

Attention-Enhanced Xception Network for Automated Microsatellite Instability Classification in Colorectal Cancer Histopathology

Yunxia Huang

Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital

Yanzong Lin

The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Teaching Hospital of Fujian Medical University

Qingyang Zhuang

Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital

Chenjun Zhang

Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital

Junxin Wu

junxinwufj@aliyun.com

Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital

Article

Keywords: Microsatellite instability, colorectal cancer, deep learning, histopathology, attention mechanism

Posted Date: March 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8106522/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background:

Microsatellite instability (MSI) is a critical molecular biomarker in colorectal cancer (CRC), guiding prognosis and immunotherapy decisions. However, standard MSI detection methods such as immunohistochemistry (IHC) and PCR remain costly, time-consuming, and inaccessible in many settings. Recent advances in computational pathology have enabled the use of deep learning models to extract clinically meaningful features from routine hematoxylin and eosin (H&E)-stained slides, offering a non-invasive and scalable alternative.

Methods:

We developed an interpretable deep learning framework integrating the Xception backbone with a Convolutional Block Attention Module (CBAM) to enhance MSI classification performance. The model was trained and tested on a large-scale public dataset containing 194,178 image patches derived from FFPE histology slides of CRC patients in the TCGA cohort. Model performance was evaluated using standard metrics including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Grad-CAM was employed to visualize model attention and assess interpretability.

Results:

Our proposed model achieved superior performance with an AUC of 0.9551, an accuracy of 88.4%, a precision of 90.7%, a recall of 84.1%, and an F1-score of 87.3%, outperforming baseline CNN architectures including DenseNet121, InceptionV3, ResNet50, and others. Grad-CAM analysis revealed that the model consistently focused on histopathological features associated with MSI, such as tumor-infiltrating lymphocytes and glandular disorganization, while suppressing non-diagnostic regions.

Conclusion:

The attention-enhanced Xception model provides a robust, interpretable, and clinically scalable solution for MSI status prediction from H&E-stained slides. It holds promise as a non-invasive alternative to traditional assays, especially in resource-constrained environments, and supports the growing role of deep learning in precision oncology.

1. Background

Colorectal cancer (CRC) ranks as the third most prevalent malignancy globally, with approximately 1.9 million new cases annually, and mortality rates second only to lung cancer(1). In China, its incidence reaches 28.2 per 100,000, posing a major public health challenge(2). Molecularly, CRC evolves through two primary pathways: chromosomal instability (CIN, 80%–85%) and microsatellite instability (MSI, 10%–15%)(3, 4). MSI-high (MSI-H) tumors, driven by deficient DNA mismatch repair (dMMR), exhibit hypermutation and high tumor mutational burden (5), rendering them uniquely responsive to

immunotherapy(6, 7). Immune checkpoint inhibitors (ICIs) such as anti-PD-1/PD-L1 antibodies have revolutionized treatment for metastatic MSI-H CRC, improving survival outcomes with response rates of 30%–50% and reducing death risk by 63% compared to chemotherapy(8). Routine MSI testing remains inaccessible to 60–70% of CRC patients globally due to methodological limitations. Immunohistochemistry exhibits interpretive variability (5–10% discordance) and high tissue consumption, while PCR (polymerase chain reaction)-based capillary electrophoresis requires specialized equipment with prolonged turnaround. Next-generation sequencing (NGS), though accurate, is cost-prohibitive and infrastructure-dependent(9, 10). These barriers disproportionately affect resource-limited regions where molecular testing penetration falls below 20%(9).

Pathomics, the high-throughput quantitative analysis of digitized histopathology images—transforms routine H&E slides into mineable data repositories capturing molecular phenotype-reflective morphological, topological, and textural patterns(11). Integrated with deep learning (DL) architectures like convolutional neural networks, this approach enables accurate microsatellite instability prediction through three key advantages: cost-effectiveness via existing slides without additional reagents; enhanced accuracy through identification of subtle histological biomarkers; and clinical scalability using conventional scanners, particularly valuable in resource-limited settings(12, 13). This computational methodology demonstrates robust performance in multicenter validations while maintaining compatibility with archival specimens. Recent studies further demonstrate DL's ability to integrate multimodal data (e.g., clinical variables, mutation profiles) and provide biologically interpretable insights(14, 15).

Pathomics-DL integration represents a pragmatic strategy to democratize MSI testing, particularly in economically disadvantaged regions. By transforming ubiquitous H&E slides into predictive biomarkers, this approach bridges the gap between precision immunotherapy and global accessibility. Future efforts should focus on prospective validation and standardization of analytical pipelines to facilitate clinical adoption.

To address the limitations of conventional MSI testing and fully leverage the diagnostic potential embedded in routine histopathological slides, we propose a novel deep learning-based pathomics framework for the non-invasive prediction of MSI status in colorectal cancer. Specifically, our model integrates the Xception architecture with a Convolutional Block Attention Module (CBAM) to enhance feature extraction from complex tissue morphologies by adaptively focusing on discriminative microstructural patterns. This design aims to improve the interpretability and robustness of MSI prediction in real-world settings, particularly where molecular testing resources are limited. Through rigorous validation on annotated histology datasets, we demonstrate that our model can achieve high predictive accuracy while maintaining low computational complexity, offering a promising alternative to traditional molecular assays. This study provides a practical and scalable AI-driven strategy to democratize precision oncology and support timely treatment decision-making for colorectal cancer patients.

