

Article

Not peer-reviewed version

Fiber Bundle Learning: A Topological Framework for Classification Using Homology and Discrete Connections

[Arturo Tozzi](#)*

Posted Date: 28 November 2025

doi: 10.20944/preprints202511.2221.v1

Keywords: topological machine learning; fiber bundles; holonomy; geometric learning; invariants; classification algorithms



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Fiber Bundle Learning: A Topological Framework for Classification Using Homology and Discrete Connections

Arturo Tozzi

ASL Napoli 1 Centro, Distretto 27, Via Comunale del Principe 13/a, 80145 Naples, Italy; tozziarturo@libero.it

Abstract

Many machine-learning tasks involve structured data whose geometry, local feature distributions and global organization interact in ways that are not well captured by existing methods based on vectorization, graph metrics or homological signatures. We introduce Fiber Bundle Learning (FBL), a topological framework that represents each data sample as a discrete fiber bundle and extracts a classification signature combining persistent homology, local feature geometry and gluing structure. FBL builds a base space from the coarse geometry of each object, models local feature patches as fibers and estimates transition maps between neighboring fibers to construct a discrete connection. From this representation, FBL computes a set of invariants: persistent homology of the base, fibers and total space; holonomy obtained by transporting fiber states along cycles; curvature-like quantities measuring transition inconsistency; discrete analogues of characteristic classes. These components are assembled into a fixed-length feature vector that can be used with any standard classifier. We show that FBL yields a signature with three desirable theoretical properties: **stability** under perturbations of geometry and local features; **invariance** under isometries and global fiber reparameterizations; **robustness** to sampling noise. Our synthetic experiments show that FBL distinguishes twisted from untwisted bundles with identical homology, a distinction classical topological methods fail to capture. Additional tests quantify the system's resistance to noise, its invariance to geometric transformations and the contribution of each signature component. Taken together, our results indicate that representing data through fiber-bundle structure may provide an effective tool for classifying complex, multi-level objects.

Keywords: topological machine learning; fiber bundles; holonomy; geometric learning; invariants; classification algorithms

1. Introduction

Several areas of machine learning and applied topology have sought to represent structured data by exploiting geometric, combinatorial or topological properties. Three major traditions provide the necessary context. Topological data analysis extracts global shape information through persistent homology and related vectorizations, but treats data as a single undifferentiated space and therefore cannot distinguish objects that share identical homology yet differ in how local feature distributions vary over the underlying geometry (Laubenbacher and Hastings 2019; Bukkuri, Andor, and Darcy 2021; Chazal and Michel 2021; Xu, Drougard, and Roy 2021; Skaf and Laubenbacher 2022; El-Yaagoubi, Chung, and Ombao 2023; Smith et al. 2024). Geometric and manifold-based approaches capture local relationships through graph, spectral or diffusion operators but do not explicitly represent how feature spaces change along cycles or how local patches are glued together (Paulsen, Gramstad, and Collas 2015; Tsamir-Rimon and Borenstein 2023; Hu et al. 2024; Sekmen and Bilgin 2024; Song et al. 2024; Gao et al. 2025). Gauge-inspired constructions introduce parallel transport and equivariant operations, yet typically enforce symmetry constraints without computing homological invariants, characteristic-class-like quantities or measures of connection inconsistency (Maîtrejean

2008; Sauer 2022; Maîtrejean 2024; Nagai, Ohno, and Tomiya 2025). Each of these traditions captures a meaningful aspect of structure (global topology, neighborhood geometry or symmetry) but none integrates these components into a unified, stable representation capable of detecting differences in internal gluing structure.

To address these limitations, we introduce **Fiber Bundle Learning (FBL)**, a topological framework that models a structured dataset as a discrete fiber bundle and extracts from it a set of invariants suitable for machine-learning tasks (Shao et al. 2019; Diksha and Biswas 2022; Eadie et al. 2023; Shams et al. 2023). Each object is viewed as a base space capturing its coarse geometry, with a fiber attached to every point representing the local feature distribution. Transition maps between neighboring fibers are estimated and combined to form a discrete connection, enabling computation of holonomy along cycles, curvature-like quantities that measure transition inconsistency and discrete analogues of characteristic classes. Together with the persistent homology of the base, fibers and total space, these invariants may define a fixed-length signature that can be used with any classifier.

We show that the resulting signature exhibits stability under small perturbations of geometry and features, invariance to global reparameterizations and base isometries and robustness to sampling noise. Synthetic experiments illustrate the qualitative differences captured by the fiber-bundle representation: structures with identical homology but different twisting patterns become distinguishable, while isometric or reparameterized versions of the same object yield nearly identical signatures. This could provide a mechanism for capturing how local feature geometry and global structural consistency jointly determine the identity of a data object.

We will proceed as follows: the methodology defines the discrete fiber-bundle construction and signature; theoretical guarantees establish stability and invariance; experiments examine discriminability, robustness and component contributions; and the final discussion integrates these findings into a unified perspective.

2. Methods

We describe here how data samples are represented as discrete fiber bundles, how topological and bundle theoretic invariants are computed from them and how these invariants are assembled into a feature representation used in a standard learning pipeline.

2.1. Data Representation and Notation

The starting point is the representation of each structured sample as a finite set of points endowed with geometric coordinates and local features. Formally, a sample is written as $X = \{(x_i, f_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ encodes geometry and $f_i \in \mathbb{R}^p$ encodes a feature vector. From this point cloud we construct a discrete approximation of a base space B , a collection of fibers $\{F_i\}$ and a total space E . The base space is modeled as a geometric graph $G_B = (V, E_B)$ with vertex set $V = \{1, \dots, n\}$ and edge set $E_B \subseteq V \times V$. The graph is obtained by connecting vertices i and j if $\|x_i - x_j\|_2 \leq \epsilon$ or if j belongs to the k -nearest neighbors of i . We denote the adjacency matrix of G_B by $A \in \{0,1\}^{n \times n}$, where $A_{ij} = 1$ if $(i,j) \in E_B$. For each vertex i , the associated fiber F_i is defined as a finite subset of feature space, for example $F_i = \{f_j : j \in \mathcal{N}_r(i)\}$ where $\mathcal{N}_r(i) = \{j : \|x_i - x_j\|_2 \leq r\}$ is a spatial neighborhood. The total space is the set $E = \{(x_i, f_i)\}_{i=1}^n \subset \mathbb{R}^{d+p}$. All subsequent constructions are applied to these discrete objects.

