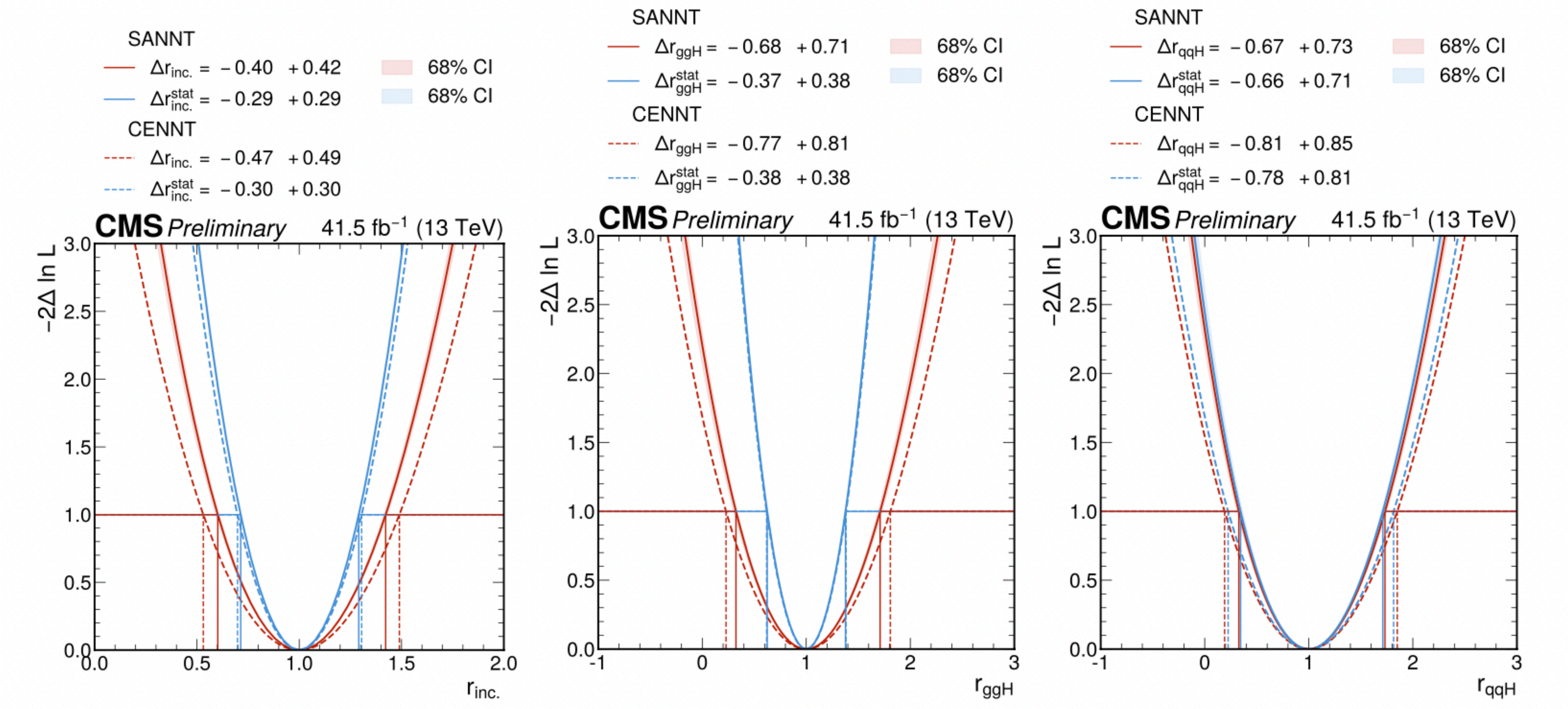
- CMS demonstrate a neural network training, capable of accounting for the effects of systematic variations, and describe its extension towards multiclass classification
  - Key: choosing gradients for histogram operation



