Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Deep Learning-Driven Subtype Classification of Colorectal Cancer Using Histopathology Images for Personalized Medicine

Nghia Trong Le Phan[1] | Mqondisi Fortune Mavuso[2] | Phuc Nguyen Thien Dao[3] | Huy Ngoc Anh Nguyen[4] | Tien Thuy Mai[4] | Kiep Thi Quang[4] | Thang Truong Le[2,5]

[1]FPT University, Can Tho City, Vietnam
[2]Smart Medicine and Health Informatics Program, International College, National Taiwan University, Taipei, Taiwan
[3]Faculty of Biology-Biotechnology, University of Science, VNUHCM, Ho Chi Minh City, Vietnam
[4]Hung Vuong High School for the Gifted, Pleiku, Gia Lai, Vietnam
[5]School of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

Correspondence
*Corresponding author: Thang Truong Le, Email: letruongthangg@gmail.com

Present Address
Smart Medicine and Health Informatics Program, International College, National Taiwan University, Taipei, Taiwan

Summary

Precise and timely diagnosis of colorectal cancer (CRC) is vital for enhancing patient outcomes. Histopathological examination of tissue samples remains the gold standard for CRC diagnosis, but it is a time-consuming and subjective approach tending to inter-observer variability. This study explores the use of deep learning, particularly ResNet architectures, for the automated classification of CRC using histopathology images. Our research focuses on evaluating different ResNet models (ResNet-18, ResNet-34, ResNet-50) to extract relevant visible characteristics. Additionally, we use Grad-CAM heatmaps to understand the model's areas of focus, ensuring alignment with established diagnostic criteria. To address limited data availability, we examine data augmentation techniques to improve the model's adaptability. Our analysis indicates that ResNet-34 strikes a balance between model complexity and performance, demonstrating notable accuracies of 91.10%, 99.11%, and 100.00% for overall, top-2, and top-3 accuracy classifications, respectively, outperforming both shallower (ResNet-18) and deeper (ResNet-50) models. This indicates that a moderate depth is practical in capturing features of CRC images. Our findings have significant importance for the development of an interpretable AI-assisted diagnostic tool for CRC, with the potential to improve the efficiency and accuracy of pathologists. This approach aims to automate image analysis, offer insights into model decisions, and ultimately improve diagnostic consistency and patient care in oncology.

KEYWORDS:
Colorectal Cancer (CRC) classification, Deep learning, ResNet architecture, Histopathology Images

1 | INTRODUCTION

Colorectal cancer stands as the third most commonly diagnosed malignancy worldwide, with approximately 1.9 million new cases and 900,000 fatalities reported in 2020 alone. Representing 10% of all new cancer diagnoses and accounting for 9.4% of all cancer deaths globally, it trails only behind lung cancer in mortality rates. The burden of this disease is expected to escalate, with projections indicating a rise to 3.2 million new cases by 2040. This increase is driven by aging populations, which are more susceptible to malignancies, and socio-economic development that influences lifestyle factors such as diet and physical activity [1].

The classification of colon and rectal cancers under the umbrella term "colorectal cancer" is based on the similarities in their pathological and clinical profiles[2]. Histopathological examination is pivotal, not only for diagnosing colorectal cancer but also for determining the appropriate treatment strategies. This involves meticulous diagnostic imaging and histological analysis of primary or metastatic sites to confirm diagnoses before treatment initiation[3 4].

In recent years, Artificial Intelligence (AI), particularly through deep learning algorithms, has reshaped the healthcare landscape by enhancing diagnostic accuracy, streamlining drug research and development, and improving patient management with applications such as telemedicine and automated medical record transcription. The capability of AI to parallel human accuracy in healthcare tasks suggests a future where it might assume roles traditionally held by healthcare professionals[5]. This transition is particularly evident in early-stage colorectal cancer management, where AI-driven interventions have significantly improved survival outcomes[6 7].

Understanding the heterogeneity of colorectal cancer is paramount as it directly influences treatment decisions and patient outcomes. The disease manifests in various histopathological subtypes, each with unique molecular signatures and clinical behaviors, which necessitate different therapeutic approaches. The traditional one-size-fits-all treatment approach is rapidly becoming obsolete as targeted therapies show significantly higher efficacy when matched with specific cancer subtypes. For instance, patients with tumors exhibiting high microsatellite instability may benefit more from immunotherapy compared to those with microsatellite-stable tumors, who may respond better to chemotherapeutic agents[8].

The ability to accurately classify colorectal cancer into its respective subtypes using machine learning and histopathology images marks a significant step towards personalized medicine. By integrating AI into the diagnostic process, clinicians can achieve a more precise understanding of tumor behavior at a granular level, leading to optimized treatment plans tailored to the individual characteristics of each cancer patient. This not only enhances the potential for successful treatment outcomes but also minimizes the risk of adverse effects associated with treatments that are less likely to be effective. Furthermore, the application of machine learning in this field can substantially reduce the time and cost associated with traditional diagnostic methods, making high-quality healthcare more accessible and efficient[9].

The complexity of colorectal cancer treatment is compounded by the heterogeneity of its subtypes, each with distinct biological characteristics, treatment responses, and prognoses. For instance, adenocarcinomas typically exhibit slower growth and respond favorably to conventional treatments such as chemotherapy and radiation, whereas squamous cell carcinomas present more aggressive behavior and resistance to these therapies[10].

This study leverages the EBHI-Seg dataset, encompassing six distinct tumor differentiation stages tissue, polyps, low-grade intraepithelial neoplasia, high-grade intraepithelial neoplasia, serrated adenoma, and adenocarcinoma. Utilizing advanced machine learning algorithms, we aim to develop a robust model capable of classifying colorectal cancer subtypes based solely on histopathology images. Our approach not only seeks to augment diagnostic accuracy but also to tailor treatment protocols more effectively, thereby enhancing patient outcomes and facilitating personalized medicine. Notably, the use of machine learning and deep learning techniques to classify subtypes of colorectal cancer using HE images has not been previously reported in the literature, marking a significant step forward in the field of medical diagnostics.

The research article is structured into four main sections. In section (2), we provide a collection of related research used as references. Section (3) outlines the methodology, detailing all processes utilized in the study. Section (4) clarifies the experiments, conduction, and the precision calculation of the proposed architecture. Finally, in section (5), we conclude and reconsider the critical domains linked to the research.

2 | RELATED RESEARCH

Colorectal cancer remains a leading cause of morbidity and mortality worldwide, affecting both men and women with varying incidence rates across different regions. Recognizing colorectal cancer in its early stages holds crucial importance today. Technological advancements have contributed to the development of various methodologies encouraging the classification of colorectal cancer images, thereby simplifying the diagnostic process for medical doctors.

In the study performed by Devvi Sarwinda[11], a deep learning method using the ResNet architecture was analyzed to detect colorectal cancer. ResNet-18 and ResNet-50 were trained on colon gland images to distinguish between benign and malignant colorectal cancer. The results illustrated that ResNet-50 demonstrated prominent performance in terms of accuracy (above 80%), sensitivity (above 87%), and specificity (above 83%) compared to ResNet-18. Ahmad[12] presented innovative criteria for diagnosing colorectal cancer cells by addressing issues related to color variations in microscope-captured images. The utilization

of HSV color enabled the reduction of elements outside the nucleus area, and subsequent gland tracking boundary and segmentation were applied to remove gland shapes. The detection of gland shapes was performed using a Multilayer Perceptron. The integration of gland shape and nucleus size in the image classification process yielded an impressive 94% accuracy using those proposed methods. Bychkov[13] operated a combination of convolutional and recurrent architectures to train a deep network predicting colorectal cancer outcomes based on tumor tissue sample images. Notably, this approach directly predicted patient results without any middle tissue classification. The research involved 420 colorectal cancer patients with clinicopathological and outcome data. The results showed that the deep learning-based outcome prediction, using only small tissue areas as input, outperformed visual histological assessments conducted by human experts at both TMA spot and whole-slide levels in stratifying patients into low-risk and high-risk categories.

