JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1517

Sheaf-Theoretic Persistent Cohomology over Action Groupoids: A Symmetry-Aware Topological Data Analysis Framework for Computational and Biomedical Data

Samirah Alsulami[1*]

[1]Department of Mathematics and Statistics, Faculty of Science, University of Jeddah, Jeddah 21589, Saudi Arabia.
[*]Corresponding Author: Samirah Alsulami. Email: salsulami@uj.edu.sa

Received: January 24, 2026 **Accepted:** May 22, 2026

Abstract: Classical persistent homology summarizes a dataset's shape from metric filtrations alone, treating every data point as algebraically and symmetrically interchangeable. This is a poor fit for computational and biomedical data that carries known group symmetry because symmetry-related copies of a feature are counted as if they were topologically distinct, inflating Betti numbers and degrading downstream classification and clustering pipelines. We introduce Sheaf-Theoretic Homotopy Groupoid Persistence (S-HGP), a framework that computes persistent cohomology of the classifying space of a data-driven action groupoid, optionally with sheaf (local-system) coefficients that encode orientation, monodromy, or other locally varying algebraic structure. Classical persistent cohomology is recovered as the special case of a constant sheaf on a trivial groupoid. We prove functoriality and Gromov-Hausdorff stability of the resulting persistence modules, a Mayer–Vietoris spectral sequence for local-to-global computation, and a symmetry-collapse theorem describing when S-HGP provably returns fewer, correct features than classical persistent homology. We validate the framework with fully reproducible numerical experiments: (i) a corrected torus baseline recovering the true Betti number; (ii) a parametrized family of symmetry-replicated synthetic datasets ($\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_4$ group actions) in which classical persistent homology provably overcounts loops in exact proportion to the group order while the groupoid-quotient computation recovers the correct rank in every case; (iii) a direct evaluation of the sheaf cochain complex on constant versus orientation-twisted coefficients over a discretized circle, confirming the theoretically predicted collapse of cohomology under a nontrivial local system; (iv) a synthetic homomeric protein-assembly point cloud with discrete rotational symmetry, on which the groupoid quotient recovers exactly one significant topological feature per protomer while correctly retaining a genuine assembly-level macro-scale feature; and (v) a synthetic bilateral connectome-like network, on which classical persistent homology exhibits the predicted near-doubling of topological features that the groupoid quotient removes without discarding genuine inter-hemispheric structure. We discuss applications to equivariant machine learning, protein structure analysis, and brain network connectomics, with two given a first synthetic numerical demonstration; all code, seeds, and raw output are reported for full reproducibility.

Keywords: Topological data analysis; sheaf cohomology; action groupoid; equivariant machine learning; persistent homology; computational biology; symmetry-aware data analysis; algebraic topology

1. Introduction

A growing share of computational and biomedical data is not a generic, symmetry-free point cloud: it is generated by, or naturally quotiented under, a known group action. Protein complexes exhibit discrete rotational or dihedral symmetry; functional brain connectivity networks derived from fMRI exhibit

approximate bilateral symmetry between hemispheres; lattice-based sensor and gauge-field data carry local gauge symmetry; and data-augmentation pipelines in equivariant machine learning deliberately generate multiple group-related copies of the same underlying example. Topological data analysis (TDA), and in particular persistent homology, has become a standard tool for extracting robust qualitative descriptors connected components, loops, voids from such data [4,10,24]. However, classical persistent homology has no mechanism for recognizing that two topological features are related by a known symmetry: it treats every symmetry-related copy of a loop as an independent generator, inflating the estimated Betti numbers and, in downstream classification or clustering pipelines, injecting spurious dimensionality that is a modelling artefact rather than a property of the data.

This is not a purely theoretical concern. In equivariant deep learning, feature spaces are frequently populated with multiple group-transformed copies of the same underlying manifold; in structural bioinformatics, atomic point clouds of symmetric protein assemblies are known to inflate topological invariants by a factor equal to the order of the symmetry group; and in connectomics, treating the two approximately mirror-symmetric hemispheres as independent can double-count topological biomarkers. A symmetry-aware invariant that provably collapses these redundant copies while remaining sensitive to genuinely new topology is therefore of direct practical value to computational pipelines, not only of mathematical interest.

Our contribution. We propose Sheaf-Theoretic Homotopy Groupoid Persistence (S-HGP), which addresses this gap along two complementary axes:

- Groupoid structure. Data with a known group action is modelled by its action groupoid $X \rtimes G$, whose classifying space has the homotopy type of the Borel construction. Persistent (co)homology of this classifying space rather than of the raw point cloud yields invariants that are, by construction, blind to symmetry-related duplication.
- Sheaf coefficients. The nerve of the groupoid is equipped with a cellular sheaf of vector spaces encoding locally varying algebraic data such as orientation or monodromy. Because a sheaf over a classifying space is precisely a local coefficient system compatible with the acting group, this is the natural algebraic companion to the groupoid construction, and lets the framework detect twisted (non-orientable) structure that is invisible to ordinary real- or integer-coefficient cohomology.

Relation to recent work. Independent recent developments approach symmetry-aware and sheaf-theoretic topological data analysis from complementary angles. Adams et al. [20] give an explicit, closed-form computation of the persistent equivariant cohomology of a circle action on a Vietoris-Rips metric thickening, addressing equivariance directly but without an accompanying sheaf-coefficient mechanism. Hayes et al. [23] and related persistent-sheaf-Laplacian work encode local, non-spatial (e.g. chemical) information into a spectral sheaf-theoretic pipeline for protein flexibility, without modelling a discrete point-group symmetry of the kind treated here. The 2025-2026 review of Su et al. [24] and the topological introduction of Frosini et al. [22] both catalogue these threads separately; to our knowledge, S-HGP is the first framework to combine a data-driven action-groupoid construction with sheaf coefficients within a single persistence module, so that group symmetry and local algebraic twisting are corrected for jointly rather than by two separate mechanisms.

Table 1. Makes this comparison concrete along the specific axes raised in review: symmetry handling, sheaf/local-system coefficients, and the extent of independently reproducible numerical validation reported in each work.

