



北京大学
PEKING UNIVERSITY

“HADRON-IN-FAT-JET” AI TAGGING TO DETECT RARE DECAYS

[Phys. Rev. D 114, 056019 \(2026\)](#)

Linrui Chen, Tianyi Yang, Zixun Kou, Zijian Wang, Youpeng
Wu, Leyun Gao and Qiang Li

QCD@LHC 2026 University of Oxford

Sep 28, 2026

Rare Hadronic Decays of the W Boson

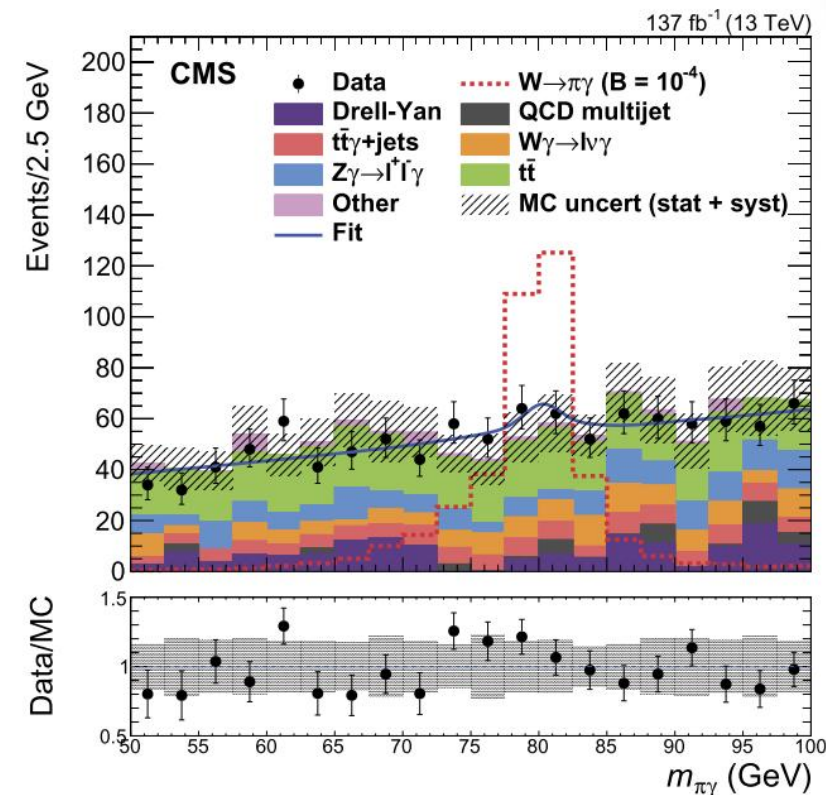
- Theoretical significance: A unique laboratory to test the SM at the interface of perturbative and non-perturbative QCD.
- QCD factorization: Serves as a gold-standard “test bench” for meson form factors at large energy scales.
- Extreme rarity: The SM predicted BF for $W^\pm \rightarrow \pi^\pm \gamma$ is as low as $\mathcal{O}(10^{-9})$.

Experimental Searches at the LHC

- Explored by **ATLAS**, with current the most stringent upper limits constrained at $\mathcal{O}(10^{-6})$. [Phys. Rev. Lett. 133 (2024) 161804]
- **CMS** reported its latest result of the upper limit in 2021 as 1.50×10^{-5} with 137 fb^{-1} proton-proton collisions data at $\sqrt{s} = 13 \text{ TeV}$ with lepton trigger in $t\bar{t}$ events.

Machine Learning Gap

- Advanced AI taggers are rarely applied to exclusive/pure hadronic rare decays.



[arXiv:2011.06028v2]

Introduction of Sophon: a Jet-tagging Model

[arXiv:2405.12972v1]

Overview of the Sophon Model

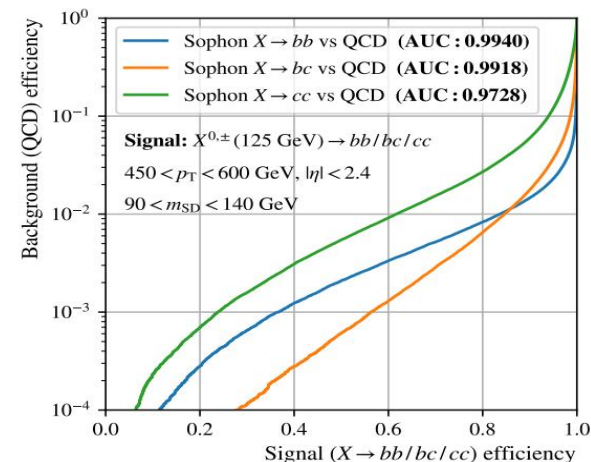
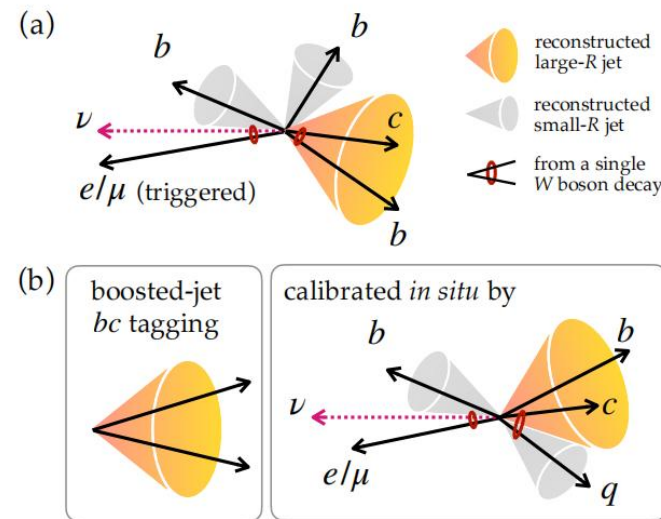
- A next-generation, deep-learning-based jet tagger designed for optimal performance across the vast majority of possible jet final-states.
- Built upon the **Particle Transformer (ParT)** architecture and pre-trained on the comprehensive JetClass-II dataset [Hugging face].
- Output a high-dimensional discriminant (**188 categories**), encoding both perturbative and non-perturbative jet substructures.

