HOMOLOGY FEATURES FOR X-RAY HYPERSPECTRAL IMAGES SEGMENTATION

Fethi Harkat[*] [†], *Guillaume Gey*[*], *Valérie Perrier*[†], *Kévin Polisano*[†],

[*]Detection Technology Plc, A Grid, Otakaari 5A, 02150 Espoo, Finland
[†] Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

ABSTRACT

Characterizing Multi-Energy X-ray Transmission (XRT) hyperspectral images presents significant challenges for traditional classification frameworks due to high-dimensional redundancy and non-linear spectral overlaps. Although manifold learning techniques such as UMAP enhance spectral representation by preserving non-linear correlations, they do not explicitly utilize the geometric structure of the embedded manifold for segmentation. This study introduces a hybrid architecture that integrates topological features derived from persistent homology with convolutional and graph neural networks. The approach computes Vietoris-Rips persistent homology on UMAP-projected superpixel point clouds to capture H_1 cyclic structures that encode spectral heterogeneity. These topological descriptors are combined with pixel-level CNN features through a broadcasting mechanism that propagates superpixel-level graph representations to full spatial resolution. Experimental results on XRT multi-energy imagery demonstrate that the proposed hybrid model significantly outperforms both CNN-only and superpixel-pooling fusion baselines, suggesting that topological features provide complementary discriminative information to local spectral-spatial descriptors.

***Index Terms*—** Hyperspectral imaging, X-ray detection, Topology, Manifold learning, Dimensionality reduction, Deep learning, Feature extraction, Computer Vision.

1. INTRODUCTION

Previous research on manifold learning for Hyperspectral Images (HSI) demonstrated that linear projection techniques, such as Principal Component Analysis (PCA), often fail to represent the complex intrinsic geometry of XRT datasets effectively [1, 2]. By applying the Uniform Manifold Approximation and Projection (UMAP) algorithm [3, 4], it was established in [5] and in our previous work [6] that the underlying data manifold exhibits non-linear correlations that can be characterized through topological approximation. These results indicate that the topology of the data contains discriminative information that is not captured by standard spectral-spatial feature extractors.

Building upon this foundation, we propose a hybrid architecture that integrates tools from Topological Data Analysis (TDA), specifically persistent homology [7, 8], with Graph Neural Networks (GNNs) [9, 10] and Convolutional Neural Networks (CNNs). The high dimensionality and spectral redundancy of HSI data have motivated a shift from pixel-wise classification to region-based reasoning through graph representations [11]. In this framework, the image is abstracted as a graph where nodes correspond to image regions and edges encode spatial or spectral adjacency. Early approaches constructed graphs at the pixel level, treating each spectral vector as a node and connecting k-nearest neighbors in spectral space [12]. While preserving fine-grained spatial resolution, this approach suffers from computational intractability and susceptibility to spectral noise, as the graph connectivity becomes unreliable in high-dimensional, noisy spectral spaces [12]. Graph Convolutional Networks (GCNs) [13] operate on this structure through message-passing mechanisms that propagate information across the graph, enabling the integration of local spectral features with global spatial context.

Computing persistent homology requires a sufficient number of data points to form meaningful topological structures. Operating at the pixel scale is impractical because individual pixels lack adequate spectral information to construct stable persistence diagrams. We therefore adopt a superpixel-based strategy. Following our previous methodology in [6], we first perform a UMAP projection on the image before grouping pixels into perceptually coherent superpixels using Simple Linear Iterative Clustering (SLIC) [14]. Each superpixel's spectral point cloud constitutes an independent metric space on which Vietoris-Rips persistent homology is computed via Ripser [15]. The resulting persistence statistics, combined with spatial descriptors, form node features for a GCN operating on the superpixel adjacency graph. A broadcasting mechanism propagates superpixel-level graph features back to individual pixels, where they are concatenated with pixel-level CNN features for segmentation.

