

Hadron-in-fat-jet AI tagging to detect rare decays such as $W^\pm \rightarrow \pi^\pm \gamma$

Linrui Chen^{*,} Tianyi Yang,[†] Zixun Kou, Zijian Wang, Youpeng Wu, Leyun Gao[,] and Qiang Li[‡]
School of Physics, Peking University, Beijing, People's Republic of China



(Received 11 June 2026; accepted 3 September 2026; published 21 September 2026)

We investigate a novel class of boosted-object signatures at the LHC, where a high- p_T fat jet contains an identifiable hadron or quarkonium state originating from rare or semiexclusive decays. Unlike conventional boosted jet studies, which focus on multiprong partonic substructure, our approach probes hybrid configurations such as $W^\pm \rightarrow \pi^\pm \gamma$, where a localized hadronic or quarkonium signal is embedded within a collimated jet. By fine-tuning the signature-oriented, pretrained Sophon artificial intelligence model optimized for large-radius jets, and combining it with an event-level boosted decision tree and a soft-drop-mass shape fit, we obtain an expected 95% confidence level upper limit of $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 2.78 \times 10^{-5}$ for 450 fb^{-1} in our nominal setup. This study serves as a first proof-of-principle demonstration of the hadron-in-fat-jet paradigm; substantial gains in sensitivity are expected from improved trigger strategies, additional production channels, and dedicated taggers, while the methodology itself is broadly applicable to a wide range of rare Standard Model processes and searches for light or exotic resonances at present and future collider experiments.

DOI: [10.1103/z3j5-12p5](https://doi.org/10.1103/z3j5-12p5)

I. INTRODUCTION

The LHC provides a high-energy and high-luminosity environment for precision tests of the Standard Model (SM) and searches for rare processes. In the boosted regime, the decay products of heavy particles are collimated and can be reconstructed as single large-radius jets. Jet-substructure techniques have therefore become essential tools for identifying such objects, primarily targeting partonic decays such as $W/Z \rightarrow q\bar{q}$, $H \rightarrow b\bar{b}$, and top quark decays, where the relevant information is encoded in multiprong radiation patterns [1–4].

However, much less attention has been paid to exclusive or semiexclusive decay modes, in which identifiable hadrons emerge inside the jet and carry additional flavor-specific information beyond conventional prong-based observables. Exclusive hadronic decays of the W boson constitute a particularly attractive example of this class, providing a unique laboratory for studying rare electroweak processes in the boosted regime. Their amplitudes involve both short-distance electroweak dynamics and long-distance QCD effects, making them sensitive probes of factorization,

meson distribution amplitudes, and hadronic form factors at the electroweak scale [5,6]. Phenomenological studies have also emphasized the potential of high-statistics LHC samples, in particular $t\bar{t}$ events, for probing rare exclusive W decays [7].

In the radiative decay $W^\pm \rightarrow \pi^\pm \gamma$, the W boson produces a light quark-antiquark pair that hadronizes into a charged pion, while a photon is radiated from the charged lines. The large W -boson mass makes this process a clean test bench for QCD factorization [5,8]. In the Standard Model, the branching fraction is predicted to be as low as $\mathcal{O}(10^{-9})$ [5]. When the W boson is produced with large transverse momentum, the $\pi^\pm$ and γ are emitted nearly collinearly, leading to a hadron-in-fat-jet topology with asymmetric energy sharing and nonperturbative internal structure. The leading-order Feynman diagrams contributing to the radiative exclusive decay $W^\pm \rightarrow \pi^\pm \gamma$ are illustrated in Fig. 1.

Experimentally, existing searches have reconstructed $W^\pm \rightarrow \pi^\pm \gamma$ as a resolved charged-hadron-plus-photon final state. Collider detector at fermilab, compact muon solenoid (CMS), and A Toroidal LHC ApparatuS (ATLAS) have all performed dedicated searches [9–11]. The CMS search used $t\bar{t}$ events, where one W boson provides a leptonic tag and the other is tested for the $\pi^\pm \gamma$ final state, and set $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 1.50 \times 10^{-5}$ at 95% confidence level (CL) [10]. The ATLAS search employed dedicated track-photon and diphoton triggers, obtaining $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 1.9 \times 10^{-6}$, $\mathcal{B}(W^\pm \rightarrow K^\pm \gamma) < 1.7 \times 10^{-6}$, and $\mathcal{B}(W^\pm \rightarrow \rho^\pm \gamma) < 5.2 \times 10^{-6}$ at 95% CL [11].

*Contact author: chenlinrui@whu.edu.cn

†Contact author: tyang99@pku.edu.cn

‡Contact author: qliphy0@pku.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

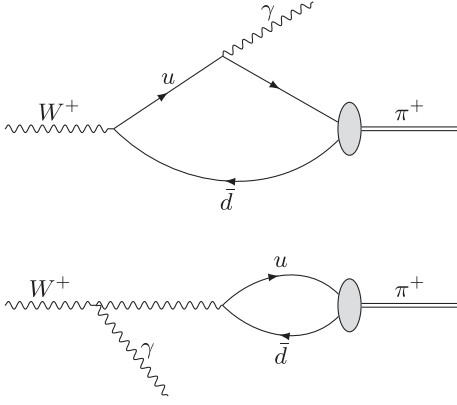


FIG. 1. Leading-order Feynman diagrams for the radiative exclusive decay $W^\pm \rightarrow \pi^\pm \gamma$.

While these resolved analyses have established the current experimental bounds, boosted topologies provide a complementary search regime. When the W boson carries a sufficiently large transverse momentum, the pion and photon become highly collimated and may be reconstructed within a single large-radius jet. In this case, jet-substructure techniques offer an alternative approach that exploits the internal structure of the merged jet, potentially extending the reach of rare hadronic decay searches and enabling sensitivity to a broader class of hadron-in-fat-jet signatures.