2. Methods

2.1 Data Collection

This study employed a publicly accessible digital pathology dataset retrieved from the Zenodo repository (<https://zenodo.org/records/2530835>), encompassing a total of 411,890 unique image patches derived from histological slides of colorectal and gastric cancer patients in The Cancer Genome Atlas (TCGA) cohort. All samples were obtained from formalin-fixed paraffin-embedded (FFPE) diagnostic slides, denoted as “DX” in the GDC data portal (<https://portal.gdc.cancer.gov/>). The original whole-slide SVS images, which form the basis of this dataset, are freely available through the GDC platform, and their preprocessing methodology has been well documented in prior technical reports.

To ensure uniformity and diagnostic relevance, all whole-slide images (WSIs) underwent a standardized preprocessing pipeline prior to patch extraction. Tumor regions were first automatically detected, followed by cropping of relevant areas into image patches of 224×224 pixels at a resolution of $0.5 \mu\text{m}$ per pixel. Color normalization was conducted using the Macenko method, which corrects for staining variability across slides and preserves morphological features essential for downstream analysis. Based on curated molecular annotations from TCGA, each patient case was assigned a microsatellite stability status, categorized as either microsatellite stable (MSS) or microsatellite instable/mutated (MSIMUT). Subsequently, patients were randomly divided into training and testing cohorts at the patient level, thereby ensuring that no overlap existed between images of the same patient across different splits. To address class imbalance in the training cohort, excess MSS patches were randomly removed, resulting in a balanced training distribution.

The dataset includes image patches from both colorectal cancer (CRC) and gastric cancer (STAD) patients, with well-matched distributions of MSS and MSIMUT samples across training and testing sets. Specifically, the CRC cohort consists of 46,704 MSS and 46,704 MSIMUT patches for training, and 70,569 MSS and 29,335 MSIMUT patches for testing. Similarly, the STAD cohort comprises 50,285 MSS and 50,285 MSIMUT training patches, along with 90,104 MSS and 27,904 MSIMUT testing patches. In total, 194,178 CRC patches and 217,712 STAD patches were curated, all of which were derived from FFPE samples, providing a diverse and large-scale dataset for developing deep learning models in histopathological image analysis. This dataset forms the foundation of our study and serves as a benchmark for evaluating model performance in automated MSI prediction. The CRC section was used in this study.

2.2 Data Preprocessing

Prior to model training, all image patches were standardized to a fixed resolution of 224×224 pixels with a spatial resolution of $0.5 \mu\text{m}/\text{px}$ and normalized to the $[0, 1]$ range. To enhance generalizability and reduce overfitting, standard data augmentation techniques were applied to the training set, including random flips, rotations, and color jittering. These augmentations simulate common staining and orientation variations in histological slides. The test set remained unaltered to ensure evaluation

consistency. All data were processed using a custom pipeline enabling efficient loading and on-the-fly augmentation during model training.

2.3 Model Architecture

The proposed deep learning model adopts a hybrid architecture that combines the Xception backbone with a Convolutional Block Attention Module (CBAM) to enhance spatial and channel-wise feature extraction from histopathological image patches. The model takes as input RGB patches of size $299 \times 299 \times 3$, which are resized from the original 224×224 patches to meet the default input requirement of the Xception network. The overall architecture of the model proposed in this study is shown in Figure 1.

Figure 1. Schematic diagram of the overall architecture of the model proposed in this study.

The backbone consists of three major flows: Entry Flow, Middle Flow, and Exit Flow. The Entry Flow begins with two standard convolutional layers using 3×3 kernels with 32 and 64 filters respectively, followed by three depthwise separable convolutional blocks. Each of these blocks employs 3×3 kernels with 128 filters and stride settings of 1 or 2 to progressively reduce spatial resolution while preserving semantic richness. Downsampling is achieved through parallel 1×1 convolutions with stride 2 in the residual connections.

The Middle Flow consists of eight identical residual blocks, each containing three depthwise separable convolutional layers with 3×3 kernels and 728 filters, all using a stride of 1. This stage enables deeper hierarchical feature abstraction with minimal parameter overhead, owing to the efficient design of separable convolutions. The Exit Flow contains two blocks with increasing depth and complexity, including 3×3 convolutions with 728 and 1024 filters, followed by a final separable convolution with stride 2 to further condense spatial information. Residual connections continue with 1×1 convolutions and stride 2.

Following the Xception backbone, we integrate a CBAM attention module to refine the feature representations. The CBAM includes two sequential submodules: a Channel Attention Module (CAM) and a Spatial Attention Module (SAM). The CAM leverages global average pooling and max pooling across spatial dimensions, followed by two shared fully connected layers (with a reduction ratio of 16), to produce a channel-wise attention map. This map is used to rescale the input feature maps, emphasizing informative channels. The SAM then aggregates spatial information by applying average and max pooling along the channel axis, concatenating them, and passing the result through a 7×7 convolutional layer and sigmoid activation to generate a spatial attention map. This spatial map is used to further modulate the refined features.

