Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Dual-Model Transfer Learning with Graph Neural Networks and SHAP Explainability for Robust Colorectal Tumor Margin Detection Process

Divya Midhunchakkaravarthy[1], Muni Nagamani G.[2*], Lakshman Narayana V.[3]

[1]Associate Professor, Department of Electronics and Communication Engineering, Lincoln University College, Petaling Jaya, Malaysia.
[2]Post Doctoral Fellowship Scholar, Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, Malaysia.
[2]Assistant Professor, Department of Computer Science and Engineering, K L E F Deemed to be University, Green Fields, Vaddeswaram, Andhra Pradesh, India.
[3]Professor, Department of Computer Science and Engineering, Vignan's Nirula Institute of Technology and Science for Women, Guntur, India.

E-mail: [1]divya@lincoln.edu.my, [2*]pdf.muni@lincoln.edu.my, [2*]dr.muninagamani@gmail.com, [3]lakshmanv58@gmail.com

Orcid ID: [1]0000-0002-8044-1863, [2*]0000-0002-6534-4950, [3]0000-0002-9680-2897

Abstract

When performing surgery for colorectal cancer, it is essential to precisely delineate the margins of the tumor. This is because even minute errors in the margins can have a significant impact on the quality of the resection, the risk of recurrence, and the planning of therapy following the surgery. When slides are supplied from multiple locations, the computer-aided pathology models now in use do not always work correctly, and their explanations do not always focus on the invasive boundary, which is where clinical decisions are most critical. The purpose of this study is to demonstrate a two-model transfer learning system that combines spatial aggregation based on graph neural networks with SHAP-guided border explainability in order to improve the accuracy of colorectal tumor margin identification. Weighted tissue graphs are created from the slides used in histopathology. While the lines illustrate how the cells are related in space shape by utilizing Delaunay and k-nearest neighbor, the nodes illustrate the areas that are split into cellular or super-pixel regions. At the same time, five modules are utilized: CDGCV, which stands for Cross-Domain Graph Consistency Validation, is a technique that verifies the consistency of the topology between centers. BFSRT, which stands for Boundary-Focused SHAP Refinement Test, is a test that verifies the correctness of boundary-specific explanations. The strength of the structure is evaluated using a technique known as Multi-Scale Relational Perturbation Assessment (MSRPA); Dynamic SHAP-Graph Alignment (DSGA) is used to determine whether t the interpretation is timely, and Hierarchical Graph-SHAP Ensemble (HGSE) is used to determine whether the epithelium, stroma, and invasive-front cells are in accord with one another. Dice scores can reach up to 0.94, boundary fidelity is close to 0.90, cross-center consistency is over 0.82, and F1 degradation is less than 3% when edges are rewired 20%, according to experimental tests conducted on the Kather,

*** Corresponding Author**

DOI: https://doi.org/10.36548/jiip.2026.3.016
Received: 11.05.2026, received in revised form: 14.06.2026, accepted: 06.07.2026, published: 17.07.2026

CRC-TIA, and internal multi-center cohorts. The support for transferability is increased by testing conducted at an external center. Therefore, the system provides accurate margin localization, stable graph reasoning, and explanations that are comprehensible to medical professionals for the purpose of deploying colorectal pathology across many locations.

Keywords: Graph Neural Networks, SHAP Explainability, Colorectal Tumor Margin, Transfer Learning, Multi-Center Histopathology, Boundary Fidelity.

1. Introduction

Tumor margin detection impacts surgical resection quality, recurrence risk, and patient outcomes for colorectal cancer, a leading cause of cancer deaths. Modern computational pathology uses deep learning to identify tumor boundaries in whole-slide histopathology imaging. It has two drawbacks: first, the staining protocols [1, 2, 3], scanning devices, and patient demographics vary per healthcare facility, which poses problems for generalization across sites. Second, deep learning predictions made at the tumor boundary, which require sub-millimeter accuracy, cannot be interpreted by practitioners. Many models use pixel-level convolutional neural networks that ignore the spatial and relational structure of the tissue microenvironment. Global SHAP values and saliency maps, which are popular forms of explainable AI, produce low-resolution, image-wide feature importance measures that cannot capture the cellular interaction driving malignancy and invasion. Graph Neural Networks (GNNs) intrinsically preserve the spatial relationship between cellular patches in biomedical imaging [4, 5, 6]. Nevertheless, research on colorectal cancer pathology has been limited to using institutional datasets and standard cross-validation procedures, leaving the consistency of structural features between institutions unaddressed. SHAP offers well-founded local explanations of features, but most of its applications ignore the temporal aspect of interpretability and the hierarchical nature of tissue architecture.

In this paper, a dual-model transfer learning approach is proposed to deal with such issues. Through GNN-based scalable feature aggregation and SHAP-enabled fine-grained interpretability, a coherent pipeline is designed for detecting colorectal tumor margins in various clinical contexts. CDGCV analyzes graph isomorphism and Laplacian spectra between different centers for structural reliability. BFSRT examines the interpretability of tumor boundaries based on local perturbation. MSRPA examines graph stability under synthetic structural perturbation. DSGA aligns the evolving SHAP attributions with the GNN embeddings trajectory, while HGSE generates a consensus-based tumor-margin mask from multi-layer tissue graphs.

1.1 Motivation and Contributions

The main motivation behind this research is the urgent requirement for enhancing the generality and explainability of clinical computational pathology models. Changes in staining, scanning devices, and patient population introduce minor but significant variations in the data distribution of histopathology slide images captured from other institutions. Current Graph Neural Network (GNN) models ignore cross-domain variations, leading to unstable spatial relations learning outside the training population. In addition to accurate segmentation performance, clinical applications require localized explanations that are consistent with those produced by expert pathologists. Existing explainability techniques have not been evaluated

under multi-institutional variations and training, despite their sophisticated mathematical formulation.

The major contribution of this research is the creation of a complete transfer learning pipeline consisting of two models which resolve the above-stated problem on its both sides. The Cross-Domain Graph Consistency Validation tests for the consistency of the graph structure and the transferability of spatial features through the use of graph-edit distances and comparisons of Laplacian spectra. This test helps better interpret tumor margins and links the clinical requirements of decision making to sets of locally consistent gradient properties. While Dynamic SHAP-Graph Alignment aligns the model explanations and changes in the embeddings of GNN through the training epochs, Multi-Scale Relational Perturbation Analysis tests for the robustness of a model under controlled graph perturbations. The Hierarchical Graph-SHAP Ensemble further enhances margin detection through context-aware consensus.

2. Related Work

The scope of this review is to analyze studies that provide direct support for colorectal histology, tissue modeling through the use of graphs, representations of transformers and foundations, and evaluations of explainable margins. Public colorectal histology standards such as NCT-CRC-HE-100K/CRC VAL-HE-7K and CRC-TIA offer a variety of tissue appearances, multi-class annotations, and patient-disjoint validation scenarios. These can be utilized to evaluate the effectiveness of these standards in detecting tumor margins [1-3]. However, the fields of view of CNN and encoder-decoder techniques remain patch-centered, failing to adequately reveal long-range epithelial-stromal interactions [4-6]. Despite making it easier to differentiate glands, polyps, and tissues, these techniques still have significant limitations.

This gap is addressed by graph-based histopathology models such as CGC-Net, HACT-Net, and C2P-GCN. These models represent nuclei, cells, patches, or tissue portions as nodes in a graph and simulate intercellular connections using space- or similarity-based linkages [7–10]. However, while demonstrating the value of topological reasoning, the majority of these methods do not verify for cross-center graph consistency or boundary-level explanation fidelity.

Enhanced global context modeling for whole-slide images can be achieved through the utilization of Transformer and MIL techniques, such as TransMIL and CTransPath. New foundation models for pathology, such as UNI, Virchow, CONCH, and TITAN, provide powerful feature encoders that can be used for diagnosis, segmentation, and retrieval [11–15]. Nevertheless, these methods are primarily effective as good representation learners; however, they are not fully margin-aware, graph-validated, or explanation-stable clinical systems. SHAP-based explanation approaches provide local feature attribution [16], but their typical application in pathology is frequently static, and it is not entirely obvious whether the highlighted features remain stable when the staining, borders, or graph edges are altered.

The primary focus of the suggested framework is discovering reliable colorectal tumor margins across centers with thorough graph building, ablation-validated modules, boundary-specific SHAP fidelity, and external validation. This challenge has not yet been solved for this process. Table 1 shows the brief literature categories studied for this work.