In modern years, the increasing volume of medical images yields a problematic challenge in practical classification. Deep learning, particularly Convolutional Neural Networks (CNNs), appears as a advantageous solution for handling this challenge. Improving classification precision and optimizing network efficiency represent primary focal matters when operating deep learning for the classification of biological images. Mehdi[14] utilized fine-tuned pre-trained deep neural networks to predict breast cancer images. The CNN architecture reached an impressive average accuracy of 99.8% across four cancer types (breast, bladder, lung, and lymphoma) using the ResNet-50 pre-trained model. Moreover, ResNet-50 and ResNet-152 demonstrated high accuracy in classifying benign and malignant subtypes (94.8% and 96.4%, respectively) for specific breast cancer classifications. Jeyaraj[15] developed a deep learning algorithm for an automated, computer-aided oral cancer detection approach. Their approach presented a novel structure of a partitioned deep Convolutional Neural Network (CNN) with two partitioned layers for labeling and classifying regions of interest in multidimensional hyperspectral images. The research achieved classification accuracy of 91.4% for benign tissue and 94.5% for normal tissue, using 100 and 500 training patterns, respectively. Tsaku[16] implemented deep convolutional neural networks in histopathological image research, introducing a cancer texture-based deep neural network named CAT-Net. This network comprehended scalable morphological features from histopathological Whole Slide Images (WSIs), capturing uniform spatial patterns through convolutional layers and enhancing predictive performance while reducing model complexity. CAT-Net surpassed existing benchmark methodologies in terms of accuracy, precision, recall, and F1-score, achieving the highest accuracy of 0.9613 ± 0.0029 (mean $\pm$ std), precision of 0.9653 ± 0.0052, recall of 0.9617 ± 0.0031, and F1-score of 0.9625 ± 0.0055, respectively.

Further advancing the field, "Color-CADx" represents a novel breakthrough in computer-aided diagnostics (CAD) for colorectal cancer (CRC), utilizing a sophisticated combination of deep learning models like ResNet50, DenseNet201, and AlexNet[17]. Employing ResNet50, DenseNet201, and AlexNet, the system enhances classification through end-to-end training at various ratios. Feature dimensionality is reduced via discrete cosine transform (DCT), which also aids in extracting and selecting significant spectral features. The analysis of variance (ANOVA) is used further to refine feature selection, achieving high classification accuracies of 99.3% on the NCT-CRC-HE-100 K dataset and 96.8% on the Kather_texture_2016_image_tiles dataset. These results underscore the efficiency of combining multiple CNNs and DCT in CRC diagnostics.

In personalized medicine, merely distinguishing between cancerous and non-cancerous tissues or between primary and malignant tumors addresses only a fraction of cancer management; it is crucial to classify cancer subtypes with distinct clinical outcomes to tailor and optimize treatment strategies effectively. Particularly with the use of machine learning (ML) and deep learning (DL) models, there has been a notable focus on leveraging histopathological data for precision medicine. High-grade serous ovarian cancer (HGSOC), characterized by its poor prognosis and diverse histopathological appearances, presents significant challenges in treatment decisions. Recent advances have demonstrated that ML and DL models can significantly improve the accuracy of HGSOC subtype classification[18]. For instance, studies have successfully used whole slide imaging (WSI) to digitally analyze tumor samples, facilitating the histopathological classification of HGSOC into subtypes such as mesenchymal, proliferative, immunoreactive, and differentiated. This approach has yielded impressive results, with the classification models achieving high accuracies, which are crucial for selecting appropriate, subtype-specific therapies that can potentially improve patient outcomes. In one highlighted study, the application of digital pathology combined with ML techniques led to a classification accuracy of up to 99.3% on specific datasets, such as the NCT-CRC-HE-100 K. In lung cancer (LC) diagnostics, a notable study has enhanced the precision of machine learning (ML) in classifying lung cancer subtypes, including the less common lung adenosquamous carcinoma (ASC)[19]. The study utilized a comprehensive dataset of 121 histopathological images, applying the Relief algorithm for feature selection and a Support Vector Machines (SVMs) classifier for subtype differentiation. This approach yielded results: 73.91% accuracy for LUSC-ASC, 83.91% for LUSC-SCLC, and 73.67% for ASC-SCLC. The performance of the classifier model, assessed through the receiver operating characteristic (ROC) curve and area under the curve (AUC), demonstrated its robust generalization capability. This advancement underscores the potential of ML to significantly improve diagnostic

accuracy and efficiency in lung cancer, thereby supporting more personalized treatment strategies.These advancements not only enhance the precision of subtype classification but also support the broader implementation of personalized therapy protocols in clinical settings, demonstrating the transformative potential of ML and DL in oncological diagnostics.

3 | METHODOLOGY

3.1 | The research implementation procedure

In our research, the implementation process will consist of eleven primary steps. These steps will be outlined in the accompanying flowchart (Fig. 1) and further elaborated upon below.

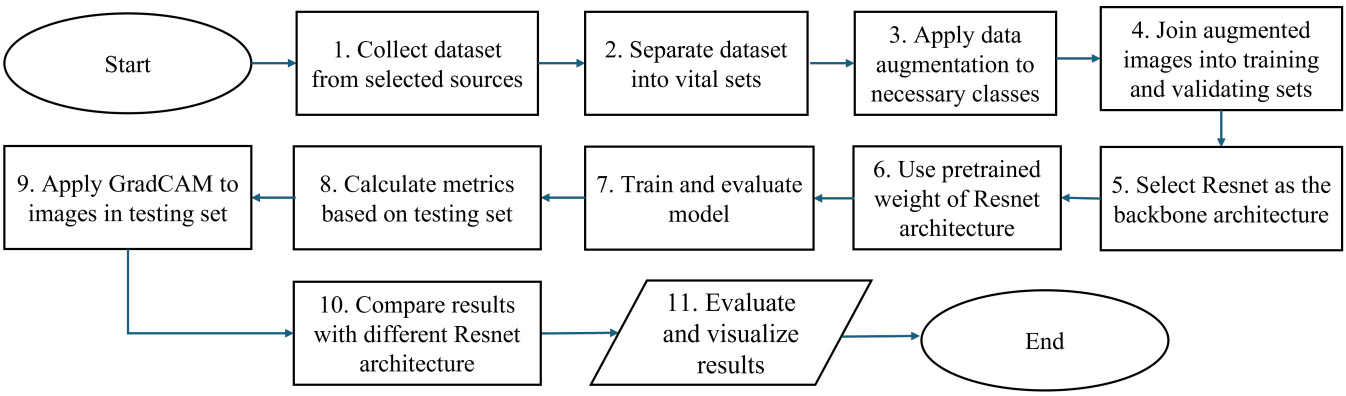

FIGURE 1 The implementing procedure flowchart

The implementation of the research involves the following steps:

1. Collect Dataset from Selected Sources: The initial step entails procuring histopathology images. Collaboration with hospitals or pathology labs is important to get de-identified image data of colorectal cancer cases and healthy controls. Alternatively, publicly available datasets may be used, provided diversity and meticulous annotation by experienced pathologists.

2. Separate Dataset into Vital Sets: Upon collection, the dataset needs to be partitioned into three vital sets:

   - Training Set: This constitutes the largest portion of the data, used to train the model and enable it to recognize patterns and features associated with colorectal cancer based on the information presented in the images.
   - Validation Set: This set facilitates fine-tuning of the model during training. The model's performance is assessed on the validation set to prevent overfitting and ensure robust generalization to unseen data.
   - Testing Set: This set is kept separate from the training and validation sets and is used for the final evaluation of the model's performance on unseen data subsequent to completion of training.

3. Apply Data Augmentation to Necessary Classes: To address limited data availability, augmentation techniques are specifically applied to classes with fewer images to create a more balanced dataset. This facilitates effective learning from all samples.

4. Join Augmented Images into Training and Validating Sets: The augmented images are merged into the training and validation sets, thereby expanding the size and diversity of the training data and enhancing the robustness of the model.