Prior work has integrated topological summaries into machine learning pipelines through persistence images [16], landscapes [17], kernel embeddings [18], and direct graph

AUTHOR EMAIL: FETHI.HARKAT@UNIV-GRENOBLE-ALPES.FR

architectures [19, 20]. However, the application of localized persistent homology to superpixel-based representations of hyperspectral data remains unexplored. Furthermore, existing hybrid CNN-GNN frameworks treat the two branches of the two neural networks as independent feature extractors whose outputs are combined only at the final layer, missing opportunities for intermediate information exchange [11, 12]. None of these approaches explicitly incorporate topological features into the GNN branch.

The key contributions of this paper are threefold. First, we demonstrate that persistent homology features computed on UMAP-projected superpixel point clouds provide complementary discriminative signal to standard spectral-spatial descriptors. Second, through systematic ablation, we identify a minimal subset of H_1 persistence statistics that optimizes the trade-off between feature dimensionality and segmentation performance. Third, we propose a broadcasting fusion strategy that preserves pixel-level spatial resolution and outperforms superpixel-pooling alternatives by a substantial margin. Experiments on XRT multi-energy imagery validate the proposed framework.

The remainder of this paper is organized as follows. Section 2 details the proposed methodology, including UMAP projection, superpixel segmentation, topological feature extraction, and the hybrid CNN-GNN architecture. Section 3 describes the experimental setup and datasets, then presents quantitative and qualitative results. Section 4 concludes with directions for future work.

2. METHODOLOGY

2.1. Overview

The proposed framework, illustrated in Fig. 1, consists of four stages. First, raw XRT hyperspectral cubes are projected onto a low-dimensional manifold using UMAP [3], fitted on a sample of training pixels, following the approach in [6]. Second, the projected image is segmented into superpixels. Third, topological and spatial features are extracted from each superpixel's spectral point cloud. Fourth, a dual-branch hybrid architecture fuses pixel-level CNN features with superpixel-level GNN features via a broadcasting mechanism, followed by a final classification layer. The first three stages are preprocessing steps; only the two branches and the classifier are trained. We detail below these different steps.

2.2. Superpixel Segmentation and Graph Construction

Raw ME-XRT images are acquired as three-dimensional cubes of shape $[H, W, C]$, where H and W denotes spatial dimensions, the detector pixel axis, and C the energy band axis [6]. UMAP [3], fitted on a sample of training pixels, projects each pixel's C-dimensional spectral signature onto a D-dimensional manifold ($D \ll C$; here $C = 16$ and $D = 5$),

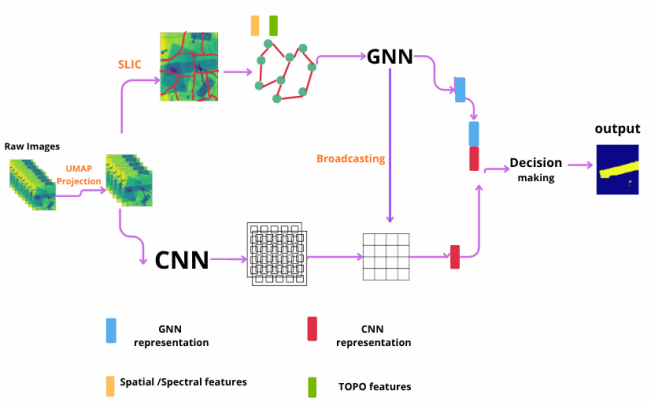


Fig. 1: Pipeline of the proposed hybrid CNN+GNN method. The CNN branch (bottom) processes the UMAP-projected image at full pixel resolution, while the GNN branch (top) operates on superpixel-level topological and spatial features. A broadcasting mechanism fuses both representations before the final classification layer.

yielding projected images of shape $[H, W, D]$ that preserve non-linear spectral correlations.