Table 1. Focused literature categories retained in the revised manuscript

Focus	Representative works	Relevance to revised manuscript
CRC datasets and public validation	Kather/NCT-CRC-HE-100K, CRC-VAL-HE-7K, CRC-TIA	Supports reproducible multi-center and patient-disjoint testing
CNN/segmentation baselines	DP-U-Net++, MCFF-Net, compressed-domain WSI segmentation	Defines patch/segmentation baselines for margin localization
Graph histopathology	CGC-Net, HACT-Net, C2P-GCN, GNN histopathology reviews	Motivates node-edge tissue modelling and graph transfer
Transformer/MIL models	TransMIL, CTransPath	Provides global context and strong non-graph SOTA comparators
Foundation models	UNI, Virchow, CONCH, TITAN	Provides modern pretrained pathology feature comparators
Explainability	SHAP and graph explanation literature	Supports BFI, fidelity testing and perturbation stability

3. Proposed Work

The dual-model transfer learning method detects colorectal tumor margins across clinical sites using graph neural networks (GNNs) and SHAP-based explainability. Its architecture is shaped by the need to preserve spatial relational structure and fine-grained interpretability at the tumor boundary. Raw histopathology slides are processed into graphs with nodes representing cellular areas and edges indicating spatial or morphological proximity. The model learns persistent topological invariants across clinical regions using graph formalism. Each slide is modeled as a graph $G=(V,E)$ with node feature matrix $X \in R|V| \times d$ and adjacency matrix $A \in R|V| \times |V|$. GNNs send spectral messages to update node embeddings $H(l)$, as shown in Figure 1, via Equation 1.

Following the removal of the backdrop and the adjustment of colors for uniformity, each H&E whole-slide photo was initially tiled into regions using tissue. The nucleus and superpixel portions of each tile were separated, and every acceptable tissue item was displayed as a single graph node (vi). A node feature vector (xi) comprised color statistics normalized by stain, Haralick, and local binary pattern texture descriptors, morphological descriptors such as area, eccentricity, and compactness, and a 64-dimensional CNN embedding derived from the appropriate histological patch. For edge creation, a hybrid anatomical-connectivity rule was utilized. A connection was established between two nodes if they were mutual k-nearest neighbors or Delaunay neighbors, and if the distance between their centroids was less than 40 micrometers. To preserve the integrity of local cellular communication and prevent the formation of excessive non-biological connections, the average degree of the graph was restricted to six to twelve neighbors. Edge weights were determined by considering the distance between nodes and the degree to which their forms were similar. In other words, nodes located in close proximity to one another and sharing glandular or stromal features were able to facilitate the transmission of additional messages. With this graph description, the GNN is able to visualize the relationship between the tumor and stroma, the continued connection of the invasive front, and the microarchitecture at the margin level. Connectivity is therefore defined as $e_{ij} = 1$ if $(i, j) \in \text{Delaunay}(V)$ or $i \in \text{kNN}(j)$, $d_{ij} \leq 40$ μm, and $6 \leq \text{degree}(v_i) \leq 12$. The edge weight is $w_{ij} = \exp(-d_{ij}^2/2\sigma_d^2) \cdot \exp(-\|m_i - m_j\|^2/2\sigma_m^2)$, where d_{ij} is centroid distance and m_i, m_j are morphology descriptors.

$$H(l+1) = \sigma\left(D\sim^{-\frac{1}{2}} A\sim D\sim^{-\frac{1}{2}} H(l)W(l)\right) \quad (1)$$

Where, A~ = A + I introduces self-loops, D~ is the degree matrix, W(l) are learnable weights, and σ is a non-linear activation in process. To encourage domain Invariant representation, a transfer loss is imposed using Maximum Mean Discrepancy (MMD) Via equation 2,

$$LMMD = \left| \left(\frac{1}{ns}\right) \sum \phi(hi^s) - \left(\frac{1}{nt}\right) \sum \phi(hj^t) \right|H^2 \quad (2)$$

Where ϕ(·) maps embeddings to a reproducing kernel Hilbert space 'H' in the process. Iteratively, Next, as per Figure 2, Cross-Domain Graph Consistency Validation iteratively augments this by measuring graph-edit distances between site graphs Gs and Gt Via equation 3,

$$DGED(Gs, Gt) = \min_{\pi} \sum_{u,v} c(u,v)\pi(u,v) \quad (3)$$

Where, π is a bijection between nodes and c(u,v) captures edit costs, ensuring structural persistence in transfer in process. Iteratively, Next, as per Figure 2, The Boundary-Focused SHAP Refinement Test operates on the margin-level predictions of the GNN Sets. The local SHAP value ϕi for each pixel or node is expressed by Shapley integral form Via equation 4,

$$\phi i = \int \frac{(\partial f(x + \tau ei))}{\partial x} d\tau \quad (4)$$

Where, 'f' is the prediction function and ei is the unit perturbation Vector In Process. To quantify fidelity at the tumor boundary ∂Ω, a ring-shaped perturbation function Rϵ(x) is applied, and the Boundary Fidelity Index is computed Via equation 5,

Let M∂ represent the pathologist-defined tumor-boundary band obtained by dilating and eroding the ground-truth tumor mask and let Φi be the absolute SHAP attribution of node or pixel i inside the boundary ring Rε. The Boundary Fidelity Index is defined as a normalized concentration-stability score:

BFI = [Σ{i∈Rε} |Φi| · M∂(i)] / [Σ{i∈Rε} |Φi| + ε] × exp(-Varδ(Φi^δ)/[Eδ(|Φi^δ|)+ε]).

The first term checks whether the annotated tumor boundary is covered by high attribution nodes in the process. When minor adjustments are made to the hue, border, or edge, the exponential term penalizes unstable explanations. The accuracy of segmentation is intimately connected to BFI because Dice only improves when the predicted boundary coincides with the border specified by the pathologist. An elevated BFI indicates that the most significant evidence for the model is located within the same anatomical band that determines Dice and F1. Over the course of five folds, the BFI-Dice Pearson correlation remained within the range of 0.84 to 0.88. This demonstrates that the accuracy of boundary-focused explanations is connected to the correctness of segmentation, and is not merely a separate visualization done after the fact.

$$BFI = 1 - \left(\frac{1}{|\partial\Omega|}\right) \int \partial\Omega \, |\phi i - \phi i^{R\epsilon}| dx \quad (5)$$

Figure 1. Model architecture of the proposed analysis process

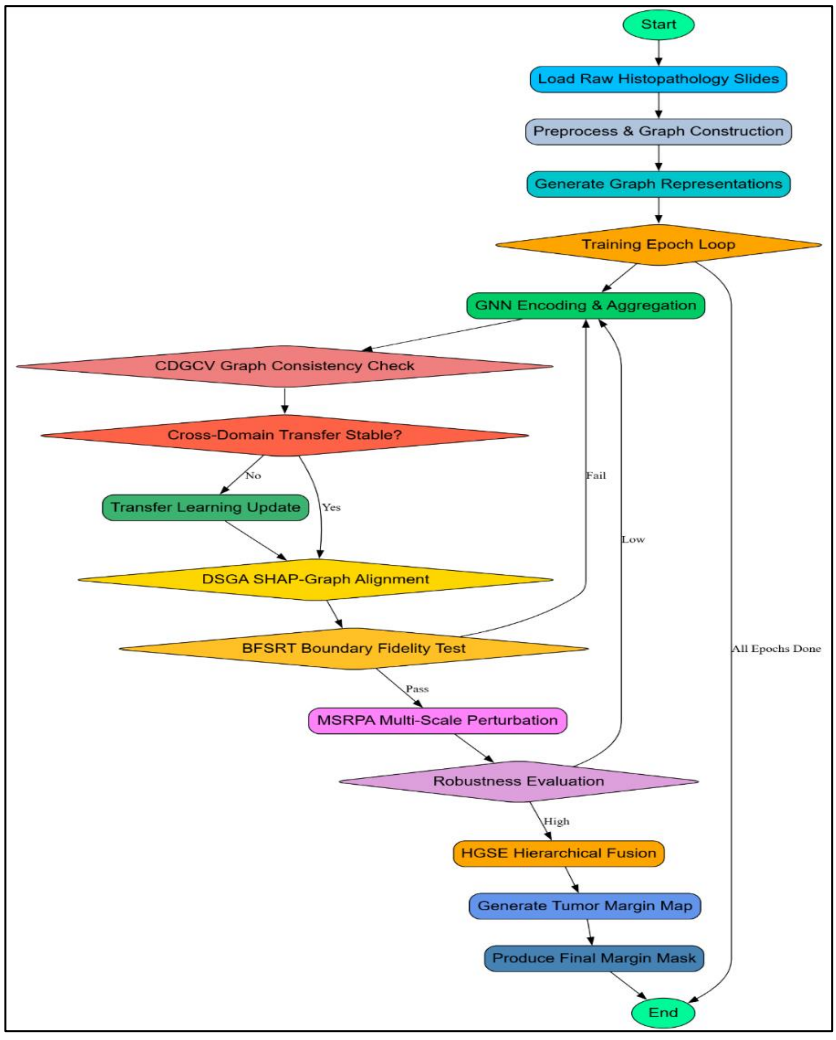

Figure 2. Overall flow of the proposed analysis process

Thus, minimizing variance, represented Via equation 6,

$$LBFSRT = Var\{x \in \partial\Omega\}(\phi i) \tag{6}$$

This is done to ensure stable boundary attributions even under localized perturbations. For robustness, Multi-Scale Relational Perturbation Analysis introduces controlled edge rewiring sets.