In this work, we study $W^\pm \rightarrow \pi^\pm \gamma$ in W + jets production, where the large inclusive production rate, combined with recoil against QCD radiation, provides access to boosted W bosons. We treat the merged $\pi^\pm \gamma$ system as a large-radius jet with radius parameter $R = 0.8$, and use jet-level information to identify the rare semiexclusive decay. Specifically, we use $W^\pm \rightarrow \pi^\pm \gamma$ as a benchmark to explore the fine-tuning of the existing Sophon framework for hadron-in-fat-jet signatures, and to provide a proof-of-principle demonstration of hadron-in-fat-jet tagging for the study of rare decay processes. We also provide a first phenomenological sensitivity estimate for boosted $W^\pm \rightarrow \pi^\pm \gamma$ based on realistic jet-level fast simulation. This boosted channel could be combined with existing resolved searches in future experimental analyses.

II. EXPERIMENTAL SETUP

Our study is based on simulated datasets for the LHC pp collisions at c.m. $\sqrt{s} = 13$ TeV. We defined the signal phase space by selecting events featured by a high- p_T fat jet (FJ) with a large electromagnetic energy fraction, optimized for capturing the boosted $W^\pm \rightarrow \pi^\pm \gamma$ signature. Our simulated datasets contained the boosted $W^\pm \rightarrow \pi^\pm \gamma$ signal from W + jets and various backgrounds including $t\bar{t}$, $W \rightarrow \tau\nu$, W + jets with hadronic and leptonic W decays, Z + jets with Z to l^+l^- , $\nu\bar{\nu}$, $q\bar{q}$, $W + \gamma$ and QCD multijet. Hard-scattering events were first generated at parton level with MadGraph5_aMC@NLO [12], which provides matrix-element

calculations and parton-level kinematics for both signal and background processes. To accurately model the proton structure, we employed the neural network parton distribution functions 3.1 next-to-next-to-leading-order (NNLO) parton distribution function (PDF) set [13]. The generated events were then interfaced to Pythia8 for parton showering, hadronization and particle decays, modeling the transition from a few-body final state to a realistic multihadron configuration [14].

For the fast detector simulation, we employed a custom Delphes 3.5.1 [15] configuration explicitly optimized for the boosted topology and machine learning applications, following the JetClass-II framework [16]. Standard isolated object identification and flavor-tagging algorithms were bypassed. Instead, the simulation incorporates an average of 50 pileup interactions and utilizes the Pileup Per Particle Identification (PUPPI) algorithm [17] to mitigate pileup contamination. The track resolutions are realistically smeared to match the Phase I CMS detector conditions. The PUPPI-weighted energy-flow objects are then clustered into large-radius ($R = 0.8$) fat jets using the anti- k_T algorithm named “AK8-jet” with the minimum p_T thresholds of 200 GeV, while jets with low radius ($R = 0.4$) are also maintained as “AK4-jet.” Crucially, the low-level kinematic and identity features of constituent particles within these fat jets are retained, providing the essential granular representation required for the jet-tagging model.

Table I details the number of generated events and the corresponding production cross sections used for the signal and background normalization. To enrich the boosted phase space and emulate the high- H_T trigger region, we impose a common reconstruction-level skimming requirement, $H_T^{\text{reco}} > 1050$ GeV, where H_T^{reco} is defined as the scalar sum of the transverse momenta of reconstructed jets [18]. This requirement is applied to all samples used in the analysis, including QCD multijet. For the signal and the non-QCD background samples, the selection is applied as

TABLE I. Summary of generated MC samples, detailing the decay processes, number of generated events (N_{gen}), and their corresponding cross sections.

Decay processes	$N_{\text{gen}} (\times 10^5)$	Cross section (mb)
Signal ($W^\pm \rightarrow \pi^\pm \gamma$)	99	1.632×10^{-8}
QCD	9902.5	9.047×10^{-6}
$W^\pm \rightarrow \tau^\pm \nu$	20	1.629×10^{-8}
$pp \rightarrow t\bar{t}$	201	1.523×10^{-8}
$pp \rightarrow W\gamma$ (initial state radiation & final state radiation)	1	2.587×10^{-11}
$pp \rightarrow W[\rightarrow l\nu] + \text{jets}$	40	4.576×10^{-9}
$pp \rightarrow W[\rightarrow q\bar{q}] + \text{jets}$	301	3.461×10^{-8}
$pp \rightarrow Z[\rightarrow l^+l^-] + \text{jets}$	5	4.922×10^{-10}
$pp \rightarrow Z[\rightarrow \nu\bar{\nu}] + \text{jets}$	10	1.189×10^{-9}
$pp \rightarrow Z[\rightarrow q\bar{q}] + \text{jets}$	100	1.407×10^{-8}

an offline skim after event generation and detector simulation. For the QCD multijet sample, whose inclusive rate is much larger, the same requirement is applied during the reconstruction/skimming workflow so that only events passing the requirement are retained. The finite Monte Carlo (MC) statistics, especially for the QCD sample, are accounted for by the background-yield smoothing and shape-extrapolation procedure described in Sec. V.

III. SOPHON MODEL FINE-TUNING

The Signature-Oriented Pre-training for Heavy-resonance ObservationN (Sophon) framework introduces a new paradigm for resonance searches at the LHC [16,19]. Conventional jet taggers are typically optimized for specific signal hypotheses, limiting their applicability to a narrow class of final states [20,21]. In contrast, Sophon employs a signature-oriented strategy, in which a deep neural network is pretrained on a large and diverse set of jet topologies to learn a universal representation of jet substructure.

The model is trained on a high-dimensional classification task spanning 188 jet categories, enabling it to encode both perturbative radiation patterns and nonperturbative features within a common latent space. This representation can be reused across analyses via transfer learning or by constructing discriminants directly from the model outputs, thereby providing a flexible and model-agnostic search framework.

