ABD-TDA: ATTENTION-BASED DEEP AND TOPOLOGICAL FEATURE FUSION FOR HISTOPATHOLOGICAL CANCER CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate histopathological image classification remains challenging due to substantial morphological and structural variability across cancer types. While Vision Transformers (ViTs) capture global visual information through self-attention, they do not explicitly represent the topological organization of tissue structures. We propose **ABD-TDA**, a deep and topological feature fusion framework that combines transformer-based visual features with complementary topological descriptors. A pretrained ViT extracts a 768-dimensional visual representation, while a dedicated topological branch processes 1,200 Betti-0 and Betti-1 features derived from RGB and HSV color spaces. The resulting 64-dimensional topological representation is fused with the ViT features for classification.

We evaluate ABD-TDA on the ICIAR2018 breast cancer and UT-Osteosarcoma bone cancer datasets, containing four and three classes, respectively. The framework achieves 94.10% and 96.54% accuracy, with macro-AUC scores of 99.42% and 99.65%, respectively. Ablation results show consistent improvements from feature fusion, increasing accuracy from 92.67% with ViT alone to 94.10% on ICIAR2018 and from 94.15% to 96.54% on UT-Osteosarcoma. These results demonstrate that topological descriptors provide complementary structural information to transformer-based visual features, leading to more discriminative representations for histopathological cancer classification.

1 INTRODUCTION

Cancer remains a major global health challenge, with a substantial and continually increasing burden on healthcare systems worldwide. Recent GLOBOCAN estimates reported more than 20 million new cancer cases and approximately 9.8 million cancer-related deaths globally in 2024, emphasizing the need for accurate and efficient diagnostic technologies Sung et al. (2026). Among different cancer types, breast and bone cancers pose considerable diagnostic challenges due to heterogeneous morphological patterns, complex tissue organization, and variations in cellular and structural characteristics. Accurate histopathological classification is therefore important for diagnosis, treatment planning, and disease management.

Histopathological examination remains a fundamental component of cancer diagnosis. In conventional practice, stained tissue specimens are digitized and examined microscopically by expert pathologists. However, manual analysis is time-consuming and requires substantial domain expertise, while subtle morphological differences can introduce interobserver variability. The increasing availability of whole-slide imaging (WSI) has therefore encouraged the development of computational pathology methods that can assist pathologists through automated and reproducible image analysis Komura & Ishikawa (2018); Tschandl et al. (2020).

Deep learning (DL), particularly convolutional neural networks (CNNs), has significantly advanced automated histopathological image analysis. CNN-based approaches can learn hierarchical representations directly from tissue images and have been successfully applied to cancer detection, classification, grading, and prognosis prediction Litjens et al. (2017); Esteva et al. (2019). Nevertheless, conventional CNNs primarily focus on local spatial patterns, which can limit their ability to capture long-range dependencies and global tissue organization. This limitation is particularly relevant to

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

histopathological images, where diagnostically meaningful information may occur across multiple spatial scales.

Vision Transformers (ViTs) have emerged as an effective alternative to convolution-based architectures for computational histopathology. Through self-attention, ViTs model relationships among image patches and capture global contextual information across the input image. Recent studies have demonstrated the effectiveness of transformer-based architectures for histopathological classification and other computational pathology tasks Dosovitskiy et al. (2021); Xu et al. (2023). More recently, pathology foundation models such as UNI, CONCH, and Virchow have further highlighted the potential of transformer-based representations for learning transferable features across tissues and diagnostic tasks Chen et al. (2024); Lu et al. (2024); Vorontsov et al. (2024). Despite these advances, image-based deep representations do not explicitly encode the structural and topological properties of tissue organization. Related studies have also explored transfer learning, transformer architectures, and alternative visual representations for medical and histopathological image analysis Ahmed (2025b; 2026e); Ahmed & Uddin (2025); Ahmed (2025c;d; 2026g); Rawat & Ahmed (2025); Ahmed (2025e; 2026b;c).

Topological data analysis (TDA) offers a complementary framework for characterizing the structural organization of complex data. Unlike conventional appearance-based descriptors, topological representations capture properties related to connectivity, components, and higher-order structures that can remain informative under certain geometric variations. TDA has increasingly been investigated in medical image analysis, including retinal imaging, brain imaging, and histopathological analysis Ahmed (2026d); Ahmed et al. (2025; 2023); Ahmed (2023); Ahmed & Coskunuzer (2023); Ahmed (2025a); Yadav et al. (2023); Ahmed & Alfrad Nobel Bhuiyan (2025); Ahmed (2026h;a;f). For histopathological images, such information can complement deep visual features by providing explicit descriptors of tissue organization. In particular, Betti-0 characterizes connected components, whereas Betti-1 captures one-dimensional topological structures such as loops or holes. These properties make TDA a promising complementary representation for histopathological classification.

Motivated by the complementary characteristics of deep visual and topological representations, we propose **ABD-TDA**, a deep-topological feature fusion framework for histopathological cancer classification. The framework combines high-level visual features extracted using a Vision Transformer with explicit topological descriptors derived from histopathological images. Specifically, we extract 1,200 Betti-0 and Betti-1 features from RGB and HSV representations. These features are processed through a dedicated multilayer perceptron (MLP) branch and fused with the 768-dimensional ViT representation. The resulting joint representation enables the classifier to exploit complementary visual and structural information.

We evaluate ABD-TDA on two publicly available benchmarks representing distinct cancer types: ICIAR2018 for four-class breast cancer classification and UT-Osteosarcoma for three-class bone cancer classification. In comparison with existing methods, ABD-TDA achieves 94.10% accuracy on ICIAR2018 and 96.54% on UT-Osteosarcoma. The ablation study further shows that fusing ViT and TDA improves performance over the individual branches, achieving 94.10% and 96.54% accuracy on the breast and bone datasets, respectively. These results demonstrate that explicit topological information can provide complementary structural cues to transformer-based visual representations.

The main contributions of this study are summarized as follows:

- We propose **ABD-TDA**, a deep-topological feature fusion framework that combines Vision Transformer features with explicit topological descriptors for histopathological cancer classification.
- We extract 1,200 Betti-0 and Betti-1 features from RGB and HSV representations to capture complementary structural and topological characteristics of histopathological images.
- We employ a dedicated TDA feature-processing branch and fuse its learned representation with the 768-dimensional ViT representation, enabling joint learning from visual and topological information.
- We evaluate the proposed framework on two histopathological benchmarks, ICIAR2018 breast cancer and UT-Osteosarcoma bone cancer, demonstrating its applicability across different tissue types and classification tasks.