CMS-PAS-MLG-23-005

Systematic-aware neural network

- Based on a comprehensive data model with 86 nontrivial shape-altering systematic variations for H→ττ analysis
  - with respect to a conventional training, observe improvements of 12% and 16%, for the sensitivity in r_{ggH} and r_{qqH}

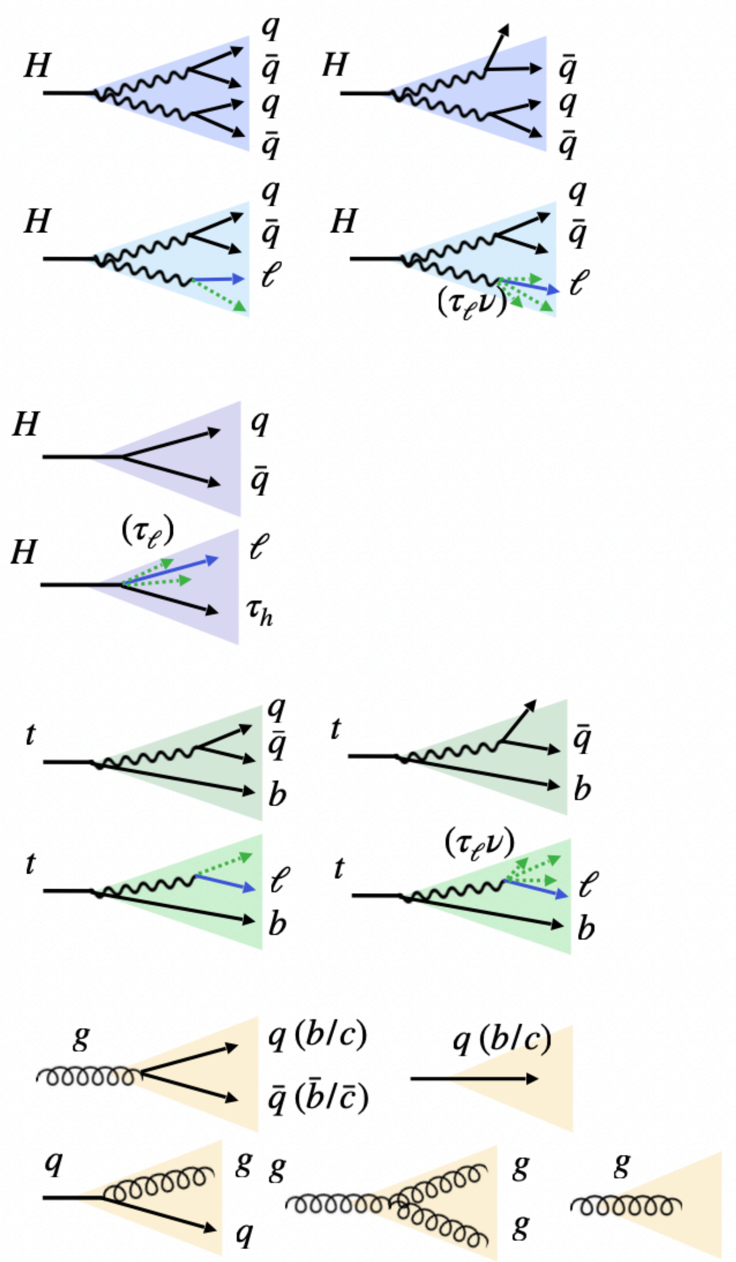


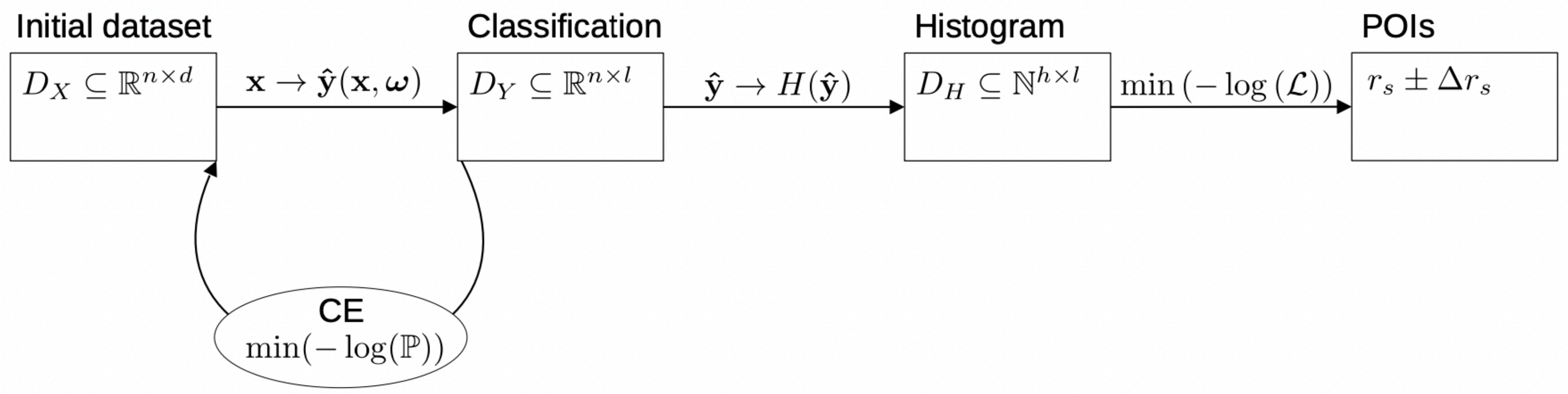
CMS-PAS-MLG-23-005