5. Select ResNet as the Backbone Architecture: ResNet architectures are chosen as the foundation for the model due to their demonstrated success in image recognition tasks and their capability to address the vanishing gradient problem, enabling efficient training of deep convolutional neural networks (CNNs).

6. Use pre-trained Weight of the Resnet Architecture: Pre-trained ResNet models are utilized as a starting point, having been trained on extensive image datasets like ImageNet. The final classification layers of these pre-trained models are removed, and new layers specific to the colorectal cancer classification task are added. This enables the model to leverage pre-learned features from general image recognition tasks and fine-tune them for the specific task of identifying colorectal cancer in histopathology images.

7. Train and Evaluate Model: The model is then trained on the augmented training set, learning to identify features in the images. The validation set is utilized to monitor the model's performance and prevent overfitting.

8. Calculate Metrics Based on Testing Set: Upon completion of training, the model's performance is evaluated on the unseen testing set. Metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate the model's ability to accurately categorize colorectal cancer cases.

9. Apply Grad-CAM to Images in Testing Set: Grad-CAM is used to generate heatmaps highlighting the image regions on which the model focuses for classification, providing insights into the model's decision-making process.

10. Compare Results with Different ResNet Architecture: The training and evaluation process is conducted with different ResNet architectures, such as ResNet-18, ResNet-34, and ResNet-50. By comparing performance metrics and Grad-CAM visualizations, the goal is to identify the optimal ResNet architecture for the specific dataset and classification task.

11. Evaluate and Visualize Results: Lastly, the research analyzes the overall performance of the models, considering accuracy, generalizability, and interpretability through Grad-CAM visualizations. These results are presented visually using graphs, tables, and heatmaps to effectively demonstrate the effectiveness of the proposed model with ResNet architectures for colorectal cancer classification using histopathology images.

3.2 | Data Augmentation

Training deep learning models for colorectal cancer classification using histopathology images poses a significant challenge due to the limited availability of real-world data. This lack can result in overfitting, wherein the model remembers specific training data instead of acquiring generalizable features for precise classification of unseen images. To mitigate this challenge, data augmentation techniques are appropriate in artificially expanding the training dataset and enhancing model performance. Data augmentation contains the generation of new variations of existing images within the dataset, simulating natural variations present in real-life histopathology images. Common data augmentation techniques will be explained in Fig. 2.

One common technique for image augmentation is image rotation, which can enhance our model's ability to handle changes in the orientation of objects. Suppose we have a point $A(x_0, y_0)$ representing any pixel in an image. After rotating the image by an angle θ, the point A will have a new coordinate (x_1, y_1). The new coordinates x_1 and y_1 can be calculated using the formulas specified in Equation 1 and 2 below.

$$x_1 = a \times \cos\theta - b \times \sin\theta \tag{1}$$

$$y_1 = a \times \sin\theta + b \times \cos\theta \tag{2}$$

In the domain of image manipulation, flipping is a fundamental operation that extends the concept of rotation. It enables us to horizontally and vertically mirror an image. Consider a point $B(x_0, y_0)$ representing a pixel with coordinates within an image of width w_i and height h_i. Upon performing a flip operation, the coordinates of point B are transformed to (x_1, y_1). For a horizontal flip, the new coordinates can be computed using Equation 3, while a vertical flip involves the point transformation defined by Equation 4. These principles are instrumental in image processing and manipulation.

$$\begin{cases} x_1 = w_i - 1 - a_1 \\ y_1 = b_1 \end{cases} \tag{3}$$

$$\begin{cases} x_1 = a_1 \\ y_1 = h_i - 1 - b_1 \end{cases} \tag{4}$$

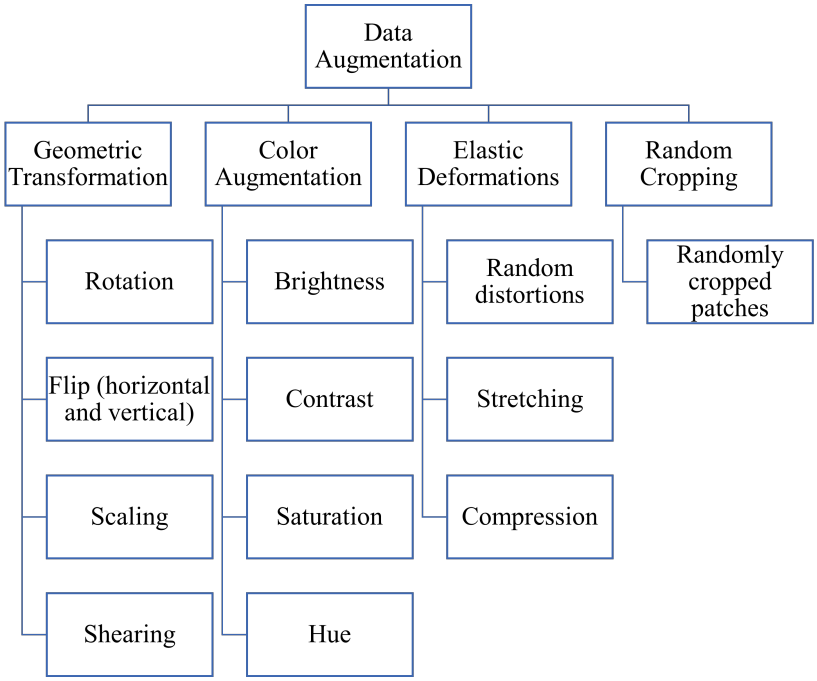

FIGURE 2 Common data augmentation techniques

3.3 | Grad-CAM for model explanation

The utilization of pre-trained weights from the ResNet architecture presents a powerful method for the classification of colorectal cancer in histopathology images. However, there remains a critical need for interpreting the decision-making process of these models. Grad-CAM (Gradient-weighted Class Activation Mapping) serves as a practical visualization technique for gaining insights into the regions of interest within an image that exert the most influence on the model's classification of colorectal cancer[20].

Grad-CAM works by examining the gradients of the final convolutional layer output concerning the class activation for colorectal cancer. Gradients essentially quantify the contribution of each pixel in the feature map to the final classification score. By directing attention to areas with high gradients, Grad-CAM produces a heatmap superimposed on the original image. This heatmap effectively highlights the image regions that the model identifies as being most significant for the accurate identification of colorectal cancer.

3.4 | Proposed CNN Architecture

Traditional CNNs face challenges in training very deep architectures due to the vanishing gradient problem, where gradients become extremely small or even zero as they propagate through the network, particularly in earlier layers. Consequently, the network struggles to effectively learn and update weights in those layers.

ResNet presents an impactful architecture for classifying EBHI-Seg classes, distinguished by its efficient feature extraction and reduced training complexity. ResNet addresses the vanishing gradient problem by introducing residual connections, which bypass some convolutional layers within a block and directly add the input of the block to its output. This approach establishes a shortcut path for information flow, ensuring that gradients do not vanish and can propagate effectively during training. Compared to traditional CNNs, residual connections enable the construction of deeper networks, allowing for the capture of more complex and hierarchical features from the EBHI-Seg images, potentially leading to more accurate classifications.

The introduction of residual connections for building deeper networks does not introduce additional trainable parameters, as the shortcut path simply adds the input to the output without learning a new transformation. This characteristic contributes to reducing the overall training complexity, especially in comparison to traditional CNNs with an equivalent number of layers.

In our study, we propose a set of methodologies for addressing the following:

- Utilization of Pre-trained ResNet on ImageNet: We employ a pre-trained ResNet model, which has acquired significant features from the extensive ImageNet dataset. These features are instrumental in recognizing general object shapes and textures, potentially transferable to the EBHI-Seg domain.
- Extraction of Overall Object Shapes: Leveraging the learned weights in the convolutional layers of the pre-trained ResNet enables us to capture overall object shapes within the EBHI-Seg ROI (Region of Interest). These pre-trained convolutional layers function as feature extractors, laying a robust foundation for our classification task.
- Introduction of New Layers and Classification: We introduce new fully-connected layers on top of the pre-trained model, specifically on the final convolutional block. These additional layers are responsible for learning the specific relationships between the extracted features and the EBHI-Seg classes. A "Flatten" layer amalgamates the pre-trained features with the newly added layers, while a final "Dense" layer with "softmax" activation undertakes the classification task, assigning probabilities to each EBHI-Seg class for a given image.

