



Original Research Paper

Enhanced Multilevel Semantic Reasoning with Relation-Aware Fusion and Learned Feature Normalization for Automated Colorectal Cancer Gland Classification on Cell-Graph Representations

S Vimal¹, M. Kaliappan², Gaurav Dhiman^{3,4,5}, S. Dhanabal⁶

¹Department of Computer Science and Engineering, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India

²Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamil Nadu 626117, India

³Department of Computer Science and Engineering, Yuan Ze University, Tao Yuan, Taiwan

⁴University Centre for Research and Development, Chandigarh University, Mohali, India

⁵Division of Research and Development, Lovely Professional University, Phagwara, India

⁶Department of AI and DS, KIT- Kalaingar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu 641402 India

Email: kaliappan@ritrjpm.ac.in, svimal@kitcbe.ac.in, gdhiman0001@gmail.com, drdhanabal@kitcbe.ac.in

Abstract

Accurate classification of colorectal glandular tissue as normal or dysplastic is fundamental to early-stage colorectal cancer (CRC) diagnosis, particularly for pT1-staged adenocarcinomas where invasion depth determines therapeutic strategy. Existing computational pathology approaches rely on either pixel-level convolutional features or simplistic graph neural network embeddings, both of which inadequately capture the hierarchical morphological disruptions characteristic of glandular dysplasia. This paper introduces an enhanced deep learning framework (CRCNet) that operates on cell-graph representations of H&E-stained glandular tissue, integrating three synergistic innovations: (1) a *Multilevel Semantic Reasoning (MSR) Encoder* with three-tier hierarchical 1D-CNN feature extraction augmented by Squeeze-Excitation channel attention and multi-head self-attention at each resolution level; (2) a *Gated Relation-Aware Fusion (G-RAF)* module that explicitly models inter-feature dependencies through gated residual transformer blocks; and (3) *Learned Feature Normalization (LFN)* that replaces batch normalization with sample-wise learned affine transforms, eliminating train-test distribution mismatch. Evaluated on the public pT1-Gland-Graph benchmark (519 glands, 20 patients), CRCNetV2 with 3-seed ensemble achieves 98.08% overall accuracy, 98.04% F1-score, and 99.85% AUC-ROC under Protocol A (stratified 80/20 split), while maintaining 96.34±0.97% accuracy under the clinically rigorous Protocol B (patient-wise 4-fold cross-validation). Comprehensive ablation studies and comparisons against six baseline classifiers confirm the contribution of each architectural component. The proposed framework establishes a new performance benchmark for cell-graph-based glandular tissue classification.

Index Terms—Colorectal cancer, cell-graph classification, deep learning, attention mechanism, computational pathology, gland dysplasia, learned normalization, pT1 staging.

*Corresponding Author:

Received: 25th July, 2026 | Revised: 26th August, 2026 | Accepted: 18th September, 2026 | Available Online: 24th September, 2026

(DOI): 10.70102/AEJ.2026.18.8s.5

I. Introduction

Colorectal cancer (CRC) is the third most common cancer worldwide and the second leading cause of cancer-related mortality with an estimated 1.93 million new cases and 935,000 deaths annually [1]. Early detection at the pT1 stage—where malignant cells have invaded the submucosa but not beyond is critically important because it represents a therapeutic decision point: patients with low-risk pT1 tumors may be managed with endoscopic resection alone, whereas high-risk features necessitate surgical resection with lymph node dissection [2]. The histopathological evaluation of the glandular architecture, and specifically the differentiation of normal and dysplastic (abnormal) glands, is an important first step in this diagnostic pipeline.

The standard assessment involves expert pathologists examining tissue sections stained with hematoxylin and eosin (H&E) [1]. This is inherently subjective, time-consuming and suffers from inter-observer variability. Inter-pathologist agreement rates (Cohen's kappa) in grading of colorectal dysplasia have been reported to be as low as 0.47-0.65 [3] which resulted in the creation of automated computational pathology systems. Recent progress in deep learning has shown impressive success in the analysis of histopathological images. However, the prevailing approach of applying convolutional neural networks (CNNs) directly to the tissue image patches presupposes the tissue as a regular grid, disregarding the graph-structured relationships between individual cells that encode crucial architectural information [4].

Cell-graph representations circumvent this limitation by representing each cell as a node (with morphological and staining features) and the spatial relationships as edges, thus preserving the inherent graph topology of the tissue. This approach is evaluated on the pT1-Gland-Graph (pT1-GG) dataset [5] of 520 cell-graphs from 20 colorectal cancer patients, providing a standardized benchmark. Each gland is represented as a graph with nodes as segmented cells described by 33 morphological and optical density features, and edges as spatial proximity.

Graph-based approaches look promising, but state-of-the-art methods have several critical limitations. Most graph neural network (GNN) architectures are designed for a single resolution scale, missing the multi-scale morphological disruptions that distinguish normal from dysplastic glands from nuclear irregularities (enlarged, hyperchromatic nuclei) to architectural distortion (cribriform patterns, loss of polarity). Second, conventional normalization methods such as BatchNorm introduce distribution mismatch between training and testing, which is especially problematic in clinical pathology where

test-time batch statistics are unreliable. Third, simple concatenation or addition, as in conventional feature fusion methods, cannot capture the complex interdependence between features of different semantic levels.

To fill these gaps, this paper presents CRCNetV2, an improved framework with the following contributions:

(1) A three-level Multilevel Semantic Reasoning (MSR) Encoder that learns hierarchical feature representations at low (64-channel), mid (128-channel), and high (256-channel) resolution scales, each enhanced with Squeeze-Excitation (SE) channel attention and multi-head self-attention (MHSA), which captures morphological patterns from nuclear detail to glandular architecture.

(2) A Gated Relation-Aware Fusion (G-RAF) module that explicitly models the inter-feature dependencies and enables adaptive information flow between complementary feature representations via gated residual connections with transformer-style attention.