This approach is particularly well suited to nonstandard jet topologies, such as the hadron-in-fat-jet configurations considered in this work, where an identified hadron or quarkonium is embedded within a boosted fat jet. Such signatures involve asymmetric energy sharing and nonperturbative dynamics that are not fully captured by traditional prong-based observables. In this context, the generalized feature space learned by Sophon offers a promising avenue to enhance sensitivity and systematically explore these novel signatures.

To identify the hadron-in-fat-jet topology, we fine-tune the original Sophon model, which initially featured $\mathcal{O}(100)$ categories but lacked specific hadronic final-state nodes. We explicitly incorporate 18 new categories (yielding a total of 206 categories for the fine-tuned model). These newly added nodes include signal and background final states that mimic the boosted $W^\pm \rightarrow \pi^\pm \gamma$ signature, such as $\tau_h \nu$,¹ $\pi \gamma q$ and $\pi \gamma \tau_h \nu$, along with other topologies susceptible to being misidentified as the signal. The training dataset was 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. This broad and uniform mass sampling is deliberately implemented to prevent the Sophon model from learning kinematic features correlated with the resonance mass. Consequently, it forces the model to strictly rely on the internal multiprong substructure and energy distribution of the fat jet for classification, ensuring a mass-decorrelated tagging performance.

The fine-tuning procedure is executed in two stages. First, the network backbone is frozen, while only the final classification head is updated. This stage employs a relatively large learning rate over a few epochs to warm up the weights of the newly introduced categories, ensuring compatibility with the pretrained weights. Subsequently, the entire network is unfrozen and trained with a reduced learning rate over a longer duration. This enables the model to adapt to the new categories while preserving its discriminative power for the original ones.

IV. ANALYSIS STRATEGIES

To isolate this rare signal from overwhelming Standard Model backgrounds, we employ a multistep selection strategy focused on both jet-level properties and event-level kinematics, paired with a robust background model. First, candidate large-radius jets are required to fall within a kinematic window of $p_T > 200$ GeV and $|\eta| < 2.5$. To ensure the jet substructure is consistent with a W boson decay, a minimum soft-drop-mass [22,23] requirement of 50 GeV is imposed. This threshold effectively suppresses low-mass QCD jets while retaining the majority of the W boson signal.

The core of our discrimination power resides in a fine-tuned Sophon model. Leveraging its signature-oriented architecture, we construct a dedicated, highly optimized discriminant, $D_{\pi\gamma}$, to maximize the separation between the $\pi\gamma$ signal and various background categories. The discriminant is formulated as

$$D_{\pi\gamma} = \frac{g_{X \rightarrow \pi\gamma}}{g_{X \rightarrow \pi\gamma} + g_{X \rightarrow \pi\gamma q} + g_{X \rightarrow \tau_h \nu} + g_{\text{QCD}} + \sum_{P_1} g_{P_1}}, \quad (1)$$

where g_i represents the output scores from the 206-category fine-tuned Sophon model. In this optimized formulation, the numerator is strictly driven by the pure $X \rightarrow \pi\gamma$ category to ensure high signal purity. The denominator acts as a targeted background penalty, comprising the signal node itself alongside primary mimicking topologies: the $X \rightarrow \pi\gamma q$ decays, hadronic τ decays $\tau_h \nu$, QCD multijet, and a collectively optimized subset of background categories denoted as P_1 .

The composition of the P_1 subset was systematically determined through a data-driven evaluation of background leakage in the high- $D_{\pi\gamma}$ -score region. In an earlier baseline iteration, the discriminant's penalty term was limited to the signal-like nodes, QCD multijet, and hadronic tau ($\tau_h \nu$) categories. While this baseline effectively suppressed the

¹Here $\tau_h \nu$ denotes an auxiliary AK8 fat-jet category used in the fine-tuning label space, rather than a standard narrow-cone hadronic- τ identification category such as those used in hadrons-plus-strips/DeepTau-style reconstruction. Ordinary boosted $\tau_h \nu$ configurations are largely suppressed by the soft-drop-mass requirement; this auxiliary node is included to penalize residual τ -like topologies that can leak into the $\pi\gamma$ signal-like region.

TABLE II. Summary of the first-priority background categories (P_1) incorporated into the denominator of the optimized $D_{\pi\gamma}$ discriminant. These specific two-prong and three- or four-prong decay topologies were systematically identified by analyzing the surviving Z + jets, W + jets and $t\bar{t}$ background events that bypassed initial baseline selections. Explicitly penalizing these nodes efficiently eliminates the dominant Standard Model configurations most prone to mimicking the boosted $W^\pm \rightarrow \pi^\pm\gamma$ signature.

First priority labels P_1			
$X \rightarrow$ two-prong	$X \rightarrow ee$	$X \rightarrow \mu\mu$	
$X \rightarrow$ three- or four-prong	$X \rightarrow \pi\gamma qq$	$X \rightarrow \pi\gamma qs$	$X \rightarrow \pi\gamma cq$
	$X \rightarrow \pi\gamma cs$	$X \rightarrow qqee$	$X \rightarrow qq\mu\mu$
	$X \rightarrow see$	$X \rightarrow qee$	$X \rightarrow gee$
	$X \rightarrow q\mu\mu$	$X \rightarrow bqev$	$X \rightarrow sqev$
	$X \rightarrow qqev$	$X \rightarrow bq\mu\nu$	$X \rightarrow sq\mu\nu$
	$X \rightarrow qq\mu\nu$		

multijet and $\tau_h\nu$ topologies, rigorous sensitivity scans—specifically evaluating background events that survived a stringent threshold of $D_{\pi\gamma}^{\text{baseline}} > 0.9$ —revealed significant residual contamination from Z + jets, W + jets, and $t\bar{t}$ processes in the signal region.