2.2. Base Space Construction and Homology Computation

The construction of the base space proceeds by thickening the graph into a simplicial complex and computing persistent homology on the resulting filtration. Given the point cloud $\{x_i\}$, we define a Vietoris Rips complex $VR(X, \alpha)$ at scale $\alpha > 0$ as the abstract simplicial complex whose vertices are the indices $1, \dots, n$ and where a finite set $\sigma \subseteq \{1, \dots, n\}$ forms a simplex if $\|x_i - x_j\|_2 \leq \alpha$ for all $i, j \in \sigma$ (Wang and Wei 2016; Wang et al. 2023; Zhang et al. 2023; Aggarwal and Periwai 2024). The

filtration $\{\text{VR}(X, \alpha)\}_{\alpha \geq 0}$ is obtained by increasing α . For each homological degree $k \in \{0, 1, \dots, k_{\max}\}$, persistent homology computes a multiset of intervals $\mathcal{B}_k(B)$ representing the birth and death times of homology classes in $H_k(\text{VR}(X, \alpha))$. We denote the barcode by $\mathcal{B}(B) = \{\mathcal{B}_k(B)\}_{k=0}^{k_{\max}}$. To obtain a numerical representation, a vectorization map Φ_{PH} is applied, for instance a persistence landscape or persistence image. A persistence landscape $\lambda_k: \mathbb{R} \rightarrow \mathbb{R}$ is defined from the intervals $[b_\ell, d_\ell)$ in $\mathcal{B}_k(B)$ by setting $\lambda_k(t) = \sup_\ell \max(0, \min(t - b_\ell, d_\ell - t))$. A finite dimensional approximation is obtained by sampling λ_k on a grid $\{t_1, \dots, t_m\}$ and stacking the values, yielding a feature vector $h_B = \Phi_{\text{PH}}(\mathcal{B}(B)) \in \mathbb{R}^{m(k_{\max}+1)}$. In practice, Vietoris Rips complexes and persistent homology are implemented using Python with the GUDHI or Ripser libraries, while k nearest neighbor graphs are built using scikit-learn and NetworkX.

2.3. Fiber Construction and Fiber Homology Aggregation

The modeling of local feature structure is based on computing persistent homology on each fiber patch and then aggregating these local descriptors. For each base vertex i , the fiber $F_i = \{f_j: j \in \mathcal{N}_r(i)\}$ is treated as a point cloud in \mathbb{R}^p . We define a local Rips complex $\text{VR}(F_i, \beta)$ in the same fashion as for the base, using a feature scale parameter β . The associated filtration $\{\text{VR}(F_i, \beta)\}_{\beta \geq 0}$ yields barcodes $\mathcal{B}_k(F_i)$ for each degree k . To avoid high dimensionality, each local barcode is vectorized with the same map Φ_{PH} , for instance by constructing a small persistence landscape or persistence image with fixed resolution. Denoting the vectorization for fiber i by $v_i = \Phi_{\text{PH}}(\mathcal{B}(F_i)) \in \mathbb{R}^q$, we obtain a family $\{v_i\}_{i=1}^n$. The global fiber homology descriptor is then defined through aggregation statistics such as empirical mean and variance. Explicitly, the mean vector is $\mu_F = \frac{1}{n} \sum_{i=1}^n v_i$ and the covariance matrix is $\Sigma_F = \frac{1}{n} \sum_{i=1}^n (v_i - \mu_F)(v_i - \mu_F)^\top$. To keep the representation finite and tractable, we either vectorize Σ_F by taking its upper triangular entries or we restrict to diagonal elements, yielding a feature vector $h_F = [\mu_F, \text{diag}(\Sigma_F)] \in \mathbb{R}^{2q}$ or similar. Local neighborhoods are extracted using kd tree queries from scikit-learn and persistent homology on fiber point clouds is again computed with GUDHI or Ripser, while aggregation is implemented in NumPy.

2.4. Total Space Embedding and Persistent Homology

The construction of the total space descriptor incorporates interactions between geometry and features by embedding the points into a joint space and computing persistent homology on that combined representation. Each pair (x_i, f_i) is mapped to a concatenated vector $z_i = (x_i, \gamma f_i) \in \mathbb{R}^{d+p}$, where $\gamma > 0$ is a scaling parameter that balances the relative influence of geometry and features in the joint metric. The set $Z = \{z_i\}_{i=1}^n$ is used to define a Rips complex $\text{VR}(Z, \rho)$ with respect to the Euclidean distance $\|z_i - z_j\|_2$. As before, the filtration $\{\text{VR}(Z, \rho)\}_{\rho \geq 0}$ yields barcodes $\mathcal{B}_k(E)$ that are vectorized into a feature vector $h_E = \Phi_{\text{PH}}(\mathcal{B}(E))$. The explicit dependence on γ allows tuning the method to regimes where geometric structure or feature structure dominates; in practice γ is selected through cross validation. From an implementation standpoint, the computation is identical to that of the base but applied to a higher dimensional point cloud. The same persistent homology tools and linear algebra libraries are used and all parameters $(k, \epsilon, r, \beta, \rho, \gamma)$ are stored for reproducibility.

2.5. Discrete Connection and Transition Map Estimation

The estimation of a discrete connection over the base graph provides the transition maps that relate fibers at neighboring vertices. For each edge $(i, j) \in E_B$, we seek a linear map $T_{ij}: \mathbb{R}^p \rightarrow \mathbb{R}^p$ that approximates how features transform from the neighborhood of i to that of j . Let $F_i = \{f_{i\ell}\}_{\ell=1}^{m_i}$ and $F_j = \{f_{j\ell}\}_{\ell=1}^{m_j}$ be sampled fiber elements. We construct paired samples by matching points in F_i and F_j using nearest neighbors or an optimal transport plan. In the simplest least squares formulation, we select matching indices $\mathcal{M}_{ij} \subseteq \{1, \dots, m_i\} \times \{1, \dots, m_j\}$ and solve

$$T_{ij}^* = \arg \min_{T \in \mathbb{R}^p \times \mathbb{R}^p} \sum_{(\ell, r) \in \mathcal{M}_{ij}} \|T f_{i\ell} - f_{jr}\|_2^2 + \lambda \|T\|_F^2,$$

where $\lambda \geq 0$ is a regularization parameter and $\|\cdot\|_F$ denotes the Frobenius norm. This yields a closed form solution

$$T_{ij}^* = \left(\sum_{(\ell,r) \in \mathcal{M}_{ij}} f_{jr} f_{i\ell}^T \right) \left(\sum_{(\ell,r) \in \mathcal{M}_{ij}} f_{i\ell} f_{i\ell}^T + \lambda I_p \right)^{-1}.$$

Alternatively, the matching \mathcal{M}_{ij} can be defined via an optimal transport plan π_{ij} between empirical measures on F_i and F_j , computed with the Python Optimal Transport (POT) library and the least squares objective is weighted accordingly. Transition maps are stored in sparse fashion as $\{T_{ij}^*\}_{(i,j) \in E_B}$. Linear algebra operations are carried out using NumPy and SciPy.

2.6. Holonomy, Curvature, Characteristic Class Proxies and Signature Assembly

The discretized connection induces holonomy and curvature like quantities on cycles and triangles of the base graph, which are summarized into numerical descriptors. Given a simple cycle $\gamma = (i_1, \dots, i_m, i_1)$ in G_B , the holonomy operator is defined as the ordered product

$$H_\gamma = T_{i_m i_1}^* T_{i_{m-1} i_m}^* \cdots T_{i_1 i_2}^*.$$

To compress this matrix into a scalar descriptor we compute invariants such as the Frobenius norm distance from the identity, $\text{hol}(\gamma) = \|H_\gamma - I_p\|_F$, the spectral radius $\rho(H_\gamma)$ or the argument of the determinant $\arg \det(H_\gamma)$. A cycle basis of the graph is computed using NetworkX and descriptors are evaluated on each basis cycle, yielding a set $\{\text{hol}(\gamma_k)\}_k$. Curvature like quantities are defined on oriented triangles (i, j, k) in the flag complex of G_B as

$$C_{ijk} = T_{ik}^* - T_{jk}^* T_{ij}^*,$$

and summarized by $\text{curv}(i, j, k) = \|C_{ijk}\|_F$. Characteristic class proxies are defined by constructing Čech type cocycles on triple overlaps of vertex neighborhoods from the transition functions, for instance by forming signs $\sigma_{ijk} = \text{signdet}(T_{jk}^* T_{ij}^* T_{ki}^*)$ and aggregating parities $\prod_{(i,j,k)} \sigma_{ijk}$ or counting nontrivial loops with non identity phase. All these scalar values are summarized by empirical moments and histograms and concatenated into vectors h_{hol} , h_{curv} and h_{CC} . The final signature is the concatenation

$$\Sigma(X) = [h_B, h_F, h_E, h_{\text{hol}}, h_{\text{curv}}, h_{\text{CC}}] \in \mathbb{R}^D,$$

with dimension D determined by the chosen vectorizations and histogram resolutions. These steps are implemented with NumPy for matrix operations and NetworkX for cycle and triangle enumeration, while histogramming uses NumPy or SciPy.