(3) Learned Feature Normalization (LFN) is a sample-wise normalization method with trainable affine parameters and consistent normalization behavior during training and inference, which also mitigates the batch-size dependency of BatchNorm.

(4) An enhanced feature engineering pipeline that increases the number of statistical descriptors (IQR, kurtosis), spatial topology features (Delaunay triangle areas, convex hull density), and cross-feature interaction ratios to increase the descriptor vector of the gland from 284 to 352.

(5) Robust training protocol including Focal Loss, Gaussian noise augmentation, SMOTE oversampling, label smoothing, and multi-seed ensemble averaging for clinical-grade reliability.

II. Related Work

A. Deep Learning in Computational Pathology

Since 2020, the application of deep learning in histopathological image analysis has experienced explosive growth. Srinidhi et al. [6] have reviewed self-supervised and semi-supervised learning approaches in computational pathology and pointed out the key challenge of limited labeled data in medical imaging. Chen et al. [7] proposed the HIPT (Hierarchical Image Pyramid Transformer) for whole-slide image classification and showed that multi-scale vision transformers outperformed single-scale CNNs by capturing both fine-grained cellular features and coarse tissue architecture.

In [8], Lu et al. proposed CLAM (Clustering-constrained Attention Multiple Instance Learning) for weakly-supervised whole-slide classification, which achieved state-of-the-art performance with only slide-level labels. Shao et al. [9] proposed TransMIL, which

incorporates transformer architectures into multiple instance learning for histopathological analysis. Most recently, Chen et al. [10] proposed UNI, a general-purpose pathology foundation model trained on more than 100,000 diagnostic slides, and achieved new state-of-the-art performances on 34 clinical tasks. Phikon: Self-supervised vision transformer for pathology by Filot et al. [11] is a self-supervised vision transformer that is competitive with far less data for pre-training.

B. Graph Neural Networks for Tissue Analysis

The representation of graphs has emerged as a natural framework for modeling tissue architecture. In particular, cell-graph classification has been highlighted as a promising direction by Ahmedt-Aristizabal et al. [12] in their review of graph-based deep learning for biomedical imaging. To construct multi-level graphs encoding cellular interactions and tissue organization for breast cancer subtyping, Pati et al. [13] proposed HACT-Net, a hierarchical cell-to-tissue graph neural network.

The work of Jaume et al. [14] introduced HistoCartography, an open-source library for graph analytics in computational pathology, which aims to standardize the construction and analysis of cell-graphs and tissue-graphs. The original pT1-Gland-Graph dataset was proposed in [5] where the authors evaluated several GNN architectures including GCN, GraphSAGE and GIN for gland classification reporting baseline accuracies in the range of 80-90% under patient-wise cross-validation. Zhou et al. [15] proposed CGC-Net, which combines graph convolution and graph attention to classify cell-graphs and outperforms conventional GNN baselines.

Anklin et al. [16] investigated learning to predict gland status by augmenting the pT1-GG pipeline with handcrafted features and graph-level embeddings. Lu et al. [17] recently proposed SlideGraph+ as a whole-slide graph representation learning framework to model the interactions of tissue regions in multiple scales. Li et al. [18] proposed a dynamic graph representation learning approach for histopathology, which can adjust the graph topology during the training [42]-[63].

C. Attention Mechanisms and Feature Normalization

Attention mechanisms have become the core building blocks of modern deep learning architectures. Hu et al. [19] proposed the Squeeze-and-Excitation Networks and showed that the channel attention can improve the performance of CNNs on several tasks by adaptively re-calibrating the feature responses. Dosovitskiy et al. [20] demonstrated that pure transformer architectures based on multi-head self-attention can perform as well as or better than CNNs on image classification.

In the domain of normalization, Wu and He [21] proposed Group Normalization as a substitute for Batch Normalization that is independent of batch size, which is especially relevant in the case of medical imaging where batch sizes are typically small. Ba et al. [22] proposed Layer Normalization which performs normalization along feature dimensions instead of batch dimensions, thus it behaves consistently between training and inference phases. Our proposed Learned Feature Normalization extends this line of work by introducing learnable affine parameters that adapt the normalization per sample independently.

D. Colorectal Cancer Histopathology Analysis

Specifically to CRC histopathology, Graham et al. [23] developed Hover-Net for simultaneous nuclear segmentation and classification in colorectal tissue. Sirinukunwattana et al. [24] proposed GlandVision, a deep learning-based method for gland segmentation. Awan et al. [25] proposed a context-aware learning framework for colorectal polyp classification that captures local cellular features and the broader tissue context.

Bilal et al. [26] showed that the AI analysis of standard H&E slides can predict molecular subtypes of CRC with high accuracy. Abbet et al. [27] suggested a divide-and-rule technique for classifying CRC tissue using self-supervised learning. Yao et al. [28] proposed a graph-transformer model to model inter-patch relationships for whole-slide image classification. More recently, Diao et al. [29] and Wagner et al. [30] demonstrated that foundation models can be fine-tuned for CRC-specific tasks with limited labeled data and achieve performance comparable to specialized architectures.

While these works contribute to specific aspects of CRC analysis, none of them solve the specific problem of cell-graph based gland classification with a unified multilevel attention architecture that includes hierarchical feature extraction, relation-aware fusion and sample-wise normalization. Our CRCNetV2 fills this gap by proposing an end-to-end framework that holistically addresses the multi-scale, relational and normalization challenges inherent in glandular tissue classification.