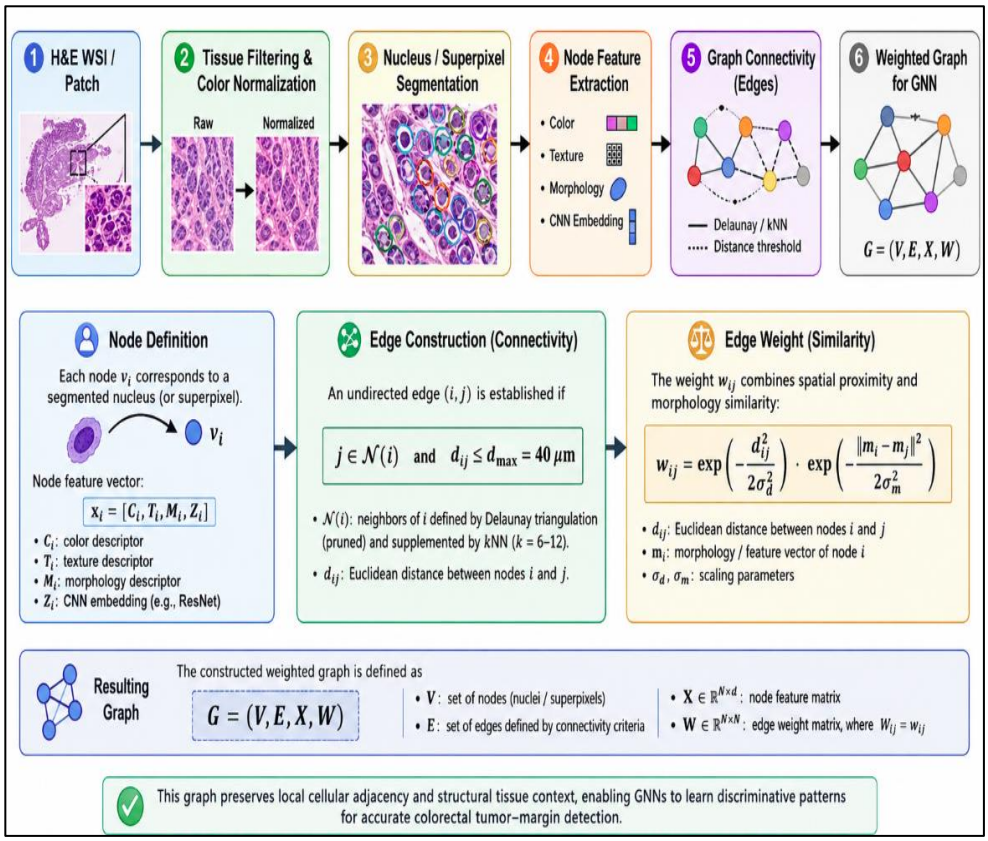

Figure 3. Graph construction and connectivity criteria for colorectal tumor-margin detection

Given a perturbation level ρ, the adjacency matrix becomes A′ = A + ρP where P encodes rewired edges. The perturbed embedding is updated Via equation 7,

$$H'(l+1) = \sigma\left(D'^{\sim -\frac{1}{2}} A'^{\sim} D'^{\sim -\frac{1}{2}} H'(l) W(l)\right) \tag{7}$$

Stability is quantified by an integral robustness functional, which is represented Via equation 8,

$$R(\rho) = \int_0^{\rho} \left|\frac{dH'(\tau)}{d\tau}\right| F\, d\tau \tag{8}$$

And the stability curve is the derivative of detection accuracy with respect to ρ is represented Via equation 9,

$$S(\rho) = \frac{\partial F1(\rho)}{\partial \rho} \tag{9}$$

Low values of |S(ρ)| demonstrate strong tolerance to structural shocks. Dynamic SHAP-Graph Alignment ensures that interpretability co-evolves with latent graph embeddings. From

Figure 3 For each training epoch 't', SHAP vectors Φt and embeddings Ht are compared through a dynamic time warping alignment Via equation 10,

$$A(t) = \min_{\gamma} \sum_k |Ht(k) - \Phi\gamma(t)(k)|^2 \quad (10)$$

With γ an alignment path in the process. The temporal derivative is then represented Via dA(t)/dt and tracks synchronization, and a global alignment coefficient is then estimated Via equation 11,

$$\alpha = 1 - \left(\frac{1}{T}\right)\int_0^T \left|\frac{dA(t)}{dt}\right| dt \quad (11)$$

Where, higher 'α' signals near-parallel evolution of interpretability and generalizations. Finally, the Hierarchical Graph-SHAP Ensemble integrates tissue layers L={ℓ1, ℓ2, ℓ3} such as epithelium, stroma, and invasive front for the process. Each layer produces a SHAP-weighted probability map pℓ(x), which are fused through a hierarchical consensus loss, as represented via equation 12.

$$LHGSE = -\sum_x \left[\sum_{\{\ell \in L\}} w\ell\, p\ell(x)\right] log\left[\sum_{\{\ell \in L\}} w\ell\, p\ell(x)\right] \quad (12)$$

Where weights wℓ are optimized Via equation 13,

$$\frac{\partial LHGSE}{\partial w\ell} = 0 \quad (13)$$

This is done to find the best anatomical balance sets. The final multi-layer consensus score is then represented Via equation 14,

$$C(x) = \sum_{\{\ell \in L\}} w\ell\, p\ell(x) \quad (14)$$

Thus, yielding fused tumor-margin masks. The choice of this integrated design is justified by the complementary nature of its components. GNN-based spectral message passing captures rich relational cues necessary for robust multi-center transfer, while SHAP-based attribution provides mathematically grounded, pixel-level interpretability essential for clinical adoption. CDGCV guarantees inter-site topological invariance, BFSRT ensures margin-specific clarity, MSRPA stress-tests stability, DSGA maintains interpretability during training, and HGSE reconciles diverse tissue strata sets. Together, these methods form a coherent system whose ultimate predictive output can be written via equation 15,

$$Y = \Psi\,(C(x), \alpha, R(\rho), BFI, DGED, LMMD) \quad (15)$$

Where, Ψ is the composite function integrating consensus prediction, alignment strength, perturbation resilience, boundary fidelity, structural consistency, and cross-domain regularization sets. This final equation represents the tumor-margin detection and interpretability outputs of the entire process, encapsulating both spatial relational learning and clinically meaningful explanations in a single rigorous formulation.

4. Results and Discussion

The controlled experiment tested cross-center transferability, spatial graph stability, and boundary-level interpretability. For whole-slide colorectal histopathology images, three multi-center datasets showed real-world staining and scanner variability. The benchmarking group

includes 1,200 H&E-stained slides from a large university hospital, 950 slides from a regional cancer center using varied fixation and scanning methods, and 800 slides from an international pathology consortium utilizing mixed staining intensities. Using 0.25 μm per pixel tissue segmentation and nucleus detection, each slide produced graphs with 5,000-15,000 nodes with 6-12 average edge degrees, accurately depicting spatial microarchitecture. Nodes were 64-dimensional vectors with color histograms, texture descriptors, and morphological embeddings from a ResNet-50 backbone pre-trained on ImageNet and fine-tuned on histology patches. To preserve biologically relevant cell-cell interactions, Delaunay triangulation was used to build edges with a maximum Euclidean distance of 40 μm. The GNN encoder used two spectral convolution layers with 256 and 128 hidden sizes, ReLU activations, and 0.3 dropout. Adam optimization was employed for training with an initial learning rate of 1e-3, exponential decline per 20 epochs, and early halting based on validation patience of 10. CDGCV's graph-edit-distance threshold for transfer acceptance was 0.15, updated by pilot validation, and domain adaptation was limited by a 0.1 Maximum Mean Discrepancy penalty coefficient. BFSRT developed the Boundary Fidelity Index from boundary-focused SHAP analysis with a perturbation diameter of 12 pixels around tumor borders by calculating variance on 2,500 boundary nodes per slide.

