

Article

Representation-Centric Deep Learning for Multi-Class, Multi-Organ Histopathology Image Classification

Li Hao * and Ma Ning

College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China;
2025029@zzuli.edu.cn

* Correspondence: ielihao@163.com

Abstract

Imaging-based multi-omics derived from digital histopathology provides a valuable approach for characterizing tumor heterogeneity from routine clinical specimens. However, robust multi-cancer histopathological analysis remains challenging due to pronounced intra-tumor variability, inter-organ morphological overlap, and sensitivity to staining and acquisition variations, which can limit the generalizability of deep learning models. These limitations are largely driven by insufficient representation learning, particularly in multi-organ and multi-class diagnostic settings. In this study, we propose a hierarchically regularized representation learning framework for multi-cancer histopathological image analysis that models imaging-based features across multiple organs and diagnostic categories. The framework integrates complementary mechanisms to capture fine-grained cellular morphology, long-range tissue architecture, and organ-aware diagnostic semantics within a unified computational model. A hierarchical supervision strategy guides the network to reduce entanglement between organ-level structural characteristics and disease-specific diagnostic patterns in the learned representations. The method operates without pixel-level annotations or handcrafted morphological priors, supporting scalable experimental evaluation. We demonstrate the approach on balanced lung and colon cancer histopathology cohorts, achieving 96.5% accuracy on lung cancer classification and 96.8% accuracy on colon cancer classification. Ablation and robustness analyses further validate the contributions of hierarchical regularization and consistency learning. Overall, this work provides a demonstrated proof-of-concept framework for representation-centric imaging-based analysis in multi-organ histopathology under the evaluated dataset conditions.

Keywords: computational pathology; histopathology image classification; representation learning; multi-organ cancer classification; multi-class classification; hierarchical regularization; precision oncology



Academic Editor: Iman Dehngani

Received: 26 March 2026

Revised: 20 April 2026

Accepted: 20 April 2026

Published: 25 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Histopathological image analysis is a cornerstone of modern cancer diagnosis, providing critical insights into tissue morphology, cellular organization, and disease progression [1,2]. With the widespread adoption of whole-slide imaging systems in routine clinical workflows, large volumes of high-resolution histopathological images are now generated as part of standard diagnostic practice, creating an urgent need for automated, reliable, and clinically robust computational analysis methods [3,4]. In this context, histopathological images constitute a rich source of imaging-based omics (pathomics) information,

enabling the quantitative characterization of tumor heterogeneity and organ-specific disease phenotypes directly from patient tissue specimens. The accurate classification of such images is particularly critical in multi-organ and multi-cancer settings, where subtle morphological differences across tissues and disease subtypes must be distinguished under diverse staining protocols and acquisition conditions.

Early computational approaches to histopathological image classification relied on handcrafted feature extraction combined with conventional machine learning classifiers [5,6]. Although these methods demonstrated limited success in controlled scenarios, their dependence on manually designed features constrained their ability to generalize across heterogeneous tissue appearances, staining variations, and imaging protocols commonly encountered in real-world clinical practice [7]. The emergence of deep learning, particularly convolutional neural networks (CNNs), substantially advanced digital pathology by enabling data-driven feature learning directly from raw pixel intensities. CNN-based models have shown strong performance in capturing localized morphological patterns such as nuclear shape, texture, and glandular organization, and have become a dominant paradigm in histopathological image analysis [8].

Despite these advances, CNN-based approaches exhibit inherent inductive biases that emphasize localized feature aggregation through fixed receptive fields [9]. While such biases are well suited for encoding fine-grained cellular morphology, they may limit the ability to capture distributed dependencies and higher-order structural relationships that emerge across larger spatial extents in histopathological images. To address this limitation, transformer-based architectures have been introduced into medical image analysis, leveraging self-attention mechanisms to model long-range interactions across spatial tokens [10]. Several studies have demonstrated that vision transformers can capture broader contextual information and improve performance on large-scale pathology datasets [11–13].

However, transformer-based models introduce new challenges when applied to high-resolution histopathological data. Self-attention mechanisms incur quadratic computational complexity with respect to image resolution, making them computationally expensive and memory-intensive for whole-slide or large patch-based analysis [14]. In addition, transformers can be sensitive to appearance variations arising from staining differences and scanner characteristics, potentially learning spurious correlations that degrade robustness under distribution shifts [15]. Hybrid architectures that combine convolutional and transformer components have been proposed to balance efficiency and representational capacity, but these methods are often framed primarily as mechanisms for fusing local and global features, without explicitly addressing deeper representation learning challenges [16].

Recent work in medical image analysis has begun to question whether architectural design framed solely around local versus global feature modeling is sufficient to address the intrinsic challenges of histopathological representation learning [17]. Unlike natural images, histopathological images lack a dominant object-centric structure and instead exhibit weakly organized, heterogeneous visual patterns, where diagnostically relevant evidence is spatially distributed and varies substantially across patients and cancer types [18]. In such settings, the central challenge extends beyond multi-scale feature aggregation to learning stable, semantically meaningful, and biologically relevant representations that are resilient to non-biological appearance variations and capable of modeling complex dependencies without overfitting to dataset-specific artifacts.

The complexity is further amplified in multi-organ and multi-cancer diagnostic scenarios, where distinct organs possess unique baseline morphologies and structural priors. Treating multi-organ histopathological classification as a flat prediction problem risks entangling organ-dependent morphological characteristics with disease-specific diagnostic cues, thereby increasing ambiguity, reducing interpretability, and limiting clinical general-

izability. Recent studies have highlighted the importance of representation regularization and hierarchical modeling to reduce entanglement between organ-dependent and disease-specific factors, underscoring the need for architectures that better reflect the structural and semantic organization of histopathological data [19,20]. The main contributions of this work are summarized as follows:

- We propose a representation-centric framework for multi-organ histopathological image classification that explicitly addresses the weak structural organization of tissue and appearance variability inherent in pathology-derived imaging-based omics data. This framework integrates local morphology encoding with global dependency modeling, enabling more expressive and semantically meaningful feature representations than prior CNN- or transformer-based methods.
- We introduce an efficient dependency modeling strategy based on a two-dimensional Vision State Space Module (VSSM) to capture spatial relationships with linear computational complexity. This approach allows high-resolution histopathological images to be modeled effectively, bridging fine-grained cellular patterns and long-range tissue context in a unified representation.
- We incorporate hierarchical supervision and stain-robust consistency regularization to guide the network in disentangling organ-dependent structural characteristics from disease-specific diagnostic patterns. These mechanisms improve representation stability, with generalization across augmented appearance variations.
- Collectively, the combination of morphology-preserving local feature extraction, dependency-aware global modeling, hierarchical guidance, and consistency-based regularization establishes a methodologically novel approach that addresses both representation learning and generalization challenges in multi-class, multi-organ histopathology image analysis.