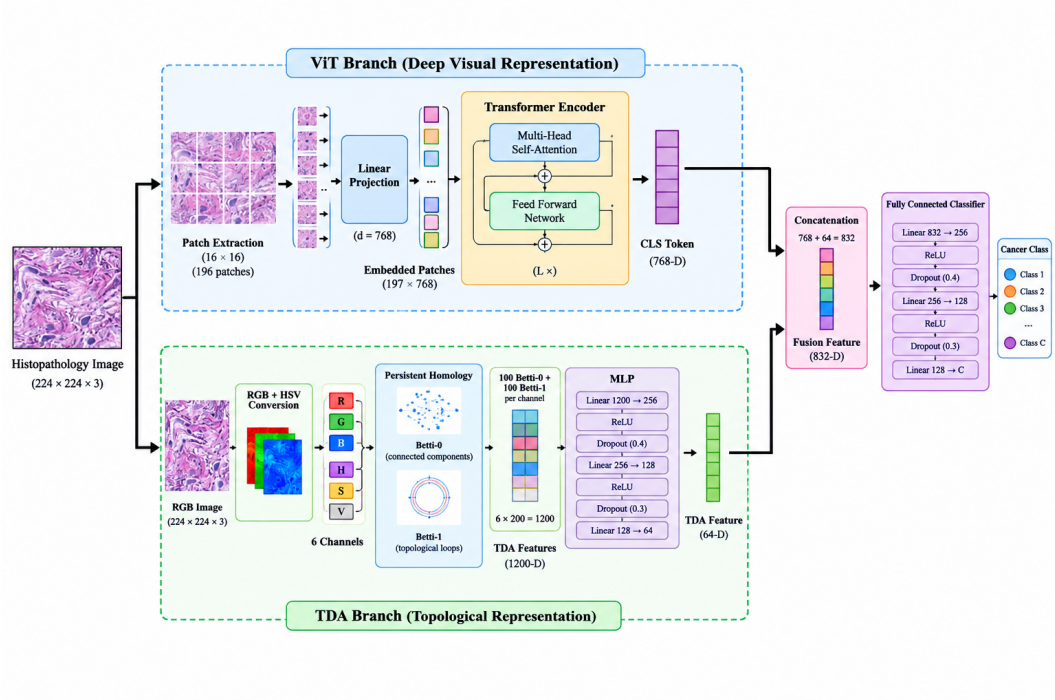


Figure 1: **Flowchart of the proposed ABD-TDA framework.** The framework consists of two complementary branches. The ViT branch extracts deep visual features from image patches using a pretrained Vision Transformer and obtains a 768-dimensional CLS representation. In parallel, the TDA branch converts the image into six RGB and HSV channels and extracts Betti-0 and Betti-1 features, producing a 1200-dimensional topological representation. The TDA features are reduced to 64 dimensions using an MLP. The 768-dimensional ViT and 64-dimensional TDA features are then concatenated to form an 832-dimensional fused representation, which is passed through a fully connected classifier to predict the cancer class.

- Ablation experiments demonstrate that integrating TDA with ViT improves precision, recall, F1-score, accuracy, and AUC compared with the individual ViT and TDA configurations.
- ABD-TDA achieves 94.10% accuracy on ICIAR2018 and 96.54% accuracy on UT-Osteosarcoma in comparison with existing methods.

2 Methodology

This section presents the proposed **ABD-TDA** framework for histopathological cancer classification. The framework combines transformer-based deep visual representations with complementary topological descriptors extracted from multiple color representations of histopathological images. As illustrated in Fig. 1, the framework consists of three main components: (i) Vision Transformer (ViT) feature extraction, (ii) multi-color-space Topological Data Analysis (TDA) feature extraction, and (iii) deep–topological feature fusion followed by classification. Unlike approaches that rely only on learned visual features, ABD-TDA incorporates structural information through Betti-0 and Betti-1 descriptors extracted independently from RGB and HSV channels.

2.1 Vision Transformer Feature Extraction

Let $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$ denote an RGB histopathological image. Each image is resized to 224×224 pixels and normalized to the interval $[0, 1]$ by

$$\mathbf{I}_{\text{norm}} = \frac{\mathbf{I}}{255}, \qquad \mathbf{I}_{\text{norm}} \in [0,1]^{224\times 224\times 3}.$$

The proposed framework employs the pretrained `google/vit-base-patch16-224` architecture Dosovitskiy et al. (2021). The normalized image is divided into non-overlapping 16×16 patches. Thus, the number of image patches is

$$N = \frac{224 \times 224}{16 \times 16} = 196.$$

The image is consequently represented as a sequence of patches,

$$\mathbf{I}_{\text{norm}} \rightarrow \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{196}\}.$$

Each patch is flattened and linearly projected into the transformer embedding space:

$$\mathbf{e}_i = \mathbf{W}_e \, \text{vec}(\mathbf{x}_i) + \mathbf{b}_e,$$

where $\mathbf{W}_e$ and $\mathbf{b}_e$ denote the learnable projection parameters. A learnable classification token $\mathbf{x}_{\text{CLS}}$ is prepended to the patch sequence, and positional embeddings are added to retain spatial information:

$$\mathbf{z}^0 = [\mathbf{x}_{\text{CLS}}, \mathbf{e}_1 + \mathbf{p}_1, \ldots, \mathbf{e}_N + \mathbf{p}_N].$$

The resulting sequence is processed through the transformer encoder layers using multi-head self-attention (MSA) and feed-forward networks. The self-attention operation is given by

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V},$$

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

Following the final transformer encoder layer, the representation of the classification token is used as the deep visual feature:

$$\mathbf{f}_{\text{ViT}} = \mathbf{z}^L_{\text{CLS}} \in \mathbb{R}^{768}.$$

Thus, each histopathological image is represented by a 768-dimensional feature vector containing high-level visual and contextual information.

2.2 Multi-Color-Space Topological Feature Extraction

To complement the learned ViT representation, Topological Data Analysis (TDA) is used to characterize structural and connectivity patterns within histopathological images. Instead of extracting topological information from a single grayscale representation, the proposed approach analyzes both RGB and HSV color spaces (Fig. 2). For an input image, the six individual channels are defined as

$$\mathcal{C} = \{R, G, B, H, S, V\}.$$

The RGB channels preserve the color information of the H&E-stained tissue, whereas HSV separates hue, saturation, and intensity-related information. Using both representations provides complementary views of the tissue structure for topological analysis.

For each channel $c \in \mathcal{C}$, let

$$\mathbf{X}^{(c)} \in \mathbb{R}^{H \times W}$$

denote the corresponding two-dimensional channel image. A cubical filtration is constructed according to the channel intensity values, resulting in a nested sequence of cubical complexes:

$$\mathcal{X}_1^{(c)} \subseteq \mathcal{X}_2^{(c)} \subseteq \cdots \subseteq \mathcal{X}_T^{(c)}.$$

For a filtration threshold t, the corresponding cubical complex is defined as

$$\mathcal{X}^{(c)}(t) = \left\{ \Delta_{ij} \subseteq \mathbf{X}^{(c)} \;\middle|\; \gamma_{ij}^{(c)} \leq t \right\},$$

where $\gamma_{ij}^{(c)}$ denotes the intensity associated with the cubical cell Δ_{ij} in channel c.