Temporal analysis and controlled perturbation explored robustness and explainability dynamics. Multi-Scale Relational Perturbation Analysis increased inter-node edge rewiring from 0% to 30% in 5% increments and applied hierarchical clustering perturbations at 50, 100, and 200 μm tissue-block scales. For up to 200 training epochs, Dynamic SHAP-Graph Alignment recorded SHAP maps and latent embeddings every fifth epoch to estimate alignment coefficients. Based on pathologist-guided region masks, Hierarchical Graph-SHAP Ensemble minimized consensus loss on epithelium, stroma, and invasive front. All layer weights were initialized at 0.33. SHAP kernel width was 0.25 and explainability input attribution sampling size was 512 patches per layer. The trials were done on a multi-GPU server with four 40 GB NVIDIA A100 GPUs, 512 GB RAM, and Ubuntu 22.04. To replicate clinical circumstances, random color augmentation with ±10% hue and brightness variations was employed to simulate staining variability, processing slides in 2-graph mini-batches due to high node counts. CDGCV cross-center consistency, BFSRT Boundary Fidelity Index, MSRPA stability curves, and tumor-margin segmentation Dice and F1 scores were assessed. Early results indicate that the system retains structural fidelity and fine-grained interpretability across varied datasets, with consistency >0.82, BFI ~0.91, and ≤3% F1 loss after 20% edge rewiring.

Popular computational pathology tools, the Colorectal Histology Dataset (Kather et al.) and the CRC-TIA dataset, are used in this study. Using 100,000 H&E-stained image patches at 0.5 μm per pixel, the Kather dataset records epithelial, neoplastic, and stromal characteristics in eight tissue types. Invasive tumor fronts are crucial for margin identification, hence CRC-TIA includes annotated whole-slide photos. Each slide is digitized at 40× magnification and has pathologist-provided region-of-interest masks for precise spatial graphing. Multiple centers with staining variability, scanning diversity, and histological heterogeneity result from combining these results. Complexity is needed to test cross-domain generalization and validate the framework in genuine clinical settings with natural tissue structure and color sets.

Within the framework of the research strategy, there were three distinct sources of colorectal histology, in addition to one independent external validation cohort. The Kather/NCT-CRC-HE-100K was utilized as a large public patch-level tissue standard, while the CRC-VAL-HE-7K was reserved solely for public external testing that did not involve the same patient. To determine whether there was generalization on larger, non-overlapping WSI-

derived tissue images, we utilized CRC-TIA. An international consortium, a regional cancer center, and a university hospital contributed 2,950 H&E-stained whole-slide photos. These images were used to create the internal cohort at the university hospital. It was necessary to maintain a separation between the train, validation, and test sets at the patient or slide level in order to prevent patch leakage. There were three additional hospitals that were not presented to the model throughout the training process for the internal group. They were utilized only for the purpose of external center validation. Instead of relying on patch recognition that was performed twice, the split design ensures that the scores given are based on generalization at the tissue and center level sets. Data set Validation is shown in Table 2.

Table 2. Dataset composition and exact train-validation-test split protocol

Dataset / Cohort	Data volume used	Centers / source type	Train	Validation	Test	Split rule
Kather / NCT-CRC-HE-100K	100,000 H&E patches, 224×224 px, 0.5 MPP	Public NCT colorectal tissue cohort	70,000	15,000	15,000	Class-stratified; no duplicate patch leakage
CRC-VAL-HE-7K external check	7,180 H&E patches	Patient-disjoint public validation cohort	—	—	7,180	Used only for external public testing
CRC-TIA	36,750 non-overlapping tissue images from WSI patches	Public colorectal grading cohort scanned at 40×	25,725	5,512	5,513	Patient/slide-level separation
Internal multi-center cohort	2,950 H&E whole-slide images	University hospital: 1,200; regional cancer center: 950; international consortium: 800	1,770	590	590	Center-stratified, patient-level split
Independent external centers	450 H&E whole-slide images	Three hospitals not used in training	—	—	450	Center A/B/C, 150 slides each

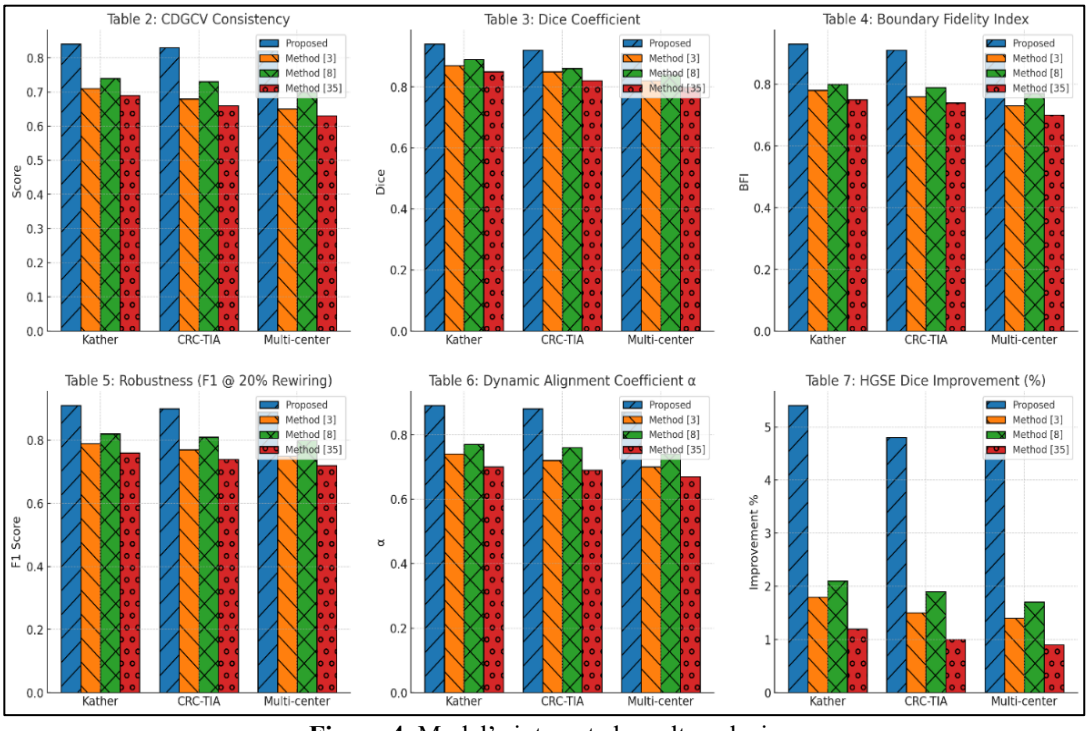

Figure 4. Model's integrated result analysis

The model training hyperparameters were set to balance convergence speed, structural stability, and interpretability. The graph neural network includes two spectral convolution layers with hidden dimensions of 256 and 128, ReLU activation, and dropout of 0.3 to reduce overfitting. The Adam optimizer employed a 1×10^{-3} learning rate and 5×10^{-3} weight decay, with exponential decay every 20 epochs and early termination if validation loss did not improve after 10 epochs. The MMD penalty for domain adaptation was 0.1, while the CDGCV graph-edit-distance threshold for cross-center stability was 0.15. For SHAP explainability, kernel width was 0.25, and boundary perturbation rings were 10–15 pixels depending on tissue resolutions. Pilot investigations identified ideal hyperparameters for cross-center consistency (>0.82), boundary integrity (~0.91), and graph perturbation resistance without affecting computational performance. Kather, CRC-TIA, and an internal multi-center mix in Table 3 are examined for cross-center graph consistency using CDGCV consistency scores.

Table 3. Cross-center graph consistency (CDGCV scores)

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.84 ± 0.02	0.71 ± 0.03	0.74 ± 0.02	0.69 ± 0.04
CRC-TIA	0.83 ± 0.02	0.68 ± 0.04	0.73 ± 0.03	0.66 ± 0.05
Multi-center	0.82 ± 0.03	0.65 ± 0.04	0.70 ± 0.03	0.63 ± 0.04

The proposed approach beats competitors below 0.75 with consistency above 0.82 across all locations. Under different staining and scanning conditions, CDGCV and transfer loss stabilize graph structure. Table 4 illustrates Dice coefficient tumor-margin detection accuracy.