To understand and mitigate this leakage, we analyzed the fine-tuned Sophon model’s multiclass output distributions for these surviving background events. We observed that they spuriously peaked in specific nontarget categories. For instance, leptonic Z + jets decays predominantly populated pure electron- or muonlike two-prong nodes (e.g., $X \rightarrow ee, \mu\mu$), whereas $t\bar{t}$ and W + jets events artificially populated complex hybrid nodes, such as $X \rightarrow \pi\gamma qq$ or heavy-flavor configurations. Consequently, we explicitly grouped these highly populated “fake” nodes into the P_1 subset and introduced them into the denominator of the refined discriminant. This targeted penalty actively and efficiently eliminated the dominant Standard Model contaminants most prone to mimicking the boosted $W^\pm \rightarrow \pi^\pm\gamma$ signature. The detailed composition of the strategically targeted P_1 subset is summarized in Table II.

To evaluate the efficacy of the Sophon model fine-tuning and the physical soundness of the $D_{\pi\gamma}$ form, receiver operating characteristic (ROC) curves were generated for the three primary background processes that most heavily mimic the $W^\pm \rightarrow \pi^\pm\gamma$ signal—namely QCD, $t\bar{t}$, and $W \rightarrow \tau\nu$ —as depicted in Fig. 2. The corresponding area under the curve metrics indicate a robust and stable background rejection capability across diverse fake-signal topologies, further validating the effectiveness of the targeted penalty terms incorporated into the $D_{\pi\gamma}$ formulation.

While the fine-tuned Sophon model provides exceptional discrimination power at the fat-jet level, the ultimate sensitivity of the analysis can be further enhanced by exploiting eventwise correlations. To this end, an event-level boosted decision tree (BDT) classifier is implemented using the XGBoost framework [24]. The BDT is trained to

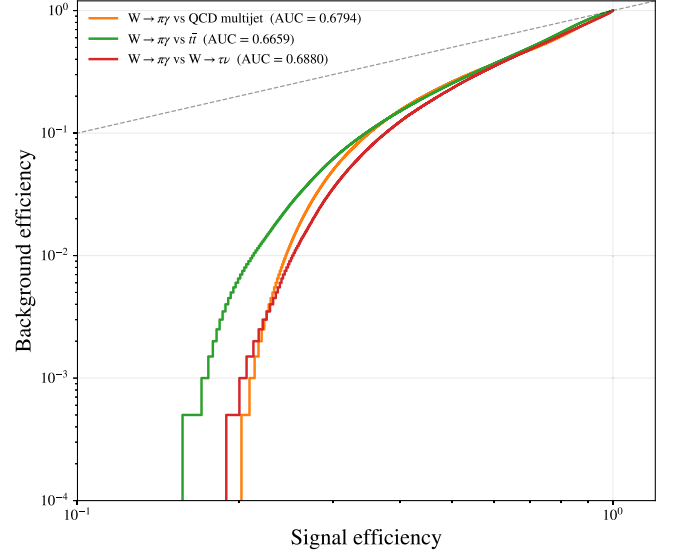


FIG. 2. ROC curves evaluating the tagging performance of the $D_{\pi\gamma}$ discriminant. The plot illustrates the signal efficiency for $W^\pm \rightarrow \pi^\pm\gamma$ against the background efficiencies of QCD multijet, $t\bar{t}$ and $W \rightarrow \tau\nu$.

separate the $W^\pm \rightarrow \pi^\pm\gamma$ signal from an inclusive mixture of all standard model backgrounds.

The input features for the BDT comprise macroscopic event kinematics, object multiplicities, and angular correlations, whose variable names and physical meanings are listed in Table III. Crucially, to facilitate reliable data-driven background estimation and in-situ calibration, these input variables are explicitly chosen to be orthogonal (i.e., highly uncorrelated) to the jet-level Sophon discriminant $D_{\pi\gamma}$. This orthogonality condition ensures that stringent selections on the BDT score will not artificially sculpt the $D_{\pi\gamma}$ background distributions, preserving the viability of control regions.

TABLE III. Summary of the input variables used for the event-level BDT classifier and their physical meanings.

Variables	Physical meaning
$p_{T,AK4}^{\text{lead}}$	p_T of the leading AK4-jet outside the fat-jet
$H_{T,AK4}$	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
$\Delta R_{\min}^{\gamma,\text{lead}}$	ΔR between any AK4-jet and the fat jet
$p_{T,\text{lead}}^{\gamma}$	p_T of the leading photon inside the leading fat jet
$\Delta\phi(l, FJ)$	$\Delta\phi$ 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
$\Delta\phi(p_T^{\text{miss}}, FJ)$	$\Delta\phi$ between the p_T^{miss} and the fat jet

Special attention is given to the event weighting during the BDT training process. Initially, both signal and background events are scaled using physical weights, corresponding to an expected integrated luminosity of 450 fb^{-1} . To prevent the overwhelmingly large background yields from dominating the gradient boosting process, a dedicated training weight is applied, normalizing the total sum of weights for both the signal and the inclusive background to unity. Conversely, the model evaluation on the testing dataset strictly utilizes the original physical weights to reflect realistic experimental conditions.

The robustness of the BDT training is rigorously evaluated to rule out any potential overtraining. Kolmogorov-Smirnov (KS) tests are performed on the BDT output score distributions, comparing the training and testing datasets [25]. As illustrated in Fig. 3, the BDT scores for both the signal and background processes exhibit excellent agreement between the training and testing samples. This confirms that the XGBoost classifier is well behaved, highly generalizable, and ready for final signal extraction.

To justify the joint application of the event-level BDT and the jet-level $D_{\pi\gamma}$ discriminant, it is essential to rigorously evaluate their mutual correlation. If these two discriminators were highly correlated, imposing a stringent threshold on the BDT score would unintentionally distort the background distributions of some feature variables. Such phase-space distortion would violate the fundamental assumption of shape invariance across different selection regions.