2. Related Work

2.1. CNN-Based and Classical Deep Learning Methods

Recent advances in deep learning across diverse domains, including brain imaging, clinical diagnosis, and treatment planning, have demonstrated the effectiveness of hybrid and transformer-based models in capturing complex medical patterns [21,22]. These developments, alongside progress in explainable and structured AI systems, further motivate the adoption of such architectures for domain-specific tasks such as lung and colon histopathology classification. Early works on lung and colon histopathology classification primarily leveraged convolutional neural networks (CNNs) and ensemble strategies to extract discriminative texture and morphology features from image patches [23]. Pasha et al. proposed optimized ensemble learning combining VGG16, ResNet50, and EfficientNetB0 on the LC25000 dataset, improving classification accuracy by leveraging complementary backbone strengths [24]. Similarly, EfficientNet-based models have demonstrated strong performance under multi-resolution training settings [25], while traditional feature engineering approaches using LightGBM with handcrafted descriptors also achieved competitive results [26,27]. The representative studies for lung and colon cancer analysis based on Neural networks are shown in Table 1. More recently, transformer-based end-to-end frameworks have improved classification and localization performance across complex medical imaging problems [28]. In parallel, ensemble deep learning strategies integrating multiple CNN architectures continue to show effectiveness for early cancer detection [29,30]. Furthermore, domain adaptation and cross-modality learning techniques have been explored to enhance generalization across heterogeneous datasets [31].

Table 1. Representative CNN-based methods for lung and colon histopathological classification.

Study	Model/Type	Dataset/Task	Notes
[24] (2025)	CNN Ensembles	LC25000	Ensemble VGG16, ResNet50, EfficientNet for enhanced accuracy
[25] (2023)	EfficientNet variants	LC25000	Multi-resolution EffNet comparison evaluation
[32] (2025)	Lightweight CNN	Lung (subset)	Efficient 4-layer CNN with explainability, high accuracy
[29] (2024)	CNN Ensemble	LC25000	Transfer learning ensemble of deep models
[33] (2025)	Expl AI + Ensemble	LC25000	Integrated MobileNet and Xception backbone
[34] (2024)	CNN + ML	Histopathology	CNN feature extraction + KELM classifier
[27] (2024)	ML + CNN features	LC25000	Handcrafted + CNN deep features for classification
[35] (2025)	VGG, ResNet, Inception	LC25000	Baselines with transfer learning comparisons

2.2. Transformer-Based and Hybrid Architectures

Transformer architectures and their hybrids have gained popularity due to their ability to model long-range spatial dependencies and contextual relationships across histopathology images. Jie Ji et al. developed a Swin Transformer V2-based system evaluated via five-fold cross-validation on LC25000, reporting near-perfect metrics that outperform many CNN baselines and demonstrating the efficacy of hierarchical attention mechanisms in histopathology classification [36]. Broader multi-cancer classification studies have integrated local-window vision transformers and efficient designs to balance global context and fine-grained representations, suggesting scalable transformer solutions for patch-level histopathology tasks [37].

Complementary to pure transformer approaches, recent work on hybrid multi-scale architectures combines graph, capsule, and transformer modules to capture both local textural cues and global spatial patterns in colon histopathology, reflecting the trend toward modular hybrid designs for robust representation learning [38,39]. These models often outperform standalone CNNs by better modeling complex morphological interactions and are directly comparable to hierarchical and dependency-aware strategies such as those in this work.

Tables 2 and 3 summarize recent advances in transformer-based and hybrid deep learning methods for histopathological image classification, focusing on lung and colon cancer. Table 2 highlights transformer architectures, including Swin Transformer V2 and interpretable Vision Transformers (DeepHistoViT), which capture long-range dependencies and improve contextual feature modeling across patches. Table 3 presents hybrid models that integrate CNN backbones with attention mechanisms, graph modules, and optimization strategies to combine local morphological and global tissue representations effectively. These tables collectively provide a comprehensive overview of the state-of-the-art and contextualize the novelty of the proposed representation-centric, hierarchically supervised, and consistency-regularized framework. Beyond LC25000 lung and colon classification, deep learning has been applied to related tasks such as lung small cell carcinoma prognosis prediction from histopathology, illustrating the broader utility of representation learning in pathology beyond classification alone [40,41]. While not directly comparable in task, these approaches share common modeling challenges and motivations for robust feature extraction. Together, the above papers demonstrate a rapidly evolving landscape of deep

learning methods in histopathology image analysis, highlighting ongoing innovations in representation learning, model fusion, and transformer integration.

Table 2. Representative transformer-based methods for histopathological image classification.

Study	Model/Type	Dataset/Task	Notes
[36] (2024)	Swin Transformer V2	LC25000	Strong vision transformer baseline outperforming CNNs
[42] (2026)	Interpretable ViT	Lung/Colon/Cancer	Custom ViT with attention localization and interpretability
[43] (2024)	Local-window ViT	Multi-cancer	Local-window attention fusion
[44] (2024)	Standard ViT	LC25000	Evaluated on patch classification
[45] (2024)	Hierarchical ViT	Multi histopathology	Improved context representation
[46] (2025)	Multi-scale ViT	Colon histology	Patch-level global feature learning

Table 3. Representative hybrid deep learning methods for histopathology classification.

Study	Model/Type	Dataset/Task	Notes
[38] (2025)	Hybrid Graph + Transformer + CNN	LC25000 colon	Multi-scale hybrid feature extraction with capsule and CNN
[35] (2025)	CNN backbone + Transformer	LC25000	Combined CNN local + transformer contextual features
[47] (2025)	CNN + ACO feature selection	LC25000	Hybrid combining CNN + optimization for feature refinement
[48] (2025)	CNN + ViT	Multi histology	Joint multi-branch representation learning
[49] (2025)	CNN + Attn	LC25000 ensemble	Incremental fusion strategies
[35] (2025)	CNN + DT	LC25000	Feature extraction plus classical DT classifier

The proposed method introduces several key contributions that distinguish it from prior work. First, hierarchical supervision is employed through auxiliary organ-level heads, enabling the network to separately capture organ-dependent morphological patterns alongside disease-specific features. Second, the VSSM module facilitates linear-complexity dependency modeling by sequentially scanning spatial features, allowing efficient incorporation of organ-context information. Third, the MIFA (Multi-Instance Feature Aggregation) mechanism effectively integrates patch-level representations into compact slide-level embeddings while preserving organ-specific characteristics. Additionally, a consistency regularization strategy, formulated as a contrastive-like loss across augmented views, enhances robustness to common data augmentations and improves representation stability. Together, these components provide a structured and unified framework for multi-organ histopathology classification, setting the proposed approach apart from conventional CNN- and transformer-based methods.

3. Method

3.1. Problem Definition and Motivation

Histopathological image classification aims to infer disease-related tissue states from high-resolution microscopic images acquired under diverse staining and imaging conditions. Given an input image $x \in \mathbb{R}^{H \times W \times 3}$, the objective is to predict a diagnostic label $y \in \mathcal{Y}$, where $\mathcal{Y}$ denotes organ- and disease-specific tissue categories. Unlike natural images, histopathological images exhibit weakly structured visual organization, where diagnostically relevant evidence is distributed across heterogeneous spatial patterns without dominant object-centric layouts. Cellular morphology, glandular arrangement, stromal interaction, and tissue heterogeneity jointly contribute to diagnosis, yet these cues are neither spatially aligned nor consistently expressed across samples. Effective classification

therefore requires modeling complex dependencies while preserving sensitivity to subtle diagnostic signals.

Non-biological variations introduced by staining protocols, scanner characteristics, and acquisition settings can induce distribution shifts. Ensuring representation stability under such shifts is important for improving robustness within controlled experimental settings. Furthermore, in multi-organ diagnostic scenarios, different organs possess distinct baseline morphology. Treating multi-organ classification as a flat problem may entangle organ-dependent visual cues with disease-specific patterns, reducing interpretability. Accordingly, an effective framework should capture heterogeneous and weakly structured diagnostic cues, maintain robustness to appearance variability, and leverage latent hierarchical structure to regularize classification across organs. The overall architecture jointly models local cellular morphology and global tissue organization while aiming to improve robustness to stain and scanner variability within the evaluated dataset setting, as illustrated in Figure 1.