The UMAP-projected image is then segmented into N perceptually coherent superpixels using SLIC [14], operating on the D manifold channels. The resulting superpixels form the nodes of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where edges connect spatially adjacent regions. Thus, the graph comprises N nodes instead of the original $H \times W$ pixels, reducing the spatial complexity from $\mathcal{O}(HW)$ to $\mathcal{O}(N)$ while preserving spatial coherence. The spectral dimension is embedded in the D-dimensional UMAP coordinates and subsequently encoded in the topological descriptors extracted per superpixel.

2.3. Node Feature Extraction

For each superpixel s_i, let $\mathcal{P}_i \subset \mathbb{R}^D$ denote its UMAP-projected spectral point cloud. Vietoris–Rips persistent homology is computed via Ripser [15], yielding persistence diagrams $\mathcal{D}_0(\mathcal{P}_i)$ and $\mathcal{D}_1(\mathcal{P}_i)$. Diagram $\mathcal{D}_0$ captures connected components (0-dimensional features), while $\mathcal{D}_1$ captures 1-dimensional loops. From these diagrams we extract a pool of 30 candidate topological descriptors.

Topological descriptors (30 candidates). For each birth–death pair (b_j, d_j) in a persistence diagram, the persistence is $\ell_j = d_j - b_j$. Let $n_k = |\mathcal{D}_k(\mathcal{P}_i)|$ $(k = 0, 1)$ denote

the number of points in $\mathcal{D}_k$. The H_1-statistics are defined as:

$$\text{MaxPers} = \max_j \ell_j, \qquad \text{MinPers} = \min_j \ell_j,$$
$$\text{MeanPers} = \tfrac{1}{n_1}\sum_j \ell_j, \qquad \text{StdPers} = \sqrt{\tfrac{1}{n_1}\sum_j (\ell_j - \bar{\ell})^2},$$
$$\text{SumPers} = \sum_j \ell_j, \qquad \text{MeanBirth} = \tfrac{1}{n_1}\sum_j b_j,$$
$$\text{MeanDeath} = \tfrac{1}{n_1}\sum_j d_j, \qquad \text{MaxBirth} = \max_j b_j,$$
$$\text{MaxDeath} = \max_j d_j, \qquad \text{Entropy} = -\sum_j p_j \log p_j,$$

where $p_j = \ell_j / \sum_k \ell_k$, and $\text{PointsCount} = n_1$ is the 11th feature. The same 11 statistics are computed on $\mathcal{D}_0$ (denoted H_0_*). Additionally, Betti numbers β_0 and β_1 are recorded at three fixed filtration scales $\epsilon_1 < \epsilon_2 < \epsilon_3$ (6 features), and two composite scores (Total_Complexity, Topology_Score) combine multi-scale topological indicators, yielding 30 candidates in total.

Spectral descriptors (25 dimensions). For each UMAP channel $d \in \{1, \ldots, D\}$ (here $D = 5$), the mean, standard deviation, and quartiles (Q1, Q2, Q3) of pixel intensities within the superpixel are computed, giving $5 \times 5 = 25$ dimensions.

Geometric descriptors (5 dimensions). A superpixel s_i is a connected set of pixels; let $\mathcal{R}_i = \{(h, w) \in [H] \times [W] \mid \text{pixel } (h, w) \text{ belongs to } s_i\}$ denote the collection of pixel coordinates belonging to superpixel s_i. From this set we extract area $|\mathcal{R}_i|$ (the pixel count), perimeter, and:

$$\text{Eccentricity} = \sqrt{1 - \frac{\lambda_{\min}}{\lambda_{\max}}},$$
$$\text{Solidity} = \frac{|\mathcal{R}_i|}{|\operatorname{Conv}(\mathcal{R}_i)|},$$
$$\text{Orientation} = \frac{1}{2}\arctan\left(\frac{2\Sigma_{12}}{\Sigma_{11} - \Sigma_{22}}\right),$$

where $\lambda_{\max}, \lambda_{\min}$ are the eigenvalues of the coordinate covariance matrix $\mathbf{\Sigma}$ of the points in $\mathcal{R}_i$, and $\operatorname{Conv}(\mathcal{R}_i)$ denotes the convex hull of $\mathcal{R}_i$ (the smallest convex polygon containing all pixels of the superpixel). Thus $|\operatorname{Conv}(\mathcal{R}_i)|$ is the area of that convex hull, and the ratio $|\mathcal{R}_i|/|\operatorname{Conv}(\mathcal{R}_i)|$ measures boundary regularity (1 for a convex shape, < 1 for irregular boundaries).