The output from the CBAM-enhanced backbone is passed through a global average pooling layer and two fully connected layers with ReLU activation (Dense–ReLU–Dense), followed by a final softmax classifier that outputs the predicted probability for MSI or MSS status. The entire model is trained using categorical cross-entropy loss and optimized using the Adam optimizer with a learning rate of 0.0001.

This attention-enhanced Xception architecture is designed to selectively focus on microstructural features critical for MSI classification, such as nuclear morphology, glandular structure, and tumor-infiltrating lymphocytes, thereby improving both classification accuracy and interpretability in complex histopathological images.

2.4 Comparative Evaluation

To evaluate the effectiveness of our proposed attention-enhanced architecture, we conducted a comprehensive comparison with six widely used convolutional neural networks (CNNs): Xception, DenseNet121, InceptionV3, MobileNetV2, ResNet50, and VGG16. All models were trained and tested on the same dataset under identical experimental settings to ensure fair comparison, including identical image input size, data splits, preprocessing pipeline, training epochs, batch size, and optimizer configuration.

2.5 Model Evaluation

We adopted a series of standard evaluation metrics widely used in binary classification tasks within medical imaging. These include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). Accuracy quantifies the overall classification correctness, while precision and recall respectively evaluate the model's ability to avoid false positives and false negatives—both of which are clinically critical in the context of MSI testing. The F1-score, as the harmonic mean of precision and recall, offers a balanced measure for imbalanced datasets. The AUC metric was used to assess the model's discriminative capacity across various threshold settings, providing a threshold-independent indicator of predictive power. In addition to quantitative metrics, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize the spatial regions within histopathological patches that contributed most significantly to the model's decision-making. Grad-CAM generates class-specific heatmaps by computing the gradient of the target class score with respect to the final convolutional layer output. These heatmaps were overlaid on the original histological images to reveal whether the model accurately localized diagnostically relevant features, such as abnormal glandular architecture or lymphocyte infiltration. This interpretability analysis not only enhanced the clinical plausibility of model predictions but also confirmed the efficacy of the attention mechanism in guiding the network's focus toward biologically meaningful regions.

3. Results

3.1 Experimental Setup

All experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB), 128 GB RAM, and an AMD Ryzen Threadripper CPU running Ubuntu 20.04. The model was implemented in Python 3.8 using the PyTorch 1.12 framework with CUDA 11.6. The dataset was split at the patient level into 70% for training and 30% for testing, following the original data partitioning protocol. No cross-validation was performed. The model was trained using the Adam optimizer with a learning rate of

0.0001, a batch size of 32, and trained for 100 epochs with early stopping based on validation loss. Hyperparameters were manually tuned based on validation performance. All models were trained under identical conditions to ensure comparability.

3.2 Quantitative Results

3.2.1 Performance of the Proposed Model

The proposed attention-enhanced Xception model demonstrated robust and consistent performance in distinguishing microsatellite instability (MSI) from microsatellite stable (MSS) cases. As shown in Figure 2A, the receiver operating characteristic (ROC) curve yielded an area under the curve (AUC) of 0.9551, indicating strong discriminative capability across various decision thresholds. This high AUC confirms the model's ability to balance sensitivity and specificity in classifying complex histopathological features. The corresponding confusion matrix (Figure 2B) illustrates that the model achieved a recall of 84.1% for MSI cases and a precision of 92.0% for MSS cases, with an overall classification accuracy of 88.4%. The relatively low false negative rate for MSI is particularly critical in clinical scenarios, where misclassification of MSI-H tumors could lead to suboptimal treatment planning.

In addition, the training dynamics are visualized in Figure 2C, where both the training and validation accuracy curves exhibit steady convergence, while the loss curves decrease consistently without signs of overfitting. By the 10th epoch, the model reached a validation accuracy exceeding 87%, confirming its stability and generalization ability on unseen data. Together, these quantitative results validate the effectiveness of integrating CBAM attention with the Xception backbone, enabling the model to learn diagnostically relevant microstructures and yield high-performance MSI classification in H&E-stained histology images.

3.2.2 Comparison with Baseline Models

We compared its performance against six widely adopted convolutional neural network (CNN) baselines: Xception, DenseNet121, InceptionV3, MobileNetV2, ResNet50, and VGG16. All models were trained under identical conditions on the same dataset, ensuring fairness in comparative evaluation.