Based on our experimental findings, we have determined several key factors essential for achieving optimal training outcomes:

- Implementation of a low learning rate: This is crucial to ensure alignment with the pre-trained model during the training process.
- Utilization of the Adam optimizer: By combining the strengths of AdaGrad and RMSProp, this optimizer facilitates efficient model optimization.
- Adoption of a power-of-two batch size: This approach optimizes memory usage during the training phase.

By harnessing the advantages of the ResNet architecture and particular hyperparameter tuning, our methodology is designed to deliver outstanding performance in EBHI-Seg classification.

4 | EXPERIMENTS

4.1 | Dataset

The EBHI-Seg dataset, collected in 2022 by researchers from the Cancer Hospital of China Medical University in Shenyang, comprises 5,170 histopathology images representing six tumor differentiation stages. Each image is accompanied by a corresponding ground truth image, enabling comprehensive analyses. The dataset is segmented into six categories: Normal, Polyp, High-grade IN, Low-grade IN, Adenocarcinoma, and Serrated adenoma. Our study focused specifically on 2,228 histopathology section images for model training and evaluation, without utilizing the provided ground truth images from the dataset. The distribution of each class is detailed in Table 1.

Each image has a resolution of 224×224 pixels and is stored in the *.PNG format. The dataset classifies the images into six primary categories, as illustrated in 3, characterized as follows:

- Normal: These refer to colorectal tissue sections typified by well-organized tubular structures devoid of signs of infection, as discerned through light microscopy observations.
- Polyp: Resembling normal tubular structures, polyps possess distinctive histological features and manifest as redundant masses on mucosal surfaces.
- Intraepithelial Neoplasia (IN): Regarded as a critical precancerous lesion, IN exhibits increased branching of glandular structures, dense arrangement, and variations in luminal size and shape. High-grade IN is discernible by more pronounced structural alterations and nuclear enlargement relative to low-grade IN.
- Adenocarcinoma: This malignant tumor of the digestive tract is characterized by irregular luminal structures and ill-defined borders. The enlarged nuclei in these tumors pose challenges in histopathological observation.
- Serrated Adenoma: These are rare and account for approximately 1% of all colonic polyps, featuring a serrated histological appearance.

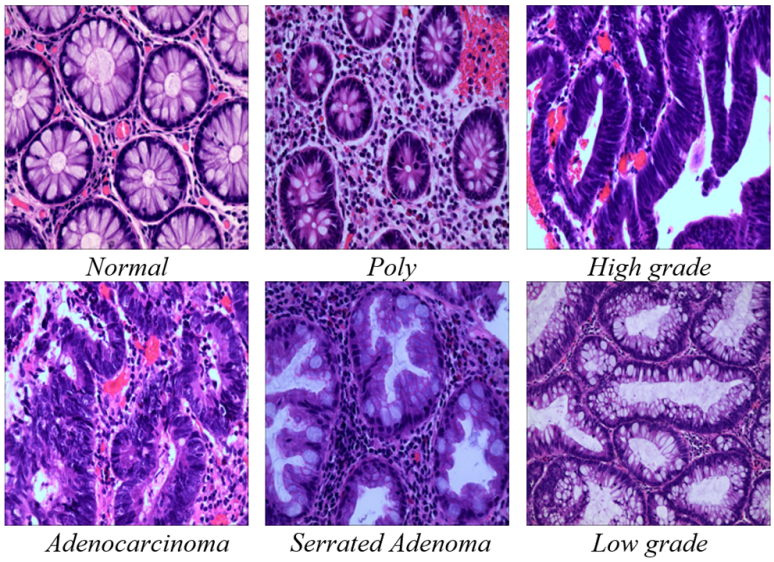

FIGURE 3 The EBHI-Seg dataset

TABLE 1 Detail information of EBHI-Seg dataset

Type of classes	Number of images
High-grade IN	186
Normal	76
Serrated	58
Low-grade IN	639
Adenocarcinoma	795
Polyp	474
Total	**2,228**

In experiment, we apply a 70-15-15 ratio to divide the base dataset into training, validating, and testing set, which will be used and applied for further training and analysis. The distribution of each class after splitting is detailed in Table 2. Because of lacking images for training and validating models, we need to apply data augmentation technique to this dataset. The testing set will not be augmented, and for each class in the training and validating sets, it will have the multiplication coefficient ratio to augment images. The detail augmented dataset will be shown in Table 3. The dataset visualization after augmenting will be illustrated in Fig. 4.

TABLE 2 Distribution of each class in the EBHI-Seg dataset after splitting into 3 sets

Base dataset	Training Set	Validating Set	Testing Set	Total Images
High-grade IN	130	28	28	**186**
Normal	53	11	12	**76**
Serrated	40	9	9	**58**
Low-grade IN	447	96	96	**639**
Adenocarcinoma	556	119	120	**795**
Polyp	331	71	72	**474**
Total Images	**1,557**	**334**	**337**	**2,228**

TABLE 3 Distribution of each class in the EBHI-Seg dataset after augmentation

Augmented dataset	Multiplication Coefficient Ratio for Training and Validation Sets	Training Set	Validating Set	Testing Set	Total Images
High-grade IN	26	3,380	728	28	**4,136**
Normal	52	2,756	572	12	**3,340**
Serrated	52	2,080	468	9	**2,557**
Low-grade IN	14	6,258	1,344	96	**7,698**
Adenocarcinoma	12	6,672	1,428	120	**8,220**
Polyp	14	4,634	994	72	**5,700**
Total Images		**25,780**	**5,534**	**337**	**31,651**

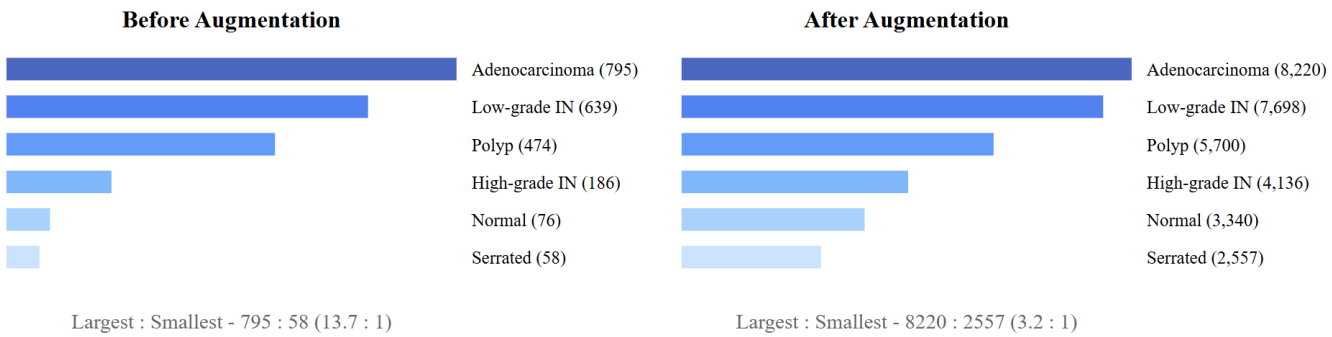

FIGURE 4 Class Distribution Before and After Augmentation

4.2 | Apply pre-trained weights on ResNet architecture

Our proposed architecture leverages transfer learning with ResNet as the backbone network, addressing the vanishing gradient problem common in deep neural networks while optimising for colorectal cancer histopathology image classification. As shown in Fig. 5, the architecture consists of three main components: feature extraction using frozen pre-trained layers, additional classification layers, and an interpretability module.