Proven Tagging Capabilities

- Validated to perform comparably to or even exceed top-tier CMS taggers (e.g. ParticleNet-MD and GloParT) in simulation.
- Achieve excellent background rejection and high AUC in established channels like bb and cc tagging.

Extension to Unexplored Region

- Successfully covers exotic configurations (e.g., novel bc -tagging) without requiring proxy phase spaces for calibration.

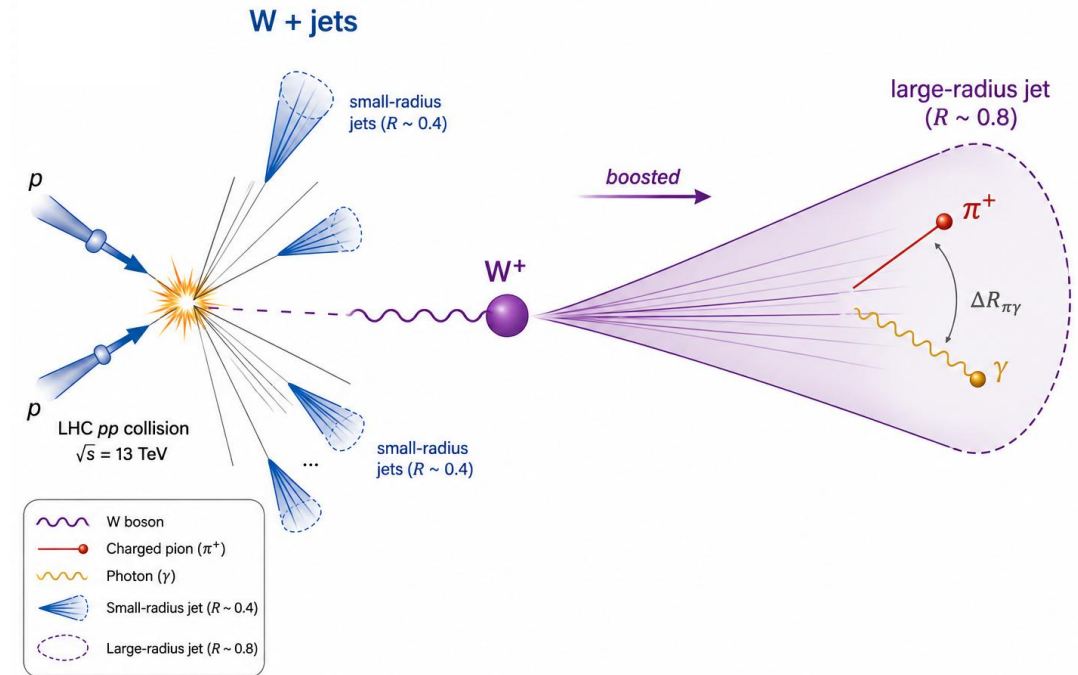


[arXiv:2503.00118v1]

Novel $W \rightarrow \pi\gamma$ Region: “Hadron-in-fat-jet”

“Hadron-in-fat-jet” $\rightarrow \Delta R_{\pi\gamma} < 0.8 \rightarrow$ **Boosted W region**

- Channel: $W + \text{jets}$
- A pion and a photon will be closer in a boosted W boson decay final-state due to Lorentz boost which provides a clean environment to apply novel “hadron-in-jet” tagging method with Sophon.
- Treat $W^\pm \rightarrow \pi^\pm\gamma$ process as a single fat-jet ($R=0.8$).
- Jet-matching with their inside substructure.



Experimental Set-up: MC Samples

- Generator: MADGRAPH5_AMC@NLO v2.9.18
- Parton-shower and hadronization: PYTHIA8
- Fast detector simulation: DELPHES 3.5.1
- Requiring $H_T > 700$ GeV at generator-level and using $H_T > 1050$ GeV trigger at reconstruction-level for all MC samples which is consistent with CMS trigger.

Fat-jet reconstruction:

- Apply Pileup Per Particle Identification (PUPPI) algorithm to mitigate pileup contamination.
- The PUPPI-weighted energy-flow objects are then clustered into large-radius ($R=0.8$) fat-jets using the anti- k_T algorithm with $p_T > 200$ GeV and $|\eta| < 2.5$.

