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

Thang Truong Le[1,2,*], Vinh-Thuyen Nguyen-Truong[3], Nghia Trong Le Phan[4], Quang Duong Van Nhat[5], Mqondisi Fortune Mavuso[1], Phuc Nguyen Thien Dao[6], Huy Ngoc Anh Nguyen[7], Tien Thuy Mai[7], Kiep Thi Quang[7]

[1] Smart Medicine and Health Informatics Program, International College, National Taiwan University, Taipei, Taiwan

[2] School of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

[3] John von Neumann Institute - Vietnam National University HCM

[4] Home Credit Vietnam

[5] International Program of Electrical Engineering and Computer Science, National Taipei University of Technology, Taipei, Taiwan

[6] Faculty of Biology-Biotechnology, University of Science, VNUHCM, Ho Chi Minh City, Vietnam

[7] Hung Vuong High School for the Gifted, Pleiku, Gia Lai, Vietnam

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

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

Abstract: Accurate and timely diagnosis of colorectal cancer is critical for improving patient outcomes. While histopathological examination remains the gold standard, it is labor-intensive and prone to inter-observer variability. This study investigates the application of deep learning, specifically ResNet and Swin Transformer V2 architectures, for automated CRC subtype classification using hematoxylin and eosin stained histopathology images. We evaluate lightweight variants-ResNet-18, ResNet-34, ResNet-50, Swin v2-T Small, Swin v2-T Tiny, and Swin v2-T Tiny w8-for their ability to extract relevant visual features. To enhance model generalizability under limited data conditions, we apply data augmentation techniques including geometric and color transformations, elastic deformations, and random cropping. Model interpretability is addressed using Grad-CAM, enabling visualization of regions contributing most to predictions. Among the models, ResNet-34 achieved the best trade-off between accuracy and complexity, with overall, top-2, and top-3 classification accuracies of 91.10%, 99.11%, and 100.00%, respectively. Swin Transformer V2 models also showed competitive results, particularly in adenocarcinoma (up to 99.17%) and polyp (95.14%) detection, though at the cost of increased computational demands. These results suggest that moderate-depth architectures can effectively capture the morphological diversity of colorectal cancer subtypes. Our approach provides an interpretable, efficient deep learning-based diagnostic tool that can support pathologists by improving classification accuracy and consistency, ultimately contributing to more personalized and timely cancer care.

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

1. INTRODUCTION

Colorectal cancer (CRC) 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].

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 [5]. 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[6].

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[7]. This transition is particularly evident in early-stage colorectal cancer management, where AI-driven interventions have significantly improved survival outcomes [8], [9]. 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 [10].

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. Sarwinda [11] explored ResNet architectures, with ResNet-50 achieving superior accuracy (above 80%), sensitivity (above 87%), and specificity (above 83%) in distinguishing benign from malignant tissues. Ahmad [12] introduced a method utilizing HSV color space and gland segmentation, reaching 94% accuracy using a Multilayer Perceptron. Bychkov [13] developed a deep learning model that directly predicted patient outcomes from tumor tissue images, outperforming human histological assessments. Given the increasing volume of medical images, CNNs have emerged as a key solution for efficient cancer classification. Mehdi [14] leveraged pre-trained CNNs to classify multiple cancer types with up to 99.8% accuracy. Jeyaraj [15] introduced a partitioned CNN for oral cancer detection, achieving over 91% accuracy. Tsaku [16] developed CAT-Net, a deep neural network for histopathological image analysis, surpassing existing methods in precision, recall, and F1-score. Recent innovations include "Color-CADx", a computer-aided diagnostic (CAD) system combining ResNet50, DenseNet201, and AlexNet for CRC classification [17]. This system employs discrete cosine transform (DCT)

and ANOVA for feature selection, achieving 99.3% accuracy on the NCT-CRC-HE-100K dataset. Beyond cancer detection, machine learning (ML) and deep learning (DL) play a vital role in precision medicine by classifying cancer subtypes. For example, ML-based whole-slide imaging (WSI) has improved the classification of high-grade serous ovarian cancer (HGSOC) into clinically relevant subtypes [18]. In lung cancer, ML techniques have enhanced subtype classification accuracy, supporting personalized treatment strategies [19]. These advancements highlight the transformative potential of AI in oncological diagnostics, enabling more precise and effective therapeutic decisions.

Our 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.

To optimize model performance while addressing computational limitations, we utilize advanced deep learning architectures such as ResNet and Swin Transformer V2. By fine-tuning these models on CRC-specific datasets, we enable efficient feature extraction and improved diagnostic accuracy, even with limited annotated data. This approach enhances the precision of CRC classification and supports more individualized treatment strategies. Our findings contribute to the growing field of AI-assisted histopathology and underscore the potential of deep learning in transforming colorectal cancer diagnostics.

2. MATERIALS AND METHODS

2.1. Data collection

The EBHI-Seg dataset [20], collected in 2022 by researchers from the Cancer Hospital of China Medical University in Shenyang, comprises 5,170 histopathology images (hematoxylin and eosin-H&E) 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.

Table 1. Six primary categories from EBHI-Seg dataset.

Subtype	Characterictics	Number
High-grade IN	High-grade IN is discernible by more pronounced structural alterations and	186

	nuclear enlargement relative to low-grade IN.	
Normal	These refer to colorectal tissue sections typified by well-organized tubular structures devoid of signs of infection, as discerned through light microscopy observations.	76
Polyp	Resembling normal tubular structures, polyps possess distinctive histological features and manifest as redundant masses on mucosal surfaces.	58
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.	639
Adenocarcinoma	This malignant tumor of the digestive tract is characterized by irregular luminal structures and illdefined borders. The enlarged nuclei in these tumors pose challenges in histopathological observation.	795
Serrated Adenoma	These are rare and account for approximately 1% of all colonic polyps, featuring a serrated histological appearance.	474

2.2. Data augmentation and preprocessing

Training deep learning models for colorectal cancer classification is hindered by limited data, which can cause overfitting, and is effectively addressed by data augmentation techniques that simulate natural histopathological variations. These techniques include geometric transformations (e.g., rotation, flipping, scaling, and shearing), color augmentation (adjusting brightness, contrast, saturation, and hue), elastic deformations (random distortions, stretching, and compression), and random cropping, which help enhance model robustness and generalizability. 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.