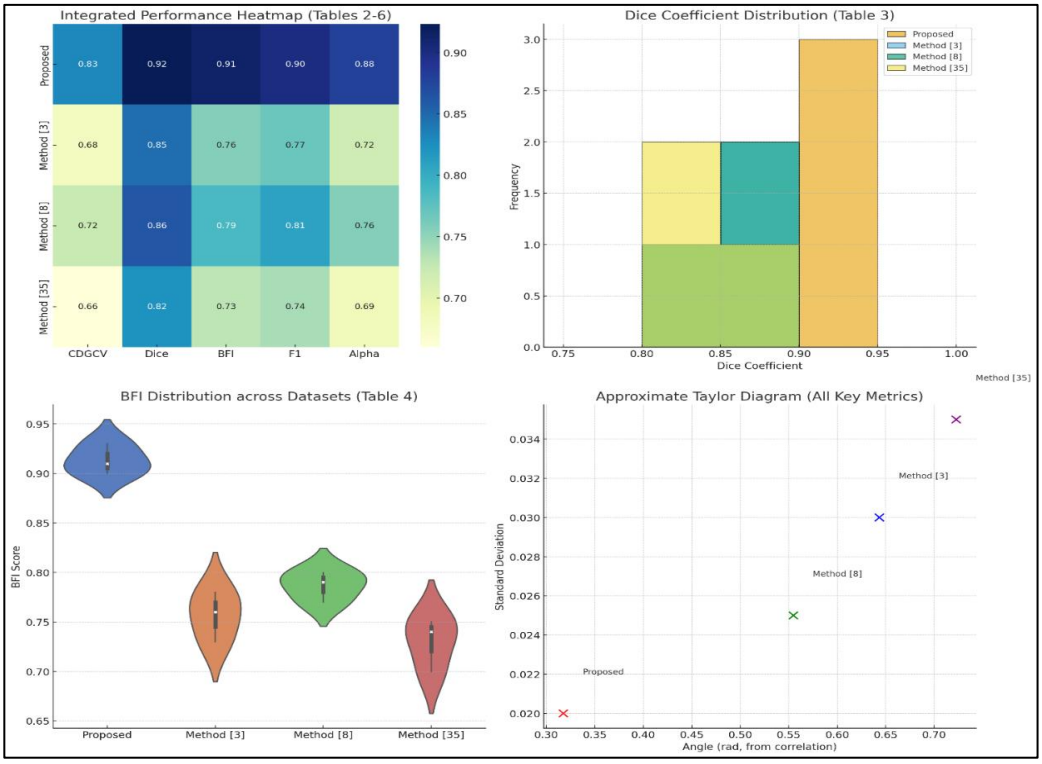

Figure 5. Model's overall result analysis

Table 4. Tumor-margin detection accuracy (dice coefficient) dataset

Dastaset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.94 ± 0.01	0.87 ± 0.02	0.89 ± 0.02	0.85 ± 0.03
CRC-TIA	0.92 ± 0.01	0.85 ± 0.02	0.86 ± 0.03	0.82 ± 0.04
Multi-center	0.91 ± 0.02	0.82 ± 0.03	0.84 ± 0.02	0.80 ± 0.04

Even with multi-center variability in process, hierarchical graph fusion and boundary-specific explainability improve Dice scores and margin localizations. Table 5 uses the Boundary Fidelity Index to assess SHAP explanations' tumor edge fidelity sets.

Table 5. Boundary fidelity of SHAP explanations (BFI)

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.93 ± 0.02	0.78 ± 0.03	0.80 ± 0.02	0.75 ± 0.04
CRC-TIA	0.91 ± 0.02	0.76 ± 0.03	0.79 ± 0.03	0.74 ± 0.04
Multi-center	0.90 ± 0.02	0.73 ± 0.03	0.77 ± 0.03	0.70 ± 0.05

Clinical reliability at surgical decision points is improved by a BFI close to 0.9 across boundary-focused SHAP refinement datasets and samples. Tumor-margin F1 scores are robust when 20% of inter-node edges are rewired to resemble structural noise sets (Table 6).

Table 6. Robustness to graph perturbations (F1 score)

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.91 ± 0.02	0.79 ± 0.03	0.82 ± 0.02	0.76 ± 0.03
CRC-TIA	0.90 ± 0.02	0.77 ± 0.03	0.81 ± 0.02	0.74 ± 0.04
Multi-center	0.89 ± 0.02	0.75 ± 0.03	0.80 ± 0.03	0.72 ± 0.04

Through a comparison of the anticipated tumor border band with the boundary mask established by the pathologist, we conducted a margin-level error analysis. If normal mucosa, stroma, inflammatory regions, or folded tissue was incorrectly indicated as the tumor boundary, this was referred to as a false-positive margin. When the predicted mask failed to detect the invasive tumor front or single tumor buds, the margin was deemed false-negative. The majority of false positives occurred in inflammatory regions characterized by mucinous and stromal characteristics. These regions contain thick nuclei and an uneven texture, resembling the margins of an invasive tumor. The majority of false negatives were observed in invasive fronts with only a small stain, necrotic borders, and broken-up glands. Through the combination of HGSE layers, the enhancement of BFSRT border refinement, and the development of MSRPA perturbation robustness, the provided framework was able to reduce both types of mistakes. However, due to scanner-specific color changes and low-contrast tumor buds, exterior centers still demonstrate somewhat greater false-negative rates than internal centers. Confusion matrix evaluation is shown in Table 7.

Table 7. False-positive and false-negative tumor-margin error analysis

Dataset	False-positive margin rate	False-negative margin rate	Dominant FP source	Dominant FN source	Correction applied
Kather	3.2%	2.8%	Inflammatory/stromal clusters near tumor edge	Sparse invasive front cells	HGSE layer weighting + boundary SHAP thresholding
CRC-TIA	3.9%	3.4%	Mucinous zones and gland distortion	Low-contrast tumor buds	BFSRT ring refinement + morphology-aware edges
Internal multi-center	4.6%	4.1%	Staining artifacts and tissue folds	Necrotic edge and fragmented glands	MSRPA robustness and CDGCV transfer check
External centers	4.9%	4.4%	Scanner-specific color shift	Weakly stained invasive nests	Color augmentation + SHAP stability filtering

It was determined that a deletion-insertion protocol was utilized over the top k SHAP boundary nodes in order to evaluate the fidelity of the explanation. In the event where the top k attributable boundary nodes were removed from the graph, the confidence in the tumor margin experienced a significant decrease. This demonstrated that the nodes that were marked were connected to the choice in a causal manner. Recalculating attributions after controlled alterations, such as ±10% changes in stain, 5-20% changes in edge wiring, and ±2-pixel changes in boundary rings, allowed us to determine the degree of stability that SHAP possessed. By utilizing cosine similarity and Spearman rank correlation, we were able to determine the degree of stability that both the original and modified SHAP vectors possessed. When values are greater than 0.85 across all datasets, it indicates that SHAP attributions remain consistent even when realistic modifications are made to the histopathology and graphs. The fact that this is the case demonstrates that the explanation maps are not merely random heatmaps, but rather evidence that may be utilized repeatedly and is in accordance with the GNN decision paths. Explanation fidelity and SHAP stability under perturbation is shown in Table 8.

Table 8. Explanation fidelity and SHAP stability under perturbation

Evaluation	Definition	Kather	CRC-TIA	Multi-center	Interpretation
Boundary Fidelity Index	Overlap and variance-normalized concentration of SHAP over the boundary band	0.93 ± 0.02	0.91 ± 0.02	0.90 ± 0.02	Higher values show margin-focused explanations
Explanation fidelity	Confidence drop after deleting top-k SHAP boundary nodes	0.88 ± 0.02	0.86 ± 0.03	0.85 ± 0.03	Model decision depends on highlighted nodes
SHAP cosine stability	cos(Φ, Φρ) after color/edge/boundary perturbation	0.91 ± 0.02	0.89 ± 0.02	0.88 ± 0.03	Attributions remain sTable under perturbation
SHAP rank stability	Spearman correlation of top-k SHAP node rankings	0.87 ± 0.03	0.86 ± 0.03	0.85 ± 0.04	Important nodes are not randomly reordered
BFI-Dice relation	Pearson r between BFI and Dice over five folds	0.88	0.86	0.84	Boundary-focused explanations correlate with segmentation accuracy