Persistent homology is computed across the filtration. For homological dimension k, the resulting persistence diagram is represented by

$$PD_k^{(c)} = \{(b_i, d_i)\}_{i=1}^m,$$

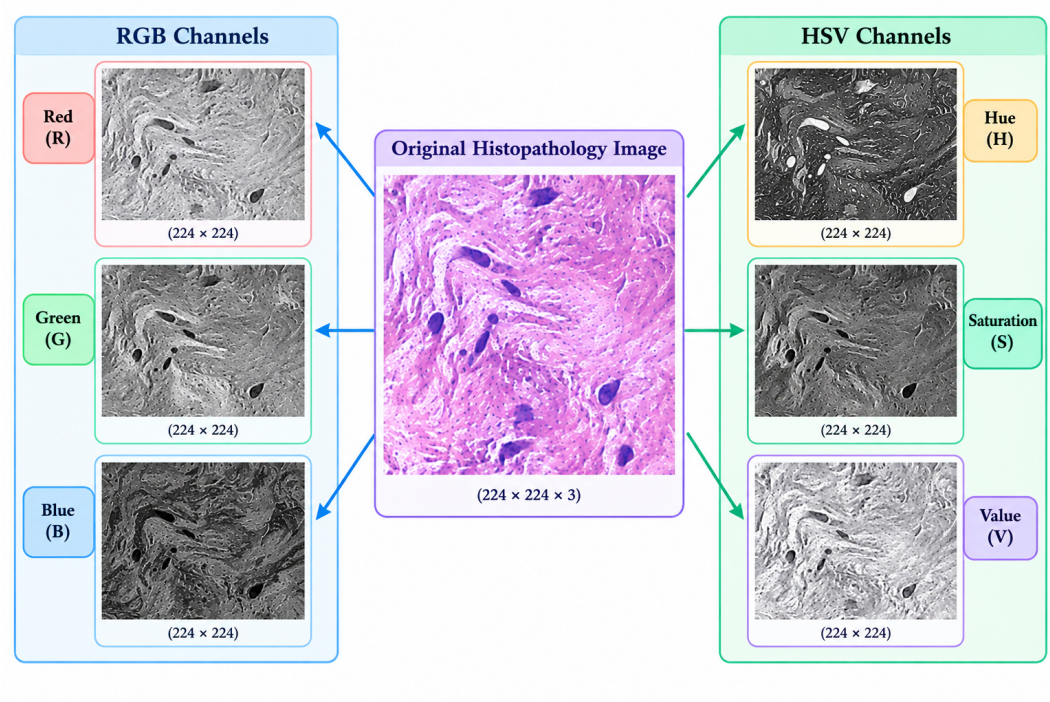


Figure 2: RGB and HSV channel extraction for topological feature generation from a histopathological image.

where b_i and d_i denote the birth and death filtration values of the ith topological structure. In this study, the two homology dimensions H_0 and H_1 are considered. Here, H_0 describes connected components, while H_1 represents one-dimensional loops or holes.

The corresponding Betti curve for homology dimension k is defined as

$$\beta_k^{(c)}(t) = \sum_{i=1}^{m} \mathbb{I}\left(b_i \leq t < d_i\right),$$

where $\mathbb{I}(\cdot)$ is the indicator function. The Betti-0 and Betti-1 curves are discretized into 100 values for each channel. Therefore, the topological representation of one channel is

$$\mathbf{f}_{\mathrm{TDA}}^{(c)} = [\boldsymbol{\beta}_0^{(c)}; \boldsymbol{\beta}_1^{(c)}] \in \mathbb{R}^{200}, \qquad \boldsymbol{\beta}_0^{(c)}, \boldsymbol{\beta}_1^{(c)} \in \mathbb{R}^{100}.$$

The six channel-wise representations are then concatenated:

$$\mathbf{f}_{\mathrm{TDA}} = [\mathbf{f}_{\mathrm{TDA}}^{(R)}; \mathbf{f}_{\mathrm{TDA}}^{(G)}; \mathbf{f}_{\mathrm{TDA}}^{(B)}; \mathbf{f}_{\mathrm{TDA}}^{(H)}; \mathbf{f}_{\mathrm{TDA}}^{(S)}; \mathbf{f}_{\mathrm{TDA}}^{(V)}] \in \mathbb{R}^{1200},$$

where

$$6 \times (100_{\beta_0} + 100_{\beta_1}) = 1200.$$

The resulting representation captures structural and connectivity variations across multiple color representations. The complete TDA feature extraction process is illustrated in Fig. 3.

2.3 TDA Feature Learning Branch

Before fusion with the ViT representation, the 1200-dimensional TDA feature vector is standardized using statistics computed from the training data. The standardized vector is then passed through a dedicated multilayer perceptron (MLP) to obtain a compact task-specific topological representation. The TDA branch consists of three fully connected layers with dimensions

$$1200 \rightarrow 256 \rightarrow 128 \rightarrow 64.$$

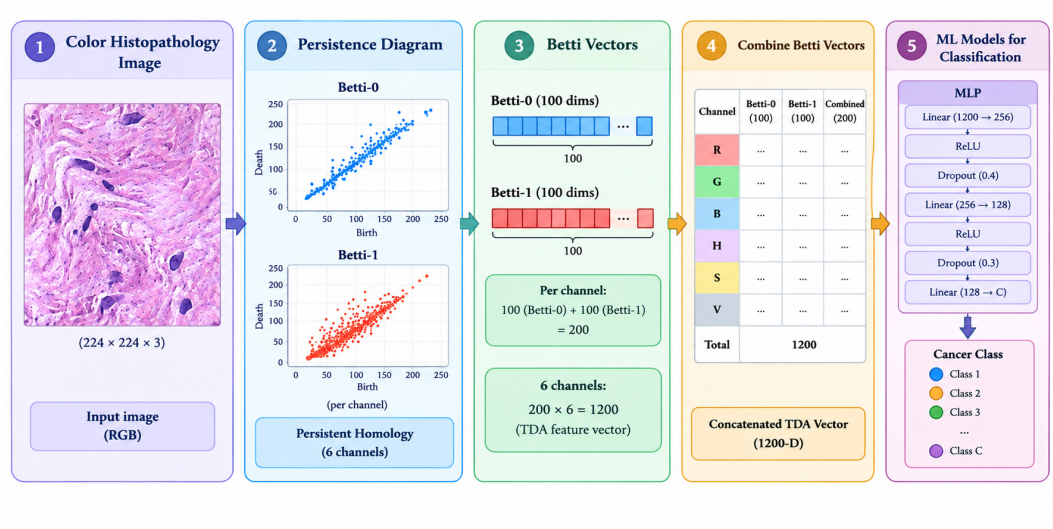


Figure 3: **TDA feature extraction flowchart.** The input histopathological image is converted into RGB and HSV channels. Persistent homology is computed for each channel, followed by Betti-0 and Betti-1 feature extraction and concatenation to form the final topological representation.

