Intrinsic geometry of collider observations and forman Ricci curvature

Jyotiranjan Beuria[*]

Indian Institute of Technology Mandi, Himachal Pradesh, India
and IKS Research Center, ISS Delhi, Delhi, India

We study the global and local topological properties of multilepton patterns reconstructed at the detectors. We investigate the sensitivity of Forman Ricci curvature distributions and persistent homology features to kinematic cuts, integrated luminosity, and scales of maximum filtration. We find that these topological properties are efficient enough in discriminating the beyond Standard Model (BSM) scenarios from the SM background, when the BSM scenarios possess a massive invisible particle. We also find that the topological properties exhibit a scaling behavior with integrated luminosity. This exploratory study suggests that the topological features can potentially supplement the traditional cut-and-count analyses in search of new physics.

I. INTRODUCTION

The Standard Model (SM) of particle physics has proven highly effective in explaining the behavior of elementary particles. Identifying the Higgs boson, a neutral scalar particle, marked a significant achievement for the SM [1,2]. Despite its success, the SM needs to address several crucial unanswered questions. Consequently, searching for models beyond the SM has prompted various upgrades to the Large Hadron Collider (LHC) and the emergence of other future colliders.

The exploration of the beyond the Standard Model (BSM) phenomenology involves constraining the parameter space of new physics models. This is achieved through collider simulations at the parton, hadron, and detector levels, incorporating signal and background analyses. Usually, it is accepted that collider events are independent and, to a great extent, identically distributed as well. However, the reconstructed objects at the detector might preferentially populate some portion of the phase space depending on the available phase space for the decaying heavy BSM particles. Thus, physicists have introduced a range of kinematic variables [3] for this purpose, relying on kinematic cuts applied on an event-by-event basis. The traditional cut-and-count method has been immensely successful in excluding several BSM scenarios. However, the intrinsic geometric structure of the phase space of the reconstructed particles featuring higher-order topological properties might add to ongoing searches at the colliders. These, coupled with traditional approaches, can potentially draw stronger exclusion limits even at low values of integrated luminosity.

In recent years, several efforts have been made to analyze the global patterns of the collider observations using techniques such as Voronoi and Delaunay tessellations [4–6], network distance metrics [7], and machine learning based classifications [7–15]. Recently, topological data analysis (TDA) [16–26] is being increasingly used in the broader field of data science for examining the intrinsic global properties of the system. However, its utility in discriminating the BSM signals from the SM background is not well understood.

TDA is a mathematical and computational approach that applies tools from algebraic topology to analyze the intrinsic topology of complex datasets. This allows for a more robust understanding of the underlying space and relationships within the dataset. As an alternative to traditional machine learning approaches, TDA is particularly effective in studying the global properties and connectivity of data points. It introduces concepts such as persistent homology, which helps identify significant topological features that persist across different scales. By representing data as a topological space, TDA enables the extraction of valuable information about clusters, voids, and other topological structures that may not be apparent through other analytical methods. In practical applications, TDA has been useful in biology, neuroscience, machine learning, and materials science. It offers a robust framework for uncovering hidden patterns and structures in diverse datasets, providing insights that traditional methods might overlook.

Since the traditional TDA technique uses unweighted simplices, a great deal of information encoded in the global geometry of the dataset is likely to be unrecoverable. By assigning weights to simplices, TDA can distinguish

---

[*]Contact author: jbeuria@iitmandi.ac.in

between strong and weak connections, emphasizing essential features and filtering out noise. This enhances the ability to extract meaningful information about the global and local structures of the dataset.

We also explore another important topological property, Ricci curvature of the simplicial complexes. Ricci curvature, a fundamental concept in differential geometry and physics, characterizes local manifold properties, such as the volume of distance balls and geodesic divergence. In general relativity, the Einstein field equations link space-time geometry to matter distribution using the Ricci curvature tensor. Ricci flow, integral to Perelman's proof of the Poincaré conjecture [27], further underscores its significance. Discrete Ricci curvature forms [28], namely Ollivier Ricci curvature (ORC) [29,30] and Forman Ricci curvature (FRC) [31,32], extend the classical concept to networks and simplicial complexes. ORC, based on Wasserstein distance, captures clustering in network structures, while FRC, derived combinatorially, reveals geodesic dispersal and topological information. Despite potential dissimilarities, ORC and FRC exhibit high correlations in complex networks. ORC proves effective for probabilistic analyses, while FRC excels in understanding combinatorial network properties. In this study, we explore the FRC distributions.

Particle colliders like LHC collect massive amounts of data featuring various production and decay cascades of elementary particles. Physicists have studied the phenomenology of several SM extensions, viz., supersymmetry [33–36] and minimal universal extra dimension [37–39], two Higgs doublet models [40,41], etc. to name a few. In a recent exploratory work [25], we have studied the topological properties of the reconstructed objects using unweighted persistent homology for a minimal extension of the SM by a real singlet scalar [42–44].

In order to illustrate the usefulness of topological patterns in distinguishing the BSM scenarios and the SM background, we choose leptons from the resonant production of light neutralino and chargino in the next-to-minimal supersymmetric Standard Model (NMSSM). One of the primary focuses of this study is to demonstrate how the topological information behaves across different kinematic cuts, scales of maximum filtration, and integrated luminosity values. We discuss three topological features: Forman Ricci curvature (FRC) distributions, persistent entropy, and persistent amplitude for three benchmark scenarios in the NMSSM. The discussion is quite generic and can be extended to any collider observations. FRC distributions capture the topological properties at a particular scale of filtration, and persistent homology studies it on multiple scales.

The organization of the paper is as follows. In Sec. II, we present some mathematical preliminaries of topological data analysis. We give the basic framework of analysis in Sec. III. We discuss the variation of the topological features

and signal-background classification accuracy with kinematic cut selection and integrated luminosity in Sec. IV. We conclude the discussion in Sec. V.

II. TOPOLOGICAL DATA ANALYSIS (TDA)

One of the prominent tools used for topological data analysis is persistent homology. The foundational geometric structure for studying persistent homology is the simplicial complex. Simplicial complexes provide a method for constructing topological spaces using fundamental combinatorial building blocks called simplices. It simplifies the treatment of the continuous geometry of topological spaces. Instead, it involves the more manageable tasks of combinatorics and counting. These elementary building blocks, known as simplices and illustrated in Fig. 1, are formed by taking the convex hull of independent points. A k-dimensional simplex is generated by this process, involving $k + 1$ points. For instance, a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a filled triangle, and a 3-simplex is a filled tetrahedron. This construction can be extended to higher-dimensional polytopes. In an n-dimensional simplex, the simplices with dimensions $k < n$ constitute its faces. Consequently, for a 2-simplex (triangle), its edges (1-simplex) serve as the faces. Throughout this work, we have used the Vietoris-Rips complex to study the topology. It constructs simplicial complexes from point cloud data, connecting points within a specified distance threshold, known as the filtration parameter or filtration scale.

A. Persistent homology features

The k-th homology group, denoted as H_k, is the quotient group, representing cycles modulo boundaries. Mathematically, it is expressed as

$$H_k = \frac{Z_k}{B_k} = \frac{\ker(\partial_k)}{\text{im}(\partial_{k+1})} \tag{1}$$

Here, $H_k(K)$ is the quotient vector space whose generators are given by k-cycles that are not boundaries of any $(k + 1)$-simplices. The rank of $H_k(K)$ is referred to as the kth Betti number, denoted as $\beta_k(K)$. The Betti number $\beta_k(K)$ signifies the number of k-dimensional holes in the simplicial complex K that are not boundaries of any $(k + 1)$-simplices. For instance, $\beta_0(K)$ represents the number of connected components in K. It is important to note that
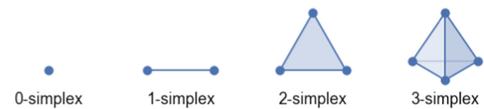

FIG. 1.   Simplices (plural of simplex) are the combinatorial building blocks of a simplicial complex. For illustration, 0-,1-,2-, and 3-simplex are shown from left to right.