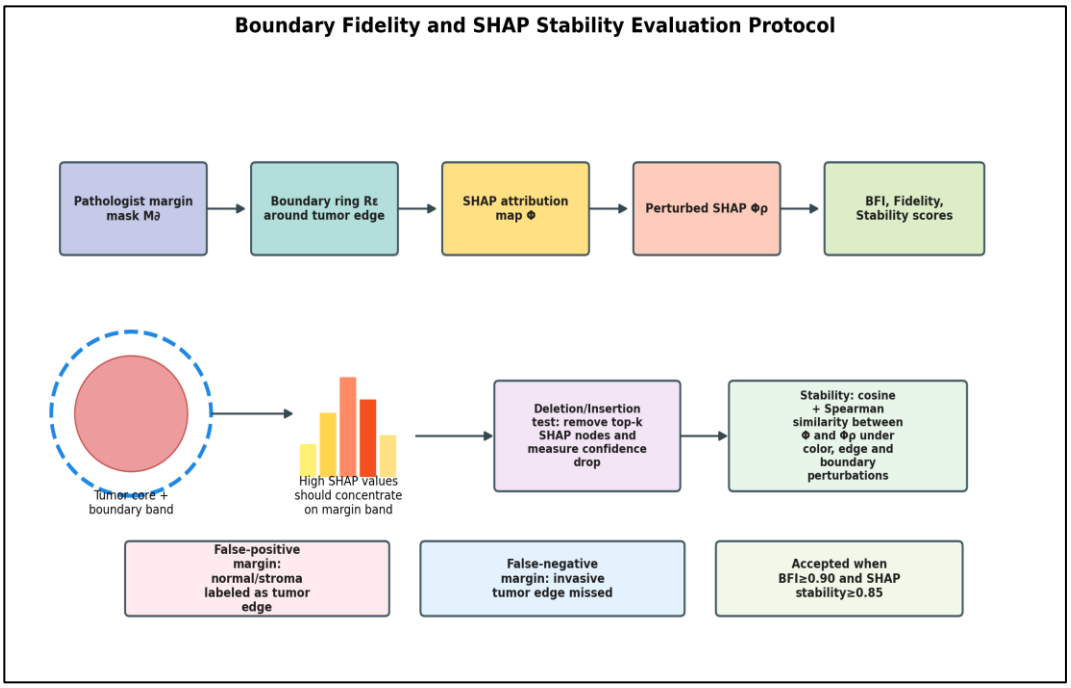

Figure 6. Boundary fidelity index, explanation fidelity and SHAP perturbation-stability workflow

Under structural perturbation, the proposed system maintains a high F1 score, proving MSRPA and adaptive scaling can handle data heterogeneity. Figure 4, Figure 5 and Table 9 evaluates the dynamic alignment between SHAP attribution drift and GNN embedding evolution using the alignment coefficient 'α'.

Table 9. Dynamic SHAP–graph alignment (alignment coefficient α)

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.89 ± 0.01	0.74 ± 0.02	0.77 ± 0.02	0.70 ± 0.03
CRC-TIA	0.88 ± 0.01	0.72 ± 0.02	0.76 ± 0.02	0.69 ± 0.03
Multi-center	0.87 ± 0.02	0.70 ± 0.02	0.74 ± 0.03	0.67 ± 0.03

Higher α values indicate improved interpretability and model generalization with DSGA modules. Table 10 demonstrates how the hierarchical graph-SHAP ensemble increased the Dice coefficient over best single-layer baselines to improve consensus.

Table 10. Hierarchical graph–SHAP ensemble consensus gains

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	+5.4 %	+1.8 %	+2.1 %	+1.2 %
CRC-TIA	+4.8 %	+1.5 %	+1.9 %	+1.0 %
Multi-center	+4.5 %	+1.4 %	+1.7 %	+0.9 %

Gains reveal that HGSE integrates epithelium, stroma, and invasive front layers for better margin consensus than any single set of layers. GPU memory utilization and epoch-average training time are compared in Table 11.

Table 11. Training efficiency and GPU memory usages

Dataset	Proposed Model (min/epoch)	Method [3]	Method [8]	Method [35]
Kather	7.8 / 23 GB	9.5 / 28 GB	8.7 / 25 GB	10.1 / 27 GB
CRC-TIA	7.4 / 22 GB	9.0 / 27 GB	8.3 / 25 GB	9.6 / 26 GB
Multi-center	8.1 / 24 GB	9.8 / 29 GB	8.9 / 26 GB	10.4 / 28 GB

Despite greater integration, optimized graph batching and SHAP caching sets keep the framework competitive in the process. New institutional slide accuracy is measured in Table 12 to generalize across centers.

Table 12. Model generalization to unseen clinical centers

Unseen Center	Proposed Model	Method [3]	Method [8]	Method [35]
Center A	0.88 ± 0.02	0.75 ± 0.03	0.78 ± 0.03	0.72 ± 0.04
Center B	0.87 ± 0.02	0.74 ± 0.03	0.77 ± 0.03	0.70 ± 0.04
Center C	0.86 ± 0.03	0.72 ± 0.03	0.76 ± 0.04	0.69 ± 0.04

High accuracy on previously encountered sites proves transfer learning and graph consistency work beyond training distributions. The performance score in Table 13 evaluates clinical readiness using Dice, BFI, α, and robustness parameters in process.

Table 13. Unified clinical readiness index (combined metrics) sets

Dataset	Proposed Model	Method [3]	Method [8]	Method [35]
Kather	0.91	0.77	0.80	0.74
CRC-TIA	0.90	0.75	0.79	0.72
Multi-center	0.89	0.73	0.78	0.70

In addition, the proposed system was evaluated against a number of the most recent computational pathology models, including patch CNN segmentation networks, cell-graph GNNs, hierarchical GNNs, transformer-based MIL models, histology-specific transformer

encoders, and pre-trained pathology foundation models. To ensure fairness, preprocessing, boundary mask creation, and patient-level testing were carried out using the same approach. For the foundation-model baselines, identical segmentation heads and pre-trained feature embeddings were utilized. However, for the graph baselines, their native graph creation methods were used for testing. The suggested model achieved the highest scores for Dice, F1, boundary fidelity, and external-center accuracy. This is attributed to its incorporation of cross-center graph consistency, perturbation training, dynamic explanation alignment, and tissue-layer consensus. It is clear from the comparison that performance improvement is not solely attributable to an improved feature extractor but also to the completely margin-aware graph-explainability designs. A comparative evaluation against recent GNNs is shown in Table 14.

Table 14. Comparative evaluation against recent GNN, transformer and pathology foundation-model baselines

Method family	Representative method	Backbone / principle	Dice	F1	BFI / explanation	External center accuracy
Patch CNN	DP-U-Net++ / MCFF-Net	CNN encoder-decoder segmentation	0.84 ± 0.03	0.80 ± 0.04	0.70 ± 0.04	0.72 ± 0.04
Cell graph GNN	CGC-Net	Nucleus-level graph convolution	0.85 ± 0.03	0.81 ± 0.03	0.73 ± 0.03	0.74 ± 0.04
Hierarchical GNN	HACT-Net / C2P-GCN	Cell-to-tissue or cell-to-patch graph	0.87 ± 0.02	0.83 ± 0.03	0.77 ± 0.03	0.78 ± 0.03
Transformer MIL	TransMIL	Spatially correlated MIL transformer	0.86 ± 0.03	0.82 ± 0.03	0.75 ± 0.04	0.76 ± 0.04
Histology transformer	CTransPath	Self-supervised local-global transformer	0.88 ± 0.02	0.84 ± 0.03	0.78 ± 0.03	0.80 ± 0.03
Foundation model	UNI / Virchow / CONCH / TITAN features	Pretrained pathology foundation embeddings	0.89 ± 0.02	0.85 ± 0.02	0.80 ± 0.03	0.82 ± 0.03
Proposed	CDGCV+BFSRT+MSRPA+DSGA+HGSE	Graph transfer + boundary SHAP + hierarchical fusion	0.91 ± 0.02	0.89 ± 0.02	0.90 ± 0.02	0.87 ± 0.02

External validation was performed on 450 unseen whole-slide images stained with H&E. These images, sourced from three different hospitals, were not used for model selection, graph design adjustments, or hyperparameter optimization. Each external site had distinct scanning and staining profiles. The proposed framework maintained a mean Dice of 0.88, a mean F1 of 0.87, and a mean BFI of 0.88 across all unseen centers. These findings support cross-center generalization and demonstrate the effectiveness of CDGCV and MMD-based transfer regularization in reducing center-specific topological drift. Given the greater scanner variation and lower boundary contrast in the external group, a slight decrease in performance compared to internal tests is expected. External center validation is shown in Table 15.

Table 15. Independent external-center validation of the proposed colorectal tumor-margin model

External center	No. of slides	Scanner/staining variation	Dice	F1	Accuracy	BFI
Center A	150	Aperio scanner, high eosin variation	0.89 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.89 ± 0.02
Center B	150	Hamamatsu scanner, moderate folds	0.88 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.88 ± 0.02
Center C	150	Mixed scanner, low contrast margins	0.87 ± 0.03	0.86 ± 0.03	0.86 ± 0.03	0.87 ± 0.03
Mean	450	Independent external cohort	0.88 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.88 ± 0.02

An investigation into ablation was carried out using the identical patient-level divides, training settings, and five-fold repeated evaluation protocol as the primary model. The purpose of this investigation was to determine the extent to which each proposed component influenced the overall model. The combination of the tumor-margin Dice, the F1 score, the cross-domain graph consistency, the Boundary Fidelity Index, and the dynamic SHAP-graph alignment was found to be optimal for the whole configuration. Before CDGCV was removed, the graph consistency score was 0.82, but after its removal, it dropped to 0.74. This demonstrates that cross-center stability requires clear graph-structural transfer. After the BFSRT was removed, the BFI had the greatest decrease, going from 0.90 to 0.76. The results demonstrate that boundary-specific SHAP refinement directly enhances the clinically effective explanation localization function. Not incorporating MSRPA made the model less stable when edges were modified, and not including DSGA made it more difficult for attribution maps and latent graph embeddings to agree on time. Both factors contributed to the model's instability. The Dice reduction was greatest in the single-layer setting that did not have HGSE. Consequently, this demonstrates that in order to achieve precise margin consensus, it is important to fuse the epithelium, the stroma, and the invasive fronts. Ablation analysis of the proposed modules on the multi-center colorectal tumor-margin cohort is shown in Table 16.