The first hidden representation is computed as

$$\mathbf{h}_1 = \mathrm{ReLU}\left(\mathbf{W}_1\mathbf{f}_{\mathrm{TDA}} + \mathbf{b}_1\right), \qquad \mathbf{h}_1 \in \mathbb{R}^{256}.$$

A dropout probability of 0.3 is applied before the second layer:

$$\mathbf{h}_2 = \mathrm{ReLU}\left(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2\right), \qquad \mathbf{h}_2 \in \mathbb{R}^{128}.$$

After a second dropout operation with probability 0.3, the final TDA embedding is obtained as

$$\mathbf{h}_{\mathrm{TDA}} = \mathrm{ReLU}\left(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3\right) \in \mathbb{R}^{64}.$$

Thus, the TDA branch transforms the original 1200-dimensional topological descriptor into a compact 64-dimensional representation for fusion with the ViT features.

2.4 Deep–Topological Feature Fusion

The main component of ABD-TDA is the integration of the deep visual representation learned by ViT with the compact topological representation learned by the TDA branch. The ViT feature and TDA embedding are given by

$$\mathbf{f}_{\mathrm{ViT}} \in \mathbb{R}^{768}, \qquad \mathbf{h}_{\mathrm{TDA}} \in \mathbb{R}^{64}.$$

They are combined through feature-level concatenation:

$$\mathbf{f}_{\mathrm{fusion}} = [\mathbf{f}_{\mathrm{ViT}}; \mathbf{h}_{\mathrm{TDA}}] \in \mathbb{R}^{832}, \qquad 832 = 768 + 64.$$

The fused representation combines the global visual information learned by the transformer with explicit structural information encoded by persistent homology.

2.5 Classification

The 832-dimensional fused representation is passed through a fully connected classification head with dimensions

$$832 \to 256 \to 128 \to C.$$

The first classification layer produces a 256-dimensional hidden representation:

$$\mathbf{z}_1 = \mathrm{ReLU}\left(\mathbf{W}_4\mathbf{f}_{\mathrm{fusion}} + \mathbf{b}_4\right), \qquad \mathbf{z}_1 \in \mathbb{R}^{256}.$$

A dropout probability of 0.4 is applied before the second fully connected layer:

$$\mathbf{z}_2 = \mathrm{ReLU}\left(\mathbf{W}_5\mathbf{z}_1 + \mathbf{b}_5\right), \qquad \mathbf{z}_2 \in \mathbb{R}^{128}.$$

A final dropout probability of 0.3 is applied before the output layer:

$$\mathbf{o} = \mathbf{W}_6\mathbf{z}_2 + \mathbf{b}_6, \qquad \mathbf{o} \in \mathbb{R}^C.$$

Here, C denotes the number of diagnostic classes. In this study,

$$C = \begin{cases} 4, & \text{for ICIAR2018 (Breast Cancer),} \\ 3, & \text{for UT-Osteosarcoma (Bone Cancer).} \end{cases}$$

The output logits are converted into class probabilities using the Softmax function:

$$\hat{y}_c = \frac{\exp(o_c)}{\sum_{j=1}^{C} \exp(o_j)}.$$

The predicted class is then determined by $\hat{c} = \arg\max_c \hat{y}_c$.

The complete network is optimized using categorical cross-entropy loss:

$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{C} y_{i,c}\log\left(\hat{y}_{i,c}\right),$$

where B denotes the batch size, $y_{i,c}$ is the ground-truth class indicator, and $\hat{y}_{i,c}$ is the predicted probability for class c.

3 Experiments

3.1 Datasets

We evaluate the proposed ABD-TDA framework on two publicly available histopathological benchmarks: the **BACH (ICIAR2018) Breast Cancer** dataset Aresta et al. (2019) and the **UT-Osteosarcoma (UT-Osteo) Bone Cancer** dataset Arunachalam et al. (2019). For fair comparison, we follow the established training–test partitions used by the corresponding comparative studies. The detailed class-wise distributions are provided in Appendix Table 4.

BACH (Breast Cancer): The BACH dataset Aresta et al. (2019) contains H&E-stained breast histopathology images from four classes: *Normal, Benign, In Situ*, and *Invasive*. The original dataset contains 400 microscopy images, with 100 images per class. We use the corresponding image-tile dataset and follow the established partition for comparison with prior work.

UT-Osteosarcoma (Bone Cancer): The UT-Osteosarcoma dataset Arunachalam et al. (2019) consists of H&E-stained osteosarcoma images collected at the University of Texas Southwestern Medical Center. The images are categorized into three classes: *Non-Tumor, Non-Viable Tumor*, and *Viable Tumor*. We follow the established training–test partition used in previous studies. The detailed sample distributions are reported in Appendix Table 4.

No data augmentation is applied to either dataset. Thus, the evaluation focuses on the learned deep and topological representations without introducing additional training samples.

3.2 Experimental Setup

To ensure a fair and consistent comparison with previously published methods, we use the same established training–test partitions adopted in the corresponding studies reported in our comparative tables. We do not introduce a new random partition of the datasets. For the UT-Osteo (bone cancer) dataset, we follow the commonly reported 70:30 training–test protocol, while for the ICIAR2018 (BACH) breast cancer dataset, we follow the established 80:20 training–test protocol. The exact class-wise numbers of training and test samples used in our experiments are reported in Appendix Table 4. The complete model configuration is provided in Appendix Table 5. We train the model using AdamW with a learning rate of 1×10^{-4}, weight decay of 1×10^{-4}, cross-entropy loss, and a batch size of 16. Early stopping is applied during training.

4 RESULTS AND DISCUSSION

The proposed **ABD-TDA** framework was evaluated on the UT-Osteosarcoma and ICIAR2018 datasets using the same train–test settings as the corresponding comparative studies. The results are summarized in Tables 1 and 2. For comparison, the TDA-LBP-Gabor method Yadav et al. (2023) uses 2,800 handcrafted features, comprising 1,600 TDA, 400 Gabor, and 800 LBP features. In contrast, our method extracts 1,200 topological features directly from the RGB and HSV channels and integrates them with ViT features.