The architecture processes (224x224x3) RGB histopathology images through a pre-trained ResNet backbone (tested with ResNet-18, ResNet-34, and ResNet-50 variants), initialised with ImageNet weights and kept frozen to preserve learned features. The feature extraction network is followed by a series of additional classification layers: a Flatten layer that converts feature maps to 1D vectors, a Batch Normalization layer (epsilon=1e-5, momentum=0.99) that normalizes layer inputs, a Dropout layer with a rate of 0.5 for regularization, and finally a Dense layer with 6 units and softmax activation for class probability outputs.

The training configuration utilizes the Adam optimizer with a learning rate of 0.0001 and sparse categorical cross-entropy as the loss function. We implement a batch size of 32 (chosen as a power-of-two for memory optimization) and train for 30 epochs with early stopping that monitors validation loss with a patience of 5. For model interpretability, we incorporate Grad-CAM visualization using the final convolutional layer to generate feature importance heatmaps.

The effectiveness of our architecture stems from several key design choices. Transfer learning with pre-trained ImageNet weights provides robust feature extraction capabilities that are transferable to histopathology domains. ResNet's residual connections facilitate gradient flow during backpropagation, enabling effective training of deep architectures. The additional classification layers are specifically designed to adapt the pre-trained features for the EBHI-Seg classification task, while the Grad-CAM visualization provides crucial insight into model decisions for clinical applications.

Training proceeds in two phases. During feature extraction, the base ResNet layers remain frozen with weights preserved from ImageNet pre-training, focusing computation on adapting the classification layers. In the classification layer training phase, we employ a low learning rate (0.0001) to ensure stable adaptation, utilize the Adam optimizer to combine adaptive gradient descent methods, apply batch normalization to stabilize training, and implement dropout to prevent overfitting.

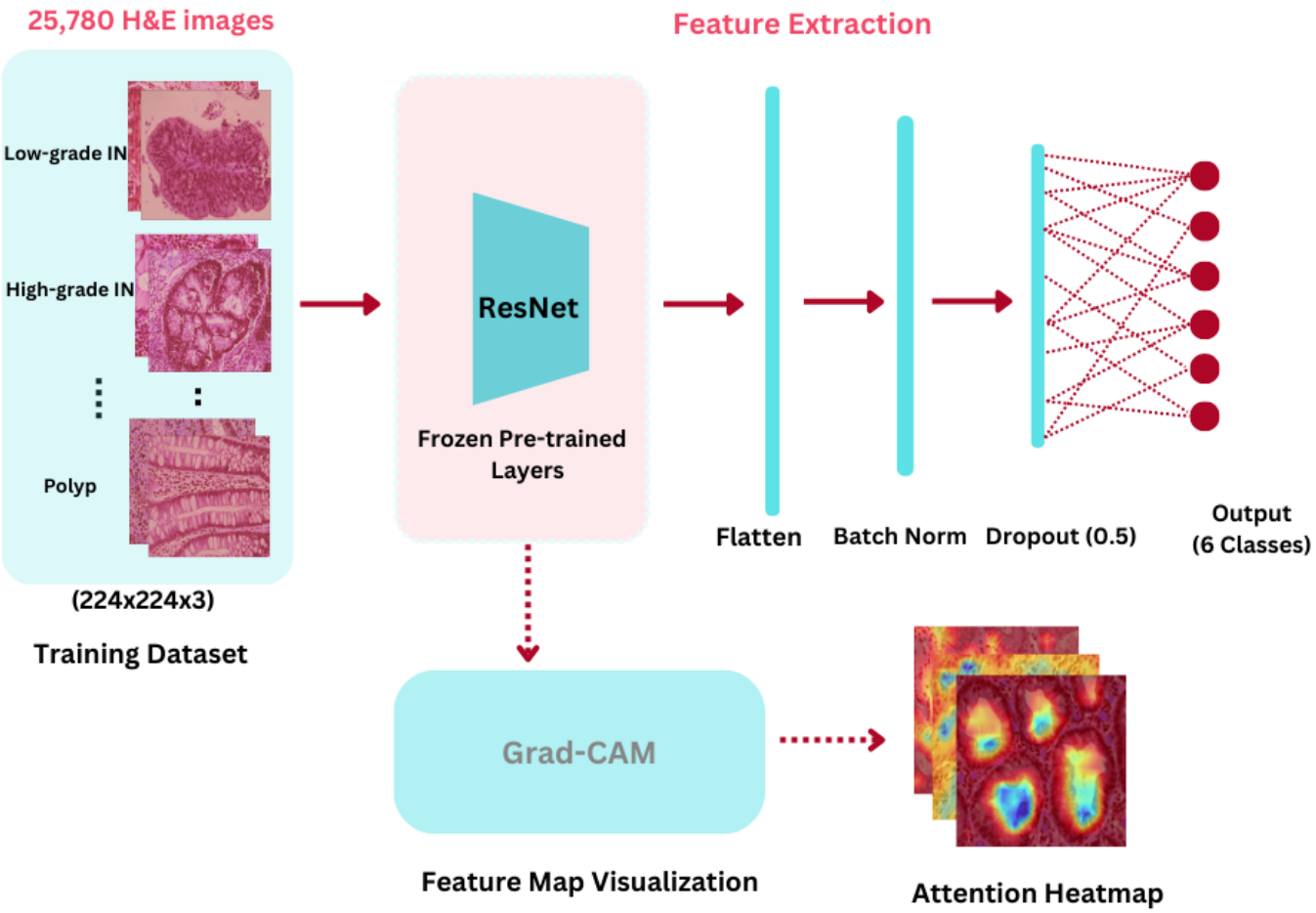

FIGURE 5 Proposed architecture with detailed layers visualization

As illustrated in Figure 5, our architecture processes 25,780 H&E images (224x224x3) through the frozen pre-trained ResNet layers and additional classification layers (Flatten, Batch Norm, Dropout) to produce probabilities for six classes. The parallel Grad-CAM branch generates attention heatmaps for model interpretability. This architecture achieves a balance between computational efficiency, model complexity, and clinical interpretability while maintaining high classification performance across different colorectal cancer subtypes.

In this study, we have employed ResNet as the foundational architecture and have utilized various versions of ResNet (ResNet-18, ResNet-34, and ResNet-50) for evaluation on the dataset. The architectural comparisons of each ResNet version will be visually presented in Fig. 6. We will assess the efficiency and effectiveness of each ResNet backbone architecture based on performance metrics, primarily accuracy and AUC scores. The metrics table detailing the testing results of these models will be included in Table 4.

TABLE 4 Testing metrics of each ResNet backbone architecture

Metrics	ResNet-18	ResNet-34	ResNet-50
Accuracy	0.9021	0.9110	0.8961
Top 2 Accuracy	0.9822	0.9911	0.9763
Top 3 Accuracy	0.9970	1.0000	0.9911
Macro-averaged ROC AUC OvO	0.9872	0.9872	0.9889
Micro-averaged ROC AUC OvR	0.9918	0.9928	0.9921
Macro-averaged ROC OVR OvR	0.9884	0.9883	0.9895

Layer Name	Output Size	18-layer	34-layer	50-layer
Convolutional 1	112×112	7×7, 64, stride 2		
Convolutional 2	56×56	3×3, max pool, stride 2		
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Convolutional 3	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Convolutional 4	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Convolutional 5	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	Average pooling layer, 1000-dfc, softmax		
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9

FIGURE 6 ResNet-18, ResNet-34, and ResNet-50 architecture

The effectiveness of the proposed ResNet-18, ResNet-34, and ResNet-50 models was extensively assessed using essential performance metrics such as Area Under the ROC Curve (AUC), accuracy, precision, recall, and F1-scores across six classes to classify colorectal cancer (CRC) from histopathology images.

Notably, the ResNet-18 model displayed impressive performance in effectively distinguishing between different classes of CRC. The ROC AUC scores ranged from 0.9764 to 0.9986 for individual classes, with a macro-averaged ROC AUC score of 0.9884 and a micro-averaged ROC AUC score of 0.9918, indicating its outstanding discriminatory capability. The overall accuracy of the model was 90.21%, with top-2 and top-3 accuracies of 98.22% and 99.70%, respectively, underscoring its frequent correct classification among the leading predictions. Precision scores for ResNet-18 spanned from 0.7778 to 1.0000, recall scores from 0.500 to 0.9917, and F1-scores from 0.6087 to 0.9558. However, the confusion matrix revealed some misclassifications, particularly in classes with limited samples, highlighting the challenges inherent in limited data and the inherent variability of histopathology images. The results of the ResNet-18 model are illustrated in Fig.7.