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**

Because the dataset exhibits a significant class imbalance, with adenocarcinoma (795 samples) being the most frequent class, while serrated lesions have only 58 samples, resulting in a largest-to-smallest ratio of 13.7:1.

Other categories, including low-grade IN (639), polyp (474), high-grade IN (186), and normal (76), also show varying levels of underrepresentation. Therefore, we need to apply data augmentation technique to this dataset. After applying data augmentation, the dataset size is significantly expanded, with adenocarcinoma increasing to 8,220 samples and serrated lesions reaching 2,557 samples, reducing the largest-to-smallest ratio to 3.2:1. This augmentation process enhances dataset balance, mitigating biases in model training and improving the generalizability of deep learning models for colorectal cancer classification.

2.3. Grad-CAM for model explanation

The use of pre-trained ResNet weights enhances colorectal cancer classification in histopathology images, but model interpretability remains crucial. Grad-CAM provides a visualization technique by analyzing gradient contributions in the final convolutional layer, generating heatmaps that highlight critical regions influencing the model's decision. This helps improve transparency and trust in deep learning-based diagnoses [21].

2.4. Deep Learning Architecture

Because traditional deep CNN architectures suffer from the vanishing gradient, where gradients become extremely small or even zero as they propagate through the network [22], cause the networks failed to learn and update the learnable parameters. In this study, two distinct deep learning architectures were employed for colorectal cancer (CRC) subtype classification: ResNet-based convolutional neural networks (CNNs) and Swin Transformer V2-based vision transformers (ViTs).

2.4.1. Resnet-based models

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 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.

ResNet has been widely utilized in medical imaging and clinical data analysis, demonstrating strong performance across various applications. In medical imaging, it has been successfully applied to X-rays for pneumonia detection, MRI and CT scans for brain tumor segmentation, and histopathological slides for cancer classification. Additionally, its adaptability extends beyond imaging to electronic health records (EHR), where it aids in disease prediction by analyzing patient demographics, clinical notes, and laboratory test results. Furthermore, ResNet has been employed in processing time-series data, such as ECG signals for cardiac arrhythmia detection and EEG recordings for neurological disorder analysis, reinforcing its versatility in medical AI [23].

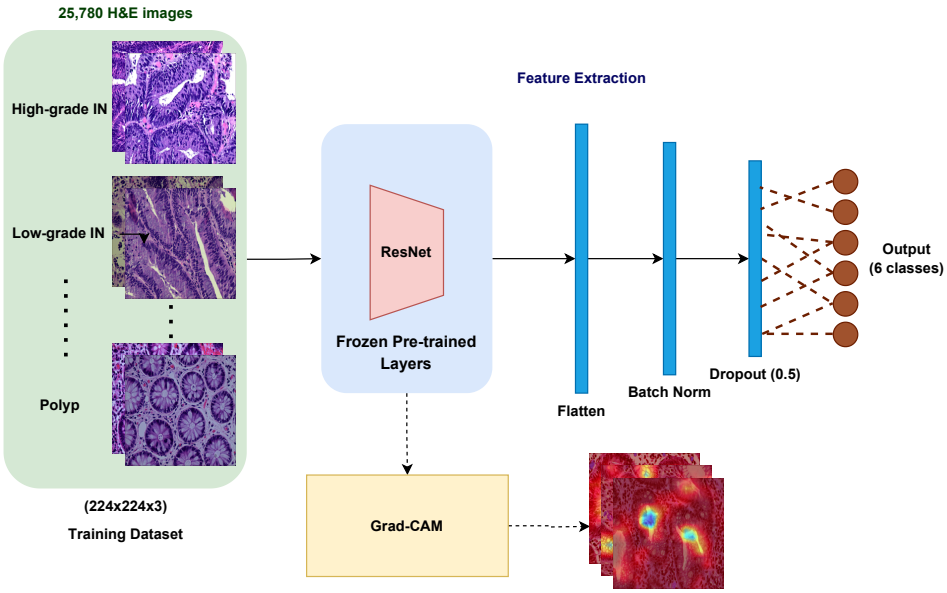

FIGURE 1. Proposed architecture with detailed layers visualization.

As illustrated in Figure 1, our architecture processes 25,780 H&E images (224×224×3t) 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.

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 Figure. 2.

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 2. ResNet-18, ResNet-34, and ResNet-50 architecture

2.4.2. Swin Transformer Model

Swin Transformer V2 enhances computational efficiency by employing a shifted window self-attention mechanism, which divides an image into smaller non-overlapping windows for localized attention computations. By shifting the windows at each layer, the model ensures cross-window interaction, allowing it to capture both local and long-range dependencies without the excessive computational cost associated with global self-attention. This hierarchical structure enables Swin V2 to process high-resolution images more effectively, making it particularly suited for medical imaging tasks such as colorectal cancer histopathology classification. Additionally, the model introduces scaled cosine attention, which normalizes attention values, reducing sensitivity to extreme activations and improving overall training stability. Through these optimizations, Swin V2 achieves superior feature extraction while maintaining computational feasibility, even for high-resolution image analysis [24].