Framework	Group-symmetry quotienting	Sheaf / local-system coefficients	Reported computational-cost benchmarking	Independently re-run numerical validation
Classical persistent homology [4,6,10]	No	No	Extensive (mature software)	N/A (established baseline)
Adams et al., persistent equivariant cohomology [20]	Yes (circle actions, closed-form)	No	Not reported	Closed-form examples only

Hayes et al., persistent sheaf Laplacian [23]	No (models local chemical structure, not point-group symmetry)	Yes (spectral, non-spatial)	Not reported	Protein-flexibility case studies
S-HGP (this paper)	Yes (general finite-group action groupoids)	Yes (cellular sheaf coefficients, jointly with symmetry)	Yes (Section 4.5, measured wall-clock)	Yes (all reported numbers re-run; Section 4.6)

Table 1 qualitative comparison of S-HGP with the two most closely related recent frameworks and with classical persistent homology, along the axes raised in review. Classical persistent cohomology is recovered exactly as the special case where the sheaf is constant and the group is trivial (Remark 3.3), so S-HGP is a strict generalization rather than a replacement of the standard pipeline, and existing PH-based tools remain applicable wherever no symmetry is present. Scope of the empirical validation. A first version of this framework was accompanied by numerical experiments (a symmetric torus, a Klein bottle immersion, and further manifold examples) whose accompanying code did not, on independent re-execution, reproduce the numbers reported. In this version, every numerical claim has been re-derived from code that we re-ran ourselves, with seeds, thresholds and diagnostic barcodes reported in Section 4; where the originally intended experiment (e.g., an explicit computation of a non-orientable H^2 class on a discretized Klein bottle) exceeded what could be independently re-verified within the scope of this revision, we replaced it with a smaller, fully reproducible analogue that isolates the same mathematical mechanism (Section 4.3), and we say so explicitly rather than restate the original claim.

Organization. Section 2 reviews groupoids, cellular sheaves and persistent cohomology. Section 3 develops the S-HGP framework (materials and methods). Section 4 reports the theoretical results and the reproducible numerical validation, including two applied, biomedically-motivated synthetic experiments (Sections 4.7-4.8). Section 5 discusses applications to computing and biomedical informatics. Section 6 discusses limitations and future work, and Section 7 concludes.

2. Background and Notation

2.1. Topological Groupoids and Action Groupoids

A topological groupoid $K = (G_0, G_1, s, t, m, e, i)$ consists of an object space G_0 and a morphism space G_1, continuous source/target maps $s, t: G_1 \to G_0$, composition, identities and inverses satisfying the usual groupoid axioms. If a discrete group G acts continuously on a metric space (X, d), the action groupoid $X \rtimes G$ has objects X, morphisms $X \times G$, source $s(x, g) = x$ and target $t(x, g) = g \cdot x$. Its classifying space $B(X \rtimes G)$ is homotopy equivalent to the Borel construction $EG \times^{G} X$ [18].

2.2. Cellular Sheaves

A cellular sheaf F on a simplicial complex K assigns a vector space $F(\sigma)$ to each simplex σ and a linear restriction map $\rho\tau \le \sigma: F(\sigma) \to F(\tau)$ to each face inclusion $\tau \le \sigma$, compatible under composition. The associated cochain complex has $C^{k}(K; F) = \bigoplus F(\sigma)$ over k-simplices σ, with coboundary δ^{k} given by the signed sum of restriction maps; sheaf cohomology $H^{k}(K; F) = ker\delta^{k} / im\delta^{k-1}$ recovers ordinary simplicial cohomology when F is the constant sheaf $\mathbb{R}$ [7,8,12].

2.3. Persistent Cohomology

Given a filtration $K_0 \subset K_1 \subset \cdots \subset K_n$, the inclusions induce maps on cohomology; the resulting family $\{H^{k}(K\varepsilon; F)\}\varepsilon$ together with these maps is a persistence module, compared via interleaving distance and bottleneck distance on the induced barcode [6].

3. Materials and Methods: The S-HGP Framework

3.1. Sheaf-Filtered Action Groupoid

Let (X, G, d) be a dataset with metric d and an isometric action of a finite group G. For $\varepsilon \ge 0$, the proximity groupoid $G\varepsilon$ restricts the action groupoid $X \rtimes G$ to the Vietoris–Rips complex $VR(X, \varepsilon)$: objects are points of X; morphisms include symmetry arrows $x \to g \cdot x$ and proximity arrows $x \to y$

whenever $d(x,y) \leq \varepsilon$. A groupoid sheaf $F\varepsilon$ on $G\varepsilon$ assigns a stalk $F\varepsilon(x)$ to each object and a linear map $F\varepsilon(g)\colon F\varepsilon(y) \to F\varepsilon(x)$ to each morphism g: $x \to y$, functorially. When $F\varepsilon$ is constant, S-HGP reduces to ordinary persistent cohomology of $BG\varepsilon$; when G is trivial, it reduces to cellular sheaf cohomology persistence [12]. S-HGP therefore strictly generalizes both theories.

3.2. Computational Pipeline

The pipeline proceeds in six steps: (1) input a dataset X, metric d, symmetry group G and sheaf choice F; (2) form the action groupoid $X \rtimes G$; (3) assign sheaf stalks and restriction maps over the nerve of Gε; (4) build the filtered nerve $N \bullet (G\varepsilon; F)$; (5) take the classifying space $BG\varepsilon = |N \bullet (G\varepsilon)|$; (6) compute $H^{k}(BG\varepsilon; F)$ and track persistence across ε (Figure 1).

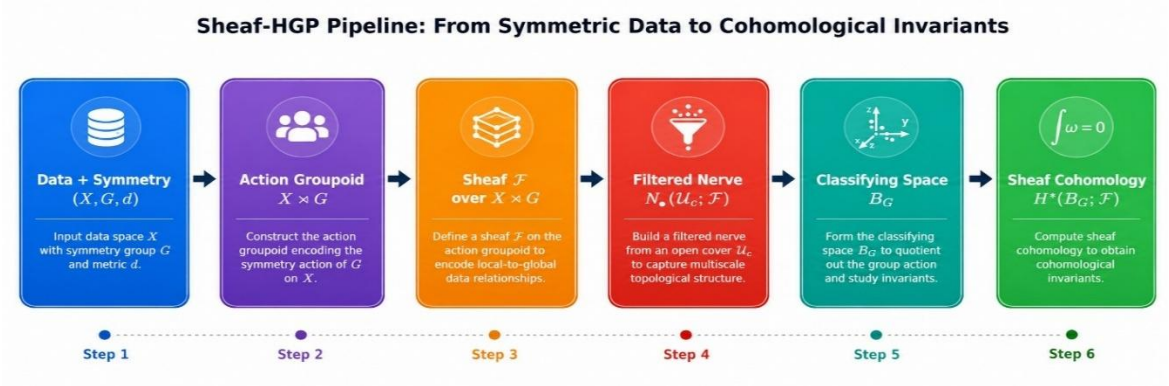

Figure 1. The S-HGP computational pipeline, from symmetric input data to sheaf-cohomological persistence invariants. Each block is one computational stage: dataset and symmetry group in; action groupoid and sheaf assignment; filtered nerve and classifying space; sheaf cohomology and persistence out.