As shown in Figure 3, our model achieved the highest area under the ROC curve (AUC) of 0.9551, outperforming all other models. Specifically, Xception and DenseNet121 also exhibited strong performance, with AUCs of 0.9407 and 0.9300 respectively, while ResNet50 and InceptionV3 reached AUCs of 0.9142 and 0.9234. MobileNetV2, although computationally efficient, recorded a lower AUC of 0.9060, indicating reduced discriminative power. VGG16, the shallowest architecture in the comparison, performed the worst with an AUC of 0.5000, barely above random chance. Detailed metrics are summarized in Table 1, where our model consistently achieved the highest values across key indicators: accuracy (88.4%), precision (90.7%), recall (84.1%), and F1-score (87.3%). Compared to the vanilla Xception model, the integration of the CBAM attention mechanism improved AUC by 1.45 percentage points and yielded better balance between sensitivity and specificity. Notably, our model exhibited higher

recall than DenseNet121 and InceptionV3, which is especially important for minimizing missed MSI-H cases in clinical decision-making. These results demonstrate that the proposed model not only preserves the efficiency of Xception’s separable convolutions but also gains enhanced interpretability and feature discrimination through attention-guided learning. This advantage becomes particularly evident in the high-resolution, texture-rich setting of histopathological image classification.

Table 1 Comparison results of various evaluation indicators of each model.

Model Name	Accuracy	Precision	Recall	F1 Score
Xception	85.70%	89.49%	80.90%	84.98%
DenseNet121	84.55%	80.77%	90.70%	85.45%
InceptionV3	83.90%	80.60%	89.30%	84.72%
MobileNetV2	81.10%	76.22%	90.40%	82.71%
ResNet50	83.00%	83.74%	81.90%	82.81%
VGG16	50.00%	50.00%	100.00%	66.67%
Ours	88.05%	85.26%	92.00%	88.50%

3.3 Grad-CAM-Based Interpretability Analysis

Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize attention regions over histopathological patches. The resulting heatmaps, as shown in Figure 4A (MSI) and Figure 4B (MSS), highlight the spatial features that contributed most significantly to the classification outcomes.

In MSI-labeled cases (Figure 4A), the model consistently focused on dense tumor-infiltrating lymphocytes (TILs), disrupted glandular architecture, and regions with cribriform or poorly differentiated growth patterns. These highlighted areas align with known morphological correlates of microsatellite instability, including increased immune activity, high mutational burden, and loss of glandular polarity—features that have been widely reported as pathological hallmarks of dMMR-driven tumors. The model’s attention to peritumoral immune infiltration further supports its potential as a proxy for immunotherapy-relevant molecular subtyping. Conversely, in MSS cases (Figure 4B), the attention maps predominantly covered well-formed glandular structures, organized epithelial boundaries, and regions with low inflammatory cell density. This aligns well with the clinical phenotype of MSS tumors, which typically exhibit more stable genomic profiles and conventional glandular differentiation. Notably, the model avoided non-diagnostic regions such as large blank spaces, adipose tissue, or stroma-dense areas, demonstrating its ability to suppress irrelevant background information.

4. Discussion

In this study, we proposed an interpretable deep learning framework for the automated prediction of microsatellite instability (MSI) status in colorectal cancer using routine H&E-stained histopathological images. By integrating the Xception backbone with a Convolutional Block Attention Module (CBAM), our model was designed to selectively enhance features relevant to MSI-related morphological patterns while suppressing irrelevant background information. Trained on a large-scale, publicly available TCGA-derived dataset of FFPE digital slides, the model achieved superior classification performance compared to six state-of-the-art convolutional neural networks, including DenseNet121, InceptionV3, and ResNet50. Quantitatively, our method reached an AUC of 0.9551 and an F1-score of 87.3%, and it demonstrated a strong ability to localize biologically meaningful regions via Grad-CAM visualizations. These findings suggest that attention-guided architectures can effectively capture the subtle histological cues associated with mismatch repair deficiency, offering a scalable and non-invasive alternative to conventional molecular testing methods.

From an algorithmic perspective, the superior performance of our model can be attributed to the synergistic combination of the Xception backbone and the CBAM attention mechanism. The Xception architecture, built upon depthwise separable convolutions, enables efficient extraction of fine-grained spatial features while maintaining a relatively low parameter count. This is particularly advantageous for histopathological images, which are inherently high-resolution and contain abundant local texture variations critical for downstream classification tasks. However, while Xception provides a strong baseline for feature representation, it lacks explicit mechanisms to guide the model's focus toward diagnostically relevant regions. The introduction of the CBAM module addresses this limitation by incorporating both channel and spatial attention in a lightweight and modular fashion. The channel attention component adaptively recalibrates feature responses across channels, emphasizing filters that capture biologically meaningful structures such as lymphocyte density, glandular irregularities, and nuclear atypia. The spatial attention module further enhances this by identifying salient regions within the tissue plane, allowing the model to prioritize informative cellular arrangements while suppressing irrelevant areas such as adipose tissue or scanning artifacts. This two-stage attention design enables the network to achieve a more precise and biologically aligned representation of MSI-associated morphological phenotypes, as confirmed by our Grad-CAM visualizations.