To verify the independence of these two classifiers, we analyze the two-dimensional distribution of the combined

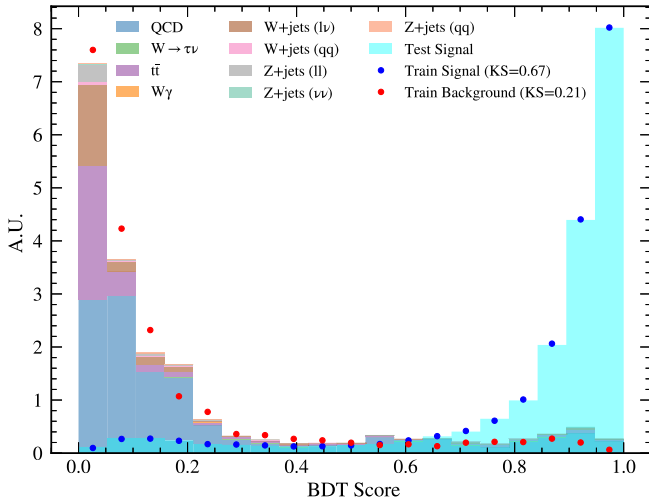


FIG. 3. Normalized BDT score distributions for the $W^\pm \rightarrow \pi^\pm \gamma$ signal and the inclusive SM background. The plot compares the outputs evaluated on the training dataset (points) and the testing dataset (histograms). The consistency between the training and testing distributions, quantitatively supported by the KS test results, demonstrates the absence of overtraining in the XGBoost classifier.

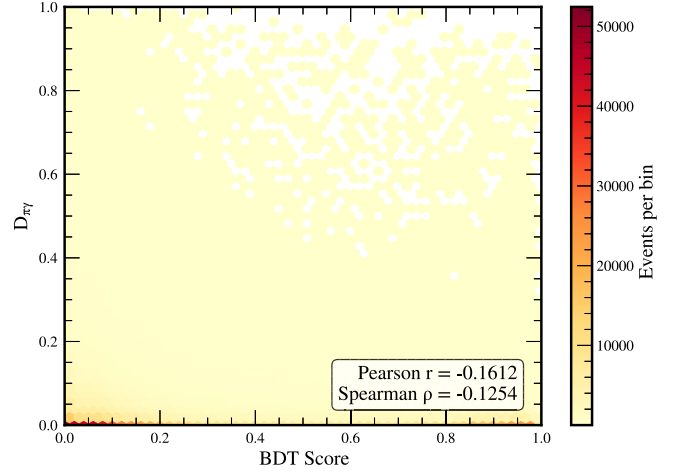


FIG. 4. Two-dimensional distribution of the combined SM backgrounds in the plane defined by the BDT score and the $D_{\pi\gamma}$ discriminant. The near-zero correlation coefficients demonstrate the approximate independence between the event-level kinematics and the jet-level Sophon discriminant, supporting their combined use in the final signal extraction.

Standard Model backgrounds in the plane of the BDT score and the Sophon discriminant $D_{\pi\gamma}$, as presented in Fig. 4. The Pearson correlation coefficient is $r = -0.1612$ [26], with a corresponding Spearman's rank correlation coefficient of $\rho = -0.1254$ [27]. These values indicate a very weak negative correlation, demonstrating that the macroscopic event kinematics captured by the BDT and the microscopic, internal jet-substructure features learned by Sophon remain largely decoupled.

This near-zero correlation ensures that applying a selection cut on the BDT score does not bias the background shape in either the m_{SD} or other feature variables. Consequently, this independence provides a solid and unbiased foundation for the subsequent signal extraction.

For the BDT score cut, we use a fixed BDT working point, $\text{BDT} > 0.85$. At this working point, most of the

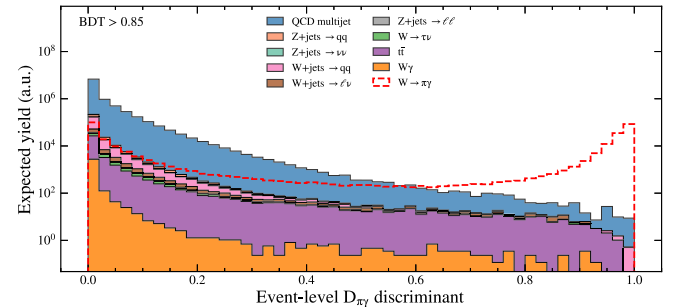


FIG. 5. Stacked $D_{\pi\gamma}$ distribution weighted by their respective physical cross sections under $\text{BDT} > 0.85$. The pronounced signal peak near $D_{\pi\gamma} \approx 1$ demonstrates the exceptional discrimination power and targeted background suppression achieved by incorporating the P_1 penalty topologies. Here we set the hypothetical branching fraction $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) = 1$.

signal is preserved. Meanwhile, to ensure sufficient statistics for the subsequent background distribution extrapolation, more background is retained compared to a more stringent BDT cut, without excessively degrading the signal significance. Under the $\text{BDT} > 0.85$ requirement, the optimized discriminant $D_{\pi\gamma}$ still possesses a powerful capability to distinguish the $W^\pm \rightarrow \pi^\pm \gamma$ signal from the dominant Standard Model backgrounds. This impressive separation performance is clearly demonstrated in Fig. 5, which presents the $D_{\pi\gamma}$ distributions for both signal and background processes, normalized to an integrated luminosity of 450 fb^{-1} . The signal events are overwhelmingly clustered in the ultrahigh score region near unity, whereas the massive QCD multijet background drops sharply by several orders of magnitude.