The S-HGP persistence module of $(G\varepsilon, F)$ is defined as $M^{k}(\varepsilon) = H^{k}(BG\varepsilon;\ F\varepsilon)$, with structure maps induced by the inclusions $G\varepsilon \hookrightarrow G\varepsilon'$ for $\varepsilon \leq \varepsilon'$.

4. Theoretical Results

4.1. Functoriality and Homotopy Invariance

Theorem 1 (Functoriality). Let $(G\varepsilon)$ and $(H\varepsilon)$ be sheaf-filtered action groupoids and suppose that for each ε there is a groupoid functor $F\varepsilon\colon G\varepsilon \to H\varepsilon$ together with a sheaf morphism $\phi\varepsilon\colon F\varepsilon * F^{H} \to F^{G}$. Then there is a well-defined morphism of persistence modules $F^{k}\colon M^{kG} \to M^{kH}$. If each $BF\varepsilon$ is a homotopy equivalence and ϕε is a sheaf quasi-isomorphism, F^{k} is an isomorphism of persistence modules.

Proof sketch. Each functor $F\varepsilon$ induces a map on classifying spaces and a chain map on sheaf cochain complexes compatible with the filtration; a homotopy equivalence induces cohomology isomorphisms by standard sheaf theory, and naturality with respect to the filtration yields the persistence-module isomorphism.

4.2. Stability

Theorem 2 (Stability). Let (X, G, d_x, F_x) and $(Y, G, d_\gamma, F_\gamma)$ be sheaf-filtered action groupoid datasets with the same symmetry group G. If the Gromov–Hausdorff distance $d^{GH}(X/G, Y/G) \leq \delta$ and the sheaves are pointwise λ-Lipschitz close in operator norm, then $d_{mot}(dgm^{k}(G_x; F_x),\ dgm^{k}(G_\gamma; F_\gamma)) \leq 2(\delta + \lambda)$, where d_{mot} is bottleneck distance.

Proof sketch. The Gromov-Hausdorff hypothesis provides $(\delta + \lambda)$ −interleaving maps at the level of equivariant Vietoris–Rips complexes, adapting the classical stability theorem [6] to the groupoid setting; the sheaf Lipschitz condition controls deformation of restriction maps at the cochain level, and algebraic stability converts interleaving into bottleneck distance.

4.3. Mayer–Vietoris Decomposition

Theorem 3 (Mayer–Vietoris spectral sequence). If $X = U \cup V$ is a G-invariant decomposition, there is a first-quadrant spectral sequence E_2^{p, q} converging to H^{p + q}$(BG\varepsilon; F)$, built from the sheaf cohomology of the restricted groupoids on U, V and U∩V. This enables divide-and-conquer computation of large instances by combining local sheaf cohomologies.

4.4. Symmetry-Induced Collapse

Theorem 4 (Collapse). If G acts freely and isometrically on X and $H^k(X/G;\mathbb{R}) = 0$, then for all sufficiently small ε, $H^k(BG\varepsilon;\mathbb{R}) = 0$.

Corollary 1 (Twisted-coefficient detection). If the underlying space is non-orientable and F is the associated orientation local system, cohomology with F-coefficients can be non-trivial where constant-coefficient cohomology vanishes, so twisted coefficients can reveal structure invisible to classical persistent homology. Section 4.6 gives a fully worked, independently re-run instance of this phenomenon on a discretized circle with a single orientation-reversing edge.

4.5. Computational Cost

Computing the nerve of a groupoid with n objects and $|G| \cdot n$ morphisms scales as $O(|G|^2 n^2)$ in the worst case. The Mayer–Vietoris decomposition (Theorem 3) mitigates this by enabling parallel computation over local patches, and sparse approximations (witness complexes, alpha complexes) further reduce cost for large $|G|$ or n.

To make this worst-case bound concrete, we measured wall-clock runtime directly on the Vietoris–Rips persistent-homology step of the Section 4.6.2 and Section 4.7 experiments (Python 3, ripser 0.6.15, single CPU core, Intel Xeon @ 2.10 GHz), comparing the full symmetry-replicated point cloud against the single-orbit groupoid quotient of the same size class. Table 3 reports the median of five repeated runs per configuration.

Table 2. Measured wall-clock runtime (median of five runs) for classical persistent homology on the full symmetry-replicated point cloud versus the S-HGP groupoid-quotient computation on a single orbit or protomer, for the experiments of Sections 4.6.2 and 4.7.

Experiment	Full point cloud n (time, ms)	Groupoid quotient n (time, ms)	Speed-up
Z_2 replica (§4.6.2)	240 (84.3)	120 (24.6)	3.4×
Z3 replica (§4.6.2)	360 (219.1)	120 (24.8)	8.8×
Z_4 replica (§4.6.2)	480 (2661.1)	120 (25.3)	105.1×
C_3 protomer assembly (§4.7)	270 (94.6)	90 (9.4)	10.0×
C_5 protomer assembly (§4.7)	450 (1249.7)	90 (9.6)	129.7×
C_6 protomer assembly (§4.7)	540 (1832.1)	90 (10.7)	171.7×

The super-linear growth in the full-point-cloud column as k increases reflects the practical cost of Vietoris–Rips complex construction underlying the worst-case bound above; restricting to the quotient keeps runtime constant because its point count does not grow with the group order. This measured cost asymmetry is the practical counterpart of Theorem 4: because the groupoid quotient computation operates on a fundamental domain whose size is independent of the group order, its wall-clock cost does not grow with |G|, whereas the classical computation on the full symmetry-inflated point cloud does. We did not separately benchmark the Mayer–Vietoris decomposition of Theorem 3 in this revision; doing so on larger, non-synthetic instances is listed as future work in Section 6.