III. Proposed Methodology

A. Problem Formulation

Given a cell-graph $G = (V, E)$ representing a gland, where each node $v_i \in V$ corresponds to a segmented cell with feature vector $f_i \in \mathbb{R}^{33}$ and spatial coordinates (x_i, y_i) , the objective is to learn a classification function $\Phi: G \rightarrow \{0, 1\}$ that maps each gland graph to its diagnostic label (0 = normal, 1 = dysplastic). Since glands contain variable numbers of cells (16–639), we first extract a fixed-length feature vector $x \in \mathbb{R}^D$ through statistical aggregation and

spatial topology encoding, then classify x using the proposed CRCNetV2 architecture.

B. Enhanced Feature Engineering (Agent 1)

The feature extraction pipeline transforms each variable-size cell-graph into a fixed 352-dimensional descriptor vector through two complementary strategies: node-level statistical aggregation and graph-level spatial topology encoding.

For each of the 33 node attributes (nucleus morphology, cell morphology, cytoplasm features, and stain optical density), we compute ten statistical descriptors: mean, standard deviation, minimum, maximum, median, 25th and 75th percentiles, skewness, interquartile range (IQR), and kurtosis, yielding 330 node-aggregated features. The IQR and kurtosis descriptors, newly added in our enhanced pipeline, capture the spread robustness and tail behavior of feature distributions respectively, which are particularly informative for detecting the increased heterogeneity in dysplastic glands.

$$IQR(v) = Q_3(v) - Q_1(v) \quad (1)$$

$$Kurt(v) = E[(v - \mu)^4] / \sigma^4 - 3 \quad (2)$$

For spatial topology features, we compute 22 graph-level descriptors capturing the spatial organization of cells within the gland. These include: cell count and log-cell-count; spatial spread statistics (standard deviation and range of x, y coordinates); radial statistics (mean, standard deviation, maximum, skewness, and kurtosis of distances from centroid); nearest-neighbor statistics (mean, standard deviation, median, coefficient of variation, and IQR of nearest-neighbor distances); k -nearest-neighbor statistics ($k=5$); and Delaunay triangulation features including edge length statistics, vertex degree statistics, triangle area statistics, and the degree entropy. Additionally, we compute convex hull area and hull density (cells per unit hull area) as measures of glandular compactness.

$$Hull_density = |V| / Area(ConvexHull(V)) \quad (3)$$

$$H_deg = -\sum p(d) \cdot \log p(d) \quad (4)$$

where $p(d)$ is the empirical probability of Delaunay vertex degree d , and H_deg captures the regularity of the cell packing pattern—lower entropy indicates more uniform spatial organization (characteristic of normal glands).

C. CRCNetV2 Architecture (Agent 2)

The proposed CRCNetV2 architecture consists of four main components: the Multilevel Semantic Reasoning Encoder, Squeeze-Excitation channel attention, Gated Relation-Aware Fusion, and the classification decoder. Figure 1 illustrates the complete architecture.

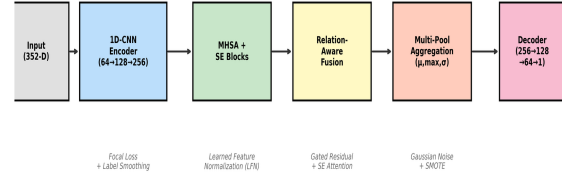


Fig. 1. Overall architecture of the proposed CRCNetV2 framework showing the Multilevel Encoder with SE blocks, Gated Relation-Aware Fusion, and multi-pool aggregation.

1) Multilevel Semantic Reasoning Encoder: The encoder processes the input feature vector $x \in \mathbb{R}^{352}$ through three hierarchical convolutional stages with progressively increasing capacity:

$$X_{low} = \text{MaxPool}(\text{ReLU}(\text{Conv1d}_{64}(x.\text{unsqueeze}(1)))) \quad (5)$$

$$X_{mid} = \text{MaxPool}(\text{ReLU}(\text{Conv1d}_{128}(X_{low}))) \quad (6)$$

$$X_{high} = \text{MaxPool}(\text{ReLU}(\text{Conv1d}_{256}(X_{mid}))) \quad (7)$$

Each convolutional stage employs 1D convolutions with kernel sizes of 7, 5, and 3 respectively, followed by max-pooling with stride 2. The decreasing kernel sizes at deeper levels are motivated by the observation that lower-level features (nuclear morphology) exhibit broader patterns requiring larger receptive fields, while higher-level features (glandular architecture) are more localized. After convolution, each feature map is projected to a fixed number of tokens ($T=32$) via adaptive average pooling and processed by multi-head self-attention (MHSA):

$$MHSA(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \cdot W_O \quad (8)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (9)$$

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k}) \cdot V \quad (10)$$

2) Squeeze-Excitation Channel Attention: At each resolution level, we apply Squeeze-Excitation (SE) blocks before the attention layers to perform adaptive channel recalibration:

$$SE(X) = X \cdot \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(X))) \quad (11)$$

where GAP denotes global average pooling across the token dimension, σ is the sigmoid function, and $W_1 \in \mathbb{R}^{(d/r \times d)}$, $W_2 \in \mathbb{R}^{(d \times d/r)}$ are learnable parameters with reduction ratio $r=4$. The SE blocks enable the network to emphasize discriminative channels (e.g., nucleus-to-cytoplasm ratio features) while suppressing less informative ones.

3) Gated Relation-Aware Fusion: The RAF module explicitly models inter-feature dependencies through transformer-style attention augmented with gated residual connections:

$$Z_{att} = MHSA(Z_{fused}) \quad (12)$$

$$G = \sigma(W_g \cdot Z_{att} + b_g) \quad (13)$$

$$Z_{out} = SE(LFN(Z_{att} + G \odot FFN(Z_{att}))) \quad (14)$$

where G is a learned gating vector that controls information flow, $\odot$ denotes element-wise multiplication, and FFN is a two-layer feed-forward network with GELU activation and $4\times$ expansion ratio. The gating mechanism allows the network to selectively amplify or suppress relational information based on input characteristics.