2.7. Learning Pipeline and Software Tools

The learning stage treats the signature as a fixed dimensional feature vector for supervised classification. Given a dataset $\{(X^{(s)}, y^{(s)})\}_{s=1}^S$, we compute $\Sigma^{(s)} = \Sigma(X^{(s)})$ for all samples, forming a feature matrix $Z \in \mathbb{R}^{S \times D}$ with rows $Z_s = \Sigma^{(s)}$. A standard linear or kernel support vector machine is then trained by solving

$$\min w, b, \xi \frac{1}{2} \|w\|_2^2 + C \sum_{s=1}^S \xi_s, \text{subject to } w^\top \phi(Z_s) + b \geq 1 - \xi_s, \xi_s \geq 0,$$

where ϕ is either the identity (linear SVM) or an implicit feature map induced by a kernel K . In experiments we used radial basis function kernels $K(u, v) = \exp(-\gamma \|u - v\|_2^2)$ and cross validation to select C and γ . The entire pipeline is implemented in Python, with NumPy and SciPy for numerical computations, scikit-learn for SVM training and model evaluation, GUDHI or Ripser for persistent homology, NetworkX for graph operations, POT for optimal transport based transition estimation and standard Python tools for data handling and reproducibility. The complete implementation, including data generation scripts, signature extraction routines and evaluation code, is provided in the Supplementary Materials (Supplementary Code Section)

Overall, we specify a complete pipeline from raw structured samples to numerical signatures and trained classifiers, including the geometric, topological and connection related computations and the software tools supporting each step.

3. Theoretical Guarantees

We formalize here the stability, invariance and robustness properties of the Fiber Bundle Learning (FBL) signature. We consider each data sample

$$X = \{(x_i, f_i)\}_{i=1}^n \subset \mathbb{R}^d \times \mathbb{R}^p$$

and denote by B the base space constructed from the geometric coordinates $\{x_i\}$, by F_i the fiber patch associated to vertex i and by E the total-space embedding obtained from the concatenated vectors $z_i = (x_i, \gamma f_i)$. The complete signature is

$$\Sigma(X) = (h_B, h_F, h_E, h_{\text{hol}}, h_{\text{curv}}, h_{\text{CC}}).$$

We analyze perturbations of the input and their effect on each component. Throughout the section, we let the perturbed sample be

$$X' = \{(x'_i, f'_i)\}_{i=1}^n$$

and assume the pointwise perturbation bounds

$$\|x_i - x'_i\|_2 \leq \varepsilon, \|f_i - f'_i\|_2 \leq \varepsilon,$$

for all i . All vectorizations of barcodes are assumed to be Lipschitz with respect to the bottleneck or p -Wasserstein distance, which is standard in topological data analysis.

Stability of homological components. We first analyse the persistent-homology features of the base, fibers and total space.

3.1. Definition of Distances

Given two finite point sets $P, Q \subset \mathbb{R}^m$, let

$$d_H(P, Q) = \max \left(\sup_{p \in P} \inf_{q \in Q} \|p - q\|_2, \sup_{q \in Q} \inf_{p \in P} \|p - q\|_2 \right)$$

denote the Hausdorff distance.

The bottleneck distance between persistence diagrams is denoted by $d_B(\cdot, \cdot)$.

Theorem 1 (Stability of base-space homology).

Let B and B' be the Vietoris–Rips filtrations associated to $\{x_i\}$ and $\{x'_i\}$. Then

$$d_B(\mathcal{B}(B), \mathcal{B}(B')) \leq d_H(\{x_i\}, \{x'_i\}) \leq \varepsilon.$$

If $h_B = \Phi_{\text{PH}}(\mathcal{B}(B))$ and $h'_B = \Phi_{\text{PH}}(\mathcal{B}(B'))$ and Φ_{PH} is L -Lipschitz, then

$$\|h_B - h'_B\| \leq L \varepsilon.$$

Theorem 2 (Stability of fiber homology aggregation).

Let F_i and F'_i denote the fiber patches. Then

$$d_B(\mathcal{B}(F_i), \mathcal{B}(F'_i)) \leq d_H(F_i, F'_i) \leq \varepsilon.$$

Let v_i, v'_i be their vectorized barcodes and assume each vectorization is L_F -Lipschitz. If

$$h_F = \frac{1}{n} \sum_i v_i, h'_F = \frac{1}{n} \sum_i v'_i,$$

then

$$\|h_F - h'_F\| \leq L_F \varepsilon.$$

Theorem 3 (Stability of total-space homology).

Let $Z = \{(x_i, \gamma f_i)\}$, $Z' = \{(x'_i, \gamma f'_i)\}$. Then

$$d_H(Z, Z') \leq \max\{\varepsilon, \gamma \varepsilon\}.$$

Hence,

$$\|h_E - h'_E\| \leq L_E \max\{\varepsilon, \gamma \varepsilon\}.$$

3.2. Stability of the Discrete Connection

Transition maps T_{ij} are computed by a least-squares or optimal-transport alignment between fibers. Their perturbation depends on the local perturbation of fibers.

Assumption (Lipschitz continuity of the estimator).

For each edge (i, j) , the transition estimate satisfies

$$\|T_{ij}(F_i, F_j) - T_{ij}(F'_i, F'_j)\|_F \leq L_T(d_H(F_i, F'_i) + d_H(F_j, F'_j))$$

for some constant $L_T > 0$.

This assumption holds for standard least-squares estimators, where perturbed design matrices produce linearly perturbed minimizers.

Theorem 4 (Stability of transition maps).

Under the estimator assumption,

$$\|T_{ij} - T'_{ij}\|_F \leq 2L_T \varepsilon.$$

3.3. Stability of Holonomy

Holonomy for a cycle $\gamma = (i_1, \dots, i_m, i_1)$ is defined by

$$H_\gamma = T_{i_m i_1} T_{i_{m-1} i_m} \cdots T_{i_1 i_2}.$$

Let the perturbed holonomy be H'_γ .

Theorem 5 (Holonomy stability).

For a cycle of length m ,

$$\|H_\gamma - H'_\gamma\|_F \leq m(2L_T \varepsilon) \max_k \|T_{i_k i_{k+1}}\|_F^{m-1} + O(\varepsilon^2).$$

In particular, if transition maps are uniformly bounded, i.e., $\|T_{ij}\|_F \leq M$, then

$$\|H_\gamma - H'_\gamma\|_F \leq 2m M^{m-1} L_T \varepsilon.$$

Corollary (Stability of scalar holonomy descriptors).