Betti numbers β_k define the Euler characteristic (χ), a topological invariant of the simplicial complex, given by

$$\chi = \sum_{k=0}^{n} (-1)^k \beta_k \tag{2}$$

Another important measure is the entropy of the points clustered in the so-called persistence diagram, also called persistence entropy. The persistence diagram shows the appearance (*birth*) and disappearance (*death*) of kth hole as the filtration parameter changes. Let $D = \{(b_i, d_i)\}$ be the set of all *birth-death* pairs associated with kth order homology group in persistence diagram with $d_i < \infty$. The kth order persistence entropy is given by

$$S_k^{pe} = S(D_k) = -\sum_i p_i \log(p_i), \tag{3}$$

where $p_i = \frac{d_i - b_i}{L_D}$ and $L_D = \sum_i (d_i - b_i)$.

We also make use of a topological feature, persistent amplitude defined on D, the set of persistent (*birth,death*) pairs as a function

$$A : D \to R,$$

for which there exists a vectorization

$$\Phi : D \to V,$$

with V, a normed space such that

$$A(x) = \|\Phi(x)\|$$

for all $x \in D$. We choose the Wasserstein metric to compute distances between persistent diagrams and amplitude. The choice of the Wasserstein metric is primarily motivated by its central stability properties [45–47].

B. Forman Ricci curvature

Let α and $\bar{\alpha}$ be k-dimensional simplices in the simplicial complex K. If there exists a simplex β in K such that $\beta > \alpha$ and $\beta > \bar{\alpha}$, α and $\bar{\alpha}$ have a common co-face β and they are termed as upper adjacent. Similarly, α and $\bar{\alpha}$ are said to be lower adjacent if they share a common face γ (a $(k-1)$-simplex), that is, $\gamma < \alpha$ and $\gamma < \bar{\alpha}$. If α and $\bar{\alpha}$ are either lower or upper adjacent, but not both, they are said to be parallel. Then, Forman Ricci curvature (FRC) is given by [31]

$$R_k(\alpha) = N(\text{Upper adjacent simplices})$$
$$+ N(\text{Lower adjacent simplices})$$
$$- N(\text{Parallel simplices}) \tag{4}$$

In case of weighted simplicial complexes with weights w, $R_k(\alpha)$ is given by

$$R_k(\alpha) = w_\alpha \left[\sum_{\beta > \alpha} \frac{w_\beta}{w_\alpha} + \sum_{\gamma < \alpha} \frac{w_\gamma}{w_\alpha} \right]$$
$$- w_\alpha \sum_{\bar{\alpha} \neq \alpha} \left[\sum_{\beta > \alpha, \bar{\alpha}} \frac{\sqrt{w_\alpha w_{\bar{\alpha}}}}{w_\beta} - \sum_{\gamma < \alpha, \bar{\alpha}} \frac{w_\gamma}{\sqrt{w_\alpha w_{\bar{\alpha}}}} \right] \tag{5}$$

For an edge, Forman Ricci curvature reduces to

$$R_1(\alpha) = w_\alpha \left(\sum_{\gamma < \alpha} \frac{w_\gamma}{w_\alpha} - \sum_{\bar{\alpha} \neq \alpha} \sum_{\gamma < \alpha, \bar{\alpha}} \frac{w_\gamma}{\sqrt{w_\alpha w_{\bar{\alpha}}}} \right) \tag{6}$$

III. FRAMEWORK OF ANALYSIS

A. Z_3 symmetric next-to-minimal supersymmetric Standard Model (NMSSM)

We choose the next-to-minimal supersymmetric Standard Model (NMSSM) to demonstrate the utility of topological information in BSM searches. The NMSSM framework incorporates an additional singlet superfield denoted as $\hat{S}$ alongside the standard MSSM superfields. In the widely studied Z_3-symmetric variant of the NMSSM, the linear and bilinear terms in $\hat{S}$ are dropped. Additionally, the Z_3 symmetry restricts the inclusion of explicit Higgsino mass term (μ-term) in the NMSSM superpotential. The μ-term is generated after the singlet scalar gets vacuume expectation value (vev) ($\mu_{\text{eff}} = \lambda v_S$). The Z_3-symmetric superpotential of the NMSSM is defined as:

$$\mathcal{W} = \mathcal{W}_{\mathcal{MSSM}}|_{\mu=0} + \lambda \hat{S} \hat{H}_u . \hat{H}_d + \frac{\kappa}{3} \hat{S}^3 \tag{7}$$

with

$$\mathcal{W}_{\mathcal{MSSM}}|_{\mu=0} = y_d \hat{H}_d \cdot \hat{Q} \hat{D}_R^c + y_u \hat{Q} \cdot \hat{H}_u \hat{U}_R^c + y_e \hat{H}_d \cdot \hat{L} \hat{E}_R^c, \tag{8}$$

where $\mathcal{W}_{\mathcal{MSSM}}|_{\mu=0}$ denotes the MSSM superpotential with the exclusion of the μ-term. The superfields $\hat{H}_u$ and $\hat{H}_d$ correspond to the doublet Higgs superfields, while $\hat{S}$ represents the gauge singlet superfield mentioned earlier. The superfields $\hat{Q}$, $\hat{U}_R$, and $\hat{D}_R$ refer to the $SU(2)$ quark-doublet, up-type $SU(2)$ singlet quark, and down-type $SU(2)$ singlet quark superfields, respectively. Additionally, $\hat{L}$ and $\hat{E}_R$ represent the $SU(2)$ doublet and singlet lepton super-fields, respectively. The symbols $y_{f=d,u,e}$ denote the corresponding Yukawa couplings.

The 5×5 symmetric neutralino mass matrix, in the basis $\{\tilde{B}, \tilde{W}, \tilde{H}_d^0, \tilde{H}_u^0, \tilde{S}\}$, is given by [36]

$$\mathcal{M}_{\chi^0} = \begin{pmatrix} M_1 & 0 & -\frac{g_1 v_d}{\sqrt{2}} & \frac{g_1 v_u}{\sqrt{2}} & 0 \\ & M_2 & \frac{g_2 v_d}{\sqrt{2}} & -\frac{g_2 v_u}{\sqrt{2}} & 0 \\ & & 0 & -\mu_{\text{eff}} & -\lambda v_u \\ & & & 0 & -\lambda v_d \\ & & & & 2\kappa v_S \end{pmatrix}, \quad (9)$$

and the chargino sector is given by

$$\mathcal{M}_{\chi^\pm} = \begin{pmatrix} M_2 & g_2 v_u \\ g_1 v_d & \mu_{\text{eff}} \end{pmatrix}, \quad (10)$$

where M_1 and M_2 represent the soft SUSY-breaking masses corresponding to the $U(1)$ ($\tilde{B}$) and $SU(2)$ ($\tilde{W}$) gauginos, respectively. The parameters g_1 and g_2 denote the respective gauge couplings associated with these gauginos. It is noteworthy that there is no direct mixing observed among the gauginos ($\tilde{B}$ and $\tilde{W}$) and the singlino ($\tilde{S}$). However, a slight mixing is indirectly introduced through the neutral Higgsino sector ($\tilde{H}_d^0$, $\tilde{H}_u^0$). Conversely, direct mixing between the Higgsinos and the singlino can occur via the off-diagonal terms of $\mathcal{M}_{\chi^0}$ that are proportionate to λ. Consequently, scenarios characterized by relatively small μ_{eff} lead to lighter neutralinos exhibiting a notable blend of singlino and Higgsino components across significant regions of the NMSSM parameter space.

We present three scenarios ($\mu_{\text{eff}} \leq 250$ GeV) in Table I such that the LSP is singlinolike and the NLSP is Higgsino-like along with a Higgsino-like chargino. Such low μ_{eff} also features a light singlet Higgs boson and the SM Higgs boson. All sfermions soft-breaking mass parameters are kept at 3 TeV. The particle spectra are calculated using NMSSMTools v6.0.2 [48].

BP1 and BP2 benchmark scenarios are excluded by recent LHC searches [49]. We choose these to suggest possible validation of using topological information in BSM searches using existing data. Also, the light charginos and Higgsinos feature a sizable cross section to illustrate topological signatures for 3ℓ signals across an extended range of integrated luminosity and kinematic cuts. However, the discussed framework is generic enough for any BSM scenarios.