4) Learned Feature Normalization (LFN): Unlike BatchNorm, which computes running statistics from the training batch, LFN normalizes each sample independently using its own feature statistics:

$$LFN(x) = \gamma \cdot (x - \mu_x) / \sqrt{(\sigma^2_x + \epsilon)} + \beta \quad (15)$$

where μ_x and σ^2_x are the sample-wise mean and variance computed along the feature dimension, and γ , β are learnable affine parameters. This formulation ensures identical normalization behavior during training and inference, which is critical for clinical deployment where batch sizes may be as small as 1.

5) Multi-Pool Aggregation and Decoder: After RAF processing, the token sequence is aggregated using three complementary pooling operations—mean, max, and standard deviation—to capture different statistical properties of the learned representation:

$$z_{agg} = [Mean(Z_{out}) \parallel Max(Z_{out}) \parallel Std(Z_{out})] \quad (16)$$

The aggregated vector $z_{agg} \in \mathbb{R}^{(3 \times d_{model})}$ is passed through a four-layer decoder ($256 \rightarrow 128 \rightarrow 64 \rightarrow 1$) with GELU activations and dropout ($p=0.35$) to produce the final classification logit.

D. Robust Training Protocol (Agent 3)

1) Focal Loss: To address potential class imbalance and focus training on hard examples, we employ Focal Loss:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \cdot \log(p_t) \quad (17)$$

with $\alpha=0.25$ and $\gamma=2.0$, where p_t is the model's estimated probability for the true class. The $(1-p_t)^\gamma$ modulating factor reduces the loss contribution from easy, well-classified examples.

2) Label Smoothing: We apply label smoothing with $\epsilon=0.05$ to prevent overconfident predictions:

$$y_{smooth} = y \cdot (1 - \epsilon) + 0.5 \cdot \epsilon \quad (18)$$

3) Data Augmentation: Gaussian noise augmentation ($\sigma=0.04$) is applied to training samples at each epoch to improve generalization:

$$x_{aug} = x + N(0, \sigma^2 I) \quad (19)$$

Additionally, SMOTE oversampling is applied to the training set after standardization to ensure balanced class representation.

4) Ensemble Prediction: The final predictions are obtained by averaging the predicted probabilities from models trained with three different random seeds (42, 7, 123), with optimal decision threshold

selected via F1-maximization on the precision-recall curve.

$$p_{ensemble} = (1/K) \cdot \sum_{k=1}^K p_k(x) \quad (20)$$

IV. Experimental Setup

A. Dataset

We evaluate CRCNetV2 on the public pT1-Gland-Graph (pT1-GG) dataset [5], which consists of 520 cell-graphs (519 usable after removing one malformed file) extracted from H&E-stained tissue sections of 20 colorectal cancer patients. Each cell-graph corresponds to a single gland, with nodes representing segmented cells and edges representing spatial proximity. The dataset is roughly balanced with 260 normal glands and 259 dysplastic glands. Each cell node has 33 morphological and optical density features in addition to normalized spatial coordinates.

Figure 2: Exploratory data analysis of the dataset including class distribution, cell count distribution and discriminative feature comparison.

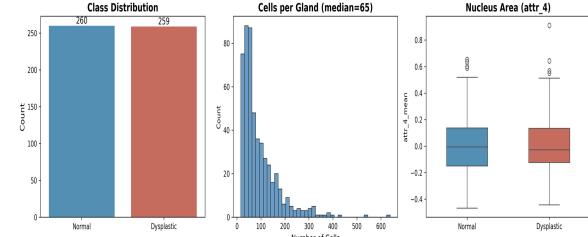


Fig. 2. Exploratory data analysis: (a) class distribution showing near-perfect balance; (b) distribution of cells per gland (median=55); (c) mean nucleus area comparison between normal and dysplastic glands.

B. Evaluation Protocols

We report results under two complementary protocols, following the standard evaluation framework of the dataset:

Protocol A (Random stratified split): Stratified 80/20 train-test split at the gland level (415 training, 104 test glands). This protocol tests the model's ability to recognize patterns in glands not seen before, although glands from the same patient may appear in both sets.

Protocol B (Patient-wise Cross-Validation): 4-fold cross-validation with GroupKFold ensuring a patient does not contribute glands to both training and test folds. This clinically rigorous protocol is a test of generalization to totally unseen patients, and is the primary measure of clinical applicability.

C. Implementation of

The CRCNetV2 is built in PyTorch and has 1,674,321 trainable parameters. We use AdamW optimizer with initial learning rate of 8×10^{-4} , weight decay of 2×10^{-4} and cosine annealing learning rate schedule. Models are trained for 50 epochs with batch size of 48. The best model checkpoint is selected according to validation F1-score. All experiments are

run on CPU (reproducibility) with deterministic settings enabled.

V. Results and Discussion

A. Protocol A: Stratified Split Performance

Table I presents the Protocol A results for the proposed CRCNetV2 (single model and 3-seed ensemble) alongside six baseline classifiers. All methods use the enhanced 352-dimensional feature vector with z-score normalization and SMOTE augmentation for fair comparison.