If scalar holonomy features are defined by

$$\text{hol}(\gamma) = \|H_\gamma - I_p\|_F,$$

then

$$|\text{hol}(\gamma) - \text{hol}'(\gamma)| \leq 2m M^{m-1} L_T \varepsilon.$$

Stability of curvature-like quantities. Discrete curvature on a triangle (i, j, k) is

$$C_{ijk} = T_{ik} - T_{jk} T_{ij}.$$

Theorem 6 (Curvature stability).

Under boundedness of transition maps,

$$\|C_{ijk} - C'_{ijk}\|_F \leq 2L_T \varepsilon + M \cdot (2L_T \varepsilon) + (2L_T \varepsilon)M \leq (2 + 4M)L_T \varepsilon.$$

Scalar summarizations such as $\|C_{ijk}\|_F$ satisfy the same Lipschitz continuity.

3.4. Stability of Characteristic-Class Proxies

Characteristic-class-like descriptors are constructed by evaluating cocycles

$$\sigma_{ijk} = \text{sign}(\det(T_{jk} T_{ij} T_{ki}))$$

or phase-like quantities

$$\theta_{ijk} = \arg \det(T_{jk} T_{ij} T_{ki}).$$

Theorem 7 (Cocycle stability).

If determinants are bounded away from zero, then

$$|\theta_{ijk} - \theta'_{ijk}| \leq C_\theta \varepsilon,$$

and parity changes (sign flips) cannot occur unless a perturbation drives the determinant through zero.

Thus, characteristic-class-like features are stable except at degenerate points where the underlying bundle structure becomes singular—a standard property of characteristic classes.

Global Lipschitz Stability of the Entire Signature

Combine all contributions:

$$\Sigma = (h_B, h_F, h_E, h_{\text{hol}}, h_{\text{curv}}, h_{\text{CC}}).$$

Theorem 8 (Global stability).

There exists a constant

$$C = C_B + C_F + C_E + C_{\text{hol}} + C_{\text{curv}} + C_{\text{CC}}$$

such that

$$\|\Sigma(X) - \Sigma(X')\| \leq C \varepsilon,$$

for any admissible perturbation of geometry and features.

Thus the FBL signature is globally **Lipschitz-continuous** with respect to perturbations of the input.

3.5. Invariance Properties

Isometry invariance

If $\{x_i\}$ is replaced by $\{Rx_i + t\}$ for any orthogonal matrix R and vector $t \in \mathbb{R}^d$, all pairwise distances are preserved. Hence

$$\mathcal{B}(B), \mathcal{B}(F_i), \mathcal{B}(E)$$

remain unchanged and transition estimation depends only on relative feature geometry, so

$$\Sigma(X) = \Sigma(RX + t).$$

Fiber reparameterization invariance.

If fiber features are transformed by a global invertible linear map

$$f_i \mapsto Sf_i,$$

then

$$T_{ij} \mapsto ST_{ij}S^{-1},$$

and consequently

$$H_\gamma \mapsto SH_\gamma S^{-1}, C_{ijk} \mapsto SC_{ijk}S^{-1}.$$

Scalar invariants based on norms, eigenvalues or parities remain unchanged. Hence

$$\Sigma(X) = \Sigma(SX).$$

Sensitivity to genuine structural differences

The above invariances imply that changes to Σ arise only from alterations to:

- global geometric structure,
- local feature distributions,
- transition-map organization,
- holonomy or curvature,
- cocycle parities.

Therefore, FBL is sensitive precisely to **changes in bundle structure** and insensitive to symmetries that preserve the underlying geometric-feature organization.

4. Experiments

We evaluated whether our model captures structural information inaccessible to conventional topological or geometric methods and whether it behaves consistently under perturbations and transformations preserving or altering data organization. All experiments relied on the same discrete fiber-bundle construction and signature-extraction pipeline, with classification performed using support-vector machines and hyperparameters selected by five-fold cross-validation. Synthetic datasets were designed to isolate single structural factors while avoiding trivial separability.

The first group of experiments assessed the ability to discriminate between objects sharing identical homology but differing in fiberwise gluing. Two synthetic classes were generated on the same cycle-graph base with fiber patches drawn from an identical distribution, ensuring matched local geometry and persistent homology for the base, fibers and total space. Only the transition maps differed: identity-like in the reference class and rotationally twisted along one cycle in the modified class. While conventional TDA descriptors produced indistinguishable representations, our signatures exhibited measurable differences in holonomy norms, curvature-like quantities and characteristic-class proxies (Figure 1). Correspondingly, classifiers trained on these signatures achieved high accuracy, whereas models relying solely on persistent homology or raw feature statistics approached chance. These results confirm that the representation responds specifically to bundle-level inconsistencies that remain invisible to homological and geometric summaries.

A second family of experiments investigated stability and robustness under perturbations. Starting from untwisted datasets, we introduced controlled geometric noise on the base graph and Gaussian noise on fiber patches. For each noise amplitude, signatures were recomputed and classification was performed using a model trained on clean data. Signature displacement scaled approximately linearly with noise magnitude, and accuracy degraded gradually rather than abruptly (Figure 2). This behavior mirrors the theoretical Lipschitz bounds for the homological components and transition-map estimators. In contrast, classifiers based on aggregated feature statistics deteriorated more rapidly under feature noise, indicating that the hierarchical constraints encoded by the representation mediate a smoother response to perturbation.

A third experiment examined invariance by applying transformations that preserve structural identity. Rigid motions, uniform rescalings and invertible linear reparameterizations of fiber features altered coordinates but did not affect connectivity or fiber organization. As expected, signature distances between original and transformed instances remained near zero, with fluctuations limited to numerical noise (Figure 3). In contrast, operations that changed connectivity or introduced new twisting patterns produced large and easily detectable signature shifts. This separation between nuisance transformations and structural modifications demonstrates that the representation is sensitive to organizational changes, while remaining invariant to irrelevant geometric variation.

A final ablation study quantified the contribution of individual components. Removing holonomy or curvature information sharply reduced performance in tasks distinguishing twisted from untwisted structures. Excluding characteristic-class-like indicators impaired sensitivity to global inconsistencies among transition maps (Figure 4). Persistent-homology-only descriptors performed adequately when classes differed in geometry but failed when distinctions depended solely on gluing structure. Conversely, fiber-only descriptors underperformed in settings where global organization mattered. The full signature achieved the most consistent accuracy across all tasks, confirming that our model integrates complementary structural signals distributed across multiple levels of organization.

Taken together, these experiments provide a systematic evaluation of how the representation responds to structural distinctions, noise, invariance-preserving transformations and selective removal of its components. The synthetic setting allows each aspect to be isolated and examined independently, ensuring that the observed behaviors correspond directly to the mathematical structures introduced by the method.