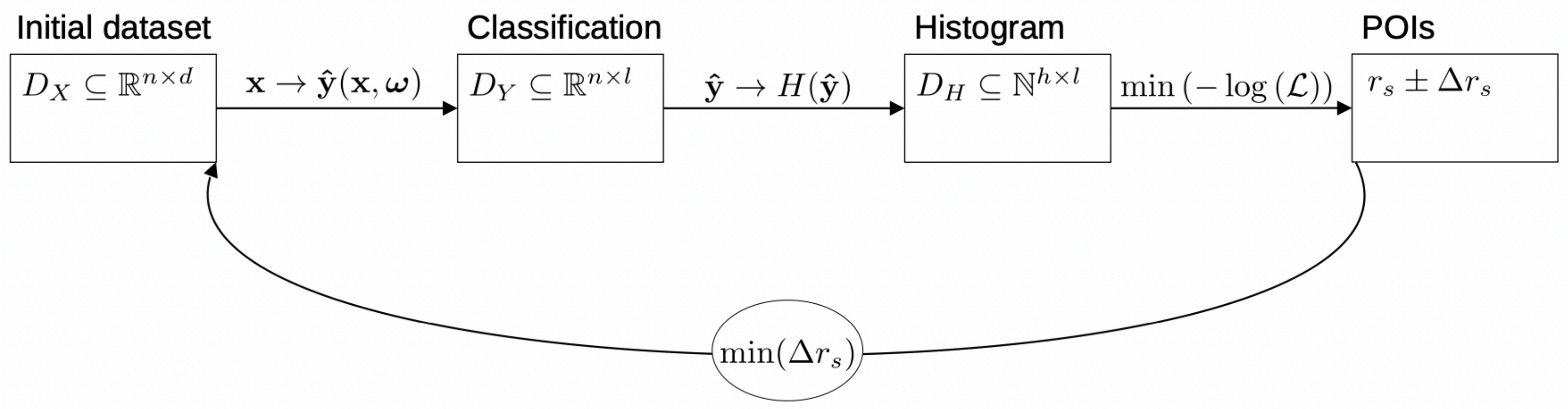
Summary

- **Machine Learning greatly enhances our ability of identifying signal from background: important for discovery of new physics**
- Lots of recent progress at CMS:
  - deep learning particle/event reconstruction
  - model-independent searches (see Roberto Seidita's talk)
  - systematic-aware neural network
  - etc.
- And there are much more to come!