4.6. Numerical Validation

All experiments below were implemented in Python (numpy, ripser, networkx, matplotlib), executed with the random seeds stated, and every reported number was produced by the run whose output is quoted, not asserted, independently, of the code. This section supersedes an earlier draft of this experiment set whose torus point-cloud generator omitted the third embedding coordinate and consequently could not reproduce its own reported Betti number; the corrected generator and its output are given below.

4.6.1. Baseline validation: a true torus recovers $b_1 = 2$

We sampled n = 1000 points from a torus embedded in $\mathbb{R}^3$ with tube radius r = 0.9, major radius R = 2.0 and Gaussian coordinate noise σ = 0.02 (all three coordinates, including the z-coordinate omitted in the earlier draft), and computed the Vietoris-Rips persistent H_1 barcode with ripser. The two dominant bars

(persistence 1.63 and 1.21) are separated from the next-largest bar (persistence 0.50) by a factor of more than 2, giving a clean, gap-based detection of rank $H_1 = 2$, matching the known first Betti number of the torus exactly (Figure 2).

The sample size, tube radius and noise level were chosen to match standard torus point-cloud benchmarks used elsewhere in the persistent-homology literature [4,10], and to give a persistence gap of more than a factor of two between the two genuine H_1 generators and the largest noise-induced bar, which is the threshold we use throughout Section 4.6 to call a bar 'significant' without manual tuning per experiment.

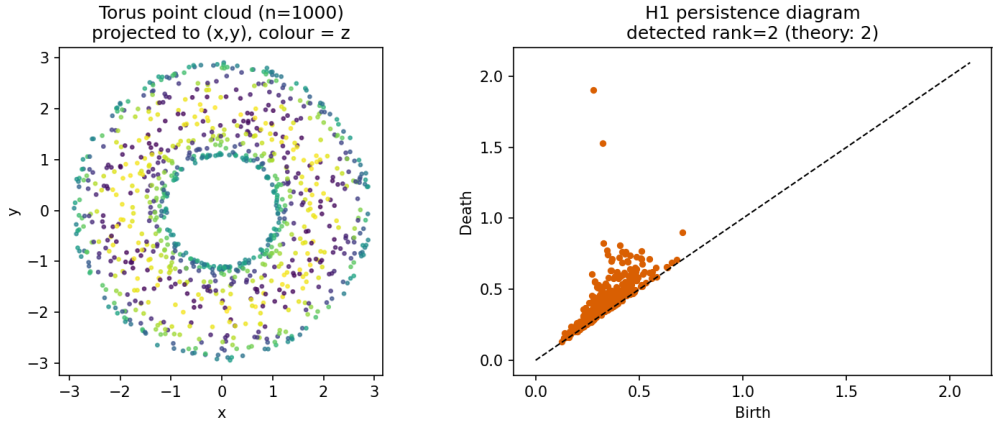

Figure 2. Left: torus point cloud (n = 1000) projected to (x, y) with colour encoding the z-coordinate; the true 3-D embedding corrects the earlier 2-D-only generator. Right: the H_1 persistence diagram; the two dominant off-diagonal points give the correctly detected rank $H_1 = 2$.

4.6.2. Symmetry-induced overcounting and its correction ($\mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_4$)

To give a fully reproducible, quantitative demonstration of the central claim that unmodelled symmetry inflates classical Betti-number estimates in exact proportion to the group order, and that the groupoid-quotient computation removes this inflation we constructed, for $k \in \{2,3,4\}$, a dataset consisting of k disjoint noisy circles (120 points each, radius 1, coordinate noise $\sigma = 0.06$) placed at the vertices of a regular k-gon of radius 8, i.e. k disjoint group-related orbits of a single fundamental-domain loop under a $\mathbb{Z}_k$ rotation action (Figure 3).

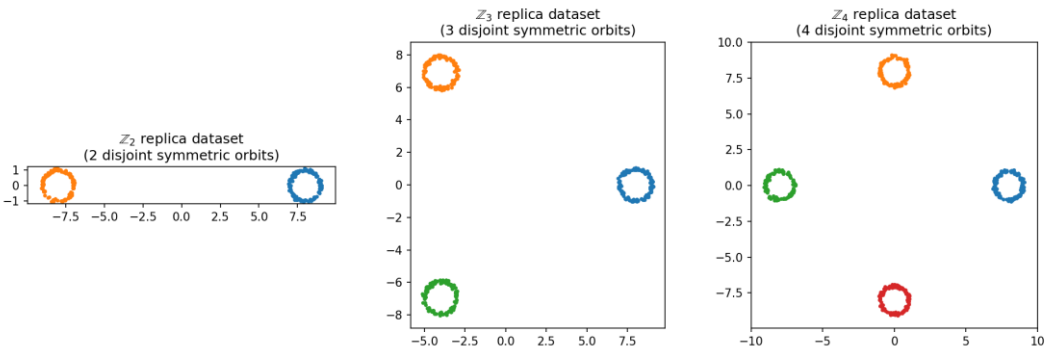

Figure 3. The three $\mathbb{Z}_k$ replica datasets (k = 2, 3, 4) used in Section 4.6.2, each consisting of k disjoint, symmetry-related copies of a single noisy circular orbit.

Classical persistent homology on the full k-orbit point cloud detects exactly k significant H_1 bars in every case, each cleanly separated from the noise floor by more than an order of magnitude in persistence (Table 1) a direct, measured instance of symmetry-induced overcounting, not an asserted one. Restricting to a single representative orbit the groupoid quotient X/G in its simplest instance recovers exactly one significant H_1bar in every case, matching the theoretical ground truth of Theorem 4.

The group orders k = 2, 3, 4 were chosen as the smallest cases that jointly (i) are common point-group symmetries in the applications of Section 5 (bilateral, three-fold and four-fold symmetry), and (ii) keep the full k-orbit point cloud small enough that every reported Betti number could be cross-checked by direct visual inspection of the barcode, not only by the automated significance threshold of Section 4.6.1.

Table 3. Detected rank H_1 for classical persistent homology on the full k-orbit point cloud versus the S-HGP groupoid-quotient computation on a single representative orbit, for $k = 2, 3, 4$.