Feature vector. Each superpixel node is initially described by a 60-dimensional vector (30 topological + 25 spectral + 5 geometric). The procedure that selects the final 11 H_1 topological features retained for the GNN is detailed in the ablation study (Section 3.2).

2.4. Hybrid CNN-GNN Architecture

As shown in Fig. 1, the architecture comprises two parallel branches whose outputs are fused before the final classification layer.

CNN Branch. The UMAP-projected image is processed by a fully convolutional network with no pooling, preserving spatial resolution. The architecture comprises three stages: a 3×3 convolution expanding the $D = 5$ UMAP channels (where D is the output dimension of the UMAP reduction from $C = 16$ spectral bands) to 32 feature maps; a 3×3 convolution with a dilation factor of 2, achieving a 5×5 effective receptive field at 64 channels; and a 1×1 convolution projecting to 64 output channels. Each convolution is followed by batch normalization, and by a ReLU activation for the first two.

GNN Branch. The superpixel graph is constructed with 41-dimensional node features: 11 topological descriptors derived from H_1 persistence statistics (see Section 3.2; maximum, mean, standard deviation, sum, and minimum persistence; mean and maximum birth and death scales; entropy; and point count) combined with 30 spatial descriptors (per-channel spectral statistics and geometric shape features). These node features are processed by a 3-layer GCN [13] ($41 \to 64 \to 64 \to 2$), with batch normalization, ReLU activation and dropout ($p = 0.3$) after the first two layers, producing a 2-dimensional output per superpixel that encodes the topological and structural signature of each region. The source code is publicly available[1].

2.5. Fusion

The CNN branch outputs pixel-level features $\mathbf{F}_{\text{cnn}} \in \mathbb{R}^{H \times W \times 64}$. To combine these with the GNN's superpixel-level outputs $\mathbf{Z}^{(3)} \in \mathbb{R}^{N \times 2}$, where the superscript (3) indicates the third (final) GCN layer and N is the number of superpixels ($N \leq 150$: SLIC is run with 150 target segments and superpixels of fewer than 5 pixels are discarded), we broadcast each superpixel's feature vector to all its constituent pixels. Let $\mathcal{S} : [H] \times [W] \to [N]$ map each pixel to its superpixel index. The broadcasted feature map $\mathbf{F}_{\text{broad}} \in \mathbb{R}^{H \times W \times 2}$ is defined by:

$$\mathbf{F}_{\text{broad},h,w,k} = \mathbf{Z}^{(3)}_{\mathcal{S}(h,w),k},$$

and set to zero for the pixels of discarded superpixels. The GNN output dimension is set to 2, the number of classes (background and target) of the binary segmentation task; these two channels are latent features learned jointly with the final classifier. The CNN and broadcasted GNN features

[1] https://gricad-gitlab.univ-grenoble-alpes.fr/harkatf/homology-features-for-x-ray-hyperspectral-images-segmentation

are concatenated channel-wise and classified by a final 1×1 convolution:

$$\widehat{\mathbf{Y}} = \mathrm{Conv}_{66\to 2}^{1\times 1}\left(\mathrm{Concat}\left[\mathbf{F}_{\text{cnn}}, \mathbf{F}_{\text{broad}}\right]\right) \in \mathbb{R}^{H\times W\times 2},$$

where $\widehat{\mathbf{Y}}$ contains the per-pixel logits of the two classes. The two branches and the final classifier are trained jointly, with gradients flowing through both branches simultaneously, whereas UMAP, SLIC and persistent homology are fixed preprocessing steps (see Fig. 1).