Table 1: Summary of MC samples

Background Processes	$N_{\text{gen}}(\times 10^4)$	Cross-Section (mb)
Signal ($W^\pm \rightarrow \pi^\pm \gamma$)	990	1.632×10^{-8}
QCD	99025	9.047×10^{-6}
$W^\pm \rightarrow \tau^\pm \nu$	200	1.629×10^{-8}
$p p \rightarrow t\bar{t}$	2010	1.523×10^{-8}
$p p \rightarrow W\gamma$ (inclusive: ISR & FSR)	10	2.587×10^{-11}
$p p \rightarrow W + \text{jets}, W \rightarrow l\nu$	400	4.576×10^{-9}
$p p \rightarrow W + \text{jets}, W \rightarrow q\bar{q}$	3010	3.461×10^{-8}
$p p \rightarrow Z + \text{jets}, Z \rightarrow l^+ l^-$	50	4.922×10^{-10}
$p p \rightarrow Z + \text{jets}, Z \rightarrow \nu\bar{\nu}$	100	1.189×10^{-9}
$p p \rightarrow Z + \text{jets}, Z \rightarrow q\bar{q}$	1000	1.407×10^{-8}

Sophon Model Fine-tuning

Training sample:

- Generated from the decays of a hypothetical spin-0 resonance, $X^{0,\pm}$, with its mass m_X uniformly sampled between 15 and 500 GeV.

Three types of training samples:

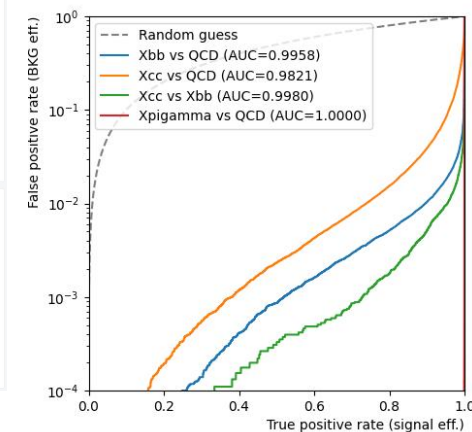
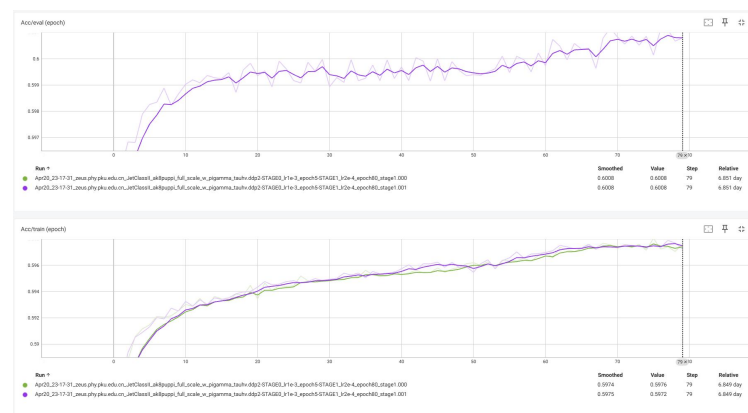
- $X \rightarrow \pi + \gamma$ (2 prong): ~24 million
- $X \rightarrow 3$ or 4 prong: ~20 million
- $X \rightarrow \tau_h + \nu$ (2 prong): ~50 million

Two steps of fine-tune:

- Step1: 5 epochs, LR: 1e-3, Freeze the whole network except for classification head.
- Step2: 80 epochs, LR: 2e-4 (gradually decayed after 70% epochs), unfreeze the entire network.