4.1 BONE CANCER CLASSIFICATION

Table 1 compares ABD-TDA with existing methods on UT-Osteosarcoma. Mishra-CNN Arunachalam et al. (2019), Mishra-SVM Arunachalam et al. (2019), VGG19 Anisuzzaman et al. (2021), and TDA-LBP-Gabor Yadav et al. (2023) achieve accuracies of 93.30%, 89.90%, 93.91%, and 94.20%, respectively. ABD-TDA achieves the highest accuracy of **96.54%**, improving over TDA-LBP-Gabor by 2.34 percentage points and VGG19 by 2.63 percentage points.

The ablation results in Table 3 further show the benefit of TDA fusion. ViT achieves 94.15% accuracy, while ViT+TDA reaches **96.54%**, with AUC increasing from 98.88% to 99.65%. These results indicate that the TDA features provide complementary structural information to the visual features learned by ViT.

Table 1: Performance comparison of the proposed method with existing state-of-the-art methods on the UT-Osteosarcoma dataset for bone cancer classification.

Method	**Train:Test**	**# Classes**	**Accuracy (%)**
Mishra-CNN Arunachalam et al. (2019)	70:30	3	93.30
Mishra-SVM Arunachalam et al. (2019)	70:30	3	89.90
VGG19 Anisuzzaman et al. (2021)	70:30	3	93.91
TDA-LBP-Gabor Yadav et al. (2023)	70:30	3	94.20
ABD-TDA (Ours)	70:30	3	**96.54**

4.2 BREAST CANCER CLASSIFICATION

Table 2 presents the results on ICIAR2018. Previous methods report accuracies ranging from 79.00% to 92.50%, while the TDA-LBP-Gabor-based method Yadav et al. (2023) achieves 91.64%. ABD-TDA obtains the highest accuracy of **94.10%**, improving over DCNN Kassani et al. (2019) by 1.60 percentage points and TDA by 2.46 percentage points.

The ablation results show that ViT achieves 92.67% accuracy, whereas ViT+TDA reaches **94.10%**. The fused model also improves macro-precision from 92.72% to 94.17%, macro-recall from 92.64% to 94.10%, macro-F1 from 92.68% to 94.13%, and AUC from 99.14% to 99.42%. Overall, the results on both datasets demonstrate that combining ViT-based visual features with Betti-0 and Betti-1 topological features provides a more effective representation for histopathological cancer classification.

5 ABLATION STUDY

An ablation study was conducted to evaluate the contribution of the ViT and TDA components in the proposed ABD-TDA framework. Three configurations were compared: MLP-TDA, which uses only topological features; ViT, which uses only deep visual features; and ViT+TDA, which combines both representations. As reported in Table 3, ViT outperforms MLP-TDA on both datasets, achieving 92.67% versus 83.50% accuracy on ICIAR2018 and 94.15% versus 86.60% on UT-Osteosarcoma. However, the TDA features still provide useful discriminative information, with AUC values of 96.45% and 96.26% for ICIAR2018 and UT-Osteosarcoma, respectively. The fusion of ViT and TDA gives the best overall performance. On ICIAR2018, ViT+TDA improves

Table 2: Performance comparison of the proposed method with existing state-of-the-art methods on the ICIAR2018 dataset for breast cancer classification.

Method	Train:Test	# Classes	Accuracy (%)
Kwok Kwok (2018)	75:25	4	79.00
Nawaz Nawaz et al. (2018)	80:20	4	81.25
Rakhlin Guo et al. (2018)	80:20	4	87.20
Vang Vang et al. (2018)	75:25	4	87.50
DCNN Kassani et al. (2019)	75:25	4	92.50
TDA-LBP-Gabor Yadav et al. (2023)	80:20	4	91.64
ABD-TDA (Ours)	80:20	4	**94.10**

accuracy from 92.67% to 94.10% and AUC from 99.14% to 99.42%. On UT-Osteosarcoma, accuracy increases from 94.15% to 96.54%, while macro-precision, macro-recall, macro-F1, and AUC reach 96.37%, 96.42%, 96.39%, and 99.65%, respectively. These consistent improvements show that TDA provides complementary structural information to the visual features learned by ViT. The results therefore support the effectiveness of combining deep visual and topological representations in ABD-TDA.

Table 3: Ablation study of the proposed deep and topological feature fusion framework on the ICIAR2018 and UT-Osteosarcoma datasets. All reported metrics are calculated using macro averaging. The best-performing configuration for each dataset is highlighted in bold.

Dataset	Model	# Classes	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	AUC (%)
ICIAR2018 (Breast)	MLP (TDA)	4	83.55	83.46	83.50	83.50	96.45
	ViT	4	92.72	92.64	92.68	92.67	99.14
	ViT + TDA	4	**94.17**	**94.10**	**94.13**	**94.10**	**99.42**
UT-Osteosarcoma (Bone)	MLP (TDA)	3	86.37	85.54	85.95	86.60	96.26
	ViT	3	93.98	94.17	94.07	94.15	98.88
	ViT + TDA	3	**96.37**	**96.42**	**96.39**	**96.54**	**99.65**

6 LIMITATIONS

Despite the promising performance of ABD-TDA, several limitations should be acknowledged. First, the evaluation is conducted on two benchmark datasets, which may not fully represent the diversity of real-world histopathological images. Second, the proposed framework introduces additional computational overhead due to the extraction and processing of topological features alongside ViT representations. Third, the current evaluation is limited to classification performance and does not assess external clinical validation or interpretability. Future work will investigate larger and more diverse multi-center datasets, computationally efficient TDA extraction, and clinically oriented validation and explainability methods.

7 CONCLUSION

In this study, we proposed **ABD-TDA**, a deep-topological feature fusion framework that integrates Vision Transformer representations with topological descriptors for histopathological cancer classification. The proposed approach combines 768-dimensional ViT features with 1,200 Betti-0 and Betti-1 features extracted from RGB and HSV representations. Experiments on the ICIAR2018 and UT-Osteosarcoma datasets demonstrated that the fusion of deep and topological representations consistently outperforms the individual ViT and TDA configurations. The proposed framework achieved accuracies of 94.10% and 96.54% in the respective SOTA comparisons, while the ablation experiments further demonstrated the effectiveness of ViT–TDA fusion, achieving 94.10% and 96.54% accuracy, respectively. These results highlight the complementary nature of learned visual and explicit topological representations and demonstrate the potential of deep-topological feature fusion for robust histopathological cancer classification.

AI USE STATEMENT

The authors used an online language-editing platform to identify and correct grammatical errors and improve the clarity and readability of the manuscript. The authors reviewed and approved all final text.

