Wavelet-Modulus Edge Enhancement and Persistent-Homology Features for Crohn's-Disease Lesion Detection in Capsule Endoscopy

TOSHIMASA OGATA[1], TERUYA MINAMOTO[2]
[1]Graduate School of Science and Engineering, Saga University, Saga, Japan
[2]Department of Computer Science, Saga University, Saga, Japan
E-mail: 24726007@edu.cc.saga-u.ac.jp, minamott@cc.saga-u.ac.jp

Abstract:

Crohn's disease triggers chronic inflammation throughout the gastrointestinal tract, with lesions often developing in the small intestine. Capsule endoscopy (CE) provides a non-invasive window for screening; however, each examination generates thousands of frames that physicians must manually inspect. We propose a training-free pipeline that couples dyadic wavelet-transform modulus (WTM) edge enhancement with topology-aware descriptors (persistence images; PI) and a fractal descriptor (fractal dimension; FD) to discriminate Crohn's-specific lesions from normal mucosa. Each image is mapped to a 1057-dimensional vector (1056 PI + 1 FD). After standardization (zero-mean, unit-variance scaling) and principal component analysis (75% cumulative variance), a two-layer multilayer perceptron is used for classification. On an augmented set of 800 images (balanced normal/lesion), the proposed method achieves 95.4% accuracy, 95.2% recall, and 96.5% specificity. Rivaling ResNet-50 (95.6%) and VGG-16 (95.3%) while requiring no pixel-level annotation and orders of magnitude fewer parameters. These findings demonstrate that topology-aware, handcrafted features can match those of deep networks and may help reduce missed lesions in clinical practice.

Keywords:

Wavelet transform; Persistent homology; Fractal dimension; Capsule endoscopy; Crohn's disease

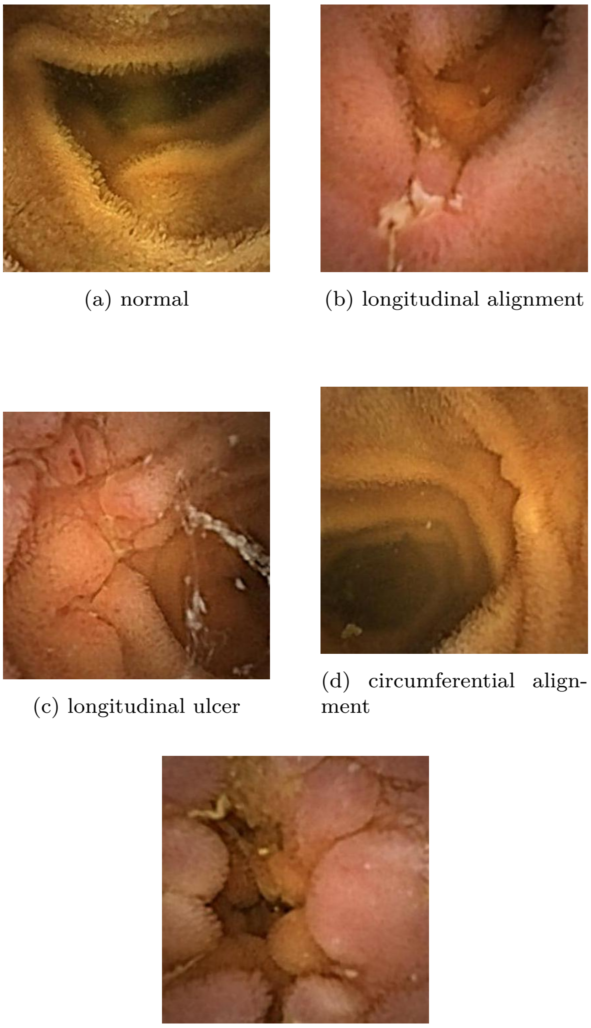
1. Introduction

Crohn's disease (CD) is an inflammatory bowel disease that can affect the entire gastrointestinal tract, with lesions most frequently occurring in the small intestine. Although its etiology remains unclear, genetic predisposition, immune dysregulation, microbiota imbalance, and environmental triggers are believed to interact in complex ways. Because CD typically manifests in young adults and follows a lifelong course, early lesion detection is critical to prevent strictures and other severe complications [1].