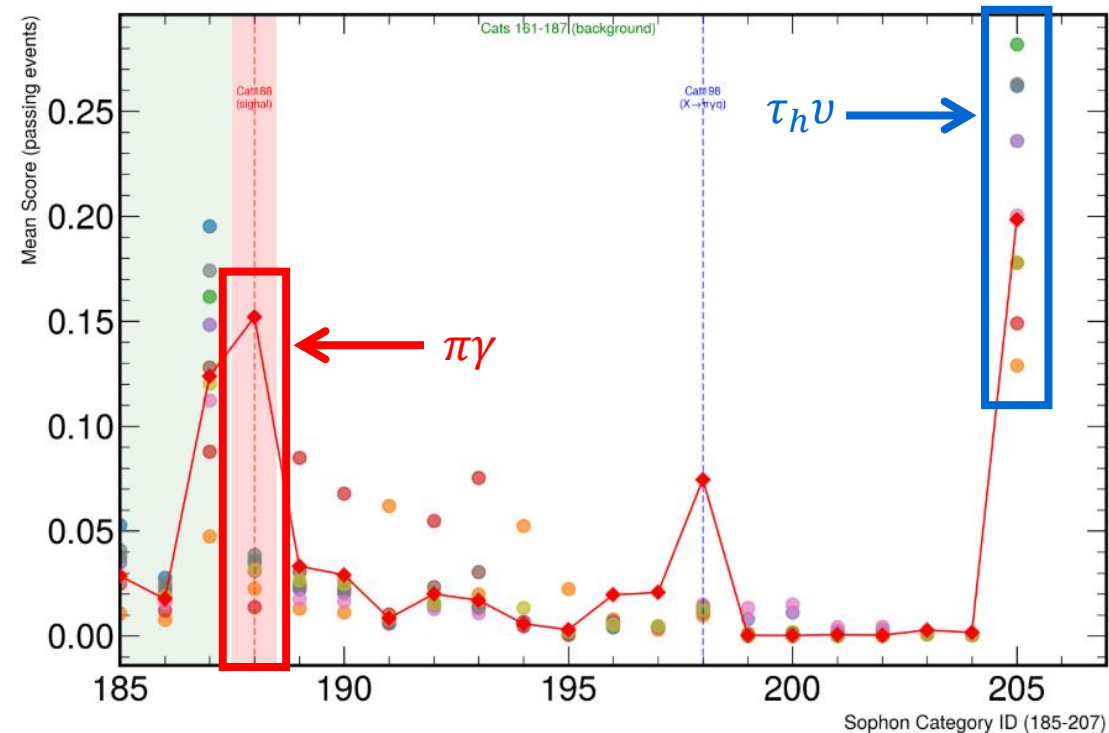
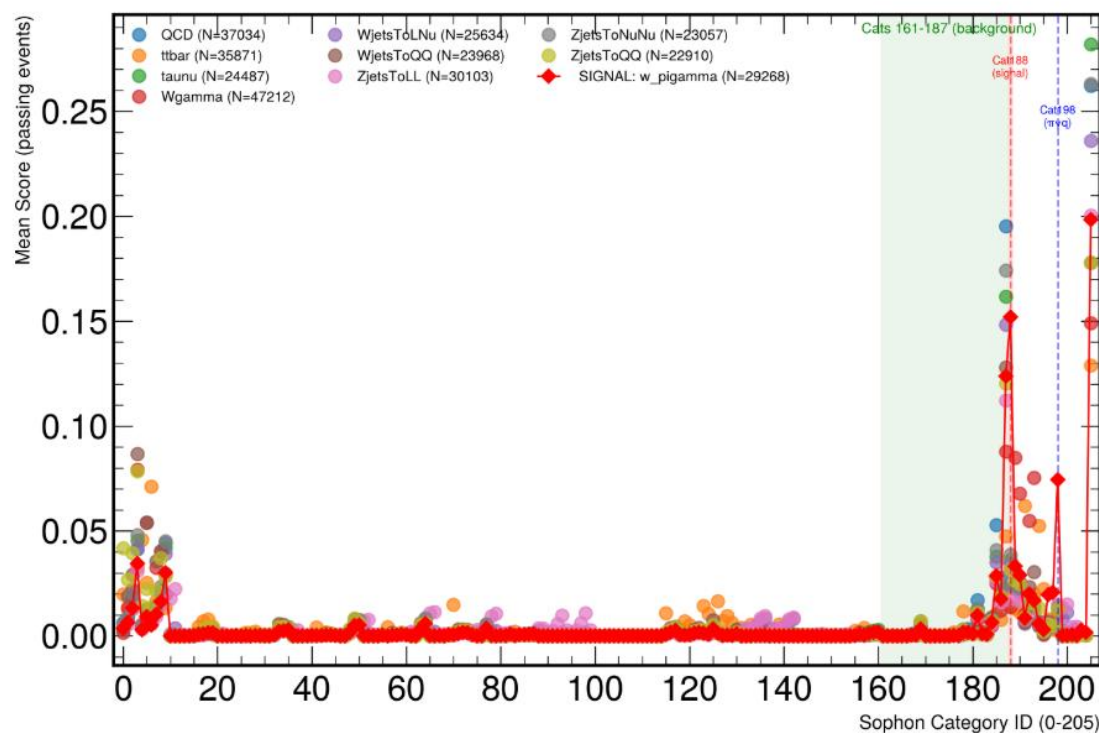
New Labels: The original 188 lack dedicated labels for the exclusive hadronic signatures targeted here.
18 NEW categories were added to Sophon model.

Major types	Index range	Label names
$X \rightarrow 2$ prong	188	$\pi\gamma$
$X \rightarrow 3$ or 4 prong	189 – 204	$\pi\gamma qq, \pi\gamma qs, \pi\gamma bq, \pi\gamma cq, \pi\gamma cs, \pi\gamma bc, \pi\gamma b, \pi\gamma c, \pi\gamma s,$ $\pi\gamma q, \pi\gamma e\nu, \pi\gamma\mu\nu, \pi\gamma\tau_e\nu, \pi\gamma\tau_\mu\nu, \pi\gamma\tau_h\nu, \pi\gamma\pi\gamma$
$X \rightarrow 2$ prong	205	$\tau_h\nu$



Sophon Model Fine-tuning

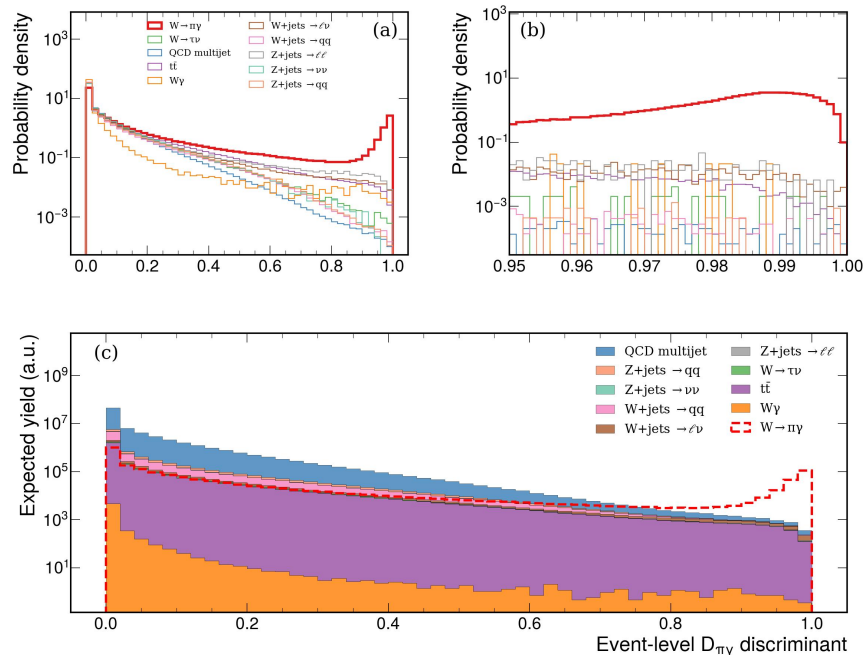
Each fat-jet passing the preselection is assigned 206 Sophon model scores, corresponding to the 206 categories, with values ranging from 0 to 1.