The ResNet-34 demonstrated superior performance compared to the other models in the study. It consistently achieved high ROC AUC scores, ranging from 0.9679 to 0.9997 for individual classes, with a macro-averaged ROC AUC score of 0.9883 and a micro-averaged ROC AUC score of 0.9928, showcasing its robustness and reliability in accurately classifying CRC images. The overall accuracy of ResNet-34 was 91.10%, with impressive top-2 and top-3 accuracies of 99.11% and 100.00%, respectively, highlighting its exceptional ability to rank the correct class with high precision. Moreover, the precision scores for ResNet-34 ranged from 0.8182 to 0.9435, recall scores from 0.6429 to 1.0000, and F1-scores from 0.6957 to 0.9590, indicating its balanced performance across various metrics. Although some minor misclassifications were observed in the confusion matrix analysis, ResNet-34 consistently and accurately classified the majority of samples. These results collectively underscore the model's suitability for practical applications. The comprehensive performance of the ResNet-34 model is visually illustrated in Fig. 8.

The ROC AUC scores of ResNet-50 ranged from 0.9659 to 0.9976, with macro-averaged and micro-averaged ROC AUC scores at 0.9895 and 0.9921, respectively. The model demonstrated an overall accuracy of 89.61%, with top-2 and top-3 accuracies of 97.63% and 99.11%, respectively. These high accuracies underscore the model's robust predictive capability. Precision scores for ResNet-50 fell within the range of 0.6471 to 0.9225, while recall scores ranged from 0.5714 to 0.9917, and F1-scores from 0.6809 to 0.9558. The confusion matrix for ResNet-50 revealed challenges in classifying certain categories, particularly those with fewer samples, mirroring the issues encountered by other models. Nevertheless, the model's consistently high-performance metrics validate its efficacy in classifying CRC images. The outcome of the ResNet-50 model is shown in Fig. 9.

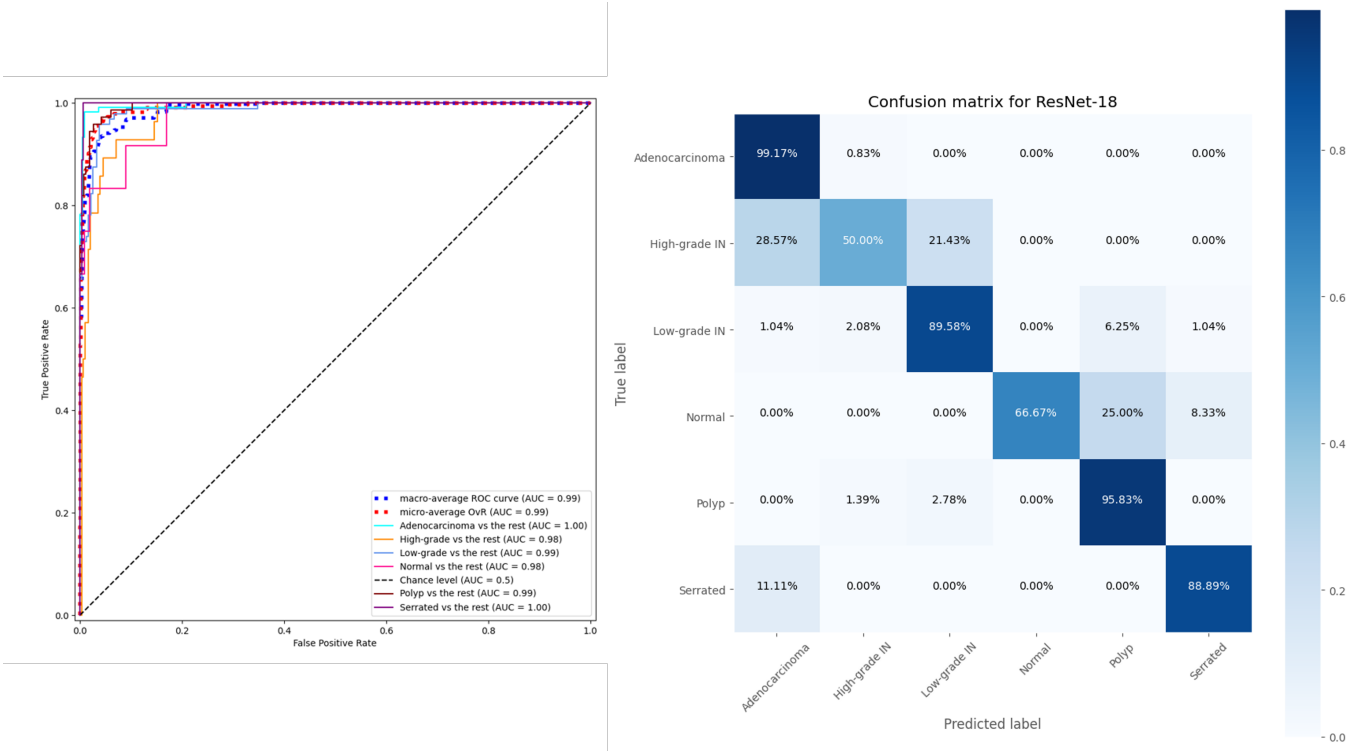

FIGURE 7 ResNet-18 classification result

4.3 | Compare models and evaluate

The remarkable ROC AUC scores and accuracy metrics demonstrated by all three models underscore their efficacy in automating the diagnosis of colorectal cancer (CRC) from histopathology images. The utilization of pre-trained ResNet architectures significantly contributed to the robustness and high performance of the models, even when confronted with limited training data. Furthermore, the integration of Grad-CAM visualization yielded valuable insights into the decision-making processes of the models, ensuring that the identified regions of interest align with the pertinent diagnostic features of CRC. The comprehensive evaluation metrics are presented in Table 5.

- ResNet-18: This model demonstrates high ROC AUC scores across all classes and perfect precision for the Normal class. It also shows good F1-scores for Adenocarcinoma and Polyp. However, it shows lower recall for High-grade IN and Normal classes, as well as lower precision for the Serrated class. Overall, considering its smaller size, the ResNet-18 displays solid performance, particularly in ROC AUC and precision for certain classes. Hence, it could be the preferred choice when model size and computational efficiency are crucial factors.

- ResNet-34: With the highest overall accuracy of 0.9110 and perfect top 3 accuracy, this model achieves the best F1-scores for most classes, including High-grade IN and Serrated. Although it exhibits slightly lower precision for the Normal class compared to the ResNet-18, it shows the most balanced and consistently high performance across metrics. Consequently, it emerges as the most effective model overall, offering the best balance of accuracy, precision, recall, and F1-scores across classes while also achieving perfect top 3 accuracy.

- ResNet-50: This model showcases the highest Macro-averaged ROC AUC scores and best ROC AUC for Adenocarcinoma and Normal classes. However, it has the lowest overall accuracy of 0.8961, along with lower recall for the Serrated class and lower precision for the Normal class. While it excels in ROC AUC metrics, it lags behind in accuracy and some class-specific measures. Despite its strength in ROC AUC metrics, it does not demonstrate a clear advantage over the smaller models in other important metrics such as accuracy and F1-scores.