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	$e\nu qq$	0c/1c	2
	$\mu\nu qq$		2
	$\tau_e\nu qq$		2
	$\tau_\mu\nu qq$		2
	$\tau_h\nu qq$		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e\tau_h$		1
	$\tau_\mu\tau_h$		1
	$\tau_h\tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	$be\nu$	1b	1
	$b\mu\nu$		1
	$b\tau_e\nu$		1
	$b\tau_\mu\nu$		1
	$b\tau_h\nu$		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1



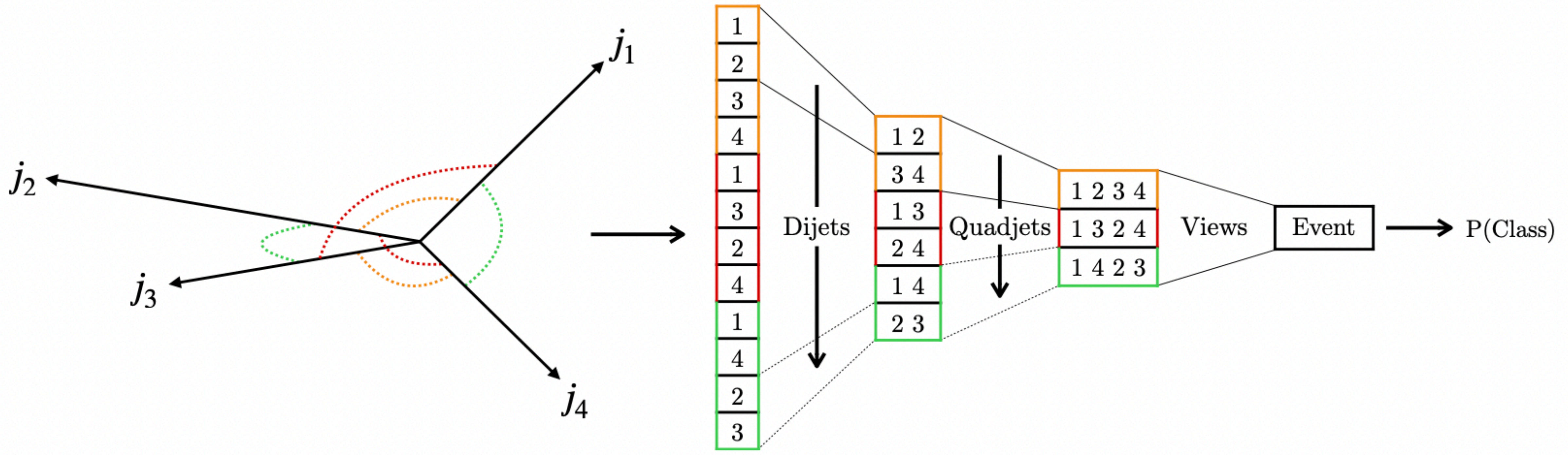




Thanks!