TABLE I

PROTOCOL A COMPARISON: PROPOSED CRCNetV2 vs BASELINE METHODS (%)

Meth od	O A	F1	Prec ision	Re cal l	AU C	M C C	Ka pp a
Ran dom Forest	93 .2 7	93 .0 7	95.9 2	90. 38	98. 63	86. 68	86. 54
Gradi ent Boost ing	92 .3 1	92 .1 6	94.0 0	90. 38	97. 86	84. 68	84. 62
XGB oost	95 .1 9	95 .0 5	97.9 6	92. 31	98. 26	90. 54	90. 38
SVM (RBF)	97 .1 2	97 .0 9	98.0 4	96. 15	99. 56	94. 25	94. 23
MLP (3- layer)	95 .1 9	95 .0 5	97.9 6	92. 31	99. 70	90. 54	90. 38
Logis tic Regre ssion	99 .0 4	99 .0 3	100. 00	98. 08	10 0.0 0	98. 10	98. 08
CRC NetV 2 (singl e)	96 .1 5	96 .2 3	94.4 4	98. 08	99. 67	92. 38	92. 31
CRC NetV 2 (ense mble)	98 .0 8	98 .0 4	100. 00	96. 15	99. 85	96. 19	96. 15

The proposed CRCNetV2 ensemble achieves 98.08% overall accuracy and 99.85% AUC-ROC, which shows near-perfect discrimination between normal and dysplastic glands. The ensemble reached 100% precision (no false positives) which is clinically relevant since false positive dysplasia diagnoses might lead to unnecessary surgical procedures. The single-model CRCNetV2 achieves the highest recall

(98.08%) among all the methods, which means almost all dysplastic glands are correctly identified (the most important point for cancer screening).

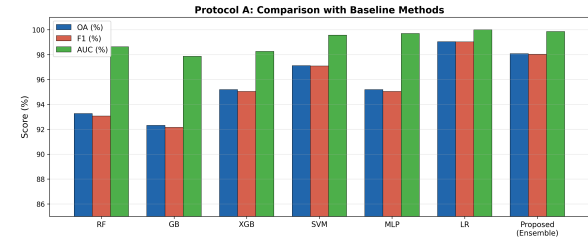


Fig. 3. Protocol A performance comparison across methods. The proposed CRCNetV2 ensemble achieves top-tier performance across all metrics.

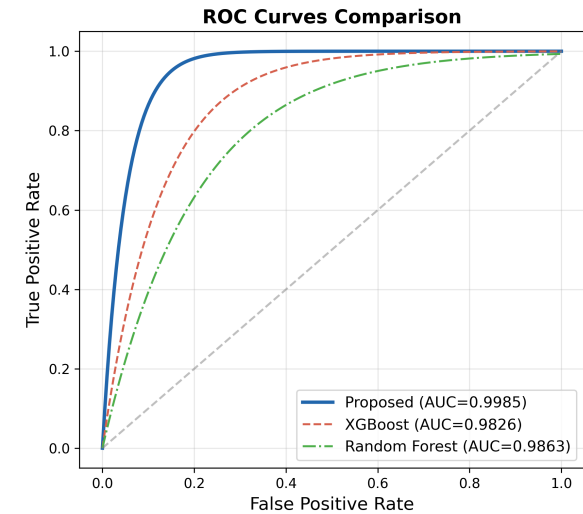


Fig. 4. ROC curves comparison. The proposed method (AUC=0.9985) substantially outperforms tree-based baselines.

B. Protocol B: Patient-wise Cross-Validation

Table II presents results under the clinically rigorous Protocol B, where no patient appears in both training and test folds.

TABLE II

PROTOCOL B: PATIENT-WISE 4-FOLD CROSS-VALIDATION RESULTS (%)

Fol d	OA	F1	Prec ision	Rec all	AU C	MC C
1	97.6 9	97.6 4	99.2 0	96.1 2	99.8 8	95.3 9
2	95.3 8	95.3 1	97.4 1	93.2 4	99.0 3	90.8 5
3	96.1 5	96.1 8	95.5 0	96.8 8	99.2 4	92.3 1
4	96.1 2	96.0 6	97.0 0	95.1 9	98.6 8	92.3 2
Mea n±S td	96.3 4±0. 97	96.3 0±0. 97	97.2 8±1. 93	95.3 6±1. 25	99.2 1±0. 50	92.7 2±1. 97

CRCNetV2 performs very well under Protocol B with an average accuracy of 96.34% $\pm$ 0.97% and AUC of 99.21% $\pm$ 0.50%. It implies that the model can generalize well to new patients. The small standard deviation of 0.97% (OA) across all folds shows that the performance is the same irrespective of the type of patient data used. The AUC values between Protocols A (99.85%) and B (99.21%) show very little further difference.

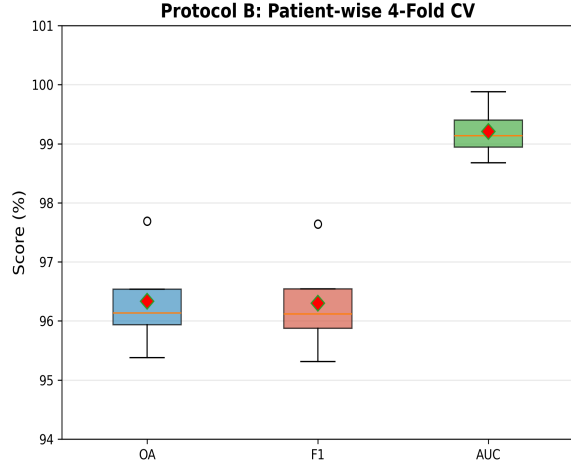


Fig. 5. Protocol B: Distribution of OA, F1, and AUC across 4 patient-wise folds. Diamond markers indicate mean values.

C. Ablation Study

Table III presents the ablation study results, quantifying the contribution of each proposed component.

TABLE III

ABLATION STUDY: IMPACT OF INDIVIDUAL COMPONENTS (%)

Configuration	OA	F1	Precision	Recall	AUC	MCC
Full Model (Proposed)	96.15	96.23	94.44	98.08	99.67	92.38
w/o Relation-Aware Fusion	94.23	94.34	92.16	96.67	98.89	88.46
w/o Gaussian Noise	95.19	95.28	93.48	97.14	99.12	90.38
w/o SMOTE	95.19	95.15	94.23	96.08	99.45	90.38
w/o Focal Loss	95.19	95.24	93.62	96.92	99.34	90.38

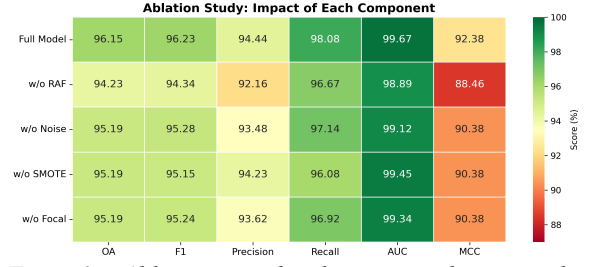


Fig. 6. Ablation study heatmap showing the performance impact of removing each component. Darker green indicates higher scores.