V. SIGNAL EXTRACTION

Due to the extremely limited background statistics under the high $D_{\pi\gamma}$ score region, we implement background shape extrapolation for the final signal extraction. We scan thresholds on $D_{\pi\gamma}$ under $\text{BDT} > 0.85$. Instead of relying on a single cut-and-count signal region, we construct a binned template in the soft-drop-mass range $50 < m_{\text{SD}} < 250 \text{ GeV}$. For each $D_{\pi\gamma}$ threshold, the total background normalization is first obtained from the weighted simulated samples and then smoothed as a function of $x = -\ln(1 - D_{\pi\gamma})$ using a Gaussian kernel, with the high-statistics, low- $D_{\pi\gamma}$ region preserved and a monotonic projection imposed, as shown in Fig. 6. The normalized m_{SD} shape is modeled by a monotone Bernstein polynomial of degree five,

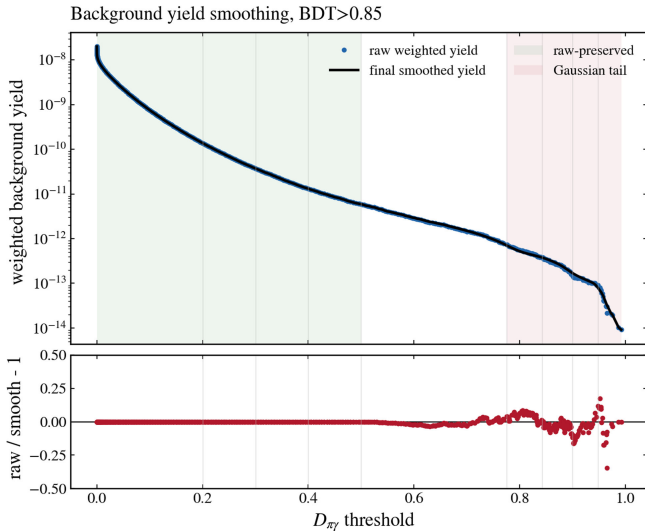


FIG. 6. Background-yield smoothing as a function of the $D_{\pi\gamma}$ threshold for the final $\text{BDT} > 0.85$ selection. The raw weighted yield is preserved in the high-statistics, low- $D_{\pi\gamma}$ region, while the sparse high- $D_{\pi\gamma}$ tail is smoothed in $x = -\ln(1 - D_{\pi\gamma})$ and projected to be monotonic.

$$f(m_{\text{SD}}|D_{\pi\gamma}) = \sum_{k=0}^5 c_k(D_{\pi\gamma}) B_k(m_{\text{SD}}), \quad (2)$$

where the coefficients satisfy $c_0 \geq c_1 \geq \dots \geq c_5 \geq 0$ through the parametrization $c_k = c_{k-1} \text{sigmoid}(\theta_k)$. The $D_{\pi\gamma}$ dependence is captured by quadratic trajectories, $\theta_k(D_{\pi\gamma}) = a_k D_{\pi\gamma}^2 + b_k D_{\pi\gamma} + d_k$, fitted globally to weighted background templates with sufficient simulated statistics. Figure 7 shows representative low- and high- $D_{\pi\gamma}$ validation points for this smooth background model.

The expected sensitivity is evaluated from the binned signal and background templates as

$$Z_{\text{shape}} = \left[\sum_i \frac{N_{S,i}^2}{N_{B,i} + \sigma_{\text{shape},i}^2} \right]^{1/2}, \quad (3)$$

where $N_{S,i}$ and $N_{B,i}$ are the expected signal and background yields in the i th m_{SD} bin for an integrated luminosity of 450 fb^{-1} [28]. The signal is normalized to a reference branching fraction $\mathcal{B}_{\text{ref}}(W^\pm \rightarrow \pi^\pm \gamma) = 10^{-6}$. The term $\sigma_{\text{shape},i}$ is obtained from toy-data refits of the monotone Bernstein function and represents the shape-fit uncertainty. The corresponding expected upper limit is estimated as $\mathcal{B}_{95} = 1.64 \mathcal{B}_{\text{ref}} / Z_{\text{shape}}$.

To avoid selecting a working point with a negligibly small expected background, we require the smoothed background prediction to satisfy $B_{\text{smooth}} \geq 10$ events. This condition is imposed on the continuous smoothed-yield model, rather than on the direct weighted yield of the sparse surviving MC events. The selected working point is found at $\text{BDT} > 0.85$ and $D_{\pi\gamma} > 0.965$. At this final working point, only five raw background MC events survive in the fit range ($50 < m_{\text{SD}} < 250 \text{ GeV}$). These events consist of two QCD multijet events, one $t\bar{t}$ event, and two $W[\rightarrow q\bar{q}] + \text{jets}$ events. The detailed information of these survived MC events are listed in Table IV. Here, N_{raw} denotes the unweighted number of simulated events surviving the final selection. No raw events from the other simulated background samples survive the full selection. After normalization to 450 fb^{-1} , the five events correspond

TABLE IV. Simulated event statistics after the final selection, $\text{BDT} > 0.85$ and $D_{\pi\gamma} > 0.965$, in the fit range $50 < m_{\text{SD}} < 250 \text{ GeV}$. The expected yields are normalized to 450 fb^{-1} .