Accurate determination of microsatellite instability (MSI) status is critical for the diagnosis and management of colorectal cancer, as it serves as a key determinant for treatment regimen selection and prognostic stratification. Currently, immunohistochemistry (IHC) and polymerase chain reaction (PCR) remain the conventional gold standards for clinical microsatellite instability (MSI) assessment (16, 17), both of which are explicitly recommended in guidelines from the American Society of Clinical Oncology (ASCO) and the National Comprehensive Cancer Network (NCCN)(16, 17). Nevertheless, these traditional approaches are associated with inherent limitations. Specifically, IHC requires the evaluation of multiple mismatch repair (MMR) proteins, involving a cumbersome workflow; moreover, its result interpretation is highly susceptible to inter-observer variability and subjective bias(18). In contrast, PCR relies on specialized laboratory equipment and well-trained technical personnel, often resulting in prolonged turnaround times and compromised efficiency in high volume clinical settings(19). In

comparison, the deep learning-based model developed in our study demonstrates distinct advantages. Regarding accuracy, this study showed that the model achieved an area under the receiver operating characteristic curve (AUC) of over 0.9 for MSI status prediction, outperforming the diagnostic accuracy of traditional IHC and PCR assays. In terms of efficiency, leveraging the high computational capacity of modern computing systems, the deep learning model can process large scale pathological omics datasets within a short timeframe, rapidly generating MSI status predictions. This not only significantly reduces the overall detection latency but also holds great potential for providing timely and precise insights to support clinical decision-making in routine practice.

A key consideration for translating the deep learning-based pathomics model for MSI prediction into clinical practice is its use in primary hospitals and resource-limited regions. Traditional IHC and PCR are costly, mainly due to expensive reagents and specialized equipment, but this model is cost-effective: after validation, its operational costs only cover affordable basic digital slide scanning and routine computing hardware, eliminating recurring consumable expenses of traditional methods. The model also standardizes MSI detection workflows—critical where experienced pathologists are scarce. Unlike IHC, which has inter-observer variability, it uses fixed algorithms for objective, reproducible MSI results, reducing reliance on specialized personnel (a common bottleneck in underserved areas). Furthermore, the model needs little complex equipment, only basic slide digitalization tools and standard computers. With low cost, standardized workflow, and simple operation, it has great value in underserved regions, bridging MSI diagnostic gaps and enabling more equitable precision care for colorectal cancer patients.

This study has key limitations: the sample size for model training/validation is modest, restricting its ability to capture heterogeneous colorectal cancer pathological features across populations and limiting real-world representativeness; additionally, the model's generalization needs verification, as its performance was only tested in specific cohorts, with adaptability to diverse regions, ethnicities, or rare subtypes unproven. To address these and boost clinical translation, future research should focus on two areas: integrating multi-modal data (clinical demographics, genomics, radiology) with pathological omics to improve diagnostic accuracy; and conducting large-scale, multi-center trials to validate the model's effectiveness, providing evidence for routine clinical use.

5. Conclusion

In conclusion, this study demonstrates that an attention-augmented deep learning model based on Xception and CBAM can accurately and interpretably predict MSI status from routine H&E-stained slides. The proposed method offers a non-invasive, cost-effective alternative to conventional molecular assays, with strong potential for clinical integration, particularly in resource-limited settings. These findings highlight the value of combining efficient architectures with attention mechanisms in computational pathology and lay the groundwork for broader applications in precision oncology.

Declarations

Author Contributions

Yunxia Huang and Yanzong Lin wrote the main manuscript text. Qingyang Zhuang was responsible for formal analysis. Chenjun Zhang prepared the figures and tables. Junxin Wu was responsible for project administration. All authors reviewed the manuscript.

Funding

This research was funded by Fujian Province Science and Technology Innovation Joint Funding Project (2024Y9617 and 2025Y9502), Fujian Cancer Hospital In-Hospital Funding Program [2025YNPT006], Xiamen Medical and Health Guidance Project (3502Z20244ZD1040 and 3502Z20244ZD1050) and the National Clinical Key Specialty Construction Program.

Institutional Review Board Statement

Data from TCGA (public, de-identified, originally IRB-approved). No additional IRB approval needed per public genomic guidelines and institutional policies.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Acknowledgments

No acknowledgments are applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

Human Ethics and Consent to Participate declarations: not applicable

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-63.
2. Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)*. 2022;135(5):584-90.
3. Grady WM, Carethers JM. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology*. 2008;135(4):1079-99.
4. Vilar E, Gruber SB. Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol*. 2010;7(3):153-62.
5. Baretta M, Le DT. DNA mismatch repair in cancer. *Pharmacol Ther*. 2018;189:45-62.