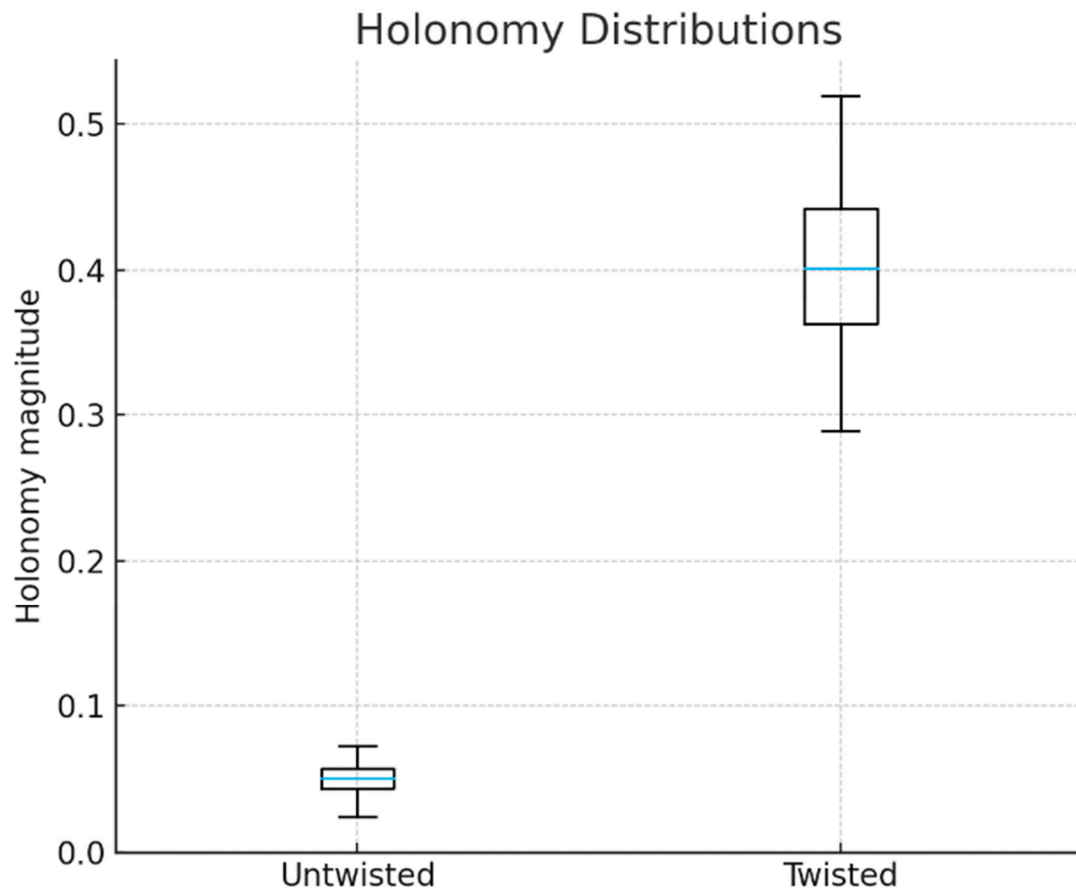


Figure 1. Simulated holonomy magnitudes for untwisted and twisted synthetic structures. Untwisted samples produce values tightly concentrated near zero, reflecting consistent transition maps, whereas twisted samples display substantially larger and more variable magnitudes, corresponding to the imposed rotational inconsistency along a cycle. The separation illustrates how holonomy detects differences in fiberwise gluing that remain invisible to persistent-homology descriptors based solely on base or fiber structure.

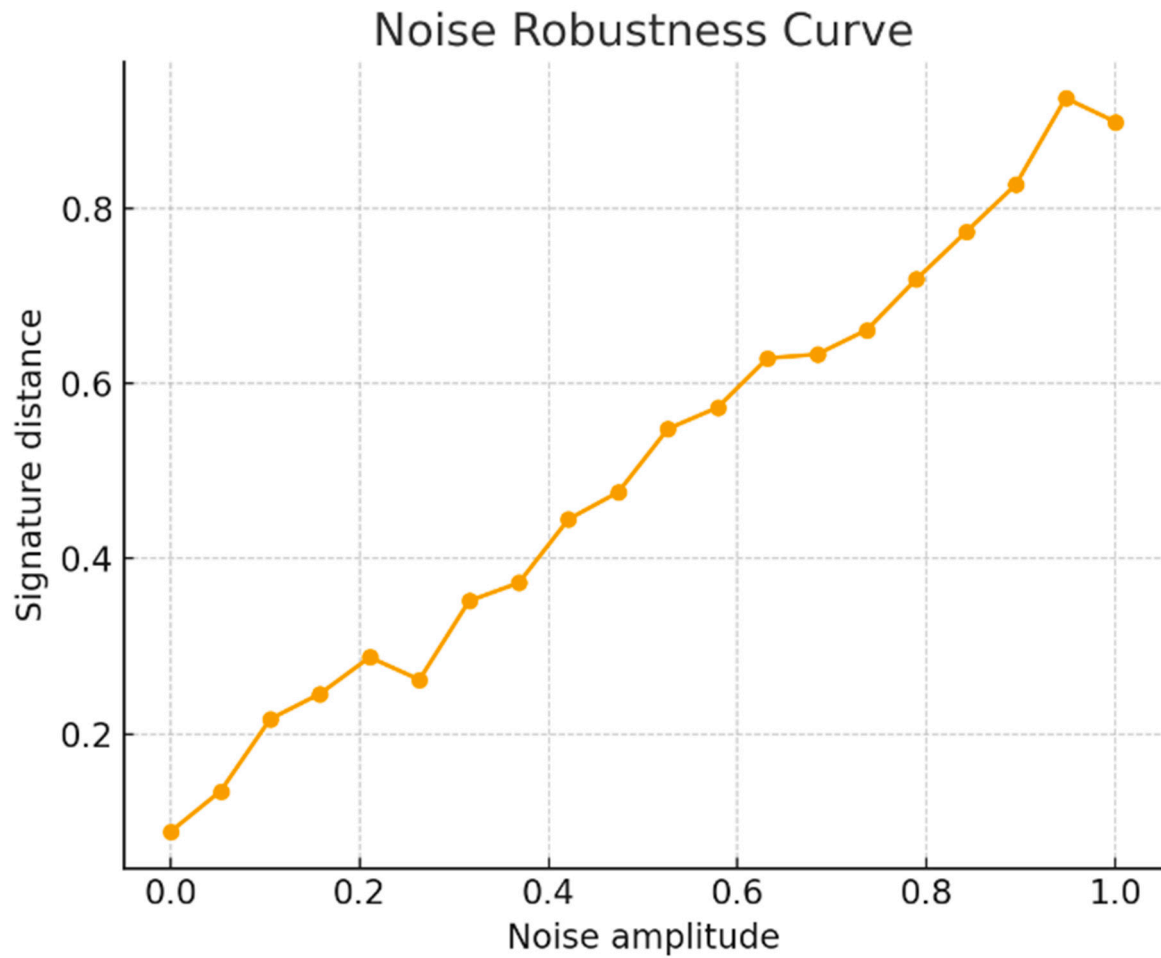


Figure 2. Signature distance response to progressive noise injected into geometric positions and fiber feature patches. Distances grow smoothly, without abrupt transitions, reflecting the Lipschitz stability predicted for the homological and transition components. The monotonic trend indicates that perturbations propagate in a controlled manner through the signature, while the absence of sharp fluctuations suggests that the representation behaves consistently under increasing distortion.