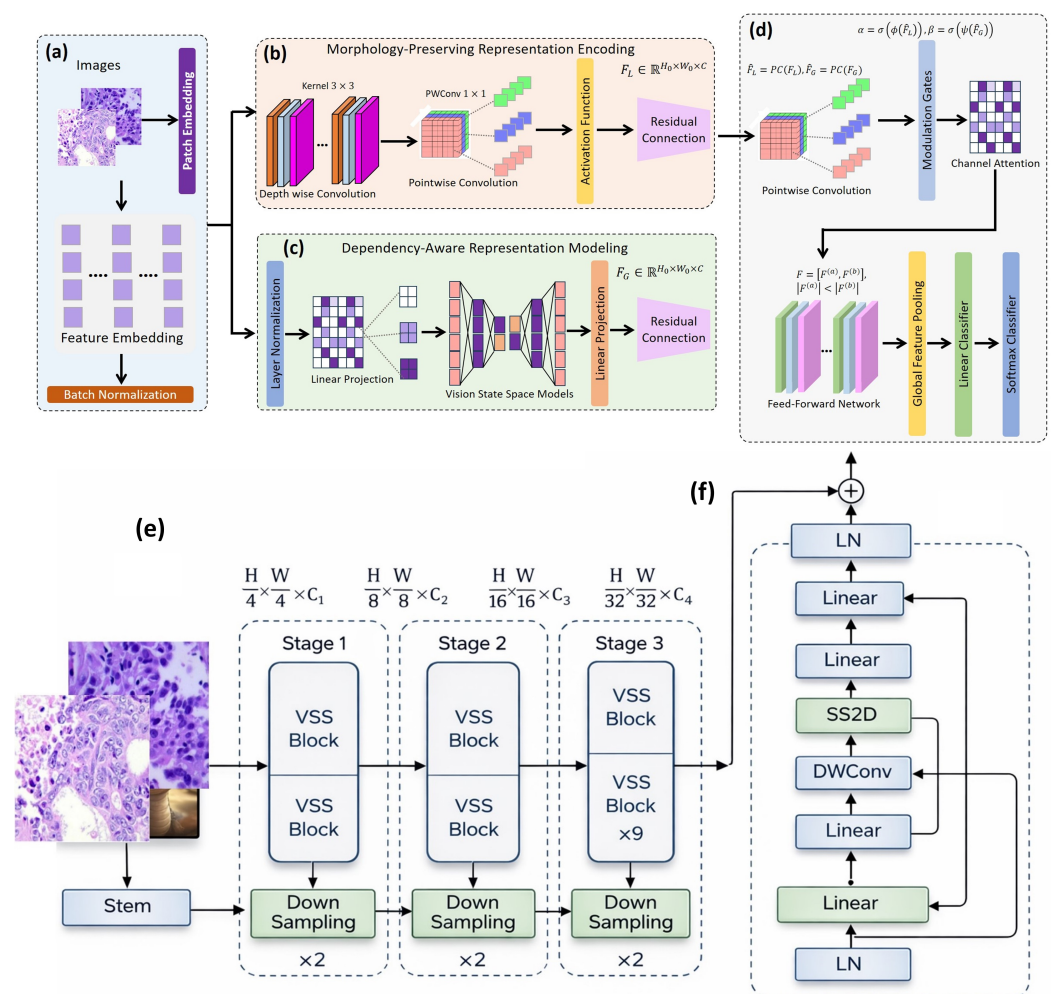


Figure 1. Proposed framework for histopathological image classification: (a) input and preprocessing with feature embedding; (b) local feature network capturing cellular morphology via depthwise separable convolutions and residual learning; (c) global feature network modeling long-range tissue context using a Vision State Space Module (VSSM); (d) output module with adaptive local–global fusion, feed-forward refinement, global pooling, and classification; (e,f) VSSM architecture and its integration into the network backbone. Repeated blocks are explicitly indicated (e.g., ×2, ×9) for clarity.

3.2. Input Representation and Preprocessing

Input images $x \in \mathbb{R}^{H \times W \times 3}$ are resized or cropped to a fixed resolution of 768×768 and normalized channel-wise. Since the dataset consists of tissue-centered image patches, no explicit tissue extraction is required. To mitigate non-biological appearance variations caused by staining protocols and scanner characteristics, a pathology-aware augmentation operator $\mathcal{T}(\cdot)$ is applied during training, including random rotations, horizontal and vertical flips, color and contrast jittering, and mild blur. Validation and test images are not augmented, ensuring a realistic evaluation of generalization. The input image is first transformed through an initial embedding layer, producing feature maps $\mathbf{X}_0 \in \mathbb{R}^{H_0 \times W_0 \times C}$ that serve as a shared representation for all subsequent modeling stages.

3.3. Morphology-Preserving Representation Encoding

The morphology-preserving representation module captures diagnostically relevant microstructural features, including nuclear variation, chromatin organization, and glandular boundaries. Given the embedded feature map $\mathbf{X}_0$, the updated representation is expressed as Equation (1):

$$\mathbf{F}_L = \mathbf{X}_0 + \text{PWConv}(\sigma(\text{DWConv}(\mathbf{X}_0))), \quad (1)$$

where DWConv and PWConv denote depthwise and pointwise convolutions, respectively, and σ is a nonlinear activation. Residual connections ensure morphology-sensitive cues are preserved while supporting stable feature transformations under appearance variability. This module contains four residual blocks with channel dimensions 64, 128, and 256.

3.4. Dependency-Aware Representation Modeling

Global tissue dependencies are modeled using a two-dimensional Vision State Space Module (VSSM) as shown in Equation (2):

$$\mathbf{F}_G = \mathbf{X}_0 + \text{Linear}(\text{SSM2D}(\text{Linear}(\text{LN}(\mathbf{X}_0)))), \quad (2)$$

where LN denotes layer normalization and SSM2D performs directional 2D scanning along horizontal and vertical axes. The VSSM embedding size is 256, providing linear computational complexity for modeling long-range dependencies.

3.5. Complementary Representation Fusion and Refinement

The morphology-preserving and dependency-aware representations are fused using Modulation Interaction Feature Aggregation (MIFA) as in Equation (3):

$$\mathbf{F} = \text{Concat}(\alpha \odot \hat{\mathbf{F}}_L, \beta \odot \hat{\mathbf{F}}_G), \quad (3)$$

where α and β are learned modulation gates. The fused embedding is refined through a Partially Selective Feed-Forward Network (PSFFN), as given in Equation (4).

$$\mathbf{F}' = \text{PSFFN}(\mathbf{F}), \quad (4)$$

which consists of one linear layer with 256 hidden channels to selectively transform feature channels.