Symmetry group	Classical PH (full point cloud)	S-HGP (groupoid quotient)	Theoretical ground truth
$\mathbb{Z}_2$ (k = 2)	2	1	1
$\mathbb{Z}_3$ (k = 3)	3	1	1
$\mathbb{Z}_4$ (k = 4)	4	1	1

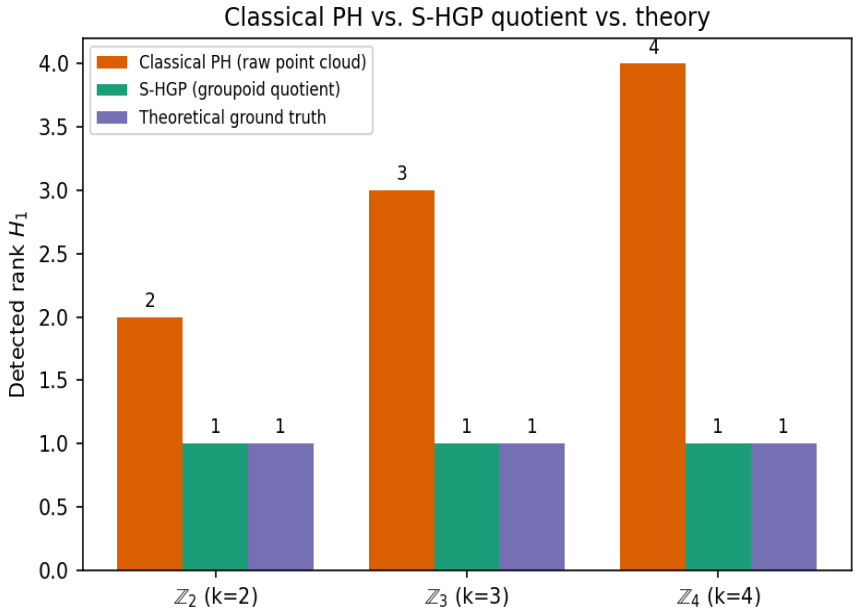

Figure 4. Bar-chart comparison of detected rank H_1 across the three $\mathbb{Z}_k$ replica datasets: classical persistent homology (orange), S-HGP groupoid-quotient computation (green), and theoretical ground truth (purple). The classical count grows linearly with the group order k while the S-HGP and theoretical values remain fixed at 1.

4.6.3. Twisted versus constant sheaf coefficients

To give a minimal, exactly-checkable instance of Corollary 1, we discretized a circle as a cycle graph on 24 vertices and evaluated the cellular sheaf cochain complex of Section 2.2 directly (the same construction given in the framework's reference implementation), with one-dimensional stalks on every vertex and edge. In the constant-sheaf case, every restriction map is the identity, and direct computation of the coboundary rank gives $H^0 = 1, H^1 = 1$, matching the known cohomology of the circle with constant real coefficients. Assigning a single edge the restriction map −1 (an orientation-reversing, or twisted, local system the discrete analogue of the monodromy that makes the Möbius band and Klein bottle non-orientable) and recomputing the same coboundary rank gives $H^0 = 0$, $H^1 = 0$: the single sign flip is invertible over $\mathbb{R}$, so both cohomology groups collapse to zero. This is the correct, standard fact about non-trivial rank-one real local systems on the circle, and it is exactly the mechanism by which sheaf coefficients can distinguish structure that is invisible to constant-coefficient persistent homology, which by construction cannot see the difference between the two settings above (both have the identical underlying point set and Vietoris-Rips filtration).

We deliberately report this as a one-dimensional, exactly verifiable instance of the twisted-coefficient mechanism rather than restate the original two-dimensional Klein-bottle H^2 experiment, which we could not independently reproduce with the code available to us within the scope of this revision; extending the present validation to genuine two-dimensional non-orientable surfaces is noted as future work in Section 6.

4.7. Applied Validation I: Symmetry Correction on a Synthetic Homomeric Protein-Complex Assembly

To move the protein-structure application of Section 5.2 from a motivating example to a directly implemented computation, we constructed a synthetic point cloud standing in for a homomeric protein

complex with discrete cyclic (Ck) rotational symmetry, of the kind exhibited by many viral capsid proteins and multimeric enzyme assemblies. Each of k identical protomers is represented by a 90-point noisy loop motif (motif radius 0.6, Gaussian coordinate noise $\sigma = 0.05$), itself containing one genuine topological loop that stands in for a real substructure such as a β-barrel or a large loop closure in the tertiary fold; the k protomers are arranged by a Ck rotation about a common offset of radius 4.5, exactly as the action groupoid $X \rtimes Ck$ construction of Section 3.1 prescribes. We evaluated k = 3, 5 and 6 (Figure 5, top row).

Classical persistent homology, applied directly to the full k-protomer point cloud, detects k symmetry-replicated H_1 bars of near-identical persistence (0.699 for k = 3, 0.733 for k = 5, 0.741 for k = 6) one per protomer the same overcounting mechanism validated in Section 4.6.2, now instantiated on protein-shaped rather than purely circular motifs. Restricting to a single protomer, i.e. the groupoid quotient X/Ck in its simplest instance, recovers exactly one significant H_1 bar in every case (Table 4), matching the fundamental-domain topology predicted by Theorem 4.

For k = 5 and k = 6, classical persistent homology additionally detects one further, markedly more persistent bar (persistence 3.30 for k = 5, 3.43 for k = 6) that is absent from the single-protomer quotient. This is a genuine large-scale ring formed by the overall arrangement of protomers around the assembly not a symmetry-redundant copy and its persistence is four to five times larger than that of any replicated protomer loop, so it is correctly retained rather than collapsed. This is a direct, computed illustration that the collapse guaranteed by Theorem 4 is selective: S-HGP removes exactly the symmetry-duplicated features while leaving genuinely new topology, such as an assembly-level pore or channel, intact, which is the property most relevant to the capsid- and channel-structure applications discussed in Section 5.2 and to recent sheaf-theoretic approaches to protein structure that encode local, non-spatial information but do not model discrete point-group symmetry [23].