B. Collider simulation

We consider resonant production of $\chi_{2,3}^0 \chi_1^\pm$ and subsequent leptonic decay via Z and $W^\pm$ bosons. This leads to $3\ell + \not{E}_T$ signature at the collider. We consider three potential contributors to the background, namely, $t\bar{t}Z$, $t\bar{t}W^\pm$, and $ZW^\pm$ via leptonic decay of Z and $W^\pm$ bosons. However, for all practical purposes, the leptonic decay of $ZW^\pm$ is the most dominant background for the $3\ell + \not{E}_T$ signature. We veto the b-tagged jets after detector-level simulation.

Event samples are generated at the lowest order (LO) in perturbation theory using MadGraph5_aMC@NLO v3.5.1 [50,51] with the nn23lo1 [52] parton distribution function at $\sqrt{s} = 13$ TeV. The generated parton-level events undergo showering with Pythia v8.309 [53]. To avoid double counting of events in the simulated samples, especially in the presence of extra hard partonic jets and the parton shower, the event generator utilizes the MLM matching technique with the variables *xqcut* and *qcut* set at appropriate values. The cross sections for all processes are estimated using an NLO K-factor of 1.2.

The FastJet (v3.3.4) [54,55] package, integrated into Delphes v3.5.0 [56], is employed for jet finding. The anti-k_T jet algorithm is utilized with a cone size of 0.5, requiring a minimum p_T^{jet} of 20 GeV and limiting the pseudorapidity to $|\eta_{\text{jet}}| < 2.5$.

Following the default parameter settings of Delphes v3.5.0, the reconstruction of leptons (electrons and muons) involves a minimum p_T^l of 10 GeV and $|\eta_{\text{jet}}| < 2.5$. For electrons and muons, the track isolation requirement entails removing jets within an angular distance $\Delta R \leq 0.5$ from the lepton. To enhance the purity of electrons, it is required that the ratio of the total p_T of stray tracks within the cones of their identification to their own p_T is less than 0.12. Similarly, the corresponding ratio for muons is set at 0.25.

We choose two values of MET, i.e., MET > 50 GeV and MET > 150 GeV to demonstrate the impact of kinematic cuts. We require at least three hard leptons (≥ 20 GeV) and veto any b-tagged jets. We require $p_T^\ell \geq 50$ GeV for the first p_T ordered lepton. The kinematic cuts are implemented using MLAnalysis [57] library. Table II presents the cross-section and cut-selection details at $\sqrt{s} = 13$ TeV and 150 fb^{-1} of integrated luminosity. The significance level is calculated using $\frac{S}{\sqrt{S+B}}$, where S and B represent the number

TABLE I. BP1, BP2 and BP3 are three benchmark scenarios in the NMSSM having two different values of μ_{eff} leading to 3ℓ signals from the resonant production of $\chi_{2,3}^0 \chi_1^\mp$. Recent LHC searches have excluded BP1 and BP2.

	μ_{eff}	λ	κ	A_λ	A_κ	$m_{h_{1,2}}$	$m_{\chi_{1,2,3}^0}$	$m_{\chi^\pm}$	$BR_{\chi_2^0 \to \chi_1^0 Z}$	$BR_{\chi_3^0 \to \chi_1^0 Z}$	$BR_{\chi_3^0 \to \chi_1^\pm W^\mp}$	$BR_{\chi_1^\pm \to \chi_1^0 W^\pm}$
BP1	180	0.6	-0.24	1000	300	95.6, 124.0	83.5, 191.3, 262.2	184.5	0.99	0.30	0.0	1.0
BP2	150	0.65	-0.30	860	310	94.1, 124.3	61.6, 163.8, 251.0	153.7	0.97	0.08	0.85	1.0
BP3	250	0.69	-0.35	1250	250	125.4, 233.0	157.1, 259.2, 369.5	255.7	1.0	0.00	0.80	1.0

TABLE II. Production cross section along the leptonic channel and cut selection (N_{cut}) of the SM background and the benchmark scenarios at 13 TeV and 150 fb^{-1} for MET > 50 GeV (> 150 GeV).

	Cross section	N_{cut}	Significance	Status
SM	435.1 fb	576 (96)		
BP1	7.90 fb	116 (39)	4.41 (3.36)	Excluded
BP2	13.53 fb	142 (41)	5.30 (3.50)	Excluded
BP3	2.41 fb	33 (12)	1.34 (1.15)	Not excluded

of the signal events and the SM background events surviving after the kinematic cuts have been applied.

C. Point cloud data from collider observations

After the fast detector level simulation, we collect the reconstructed leptons event-wise. The number of events is also normalized to the integrated luminosity multiplied by the effective cross section of the considered process. Integrated luminosity values are varied between 50 fb^{-1} to 1000 fb^{-1} to understand the evolution of topological properties and their discriminatory power in search of new physics.

The point cloud comprises all leptons passing the basic kinematic cuts mentioned above. The three momenta of the reconstructed leptons serve as the coordinate for points in the point cloud. Thus, a single collider event passing the above-mentioned cuts contributes three vertices to the Vietoris-Rips complex because of our leptonic requirements. For simplicity, we use cartesian basis for the leptonic momentum.

In the study of persistent homology of the Vietoris-Rips complex, the filtration parameter is the scale. Ricci curvature captures the local behavior of the network connectivity at a particular scale. However, persistent homology features such as the Betti curve, persistent entropy, and persistent amplitude are essentially multiscale features. They represent the global behavior of the system across different scales. Therefore, we employ two slightly different strategies while studying Ricci curvature and persistent features, as discussed below.

It is to be noted that we do not need the coordinates of vertices explicitly to find the Forman Ricci curvature distributions. All we need to know is whether an edge exists between two vertices and the weights assigned to them. We choose an edge (i,j) to exist only if $\max\{p_T^{\ell_i}, p_T^{\ell_j}\} \leq$ scale. We choose a weighting system motivated by the collider observations. For simplicity, the vertices are assigned unit weights. The edges are assigned a weight that depends on the ratio of p_T of the leptons forming the edge. It is expressed as follows.

$$w_i = 1$$

$$w_{ij} = \sqrt{\frac{p_T^{\ell_i} + p_T^{\ell_j}}{\max\{p_T^{\ell_i}, p_T^{\ell_j}\}}} \quad (11)$$

However, we have used unweighted simplicial complexes to study the persistent homology of collider observations. We use our custom code to compute Ricci curvature distributions and Giotto-ai v0.6.0 library [58] for persistent homology features.

D. Machine learning and feature extraction

Given a Vietoris-Rips complex at a particular scale, we obtain the Ricci curvature distributions for the edges and use this as an input for machine learning classification of the topological features. Thus, the input feature is a one-dimensional vector whose length is the same as the number of edges in the complex. Throughout this study, we compare the topological features of the background with the combined events of background and signal. Thus, the sample is essentially skewed. Thus, we perform random undersampling so that the training of the classifier is not biased. Without this important operation, the binary classification accuracy almost always becomes 100%. We also normalize the input data using standard scalar normalization or Z—score normalization.

We use support vector machine (SVM) with the radial basis function (RBF) kernel. SVM is a powerful supervised learning algorithm used for both classification and regression tasks. The regularization parameter, also known as the penalty parameter, C, is fixed at 1.0 throughout this study. It controls the trade-off between maximizing the margin and minimizing the classification errors. We also perform 10-fold cross-validations to reduce any possible statistical fluctuation in ML classifications.

In the latter part of this work, we use Betti number distributions across filtration scales up to 50 GeV or 250 GeV for ML classification with SVM. The Betti number distributions are also popularly called Betti curves. We determine the Betti numbers up to the second homology dimensions. Thus, the input data for SVM is three-dimensional in this case. The input parameters used in SVM are the same as before.

IV. RESULTS AND DISCUSSION

This study aims to explore the dependence of topological properties of the leptonic distributions on kinematic cut selection and integrated luminosity. Many BSM frameworks, like NMSSM, feature a massive LSP. Thus, we expect the leptonic momentum distributions for the BSM to differ from the SM background. Since the mass of the invisible particle affects the missing transverse momentum distributions, the scale of maximum filtration is likely to be correlated with it.

We also explore the signal-background discriminatory potential of topological properties in relation to kinematic cuts on MET, integrated luminosity values, and the maximum filtration scale. We find local topological features using Forman Ricci curvature and global topological properties through persistent homology.