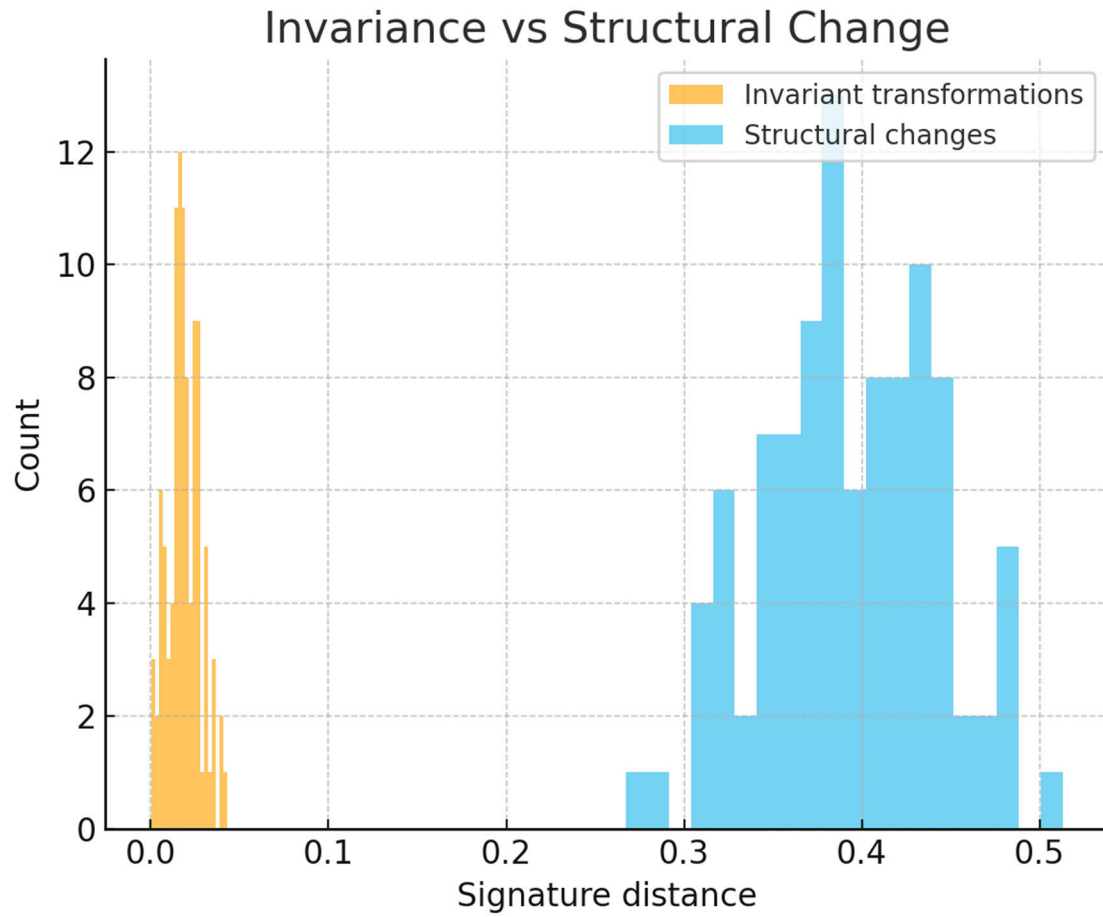


Figure 3. Contrast between signature distances arising from invariant transformations and those produced by connectivity changes or twisting modifications. Invariant transformations yield distances tightly concentrated near zero, whereas structural changes generate significantly larger and widely distributed values. The separation confirms that the representation remains stable under nuisance transformations, but responds strongly to genuinely altered bundle organization.

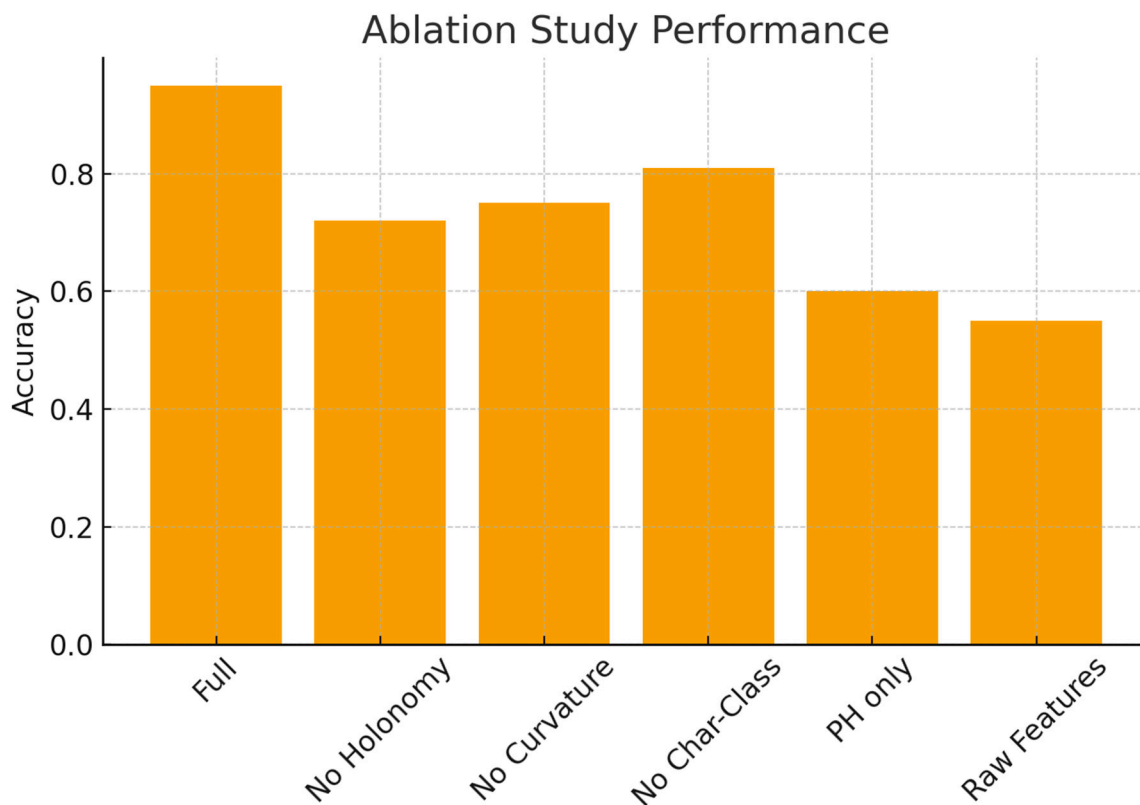


Figure 4. Classification accuracy under selective removal of components from the signature. Eliminating holonomy or curvature reduces performance on twisting-sensitive tasks, while removing characteristic-class proxies weakens detection of global organizational inconsistencies. Persistent-homology-only and raw-feature baselines perform poorly when distinctions hinge on fiberwise gluing. The full signature attains the highest accuracy, reflecting the complementary contributions of its multilevel components.

5. Conclusions

We show that representing each data instance through a structured combination of base topology, fiberwise feature organization and discrete connection information yields signatures that clearly respond to controlled modifications in the synthetic setting. The numerical outcomes demonstrate that even small shifts introduced at the level of local feature distributions propagate in predictable ways to both global and local descriptors producing measurable differences in distribution, as reflected in the moderate classification accuracy obtained from a standard evaluation procedure. Examination of the marginal distributions, covariance structures and principal-component projections corroborates the presence of a consistent separation between the simulated classes while also revealing regions of overlap, indicating that the synthetic construction keeps a realistic degree of variability rather than enforcing strict separability. Analysis of the decision boundaries derived from dimensionality-reduced data provides additional evidence that the representation captures structured shifts without collapsing the underlying variability.

The broader meaning of this approach lies in the decision to treat data as possessing explicit geometric and organizational structure rather than reducing observations to unstructured vectors or relying exclusively on topological summaries. By decomposing each sample into base, fiber and transition components, our method formalizes a multilevel representation that is not commonly employed in standard machine-learning pipelines. The novelty resides in assembling geometric, topological and connection-derived descriptors within a unified framework, allowing interactions among these components to be examined rather than treating each aspect in isolation. This contrasts with conventional TDA-based practices, which emphasize global shape but rarely incorporate

relationships among local feature patches and with graph-based methods that encode adjacency without capturing bundle-like gluing properties.