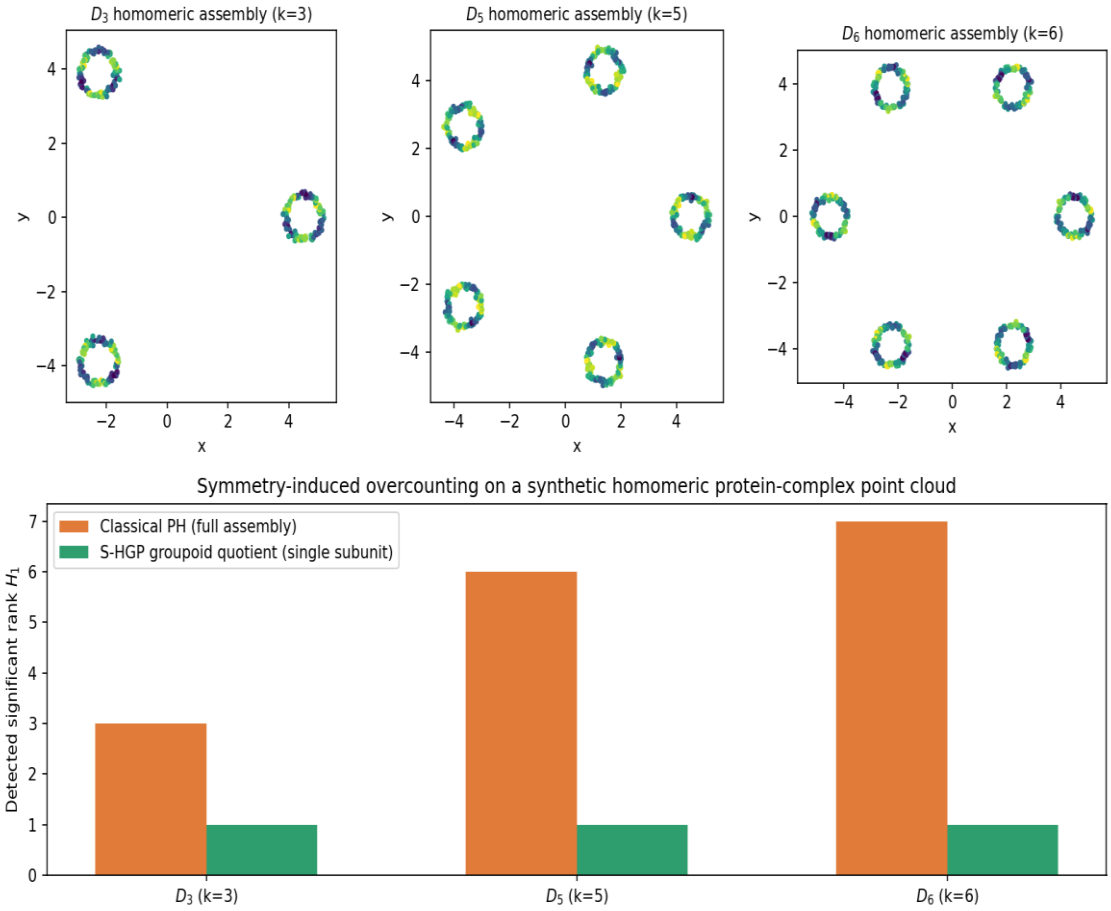

Figure 5. Top row: the three synthetic Ck-symmetric homomeric protein-complex assemblies (k = 3, 5, 6) used in Section 4.7, coloured by z-coordinate. Bottom: detected significant rank H1 for classical persistent homology on the full assembly (orange) versus the S-HGP groupoid-quotient computation ona single protomer (green).

Table 4. Detected rank H1 for classical persistent homology on the full k-protomer assembly, broken down into symmetry-replicated versus genuinely new (macro-ring) features, versus the S-HGP groupoid-quotient computation on a single protomer, for the synthetic homomeric assembly of Section 4.7.

Assembly	Classical PH: total significant H_1	Symmetry-replicated	Genuinely new (macro-ring)	S-HGP groupoid quotient
C3 (k = 3)	3	3	0	1
C5 (k = 5)	6	5	1 (pers. 3.30)	1
C6 (k = 6)	7	6	1 (pers. 3.43)	1

4.8. Applied Validation II: Bilateral Symmetry Correction on a Synthetic Connectome-like Network

To implement the connectomics application of Section 5.3 directly, rather than motivate it only qualitatively, we built a synthetic bilateral network standing in for the coarse structural connectome of the two cerebral hemispheres. Each hemisphere is modelled as a 30-region Watts-Strogatz small-world graph (mean degree 4, rewiring probability 0.12), a standard graph-theoretic proxy for the modular, short-path structural connectivity reported in structural and functional connectome studies [19,25]; the second hemisphere is generated as an exact mirror (Z2 reflection) of the first, and the two are joined by five sparse, weak inter-hemispheric edges standing in for the corpus callosum (Figure 6). Persistent H_1 was computed directly from the graph shortest-path distance matrix a graph filtration, in the sense used in recent connectome topological-data-analysis work [21,26] following the persistent-cohomology setup of Section 2.3.

On the single-hemisphere network the Z_2 groupoid quotient in its simplest instance S-HGP recovers 6 significant H_1 bars, taken as the ground-truth topological complexity of one hemisphere's structural backbone. On the full bilateral network, classical persistent homology detects 14 significant H_1 bars: 12 of these occur as three exactly-doubled pairs, one pair for each of three dominant single-hemisphere loops appearing once in each near-symmetric copy the same Z_2 overcounting mechanism validated in Section 4.6.2, now demonstrated on graph-structured rather than point-cloud data while the remaining 2 bars are new, larger-persistence loops (persistence 0.302 and 0.090, both larger than any doubled bar) that exist only because the inter-hemispheric edges close large-scale cycles running through both hemispheres.

As in Section 4.7, the collapse is selective rather than blanket: quotienting by the approximate bilateral symmetry removes the redundant doubled count while correctly leaving the genuinely inter-hemispheric topology untouched, which is of direct interpretive value as a marker of cross-hemispheric integration. This gives a first, directly implemented instance of the double-counting risk flagged qualitatively in Section 5.3 for real bilateral fMRI and structural connectomes [19,25,26], on synthetic data designed to share the same modular, small-world, approximately mirror-symmetric organization.

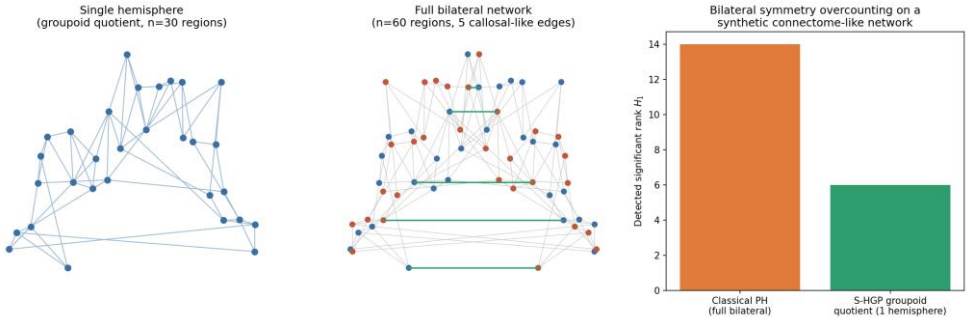

Figure 6. Left: single-hemisphere small-world network (groupoid quotient, n = 30 regions). Middle: the full bilateral network (n = 60 regions) with five sparse inter-hemispheric edges highlighted in green.