To better distinguish signal and background, we define a Discriminant as follows:

$$D_{\pi\gamma} = \frac{g_{X \rightarrow \pi\gamma}}{g_{X \rightarrow \pi\gamma} + g_{X \rightarrow \pi\gamma q} + g_{X \rightarrow \tau_h \nu} + g_{\text{QCD}} + \sum_{P_1} g_{P_1}}$$

To maximize the separation between $W \rightarrow \pi\gamma$ signal and various background categories.



First Priority labels P_1			
$X \rightarrow 2$ prong	$X \rightarrow ee$	$X \rightarrow \mu\mu$	
	$X \rightarrow \pi\gamma qq$	$X \rightarrow \pi\gamma qs$	$X \rightarrow \pi\gamma cq$
	$X \rightarrow \pi\gamma cs$	$X \rightarrow qqee$	$X \rightarrow qq\mu\mu$
$X \rightarrow 3$ or 4 prong	$X \rightarrow see$	$X \rightarrow qee$	$X \rightarrow gee$
	$X \rightarrow q\mu\mu$	$X \rightarrow bqev$	$X \rightarrow sqev$
	$X \rightarrow qqev$	$X \rightarrow bq\mu\nu$	$X \rightarrow sq\mu\nu$
	$X \rightarrow qq\mu\nu$		

Event-level selection: Fat-jet with highest p_T within each event. — Leading fat-jet.

Signal heavily concentrates at $D_{\pi\gamma} \approx 1$, while inclusive SM backgrounds are robustly pushed down.

BDT TRAINING:

- Exploit **global event kinematics** to enhance signal-to-background separation beyond the fat-jet level.
- Framework: **XGBoost classifier** (Sig vs. Inclusive Bkg)

Input Variables:

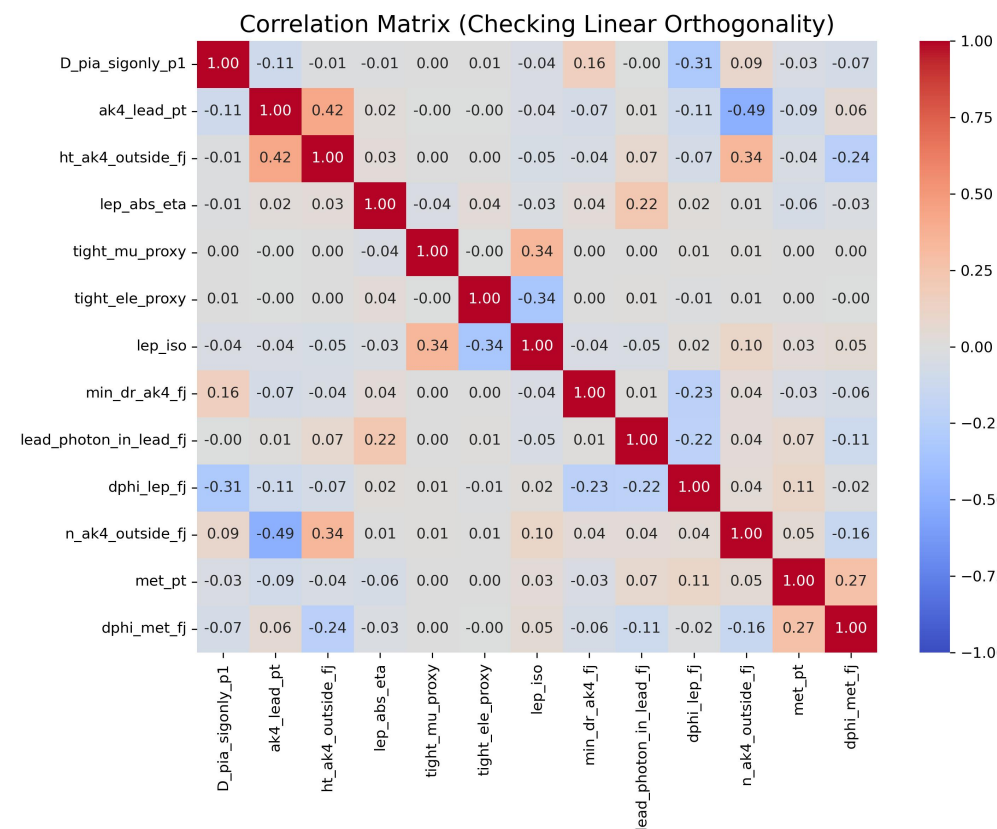
- Uses 12 macroscopic features, including missing transverse energy, ak4-jet kinematics, lepton proxies and angular correlations.