As the subsequent analysis will reveal, these intrinsic geometric properties of collider observations are efficient enough to discriminate new physics scenarios from the SM background. We have also explicitly checked that the results reported below are not very much sensitive to experimental systematic errors (at 5% level) as these fail to alter the topological structure of the dataset appreciably.

A. Forman Ricci curvatue

In Fig. 2, we present the Forman Ricci curvature R_1 distribution of the three benchmark scenarios and the SM background for two scales (50 GeV and 250 GeV)

and for two cuts on missing transverse momentum (MET > 50 GeV and MET > 150 GeV) at 150 fb^{-1} integrated luminosity. We observe that the Ricci curvature distribution for BP2 occupies most negative values, and the distribution for BP3 is very close to the SM case. This is primarily because of the cross-section distributions. It is to be reminded that we are considering a combined sample of signal and background for the benchmark scenarios.

Comparing Figs. 2(a) and 2(c), we note that with increasing MET, the Ricci curvature distributions for BP1 and BP2 become less distinct from each other. MET > 150 GeV eliminates most of the SM background. Thus, the edge connectivity is reduced significantly which results in less negative R_1 compared to the MET > 50 GeV case.

However, increasing the scale of filtration to 250 GeV increases the edge connectivity for a node. Thus, the SM distribution starts overlapping with the BSM ones. This is shown in Figs. 2(b) and 2(d). The scale at 250 GeV admits
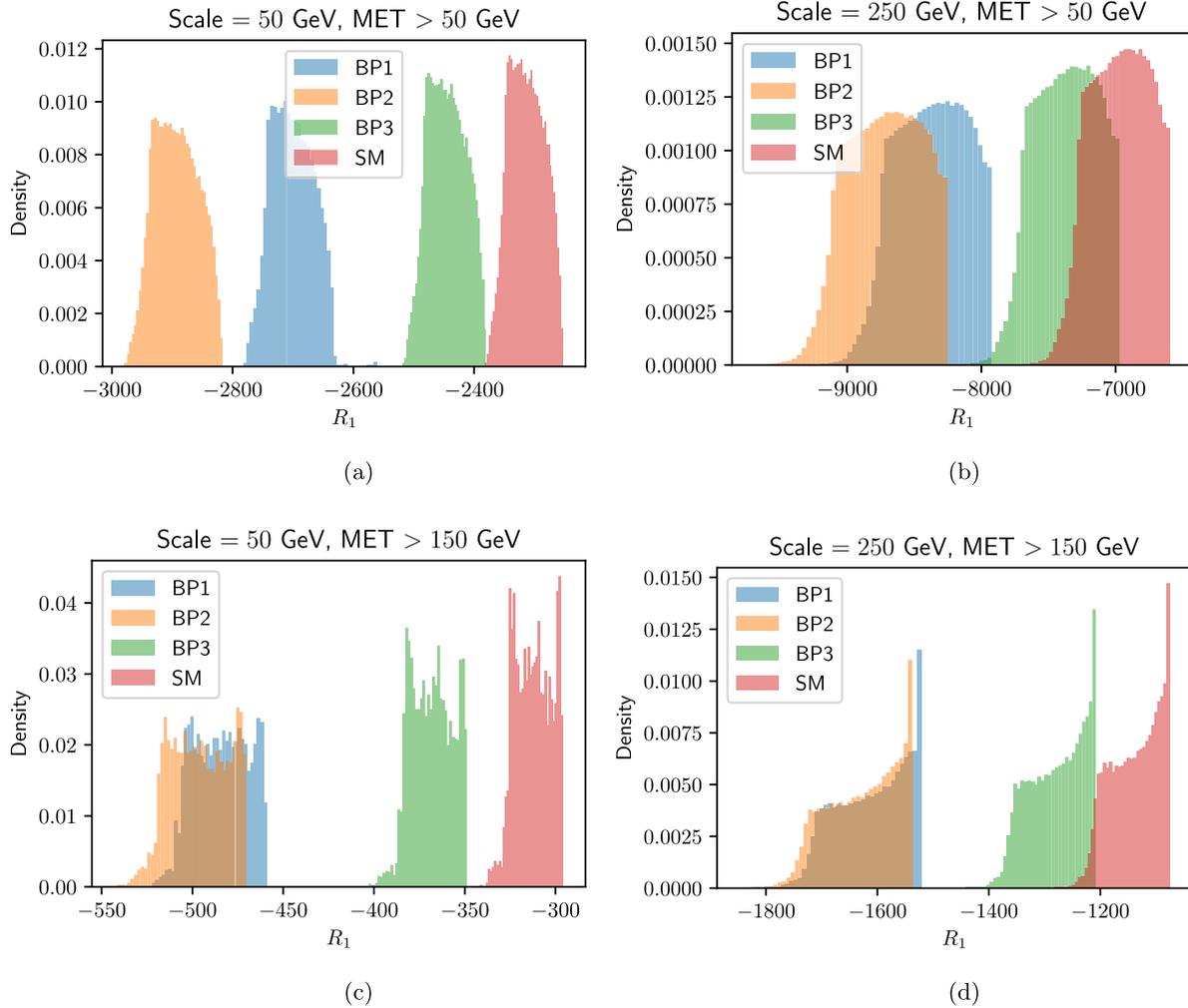

FIG. 2. R_1 distributions associated with different scales and missing transverse momentum (MET) cuts at an integrated luminosity of 150 fb^{-1}. (a) scale = 50 GeV and MET = 50 GeV, (b) scale = 250 GeV and MET = 50 GeV, (c) scale = 50 GeV and MET=150 GeV, and (d) scale = 250 GeV and MET = 150 GeV.
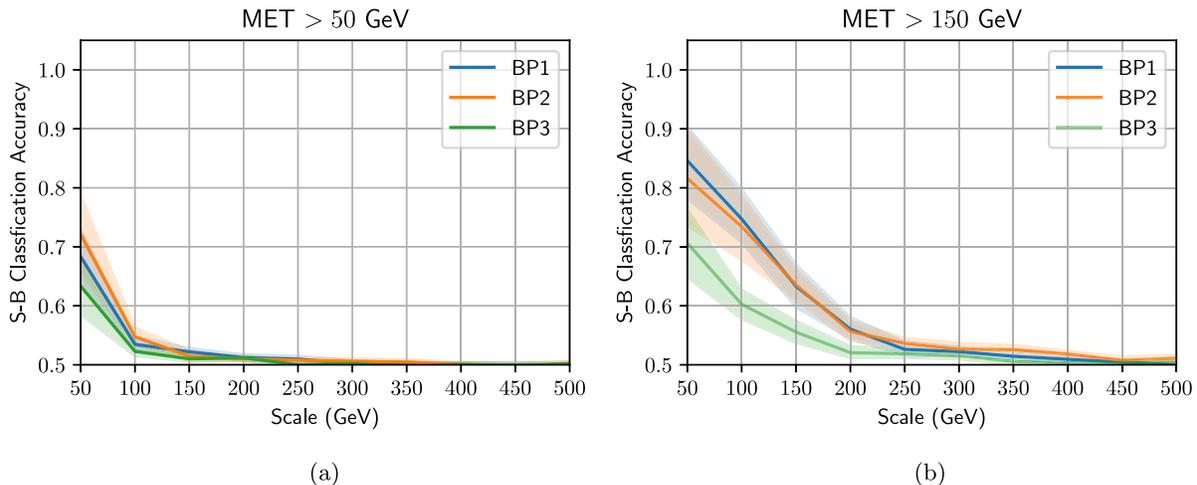
FIG. 3. Signal-background classification accuracy across different scales at 150 fb^{-1} using R_1 for (a) MET > 50 GeV and (b) MET > 150 GeV are shown. The shaded regions correspond to the 95 CL.

more edges compared to 50 GeV. Thus, for both MET > 50 GeV and MET > 150 GeV, the R_1 is more negative than their counterparts at the 50 GeV scale. As before, MET > 150 GeV reduces more SM events than BSM scenarios. Thus, Figs. 2(c) and 2(d) demonstrate qualitatively similar behavior.