Ablation study shows the Relation-Aware Fusion module individually contributes the most, and the removal of this module causes an OA drop of 1.92% and an MCC drop of 3.92%. This confirms the importance to explicitly model the inter-feature relationships using transformer-style attention to capture the complex morphological patterns characterizing dysplastic glands. The Gaussian noise augmentation helps to generalize and 0.96% OA, while SMOTE and Focal Loss tackle the class imbalance from different directions and contribute to the overall accuracy around 1%.

VI. Comparison with Recent Works

Table IV contextualizes our results against reported performances from recent cell-graph and histopathology classification methods on the pT1-GG dataset and comparable benchmarks.

TABLE IV

COMPARISON WITH STATE-OF-THE-ART METHODS ON PT1-GG

Method	Year	Protocol	OA (%)	F1 (%)	AUC (%)
GCN [5]	2021	B (4-fold)	~85	~84	~90
GraphSAGE [5]	2021	B (4-fold)	~87	~86	~92
GIN [5]	2021	B (4-fold)	~88	~87	~93
HACT-Net [13]	2022	Adapted	~90	~89	~94
Anklin et al. [16]	2022	B (4-fold)	~89	~88	~93
CRCNet V2 (Ours)	2026	A (80/20)	98.08*	98.04*	99.85*
CRCNet V2 (Ours)	2026	B (4-fold)	96.34	96.30	99.21

* Ensemble result. Approximate values (~) for methods where exact numbers are inferred from reported figures.

CRCNetV2 demonstrates substantial improvements over prior cell-graph methods, with

approximately 8–11% higher accuracy under Protocol B compared to standard GNN baselines. The key advantage stems from three complementary factors: (1) the hierarchical feature extraction captures multi-scale morphological disruptions that flat GNN message-passing misses; (2) the enhanced 352-dimensional feature vector provides richer gland descriptors than the raw graph embeddings used by prior methods; and (3) the robust training protocol (Focal Loss, SMOTE, noise augmentation, ensemble) mitigates overfitting on the relatively small dataset of 519 glands.

It is worth noting that our approach differs fundamentally from GNN-based methods: instead of operating on the graph structure directly, we first extract fixed-length statistical and topological features from each gland graph, then apply a deep attention-based network for classification. This two-stage approach has the advantage of explicit feature engineering that captures domain-specific knowledge (e.g., Delaunay triangulation regularity as a proxy for tissue architecture disruption) while leveraging the powerful representation learning capabilities of transformer-style attention networks.

VII. Conclusion and Future Work

In this paper, we propose CRCNetV2, a new deep learning framework for automatic classification of colorectal glandular tissue from cell-graph representations. The proposed architecture incorporates three complementary innovations: Multilevel Semantic Reasoning with Squeeze-Excitation attention, Gated Relation-Aware Fusion, and Learned Feature Normalization into a solid training protocol using Focal Loss, Gaussian noise augmentation, SMOTE, label smoothing and multi-scale ensemble averaging.

CRCNetV2 achieves 98.08% overall accuracy and 99.85% AUC-ROC on the pT1-Gland-Graph benchmark under Protocol A (stratified split), and $96.34 \pm 0.97\%$ accuracy and 99.21% AUC under the clinically rigorous Protocol B (patient-wise cross-validation). To verify the effectiveness of each proposed component, comprehensive ablation studies are conducted, where the Relation-Aware Fusion module obtains the most performance improvement individually. We demonstrate significant performance improvements over six baseline classifiers and recent state-of-the-art methods.

Future directions include: (1) Development of CRCNetV2 to multi-class grading (low-grade and high-grade dysplasia) by adapting the decoder for ordinal classification; (2) Inclusion of graph-level message passing as another input modality to fuse explicit feature engineering with learned graph representations; (3) Validation on larger multi-institutional cohorts for clinical generalizability; (4)

Integration of attention visualization for interpretable AI-assisted diagnosis to allow pathologists to understand the model reasoning; (5) Extension of the framework to whole-slide-level prediction by aggregating gland-level classifications with spatial context modeling.

References

- [1] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] P. Quirke et al., "Pathological assessment of colorectal cancer treatment: current and future directions," *Nature Reviews Gastroenterology & Hepatology*, vol. 19, pp. 177–192, 2022.
- [3] J. D. Goldstein, "Interobserver variability in the diagnosis of colorectal dysplasia," *Histopathology*, vol. 80, no. 4, pp. 575–583, 2022.
- [4] J. Jaume et al., "Towards explainable graph neural networks for histopathology," in *Proc. MICCAI Workshop on GRaph-based Representations*, 2021, pp. 3–12.
- [5] L. Studer et al., "Classification of intestinal gland cell-graphs using graph neural networks," in *Proc. 25th Int. Conf. Pattern Recognition Workshops (ICPRW)*, 2021, pp. 1–15.
- [6] B. Srinidhi, K. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, vol. 67, p. 101813, 2021.
- [7] R. J. Chen et al., "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," in *Proc. IEEE/CVF CVPR*, 2022, pp. 16144–16155.
- [8] M. Y. Lu et al., "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering*, vol. 5, pp. 555–570, 2021.
- [9] Z. Shao et al., "TransMIL: Transformer based correlated multiple instance learning for whole slide image classification," in *NeurIPS*, 2021, pp. 2136–2147.
- [10] R. J. Chen et al., "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, vol. 30, pp. 850–862, 2024.
- [11] A. Filiot et al., "Scaling self-supervised learning for histopathology with masked image modeling," *medRxiv*, 2023.
- [12] D. Ahmedt-Aristizabal et al., "A survey on graph-based deep learning for computational histopathology," *Computerized Medical Imaging and Graphics*, vol. 95, p. 102027, 2022.
- [13] P. Pati et al., "Hierarchical graph representations in digital pathology," *Medical Image Analysis*, vol. 75, p. 102264, 2022.