Capsule endoscopy (CE) provides a non-invasive means of visualizing the small intestine; however, a single examination can yield tens of thousands of frames that physicians must inspect manually. Automated analysis is, therefore, indispensable. Previous work has addressed binary classification between normal mucosa and a single lesion subtype e.g. circumferential alignment [2], but to our knowledge, no study has tackled comprehensive binary classification that treats all major Crohn's-specific lesion morphologies collectively (Fig. 1).

We propose a training-free pipeline that (i) enhances edge information in CE images via the dyadic wavelet-transform modulus (WTM), (ii) encodes shape characteristics through zero-/first-order persistent homology and a fractal dimension, and (iii) performs classification with a multilayer perceptron (MLP). By focusing on topological and fractal cues that conventional deep networks may overlook, the method captures microstructural differences between normal and lesion tissue, achieving superior diagnostic accuracy while reducing physician workload.

Paper organisation. Section 2 details the WTM-based edge-enhancement procedure. Section 3 describes the extraction of persistent-homology and fractal features, while Section 4 integrates these components into the overall classification pipeline. Section 5 reports the dataset, experimental settings, and comparative results. Finally, Section 6 concludes the paper and outlines future work.



(a) normal

(b) longitudinal alignment

(c) longitudinal ulcer

(d) circumferential alignment

(e) cobblestone appearance

FIGURE 1. normal and lesion images

2 Wavelet Modulus Edge Enhancement

We begin by revisiting the dyadic wavelet transform and introduce the wavelet-transform modulus (WTM), explaining how this edge-enhancement step preserves spatial resolution while accentuating subtle boundaries characteristic of Crohn's lesions.

2.1 Dyadic Wavelet Transform

The dyadic wavelet transform (DYWT) preserves shift invariance and yields frequency components of the same size as the original image. Following Mallat [3], the 2-D DYWT at decomposition level j is expressed as

$$\begin{aligned} C^{j+1}[m,n] &= \sum_k \sum_l h[k]\, h[l]\, C^j[m+2^j k, n+2^j l], \\ D^{j+1}[m,n] &= \sum_k \sum_l h[k]\, g[l]\, C^j[m+2^j k, n+2^j l], \\ E^{j+1}[m,n] &= \sum_k \sum_l g[k]\, h[l]\, C^j[m+2^j k, n+2^j l], \\ F^{j+1}[m,n] &= \sum_k \sum_l g[k]\, g[l]\, C^j[m+2^j k, n+2^j l], \end{aligned} \tag{1}$$

where $h[\cdot]$ and $g[\cdot]$ denote the low-pass and high-pass filters, respectively. C^j, D^j, E^j, and F^j correspond to the low-frequency, horizontal, vertical, and diagonal high-frequency components.

All convolutions in Eq. (1) employ the quadratic–spline analysis filters of order $m = 2$ listed in Table 5.1 of Ref. [3]. After rescaling the tabulated values by $\sqrt{2}$, the low-pass filter $h[n]$ and its quadrature-mirror high-pass counterpart $g[n]$ become

n	-1	0	1	2
$h[n]$	0.177	0.530	0.530	0.177
$g[n]$		-0.707	0.707	

The rescaling guarantees $\|h\|_2 = 1$ and $\|g\|_2 = 1$, thereby preserving energy and yielding a fully shift-invariant dyadic decomposition with no trainable parameters.

2.2 Wavelet Transform Modulus

Following Mallat [3], we define the wavelet-transform modulus (WTM) at scale j as

$$\mathrm{WTM}^j[m,n] = \sqrt{|D^j[m,n]|^2 + |E^j[m,n]|^2}, \tag{2}$$

where D^j and E^j are the horizontal and vertical detail coefficients in Eq. (1). The diagonal component F^j is not used because the edge magnitude is obtained from the gradient vector whose x- and y-components correspond to the horizontal and vertical wavelet responses.

The resulting edge map accentuates lesion boundaries that are otherwise subtle in raw capsule-endoscopy images.