Orthogonality

- All BDT input variables must be explicitly orthogonal to the jet-level Sophon discriminant ($D_{\pi\gamma}$).

Event Weighting Strategy

- Physical weight: Scaled to physical cross-section with an expected luminosity of 450 fb^{-1} → **Testing phase**
- Training weights: Re-weighted so that the sum of weights for signal and background both equal 1.0 → **Training phase**

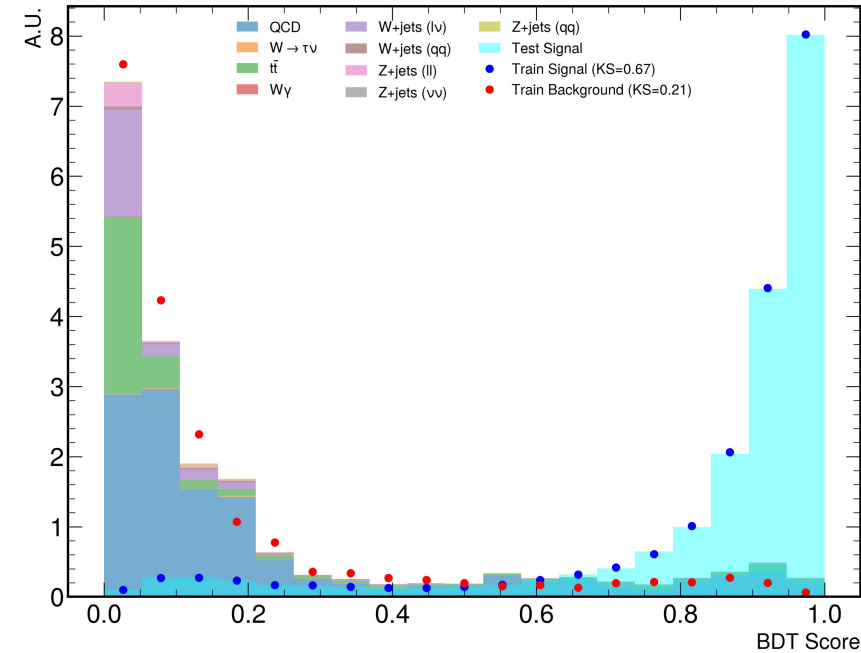


Analysis Strategies

Robustness Against Overtraining

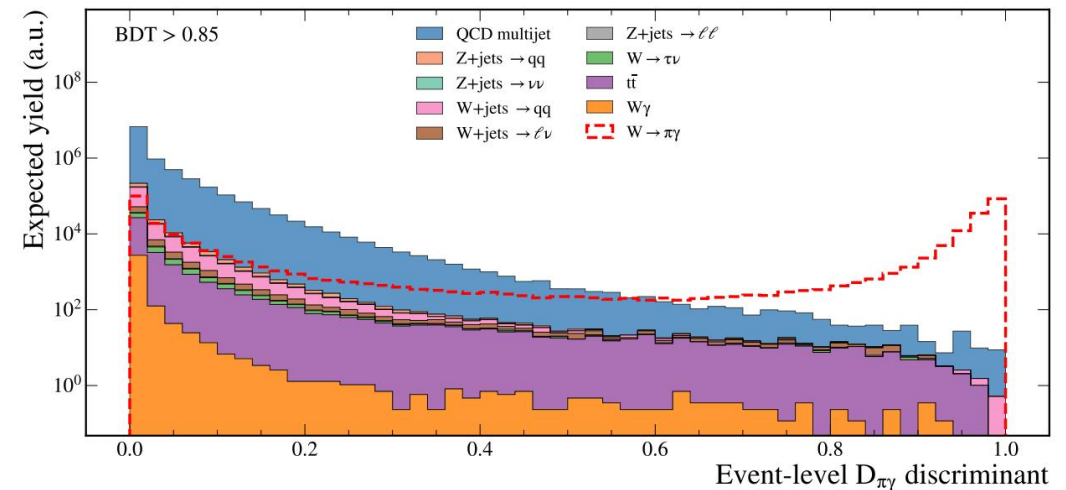
Kolmogorov-Smirnov (KS) Test

- Excellent agreement observed for both sig and bkg processes.



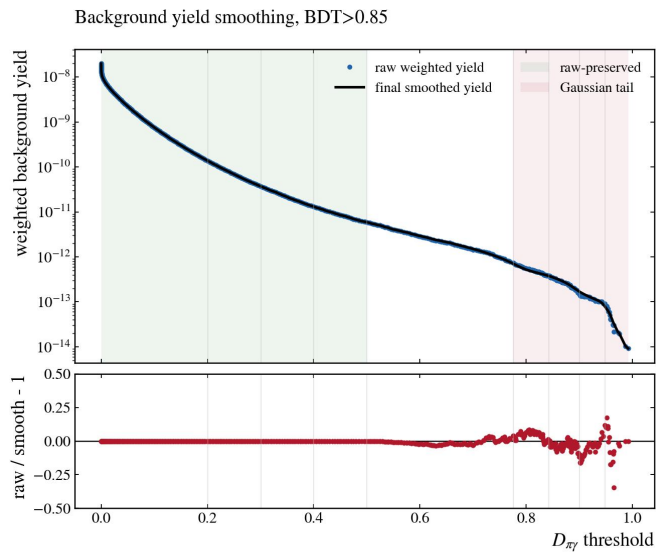
Overall Performance

- Achieves strong background rejection while maintaining high signal efficiency.
- The combined strategy (jet-level $D_{\pi\gamma}$ + event-level BDT) securely isolates the rare $W^{\pm} \rightarrow \pi^{\pm}\gamma$ topology.