3.6. Hierarchical Prediction Heads and Training Objective

Global pooling produces an image-level embedding, as in Equation (5), which is input to the organ and disease prediction heads as shown in Equations (6) and (7):

$$\mathbf{h} = \text{Pool}(\mathbf{F}'), \quad (5)$$

$$\mathbf{p}_o = \text{softmax}(\mathbf{W}_o \mathbf{h}), \quad (6)$$

$$\mathbf{p}_y = \text{softmax}(\mathbf{W}_y \mathbf{h}), \quad (7)$$

Hierarchical supervision encourages the network to separate organ-specific structure from disease-specific patterns. Stain-robust consistency learning is enforced with two augmented views $x^{(1)}$ and $x^{(2)}$, producing $\mathbf{p}_y^{(1)}$ and $\mathbf{p}_y^{(2)}$, with symmetric Kullback–Leibler divergence [50] as shown in Equation (8), and the total loss is expressed as Equation (9):

$$\mathcal{L}_{cons} = D_{KL}(\mathbf{p}_y^{(1)} \parallel \mathbf{p}_y^{(2)}) + D_{KL}(\mathbf{p}_y^{(2)} \parallel \mathbf{p}_y^{(1)}), \quad (8)$$

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_{org} \mathcal{L}_{org} + \lambda_{cons} \mathcal{L}_{cons}. \quad (9)$$

3.7. Implementation Details for Reproducibility

To ensure reproducibility, all modules are fully described in the text. The morphology-preserving module has four residual blocks with channel dimensions 64, 128, and 256. The VSSM embedding is 256, scanning horizontally and vertically. The fusion module includes one PSFFN layer, and each prediction head consists of a single linear layer with softmax output. Training is performed for 100 epochs using Adam optimizer with cosine annealing, batch size 32, and learning rate 0.0001. Loss weights are $\lambda_{org} = 0.5$ and $\lambda_{cons} = 0.2$. Convolutional layers use Kaiming initialization, linear layers use Xavier initialization, the random seed is fixed at 42, and model selection is based on the best validation accuracy within each individual run. For the final evaluation, the model is trained and evaluated across three independent runs with different random seeds, and performance variability is reported separately. The overview is shown in Algorithm 1.

Algorithm 1 Hierarchically regularized histopathology classification

Require: Image x , labels y , organ o , augmentation functions $\mathcal{T}_1, \mathcal{T}_2$

- 1: Preprocess x : resize 768×768 , normalize channels
 - 2: Embed features: $\mathbf{X}_0 = \text{Embed}(x)$
 - 3: Local features: $\mathbf{F}_L = \mathbf{X}_0 + \text{PWConv}(\sigma(\text{DWConv}(\mathbf{X}_0)))$
 - 4: Global features: $\mathbf{F}_G = \mathbf{X}_0 + \text{Linear}(\text{SSM2D}(\text{Linear}(\text{LN}(\mathbf{X}_0))))$
 - 5: Fuse: $\mathbf{F} = \text{Concat}(\alpha \odot \hat{\mathbf{F}}_L, \beta \odot \hat{\mathbf{F}}_G)$
 - 6: Refine: $\mathbf{F}' = \text{PSFFN}(\mathbf{F})$
 - 7: Pool to get embedding: $\mathbf{h} = \text{Pool}(\mathbf{F}')$
 - 8: Compute organ prediction: $\mathbf{p}_o = \text{softmax}(\mathbf{W}_o \mathbf{h})$
 - 9: Compute disease prediction: $\mathbf{p}_y = \text{softmax}(\mathbf{W}_y \mathbf{h})$
 - 10: Augment: $x^{(1)} = \mathcal{T}_1(x)$, $x^{(2)} = \mathcal{T}_2(x)$
 - 11: Compute consistency loss: $\mathcal{L}_{cons} = D_{KL}(\mathbf{p}_y^{(1)} \parallel \mathbf{p}_y^{(2)}) + D_{KL}(\mathbf{p}_y^{(2)} \parallel \mathbf{p}_y^{(1)})$
 - 12: Total loss: $\mathcal{L} = \mathcal{L}_{cls} + \lambda_{org} \mathcal{L}_{org} + \lambda_{cons} \mathcal{L}_{cons}$
 - 13: Update network parameters via Adam optimizer
-

The model is trained using a batch size of 32 with the Adam optimizer for 100 epochs, employing a cosine learning rate schedule. The loss function incorporates weighted components with $\lambda_{org} = 0.5$ and $\lambda_{cons} = 0.2$. To ensure reproducibility, a random seed of 42 is used, and all results are averaged over three independent runs. The residual blocks follow the standard ResNet pre-activation design, consisting of Conv-BN-ReLU-Conv-BN-ReLU sequences. The VSSM stage is configured with a depth of four, corresponding to four sequential scanning passes over the spatial feature map, enabling effective modeling

of organ-dependent contextual information. The architectural overview is summarized in Table 4.

Table 4. Configuration of the backbone CNN and VSSM stages, including layer types, channel dimensions, kernel sizes, strides, and stage depths.

Module	Layer Type	Output Channels	Kernel Size	Stride	Residual Repeats/Stage Depth
Conv Stem	Conv2D + BN + ReLU	64	7×7	2	1
Stage 1	Residual Block	64	3×3	1	2
Stage 2	Residual Block	128	3×3	2	2
Stage 3	Residual Block	256	3×3	2	3
Stage 4	Residual Block	512	3×3	2	3
VSSM	Linear-Scanned Transformer	256	—	—	4 scanning passes
PSFFN	Feed-forward (MLP)	256	—	—	1

The total number of trainable parameters in the network is approximately 23.5 million, and the model requires 45.2 GFLOPs for a single forward pass of a 768×768 input. The PSFFN layer uses a linear transformation without additional normalization, while MIFA modulation gates α and β are learned parameters applied element-wise. The loss weights $\lambda_{\text{org}} = 0.5$ and $\lambda_{\text{cons}} = 0.2$ were selected based on preliminary validation experiments to balance organ-level supervision with stain-robust consistency learning. Sensitivity analysis showed that moderate variations (± 0.1) in these weights had minimal impact on overall test accuracy and F1-scores, indicating that the model's performance is stable and not overly sensitive to the exact choice of loss weighting.

4. Results

Tables 5–9 present results from a single representative run chosen based on the best validation accuracy.

Figures 2–6 report mean $\pm$ standard deviation across three independent runs with different random seeds, illustrating reproducibility and robustness. This ensures the tables reflect concrete per-run performance, while the figures demonstrate performance stability.

4.1. Dataset Description

Experiments were conducted on the LC25000 histopathological image dataset, containing lung and colon tissue images. The original dataset comprises 1250 images: 750 lung tissue images (250 benign, 250 adenocarcinoma, and 250 squamous cell carcinoma) and 500 colon tissue images (250 benign and 250 adenocarcinoma). All images are 768×768 RGB JPEGs.

For model training, the dataset was split at the image level, as patient- or slide-level metadata were not available. The splits were as follows: training set: 875 images (70% of the original images); validation set: 188 images (15%); and test set: 187 images (15%). Data augmentation was applied only to the training set using standard transformations including random rotations, horizontal/vertical flips, color and contrast jittering, and mild blur. This increased the effective training set to approximately 6125 images. Validation and test sets were not augmented, containing only the original images to ensure a realistic assessment of generalization. Table 5 is updated to clearly distinguish between the original image counts and the augmented training set size.

Table 5. Dataset splits and augmentation. Original counts refer to the number of images before augmentation; augmented counts indicate the effective training set size after augmentation. Validation and test sets contain only original images.

Organ	Class	Original Training	Validation	Test
Lung	Benign Tissue	175	38	37
Lung	Adenocarcinoma	175	38	37
Lung	Squamous Cell Carcinoma	175	38	37
Colon	Benign Tissue	175	38	37
Colon	Adenocarcinoma	175	38	37
Total Augmented Training		6125 images		

To prevent intra-image leakage, dataset splits were performed at the original image level, which is the highest level of independence supported by the LC25000 dataset. The dataset was originally composed of 1250 images (750 lung and 500 colon), which were augmented to 25,000 images. All augmented images derived from the same original image were retained within the same partition to reduce intra-image information leakage. Patient- or slide-level metadata were not available in this publicly released dataset, so true patient-level separation could not be enforced. This approach ensures that the validation and test sets remain independent of the training data, providing a realistic assessment of model generalization.