Due to its superior ability to model both local and global relationships within medical images, Swin Transformer has found extensive applications in histopathological image classification, brain tumor segmentation in MRI scans, lung nodule detection in CT scans, and diabetic retinopathy screening from retinal fundus images. Additionally, its effectiveness extends to dermatology, where it has been used for skin lesion classification in dermoscopic images. Beyond imaging, Swin Transformer has also been explored for multimodal medical data integration, combining histopathological imaging with genomic data to advance personalized medicine.[25]

2.5 Training Procedure

Our proposed architecture employs a ResNet backbone network to address the vanishing gradient problem inherent in deep neural networks, while optimizing classification performance for colorectal cancer histopathology images. As illustrated in Figure 1, the architecture consists of three primary components: (1) a feature extraction module using frozen ResNet layers, (2) a series of additional classification layers, and (3) an interpretability module.

The input to the system comprises RGB histopathology images of size (224×224×3), which are processed through a ResNet backbone (tested with ResNet-18, ResNet-34, and ResNet-50 variants). These backbone networks are initialized with ImageNet weights and kept frozen during training to preserve previously acquired visual representations. Following feature extraction, the architecture incorporates a classification head consisting of a Flatten layer, a Batch Normalization layer (ε = 1e-5, momentum = 0.99), a Dropout layer (rate = 0.5) for regularization, and a Dense layer with six units and softmax activation for multiclass prediction.

Model training utilizes the Adam optimizer (learning rate = 1e-4) with sparse categorical cross-entropy as the loss function. A batch size of 32 is selected for computational efficiency, and training proceeds for 30 epochs with early stopping based on validation loss, using a patience value of 5. For interpretability, Grad-CAM is applied to the final convolutional layer, generating visual heatmaps that highlight important regions in histopathology images.

The effectiveness of this architecture stems from several core design choices. The use of ResNet initialized

with ImageNet weights allows for efficient and robust feature extraction, even with limited domain-specific data. The residual connections within ResNet improve gradient flow, facilitating the training of deeper networks. Additional classification layers are tailored to adapt extracted features for the EBHI-Seg dataset, while Grad-CAM enables interpretability, offering visual explanations aligned with clinical reasoning.

Training is conducted in two phases. In the initial phase, the ResNet base remains frozen to retain its representational capacity, while only the classification layers are updated. In the second phase, the low learning rate ensures stable convergence. Batch normalization promotes training stability, and dropout mitigates overfitting, enhancing model generalizability.

2.6. Evaluation Setup

The model's performance was assessed using a range of evaluation metrics to ensure a comprehensive analysis. Accuracy was measured to determine overall correctness, while a correlation matrix was utilized to examine relationships between variables. The area under the receiver operating characteristic curve (AUROC) provided insight into the model's discriminatory power. Additionally, precision and recall were analyzed to evaluate the balance between false positives and false negatives. The F1-score, which combines precision and recall, was also considered to offer a more holistic view of classification performance. These metrics collectively enabled a rigorous evaluation of the model's effectiveness.

Accuracy = (TP+TN)/(TP+FP+TN+FN); Precision=TP/(TP+FP);

Recall=TP/(TP+FN); F1 = 2×(Precision×Recall)/(Precision+Recall); FPR=FP/(FP+TN)

$$\text{AUROC} = \int_0^1 \text{Recall}(\text{FPR})\,\text{d}(\text{FPR})$$

Where: TP (True Positives) is correctly predicted positive cases, TN (True Negatives) is correctly predicted negative cases, FP (False Positives) is incorrectly predicted positive cases, and FN (False Negatives is incorrectly predicted negative cases.

3. RESULTS

3.1. Results of Resnet models

In this study, 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 3.

TABLE 3. 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

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 Figure 3.

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 Figure 4.

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 Figure 5.

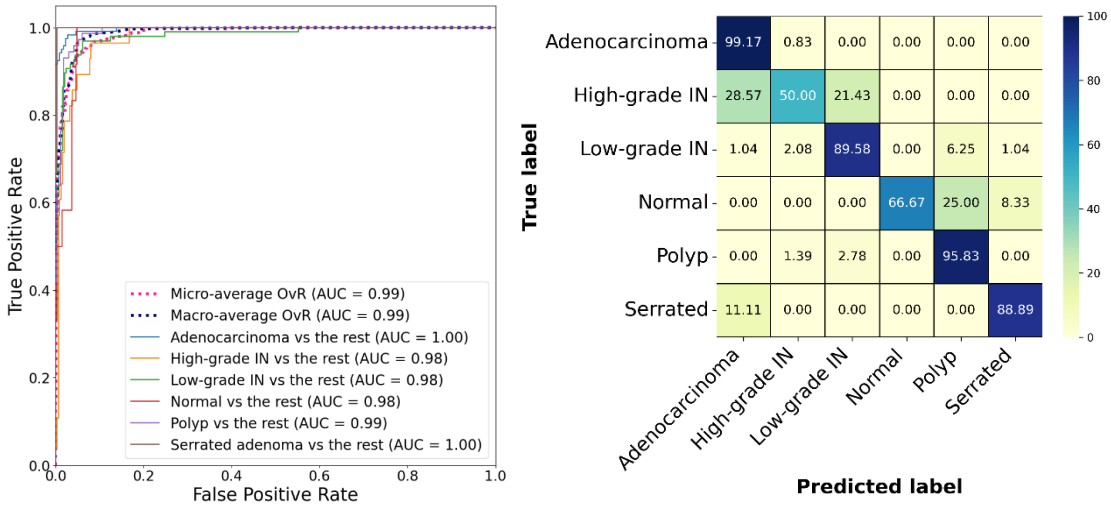

FIGURE 3. ResNet-18 classification result.