Background Yield Smearing

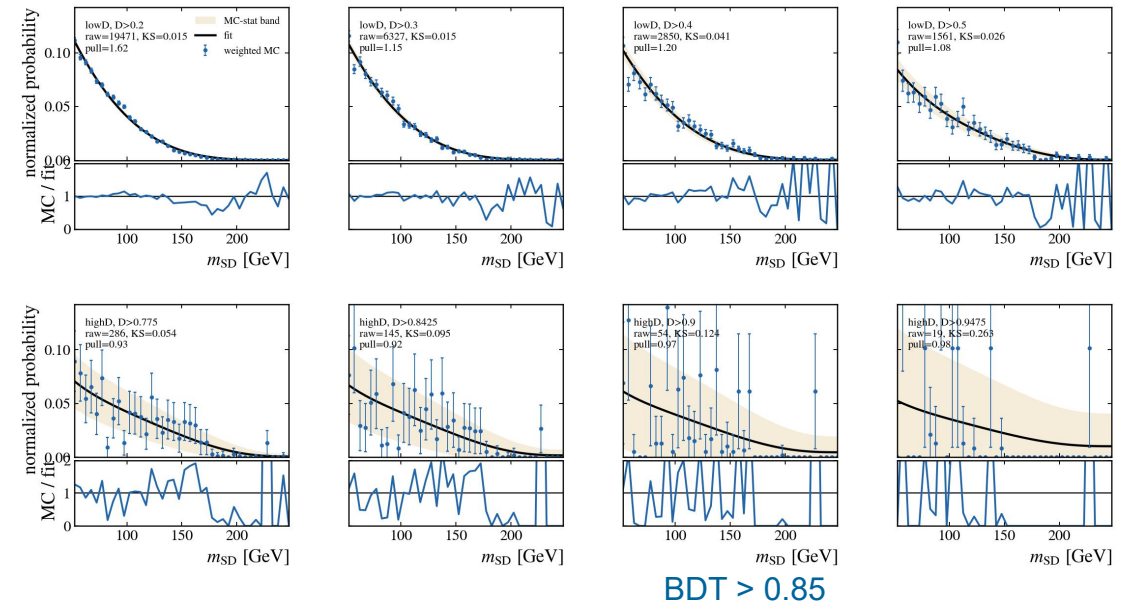
- Keep raw yield for $D_{\pi\gamma} < 0.5$ to preserve high-statistics region.
- Transition to fully smoothed tail for $D_{\pi\gamma} > 0.775$ using a cubic smoothstep.
- Enforce monotonic projection to guarantee physically consistent background suppression.



$$x = -\ln(1 - D_{\pi\gamma})$$

Mass Shape Modeling (m_{SD})

- Model the background m_{SD} shape using a 5-degree monotone Bernstein polynomial.
- Coefficients are strictly constrained via Sigmoid to prevent unphysical fake peaks.
- Shape parameters $\theta_j(D_{\pi\gamma})$ are modeled as quadratic functions, globally fitted to selected MC anchors



Shape uncertainty and sensitivity

- 80 Poisson toys and mono5 refits give a 1- σ shape band in $50 < m_{SD} < 250$ GeV.

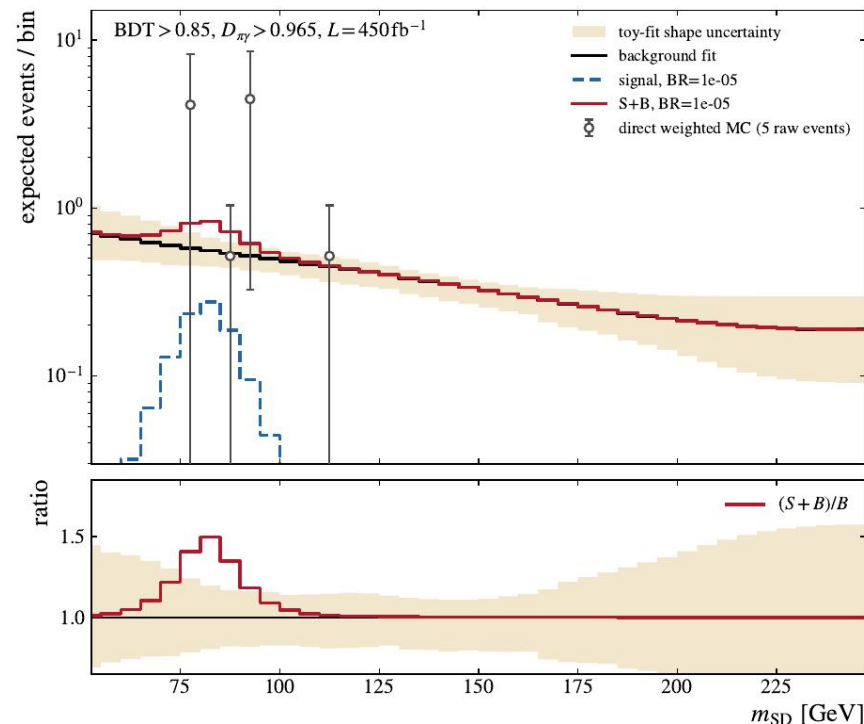
$$Z_{\text{toy}} = \sqrt{\sum_i \frac{S_i^2}{B_i + \sigma_{\text{toy},i}^2}}$$