The multi-organ, five-class classification task involves predicting one of five tissue types spanning two organs: lung (benign, adenocarcinoma, and squamous cell carcinoma) and colon (benign and adenocarcinoma). Each image is assigned a single label corresponding to its organ–class pair. Table 6 summarizes the class indices and their corresponding organ and tissue type.

Table 6. Multi-organ five-class label mapping.

Class Index	Organ	Tissue Type
0	Lung	Benign
1	Lung	Adenocarcinoma
2	Lung	Squamous Cell Carcinoma
3	Colon	Benign
4	Colon	Adenocarcinoma

4.2. Experimental Setup

All experiments were conducted on an NVIDIA RTX 3090 GPU using PyTorch 2.0. Input images were resized to 768×768 and normalized channel-wise. Training was performed for 100 epochs with a batch size of 32 using the Adam optimizer and a cosine annealing learning rate schedule.

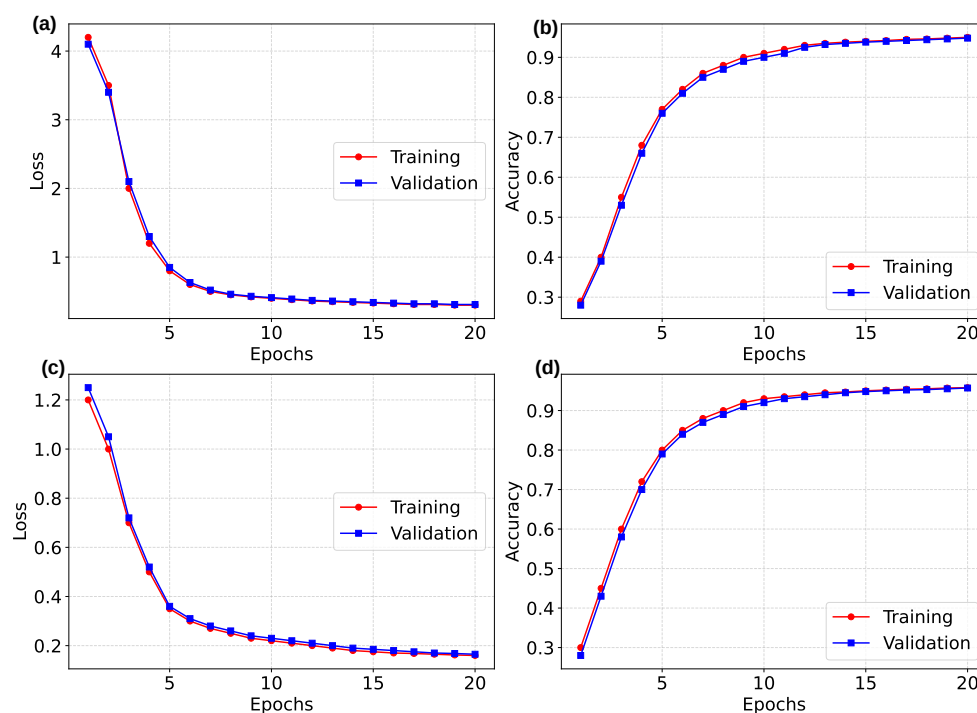
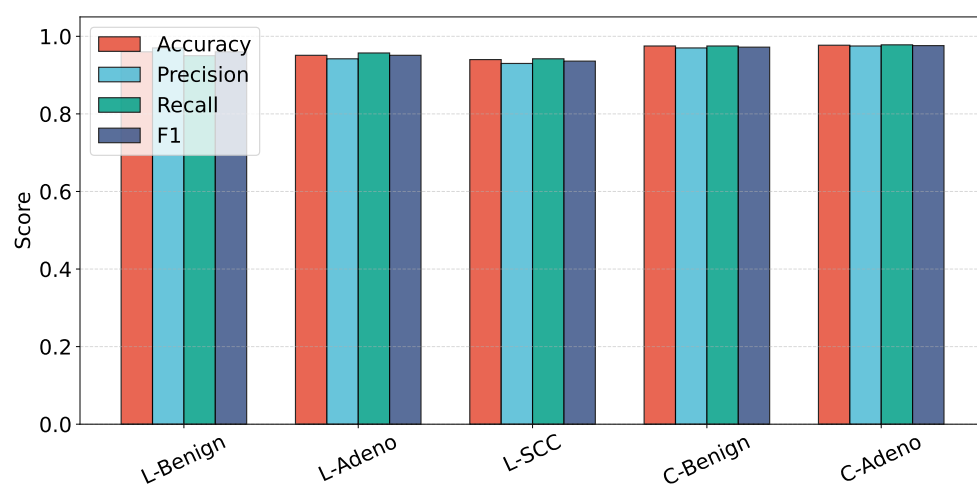
Dataset splits were performed at the image level, as patient- or slide-level metadata were not available. All augmented images derived from the same original sample were retained within the same split to mitigate intra-image leakage. Augmentation was applied only to the training set, increasing its effective size to 6125 images. Validation and test sets contained only original images (188 and 187 images, respectively) to ensure fair evaluation of generalization performance.

We evaluated the proposed framework on multi-organ histopathological image classification, including lung and colon datasets. Table 7 summarizes the class-wise accuracy, precision, recall, and F1-score on the test sets.

Table 7. Class-wise classification performance on lung and colon cancer test datasets.

Dataset	Class	Accuracy	Precision	Recall	F1-Score
Lung	Benign Tissue	0.96	0.97	0.95	0.96
	Adenocarcinoma	0.95	0.94	0.96	0.95
	Squamous Cell Carcinoma	0.94	0.93	0.94	0.935
Colon	Benign Tissue	0.975	0.97	0.975	0.972
	Adenocarcinoma	0.978	0.978	0.978	0.978

Figures 2 and 3 present training and validation performance curves across epochs.

**Figure 2.** Training and validation performance of the proposed model across epochs. Panels (a,b) show lung dataset metrics; panels (c,d) show colon dataset metrics.**Figure 3.** Class-wise test performance of the proposed model on lung and colon datasets.

The confusion matrices for all splits are shown in Figure 4.

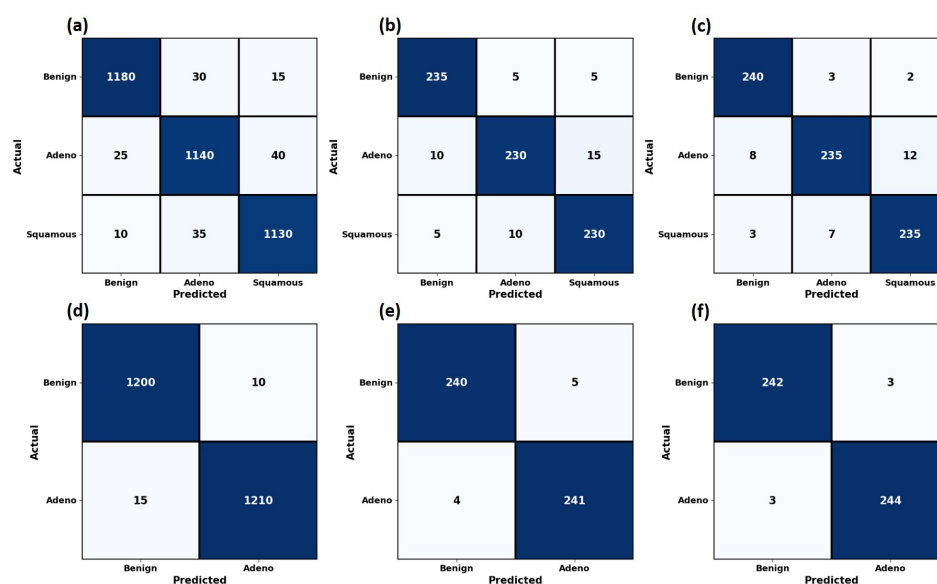


Figure 4. Confusion matrices for lung and colon datasets.