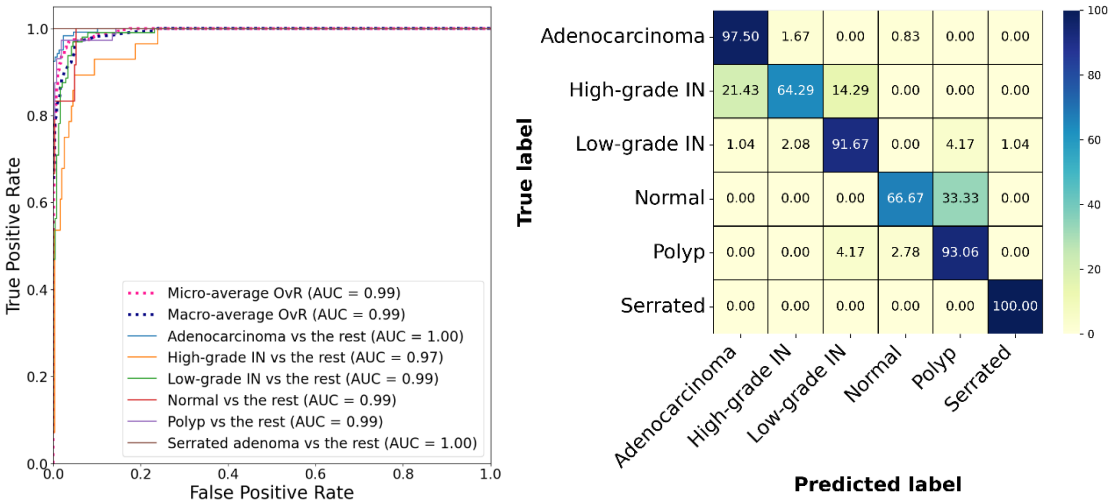

FIGURE 4. ResNet-34 classification result.

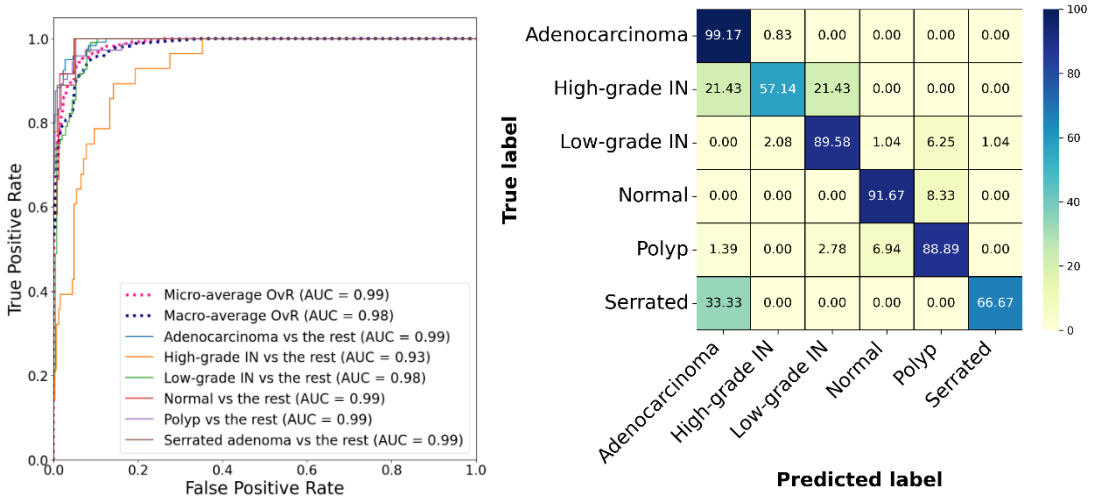

FIGURE 5. ResNet-18 classification result.

3.2. Compare ResNet models

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 4.

TABLE 4. 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

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.

3.3. Resnet model Grad-CAM visualizations

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 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 Figure 6.

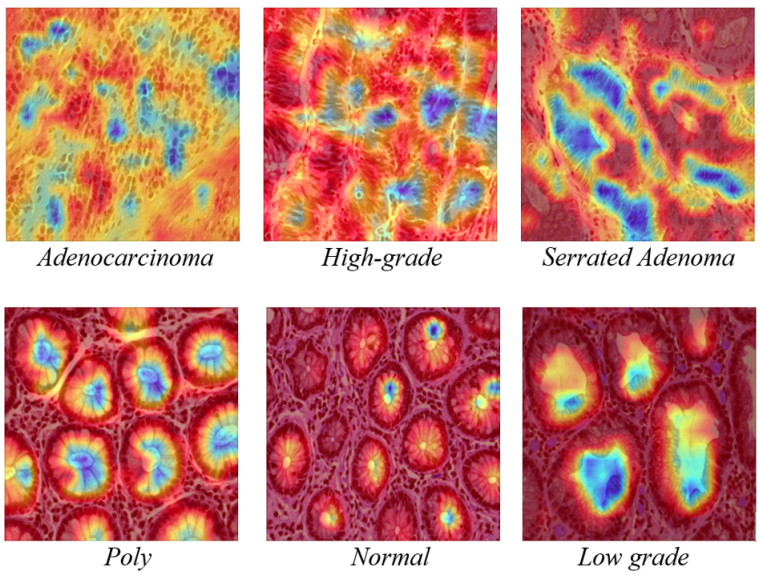

FIGURE 6. Resnet model GradCAM explanation on EBHI-Seg dataset.

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 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.

3.4. Swin Transformer Architectures for futher classification data

The result of Swin v2-T Small Evaluation on epoch 5 and 38 is illustrated in Figure 7. This architecture demonstrated robust performance in colorectal cancer classification across six distinct classes. The model

achieved a macro-average AUC of 0.98 and a micro-average AUC of 0.99, indicating strong overall discriminative capability. In class-specific performance, the model showed exceptional accuracy in identifying Adenocarcinoma with 95.00% correct classifications and only minimal misclassification into other categories. The model performed particularly well in detecting Polyp cases, achieving 95.14% accuracy. For Low-grade classifications, the model maintained high reliability with 93.23% accuracy. However, the architecture showed some limitations in High-grade detection, achieving only 51.79% accuracy with notable confusion between High-grade and other categories. The model demonstrated moderate performance in Normal and Serrated class detection, both achieving 66.67% accuracy. The ROC curve analysis revealed consistent performance across most classes with AUC values ranging from 0.94 to 0.99, with the High-grade class showing relatively lower discriminative capability.