3 Topological and Fractal Feature Extraction

This section details how persistent homology and fractal analysis transform the WTM edge maps into a compact

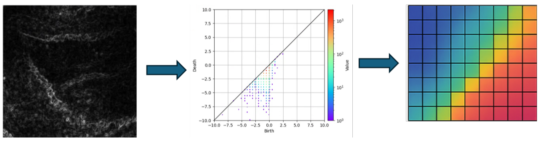


FIGURE 2. Pipeline for converting a persistence diagram into a persistence image.

yet expressive descriptor: first, zero- and first-order persistence diagrams are converted to persistence images (PIs) that encode global geometric structure, and second, a box-counting fractal dimension (FD) augments these features with fine-scale textural complexity.

3.1 Persistent Homology

Persistent homology (PH) [4] is employed to encode shape information contained in the WTM edge map. A sub-level-set filtration treats pixel intensities as scalar values and records the birth and death of topological features as the threshold decreases. Zero-order PH captures the evolution of connected components, whereas first-order PH tracks the formation and disappearance of holes.

Each birth-death pair is plotted on a persistence diagram (PD). Because PDs are not directly amenable to machine-learning models, they are vectorized into persistence images (PIs) [5]. Every point in the PD is weighted and convolved with a Gaussian kernel to form a smooth persistence surface, which is then sampled on a regular grid(Fig. 2). In this work, both zero- and first-order PDs are extracted, each yielding 528 PI features. Consequently, a total of 1056 PI features are obtained per image.

3.2 Fractal Dimension

Fractal dimension (FD) provides a scale-invariant index of structural complexity. We estimate FD on the grayscale WTM edge map by the box–counting method [6], which is conventionally defined for binary images. Accordingly, the edge map is first binarised via adaptive thresholding, yielding a foreground set $S \subset \mathbb{R}^2$. For a square grid of side length ε, let $N(\varepsilon)$ denote the number of grid squares that contain at least one foreground pixel of S. The box–counting dimension is

$$\mathrm{FD} = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log\bigl(1/\varepsilon\bigr)} \quad \approx \quad m, \tag{3}$$

where m is the slope (i.e. the gradient) of the line that best fits the log–log points $\bigl(\log(1/\varepsilon), \log N(\varepsilon)\bigr)$.

Implementation. We sample $\varepsilon \in \{1, 2, 4, \ldots, 128\}$ pixels, plot $\log N(\varepsilon)$ versus $\log(1/\varepsilon)$, and compute m by least-squares regression; this slope is reported as the FD scalar. The value complements the 1056 persistence-image features, forming a compact 1057-dimensional topological–fractal descriptor for each image.

4 Classification Pipeline

Figure 3 illustrates the proposed six-stage workflow that maps a capsule-endoscopy (CE) frame to a binary diagnosis (normal vs. Crohn's lesion comprising longitudinal alignment/ulcer, circumferential alignment, and cobblestone appearance).

1. Pre-processing: Each RGB frame is converted to grayscale and resized to 256 × 256 px for descriptor consistency.

2. Edge enhancement: The level-3 dyadic wavelet-transform modulus (WTM) defined in Eq. (2) is applied, producing an edge map that accentuates mucosal boundaries.

3. Topological–fractal feature extraction:

   - Zero- and first-order persistence diagrams are computed on the WTM edge map and rasterised to 32 × 32 persistence images (PIs); keeping the upper-triangular half yields $2 \times 528 = 1056$ PI features.
   - A single fractal-dimension (FD) value is estimated on the binarised edge map using the box-counting method.

   Concatenating the PI and FD outputs forms a 1057-dimensional descriptor $\mathbf{x}$ for each frame.

4. Standardisation: Each feature is scaled to zero mean and unit variance on the training set.

5. Dimensionality reduction: Principal-component analysis (PCA) retains the first 29 components, explaining 75 % of the cumulative variance.

6. Classification: The reduced features are fed to a two-layer multilayer perceptron (MLP; $29 \to 64 \to 32 \to 2$) with ReLU activation and softmax output to predict normal or lesion.

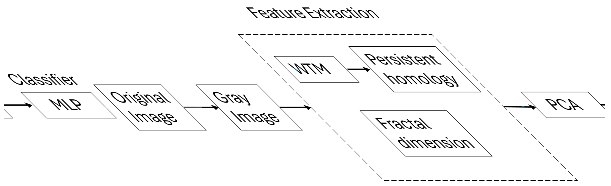


FIGURE 3. Overview of the proposed classification pipeline.