- $B_{95} = 1.64 B_{\text{ref}} / Z_{\text{toy}}$
- Minimize B_{95} with $B_{\text{smooth}} \geq 10$ events

BDT > 0.85, $D_{\pi\gamma} > 0.965$, $L = 450 \text{ fb}^{-1}$

- Yields:** $B = 14.6$, $S = 0.115$ (at benchmark $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) = 1 \times 10^{-6}$).

Expected limit at 95% CL: $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 2.78 \times 10^{-5}$.



Summary and Future Prospects

Proof of principle in the boosted regime

A fine-tuned Sophon “hadron-in-fat-jet” tagger identifies boosted $W^\pm \rightarrow \pi^\pm \gamma$.

Jet-level $D_{\pi\gamma}$ and an approximately independent event-level BDT suppress backgrounds while retaining signal. Yield smoothing and a monotone Bernstein m_{SD} model describe the sparse tail.

Expected 95% CL: $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 2.78 \times 10^{-5}$ at 450 fb⁻¹

BDT > 0.85 and $D_{\pi\gamma} > 0.965$, including toy-fit shape uncertainty

Limitations and future prospects

Limited background MC statistics and sensitivity motivate improved triggers and additional channels.

What about $t\bar{t}$ channel with lepton trigger?



Worth a try

Thanks for your attention

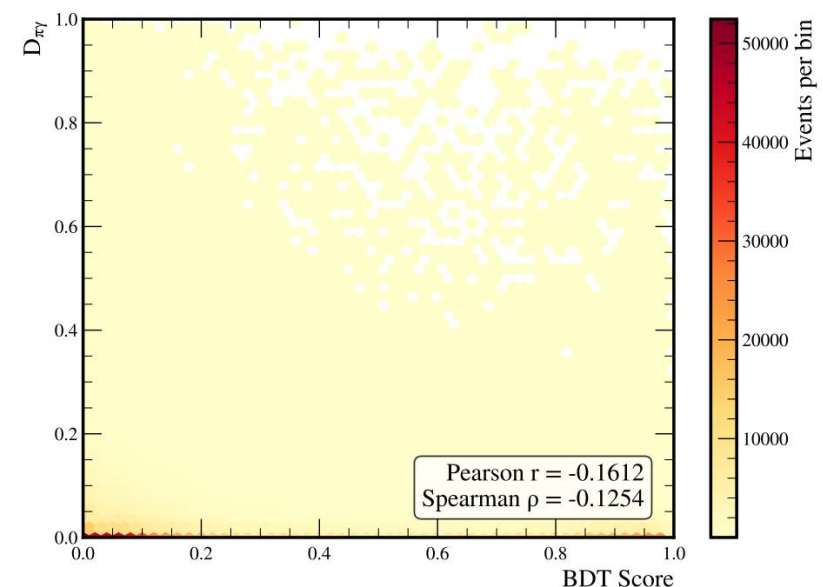
BACK UP

BDT Input Variables and Result

- Summary of the input variables used for the event-level BDT classifier and their physical meanings.
- All input variables must keep orthogonal with $D_{\pi\gamma}$.

- Two-dimensional distribution of the combined standard model backgrounds in the plane defined by the BDT score and the $D_{\pi\gamma}$ discriminant.

Variables	Physical Meaning
$p_{T, \text{AK4}}^{\text{lead}}$	p_T of the leading AK4-jet outside the fat-jet
$H_{T, \text{AK4}}$	H_T of all AK4-jets outside the fat-jet
$ \eta_l $	Absolute value of the lepton pseudorapidity
I_{μ}^{tight}	Selection flag indicating the presence of tight μ
I_e^{tight}	Selection flag indicating the presence of tight e
I_l^{rel}	Relative isolation of the lepton
ΔR_{min}	ΔR between any AK4-jet and the fat-jet
$p_{T, \text{lead}}^{\gamma, \text{lead}}$	p_T of the leading photon inside the leading fat-jet
$\Delta\phi(l, \text{FJ})$	$\Delta\phi$ between the lepton and the fat-jet
N_{AK4}	Multiplicity of AK4-jets outside the fat-jet
p_T^{miss}	Magnitude of the missing p_T
$\Delta\phi(p_T^{\text{miss}}, \text{FJ})$	$\Delta\phi$ between the p_T^{miss} and the fat-jet



Approximate independence between the event-level kinematics and the jet-level Sophon discriminant.

Background MC samples

- At BDT and $D_{\pi\gamma}$ working point, only 5 raw background MC events survive in the fit range.

Process	N_{raw}	Weighted yield	N_{eff}
QCD multijet	2	8.222	2.00
$t\bar{t}$	1	0.341	1.00
W +jets (qq)	2	1.035	2.00
Other backgrounds	0	0	0
Total direct background	5	9.598	2.67
Smoothed background model	—	14.644	—