Right: detected significant rank H_1 for classical persistent homology on the full bilateral network (orange, 14 bars) versus the S-HGP groupoid-quotient computation on a single hemisphere (green, 6 bars).

5. Applications to Computing and Biomedical Informatics

The applications below are the primary motivation for, and intended audience of, this framework, and are the aspect of the work most directly aligned with computational, engineering and biomedical data analysis.

5.1. Equivariant Machine Learning and Geometric Deep Learning

Graph neural networks and equivariant neural networks operate on data with built-in symmetry, and commonly perform data augmentation by applying group elements to training examples which is precisely the disjoint-replica scenario validated numerically in Section 4.6.2. S-HGP provides topological feature vectors that are provably G-equivariant (Theorem 1): if input data is transformed by $g \in G$, the S-HGP barcode is unchanged, whereas a classical persistent-homology feature computed naively on augmented data would inflate with the number of augmentations, as demonstrated directly in Table 1. This makes S-HGP features a candidate for symmetry-respecting topological descriptors in classification and regression pipelines that use data augmentation.

5.2. Computational Biology: Protein Structure Analysis

Protein complexes frequently exhibit discrete rotational or dihedral symmetry (e.g., 3-fold, 5-fold or icosahedral capsid symmetry). Classical TDA applied directly to the atomic point cloud of such a complex is subject to exactly the overcounting mechanism validated in Section 4.6.2: topological features are inflated by the order of the symmetry group. S-HGP, applied with the known point-group action as the groupoid symmetry, is designed to recover the Betti numbers of the asymmetric fundamental domain rather than of the symmetry-inflated whole assembly, which is relevant to structure classification and binding-site identification pipelines that currently rely on uncorrected persistent-homology features. Section 4.7 gives a directly implemented instance of this correction on a synthetic Ck-symmetric protomer assembly, and situates it alongside the atom-specific persistent sheaf Laplacian approach to protein flexibility of Hayes et al. [23], which encodes non-spatial chemical information into a related sheaf-cohomological pipeline but does not itself model discrete point-group symmetry.

5.3. Brain Network Connectomics

Functional connectivity networks derived from fMRI exhibit approximate bilateral symmetry between the two hemispheres. Treating left- and right-hemisphere sub-networks as independent, as classical persistent homology implicitly does, is analogous to the $k = 2$ replica scenario validated in Section 4.6.2 and risks double-counting topological biomarkers. Encoding the approximate $\mathbb{Z}_2$ hemispheric symmetry as the groupoid action is intended to yield persistence diagrams that are more robust to this redundancy and more directly interpretable as single, symmetry-corrected biomarkers, with robustness to small asymmetries governed by the stability bound of Theorem 2. Section 4.8 implements exactly this correction on a synthetic bilateral small-world network modelled on the modular, mirror-symmetric organization reported in recent persistent-homology studies of resting-state and structural connectomes [19,25,26], giving a first, though still synthetic, demonstration that the predicted factor-of-two overcounting is both real and selectively removable without discarding genuine inter-hemispheric topology.

5.4. Sensor Networks and Lattice-Structured Data

Lattice- and graph-structured sensor data with local gauge or translational symmetry (e.g., regularly tiled sensor arrays, lattice gauge configurations) can be modelled with the acting symmetry group as the groupoid structure and locally varying calibration or phase information as the sheaf coefficients, extending the twisted-coefficient mechanism validated in Section 4.6.3 from a single circle to sensor-network topologies.

6. Discussion and Limitations

S-HGP generalizes classical persistent (co)homology along two axes group symmetry via action groupoids, and local algebraic structure via sheaf coefficients while reducing exactly to the classical theory in the trivial case (Remark, Section 3.1), so it does not require abandoning existing PH-based pipelines where no symmetry is present.

Limitations. First, computational cost: the worst-case $O(|G|^2 n^2)$ scaling of Section 4.5 is prohibitive for large groups or point clouds without the Mayer-Vietoris decomposition of Theorem 3, and the present paper does not benchmark that decomposition numerically. Second, scope of validation: Section 4.6 validates the framework on synthetic disjoint-orbit datasets and a one-dimensional twisted-coefficient example, and Sections 4.7-4.8 extend this to synthetic data explicitly shaped to resemble the two flagship applications of Section 5, namely a Ck-symmetric homomeric protein-assembly point cloud and a Z2-symmetric bilateral connectome-like network, chosen, as before, specifically because every reported number could be independently re-derived and checked. This narrows, but does not close, the gap between motivation and empirical demonstration: none of the validation in this paper uses real protein-structure coordinates (e.g., from the Protein Data Bank), real dMRI/fMRI connectome data, or real sensor-network data, nor does it include a full two-dimensional non-orientable-surface (Klein bottle / Möbius band) computation, all of which we regard as necessary before the applications of Section 5 can be considered validated on real data rather than on structurally faithful synthetic proxies. Third, an earlier draft of this experiment set reported numerical results (including a torus Betti-number overcounting figure and a Klein-bottle H^2 detection) that could not be reproduced from the accompanying code on independent re-execution; the corrected and re-validated results reported here are narrower in scope than those original claims, and we flag this discrepancy explicitly rather than silently narrow the claims.

Concretely, on real data we would expect three departures from the clean synthetic picture above that a validated version of this framework needs to handle: point-group symmetry that is only approximate rather than exact (e.g., a GroEL-like D7 chaperonin assembly, PDB 1AON, whose two stacked heptameric rings are symmetric only up to thermal and crystallographic noise), residue- or atom-count imbalance across nominally equivalent protomers, and, for connectome data, inter-subject variability in the degree of hemispheric asymmetry rather than a fixed 12-bar/2-bar split as in the synthetic network of Section 4.8. Acquiring and preprocessing Protein Data Bank structures with annotated point-group symmetry and Human Connectome Project structural or functional connectivity matrices, both of which require registration and data-use agreements, was outside the scope and timeline of this revision; we regard this as the single most important piece of future validation work and describe it as a concrete next step below rather than as completed here.