ETHICS STATEMENT

Not applicable. This study did not involve human participants or animals, and publicly available datasets were used.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide the complete implementation of the proposed **ABD-TDA** framework, including the model architecture and experimental code, in an anonymous repository: `https://anonymous.4open.science/r/ABD-TDA-E086/README.md`. The main paper describes the overall architecture, feature extraction, and fusion strategy, while the experimental setup specifies the datasets, train–test protocols, preprocessing, optimization settings, and evaluation metrics. Detailed dataset distributions, model hyperparameters, and additional experimental results are provided in the appendix. The experiments use publicly available BACH (ICIAR2018) and UT-Osteosarcoma datasets and follow the established partitions used in the corresponding comparative studies. These resources together provide the necessary information and implementation for reproducing the reported results.

REFERENCES

Faisal Ahmed. *Topological Machine Learning in Medical Image Analysis*. PhD thesis, The University of Texas at Dallas, 2023.

Faisal Ahmed. 3d-tda: Topological feature extraction from 3d mri for alzheimer's disease classification. *Available at SSRN 5882122*, 2025a.

Faisal Ahmed. Colormap-enhanced vision transformers for mri-based multiclass (4-class) alzheimer's disease classification. *arXiv preprint arXiv:2512.16964*, 2025b.

Faisal Ahmed. Histovit: Vision transformer for accurate and scalable histopathological cancer diagnosis. *arXiv preprint arXiv:2508.11181*, 2025c.

Faisal Ahmed. Repvit-cxr: A channel replication strategy for vision transformers in chest x-ray tuberculosis and pneumonia classification. *arXiv preprint arXiv:2509.08234*, 2025d.

Faisal Ahmed. Transfer learning with efficientnet for accurate leukemia cell classification. *arXiv preprint arXiv:2508.06535*, 2025e.

Faisal Ahmed. Brain tumor classification from 3d mri using persistent homology and betti features: A topological data analysis approach on brats2020. *arXiv preprint arXiv:2603.13771*, 2026a.

Faisal Ahmed. Bridging topology and deep representation learning: A tda-vit fusion model for four-class brain tumor classification. *arXiv preprint arXiv:2606.00927*, 2026b.

Faisal Ahmed. Enhancing brain tumor classification using vision transformers with colormap-based feature representation on brisc2025 dataset. *arXiv preprint arXiv:2603.21234*, 2026c.

Faisal Ahmed. Four-stage alzheimer's disease classification from mri using topological feature extraction, feature selection, and ensemble learning. *arXiv preprint arXiv:2601.00918*, 2026d.

Faisal Ahmed. Hog-cnn: Integrating histogram of oriented gradients with convolutional neural networks for retinal image classification. *Journal of Imaging Informatics in Medicine*, pp. 1–16, 2026e.

Faisal Ahmed. Hybrid topological and deep feature fusion for accurate mri-based alzheimer's disease severity classification. *arXiv preprint arXiv:2602.00956*, 2026f.

Faisal Ahmed. Pseudocolorvit-cxr: Colormap-enhanced vision transformers for tuberculosis and pneumonia detection from grayscale chest x-ray images. In *2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, pp. 1–7. IEEE, 2026g.

Faisal Ahmed. Topological signatures vs. gradient histograms: A comparative study for medical image classification. *Biomedical Signal Processing and Control*, 125:110783, 2026h.

Faisal Ahmed and Mohammad Alfrad Nobel Bhuiyan. Robust five-class and binary diabetic retinopathy classification using transfer learning and data augmentation. *arXiv e-prints*, pp. arXiv–2507, 2025.

Faisal Ahmed and Baris Coskunuzer. Tofi-ml: Retinal image screening with topological machine learning. In *Annual Conference on Medical Image Understanding and Analysis*, pp. 281–297. Springer, 2023.

Faisal Ahmed and MD Joshem Uddin. Ocuvit: A vision transformer-based approach for automated diabetic retinopathy and amd classification. *Journal of Imaging Informatics in Medicine*, pp. 1–11, 2025.

Faisal Ahmed, Brighton Nuwagira, Furkan Torlak, and Baris Coskunuzer. Topo-CXR: Chest X-ray TB and Pneumonia Screening with Topological Machine Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2326–2336, 2023.

Faisal Ahmed, Mohammad Alfrad Nobel Bhuiyan, and Baris Coskunuzer. Topo-cnn: Retinal image analysis with topological deep learning. *Journal of Imaging Informatics in Medicine*, pp. 1–17, 2025.

DM Anisuzzaman, Hosein Barzekar, Ling Tong, Jake Luo, and Zeyun Yu. A deep learning study on osteosarcoma detection from histological images. *Biomedical Signal Processing and Control*, 69:102931, 2021.

Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019.

Harish Babu Arunachalam, Rashika Mishra, Ovidiu Daescu, Kevin Cederberg, Dinesh Rakheja, Anita Sengupta, David Leonard, Rami Hallac, and Patrick Leavey. Viable and necrotic tumor assessment from WSI of osteosarcoma using ML and DL models. *PloS one*, 14(4):e0210706, 2019.

Osama Attallah. Vision transformers in medical computer vision: A comprehensive survey. *Computers in Biology and Medicine*, 152:106407, 2023.

Jianpeng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

Andre Esteva, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. A guide to deep learning in healthcare. *Nature medicine*, 25(1):24–29, 2019.

Yao Guo, Huihui Dong, Fangzhou Song, Chuang Zhu, and Jun Liu. Breast cancer histology image classification based on deep neural networks. In *ICIAR*, pp. 827–836. Springer, 2018.

Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J Wesolowski, Kevin A Schneider, and Ralph Deters. Breast cancer diagnosis with transfer learning and global pooling. In *ICTC*, pp. 519–524. IEEE, 2019.

Daisuke Komura and Seiya Ishikawa. Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16:34–42, 2018.

Scotty Kwok. Multiclass classification of breast cancer in whole-slide images. In *International conference image analysis and recognition*, pp. 931–940. Springer, 2018.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42: 60–88, 2017.

Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature medicine*, 30(3):863–874, 2024.

Wajahat Nawaz, Sagheer Ahmed, Ali Tahir, and Hassan Aqeel Khan. Classification of breast cancer histology images using alexnet. In *ICIAR*, pp. 869–876. Springer, 2018.

Shekoofeh Azizi Rao et al. Global histopathological image classification using convolutional and vision transformer networks. *arXiv preprint arXiv:2106.15232*, 2021.

Richa Rawat and Faisal Ahmed. Efficient breast and ovarian cancer classification via vit-based preprocessing and transfer learning. *arXiv preprint arXiv:2509.18553*, 2025.

Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

Hyuna Sung, Adalberto M Filho, Mathieu Laversanne, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2024: Globocan estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA: A Cancer Journal for Clinicians*, 76(4):e70090, 2026.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning (ICML)*, pp. 10347–10357. PMLR, 2021.

Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, and Harald Kittler. Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234, 2020.

Yeeleng S Vang, Zhen Chen, and Xiaohui Xie. Deep learning framework for multi-class breast cancer histology image classification. In *ICIAR*, pp. 914–922. Springer, 2018.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature medicine*, 30(10): 2924–2935, 2024.

Xinyu Wang, Jianpeng Chen, Chong You, and Yuyin Zhou. Transformer-based methods in medical imaging: A review. *Neurocomputing*, 479:79–97, 2022.

Hongming Xu, Qi Xu, Fengyu Cong, Jeonghyun Kang, Chu Han, Zaiyi Liu, Anant Madabhushi, and Cheng Lu. Vision transformers for computational histopathology. *IEEE Reviews in Biomedical Engineering*, 17:63–79, 2023.

Ankur Yadav, Faisal Ahmed, Ovidiu Daescu, Reyhan Gedik, and Baris Coskunuzer. Histopathological cancer detection with topological signatures. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1610–1619. IEEE, 2023.

Yifan Zhao, Lu Yang, Ruoyu Kong, Ying Xu, and Shaoting Zhang. Recent advances in deep learning for digital pathology image analysis: A comprehensive review with a focus on histopathological applications. *IEEE Transactions on Medical Imaging*, 42(7):1692–1712, 2023.

Kai Zhou, Han Lin, Ruixue Wang, Shuai Zhang, and Qiang Wang. A comprehensive survey on deep learning for histopathological image analysis. *Artificial Intelligence Review*, 57:213–268, 2024.

A Appendix

B Related Work

B.1 Deep Learning for Histopathological Image Analysis

Deep learning (DL) has substantially advanced computational pathology, particularly for automated histopathological image analysis, cancer detection, grading, and classification Litjens et al. (2017); Esteva et al. (2019). Convolutional neural networks (CNNs) have been extensively adopted because of their ability to automatically learn hierarchical representations from histological images, ranging from low-level cellular patterns to higher-level tissue characteristics Shen et al. (2017). CNN-based approaches have demonstrated strong performance in tumor localization, tissue segmentation, cancer classification, and other pathology-related tasks. However, conventional CNNs primarily rely on local convolutional operations, which can make the modeling of long-range spatial dependencies and global tissue organization challenging. In addition, computational pathology systems often require large annotated datasets and may be sensitive to staining variations, tissue heterogeneity, and differences in acquisition protocols Komura & Ishikawa (2018); Tschandl et al. (2020).

The limitations of conventional CNN-based approaches have motivated the exploration of architectures capable of modeling broader contextual relationships within histopathological images. In particular, the large spatial extent and structural complexity of whole-slide and patch-based pathology images make global representation learning an important consideration. Consequently, recent research has increasingly investigated attention-based and transformer-based architectures as alternatives or complements to convolutional models.

B.2 Vision Transformers in Computational Pathology

The introduction of the Vision Transformer (ViT) demonstrated that transformer architectures can effectively learn visual representations by treating an image as a sequence of patches and modeling their relationships through self-attention Dosovitskiy et al. (2021). Unlike conventional CNNs, which progressively construct receptive fields through local convolutions, ViTs can directly model relationships between distant image regions. Data-efficient transformer variants such as DeiT further demonstrated that transformer-based visual models can be trained effectively under more limited data regimes Touvron et al. (2021).

Transformers have subsequently been adopted across medical image analysis for classification, segmentation, and multimodal learning Chen et al. (2021); Rao et al. (2021); Wang et al. (2022). Their ability to capture long-range dependencies is particularly relevant to histopathology, where diagnostic information can be distributed across cellular, glandular, and tissue-level structures. Recent studies and surveys have specifically highlighted the growing role of vision transformers in computational histopathology Attallah (2023); Zhao et al. (2023); Zhou et al. (2024); Xu et al. (2023).

The development of pathology-specific foundation models has further demonstrated the potential of transformer-based representations. For example, UNI introduced a general-purpose pathology foundation model capable of learning transferable representations from large-scale histopathology data Chen et al. (2024). CONCH extended this direction through vision-language pretraining for computational pathology Lu et al. (2024), while Virchow demonstrated the effectiveness of large-scale self-supervised learning for general-purpose pathology representation learning Vorontsov et al. (2024). These developments suggest that transformer-based architectures can provide rich semantic representations across heterogeneous tissue types and diagnostic tasks.

Several recent studies have also explored transfer learning, transformer architectures, and alternative visual representations for medical and histopathological image analysis Ahmed (2025b; 2026e);

Ahmed & Uddin (2025); Ahmed (2025c;d; 2026g); Rawat & Ahmed (2025); Ahmed (2025e; 2026b;c). Collectively, these studies demonstrate the increasing effectiveness of transformer-based representation learning for medical image classification. Nevertheless, most transformer-based approaches primarily learn appearance and semantic representations directly from image data and do not explicitly incorporate topological information describing tissue organization.

B.3 Topological Data Analysis in Medical and Histopathological Imaging

Topological Data Analysis (TDA) provides a complementary mathematical framework for characterizing structural properties of complex data. Unlike conventional intensity-, texture-, or deep feature-based representations, TDA focuses on properties such as connectivity, components, loops, and other structural relationships. Persistent homology is particularly useful for extracting robust topological descriptors from high-dimensional data, with Betti numbers providing interpretable summaries of topological structures.

TDA has increasingly been explored in medical image analysis, including applications in retinal imaging, brain imaging, and histopathological analysis Ahmed (2026d); Ahmed et al. (2025; 2023); Ahmed (2023); Ahmed & Coskunuzer (2023); Ahmed (2025a); Yadav et al. (2023); Ahmed & Alfrad Nobel Bhuiyan (2025); Ahmed (2026h;a;f). These studies demonstrate that topological descriptors can capture structural characteristics that may not be adequately represented by conventional image features alone. In particular, topological information can provide complementary descriptions of tissue organization and morphological structure, making it potentially valuable for histopathological classification.

The usefulness of TDA is especially relevant to histopathology because tissue morphology is inherently structural. Cellular organization, connected tissue regions, glandular structures, and spatial relationships can contain important diagnostic information. However, TDA-based representations alone may not capture the rich semantic and appearance-based information learned by modern deep neural networks. This motivates the integration of topological descriptors with learned visual representations rather than treating the two approaches as competing alternatives.