In Fig. 3, we present signal and background classification across different scales for maximum filtration using the SVM framework, as discussed previously. Figure 3(a) corresponds to the MET > 50 GeV scenario, and Fig. 3(b) corresponds to the MET > 150 GeV scenario. We notice that the classification accuracy remains remarkable for both at lower scales. However, with the increasing scale of maximum filtration, accuracy drops. In Fig. 3(b), the accuracy remains significant until a larger scale. We can attribute this to the better signal-background separation at higher MET. Thus, the scale of maximum filtration is

inevitably connected with the MET and, thereby, the mass of the LSP.

In Fig. 4, we present the classification of Ricci curvature distributions for the SM and the BSM processes across different values of integrated luminosity. In the traditional cut-and-count approach, the significance of discrimination increases with integrated luminosity. However, it is not very obvious for R_1 distribution-based classification. We have chosen the benchmark points BP1 and BP2, already excluded by BSM searches at the LHC. This is corroborated by very high classification accuracy, even at a low value of integrated luminosity. Also, the benchmark point BP3 attains appreciable accuracy, particularly for MET > 150 GeV. However, Table II shows that BP3 is not excluded when traditional cut-and-count analysis is employed.

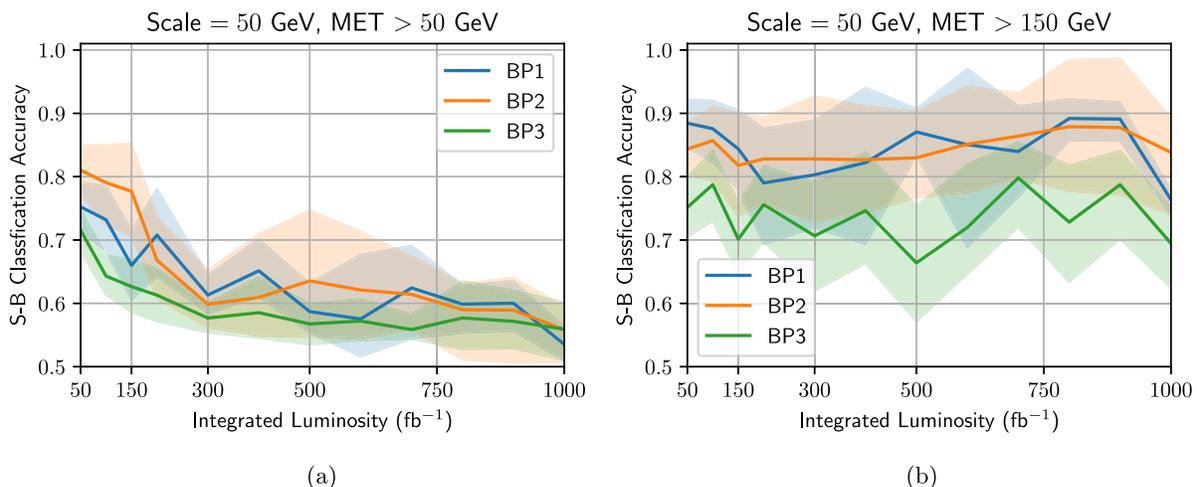The discussion of Ricci curvature entails the local topological properties of the LHC observations. Next,

FIG. 4. Classification accuracies across different integrated luminosity using R_1 for (a) MET > 50 GeV and (b) MET > 150 GeV are shown. The scale of filtration is fixed at 50 GeV. The shaded regions correspond to the 95% CL.
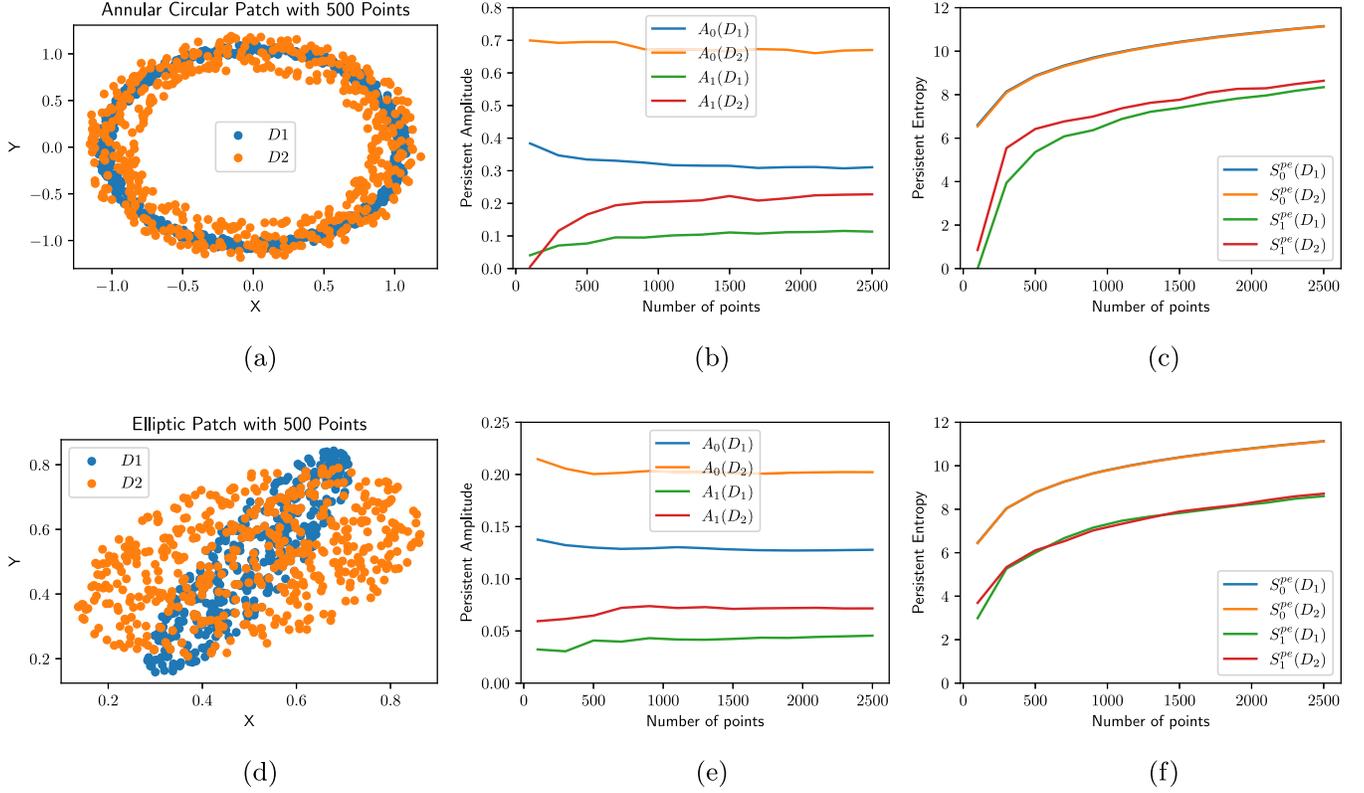
FIG. 5. Spatial distributions, persistent amplitude, and persistent entropy for two distributions (D_1 and D_2) sampled from (a, b, and c) annular regions with varying radii and (d, e, and f) elliptical patches with different widths. The number of points in D_1 and D_2 are always equal.

we delve into the discussion of persistent homology that characterizes the global geometrical properties of the LHC events.

B. Persistent homology features

Before we delve into the discussion of the persistent homology of collider observations, we first illustrate some toy examples. In Fig. 5, we present two distributions D_1 and D_2 sampled from an annular patch (first row) and an elliptical patch (second row). We vary the number of points from 100 to 2500 for each distribution. We use a maximum filtration scale of 0.2 to form the Vietoris-Rips complex. We find that as the number of sampled points increases, topological features such as persistent entropy and persistent amplitude attain constant values. This is expected because topology is the study of global features of the dataset.

The persistent amplitude is shown in Figs. 5(b) and 5(d). We observe that it is almost featureless for the elliptic patch. This is because the annular patch has a hole that remains topologically invariant. In other words, the hole cannot be removed without tearing apart the annulus. As shown in Fig. 5(a), D_2 has a smaller inner radius and larger outer radius compared to D_1. This results in a larger persistent amplitude for D_2 than D_1 across zeroth and first homology

dimensions. The situation remains the same for persistent entropy shown in Figs. 5(c) and 5(f). Persistent entropy for the zeroth homology group S_0^{pe} remains the same for D_1 and D_2 in the case of both annular and elliptic patches. However, as mentioned before, topological properties of D_1 and D_2 attain a constant ratio with rise in number of points. We will also observe similar behavior in our latter analysis of LHC events while comparing distributions of the SM background and the BSM scenarios.