Training We employ weighted cross-entropy at the pixel level to handle class imbalance:

$$\mathcal{L} = -\frac{\sum_{h,w} \omega_{Y_{h,w}} \log \sigma(\widehat{\mathbf{Y}}_{h,w})_{Y_{h,w}}}{\sum_{h,w} \omega_{Y_{h,w}}}, \tag{1}$$

where σ is the softmax over the two classes, $Y_{h,w} \in \{0, 1\}$ is the ground-truth label of pixel (h, w), and the class weights $\omega_0 = 1.0$ and $\omega_1 = 5.0$ penalize false negatives on the target class.

3. EXPERIMENTAL SETUP AND RESULTS

3.1. Dataset and Metrics

We evaluate on the cigarette dataset from our prior work [6], consisting of ME-XRT image patches depicting cigarette packs placed over luggage backgrounds, with binary ground-truth masks distinguishing target (cigarette) from background. This dataset is particularly suited for evaluating topological descriptors in XRT imagery because the partial superposition of two distinct materials creates spectral mixing regions at their interfaces, where H_1 cyclic structures in the UMAP-projected manifold encode material boundaries. Evaluation uses Accuracy and F1-score [21] at both the pixel and the superpixel level (a superpixel is positive if $\geq$70% of its pixels are target).

3.2. Feature Selection and Ablation

The 11 topological descriptors retained in Section 2.3 were selected by ablating the GNN branch *in isolation*: the CNN branch was disabled during this phase, so the reported scores reflect pure graph-based classification without pixel-level CNN features. We kept the 30 spatial features (25 spectral statistics + 5 geometric descriptors) fixed and progressively added candidate topological features from the pool of 30 defined in Section 2.3, tracking the F1-score [21] on the test set over 10 runs per configuration (the data set is split beforehand into a training and a test set, same as in [6]).

Table 1 reports the results. Topological features alone (30 candidates on raw HSI) yield the lowest performance (F1 = 0.283), because persistent homology is invariant to pixel permutations within a superpixel and thus discards absolute spectral intensities and spatial coordinates. On raw HSI, adding the 11 selected H_1 statistics to the spatial features raises F1 from 0.685 to 0.764, whereas expanding to 19 or 30 topological features degrades it to 0.752 and 0.725, respectively: the additional candidates (H_0 statistics, Betti numbers at fixed scales [7], composite scores) increase node dimensionality from 41 to 60 without yielding extra discriminative power, diluting class-specific signals during graph convolution. The same 11 features bring a similar gain on UMAP-projected data, where F1 rises from 0.778 to 0.823, confirming a significant complementary signal.

Table 1: GNN ablation on raw HSI (16 bands) and UMAP-projected data (5 components). F1 (Target) denotes the F1-score on the positive (target) class; OA denotes Overall Accuracy. All scores are obtained with the GNN branch alone (no CNN fusion).

Input	Node Features	F1 (Target)	OA
Raw HSI	Topo only (30)	0.283 ± 0.011	0.391 ± 0.003
Raw HSI	Spatial only	0.685 ± 0.005	0.912 ± 0.006
Raw HSI	Spatial + 3 Topo	0.683 ± 0.008	0.947 ± 0.006
Raw HSI	**Spatial + 11 Topo**	$\mathbf{0.764 \pm 0.007}$	0.934 ± 0.010
Raw HSI	Spatial + 19 Topo	0.752 ± 0.003	0.948 ± 0.002
Raw HSI	Spatial + 30 Topo	0.725 ± 0.005	0.933 ± 0.003
UMAP	Spatial only	0.778 ± 0.011	0.935 ± 0.009
UMAP	**Spatial + 11 Topo**	$\mathbf{0.823 \pm 0.003}$	$\mathbf{0.956 \pm 0.006}$