B.4 Deep and Topological Feature Fusion

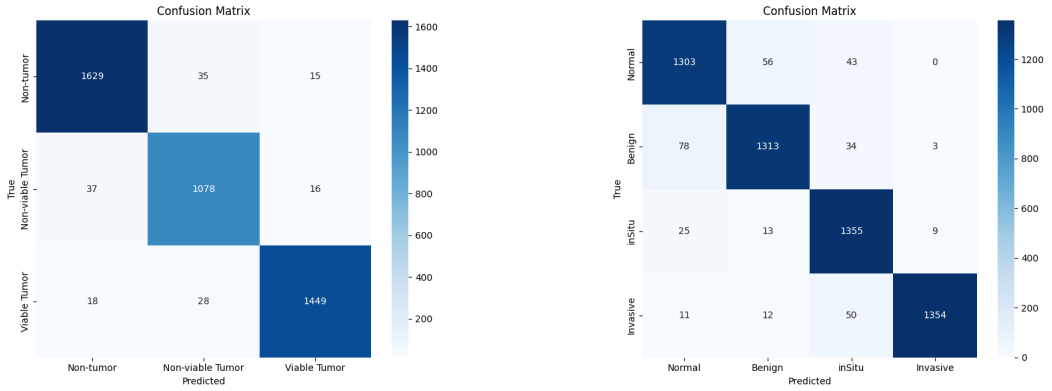
Recent advances in computational pathology suggest that combining complementary representations can improve the robustness of automated diagnosis. Deep learning models can learn highly discriminative semantic representations directly from image data, whereas hand-crafted or mathematically derived descriptors can provide additional structural information. Transformer-based pathology models further demonstrate the benefits of learning representations that capture both local morphology and broader tissue context Chen et al. (2024); Lu et al. (2024); Vorontsov et al. (2024).

Despite these advances, relatively limited attention has been given to explicitly combining transformer-based visual representations with topological descriptors for histopathological cancer classification. In particular, the integration of Betti-based features extracted from multiple color representations with transformer features remains insufficiently explored. This creates an opportunity to investigate whether explicit topological information can complement the global visual representations learned by a Vision Transformer.

B.5 Positioning of the Proposed Method

Motivated by the complementary characteristics of deep visual and topological representations, we propose **ABD-TDA**, a deep-topological feature fusion framework for histopathological cancer classification. Unlike approaches that rely exclusively on CNN or transformer representations, the proposed framework combines a fine-tuned Vision Transformer with an independent TDA branch. Specifically, 1,200 Betti-0 and Betti-1 descriptors are extracted from RGB and HSV representations and transformed through a dedicated multilayer perceptron before being fused with the 768-dimensional ViT representation.

The proposed framework is evaluated on two histopathological cancer datasets, ICIAR2018 for breast cancer and UT-Osteosarcoma for bone cancer. In addition to comparison with existing meth-



(a) Confusion matrix for Bone Cancer classification.

(b) Confusion matrix for Breast Cancer classification.

Figure 4: Confusion matrices obtained by the proposed ABD-TDA framework on the UT-Osteosarcoma (Bone Cancer) and ICIAR2018 (Breast Cancer) datasets. The matrices illustrate the distribution of true and predicted class labels, providing a detailed view of the model's classification performance across the diagnostic categories.

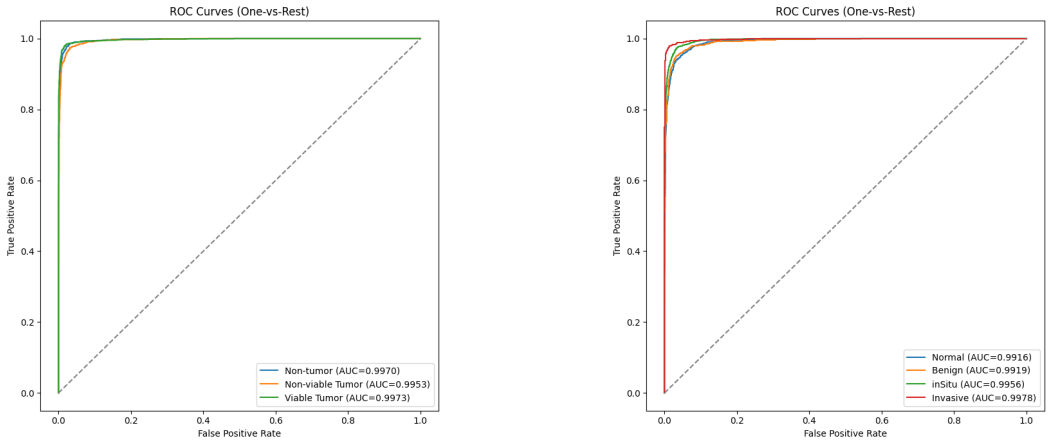
ods, an ablation study is performed to quantify the individual and combined contributions of ViT and TDA representations. The experimental results demonstrate that the fusion of deep visual and topological features consistently improves classification performance over the individual branches, supporting the hypothesis that explicit structural information provides complementary information to transformer-based visual representations.

Table 4: Class-wise distribution of training and test samples for the UT-Osteo and BACH histopathological cancer datasets.

Dataset	**Class**	**Training Samples**	**Test Samples**
UT-Osteo (Bone)	Non-tumor	3,874	1,679
	Non-viable tumor	2,696	1,131
	Viable	3,447	1,495
	Total	**10,017**	**4,305**
BACH (Breast)	Normal	3,248	1,402
	Benign	3,324	1,428
	In Situ	3,278	1,402
	Invasive	3,339	1,427
	Total	**13,189**	**5,659**

Table 5: Hyperparameter and training configuration of the proposed ABD-TDA framework.

Parameter	**Configuration**
Backbone	ViT-Base/16
Pre-trained model	`google/vit-base-patch16-224`
ViT patch size	16×16
Input image size	224×224
ViT feature dimension	768
TDA input features	1,200
TDA descriptors	Betti-0 and Betti-1
Color representations	RGB and HSV
TDA branch	MLP
TDA MLP architecture	$1200 \to 256 \to 128 \to 64$
TDA activation	ReLU
TDA dropout	0.30
Fusion strategy	Feature concatenation
Fused feature dimension	$768 + 64 = 832$
Classifier architecture	$832 \to 256 \to 128 \to 2/3/4$
Classifier activation	ReLU
Classifier dropout	0.40, 0.30
Loss function	Cross-Entropy Loss
Optimizer	AdamW
Learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Batch size	16
Maximum epochs	50
Early stopping patience	25 epochs
Image scaling	Pixel values divided by 255
Data augmentation	None
Hardware	GPU when available



(a) ROC curve for Bone Cancer classification.

(b) ROC curve for Breast Cancer classification.

Figure 5: Receiver operating characteristic (ROC) curves of the proposed ABD-TDA framework on the UT-Osteosarcoma (Bone Cancer) and ICIAR2018 (Breast Cancer) datasets. The ROC curves illustrate the discriminative performance of the proposed model across the respective diagnostic classes, with the area under the ROC curve (AUC) providing an overall measure of classification discrimination.