6. André T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *N Engl J Med*. 2020;383(23):2207-18.
7. Hu H, Kang L, Zhang J, Wu Z, Wang H, Huang M, et al. Neoadjuvant PD-1 blockade with toripalimab, with or without celecoxib, in mismatch repair-deficient or microsatellite instability-high, locally advanced, colorectal cancer (PICC): a single-centre, parallel-group, non-comparative, randomised, phase 2 trial. *Lancet Gastroenterol Hepatol*. 2022;7(1):38-48.
8. Bari S, Matejic M, Kim RD, Xie H, Sahin IH, Powers BD, et al. Practice Patterns and Survival Outcomes of Immunotherapy for Metastatic Colorectal Cancer. *JAMA Netw Open*. 2025;8(3):e251186.
9. Pedroza Matute S, Iyavoo S. Implementation of NGS and SNP microarrays in routine forensic practice: opportunities and barriers. *BMC Genomics*. 2025;26(1):541.
10. Rüschoff J, Baretton G, Bläker H, Dietmaier W, Dietel M, Hartmann A, et al. MSI testing : What's new? What should be considered? *Pathologe*. 2021;42(Suppl 1):110-8.
11. Lu MY, Chen TY, Williamson DFK, Zhao M, Shady M, Lipkova J, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature*. 2021;594(7861):106-10.
12. Yamamoto Y, Tsuzuki T, Akatsuka J, Ueki M, Morikawa H, Numata Y, et al. Automated acquisition of explainable knowledge from unannotated histopathology images. *Nat Commun*. 2019;10(1):5642.
13. Komura D, Ishikawa S. Machine Learning Methods for Histopathological Image Analysis. *Comput Struct Biotechnol J*. 2018;16:34-42.
14. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13(1):152.
15. Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*. 2022;40(8):865-78.e6.
16. Stoffel EM, Mangu PB, Gruber SB, Hamilton SR, Kalady MF, Lau MW, et al. Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines. *J Clin Oncol*. 2015;33(2):209-17.
17. Gupta S, Provenzale D, Regenbogen SE, Hampel H, Slavin TP, Hall MJ, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. *J Natl Compr Canc Netw*. 2017;15(12):1465-75.
18. Shia J. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part I. The utility of immunohistochemistry. *J Mol Diagn*. 2008;10(4):293-300.
19. Lee CJ, Shin W, Song M, Shin SS, Park Y, Srikuhnath K, et al. Comparison of digital PCR platforms using the molecular marker. *Genomics Inform*. 2023;21(2):e24.

Figures

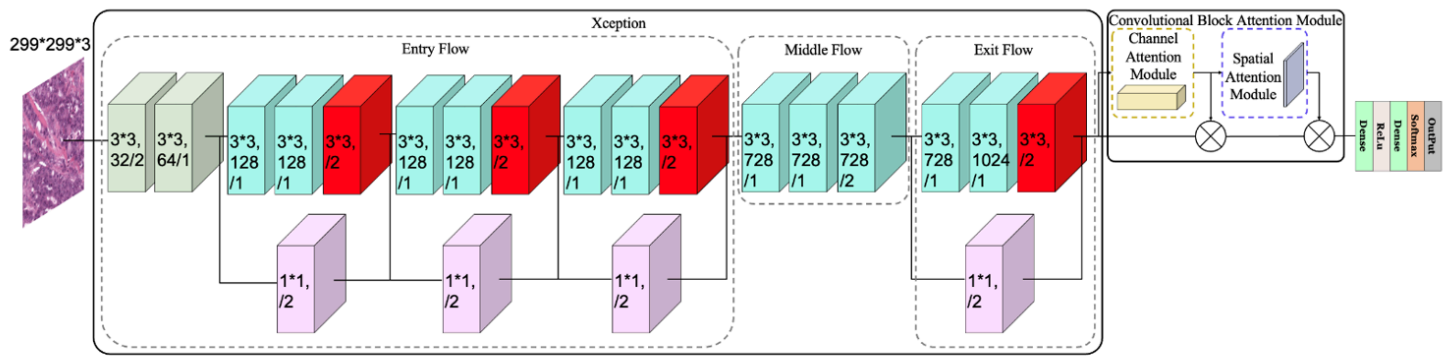


Figure 1

Schematic diagram of the overall architecture of the model proposed in this study.