The Grad-CAM visualizations of the ResNet-34 model exhibit compelling evidence of its effectiveness in discerning various classes of CRC histopathology images. The heatmaps generated by the model adeptly highlight regions of interest within the

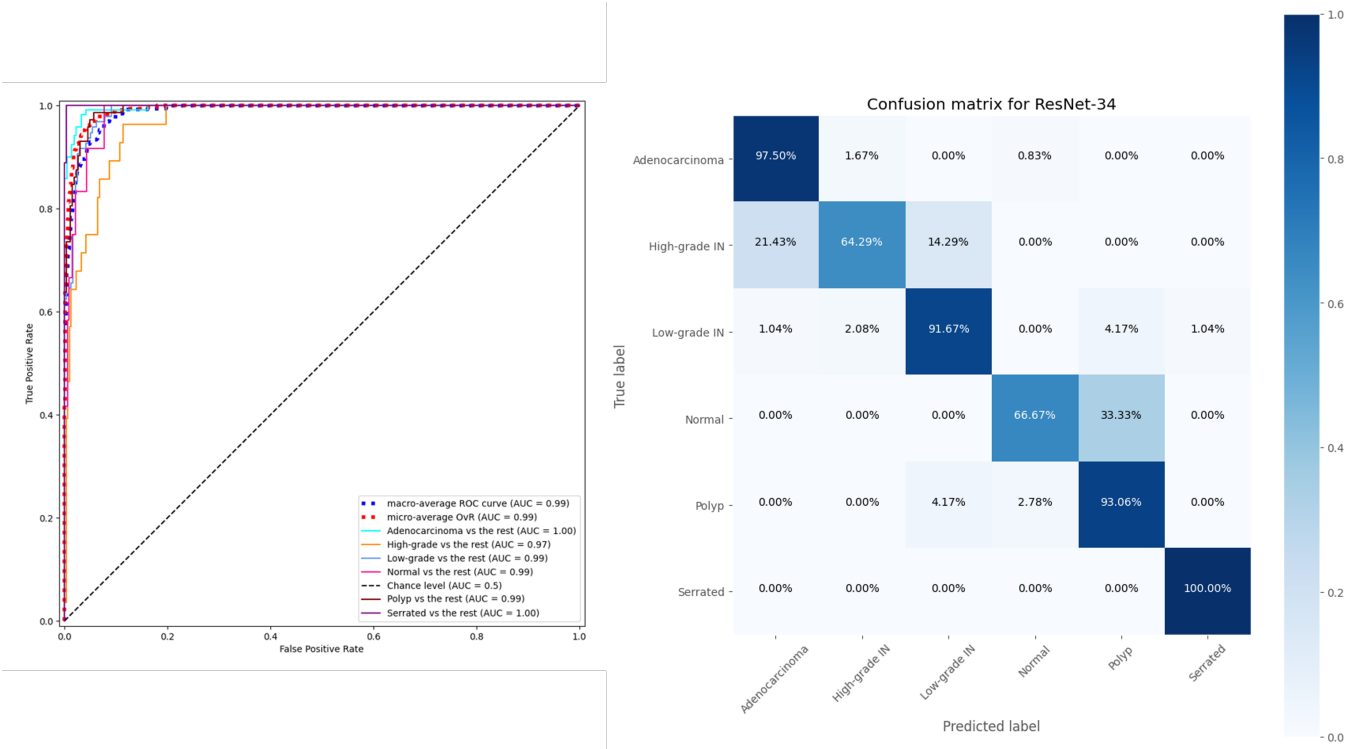

FIGURE 8 ResNet-34 classification result

tissue samples, thereby effectively identifying abnormal cell locations. These heatmaps convincingly demonstrate the model's capacity to concentrate on diagnostically significant areas, closely aligning with expert pathologist annotations. This precise highlighting of abnormal cellular regions is pivotal for the model's proficiency in distinguishing between different CRC classes, thus affirming its potential for practical clinical applications. Notably, the Grad-CAM results not only enhance the interpretability of the model's predictions but also bolster confidence in its diagnostic capabilities. The application of GradCAM to the testing set is depicted in Fig. 10.

TABLE 5 Evaluated metrics compared on each ResNet backbone architecture

Classified EBHI-SEG Classes		Adenocarcinoma	High-grade IN	Low-grade IN	Normal	Polyp	Serrated
ROC AUC OvR	ResNet-18	0.9968	0.9764	0.9877	0.9762	0.9947	0.9986
	ResNet-34	0.9963	0.9679	0.9891	0.9851	0.9918	0.9997
	ResNet-50	0.9974	0.9659	0.9887	0.9949	0.9922	0.9976
Precision	ResNet-18	0.9225	0.7778	0.9149	1.0000	0.8846	0.8000
	ResNet-34	0.9435	0.8182	0.9263	0.7273	0.8933	0.9000
	ResNet-50	0.9225	0.8421	0.9149	0.6471	0.9014	0.8571
Recall	ResNet-18	0.9917	0.5000	0.8958	0.6667	0.9583	0.8889
	ResNet-34	0.9750	0.6429	0.9167	0.6667	0.9306	1.0000
	ResNet-50	0.9917	0.5714	0.8958	0.9167	0.8889	0.6667
F1-score	ResNet-18	0.9558	0.6087	0.9053	0.8000	0.9200	0.8421
	ResNet-34	0.9590	0.7200	0.9215	0.6957	0.9116	0.9474
	ResNet-50	0.9558	0.6809	0.9053	0.7586	0.8951	0.7500

Despite their high overall performance, the models demonstrated certain limitations, particularly in misclassifying classes with smaller sample sizes. This issue may stem from the inherent variability present in histopathology images and the restricted size