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

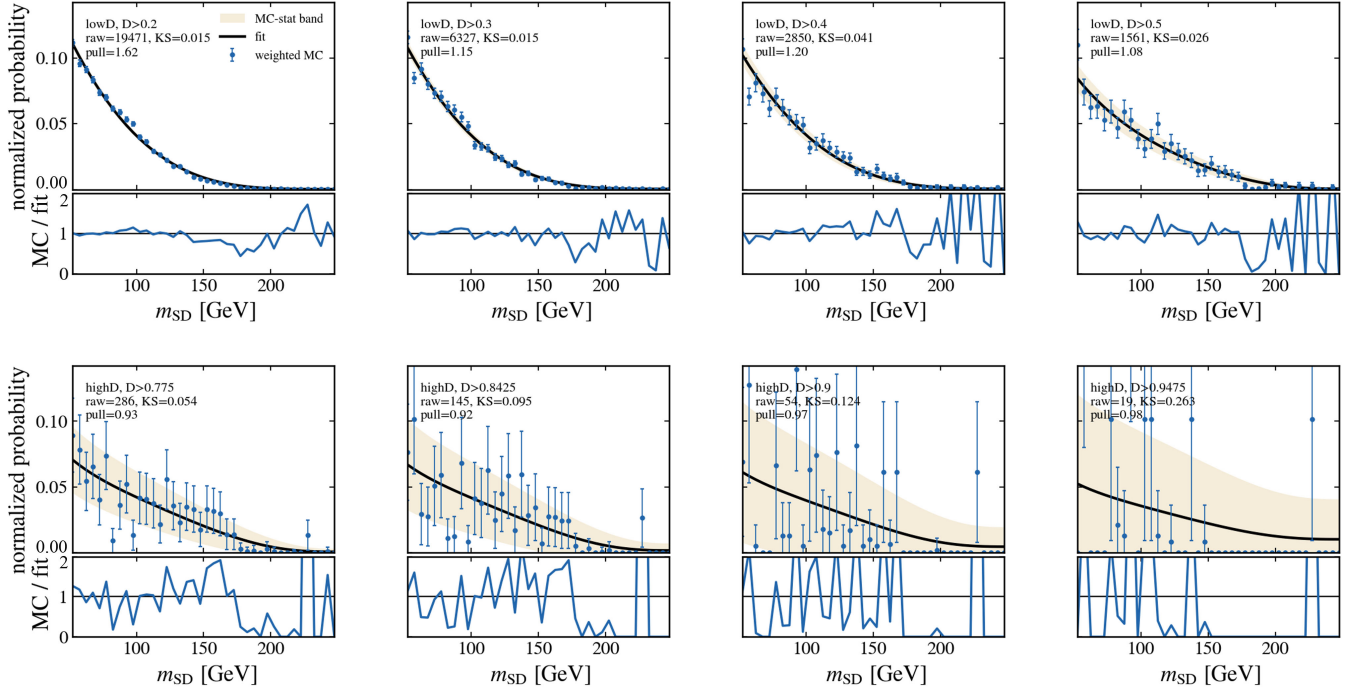


FIG. 7. Validation of the final m_{SD} background shape model at representative $D_{\pi\gamma}$ thresholds. Four low- $D_{\pi\gamma}$ and four high- $D_{\pi\gamma}$ points are shown. In each panel, the weighted simulated background template is compared with the monotone Bernstein fit, and the lower ratio panel displays the agreement relative to the fitted shape. The high- $D_{\pi\gamma}$ region has limited simulated statistics, which motivates the smooth trajectory model and the additional uncertainty treatment used in the sensitivity estimate.

to a direct weighted yield of 9.60 events and an effective MC sample size of

$$N_{\text{eff}} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} = 2.67. \quad (4)$$

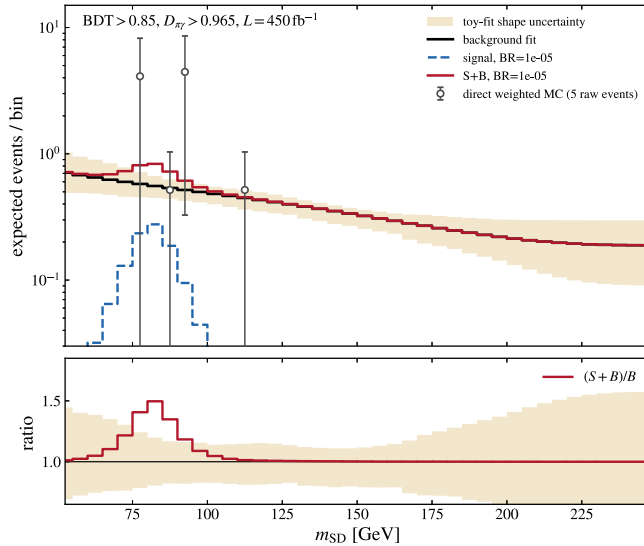


FIG. 8. Final m_{SD} signal and background templates for the $N_B \geq 10$ working point, $BDT > 0.85$ and $D_{\pi\gamma} > 0.965$. The background prediction includes the toy-fit shape uncertainty band, while the signal overlay is shown with an enhanced normalization of $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) = 10^{-5}$ for visibility.

The difference between $N_{\text{raw}} = 5$ and $N_{\text{eff}} = 2.67$ results from the process-dependent event weights, particularly the comparatively large weights of the two surviving QCD events. The finite-MC statistical uncertainty of the direct weighted yield is $\sqrt{\sum_i w_i^2} = 5.87$ events. At the same working point, the smoothed background model predicts 14.644 events. The difference of 5.046 events is therefore approximately 0.86 times the statistical uncertainty of the direct weighted MC estimate, indicating that the two estimates are compatible within the limited MC statistics. The difference is expected because the smoothed prediction is obtained from the continuous yield-smoothing procedure, rather than from a direct histogram of the five surviving events; the corresponding m_{SD} distribution is provided by the global shape-extrapolation model.

Including the toy-fit shape uncertainty, this selection gives an expected upper limit of $\mathcal{B}_{95}(W^\pm \rightarrow \pi^\pm \gamma) = 2.78 \times 10^{-5}$ for 450 fb $^{-1}$. The corresponding signal and background templates are displayed in Fig. 8.

VI. CONCLUSION

In summary, we have presented a novel study on the hadron-in-fat-jet topology at the LHC, exploring a qualitatively new regime that connects jet substructure with exclusive hadronic states. Crucially, this work represents the first phenomenological sensitivity study of the rare decay $W^\pm \rightarrow \pi^\pm \gamma$ in the boosted regime, utilizing realistic fast detector simulation to model the complex signal and

backgrounds. Unlike traditional substructure techniques targeting multiprong partonic decays, our approach tackles hybrid configurations where a localized hadronic state is embedded within a single collimated jet.