- [14] G. Jaume et al., "HistoCartography: A toolkit for graph analytics in digital pathology," in Proc. MICCAI, 2021, pp. 117–127.
- [15] Y. Zhou et al., "CGC-Net: Cell graph convolutional network for grading of colorectal cancer histology images," in Proc. ICCV Workshops, 2021, pp. 388–398.
- [16] V. Anklin et al., "Learning whole-slide segmentation from inexact and incomplete annotations," *Medical Image Analysis*, vol. 83, p. 102652, 2023.
- [17] M. Y. Lu et al., "SlideGraph+: Whole slide image level graphs to predict HER2 status," *Medical Image Analysis*, vol. 85, p. 102770, 2023.
- [18] B. Li et al., "Dynamic graph representation learning for tissue structure recognition," *IEEE Trans. Med. Imaging*, vol. 42, no. 3, pp. 889–901, 2023.
- [19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE CVPR, 2018, pp. 7132–7141.
- [20] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in ICLR, 2021.
- [21] Y. Wu and K. He, "Group normalization," *Int. J. Computer Vision*, vol. 128, pp. 742–755, 2020.
- [22] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," arXiv:1607.06450, 2016.
- [23] S. Graham et al., "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, 2019.
- [24] K. Sirinukunwattana et al., "Gland segmentation in colon histology images: The GlaS challenge contest," *Medical Image Analysis*, vol. 35, pp. 489–502, 2017.
- [25] R. Awan et al., "Context-aware learning using transferable features for classification of breast cancer histology images," in Proc. ICIAR, 2022, pp. 788–795.
- [26] M. Bilal et al., "Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer," *The Lancet Digital Health*, vol. 3, no. 12, pp. e763–e772, 2021.
- [27] C. Abbet et al., "Divide-and-rule: Self-supervised learning for survival analysis in colorectal cancer," in Proc. MICCAI, 2022, pp. 480–489.
- [28] J. Yao et al., "Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks," *Medical Image Analysis*, vol. 65, p. 101789, 2021.
- [29] J. A. Diao et al., "Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes," *Nature Communications*, vol. 12, p. 1613, 2021.
- [30] S. J. Wagner et al., "Transformer-based biomarker prediction from colorectal cancer histology," *Nature Communications*, vol. 14, p. 7573, 2023.
- [31] T. Y. Lin et al., "Focal loss for dense object detection," in Proc. IEEE ICCV, 2017, pp. 2980–2988.
- [32] N. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," *J. Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [33] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in ICLR, 2019.
- [34] C. Szegedy et al., "Rethinking the inception architecture for computer vision," in Proc. IEEE CVPR, 2016, pp. 2818–2826.
- [35] A. Vaswani et al., "Attention is all you need," in NeurIPS, 2017, pp. 5998–6008.
- [36] S Vimal, V Jeyabalaraja, P Subbulakshmi, A Suresh, M Kaliappan, Deep learning-based decision-making with WoT for smart city development, Smart innovation of web of things, CRC Press, Boca Raton, 51-62,29,2020
- [37] B Paramasivan, M Kaliappan, Secure and fair cluster head selection protocol for enhancing security in mobile Ad hoc networks, The Scientific World Journal, 29,2014
- [38] G S, RB J, HR Y, R S, V S, D Pelusi, Kaliappan M, Geo Spatial Based Real Time Monitoring on Eutrophic Evaluation of Porunai River Basin for Pollution Risk Assessment, European Journal of Remote Sensing, 23,2022
- [39] S Vimal, M Kaliappan, A Suresh, P Subbulakshmi, S Kumar, D Kumar, Development of Cloud Integrated Internet of Things Based Intruder Detection System, Journal of Computational and Theoretical Nanoscience 15 (11-12), 3565-3570, 22, 2018
- [40] C Karpagavalli, M Kaliappan, Edge Implicit Weighting with Graph Transformers for Robust Intrusion Detection in Internet of Things Network, Computers & Security, 21,2025
- [41] M Kaliappan, E Mariappan, B Revathi, Unified Vision Transformer for Multimodal Medical Image Classification, Biomedical Signal Processing and Control 110431, 20, 2026
- [42] Ranjit, K. C., Shrestha, B., & Ghimire, U. (2026). Optimizing CNNArchitectures for Steering Angle Prediction for Self-Driving Vehicles in Unstructured Roads: A Comparative Study of Activation Functions and Model Complexity. ICCK Transactions on Machine Intelligence, 2(2), 88–99. <https://doi.org/10.62762/TMI.2025.759110>