This pattern reflects the distinct nature of the descriptors. H_0 statistics correlate with superpixel size and density rather than material identity, whereas the selected H_1 descriptors quantify non-trivial cyclic structures in the spectral point cloud arising from spectral mixing in heterogeneous regions. The compact 11-feature subset thus captures the essential topological structure without redundant dimensionality. **With and without UMAP.** Table 1 also compares persistent homology computed directly in the original 16-band spectral space (raw HSI) and in the 5-dimensional UMAP space. The 11 H_1 features improve the spatial-only baseline in both spaces ($+0.079$ in F1 on raw HSI, $+0.045$ on UMAP), which shows that the cyclic structures they capture are already present in the original spectra and are not created by the embedding. UMAP nevertheless gives the best results, with F1 rising from 0.764 to 0.823 for the same 11-feature configuration: by unfolding the spectral manifold into a lower-dimensional metric space, it removes redundant and noisy dimensions and yields more stable persistence diagrams.

3.3. Hybrid Architecture: Quantitative Results

Table 2 compares the two fusion strategies against CNN-only baselines. The broadcasting hybrid achieves the highest scores across all metrics, surpassing the best CNN baseline by $+0.086$ in superpixel F1 and $+0.070$ in pixel F1. Even the pooling hybrid outperforms both CNN variants, though it trails broadcasting by -0.061 in superpixel F1.

Table 2: Hybrid CNN+GNN: fusion strategies vs. CNN baselines.

	Superpixel		Pixel	
Method	**Acc.**	**F1**	**Acc.**	**F1**
CNN (Pixel-wise)	0.9605 $\pm$0.003	0.8467 $\pm$0.007	0.9433 $\pm$0.004	0.8245 $\pm$0.008
CNN (Superpixel-wise)	0.9387 $\pm$0.005	0.8201 $\pm$0.009	0.9317 $\pm$0.006	0.8144 $\pm$0.009
Hybrid (Pooling)	0.9589 $\pm$0.003	0.8720 $\pm$0.006	0.9496 $\pm$0.004	0.8569 $\pm$0.007
Hybrid (Broadcasting)	**0.9802 $\pm$0.002**	**0.9329 $\pm$0.003**	**0.9640 $\pm$0.003**	**0.8945 $\pm$0.005**

These gains show that persistent homology features carry useful information for XRT segmentation. The CNN captures local spectral-spatial patterns in the UMAP embedding, but it does not characterize the global structure of the spectral manifold. The GNN supplies this through H_1 statistics that quantify cyclic structures in each superpixel's spectral point cloud. These cycles encode spectral mixing, the patterns that appear when the X-ray beam crosses multiple materials, and are not visible to convolutional filters alone. The hybrid model therefore outperforms both CNN baselines because topological descriptors capture patterns that local features miss.

The gap between broadcasting and pooling comes from spatial resolution. Broadcasting sends each superpixel's topological descriptor to all its pixels, letting the final classifier mix local CNN features with global topological information at full resolution. Under pooling, the CNN map is collapsed to superpixel averages before fusion, destroying fine spatial detail. Broadcasting is also more robust to segmentation errors: a bad superpixel poisons all its pixels under pooling, but under broadcasting the wrong topological vector only affects 2 of the 66 channels, which the CNN can partially correct.

These results extend our previous work [6], which relied on the UMAP projection alone: adding H_1 statistics through the GNN branch yields clear further gains, confirming that the topological structure of the spectral manifold carries discriminative information beyond the UMAP coordinates themselves.

Figure 2 supports this: superpixels with high H_1 maximum persistence (warm colors) correspond to regions where the spectral point cloud has complex structure, while homogeneous regions show low persistence (cool colors). The topological signal is strongest where spectral mixing is most pronounced.