5 Experiments and Results

The Proposed Method was conducted using Google Colab (CPU) with Python 3.11 and HomCloud 4.7.0 [7]. Additionally, ResNet-50 and VGG-16 were implemented on a workstation equipped with an Intel Core i5-12500 CPU and 16 GB RAM using MATLAB R2023b.

5.1 Dataset

The raw dataset consists of 605 small-bowel capsule-endoscopy (CE) images with a spatial resolution of 400 × 400 px: 400 normal frames and 205 Crohn's-lesion frames. Lesions are further categorised into longitudinal alignment (19 images), longitudinal ulcer (34), circumferential alignment (101) and cobblestone appearance (51).

To mitigate class imbalance and enhance robustness, only lesion images were augmented using 90° rotations and horizontal/vertical flips, yielding exactly 400 lesion samples (95, 102, 101 and 102 images in the respective sub-categories) and a balanced total of 800 images (400 normal + 400 lesion).

5.2 Handcrafted Feature Extraction and Classifier Training

A patient-wise 70 / 30 split prevents identity leakage: 560 images for training, 240 for testing.

Feature extraction. For each frame, the dyadic wavelet-transform modulus computed at level 3 (WTM-3) yields an edge-enhanced grayscale map. Zero- and first-order persistence diagrams are then computed on WTM-3 via a lower-star filtration within the birth-death window (−10, 15). Each diagram is rasterized to a 32 × 32 persistence image (PI) with a Gaussian bandwidth of $\sigma = 0.002$, after which only the upper-triangular part of the grid is retained, resulting in 528 informative bins per diagram. Stacking the 0D and 1D PIs yields $2 \times 528 = 1,056$ topological features. A box-counting fractal dimension (FD) scalar is appended, forming the final descriptor $\mathbf{x} \in \mathbb{R}^{1057}$.

Pre-processing and dimensionality reduction. All 1057 features are standardized (zero mean, unit variance) on the training set, followed by principal component analysis (PCA); the first 29 components (75 % cumulative variance) are retained.

Classifier. A two-layer multilayer perceptron (MLP; $29 \to 64 \to 32 \to 2$) with ReLU activation and softmax output is trained for 50 epochs (Adam, learning rate 0.001, batch size 32, L_2 weight decay 0.05). Early stopping monitors validation loss with patience of five epochs.

5.3 Baseline CNNs

For comparison, ResNet-50 and VGG-16 pre-trained on ImageNet [8] are fine-tuned. Both models employ stochastic gradient descent (learning rate 10^{-4}, batch size 16) for five epochs, with validation every ten iterations to curb over-fitting.

5.4 Evaluation Metrics

Model performance was quantified on the held-out test set using three classical medical-imaging indices—accuracy, recall (sensitivity) and specificity. Let TP, TN, FP and FN denote the numbers of true positives, true negatives, false positives and false negatives, respectively. The metrics are defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{5}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{6}$$

Accuracy measures overall correctness, whereas recall (sensitivity) gauges the ability to detect lesions; specificity evaluates how well normal tissue is preserved from false alarms—critical for reducing unnecessary follow-up procedures.

5.5 Quantitative Results

Table 1 shows that the topology-aware pipeline achieves a comparable overall accuracy to VGG-16 (95.4 %), while maintaining a superior balance between lesion recall (95.2 %) and normal-image specificity (96.5 %). ResNet-50 achieves the highest recall but at the expense of markedly lower specificity, indicating a tendency to

TABLE 1. Performance comparison between the proposed method and CNN baselines.

Method	Accuracy (%)	Recall (%)	Specificity (%)
Proposed (WTM-3 + PI + FD)	95.4	95.2	96.5
ResNet-50	93.3	96.7	90.0
VGG-16	95.4	91.7	99.2

TABLE 2. Ablation study showing the contribution of each handcrafted component.