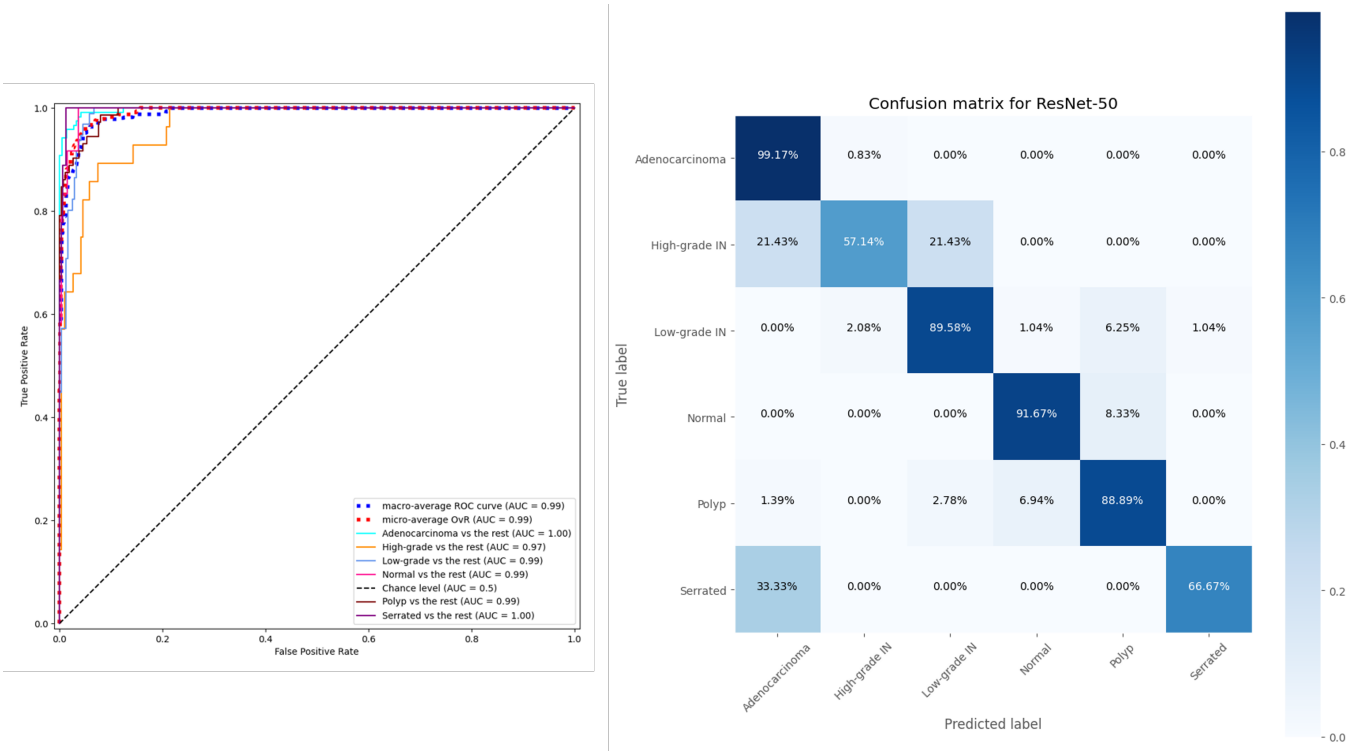

FIGURE 9 ResNet-50 classification result

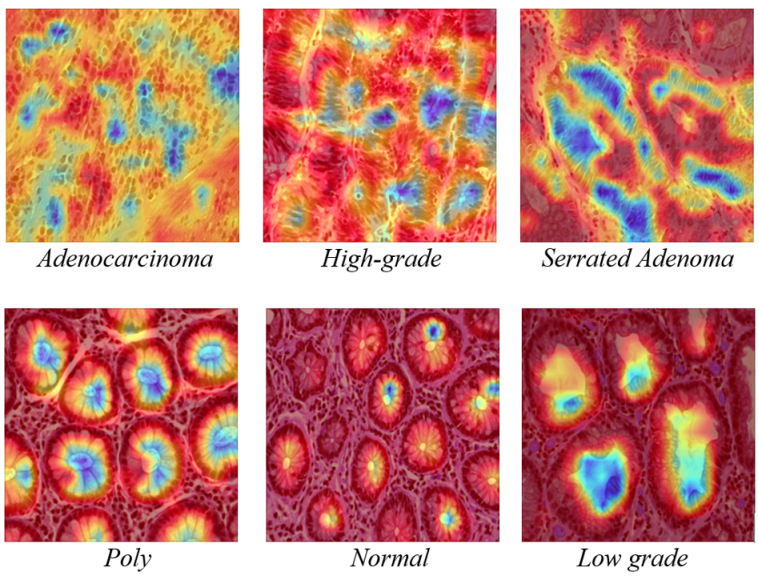

FIGURE 10 GradCAM explanation on EBHI-Seg dataset

of the training dataset. While the utilization of data augmentation techniques enhanced generalizability, there remains a potential risk of overfitting to the training data. The robustness of the models, attributed to the effective use of pre-trained architectures, highlights their promise in clinical applications. However, the observed variability in precision, recall, and F1-scores, especially in classes with limited samples, underscores the challenges posed by data variability and emphasizes the necessity for further improvements.

5 | CONCLUSION

The findings of our research underscore the considerable promise of deep learning models in automating the diagnosis of colorectal cancer (CRC) from histopathology images. With further validation and refinement, these models have the potential to significantly improve diagnostic accuracy, consistency, and efficiency, thereby enhancing patient outcomes in CRC care. Among the models assessed, ResNet-34, distinguished by its deeper architecture, emerged as the most effective, achieving an overall accuracy of 91.10%. However, its heightened complexity translates to greater computational demands, potentially limiting its applicability in resource-constrained settings. Nevertheless, its exceptional performance renders it an optimal choice in scenarios where the highest diagnostic accuracy is imperative. To optimize the outcomes of this study, several methodologies can be considered. Firstly, expanding the size and diversity of the training dataset through the inclusion of images from multiple sources and the utilization of different staining techniques can bolster the model's robustness and generalizability. Secondly, exploring alternative deep learning architectures and ensemble methods holds promise for further enhancing classification performance. Thirdly, collaborating with pathologists to seamlessly integrate these models into clinical workflows, and providing real-world validation and refinement will ensure alignment with the practical requirements of healthcare professionals. Through ongoing development, this research has the potential to markedly enhance patient outcomes in CRC care by enabling swifter, more precise, and consistent diagnoses.

ACKNOWLEDGEMENT

We thank Nguyen-Truong Vinh-Thuyen (Research Engineer at the National University of Singapore) for his support in content consultation and coding. We also acknowledge financial support from Mr. Thang Truong Le and Ms. Kiep Thi Quang.

References

1. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol* 2021; 14(10): 101174. doi: 10.1016/j.tranon.2021.101174

2. Lee Y, Lee Y, Chuang J, Lee J. Differences in survival between colon and rectal cancer from SEER data. *PloS One* 2013; 8(11): e78709. doi: 10.1371/journal.pone.0078709

3. Thijs J, Zwet vA, Thijs W, et al. Diagnostic tests for Helicobacter pylori: a prospective evaluation of their accuracy, without selecting a single test as the gold standard. *Am J Gastroenterol* 1996; 91(10): 2125-2129.

4. Cervantes A, Adam R, Roselló S, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* 2023; 34(1): 10-32. doi: 10.1016/j.annonc.2022.10.003

5. Wang Y, He X, Nie H, Zhou J, Cao P, Ou C. Application of artificial intelligence to the diagnosis and therapy of colorectal cancer. *Am J Cancer Res* 2020; 10(11): 3575-3598.

6. Alinia S, Ahmadi S, Mohammadi Z, et al. Exploring the impact of stage and tumor site on colorectal cancer survival: Bayesian survival modeling. *Sci Rep* 2024; 14: 4270. doi: 10.1038/s41598-024-54943-8

7. Mitsala A, Tsalikidis C, Pitiakoudis M, Simopoulos C, Tsaroucha A. Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era. *Curr Oncol* 2021; 28(3): 1581-1607. doi: 10.3390/curroncol28030149

8. Saoudi González N, Salvà F, Ros J, et al. Unravelling the Complexity of Colorectal Cancer: Heterogeneity, Clonal Evolution, and Clinical Implications. *Cancers* 2023; 15(16): 4020. doi: 10.3390/cancers15164020

9. Yin Z, Yao C, Zhang L, Qi S. Application of artificial intelligence in diagnosis and treatment of colorectal cancer: A novel Prospect. *Front Med* 2023; 10: 1128084. doi: 10.3389/fmed.2023.1128084

10. Singh M, Rai S, Pandey A, Singh N, Srivastava S. Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine. *Genes Dis* 2019; 8(2): 133-145. doi: 10.1016/j.gendis.2019.10.013

11. Sarwinda D, Paradisa RH, Bustamam A, Anggia P. Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Computer Science* 2021; 179: 423-431. 5th International Conference on Computer Science and Computational Intelligence 2020doi: https://doi.org/10.1016/j.procs.2021.01.025

12. Ahmad MY, Mohamed A, Mohd Yusof YA, Md. Ali SA. Colorectal cancer image classification using image pre-processing and multilayer Perceptron. 2012; 1: 275-280. doi: 10.1109/ICCISci.2012.6297253

13. Bychkov LNTRea. Deep learning based tissue analysis predicts outcome in colorectal cancer. 2018; 8. doi: 10.1038/s41598-018-21758-3

14. Motlagh MH, Jannesari M, Aboulkheyr H, et al. Breast Cancer Histopathological Image Classification: A Deep Learning Approach. *bioRxiv* 2018. doi: 10.1101/242818

15. Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *Journal of Cancer Research and Clinical Oncology* 2019; 145(4): 829-837. doi: 10.1007/s00432-018-02834-7

16. Tsaku NZ, Kosaraju SC, Aqila T, et al. Texture-based Deep Learning for Effective Histopathological Cancer Image Classification. 2019: 973-977. doi: 10.1109/BIBM47256.2019.8983226

17. Sharkas M, Attallah O. Color-CADx: a deep learning approach for colorectal cancer classification through triple convolutional neural networks and discrete cosine transform. *Sci Rep* 2024; 14: 6914. doi: 10.1038/s41598-024-56820-w

18. Miyagawa C, Nakai H, Otani T, et al. Histopathological subtyping of high-grade serous ovarian cancer using whole slide imaging. *J Gynecol Oncol* 2023; 34(4): e47. doi: 10.3802/jgo.2023.34.e47

19. Li M, Ma X, Chen C, et al. Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images. 2021; 9: 53687-53707. doi: 10.1109/ACCESS.2021.3071057

20. Gildenblat J, contributors . PyTorch library for CAM methods. https://github.com/jacobgil/pytorch-grad-cam; 2021.