Now, we return to the discussion on collider observations. As mentioned before, we use unweighted simplicial complexes to study the persistent homology of the SM background and the benchmark scenarios. The qualitative nature of the conclusions presented here will not change even if we assign weights to edges and vertices in the simplicial complex. In Fig. 6, we present the variation of persistent amplitude (A_0 and A_1) as integrated luminosity increases for two scales of maximum filtration (50 GeV and 250 GeV) while keeping MET > 150 GeV.

As noted in the above-mentioned illustrative example, A_0 and A_1 for the SM background and the benchmark scenarios attain an approximate constant scaling. The SM background always features a larger magnitude of amplitude because it contributes a larger number of vertices to the simplicial complex. We find that the scale of maximum filtration does have an impact on the persistent
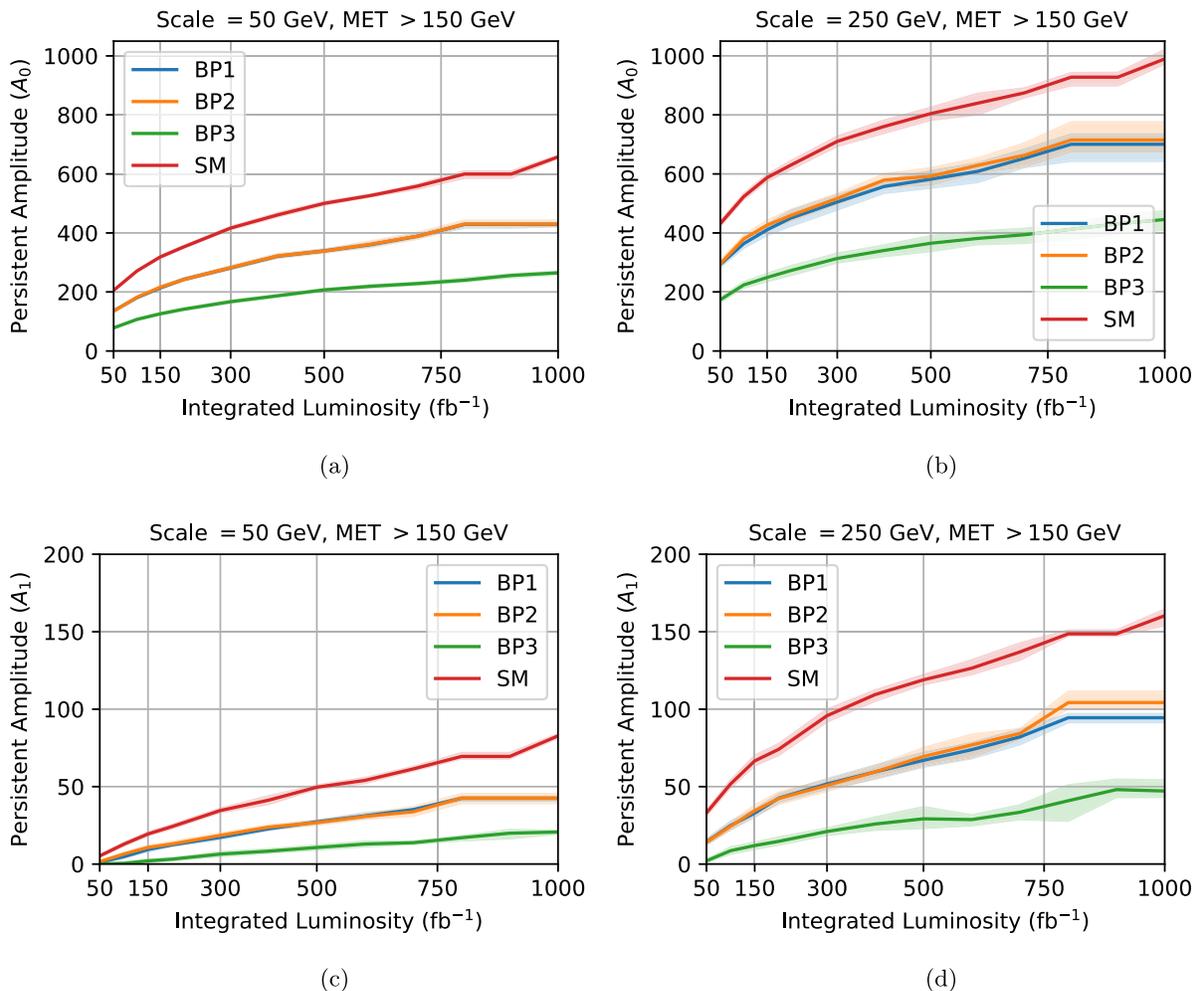
FIG. 6. Persistent amplitude distributions across different values of integrated luminosity are shown. (a) scale = 50 GeV and MET > 150 GeV for the zeroth homology group, (b) scale = 250 GeV and MET > 150 GeV for the zeroth homology group, (c) scale = 50 GeV and MET > 150 GeV for the first homology group, and (d) scale = 250 GeV and MET > 150 GeV for the first homology group. The shaded regions correspond to the 95% CL.

amplitude. A scale of 250 GeV favors larger A_0 and A_1 than 50 GeV scale. This can possibly be attributed to the inclusion of more edge connectivity in the simplicial complex.

In Fig. 7, we present the variation of persistent entropy (S_0^{pe} and S_1^{pe}) as integrated luminosity increases. The variation across luminosity is similar to Fig. 6. While S_0^{pe} and S_1^{pe} for BP1 and BP2 behave almost similarly, the ratio with the SM background attains an almost constant scaling across luminosity. The situation is slightly different for BP3, and persistent entropy for BP3 is always the lowest. This is mostly because of its lower cross section than the other benchmark points.

It is important to note that this scaling of topological properties does not tell us much about the signal-background discrimination potential of persistent homology. We show that the Betti number distribution of the SM process is different from that of the BSM processes. We use

Betti number distributions across filtration parameters at a particular scale to train the SVM classifier. The variation of the classification accuracy with integrated luminosity at two different scales of maximum filtration (50 GeV and 250 GeV) while keeping MET > 150 GeV is presented in Fig. 8.