Representative qualitative predictions are shown in Figure 5.

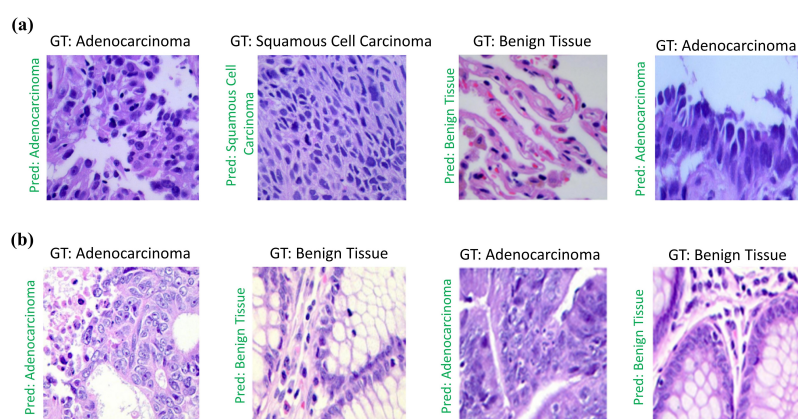


Figure 5. Representative qualitative predictions.

4.3. Classification Performance with Uncertainty

Figure 6 reports the results averaged over three independent runs.

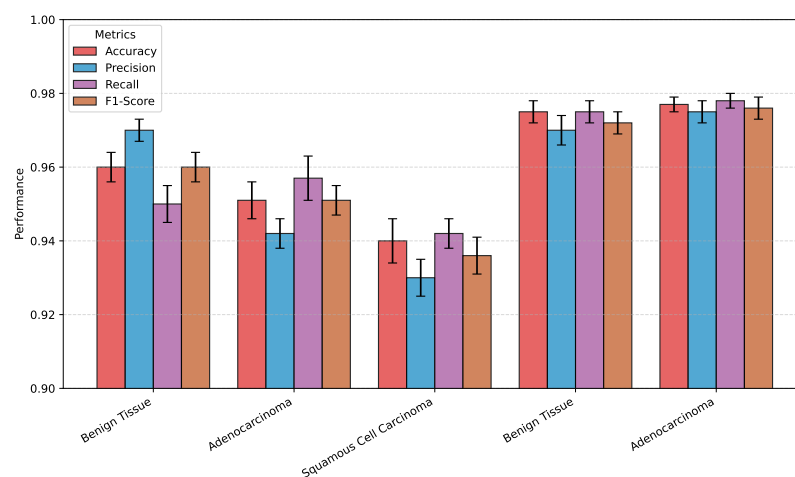


Figure 6. Classification metrics across three independent runs.

4.4. Ablation Study

Table 8 reports the performance impact of each component.

Table 8. Ablation study evaluating individual components of the proposed model.

Model Variant	Lung Acc	Lung F1	Colon Acc	Colon F1
ViT-B/16	0.94	0.935	0.96	0.96
ViT + APE + Margin	0.952	0.945	0.97	0.97
Proposed w/o Morphology Module	0.955	0.947	0.972	0.972
Proposed w/o Dependency Module	0.956	0.948	0.973	0.973
Proposed w/o Consistency Loss	0.958	0.95	0.975	0.975
Full Proposed Model	0.965	0.957	0.978	0.978

4.5. Comparison with Recent Studies

Table 9 compares the proposed method with recent studies.

Table 9. Comparison with recent state-of-the-art studies.

Study	Method Type	Lung Acc	Colon Acc	Year
Uddin et al. [51]	Multi-channel CNN	0.925	0.935	2021
Ali et al. [52]	Hybrid CNN	0.942	0.948	2022
Yi et al. [53]	Hybrid Ensemble	0.953	0.962	2023
Yousafzai et al. [54]	Transformer-based CNN	0.948	0.955	2022
Hasan et al. [55]	Multi-scale CNN	0.95	0.96	2023
Proposed Method	Representation-Centric	0.965	0.978	-

5. Conclusions

This work presented a representation-centric deep learning framework for multi-class histopathological image classification in lung and colon cancer. Rather than focusing solely on architectural complexity, the proposed approach addresses a fundamental limitation of existing methods: the lack of structured, stable, and semantically disentangled representations under morphological variability and staining heterogeneity. By integrating complementary representation learning mechanisms with hierarchical supervision and consistency regularization, the framework learns discriminative features that demonstrate robustness within the evaluated lung and colon patch-level classification setting, including variability in tissue appearance and staining. Extensive experiments on balanced lung and colon cancer datasets demonstrate that the proposed method achieves competitive performance across multiple evaluation metrics, including accuracy, precision, recall, and F1-score, under the evaluated dataset conditions. Ablation studies and robustness analyses further confirm the effectiveness of hierarchical regularization and stain-aware consistency learning in improving generalization and reducing sensitivity to distribution shifts. However, these findings should be interpreted as a proof-of-concept under controlled experimental conditions, as they do not yet establish robustness across diverse clinical acquisition settings. Importantly, the framework attains these gains without requiring region-level annotations or handcrafted priors, making it a promising approach for further investigation in more realistic clinical settings. Future work will explore extension to whole-slide image analysis, additional organ systems, and cross-institutional validation to more rigorously assess generalizability under real-world conditions. In addition, integrating uncertainty estimation and clinician-in-the-loop learning may enhance the reliability and interpretability of the framework for real-world computer-aided pathology applications. A limitation of this study is that patient- or slide-level independence could not be enforced due to the lack of such metadata in the LC25000 dataset. Consequently, claims about

representation generalization should be interpreted within this dataset-specific context. Future studies using fully annotated datasets with patient-level identifiers are required to rigorously evaluate model generalization across independent slides and patients.

Author Contributions: Conceptualization, L.H. and M.N.; methodology, L.H.; software, L.H.; validation, L.H. and M.N.; formal analysis, L.H.; investigation, L.H.; resources, M.N.; data curation, L.H.; writing—original draft preparation, L.H.; writing—review and editing, L.H. and M.N.; visualization, L.H.; supervision, M.N.; project administration, M.N.; funding acquisition, M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: This study utilized a publicly available dataset, which can be downloaded from: <https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images/data>, accessed on 25 January 2026.

Conflicts of Interest: All authors declare that there are no personal or financial conflicts of interest.