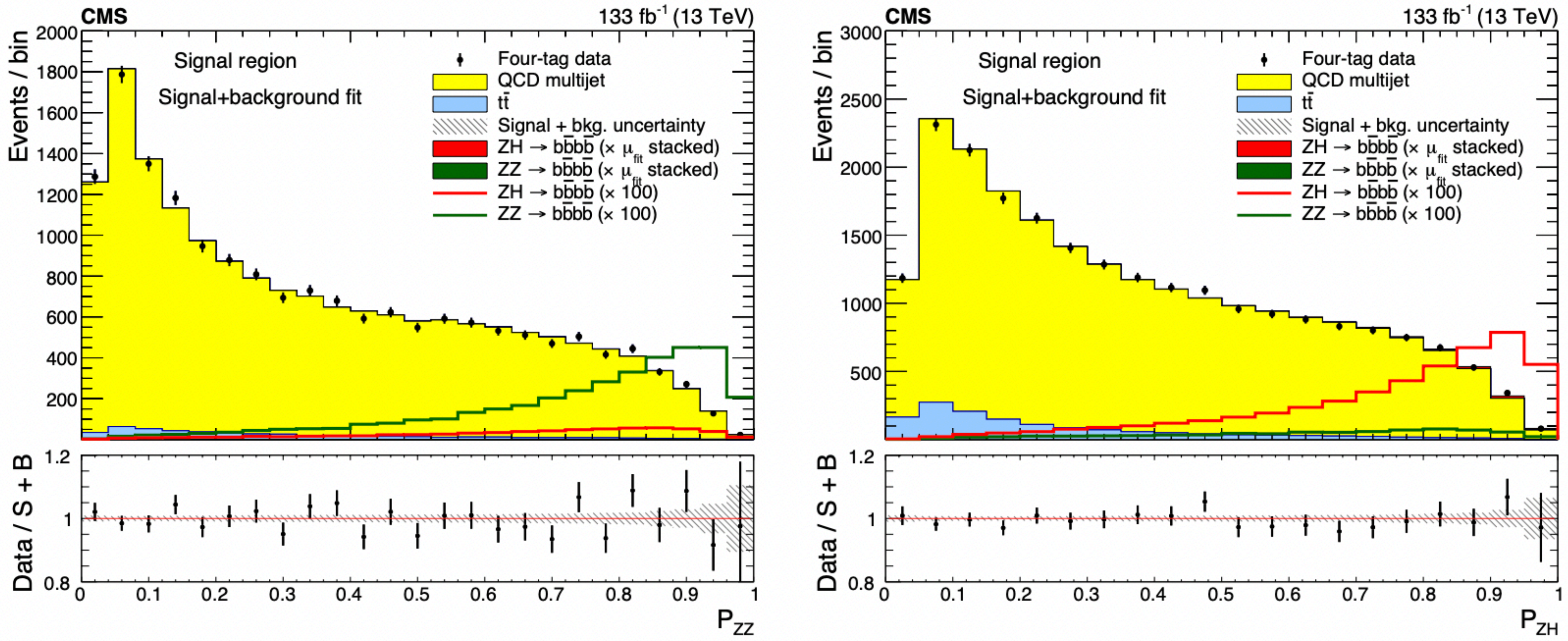
ZZ/ZH → 4b

- Search for ZZ and ZH production in 4b final state



- Benefits from a multiclass multivariate classifier, which uses convolutions to solve combinatoric jet pairing problem, and has been designed with an architecture customized to 4b final state

- Observed (expected) upper limits on ZZ → 4b and ZH → 4b production cross sections correspond to 3.8 (3.8) and 5.0 (2.9) times SM prediction, respectively