We notice that the signal-background classification accuracy drops for BP1 and BP2 as integrated luminosity increases when the scale of maximum filtration is 50 GeV. Interestingly, the curve for BP3 remains almost constant. We also notice that the classification accuracy remains almost constant for all benchmark points as integrated luminosity increases for the 250 GeV scale. Thus, we can conclude that MET > 150 GeV likely prefers a higher scale of maximum filtration. This observation is crucial and conveys the importance of the suitable scale while studying the topological properties of collider observations.
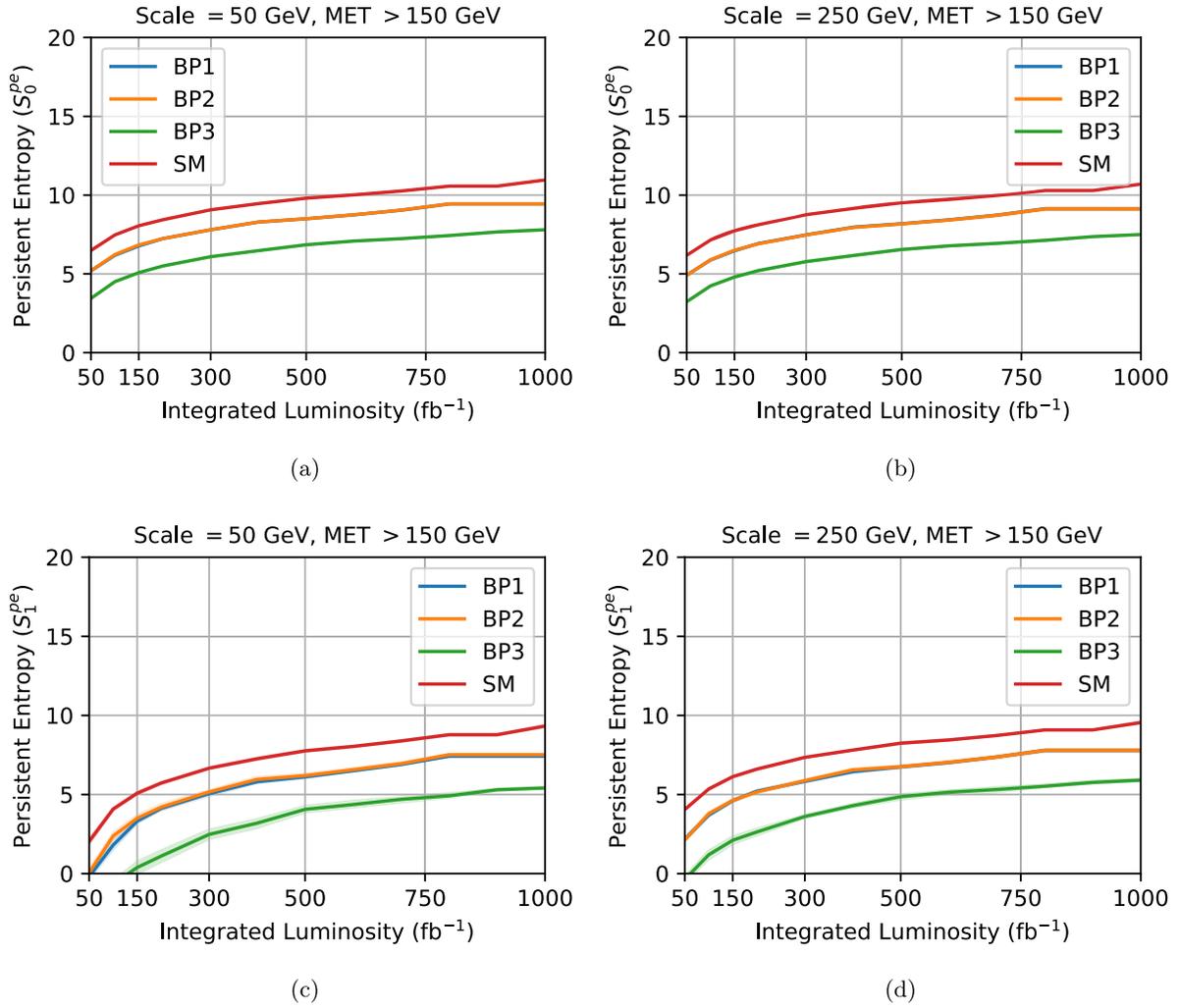
FIG. 7. Persistent entropy distributions across different values of integrated luminosity are shown. (a) scale = 50 GeV and MET > 150 GeV for the zeroth homology group, (b) scale = 250 GeV and MET > 150 GeV for the zeroth homology group, (c) scale = 50 GeV and MET > 150 GeV for the first homology group, and (d) scale = 250 GeV and MET > 150 GeV for the first homology group. The shaded regions correspond to the 95% CL.
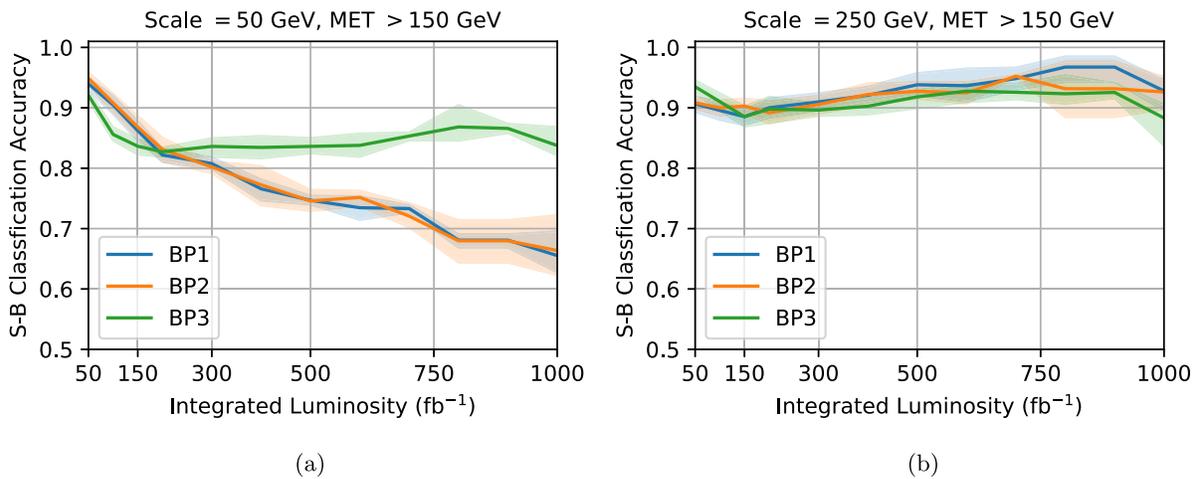

FIG. 8. Signal-background classification accuracy for (a) scale = 50 GeV and (b) scale = 250 GeV at MET > 150 GeV across different values of integrated luminosity. The shaded regions correspond to the 95% CL.

V. CONCLUSION

In conclusion, our study has provided insights into the distinction between the topological properties of the BSM signal and the SM background events. The framework discussed in this work is quite generic and can be applied to other BSM searches at colliders. Through the analysis of Forman Ricci curvature and persistent homology, we have explored the dependence of topological features on kinematic cut selection, integrated luminosity, and the scale of maximum filtration.

Our results highlight the efficacy of topological properties, such as Forman Ricci curvature and persistent homology, in capturing intrinsic geometric structures present in collider observations. This study also emphasizes the impact of MET cuts and the scale of maximum filtration while studying topological properties. Furthermore, our investigation into persistent homology features reveals an approximate scaling behavior across integrated luminosity, suggesting robustness in topological properties irrespective of the dataset size.

Crucially, our study underscores the interconnectedness between the scale of maximum filtration, MET cuts, and the mass of the LSP, highlighting the importance of considering these factors collectively in topological analyses aimed at uncovering new physics scenarios. We believe using topological discriminators with a traditional kinematic cut-based approach can yield stronger exclusion limits even at low integrated luminosity.

Overall, our findings deepen the understanding of the intrinsic geometric structures in collider observations and provide valuable implications for future experimental studies to uncover novel physics phenomena beyond the Standard Model.

The Python code used in this work is available at Ref. [59].

[1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. A. Khalek, A. A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. AbouZeid *et al.*, 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**, 1 (2012).

[2] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan *et al.*, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B **716**, 30 (2012).

[3] R. Franceschini, D. Kim, K. Kong, K. T. Matchev, M. Park, and P. Shyamsundar, Kinematic variables and feature engineering for particle phenomenology, Rev. Mod. Phys. **95**, 045004 (2023).

[4] D. Debnath, J. S. Gainer, D. Kim, and K. T. Matchev, Edge detecting new physics the Voronoi way, Europhys. Lett. **114**, 41001 (2016).

[5] D. Debnath, J. S. Gainer, C. Kilic, D. Kim, K. T. Matchev, and Y.-P. Yang, Identifying phase space boundaries with Voronoi tessellations, Eur. Phys. J. C **76**, 645 (2016).

[6] K. T. Matchev, A. Roman, and P. Shyamsundar, Finding wombling boundaries in LHC data with Voronoi and Delaunay tessellations, J. High Energy Phys. 12 (2020) 137.

[7] A. Mullin, S. Nicholls, H. Pacey, M. Parker, M. White, and S. Williams, Does SUSY have friends? A new approach for LHC event analysis, J. High Energy Phys. 02 (2021) 160.

[8] D. Guest, K. Cranmer, and D. Whiteson, Deep learning and its application to LHC physics, Annu. Rev. Nucl. Part. Sci. **68**, 161 (2018).

[9] S. R. Qasim, J. Kieseler, Y. Iiyama, and M. Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks, Eur. Phys. J. C **79**, 1 (2019).

[10] M. Abdughani, J. Ren, L. Wu, and J. M. Yang, Probing stop pair production at the LHC with graph neural networks, J. High Energy Phys. 05 (2019) 055.