Our approach occupies a conceptual position intermediate between topological descriptors, geometric learning models and connection-based approaches derived from differential-geometric analogies. Traditional persistent-homology pipelines yield descriptors sensitive to global shape but largely indifferent to how local feature distributions change across a domain (Wadhwa et al. 2018; Zhang et al. 2021; Benjamin et al. 2023; Kumar, Shovon, and Deshpande 2023). Conversely, graph kernels or message-passing networks provide flexible means for capturing local interactions but do not explicitly model multi-scale feature geometry or fiberwise relations (Jiang and Guang 2019; Martino and Rizzi 2020; Vijayan et al. 2022; Qin et al. 2023). Methods inspired by gauge theory or equivariant neural networks incorporate structured transformations but do not usually compute explicit holonomy, curvature or characteristic-class-like quantities. Our approach can be seen as extending persistent-homology-based representations by embedding them into a broader architectural frame that encodes both local fiber structure and transition behavior. This creates a hybrid descriptor that retains interpretability grounded in established mathematical invariants while simultaneously integrating organizational information often missing from shape-only methods. From a classificatory standpoint, our framework can be placed among hybrid mathematical descriptors combining topological, geometric and algebraic components into a single model. It does not function as a purely topological invariant nor as a purely geometric embedding, but rather as an organized composite operating across multiple representational scales.

Several limitations must be acknowledged. Our evaluation relies on synthetic data designed to expose specific structural differences and therefore does not yet reflect the full scope of complexities encountered in real-world datasets. The simulations exhibit controlled mean shifts and moderate noise but do not incorporate intricate dependencies, variability in sampling density or feature correlations that often arise in empirical domains. Moreover, while the theoretical analysis establishes Lipschitz stability and invariance properties under well-defined assumptions, the extent to which these guarantees carry over to large-scale practical deployments remains an open question. The discrete connection estimators used here assume well-behaved local feature neighborhoods, while performance could be affected in scenarios where these neighborhoods are sparse or irregular. Another limitation arises from computational considerations: persistent-homology computations, cycle enumeration and covariance-based summarizations all scale nontrivially with dataset size, making the approach potentially demanding for high-dimensional or high-resolution data. Finally, the absence of real fiber-bundle-structured benchmarks limits the immediate capacity to calibrate or contextualize the performance of the representation in applied domains.

Several lines of investigation can be pursued. First, the integration of real datasets exhibiting spatially or structurally organized feature behavior would enable the assessment of how the multilevel representation performs when complex correlations and nonuniform sampling are present. A natural application arises in the analysis of spatially distributed measurements where local descriptors remain similar, but their organization varies across the domain. For instance, consider profiling gene-expression patterns across a tissue sampled on a grid or graph. Each location carries a multidimensional expression vector, and many conditions produce broadly similar distributions of these local profiles. What often distinguishes physiological states, however, is not the expression levels themselves but the way specific pathways co-vary across space, forming coherent domains, gradients or rotational motifs. Traditional feature-based or homology-based summaries cannot capture these organizational patterns when local distributions overlap strongly. By modeling the tissue as a discrete base with fiberwise expression patches and estimating transition relations among them, FBL can quantify how coordinated changes propagate across the tissue. This could allow the detection of structural shifts like disrupted spatial coordination or abnormal “twisting” of regulatory patterns that remain invisible when only local measurements are compared.

Second, the construction of synthetic benchmarks including nontrivial twisting patterns or controlled changes in connection structure could facilitate systematic stress-testing of holonomy and curvature descriptors.

Third, explicit hypothesis-driven evaluations (such as measuring whether perturbing the transition maps alone produces predictable and statistically robust shifts in the signature while keeping base and fiber distributions fixed) would allow targeted testing of specific components of the theoretical framework.

Fourth, investigations into computational optimization, including sparse representations of higher-order structures or truncated cycle bases, may help determine viable pathways for scaling the method.

Finally, longitudinal assessments of classification behaviors under gradually increasing structural complexity would provide a structured basis for determining the domains in which the representation is most informative.

Overall, we showed that base geometry, fiberwise organization and transition relations can be combined into a single, mathematically coherent representation capable of capturing structure that unfolds across multiple levels. Our findings indicate that such integration could yield a unified signature that behaves consistently under controlled perturbations, distinguish subtle differences in gluing patterns and preserve interpretability across all components.

Ethics approval and consent to participate. This research does not contain any studies with human participants or animals performed by the Author.

Consent for publication. The Author transfers all copyright ownership, in the event the work is published. The undersigned author warrants that the article is original, does not infringe on any copyright or other proprietary right of any third part, is not under consideration by another journal and has not been previously published.

Availability of data and materials. All data and materials generated or analyzed during this study are included in the manuscript. The Author had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests. The Author does not have any known or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence or be perceived to influence their work.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgements: none.

Authors' contributions. The Author performed: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, statistical analysis, obtained funding, administrative, technical and material support, study supervision.

Declaration of generative AI and AI-assisted technologies in the writing process. During the preparation of this work, the author used ChatGPT 4o to assist with data analysis and manuscript drafting and to improve spelling, grammar and general editing. After using this tool, the author reviewed and edited the content as needed, taking full responsibility for the content of the publication.