Method	Accuracy (%)	Recall (%)	Specificity (%)
WTM-3 + PI + FD	95.4	95.2	96.5
WTM-3 + PI	94.6	95.2	93.9
WTM-3 + FD	47.9	0.0	100
PI + FD	59.2	68.0	49.6
PI	62.5	80.0	43.5
FD	52.5	33.6	73.0

overdetect lesions. The handcrafted WTM-3+PI+FD descriptor, therefore, provides competitive performance without large-scale annotated data, underscoring its suitability for clinical decision support.

The six-way ablation in Table 2 elucidates the contribution of each handcrafted component.

- WTM-3 + PI + FD (full model) obtains the best balance: 95.4 % accuracy, 95.2 % recall and 96.5 % specificity.
- WTM-3 + PI (–FD) drops accuracy by only 0.8 percentage points (95.4 % → 94.6 %) and lowers specificity (96.5 → 93.9), indicating that the FD scalar mainly sharpens normal–tissue discrimination.
- WTM-3 + FD (–PI) collapses to 47.9 % accuracy and 0 % recall, confirming that PI features are indispensable for lesion detection.
- PI + FD (–WTM) attains 59.2 % accuracy: without WTM edge enhancement, both recall and specificity degrade, demonstrating that WTM concentrates topological cues along mucosal boundaries.
- PI only improves over FD only but still lags far behind the full model (62.5 % vs. 95.4 %), showing that PI captures lesion geometry yet misses textural detail.
- FD only yields the lowest recall (33.6 %) and a modest specificity (73.0 %), revealing that fractal texture alone cannot separate lesions from normal mucosa.

In summary, PI features provide the core discriminative power, WTM delivers boundary-focussed input that maximises PI efficacy, and the single FD scalar fine-tunes specificity. Their combination recreates state-of-the-art accuracy without deep-network fine-tuning, validating the complementary nature of wavelet, topological and fractal information.

6 Conclusion

We proposed a training-free pipeline that fuses dyadic wavelet-transform modulus (WTM) edge enhancement with topology-aware persistence images (PI) and a box-counting fractal-dimension (FD) scalar for binary classification of capsule-endoscopy frames. On a balanced 800-image dataset the complete WTM-3 + PI + FD descriptor achieved 95.4 % accuracy, 95.2 % recall, and 96.5 % specificity—matching VGG-16 in overall accuracy while surpassing ResNet-50 in specificity, all without pixel-level annotation or network fine-tuning.

A six-way ablation revealed the complementary roles of the handcrafted components: WTM focuses topology on mucosal boundaries, PI supplies the essential geometric signal, and the FD scalar fine-tunes normal-tissue specificity. Removing FD reduced accuracy by only 0.8 percentage points, whereas removing PI collapsed recall to zero, underscoring the indispensability of topology and the added value of fractal texture.

Future work will extend the framework to multi-class lesion sub-typing.

Acknowledgements

This research was partially supported by JSPS KAKENHI Grant No. 25K07119.

References

[1] Japan Intractable Diseases Information Center, ”Crohn’s Disease,” 2023. https://www.nanbyou.or.jp/entry/81

[2] T. Suka, H. Omura, and T. Minamoto, ”A Detection Method for Circumferential Alignment of Diminutive Lesions Using Wavelet Transform Modulus Maxima and Higher-Order Local Autocorrelation,” Advances in Intelligent Systems and Computing, vol. 1456, pp. 447–456, 2024,doi:10.1007/978-3-031-56599-1_56.

[3] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd ed. Academic Press, 2009.

[4] H. Edelsbrunner, D. Letscher, and A. Zomorodian, “Topological Persistence and Simplification,” in Proc. 41st IEEE Symp. Foundations of Computer Science (FOCS), 2000, pp. 454–463.

[5] H. Adams et al., “Persistence Images: A Stable Vector Representation of Persistent Homology,” Journal of Machine Learning Research, vol. 18, no. 8, pp. 1–35, 2017.

[6] J. Feder, Fractals. Springer, 1988, chap. ”Random Walks and Fractals,” pp. 163–183.

[7] I. Obayashi, T. Nakamura, and Y. Hiraoka, “Persistent Homology Analysis for Materials Research and Persistent Homology Software: HomCloud,” Journal of the Physical Society of Japan, vol. 91, no. 9, 091013, 2022, doi:10.7566/JPSJ.91.091013.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.