To isolate this signature, we developed a comprehensive discrimination strategy that marks the first exploration of the Sophon model's fine-tuning methodology and its novel application to rare decay searches via jet tagging. By combining the fine-tuned, signature-oriented jet-level discriminant $D_{\pi\gamma}$ with an orthogonal event-level XGBoost BDT and a data-driven soft-drop-mass shape fit, we established an expected 95% CL upper limit of $\mathcal{B}(W^\pm \rightarrow \pi^\pm \gamma) < 2.78 \times 10^{-5}$ for an integrated luminosity of 450 fb^{-1} .

This framework enables further access to rare decay modes that are challenging in resolved topologies due to limited acceptance or overwhelming backgrounds, potentially enhancing sensitivity to both Standard Model measurements and new physics scenarios. Further improvements are expected by exploiting semileptonic $t\bar{t}$ events selected with single-lepton triggers, which provide a clean and efficiently triggered source of boosted W bosons while substantially suppressing multijet backgrounds.

The additional kinematic constraints from the top-quark system are anticipated to enhance signal purity and overall sensitivity. Beyond the specific $W^\pm \rightarrow \pi^\pm \gamma$ benchmark, the proposed hadron-in-fat-jet strategy opens a new avenue for searching for light resonances decaying into a pion and a photon, whose highly boosted decay products may merge into a single jet and evade conventional reconstruction techniques. Our results demonstrate the potential of hadron-in-fat-jet tagging as a complementary paradigm to traditional boosted-object identification and establish a general framework for probing rare electroweak processes and light resonances at the LHC.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China under Grant No. 12325504.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

-
- [1] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, Substructure of high- p_T Jets at the LHC, *Phys. Rev. D* **79**, 074017 (2009).
 - [2] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-object Phenomenology*, Vol. 958 (Springer, New York, 2019).
 - [3] A. J. Larkoski, I. Moult, and B. Nachman, Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning, *Phys. Rep.* **841**, 1 (2020).
 - [4] R. Kogler *et al.*, Jet substructure at the Large Hadron Collider: Experimental review, *Rev. Mod. Phys.* **91**, 045003 (2019).
 - [5] Y. Grossman, M. König, and M. Neubert, Exclusive radiative decays of W and Z bosons in QCD factorization, *J. High Energy Phys.* **04** (2015) 101.
 - [6] T. Melia, Exclusive hadronic W decay: $W \rightarrow \pi\gamma$ and $W \rightarrow \pi^+\pi^+\pi^-$, *Nucl. Part. Phys. Proc.* **273**, 2102 (2016).
 - [7] M. Mangano and T. Melia, Rare exclusive hadronic W decays in a $t\bar{t}$ environment, *Eur. Phys. J. C* **75**, 258 (2015).
 - [8] G. P. Lepage and S. J. Brodsky, Exclusive processes in quantum chromodynamics: Evolution equations for hadronic wave functions and the form-factors of mesons, *Phys. Lett.* **87B**, 359 (1979).
 - [9] T. Aaltonen *et al.* (CDF Collaboration), Search for the rare radiative decay: $W \rightarrow \pi\gamma$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$, *Phys. Rev. D* **85**, 032001 (2012).
 - [10] A. M. Sirunyan *et al.* (CMS Collaboration), Search for the rare decay of the W boson into a pion and a photon in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$, *Phys. Lett. B* **819**, 136409 (2021).
 - [11] G. Aad *et al.* (ATLAS Collaboration), Search for the exclusive W boson hadronic decays $W^\pm \rightarrow \pi^\pm \gamma$, $W^\pm \rightarrow K^\pm \gamma$ and $W^\pm \rightarrow \rho^\pm \gamma$ with the ATLAS detector, *Phys. Rev. Lett.* **133**, 161804 (2024).
 - [12] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *J. High Energy Phys.* **07** (2014) 079.
 - [13] R. D. Ball *et al.*, Parton distributions from high-precision collider data, *Eur. Phys. J. C* **77**, 663 (2017).
 - [14] T. Sjostrand, S. Mrenna, and P. Z. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
 - [15] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2017) 057.
 - [16] C. Li *et al.*, Accelerating resonance searches via signature-oriented pre-training, *arXiv:2405.12972*.
 - [17] D. Bertolini, P. Harris, M. Low, and N. Tran, Pileup per particle identification, *J. High Energy Phys.* **10** (2014) 059.

- [18] A. Hayrapetyan, A. Tumasyan, W. Adam, J. Andrejkovic, L. Benato, T. Bergauer, S. Chatterjee, K. Damanakis, M. Dragicevic, P. Hussain *et al.*, Performance of the cms high-level trigger during lhc run 2, *J. Instrum.* **19**, P11021 (2024).
- [19] Y. Zhao, C. Li, A. Agapitos, D. Fu, L. Gao, Y. Mao, and Q. Li, Novel $|V_{cb}|$ extraction method via boosted bc -tagging with In-situ calibration, [arXiv:2503.00118](#).
- [20] C. Collaboration *et al.*, Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, [arXiv:2004.08262](#).
- [21] H. Qu and L. Gouskos, Jet tagging via particle clouds, *Phys. Rev. D* **101**, 056019 (2020).
- [22] M. Dasgupta, A. Fregoso, S. Marzani, and G.P. Salam, Towards an understanding of jet substructure, *J. High Energy Phys.* **09** (2013) 029.
- [23] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft drop, *J. High Energy Phys.* **05** (2014) 146.
- [24] T. Chen and C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), pp. 785–794.
- [25] F. Takahashi *et al.* (Particle Data Group), Review of particle physics, *Int. J. Mod. Phys. A* **41**, 2630011 (2026).
- [26] K. Pearson, VII. Mathematical contributions to the theory of evolution—III. Regression, heredity, and panmixia, *Phil. Trans. R. Soc. A* **187**, 253 (1896).
- [27] C. Spearman, The proof and measurement of association between two things, *Am. J. Psychol.* **15**, 72 (1904).
- [28] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71**, 1554 (2011).