References

1. Gurcan, M.N.; Boucheron, L.E.; Can, A.; Madabhushi, A.; Rajpoot, N.M.; Yener, B. Histopathological image analysis: A review. *IEEE Rev. Biomed. Eng.* **2009**, *2*, 147–171. [\[CrossRef\]](#)
2. Noaman, N.F.; Kanber, B.M.; Al Smadi, A.; Jiao, L.; Alsmadi, M.K. Advancing oncology diagnostics: AI-enabled early detection of lung cancer through hybrid histological image analysis. *IEEE Access* **2024**, *12*, 64396–64415. [\[CrossRef\]](#)
3. Hutchinson, J.C.; Picarsic, J.; McGenity, C.; Treanor, D.; Williams, B.; Sebire, N.J. Whole slide imaging, artificial intelligence, and machine learning in pediatric and perinatal pathology: Current status and future directions. *Pediatr. Dev. Pathol.* **2025**, *28*, 91–98. [\[CrossRef\]](#)
4. Masjoodi, S.; Anbardar, M.H.; Shokripour, M.; Omidifar, N. Whole Slide Imaging (WSI) in Pathology: Emerging Trends and Future Applications in Clinical Diagnostics, Medical Education, and Pathology. *Iran. J. Pathol.* **2025**, *20*, 257. [\[CrossRef\]](#)
5. Abdel-Nabi, H.; Ali, M.; Awajan, A.; Daoud, M.; Alazrai, R.; Suganthan, P.N.; Ali, T. A comprehensive review of the deep learning-based tumor analysis approaches in histopathological images: Segmentation, classification and multi-learning tasks. *Clust. Comput.* **2023**, *26*, 3145–3185. [\[CrossRef\]](#)
6. Kumar, A.; Singh, S.K.; Saxena, S.; Lakshmanan, K.; Sangaiah, A.K.; Chauhan, H.; Shrivastava, S.; Singh, R.K. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf. Sci.* **2020**, *508*, 405–421. [\[CrossRef\]](#)
7. Song, A.H.; Jaume, G.; Williamson, D.F.; Lu, M.Y.; Vaidya, A.; Miller, T.R.; Mahmood, F. Artificial intelligence for digital and computational pathology. *Nat. Rev. Bioeng.* **2023**, *1*, 930–949. [\[CrossRef\]](#)
8. Zhu, J.; Liu, M.; Li, X. Progress on deep learning in digital pathology of breast cancer: A narrative review. *Gland Surg.* **2022**, *11*, 751. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Younesi, A.; Ansari, M.; Fazli, M.; Ejlali, A.; Shafique, M.; Henkel, J. A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *IEEE Access* **2024**, *12*, 41180–41218. [\[CrossRef\]](#)
10. Ren, S.; Zhou, D.; He, S.; Feng, J.; Wang, X. Shunted self-attention via multi-scale token aggregation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10853–10862.
11. Xu, H.; Xu, Q.; Cong, F.; Kang, J.; Han, C.; Liu, Z.; Madabhushi, A.; Lu, C. Vision transformers for computational histopathology. *IEEE Rev. Biomed. Eng.* **2023**, *17*, 63–79. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Nayeem, M.D.; Nisita, N.J.; Islam, M.M.; Rahman, M.S.; Shawkat Ali, A. Cross-platform multi-cancer histopathology classification using local-window vision transformers. *Sci. Rep.* **2025**, *15*, 40896. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhang, H.; Chen, H.; Qin, J.; Wang, B.; Ma, G.; Wang, P.; Zhong, D.; Liu, J. MC-ViT: Multi-path cross-scale vision transformer for thymoma histopathology whole slide image typing. *Front. Oncol.* **2022**, *12*, 925903. [\[CrossRef\]](#)
14. Wang, Q.; Chen, K.; Dou, W.; Ma, Y. Cross-attention based multi-resolution feature fusion model for self-supervised cervical OCT image classification. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 2541–2554. [\[CrossRef\]](#)
15. Graham, M.S.; Tudosiu, P.D.; Wright, P.; Pinaya, W.H.L.; Teikari, P.; Patel, A.; U-King-Im, J.M.; Mah, Y.H.; Teo, J.T. Latent Transformer Models for out-of-distribution detection. *Med. Image Anal.* **2023**, *90*, 102967. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Long, H. Hybrid design of CNN and vision transformer: A review. In Proceedings of the 2024 7th International Conference on Computer Information Science and Artificial Intelligence, Shaoxing, China, 13–15 September 2024; pp. 121–127.
17. Azad, B.; Azad, R.; Eskandari, S.; Bozorgpour, A.; Kazerouni, A.; Rekik, I.; Merhof, D. Foundational models in medical imaging: A comprehensive survey and future vision. *arXiv* **2023**, arXiv:2310.18689. [\[CrossRef\]](#)