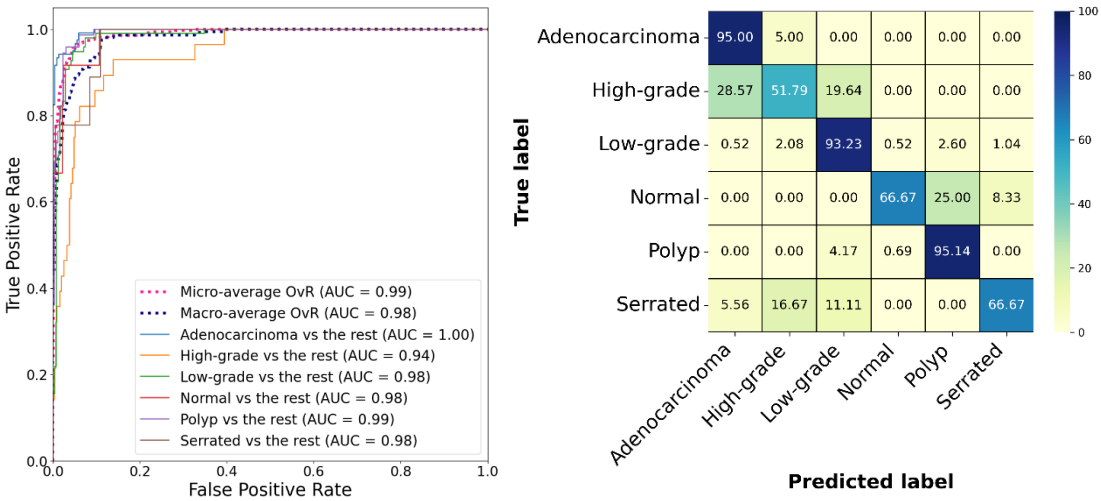

FIGURE 7. Swin v2-T Small Evaluation on epoch 5 and 38.

The result of Swin v2-T Tiny Evaluation on epoch 4 and 14 is illustrated in Figure 8. This architecture variant exhibited impressive performance characteristics despite its reduced parameter count. The architecture achieved perfect classification capability (AUC = 1.00) for both Adenocarcinoma and Serrated classes, demonstrating exceptional specificity in these critical categories. The model's performance in Adenocarcinoma detection was particularly noteworthy, achieving 98.75% accuracy in direct classi- fication. The architecture showed improved capabilities in High-grade classification compared to its larger counterpart, reaching 58.93% accuracy. The model maintained strong performance in Polyp detection with 94.44% accuracy and showed robust Low- grade classification at 91.67%. However, the architecture demonstrated limitations in Normal class discrimination, achieving only 37.50% accuracy with significant misclassification into the Polyp category. The Serrated class detection showed marked improvement over other architectures, reaching 88.89% accuracy. ROC analysis confirmed the model's strong discriminative capabilities with most classes achieving AUC values of 0.97 or higher.

The Window-8 variant of Swin v2-T Tiny presented unique performance characteristics attributed to its modified attention mechanism, which its result on epoch 7 and 20 is illustrated in Figure 9. This architecture

achieved superior performance in Ade- nocarcinoma detection with 99.17% accuracy, the highest among all variants. The model maintained consistent performance in High-grade classification at 51.79% accuracy, matching the base model's capabilities in this challenging category. Low-grade classification remained robust at 92.71% accuracy, demonstrating reliable discrimination capabilities. The architecture showed particular strength in Polyp classification, achieving 92.36% accuracy with minimal misclassification into other categories. The model's performance in Normal class detection reached 45.83%, showing improvement over the base Tiny variant but still presenting challenges in accurate discrimination. Serrated class detection achieved 72.22% accuracy, positioning this variant between the Small and Tiny architectures in terms of performance for this specific class. ROC curve analysis revealed excel- lent discriminative capabilities across all classes, with AUC values ranging from 0.95 to 1.00, and notably achieving perfect discrimination (AUC = 1.00) for the Serrated class.

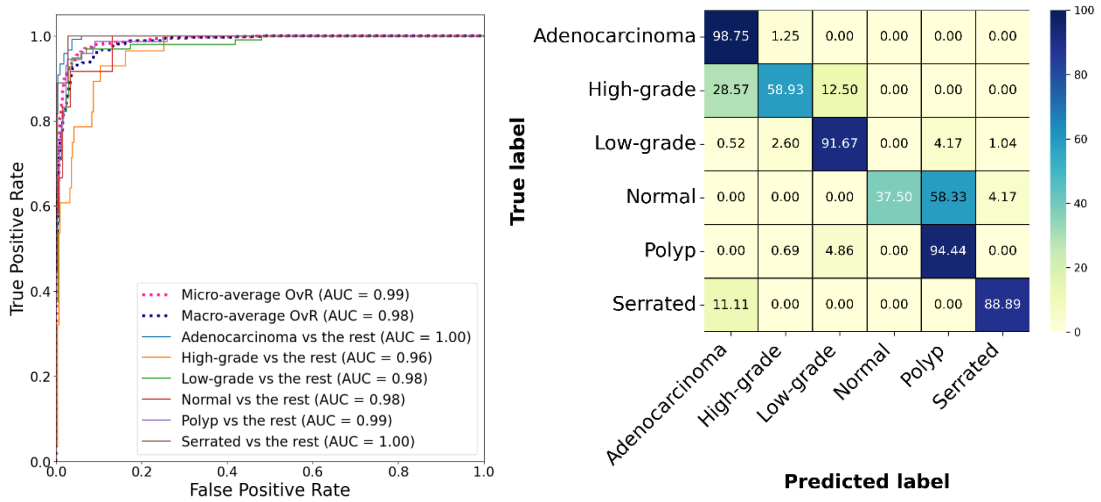

FIGURE 8. Swin v2-T Tiny Evaluation on epoch 4 and 14.