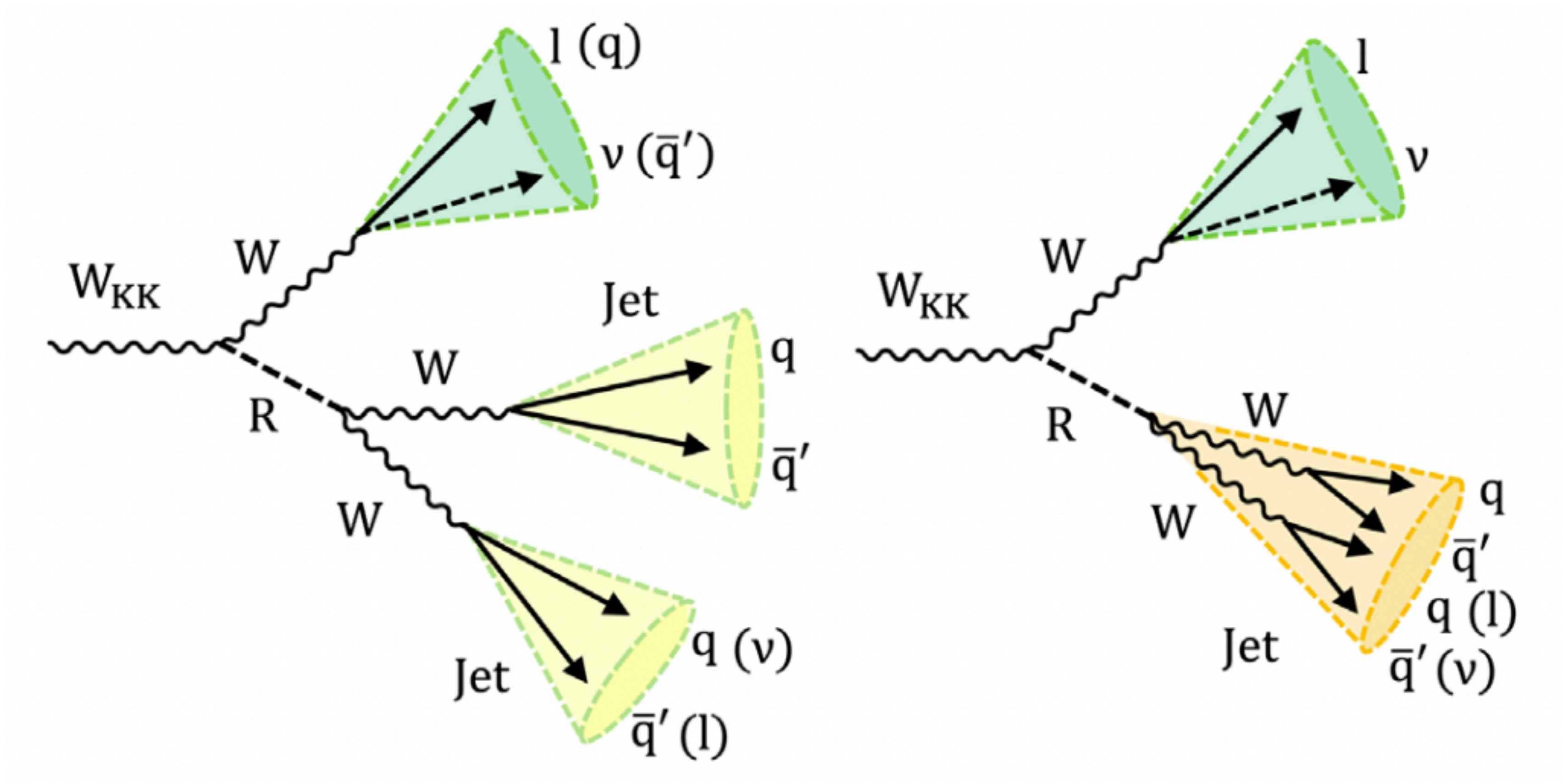


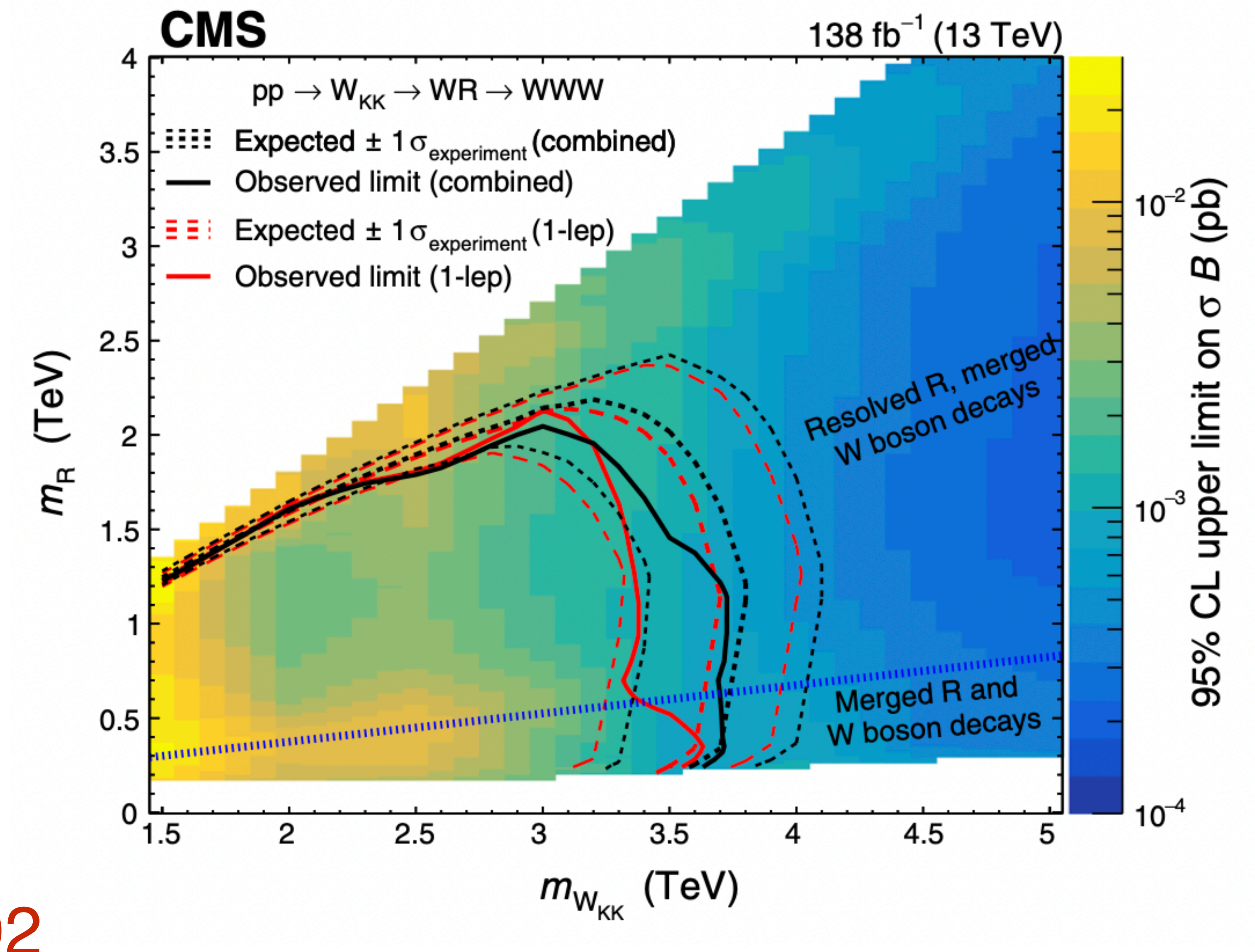
- Analysis techniques directly applicable to the HH → 4b analysis

arxiv:2403.20241

Search for Resonances Decaying to Three W Bosons

- Radion decay configurations are simultaneously probed by combining outputs of DEEPAK8 algorithm
  - call the resulting discriminants for merged radion decays “deep-WH” and for W bosons “deep-W”
- Using jet mass and deep-WH & deep-W discriminants, selected events are split into six SRs
  - Limits are set on an extended warped extra-dimensional model
- The novel radion identification techniques are also applicable to Lorentz-boosted Higgs boson decays