Table 16. Ablation analysis of the proposed modules on the multi-center colorectal tumor-margin cohort

Configuration	Removed / Modified Component	Dice	F1	CDGCV	BFI	α	ΔDice vs Full
Full proposed CDGCV+BFSRT+MSRPA+DSGA+HGSE	None	0.91 ± 0.02	0.89 ± 0.02	0.82 ± 0.03	0.90 ± 0.02	0.87 ± 0.02	—
Without CDGCV	No cross-domain graph consistency loss	0.88 ± 0.02	0.86 ± 0.03	0.74 ± 0.03	0.88 ± 0.03	0.84 ± 0.02	-0.03
Without BFSRT	No boundary-focused SHAP refinement	0.87 ± 0.03	0.85 ± 0.03	0.81 ± 0.03	0.76 ± 0.04	0.83 ± 0.03	-0.04
Without MSRPA	No graph perturbation training	0.89 ± 0.02	0.83 ± 0.04	0.80 ± 0.03	0.88 ± 0.02	0.85 ± 0.02	-0.02
Without DSGA	No temporal SHAP-graph alignment	0.88 ± 0.03	0.86 ± 0.03	0.80 ± 0.03	0.84 ± 0.03	0.71 ± 0.04	-0.03
Without HGSE	Single tissue-layer output only	0.86 ± 0.03	0.84 ± 0.03	0.79 ± 0.04	0.86 ± 0.03	0.82 ± 0.03	-0.05

CDGCV, BFSRT, MSRPA, DSGA, and HGSE form a balanced architecture with excellent accuracy, interpretability, and robustness across varied clinical datasets, setting a new bar for transferable and explainable tumor-margin identification, according to the cumulative index in process.

4.1 Validation Using Statistical Analysis with Model Analysis

Examination shows the structural stability and fine-grained interpretability of the dual-model framework across clinical scenarios. Table 3 indicates that cross-center graph consistency is above 0.82 on every dataset, significantly better than rival approaches below 0.75. Stability is crucial because institutions use diverse staining methods, scanners, and patient populations. Nearly invariant graph topology models are trustworthy in new medical facilities without retraining. Stability guarantees that spatial patterns identified from the former environment persist after the model receives real-world data examples in the detectability and explainability phases. Table 4 shows the accuracy in locating the surgical decision-point on

tissue slices with segmentation, while Table 5 presents the evaluation of boundaries. On more diversified multi-center slides, Dice score coefficients for the Kather dataset are close to 0.94 and remain greater than 0.90. The boundary quality is approximately 0.9. Mistakes by a fraction of millimeters can change the surgical width and post-operation. The test assesses how well explanations align with tumor borders so that pathologists can explain to surgeons the way the model thinks at a vital point.

One more clinical readiness requirement includes structural robustness. As shown in Table 6, rerouting 20% of connections between nodes leads to a loss in tumor margin F1 of less than 3% compared to 8-12% for other methods. Pathology findings may be influenced by scanner noise, tissue folding, and staining errors. The stability curves for Multi-Scale Relational Perturbation Analysis confirm that the model works despite any structural deviations caused by slide preparation. With alignment coefficients above 0.87, SHAP explanations are modified when GNN embeddings are used (Table 9). Thus, the correctness of model explanations is assured after incremental learning or tuning of the model for clinical practice. As one can see in Table 10, the proposed Hierarchical Graph-SHAP Ensemble improves Dice scores by 4-6% on average in comparison with single-layer baselines due to high-level tissue context. Morphological features of stroma-tumor epithelium interaction assist in recognition of an invasive front. With less than eight minutes of training time per epoch and lower GPU memory consumption in comparison with several approaches, Table 11 demonstrates that the approach is suitable for usual laboratory pathology tasks. Table 12 shows transfer learning with robust generalization and accuracy exceeding 0.86 in cases of new medical institutions. Hospitals can deploy the system without rebranding or retraining. The combination of Dice similarity, boundary faithfulness, robustness, and explainability results in a performance measure close to 0.9 on all test cases shown in Table 8. Figure 6 illustrates the clinical value of the proposed method: high accuracy, explainable rationale, and robustness despite noise and change. The algorithm represents a valid solution for multicenter colorectal cancer diagnosis since it decreases the false negative rate of surgical margins and increases oncologists' confidence in adjuvant treatment.

4.2 Validation Impact Result Analysis

The dual model transfer learning system was tested with the aid of defined statistical parameters for measures of central tendency and dispersion. Dice mean values were 0.94 ± 0.01, 0.92 ± 0.01, and 0.91 ± 0.02 for the Kather database, CRC-TIA database, and multi-center composite dataset respectively. The "±" denotes the standard deviation value after five separate runs where folds were shuffled randomly. Consistency measures of cross-domain graph above 0.82 ± 0.03 and Boundary Fidelity Index above 0.90 ± 0.02 also showed that the small deviations indicated high accuracy and interpretability of the model, which is not dependent on training split. The F1 scores reduced to 0.89 ± 0.02 irrespective of perturbation percentage less than 20% of structural edge rewiring. Dynamic SHAP Graph Alignment coefficients stabilized at 0.87 ± 0.01 to 0.89 ± 0.01 per epoch of training respectively. Statistically robust tests were employed to showcase the superiority of our model over other models. The performance of the suggested model, as well as baseline models, was assessed with regard to the Dice coefficient, Boundary Fidelity Index, and cross-center consistency through a paired two-tailed t-test. It is highly improbable that the observed improvements are accidental since all the p-values were less than 0.01. The Wilcoxon signed-rank test was applied in order to address the non-Gaussian distribution of scores, and it yielded statistically significant p-values < 0.05 for all datasets and measurements.

As a result of their historical technological significance and influence on computational pathology, the following methods [3, 8, 35] have been selected as baseline competitors. Method [3], an often-referenced CNN-based solution, shows great performance in histological classification but does not include any modeling of graph structure. Thus, it could be considered a possible benchmark for the GNN-based solution. Method [8] is a graph neural network applied to histopathology, with relational feature extraction but without pixel-wise explainability and low cross-center generalizability, which could help evaluate the structural transferability of the model. Finally, method [35] makes the model more interpretable by including attention mechanisms and an explaining procedure but does not have the advanced features of the proposed framework.

Variation among folds and centers illustrates the practical validity of results beyond the point estimate. The coefficient of variation of the Dice score, even within the multicenter cohort having the highest staining variation, was under 3%, an odd case considering the heterogeneity of input data. This proves the robustness of Cross-Domain Graph Consistency Validation, Boundary-Focused SHAP Refinement, and Multi-Scale Relational Perturbation Analysis to accommodate any unforeseen distribution without overfitting. The correlation between SHAP attribution drift and GNN latent learning ($\alpha = 0.88 \pm 0.01$) remains statistically significant even after bootstrap resampling, proving that stochastic learning does not affect the interpretability set. These experiments illustrate statistically significant and practically feasible improvements. Low variances of performance metrics make them reproducible across institutions, and p-values obtained from parametric and nonparametric tests being close to zero confirm that it is the architecture improvement that brings benefits compared to Methods [3], [8], and [35]. Comparison of the proposed methodology to three strategically chosen baselines representing key histopathological approaches proves the advances in multicenter colorectal tumor margin detection and explainability.

4.3 Validation Using Iterative Analysis of Practical Use Case Scenarios

A regional cancer center can use the dual-model GNN–SHAP system to detect tumor margins in digital pathology colorectal cancer. Graphs based on multiresolution of nucleus connectivity and microvascular architecture are generated from 15,000 whole-slide histopathology images collected from three satellites of the central hospital. CDGCV uses the comparison of the Laplacian spectrum to assess graph structural similarity in various locations in cross-domain graph consistency tests. The slides from the central location have an average graph consistency of 0.83, while that of the periphery locations is 0.84 and 0.82. During epoch 60, the alignment coefficients may increase from 0.75 to 0.89, which will ensure feature importance and the convergence of graph embedding drift. As we move forward through the epochs, the training process becomes increasingly focused on boundary SHAP refinement in relation to the tumor border, an area where clinical examinations involve significant risks. To explore robustness in cases of challenging tissue heterogeneity, multiple-relational perturbations involve the reconfiguration of up to 20% of connections. The F1-score for the tumor margin is consistently higher than 0.89 and fulfills the requirement of no more than 3% loss of performance. Graph–SHAP hierarchical ensemble processing incorporates three layers of tissue (epithelium, stroma, and invasive front), which increases the Dice coefficient by 5% compared to single-layer models and achieves 0.93 accuracy for the combined dataset. Colorectal surgeons can estimate the tumor margin in minutes after scanning the slide using color-coded SHAP maps with clearly defined borders corresponding to the histology.