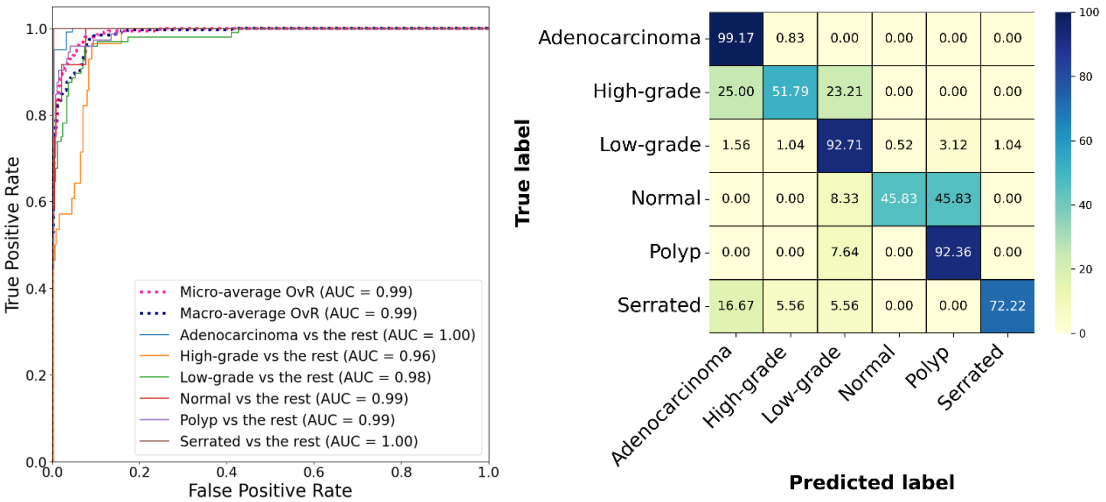

FIGURE 9. Swin v2-T Tiny (Window-8) Evaluation on epoch 7 and 20.

3.5. Compare Swin Transformer

TABLE 5. Swin Transformer Models Performance Comparison (Recall/Sensitivity %).

Class	Model		
	Swin v2-T Small	**Swin v2-T Tiny**	**Swin v2-T Tiny (W8)**
Adenocarcinoma	95.00	98.75	99.17
High-grade IN	51.79	58.93	51.79
Low-grade IN	93.23	91.67	92.71
Normal	66.67	37.50	45.83
Polyp	95.14	94.44	92.36
Serrated	66.67	88.89	72.22

TABLE 6. ROC Curve Analysis (AUC Values)

Class	Model		
	Swin v2-T Small	**Swin v2-T Tiny**	**Swin v2-T Tiny (W8)**
Adenocarcinoma	0.99	1.00	0.99
High-grade IN	0.99	0.97	0.95
Low-grade IN	0.94	0.98	0.98
Normal	0.98	0.97	0.99
Polyp	0.98	0.98	0.99
Serrated	0.99	1.00	1.00
Macro-average	0.98	0.98	0.98
Micro-average	0.99	0.99	0.99

When comparing the three Swin Transformer variants, each architecture demonstrated distinct advantages and trade-offs in colorectal cancer classification (Table 5, Table 6). The Swin v2-T Small architecture provided the most balanced performance across all classes, making it suitable for general-purpose applications where consistent reliability is crucial. However, both Tiny variants achieved superior performance in specific aspects while maintaining a smaller parameter footprint.

Notably, while the Swin v2-T Small achieved 95.00% accuracy in Adenocarcinoma detection, both Tiny variants surpassed this performance, with the Window-8 variant reaching an impressive 99.17% accuracy. The window-based attention mechanism in the Window-8 variant particularly enhanced its discriminative capabilities, as evidenced by its consistently high AUC values (0.95-1.00) across all classes. However, all three architectures showed similar challenges with High-grade classification, with accuracies ranging from 51.79% to 58.93%, suggesting this remains a challenging category regardless of model architecture.

An interesting observation was the trade-off in Normal class detection, where Swin v2-T Small performed better (66.67%) com- pared to both Tiny variants (37.50% and 45.83% respectively). Conversely, the Tiny variants

showed superior performance in Serrated class detection, with the base Tiny variant achieving 88.89% accuracy compared to the Small variant's 66.67%. This suggests that model complexity does not uniformly correlate with improved performance across all classes, and architectural choices may need to be guided by specific clinical priorities in colorectal cancer detection.

These findings indicate that while the Swin v2-T Small architecture offers reliable general-purpose performance, the Tiny variants, particularly with the Window-8 attention mechanism, present compelling alternatives when computational efficiency and specific class detection are prioritized. The choice between these architectures should be guided by the specific requirements of the clinical application, balancing the trade-offs between model size, computational efficiency, and class-specific performance requirements.

An interesting observation was the trade-off in Normal class detection, where Swin v2-T Small performed better (66.67%) compared to both Tiny variants (37.50% and 45.83% respectively). Conversely, the Tiny variants showed superior performance in Serrated class detection, with the base Tiny variant achieving 88.89% accuracy compared to the Small variant's 66.67%. This suggests that model complexity does not uniformly correlate with improved performance across all classes, and architectural choices may need to be guided by specific clinical priorities in colorectal cancer detection.

These findings indicate that while the Swin v2-T Small architecture offers reliable general-purpose performance, the Tiny variants, particularly with the Window-8 attention mechanism, present compelling alternatives when computational efficiency and specific class detection are prioritized. The choice between these architectures should be guided by the specific requirements of the clinical application, balancing the trade-offs between model size, computational efficiency, and class-specific performance requirements.