[11] Y.-L. Du, K. Zhou, J. Steinheimer, L.-G. Pang, A. Motornenko, H.-S. Zong, X.-N. Wang, and H. Stöcker, Identifying the nature of the QCD transition in relativistic collision of heavy nuclei with deep learning, Eur. Phys. J. C **80**, 1 (2020).

[12] F. Flesher, K. Fraser, C. Hutchison, B. Ostdiek, and M. D. Schwartz, Parameter inference from event ensembles and the top-quark mass, J. High Energy Phys. 09 (2021) 058.

[13] S. Chang, T.-K. Chen, and C.-W. Chiang, Distinguishing W' signals at hadron colliders using neural networks, Phys. Rev. D **103**, 036016 (2021).

[14] B. Nachman and J. Thaler, Learning from many collider events at once, Phys. Rev. D **103**, 116013 (2021).

[15] M. Knipfer, S. Meier, J. Heimerl, P. Hommelhoff, and S. Gleyzer, Deep learning-based spatiotemporal multi-event reconstruction for delay line detectors, arXiv:2306.09359.

[16] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, Topological data analysis of contagion maps for examining spreading processes on networks, Nat. Commun. **6**, 7723 (2015).

[17] C. M. Topaz, L. Ziegelmeier, and T. Halverson, Topological data analysis of biological aggregation models, PLoS One **10**, e0126383 (2015).

[18] S. Lloyd, S. Garnerone, and P. Zanardi, Quantum algorithms for topological and geometric analysis of data, Nat. Commun. **7,** 10138 (2016).

[19] M. Gidea and Y. Katz, Topological data analysis of financial time series: Landscapes of crashes, Physica (Amsterdam) **491A,** 820 (2018).

[20] M. Saggar, O. Sporns, J. Gonzalez-Castillo, P. A. Bandettini, G. Carlsson, G. Glover, and A. L. Reiss, Towards a new approach to reveal dynamical organization of the brain using topological data analysis, Nat. Commun. **9,** 1399 (2018).

[21] A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett, The importance of the whole: Topological data analysis for the network neuroscientist, Netw. Neurosci. **3,** 656 (2019).

[22] J. Murugan and D. Robertson, An introduction to topological data analysis for physicists: From LGM to FRBS, arXiv:1904.11044.

[23] A. Cole and G. Shiu, Topological data analysis for the string landscape, J. High Energy Phys. 03 (2019) 054.

[24] F. Chazal and B. Michel, An introduction to topological data analysis: Fundamental and practical aspects for data scientists, Front. Artif. Intell. **4,** 108 (2021).

[25] J. Beuria, Persistent homology of collider observations: When (w)hole matters, Phys. Lett. B **846,** 138188 (2023).

[26] K. V. Gupta, J. Beuria, and L. Behera, Characterizing EEG signals of meditative states using persistent homology and Hodge spectral entropy, Biomed. Signal Process. Control **89,** 105779 (2024).

[27] G. Perelman, Ricci flow with surgery on three-manifolds, arXiv:math/0303109.

[28] A. Samal, R. Sreejith, J. Gu, S. Liu, E. Saucan, and J. Jost, Comparative analysis of two discretizations of Ricci curvature for complex networks, Sci. Rep. **8,** 8650 (2018).

[29] Y. Ollivier, Ricci curvature of metric spaces, C. R. Math. **345,** 643 (2007).

[30] Y. Ollivier, Ricci curvature of Markov Chains on metric spaces, J. Funct. Anal. **256,** 810 (2009).

[31] R. Forman, Bochner's method for cell complexes and combinatorial Ricci curvature, Discrete Comput. Geom. **29,** 323 (2003).

[32] R. Sreejith, K. Mohanraj, J. Jost, E. Saucan, and A. Samal, Forman curvature for complex networks, J. Stat. Mech. 2016, 063206.

[33] C. Csaki, The minimal supersymmetric standard model, Mod. Phys. Lett. A **11,** 599 (1996).

[34] S. P. Martin, A supersymmetry primer, in *Perspectives on Supersymmetry* (World Scientific, Singapore, 1998), pp. 1–98.

[35] H. Baer and X. Tata, *Weak Scale Supersymmetry: From Superfields to Scattering Events* (Cambridge University Press, Cambridge, England, 2006).

[36] U. Ellwanger, C. Hugonie, and A. M. Teixeira, The next-to-minimal supersymmetric standard model, Phys. Rep. **496,** 1 (2010).

[37] J. A. Cembranos, J. L. Feng, and L. E. Strigari, Exotic collider signals from the complete phase diagram of minimal universal extra dimensions, Phys. Rev. D **75,** 036004 (2007).

[38] A. Datta, K. Kong, and K. T. Matchev, Minimal universal extra dimensions in CalcHep/CompHep, New J. Phys. **12,** 075017 (2010).

[39] J. Beuria, A. Datta, D. Debnath, and K. T. Matchev, LHC collider phenomenology of minimal universal extra dimensions, Comput. Phys. Commun. **226,** 187 (2018).

[40] S. Davidson and H. E. Haber, Basis-independent methods for the two-Higgs-doublet model, Phys. Rev. D **72,** 035004 (2005).

[41] G. C. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, Theory and phenomenology of two-Higgs-doublet models, Phys. Rep. **516,** 1 (2012).

[42] S. Ham, Y. Jeong, and S. Oh, Electroweak phase transition in an extension of the standard model with a real Higgs singlet, J. Phys. G **31,** 857 (2005).

[43] V. Barger, P. Langacker, M. McCaskey, M. J. Ramsey-Musolf, and G. Shaughnessy, CERN LHC phenomenology of an extended standard model with a real scalar singlet, Phys. Rev. D **77,** 035005 (2008).

[44] W.-L. Guo and Y.-L. Wu, The real singlet scalar dark matter model, J. High Energy Phys. 10 (2010) 083.

[45] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko, Lipschitz functions have l p-stable persistence, Found. Comput. Math. **10,** 127 (2010).

[46] P. Skraba and K. Turner, Wasserstein stability for persistence diagrams, arXiv:2006.16824.

[47] T. Songdechakraiwut, B. M. Krause, M. I. Banks, K. V. Nourski, and B. D. Van Veen, Wasserstein distance-preserving vector space of persistent homology, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, New York, 2023), pp. 277–286.

[48] D. Das, U. Ellwanger, and A. M. Teixeira, NMSDECAY: A Fortran code for supersymmetric particle decays in the next-to-minimal supersymmetric standard model, Comput. Phys. Commun. **183,** 774 (2012).

[49] G. Aad *et al.* (ATLAS Collaboration), Searches for new phenomena in events with two leptons, jets, and missing transverse momentum in 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, Eur. Phys. J. C **83,** 515 (2023).

[50] 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.

[51] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, The automation of next-to-leading order electroweak calculations, J. High Energy Phys. 07 (2018) 185; J. High Energy Phys. 11 (2021) 85.

[52] R. D. Ball, V. Bertone, S. Carrazza, L. D. Debbio, S. Forte, P. Groth-Merrild, A. Guffanti, N. P. Hartland, Z. Kassabov, J. I. Latorre *et al.*, Parton distributions from high-precision collider data: NNPDF Collaboration, Eur. Phys. J. C **77,** 663 (2017).

[53] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel,

C. T. Preuss *et al.*, A comprehensive guide to the physics and usage of Pythia 8.3, SciPost Phys. Codebases 8 (2022).

[54] M. Cacciari and G. P. Salam, Dispelling the N^3 myth for the k_t jet-finder, Phys. Lett. B **641**, 57 (2006).

[55] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, Eur. Phys. J. C **72**, 1896 (2012).

[56] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, J. High Energy Phys. 02 (2014) 057.

[57] Y.-C. Guo, F. Feng, A. Di, S.-Q. Lu, and J.-C. Yang, MLAnalysis: An open-source program for high energy physics analyses, Comput. Phys. Commun. **294**, 108957 (2024).

[58] G. Tauzin, U. Lupo, L. Tunstall, J. B. Pérez, M. Caorsi, A. M. Medina-Mardones, A. Dassatti, and K. Hess, giotto-tda: A topological data analysis toolkit for machine learning and data exploration, J. Mach. Learn. Res. **22**, 1 (2021).

[59] https://github.com/jbeuria/FRC-PH-LHC/