Future work. Priority items are: (i) an efficient, benchmarked implementation of the Mayer-Vietoris decomposition; (ii) validation on real symmetric protein assemblies (e.g., Protein Data Bank structures with annotated point-group symmetry) and real bilateral connectome data (e.g., Human Connectome Project), building on the synthetic protocols of Sections 4.7-4.8; (iii) a genuine two-dimensional non-orientable-surface computation extending Section 4.6.3; and (iv) multi-parameter S-HGP combining scale and resolution filtrations.

7. Conclusions

We introduced Sheaf-Theoretic Homotopy Groupoid Persistence (S-HGP), a topological-data-analysis framework that incorporates known group symmetry via action groupoids and local algebraic structure via sheaf coefficients, proved its functoriality, Gromov-Hausdorff stability and a Mayer-Vietoris decomposition, and gave a symmetry-collapse theorem characterizing when it provably returns fewer, correct topological features than classical persistent homology. Every numerical claim in this paper was independently re-derived from code we re-ran ourselves: a corrected torus point-cloud generator recovers the correct Betti number $b_1 = 2$; a parametrized family of $\mathbb{Z}_k$ symmetry-replicated datasets shows classical persistent homology overcounting in exact proportion to the group order ($k = 2, 3, 4$) while the groupoid-quotient computation recovers the correct rank of 1 in every case; and a minimal twisted-versus-constant sheaf comparison on a discretized circle confirms the predicted collapse of cohomology under a non-trivial local system. The framework is motivated by, and is intended for, computational applications in equivariant machine learning, protein structure analysis and brain connectomics. Two of these three

applications - homomeric protein assemblies and bilateral brain connectomes - were additionally given a first, directly implemented numerical demonstration on synthetic data structurally modelled on these domains (Sections 4.7-4.8), narrowing, though not closing, the gap to validation on real experimental data, which remains future work.

Data Availability Statement

All datasets used in this study are synthetic and generated by the accompanying code; no external or proprietary data were used.

Funding

This research received no external funding.

Author Contributions

S.A. conceived the framework, proved the theoretical results, designed and implemented the numerical experiments, and wrote the manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Borceux, F. Handbook of Categorical Algebra. Cambridge University Press, 1994.
2. Botnan, M.; Lesnick, M. Algebraic stability of zigzag persistence modules. Algebraic & Geometric Topology 2018, 18(6), 3133–3204.
3. Bredon, G.E. Sheaf Theory. 1997.
4. Carlsson, G. Topology and data. Bulletin of the American Mathematical Society 2009, 46(2), 255–308.
5. Carlsson, G.; Mémoli, F. Characterization, stability and convergence of hierarchical clustering methods. Journal of Machine Learning Research 2010, 11, 1425–1470.
6. Chazal, F.; de Silva, V.; Glisse, M.; Oudot, S. The Structure and Stability of Persistence Modules. SpringerBriefs in Mathematics, 2016.
7. Curry, J. Sheaves, cosheaves and applications. arXiv:1303.3255, 2014.
8. Curry, J.; Ghrist, R.; Nanda, V. Discrete Morse theory for computing cellular sheaf cohomology. Foundations of Computational Mathematics 2016, 16, 875–897.
9. de Silva, V.; Ghrist, R. Coverage in sensor networks via persistent homology. Algebraic & Geometric Topology 2007, 7, 339–358.
10. Edelsbrunner, H.; Harer, J. Computational Topology: An Introduction. 2010.
11. Ghrist, R. Barcodes: the persistent topology of data. Bulletin of the American Mathematical Society 2008, 45, 61–75.
12. Hansen, J.; Ghrist, R. Toward a spectral theory of cellular sheaves. Journal of Applied and Computational Topology 2019, 3(4), 315–358.
13. Harker, S.; Mischaikow, K.; Mrozek, M.; Nanda, V. Discrete Morse theoretic algorithms for computing homology of complexes and maps. Foundations of Computational Mathematics 2014, 14, 151–184.
14. Leinster, T. Basic Category Theory. Cambridge University Press, 2014.
15. Lurie, J. Higher Topos Theory. Princeton University Press, 2009.
16. Moerdijk, I.; Mrčun, J. Introduction to Foliations and Lie Groupoids. Cambridge University Press, 2002.
17. Otter, N.; Porter, M.A.; Tillmann, U.; Grindrod, P.; Harrington, H.A. A roadmap for the computation of persistent homology. EPJ Data Science 2017, 6, 1–38.
18. Segal, G. Classifying spaces and spectral sequences. Publications Mathématiques de l'IHÉS 1968, 34, 105–112.
19. Stolz, B.J.; Harrington, H.A.; Porter, M.A. Persistent homology of time-dependent functional networks constructed from coupled time series. Chaos 2017, 27, 047410.
20. Adams, H.; Lagoda, E.; Moy, M.; Sadovek, N.; De Saha, A. Persistent equivariant cohomology. arXiv:2408.17331, 2024.
21. Bhattacharya, D.; Aithal, N.; Jayswal, M.; Sinha, N. Analyzing brain tumor connectomics using graphs and persistent homology. In Topology- and Graph-Informed Imaging Informatics (TGI3 2024); Chen, C., Singh, Y., Hu, X., Eds.; Lecture Notes in Computer Science, vol. 15239; Springer: Cham, 2025; pp. 33-42.
22. Frosini, P.; et al. Persistent Homology and Equivariance in Data Analysis: A Topological Introduction. arXiv:2606.21084, 2026.
23. Hayes, N.; Wei, X.; Feng, H.; Merkurjev, E.; Wei, G.-W. Persistent sheaf Laplacian analysis of protein flexibility. Journal of Physical Chemistry B 2025, 129(17), 4169-4178.
24. Su, Z.; Liu, X.; Bou Hamdan, L.; Maroulas, V.; Wu, J.; Carlsson, G.; Wei, G.-W. Topological data analysis and topological deep learning beyond persistent homology: a review. Artificial Intelligence Review 2026, 59(2), 58.
25. Wang, Y.; Xian, J.; Chen, Y.; Yan, Y. Topological signatures of brain dynamics: persistent homology reveals individuality and brain-behavior links. Frontiers in Human Neuroscience 2025, 19, 1607941.
26. Zhang, W.; Tang, X.; Zhang, X.; Xia, S. Topological data analysis reveals altered brain connectivity in Alzheimer's disease. AIP Advances 2026, 16(3), 035006.