Phys. Rev. Lett. 129 (2022) 021802

Systematic-aware neural network

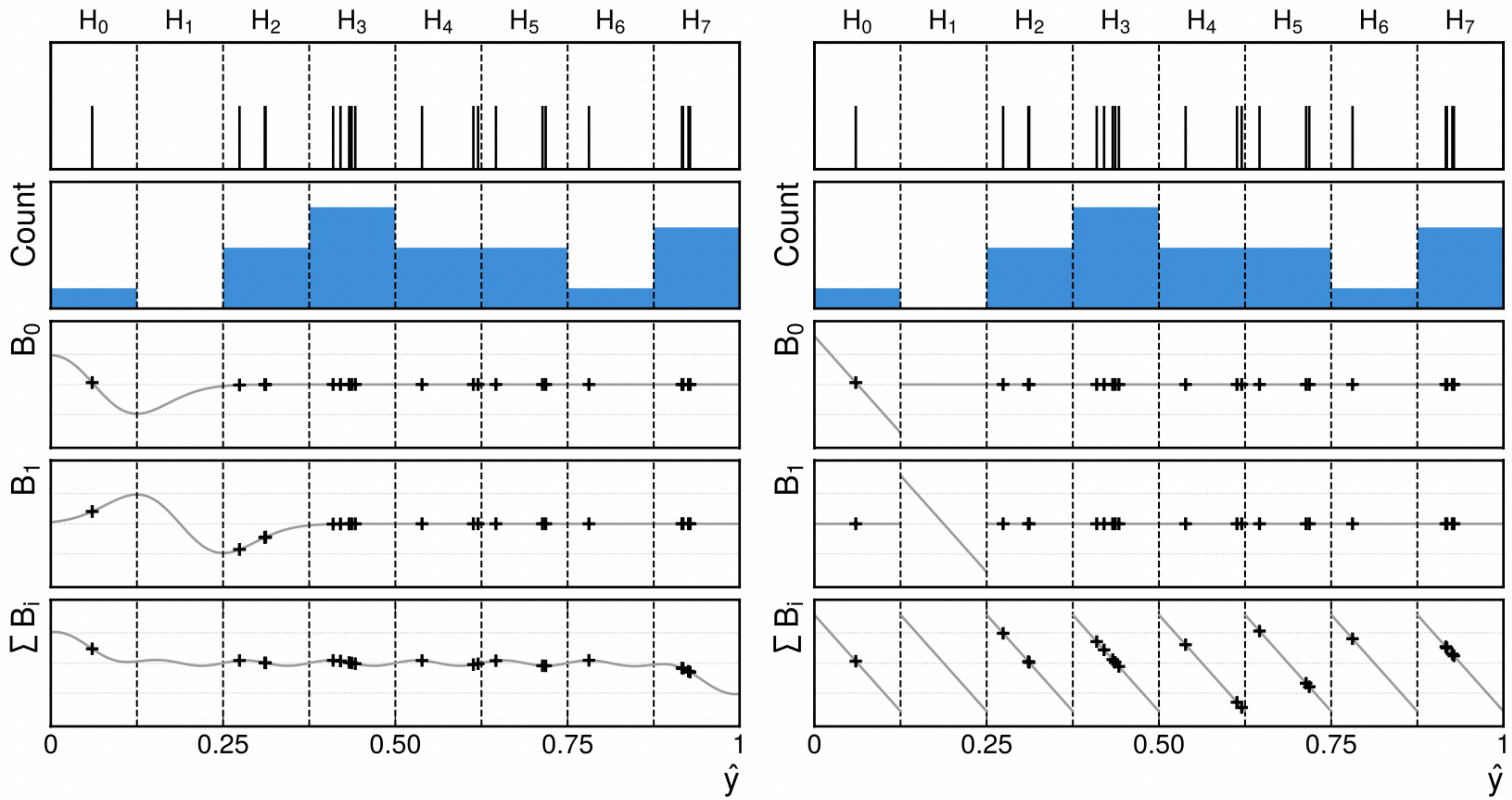


Figure 2: Custom functions $\mathcal{B}_i$ for the backward pass of the backpropagation algorithm, as used (left) in Ref. [5] and (right) in this paper. In the first row of each sub-figure the same 20 random samples of a simple setup of pseudo-experiments, as described in Section 3.2 are shown. In the second row the resulting histogram H, in the third and fourth rows the functions B_0 andB_1 for the individual bins H_0 and H_1, and in the last row the collective effect of $\sum \mathcal{B}_i$ are shown.

CMS-PAS-MLG-23-005