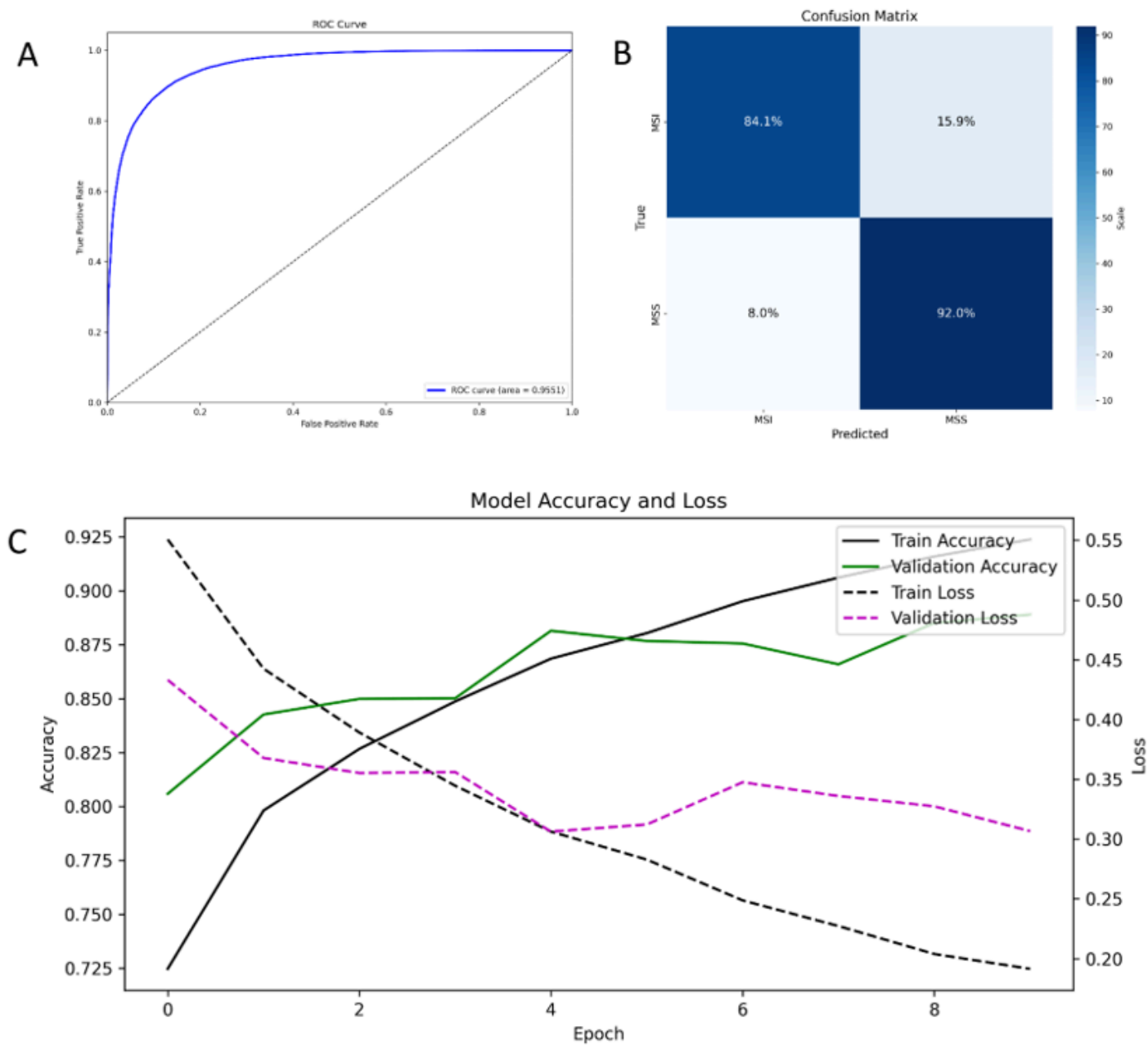


Figure 2

Performance of the model proposed in this study. (A) shows the ROC curve in the test set; (B) shows the corresponding confusion matrix in the test set; and (C) shows the training curve.