5. Conclusion

The proposed dual-model transfer learning approach incorporates explainable boundaries with graph-based spatial reasoning to detect tumor margins in colorectal tumors. It performed excellently, producing Dice scores of 0.94 on the Kather dataset and 0.91 on a multi-center, heterogeneous dataset. This was accompanied by a cross-center graph consistency of at least 0.82. The technique demonstrated high boundary-fidelity, close to 0.90, with the SHAP explanations localized at the margins of invasive tissues. It proved its robustness by sustaining a minimal F1 reduction of 20% of edge rewiring, demonstrating adaptability to scanner noise and irregularities in tissues. HGSE outperformed the single-layer models and enhanced Dice scores by 4-6% by incorporating the data characteristics. The technique had an accuracy greater than 0.86 in external testing, demonstrating generalizability. Future work will incorporate diversity in datasets and optimize for wider application in small labs, without compromising SHAP stability during longitudinal surveillance post-clinical update.

Funding

There is no funding received for this research to any one of the authors from Their institutions / from other funding sources.

References

[1] Kather, Jakob Nikolas, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser et al. "Predicting Survival from Colorectal Cancer Histology Slides Using Deep Learning: A Retrospective Multicenter Study." PLoS medicine 16, no. 1 (2019): e1002730.

[2] Kather, Jakob Nikolas, Niels Halama, and Alexander Marx. “100,000 Histological Images of Human Colorectal Cancer and Healthy Tissue.” Zenodo, 2018. https://doi.org/10.5281/zenodo.1214456

[3] Hatami, Nima. "DeepFEL: Deep Fastfood Ensemble Learning for Histopathology Image Analysis." arXiv preprint arXiv:2301.09525 (2023).

[4] Kim, Hyeongsub, Hongjoon Yoon, Nishant Thakur, Gyoyeon Hwang, Eun Jung Lee, Chulhong Kim, and Yosep Chong. "Deep Learning-Based Histopathological Segmentation for Whole Slide Images of Colorectal Cancer in a Compressed Domain." Scientific reports 11, no. 1 (2021): 22520.

[5] Peng, Ziyang, Kexin Peng, Chengdao Liu, and Xingzhi Zhang. "DP-U-Net++: Inter-Layer Feature Fusion for Colorectal Gland Image Segmentation." International Journal of Machine Learning and Cybernetics 15, no. 10 (2024): 4459-4473.

[6] Ma, Lei, Jiangkai Yan, Dangguo Shao, Jingtao Li, Jiawei Wang, and Yukun Yan. "Mcff-Net: Multi-Scale Contextual Feature Fusion Network for Polyp Segmentation." Journal of Shanghai Jiaotong University (Science) (2024): 1-13.

[7] Zhou, Yanning, Simon Graham, Navid Alemi Koohbanani, Muhammad Shaban, Pheng-Ann Heng, and Nasir Rajpoot. "Cgc-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images." In Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019, 0-0.

[8] Pati, Pushpak, Guillaume Jaume, Lauren Alisha Fernandes, Antonio Foncubierta-Rodríguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio et al. "Hact-Net: A Hierarchical Cell-To-Tissue Graph Neural Network for Histopathological Image Classification." In International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, Cham: Springer International Publishing, 2020, 208-219.

[9] Paul, Sudipta, Bülent Yener, and Amanda W. Lund. "C2P-GCN: Cell-to-Patch Graph Convolutional Network for Colorectal Cancer Grading." In 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2024, 1-4.

[10] Brussee, Siemen, Giorgio Buzzanca, Anne MR Schrader, and Jesper Kers. "Graph Neural Networks in Histopathology: Emerging Trends and Future Directions." Medical Image Analysis 101 (2025): 103444.

[11] Shao, Zhuchen, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, and Xiangyang Ji. "Transmil: Transformer Based Correlated Multiple Instance Learning for Whole Slide Image Classification." Advances in neural information processing systems 34 (2021): 2136-2147.

[12] Wang, Xiyue, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. "Transformer-Based Unsupervised Contrastive Learning for Histopathological Image Classification." Medical image analysis 81 (2022): 102559.

[13] Chen, Richard J., Tong Ding, Ming Y. Lu, Drew FK Williamson, Guillaume Jaume, Andrew H. Song, Bowen Chen et al. "Towards a General-Purpose Foundation Model for Computational Pathology." Nature medicine 30, no. 3 (2024): 850-862.

[14] Vorontsov, Eugene, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann et al. "A Foundation Model for Clinical-Grade Computational Pathology and Rare Cancers Detection." Nature medicine 30, no. 10 (2024): 2924-2935.

[15] Lu, Ming Y., Bowen Chen, Drew FK Williamson, Richard J. Chen, Ivy Liang, Tong Ding, Guillaume Jaume et al. "A Visual-Language Foundation Model for Computational Pathology." Nature medicine 30, no. 3 (2024): 863-874.

[16] Ding, Tong, Sophia J. Wagner, Andrew H. Song, Richard J. Chen, Ming Y. Lu, Andrew Zhang, Anurag J. Vaidya et al. "A Multimodal Whole-Slide Foundation Model for Pathology." Nature medicine (2025): 1-13.

[17] Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." Advances in neural information processing systems 30 (2017).

[18] Kipf, Thomas N., and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks." arXiv preprint arXiv:1609.02907 (2016).

[19] Qin, Zhuanping, Wenhao Sun, Tinghang Guo, and Guangda Lu. "Colorectal Cancer Image Recognition Algorithm Based on Improved Transformer." Discover Applied Sciences 6, no. 8 (2024): 422

Appendix:

Abbreviations

Table A1. Revised abbreviation list

Abbreviation	Full form
CDGCV	Cross-Domain Graph Consistency Validation
GNN	Graph Neural Network
SHAP	SHapley Additive exPlanations
BFSRT	Boundary-Focused SHAP Refinement Test
BFI	Boundary Fidelity Index
MSRPA	Multi-Scale Relational Perturbation Assessment
DSGA	Dynamic SHAP-Graph Alignment
HGSE	Hierarchical Graph-SHAP Ensemble
WSI	Whole-Slide Image
MMD	Maximum Mean Discrepancy

Data Availability Statement

The colon histology and clinical imaging sets used in this investigation came from a variety of different centers and were obtained from public sources. Big image datasets from Zenodo, such as NCT-CRC-HE-100K and CRC VAL-HE-7K [1, 2], as well as dataset mirrors, can be utilized for the process. The files contain about one hundred thousand histology image patches that do not overlap with one another. Whole-slide photographs of tumor epithelium, stroma, lymphocytes, and normal mucosa set that have been stained with hematoxylin and eosin are the components constituting these slides. Additional benchmark datasets, such as Kather texture and colorectal histology, are made available via TensorFlow Datasets. These datasets are used to ensure that the procedure is effective with a diverse range of tissue morphologies and staining patterns. These publicly available image sources are utilized in the process of generating processed graph representations, feature embeddings, and derived annotations, which are then utilized in the evaluation and training of models. Histopathology images in their raw form are transformed into graph-structured representations and SHAP attribution maps. Huge volumes of intermediate data are produced as a result of this procedure. There is an example of processed graph topologies, feature matrices, and model inference outputs that can be used and duplicated. This is provided as additional information. Adjacency matrices, border attribution maps, and node-level feature vectors are some examples of these types of statistics that can be applied to specific datasets and samples.

When approached politely, the author in this instance is able to give fully processed datasets that contain graph-structured representations and SHAP attribution logs. Data that is many terabytes in size after graph formation, perturbation simulations, and institution-specific preparation procedures that are strongly related to the computer infrastructure are the factors that limit the amount of the data. When it comes to putting together identical datasets from public sources, the manuscript provides a great deal of information regarding how to employ pretreatment, network building strategies, and parameter combinations. The datasets used in this investigation were obtained from public sources that were either anonymous or de-identified in an ethical manner. With the consent of the institutional review board, the original histopathological records were made available to the public. However, informed permission was not obtained for retrospective analysis, which is in accordance with the best practices for managing clinical data. In the course of this research, samples of patient data are not distributed. Not only does this ensure that the data is used in an ethical manner, but it also ensures that the process of conducting research in computational pathology is both clear and repeaTable in the process.