3.6. Swin transformer model Grad-CAM visualizations

The Grad-CAM visualizations of the Swin Transformer V2 model provide strong evidence of its ability to accurately distinguish various classes of CRC histopathology images. The attention-based mechanism of the transformer allows for adaptive focus on diagnostically relevant regions, effectively capturing both local and global contextual information within tissue samples. The generated heatmaps highlight critical areas of interest, demonstrating a strong alignment with expert pathologist annotations and reinforcing the interpretability of the model's predictions. This capability is particularly beneficial in identifying subtle morphological differences between CRC subtypes, which may be challenging for conventional CNN-based models. The hierarchical self-attention mechanism in Swin V2 enables the model to process high-resolution histopathology images with remarkable precision, making it well-suited for clinical applications. Figure 10 illustrates the application of Grad-CAM on the testing set, showcasing the model's attention patterns.

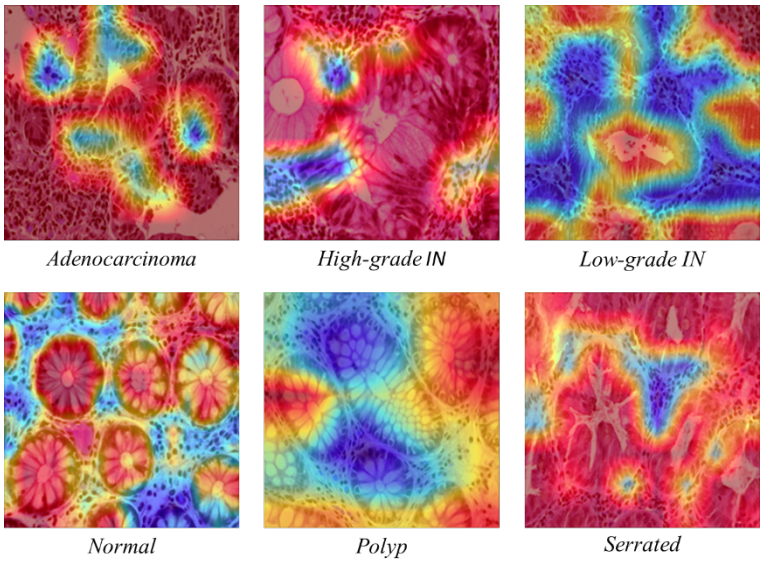

FIGURE 10. Swin Transformer model GradCAM explanation on EBHI-Seg dataset.

4. DISCUSSION

This study demonstrates the efficacy of deep learning models, particularly ResNet and Swin Transformer V2 architectures, in classifying colorectal cancer (CRC) subtypes from histopathology images. Among the ResNet variants, ResNet-34 offered the best balance of performance and computational efficiency, achieving the highest overall accuracy (91.10%) and perfect top-3 accuracy, outperforming both shallower (ResNet-18) and deeper (ResNet-50) networks. While ResNet-50 achieved slightly higher AUC scores in certain classes, its increased complexity did not translate into improved overall accuracy, highlighting that moderate network depth may be optimal for CRC classification tasks.

To address the challenge of data imbalance, we applied a comprehensive set of data augmentation techniques, which significantly improved the class distribution (from a 13.7:1 to 3.2:1 ratio) and enhanced model generalizability. This step proved critical, especially in boosting performance for underrepresented classes such as serrated and normal tissues. Additionally, Grad-CAM visualizations provided meaningful insights into the models' decision-making by highlighting regions relevant to CRC diagnosis. These heatmaps aligned well with pathological features, supporting the clinical interpretability of our models.

The Swin Transformer V2 architectures further validated the feasibility of vision transformers for medical imaging. Notably, the Tiny variant with Window-8 attention achieved the highest class-specific performance in adenocarcinoma (99.17%) and serrated (88.89%), though challenges remained in normal and high-grade IN classifications across all Swin models. These findings suggest that transformer-based models, though more computationally intensive, offer strong alternatives to traditional CNNs, especially in precision-focused clinical applications.

Despite these promising results, misclassifications were observed in classes with limited data, underscoring the ongoing need for larger, more diverse datasets. Future research should focus on dataset expansion, model ensembling, and real-world validation through collaboration with pathologists to fully integrate AI tools into

clinical workflows.

5. CONCLUSION

In this study, we explored the use of deep learning architectures-ResNet and Swin Transformer V2-for the classification of colorectal cancer subtypes from histopathology images. Our findings indicate that ResNet-34 delivers the most balanced performance, achieving high accuracy and strong generalization across multiple CRC classes. The integration of data augmentation techniques effectively mitigated class imbalance, improving model robustness, especially for underrepresented categories. Furthermore, Grad-CAM visualizations enhanced model interpretability by highlighting diagnostically relevant regions, supporting potential clinical integration.

In optimizing model training, we identified several crucial factors for achieving superior performance: the use of a low learning rate to maintain alignment with pre-trained weights, the implementation of the Adam optimizer for efficient convergence, and the adoption of a power-of-two batch size to maximize memory efficiency. Together, these strategies contribute to a robust and computationally efficient pipeline for automated CRC classification. Our results demonstrate the potential of AI-assisted diagnostics to enhance accuracy, interpretability, and clinical utility in colorectal cancer care.

Authorship

Conceptualization: Thang Truong Le, Vinh-Thuyen Nguyen-Truong. Experiment design: Thang Truong Le. Data collection and analysis: Nghia Trong Le Phan, Quang Duong Van Nhat, Mqondisi Fortune Mavuso, Phuc Nguyen Thien Dao, Huy Ngoc Anh Nguyen, Tien Thuy Mai, Kiep Thi Quang. Writing: Nghia Trong Le Phan, Mqondisi Fortune Mavuso, Quang Duong Van Nhat. Review and edit: Thang Truong Le, Vinh-Thuyen Nguyen-Truong. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENT

We thank to Nguyen-Truong Vinh was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.ThS.JVN.09, and his support in content consultation and coding. We also acknowledge financial support from Mr. Thang Truong Le and Ms. Kiep Thi Quang.