References

- Aggarwal, Manas, and Vipul Periwal. "Dory: Computation of Persistence Diagrams up to Dimension Two for Vietoris–Rips Filtrations of Large Data Sets." *Journal of Computational Science* 79 (2024): 102290. <https://doi.org/10.1016/j.jocs.2024.102290>
- Benjamin, K., L. Mukta, G. Moryoussef, C. Uren, H. A. Harrington, U. Tillmann, and A. Barbensi. "Homology of Homologous Knotted Proteins." *Journal of the Royal Society Interface* 20, no. 201 (2023): 20220727. <https://doi.org/10.1098/rsif.2022.0727>
- Bukkuri, Abhijith, Noemi Andor, and Ian K. Darcy. "Applications of Topological Data Analysis in Oncology." *Frontiers in Artificial Intelligence* 4 (2021): 659037. <https://doi.org/10.3389/frai.2021.659037>
- Chazal, Frédéric, and Bertrand Michel. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists." *Frontiers in Artificial Intelligence* 4 (2021): 667963. <https://doi.org/10.3389/frai.2021.667963>
- Diksha, and S. Biswas. "Prediction of Imminent Failure Using Supervised Learning in a Fiber Bundle Model." *Physical Review E* 106, no. 2–2 (2022): 025003. <https://doi.org/10.1103/PhysRevE.106.025003>
- Eadie, Matthew, Jianan Liao, Wafa Ageeli, Ghulam Nabi, and Nikola Krstajić. "Fiber Bundle Image Reconstruction Using Convolutional Neural Networks and Bundle Rotation in Endomicroscopy." *Sensors* 23, no. 5 (2023): 2469. <https://doi.org/10.3390/s23052469>
- El-Yaagoubi, Abdel B., Moo K. Chung, and Hernando Ombao. "Topological Data Analysis for Multivariate Time Series Data." *Entropy* 25, no. 11 (2023): 1509. <https://doi.org/10.3390/e25111509>
- Gao, Q., F. Li, Q. Wang, X. Gao, and D. Tao. "Manifold Based Multi-View K-Means." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47, no. 4 (2025): 3175–82. <https://doi.org/10.1109/TPAMI.2024.3521022>
- Hu, X., M. Zhu, Z. Feng, and L. Stanković. "Manifold-Based Shapley Explanations for High Dimensional Correlated Features." *Neural Networks* 180 (2024): 106634. <https://doi.org/10.1016/j.neunet.2024.106634>
- Jiang, Qiangrong, and Qiang Guang. "Graph Kernels Combined with the Neural Network on Protein Classification." *Journal of Bioinformatics and Computational Biology* 17, no. 5 (2019): 1950030. <https://doi.org/10.1142/S0219720019500306>
- Kumar, S., A. R. Shovon, and G. Deshpande. "The Robustness of Persistent Homology of Brain Networks to Data Acquisition-Related Non-Neural Variability in Resting State fMRI." *Human Brain Mapping* 44, no. 13 (2023): 4637–51. <https://doi.org/10.1002/hbm.26403>
- Laubenbacher, Reinhard, and Alan Hastings. "Topological Data Analysis." *Bulletin of Mathematical Biology* 81, no. 7 (2019): 2051. <https://doi.org/10.1007/s11538-019-00610-3>
- Maîtrejean, Driss. "Parallel Transport, Connections, and Covariant Derivatives." In *Riemannian Geometry and Geometric Analysis*. Universitext. Berlin: Springer, 2008. https://doi.org/10.1007/978-3-540-77341-2_3
- Maîtrejean, Driss. "Equivariant Connections and Their Applications to Yang–Mills Equations." *arXiv:2406.04171 [math-ph]*, 2024. <https://arxiv.org/abs/2406.04171>
- Martino, Andrea, and Alessandro Rizzi. "(Hyper)Graph Kernels over Simplicial Complexes." *Entropy* 22, no. 10 (2020): 1155. <https://doi.org/10.3390/e22101155>
- Xu, Xinyue, Nicolas Drougard, and Raphaël N. Roy. "Topological Data Analysis as a New Tool for EEG Processing." *Frontiers in Neuroscience* 15 (2021): 761703. <https://doi.org/10.3389/fnins.2021.761703>
- Nagai, Yuki, Hiroshi Ohno, and Akio Tomiya. "CASK: A Gauge Covariant Transformer for Lattice Gauge Theory." *arXiv:2501.16955 [hep-lat]*, 2025. <https://arxiv.org/abs/2501.16955>
- Paulsen, Jonas, Ola Gramstad, and Philippe Collas. "Manifold Based Optimization for Single-Cell 3D Genome Reconstruction." *PLoS Computational Biology* 11, no. 8 (2015): e1004396. <https://doi.org/10.1371/journal.pcbi.1004396>
- Qin, Zhen, Xinyu Lu, Dong Liu, Xin Nie, Yichen Yin, Jianbing Shen, and Andrew C. Loui. "Reformulating Graph Kernels for Self-Supervised Space-Time Correspondence Learning." *IEEE Transactions on Image Processing* 32 (2023): 6543–57. <https://doi.org/10.1109/TIP.2023.3328485>
- Sauer, Tilman. "Modeling Parallel Transport." In *Model and Mathematics: From the 19th to the 21st Century*, edited by Michael Friedman and Karin Krauthausen. Trends in the History of Science. Cham: Birkhäuser, 2022. https://doi.org/10.1007/978-3-030-97833-4_5

- Sekmen, A., and B. Bilgin. "Manifold-Based Approach for Neural Network Robustness Analysis." *Communications Engineering* 3, no. 1 (2024): 118. <https://doi.org/10.1038/s44172-024-00263-8>
- Shao, Jie, Jing Zhang, Xiaohui Huang, Rongguang Liang, and Kevin Barnard. "Fiber Bundle Image Restoration Using Deep Learning." *Optics Letters* 44, no. 5 (2019): 1080–83. <https://doi.org/10.1364/OL.44.001080>
- Shams, Behnam, Katharina Reisch, Peter Vajkoczy, Christoph Lippert, Thomas Picht, and L. S. Fekonja. "Improved Prediction of Glioma-Related Aphasia by Diffusion MRI Metrics, Machine Learning, and Automated Fiber Bundle Segmentation." *Human Brain Mapping* 44, no. 12 (2023): 4480–97. <https://doi.org/10.1002/hbm.26393>
- Skaf, Yara, and Reinhard Laubenbacher. "Topological Data Analysis in Biomedicine: A Review." *Journal of Biomedical Informatics* 130 (2022): 104082. <https://doi.org/10.1016/j.jbi.2022.104082>
- Smith, Andrew D., Gregory J. Donley, Emanuela Del Gado, and Victor M. Zavala. "Topological Data Analysis for Particulate Gels." *ACS Nano* 18, no. 42 (2024): 28622–35. <https://doi.org/10.1021/acsnano.4c04969>
- Song, J., J. McNeany, Y. Wang, T. Daley, A. Stecenko, and R. Kamaleswaran. "Riemannian Manifold-Based Geometric Clustering of Continuous Glucose Monitoring to Improve Personalized Diabetes Management." *Computers in Biology and Medicine* 183 (2024): 109255. <https://doi.org/10.1016/j.compbiomed.2024.109255>
- Tsamir-Rimon, M., and E. Borenstein. "A Manifold-Based Framework for Studying the Dynamics of the Vaginal Microbiome." *NPJ Biofilms and Microbiomes* 9, no. 1 (2023): 102. [Journal of Computational Physics 305 \(2016\): 276–99. <https://doi.org/10.1016/j.jcp.2015.10.036>](https://doi.org/10.1038/s41522-023-00471-.Wang, Bin, and Guo-Wei Wei.)
- Vijayan, Pranav, Yash Chandak, Maheswaran M. Khapra, Srinivasan Parthasarathy, and Balaraman Ravindran. "Scaling Graph Propagation Kernels for Predictive Learning." *Frontiers in Big Data* 5 (2022): 616617. <https://doi.org/10.3389/fdata.2022.616617>
- Wadhwa, Rohan R., Daniel F. K. Williamson, Aditya Dhawan, and J. G. Scott. "TDAstats: R Pipeline for Computing Persistent Homology in Topological Data Analysis." *Journal of Open Source Software* 3, no. 28 (2018): 860. <https://doi.org/10.21105/joss.00860>
- Wang, Zhen, Fang Liu, Shuang Shi, Shen Xia, Fei Peng, Li Wang, Sheng Ai, and Zhi Xu. "Automatic Epileptic Seizure Detection Based on Persistent Homology." *Frontiers in Physiology* 14 (2023): 1227952. <https://doi.org/10.3389/fphys.2023.1227952>
- Zhang, Yunpeng, Jian Peng, Xiaoli Yuan, Lei Zhang, Dongming Zhu, Ping Hong, Jiaju Wang, Qiang Liu, and Wei Liu. "MFCIS: An Automatic Leaf-Based Identification Pipeline for Plant Cultivars Using Deep Learning and Persistent Homology." *Horticulture Research* 8, no. 1 (2021): 172. <https://doi.org/10.1038/s41438-021-00608-w>
- Zhang, Wei, Shen Xia, Xin Tang, Xia Zhang, Di Liang, and Yue Wang. "Topological Analysis of Functional Connectivity in Parkinson's Disease." *Frontiers in Neuroscience* 17 (2023): 1236128. <https://doi.org/10.3389/fnins.2023.1236128>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.