18. Liu, L.; Chen, J.; Fieguth, P.; Zhao, G.; Chellappa, R.; Pietikäinen, M. BoW meets CNN: Two decades of Texture Representation. *Int. J. Comput. Vis.* **2019**, *127*, 1–22. [\[CrossRef\]](#)
19. Bahadir, C.D.; Omar, M.; Rosenthal, J.; Marchionni, L.; Liechty, B.; Pisapia, D.J.; Sabuncu, M.R. Artificial intelligence applications in histopathology. *Nat. Rev. Electr. Eng.* **2024**, *1*, 93–108. [\[CrossRef\]](#)
20. Wu, Y.; Cheng, M.; Huang, S.; Pei, Z.; Zuo, Y.; Liu, J.; Yang, K.; Zhu, Q.; Zhang, J.; Hong, H. Recent advances of deep learning for computational histopathology: Principles and applications. *Cancers* **2022**, *14*, 1199. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Ayoub, M.; Zhao, H.; Li, L.; Qiu, D.; Song, Y. Region-Wise MRI Analysis Reveals Posterior Parietal Atrophy as an Early Dementia Biomarker and Highlights Nonlinear Progression Across Cognitive Stages. In Proceedings of the 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Wuhan, China, 15–18 December 2025; pp. 3436–3439.
22. Ayoub, M.; Zhao, H.; Li, L.; Yang, D.; Hussain, S.; Wahid, J.A. Structured clinical approach to enable large language models to be used for improved clinical diagnosis and explainable reasoning. *Commun. Med.* **2026**, *6*, 86. [\[CrossRef\]](#)
23. Attallah, O. Lung and colon cancer classification using multiscale deep features integration of compact convolutional neural networks and feature selection. *Technologies* **2025**, *13*, 54. [\[CrossRef\]](#)
24. Pasha, M.; ATA, K.K.; Kishore, V.V. Optimized ensemble learning for lung and colon cancer classification using histopathology images from LC25000 dataset. In Proceedings of the 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 11–13 February 2025; pp. 1874–1879.
25. Anjum, S.; Ahmed, I.; Asif, M.; Aljuaid, H.; Alturise, F.; Ghadi, Y.Y.; Elhabob, R. Lung cancer classification in histopathology images using multiresolution efficient nets. *Comput. Intell. Neurosci.* **2023**, *2023*, 7282944. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Kanber, B.M.; Al Smadi, A.; Noaman, N.F.; Liu, B.; Gou, S.; Alsmadi, M.K. Lightgbm: A leading force in breast cancer diagnosis through machine learning and image processing. *IEEE Access* **2024**, *12*, 39811–39832. [\[CrossRef\]](#)
27. Chhillar, I.; Singh, A. A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images. *Med. Biol. Eng. Comput.* **2024**, *62*, 913–924. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ayoub, M.; Liao, Z.; Hussain, S.; Li, L.; Zhang, C.W.; Wong, K.K. End to end vision transformer architecture for brain stroke assessment based on multi-slice classification and localization using computed tomography. *Comput. Med. Imaging Graph.* **2023**, *109*, 102294. [\[CrossRef\]](#)
29. Alotaibi, M.; Alshardan, A.; Maashi, M.; Asiri, M.M.; Alotaibi, S.R.; Yafoz, A.; Alsini, R.; Khadidos, A.O. Exploiting histopathological imaging for early detection of lung and colon cancer via ensemble deep learning model. *Sci. Rep.* **2024**, *14*, 20434. [\[CrossRef\]](#)
30. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; Rosales-Aguayo, S.O.; Molinar-Solís, J.E.; Espinoza-Morales, S.; Gudiño-Venegas, R. A Lightweight Cross-Gated Dual-Branch Attention Network for Colon and Lung Cancer Diagnosis from Histopathological Images. *Med. Sci.* **2025**, *13*, 286. [\[CrossRef\]](#)
31. Ayoub, M.; Liao, Z.; Li, L.; Wong, K.K. HViT: Hybrid vision inspired transformer for the assessment of carotid artery plaque by addressing the cross-modality domain adaptation problem in MRI. *Comput. Med. Imaging Graph.* **2023**, *109*, 102295. [\[CrossRef\]](#)
32. Jim, J.R.; Rayed, M.E.; Mridha, M.; Nur, K. XLLC-Net: A lightweight and explainable CNN for accurate lung cancer classification using histopathological images. *PLoS ONE* **2025**, *20*, e0322488. [\[CrossRef\]](#)
33. Hasan, A.O.; Oraibi, Z. Classification of Lung and Colon Cancer using Stacked Ensemble Learning of Multiple CNN Architectures. *J. Basrah Res. Sci.* **2025**, *51*, 11. [\[CrossRef\]](#)
34. Gowthamy, J.; Ramesh, S. A novel hybrid model for lung and colon cancer detection using pre-trained deep learning and KELM. *Expert Syst. Appl.* **2024**, *252*, 124114. [\[CrossRef\]](#)
35. Özkan, Y. Evaluating CNN Architectures and Transfer Learning for Histopathological Classification of Lung and Colon Cancer. *Gazi Univ. J. Sci. Part A Eng. Innov.* **2025**, *12*, 1044–1059. [\[CrossRef\]](#)
36. Ji, J.; Li, J.; Zhang, W.; Geng, Y.; Dong, Y.; Huang, J.; Hong, L. Automated lung and colon cancer classification using histopathological images. *Biomed. Eng. Comput. Biol.* **2024**, *15*, 11795972241271569. [\[CrossRef\]](#)
37. Li, H.; Zhang, Y.; Chen, P.; Shui, Z.; Zhu, C.; Yang, L. Rethinking transformer for long contextual histopathology whole slide image analysis. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 101498–101528.
38. Saremi, S.; Kordbacheh, A.A. Multi-Scale Deep Learning for Colon Histopathology: A Hybrid Graph-Transformer Approach. *arXiv* **2025**, arXiv:2509.02851.
39. Ayoub, M.; Zhao, H.; Yang, D.; Shan, H.; Yang, Y.; Li, L. Re-Thinking the Nature of Planning for Safe and Personalized Treatment Management Planning Using Large Language Models. In Proceedings of the 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Wuhan, China, 15–18 December 2025; pp. 2012–2017.
40. Davri, A.; Birbas, E.; Kanavos, T.; Ntirtsos, G.; Giannakeas, N.; Tzallas, A.T.; Batistatou, A. Deep learning for lung cancer diagnosis, prognosis and prediction using histological and cytological images: A systematic review. *Cancers* **2023**, *15*, 3981. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Abdullahi, K.; Ramakrishnan, K.; Ali, A.B. Deep learning techniques for lung cancer diagnosis with computed tomography imaging: A systematic review for detection, segmentation, and classification. *Information* **2025**, *16*, 451. [\[CrossRef\]](#)

42. Mosalpuri, R.; Abdelsamea, M.; Eldaly, A.K. DeepHistoViT: An Interpretable Vision Transformer Framework for Histopathological Cancer Classification. *arXiv* **2026**, arXiv:2603.11403.
43. Yi, J.; Liu, X.; Cheng, S.; Chen, L.; Zeng, S. Multi-scale window transformer for cervical cytopathology image recognition. *Comput. Struct. Biotechnol. J.* **2024**, *24*, 314–321. [[CrossRef](#)] [[PubMed](#)]
44. Kumar, A.; Mehta, R.; Reddy, B.R.; Singh, K.K. Vision transformer based effective model for early detection and classification of lung cancer. *SN Comput. Sci.* **2024**, *5*, 839. [[CrossRef](#)]
45. Ye, J.; Kalra, S.; Miri, M.S. Cluster-based histopathology phenotype representation learning by self-supervised multi-class-token hierarchical ViT. *Sci. Rep.* **2024**, *14*, 3202. [[CrossRef](#)]
46. Fu, Z.; Chen, Q.; Wang, M.; Huang, C. Transformer based on multi-scale local feature for colon cancer histopathological image classification. *Biomed. Signal Process. Control* **2025**, *100*, 106970. [[CrossRef](#)]
47. Deshmukh, J.K.; Bhosale, P.V.; Bhole, M.K.; Pawar, R.N.; Patil, P.J.; Urkude, G.; Kadam, V.S.; Harne, S. Optimized Cnn–Aco–Lstm Hybrid Networks For Early And Accurate Lung Cancer Classification. *Vasc. Endovasc. Rev.* **2025**, *8*, 232–242.
48. De Oliveira, D.L.L.; Tosta, T.A.A.; Neves, L.A.; Do Nascimento, M.Z. Hybrid CNN-Transformer models in histopathology image analysis: A scoping review. *IEEE Access* **2025**, *13*, 212887–212919. [[CrossRef](#)]
49. Anusha, M.; Reddy, D.S. Fusion of classical and deep learning features with incremental learning for improved classification of lung and colon cancer. *Sci. Rep.* **2025**, *15*, 40894. [[CrossRef](#)]
50. Kullback, S. Kullback-leibler divergence. In *Encyclopedia of Machine Learning*; Springer: Berlin/Heidelberg, Germany, 1951; pp. 581–583.
51. Uddin, A.H.; Chen, Y.L.; Akter, M.R.; Ku, C.S.; Yang, J.; Por, L.Y. Colon and lung cancer classification from multi-modal images using resilient and efficient neural network architectures. *Heliyon* **2024**, *10*, e30625. [[CrossRef](#)] [[PubMed](#)]
52. Ali, N.J. A Deep Hybrid CNN-transformer Model for Multi-modal Tumour Detection and Segmentation Across Medical Imaging Applications. *Int. J. Intell. Eng. Syst.* **2026**, *19*, 138.
53. Yıldız, G.; Yakut, Ö. Multi-class cancer diagnosis on histopathological images with deep ensemble learning model. *Comput. Biol. Med.* **2026**, *200*, 111381. [[CrossRef](#)]
54. Yousafzai, S.N.; Nasir, I.M.; Mansour, S.; Negm, N.; Alhashmi, A.A.; Alharbi, M.A.; Kim, E. A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. *Sci. Rep.* **2026**, *ahead of print*. [[CrossRef](#)] [[PubMed](#)]
55. Hasan, M.A.; Haque, F.; Sabuj, S.R.; Sarker, H.; Goni, M.O.F.; Rahman, F.; Rashid, M.M. An end-to-end lightweight multi-scale CNN for the classification of lung and colon cancer with XAI integration. *Technologies* **2024**, *12*, 56. [[CrossRef](#)]

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