REFERENCES

[1] Y. Xi and P. Xu, "Global colorectal cancer burden in 2020 and projections to 2040," *Transl. Oncol.*, vol. 14, no. 10, p. 101174, Oct. 2021, doi: 10.1016/j.tranon.2021.101174.

[2] Y.-C. Lee, Y.-L. Lee, J.-P. Chuang, and J.-C. Lee, "Differences in survival between colon and rectal cancer from SEER data," *PloS One*, vol. 8, no. 11, p. e78709, 2013, doi: 10.1371/journal.pone.0078709.

[3] J. C. Thijs *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.*, vol. 91, no. 10, pp. 2125–2129, Oct. 1996.

[4] A. Cervantes *et al.*, "Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up," *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, vol. 34, no. 1, pp. 10–32, Jan. 2023, doi: 10.1016/j.annonc.2022.10.003.

[5] N. Saoudi González *et al.*, "Unravelling the Complexity of Colorectal Cancer: Heterogeneity, Clonal Evolution, and Clinical Implications," *Cancers*, vol. 15, no. 16, p. 4020, Aug. 2023, doi: 10.3390/cancers15164020.
[6] M. P. Singh, S. Rai, A. Pandey, N. K. Singh, and S. Srivastava, "Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine," *Genes Dis.*, vol. 8, no. 2, pp. 133–145, Mar. 2021, doi: 10.1016/j.gendis.2019.10.013.
[7] Y. Wang, X. He, H. Nie, J. Zhou, P. Cao, and C. Ou, "Application of artificial intelligence to the diagnosis and therapy of colorectal cancer," *Am. J. Cancer Res.*, vol. 10, no. 11, pp. 3575–3598, Nov. 2020.
[8] S. Alinia *et al.*, "Exploring the impact of stage and tumor site on colorectal cancer survival: Bayesian survival modeling," *Sci. Rep.*, vol. 14, no. 1, p. 4270, Feb. 2024, doi: 10.1038/s41598-024-54943-8.
[9] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial Intelligence in Colorectal Cancer Screening, Diagnosis and Treatment. A New Era," *Curr. Oncol. Tor. Ont*, vol. 28, no. 3, pp. 1581–1607, Apr. 2021, doi: 10.3390/curroncol28030149.
[10] Z. Yin, C. Yao, L. Zhang, and S. Qi, "Application of artificial intelligence in diagnosis and treatment of colorectal cancer: A novel Prospect," *Front. Med.*, vol. 10, p. 1128084, Mar. 2023, doi: 10.3389/fmed.2023.1128084.
[11] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," *Procedia Comput. Sci.*, vol. 179, pp. 423–431, Jan. 2021, doi: 10.1016/j.procs.2021.01.025.
[12] M. Y. Ahmad, A. Mohamed, Y. A. Mohd Yusof, and S. A. Md. Ali, "Colorectal cancer image classification using image pre-processing and multilayer Perceptron," *2012 Int. Conf. Comput. Inf. Sci. ICCIS*, pp. 275–280, Jun. 2012, doi: 10.1109/ICCISci.2012.6297253.
[13] D. Bychkov *et al.*, "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, p. 3395, Feb. 2018, doi: 10.1038/s41598-018-21758-3.
[14] M. H. Motlagh *et al.*, "Breast Cancer Histopathological Image Classification: A Deep Learning Approach," Jan. 04, 2018, *bioRxiv*. doi: 10.1101/242818.
[15] P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *J. Cancer Res. Clin. Oncol.*, vol. 145, no. 4, pp. 829–837, Apr. 2019, doi: 10.1007/s00432-018-02834-7.
[16] N. Zange Tsaku *et al.*, "Texture-based Deep Learning for Effective Histopathological Cancer Image Classification," *2019 IEEE Int. Conf. Bioinforma. Biomed. BIBM*, Nov. 2019, doi: 10.1109/BIBM47256.2019.8983226.
[17] M. Sharkas and O. Attallah, "Color-CADx: a deep learning approach for colorectal cancer classification through triple convolutional neural networks and discrete cosine transform," *Sci. Rep.*, vol. 14, no. 1, p. 6914, Mar. 2024, doi: 10.1038/s41598-024-56820-w.
[18] C. Miyagawa *et al.*, "Histopathological subtyping of high-grade serous ovarian cancer using whole slide imaging," *J. Gynecol. Oncol.*, vol. 34, no. 4, p. e47, Jul. 2023, doi: 10.3802/jgo.2023.34.e47.
[19] M. Li *et al.*, "Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images," *IEEE Access*, vol. 9, pp. 53687–53707, 2021, doi: 10.1109/ACCESS.2021.3071057.
[20] L. Shi *et al.*, "EBHI-Seg: A novel enteroscope biopsy histopathological hematoxylin and eosin image dataset for image segmentation tasks," *Front. Med.*, vol. 10, p. 1114673, 2023, doi: 10.3389/fmed.2023.1114673.
[21] J. Gildenblat, *jacobgil/pytorch-grad-cam*. (Mar. 17, 2025). Python. Accessed: Mar. 17, 2025. [Online]. Available: https://github.com/jacobgil/pytorch-grad-cam
[22] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, Dec. 2020, doi: 10.1007/s10462-020-09825-6.
[23] W. Xu, Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Comput. Methods Programs Biomed.*, vol. 240, p. 107660, Oct. 2023, doi: 10.1016/j.cmpb.2023.107660.

[24] Z. Liu *et al.*, “Swin Transformer V2: Scaling Up Capacity and Resolution,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 11999–12009. doi: 10.1109/CVPR52688.2022.01170.
[25] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, “Transformers in medical image segmentation: A review,” *Biomed. Signal Process. Control*, vol. 84, p. 104791, Jul. 2023, doi: 10.1016/j.bspc.2023.104791.