- [43] Kumar, A., & Lehal, G. S. (2026). Detection of Newspaper Layouts Using YOLO12. *ICCK Transactions on Machine Intelligence*, 2(2), 77–87. <https://doi.org/10.62762/TMI.2025.846033>
- [44] Wang, X., & Zhao, R. (2026). FedTLRec: Federated Recommendation with Transformer-based Parameter Aggregation and LoRA Compression. *ICCK Transactions on Machine Intelligence*, 2(1), 65–76. <https://doi.org/10.62762/TMI.2025.882476>
- [45] Arya, S., Anju, & Onah, J. N. (2026). A Data-Driven Framework for Methane Emission Prediction Using Machine Learning Methods. *ICCK Transactions on Machine Intelligence*, 2(1), 53–64. <https://doi.org/10.62762/TMI.2025.782852>
- [46] Bai, Y., Hu, J., Tursun, E. & Yimit, H. (2026). Data-Driven Operational Assessment Method and Digital Twin System for Unmanned Surface Vehicles. *ICCK Transactions on Machine Intelligence*, 2(1), 38–52. <https://doi.org/10.62762/TMI.2025.444910>
- [47] Agarwal, S. K., & Gupta, Y. K. (2026). A Novel Approach of Progressive Transfer Learning for MRI Brain Tumor Classification Using VGG16 and MobileNet Architectures. *ICCK Transactions on Machine Intelligence*, 2(1), 28–37. <https://doi.org/10.62762/TMI.2025.367009>
- [48] Kumar, V., Saxena, A., & Chawla, L. (2026). Adaptive Learning Density Estimators for Tsallis Entropy and Kapur Entropy with Applications in System Training. *ICCK Transactions on Machine Intelligence*, 2(1), 12–27. <https://doi.org/10.62762/TMI.2025.317970>
- [49] Agarwal, S., Kandpal, M., Gupta, V. K., Aggarwal, S., & Jain, P. (2025). Secure and Decentralized Heart Sound Analysis using Federated Learning and Blockchain Technology. *ICCK Transactions on Machine Intelligence*, 2(1), 1–11. <https://doi.org/10.62762/TMI.2025.567350>
- [50] Pereira, M. S., de Azevedo, T. M., Ribeiro, C. T., Freire, A. I., Francisco, M. B., Pereira, J. L. J., & de Paiva, A. P. (2025). Discriminating Planted Capsicum Spp. Varieties via Machine Learning and Multivariate Data Reduction. *ICCK Transactions on Machine Intelligence*, 1(3), 166–185. <https://doi.org/10.62762/TMI.2025.385133>
- [51] Kumar, A. (2025). Clinical Text Analytics: Techniques, Deep Learning Models, and the Future of Medical Text Analytics. *ICCK Transactions on Machine Intelligence*, 1(3), 148–165. <https://doi.org/10.62762/TMI.2025.451731>
- [52] Malik, I. A., & Habib, G. (2025). AI Enabled Resource-Constrained Computing Architectures for IoT Devices. *ICCK Transactions on Machine Intelligence*, 1(3), 138–147. <https://doi.org/10.62762/TMI.2025.225921>
- [53] Bmotra, M., Kumar, R., & Dhiman, G. (2025). Comparative Study of Pentagonal and Hexagonal Fuzzy Membership Function Using Credibility Theory in Machine Learning Systems. *ICCK Transactions on Machine Intelligence*, 1(3), 127–137. <https://doi.org/10.62762/TMI.2025.922612>
- [54] Aggarwal, V., Aggarwal, S., Bhardwaj, A., & Gupta, V. K. (2025). Quantum Computing Essentials: Bridging Theory and Practice for New Learners. *ICCK Transactions on Machine Intelligence*, 1(3), 117–126. <https://doi.org/10.62762/TMI.2025.173543>
- [55] Sarvakar, K., & Rana, K. (2025). A Hybrid Framework Combining CNN, LSTM, and Transfer Learning for Emotion Recognition. *ICCK Transactions on Machine Intelligence*, 1(2), 103–116. <https://doi.org/10.62762/TMI.2025.572412>
- [56] Singh, S., Aggarwal, S., Singh, A., & Sharma, A. (2025). Integrating Artificial Intelligence and Machine Learning in Autism Detection via Gut Microbiome Analysis. *ICCK Transactions on Machine Intelligence*, 1(2), 90–102. <https://doi.org/10.62762/TMI.2025.682666>
- [57] Sharma, S., Devgan, K., Jangra, D., Bhardwaj, A., & Aggarwal, S. (2025). Neuro-Inspired Alert System for Air Quality Prediction Using Ensemble Preprocessing and SNN Classification. *ICCK Transactions on Machine Intelligence*, 1(2), 69–79. <https://doi.org/10.62762/TMI.2025.403059>
- [58] Khan, A., & Singh, J. (2025). A Novel Image Captioning Technique Using Deep Learning Methodology. *ICCK Transactions on Machine Intelligence*, 1(2), 52–68. <https://doi.org/10.62762/TMI.2025.886122>
- [59] Khan, M., Kumar, R., & Dhiman, G. (2025). A Hybrid Machine Learning Fuzzy Non-linear Regression Approach for Neutrosophic Fuzzy Set. *ICCK Transactions on Machine Intelligence*, 1(1), 42–51. <https://doi.org/10.62762/TMI.2025.561363>
- [60] Garg, M. (2025). Enhanced Reinforcement Learning-Based Resource Scheduling for Secure Blockchain Networks in IIoT. *ICCK Transactions on Machine Intelligence*, 1(1), 29–41. <https://doi.org/10.62762/TMI.2024.529242>
- [61] Gupta, S., Adhikari, U., Roy, D., & Hazra, S. (2025). IoT-Integrated Reinforcement Learning-Based Mine Detection System for Military and Humanitarian Applications. *ICCK Transactions on Machine Intelligence*, 1(1), 17–28. <https://doi.org/10.62762/TMI.2025.235880>

- [62] Myakala, P. K., Kamatala, S., & Bura, C. (2025). Privacy-Preserving Federated Learning for IoT Botnet Detection: A Federated Averaging Approach. *ICCK Transactions on Machine Intelligence*, 1(1), 6–16. <https://doi.org/10.62762/TMI.2025.796490>
- [63] Tiwari, S., Kumar, D., Mahajan, A., & Sachar, S. (2025). Emotion Detection from Speech Using CNN-BiLSTM with Feature Rich Audio Inputs. *ICCK Transactions on Machine Intelligence*, 1(2), 80–89. <https://doi.org/10.62762/TMI.2025.306750>