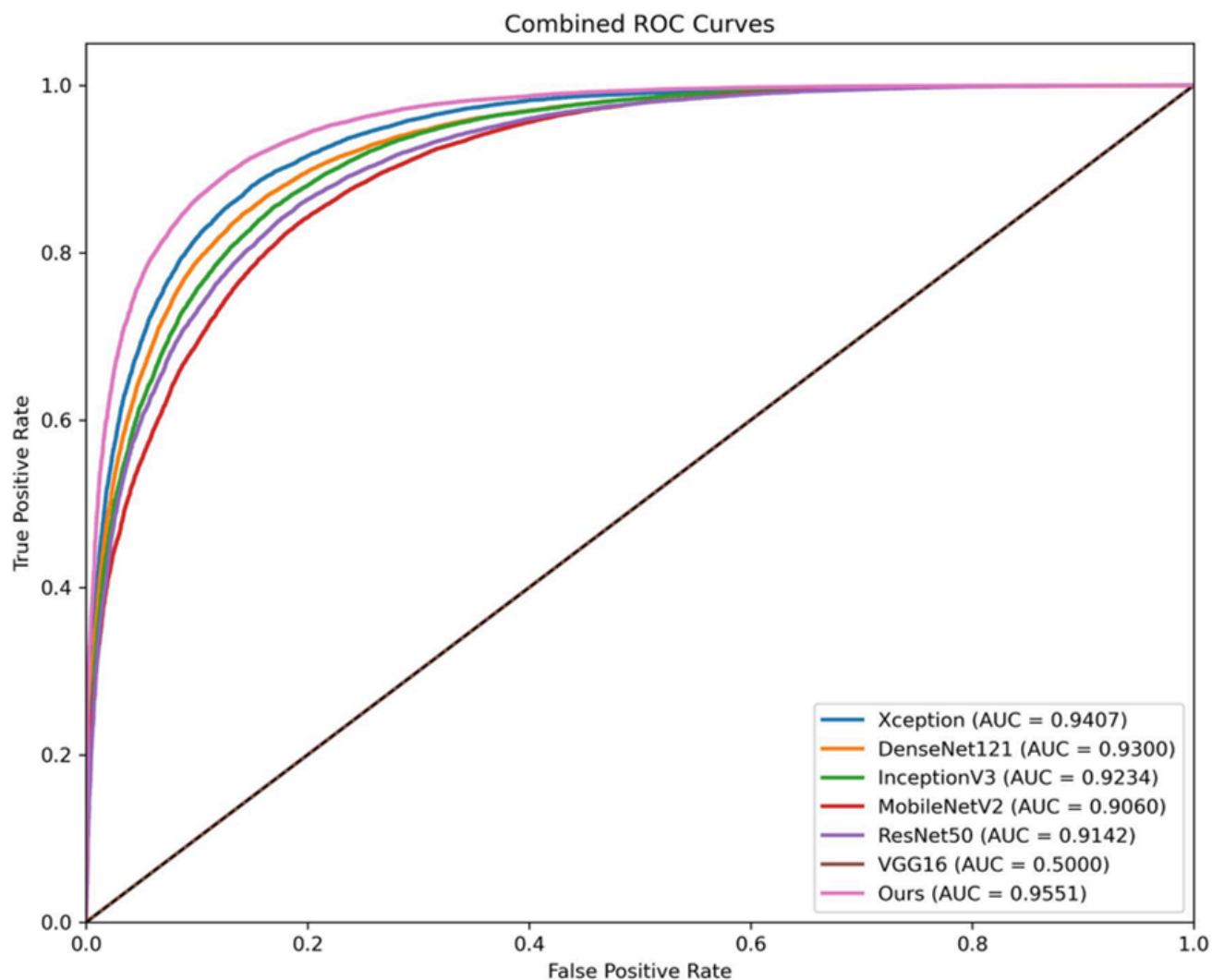


Figure 3

Comparison of ROC curves of various models in the test set.

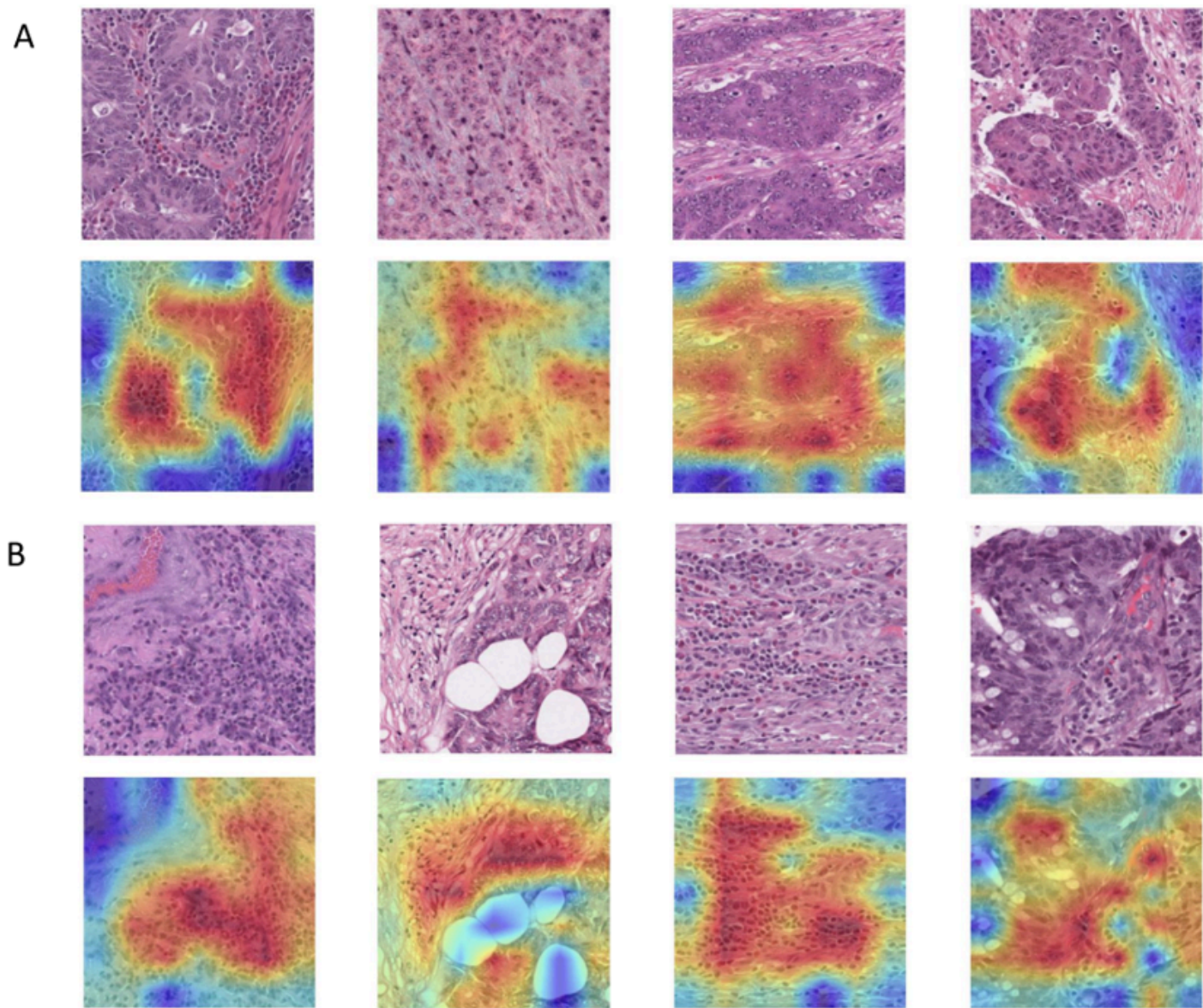


Figure 4

Comparison of Grad-CAM visualization and original image. (A) and (B) represent examples of MSI and MSS, respectively.