4. CONCLUSION

This study investigated the integration of persistent homology features into a hybrid CNN-GNN framework for multi-energy XRT hyperspectral image segmentation. Our findings demonstrate that 1-dimensional persistent homology statistics derived from UMAP-projected superpixel point clouds

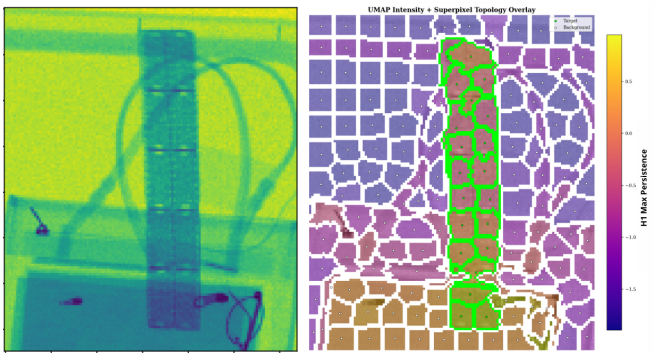


Fig. 2: H_1 maximum persistence per superpixel overlaid on UMAP intensity. Warm colors indicate high persistence in heterogeneous regions; cool colors indicate homogeneous regions.

effectively encode the non-linear structure and spatial complexity of the spectral manifold. In particular, these topological descriptors capture local spectral mixing at material interfaces, providing complementary discriminative information that standard spectral-spatial features miss. Future works may focus on a deeper theoretical and empirical exploration of topological descriptors in X-ray hyperspectral imaging, investigating why persistent homology is particularly suited to modeling material boundaries, extending the framework to complex multi-class scenarios, and exploring adaptive superpixel partitioning strategies.

5. REFERENCES

[1] Ian T. Jolliffe and Jorge Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, pp. 20150202, 2016.

[2] Charles M Bachmann, Thomas L Ainsworth, and Robert A Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.

[3] Leland McInnes, John Healy, and James Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[4] John Healy and Leland McInnes, "Uniform manifold approximation and projection," *Nature Reviews Methods Primers*, vol. 4, no. 1, pp. 82, 2024.

[5] Chen Lou, Mohammed AA Al-qaness, Sike Ni, Dalal Al-Alimi, Robertas Damaševičius, and Saeed Hamood Alsamhi, "Hyperspectral image classification using uniform manifold approximation and projection with fu-

sion deep learning network," *Applied Soft Computing*, vol. 180, pp. 113371, 2025.

[6] Fethi Harkat, Guillaume Gey, Valérie Perrier, Kévin Polisano, and Tiphaine Debeuret, "Manifold learning for hyperspectral images," in *2025 15th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2025, pp. 1–7.

[7] Herbert Edelsbrunner and John Harer, *Computational topology: an introduction*, American Mathematical Soc., 2010.

[8] Gunnar Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[9] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020.

[10] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[11] Danfeng Hong, Lianru Gao, Naoto Yokoya, Jing Yao, Jocelyn Chanussot, Qian Du, and Bing Zhang, "Graph convolutional networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 5966–5978, 2020.

[12] Qinghui Liu, Liangpei Xiao, Jie Yang, and Zhongyi Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 5966–5978, 2020.

[13] Thomas N. Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

[14] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[15] Ulrich Bauer, "Ripser: Efficient computation of vietoris-rips persistence barcodes," *Journal of Applied and Computational Topology*, vol. 5, no. 3, pp. 391–423, 2021.

[16] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier, "Persistence images: A stable vector representation of persistent homology," *Journal of Machine Learning Research*, vol. 18, no. 8, pp. 1–35, 2017.

[17] Peter Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.

[18] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt, "A stable multi-scale kernel for topological machine learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 3071–3078.

[19] Mathieu Carrière, Theo Lacombe, Martin Glisse, and Steve Oudot, "Perslay: A neural network layer for persistence diagrams and new graph topological signatures," in *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, Palermo, Sicily, Italy, 2020, vol. 108 of *Proceedings of Machine Learning Research*, pp. 2786–2796.

[20] Max Horn, Edward De Brouwer, Michael Moor, Yves Moreau, Bastian Rieck, and Karsten Borgwardt, "Topological graph neural networks," *ICLR 2022*, 2022.

[21] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.